TAMPEREEN TEKNILLINEN YLIOPISTO
TAMPERE UNIVERSITY OF TECHNOLOGY

Jukka Huhtamäki

Ostinato Process Model for Visual Network Analytics

Experiments in Innovation Ecosystems

Julkaisu 1425 • Publication 1425

Tampere 2016

Tampereen teknillinen yliopisto. Julkaisu 1425
Tampere University of Technology. Publication 1425

Jukka Huhtamäki

Ostinato Process Model for Visual Network Analytics
Experiments in Innovation Ecosystems

Thesis for the degree of Doctor of Science in Technology to be presented with due permission for public examination and criticism in Kampusareena Building, Kampusklubi, at Tampere University of Technology, on the 18th of November 2016, at 12 noon.

Tampereen teknillinen yliopisto - Tampere University of Technology
Tampere 2016

ISBN 978-952-15-3833-9 (printed)
ISBN 978-952-15-3846-9 (PDF)
ISSN 1459-2045

Supervisor

Emeritus Professor Seppo Pohjolainen
Tampere University of Technology

Pre-examiners

Professor Paavo Ritala
Lappeenranta University of Technology

Professor Elliot Bendoly
The Ohio State University

Opponent

Associate Professor Sean P. Goggins
University of Missouri

Tiivistelmä

Innovaatiot ylittävät organisaatioiden rajat useammin kuin koskaan aiemmin ja innovaatiotoiminta siirtyy organisaatioiden ulkopuolelle ja väliin. Tähän uuteen kontekstiin viitataan innovaatioekosysteemin käsitteellä. Avoin innovaatio, yhteiskehittely, käyttäjälähtöisyys, API- ja alustatalous ja liiketoiminnan ekosysteemit ovat muutoksen keskeisiä ajureita. Innovaatioekosysteemit ylittävät organisaatioiden ja maantieteelliset rajat kokonaisvaltaisina avoimina dynaamisina järjestelminä. Niillä on sekä poliittinen ja taloudellinen että teknologinen ulottuvuus. Lahjakkailla yksilöillä on aivan keskeinen rooli ekosysteemisessä innovaatiotoiminnassa. Liiketoimintaekosysteemeistä kumpuava innovaatioekosysteemien teoria määrittelee uuden viitekehyksen innovaatiotoiminnan analyysille ja tutkimukselle ja siten innovaatioiden mittaamiselle.

Innovaation mittaaminen ja visualisointi on haastavaa. Innovaatioekosysteemeissä haasteet lisääntyvät entisestään innovaatiotoiminnan kompleksisuuden takia takia. Jopa olennaisten toimijoiden ja sidosryhmien kattava tunnistaminen on vaikeaa. Samalla innovaatioekosysteemin ekosysteemitason rakenteen analyysit ovat välttämättömiä kolmelle ryhmälle: innovaatioekosysteemien tutkijoille, politiikan ja innovaatiotoiminnan päätöksentekijöille sekä innovaatioekosysteemien toimijoille. Digitaalisen datan uusia lähteitä on saatavilla innovaatiotoimintaan aikaisempaa enemmän ja siten periaatteelliset mahdollisuudet ekosysteemitason analyysiin ovat parantuneet.

Tässä väitöskirjatyössä tavoitteenamme on edistää digitaalisen datan soveltamismahdollisuuksia innovaatioekosysteemien rakenteellisessa analyysissä ekosysteemitasolla. Kehitämme tässä toimintasuunnittelututkimuksen otteella toteutetussa väitöskirjassa menetelmän innovaatioekosysteemien rakenteellisten ominaisuuksien tutkimiseen ekosysteemitasolla käyttämällä visuaalista verkostoanalytiikkaa. Keskeinen lähtökohta tutkimustyöllemme on havainto siitä, että toimijoiden väliset kytkökset ovat avainasemassa innovaatioekosysteemeissä. Luontevan keinon innovaatioekosysteemien rakenteellisen analyysin toteuttamiseen tarjoaa verkostoanalyysi. Se antaa innovaatioekosysteemien tutkijoille ja toimijoille mahdollisuuden tehdä havaintoja innovaatioekosysteemien rakenteesta ja toimijoiden rakenteellisista rooleista. Tutkimme tässä työssä joukon erilaisia innovaatioekosysteemejä alustapohjaisesta kansalliseen ja kansainväliseen sekä kasvua tukevaan ohjelmatoimintaan. Tavoitteenamme oli tunnistaa tapoja innovaatioekosysteemien mallintamiseen ja analysointiin verkostoina.

Visuaalinen verkostoanalyysi tuo lisäarvoa innovaatioekosysteemien ekosysteemitason rakenteen kartoittamiseen ja tutkimiseen. Ehdottamassamme lähestymistavassa innovaatioekosysteemien toimijoita ja heidän vuorovaikutustaan edustavaa mikrodataa kerätään moninaisista digitaalisista lähteistä. Innovaatioekosysteemin toimijat esitetään verkoston solmuina, jotka kytketään toisiinsa vuorovaikutuksen ja muiden yhteyksien perusteella. Sijoitukset, yrityshankinnat ja erilaiset sopimukset sekä neuvonantajana, perustajana tai keskeisenä työntekijänä toimiminen ovat esimerkkejä yhteyksistä toimijoiden välillä. Verkostoanalyysin tunnusluvut mahdollistavat erilaisten toimija- ja ekosysteemitason rakenteellisten ominaisuuksien esittämisen lukuarvoina. Menetelmässä verkostot visualisoidaan kartoittavan analyysin ja tulosten raportoimisen tueksi vuorovaikutteisilla välineillä, jotka mahdollistavat joustavan liikkumisen sekä yksityiskohtia että kokonaisuuksia valottavien näkymien välillä.

Tässä väitöskirjatutkimuksessa edistämme innovaatioekosysteemien datalähtöisen visuaalisen verkostoanalyysin teoriaa ja käytäntöä useilla tavoilla. Väitöskirjatutkimuksen keskeinen tulos on ostinato-prosessimalli, joka mahdollistaa toistamiseen perustuvan, käyttäjäkeskeisen tavan toteuttaa automatisoituja datalähtöisen visuaalisen verkostoanalyysin prosesseja. Ostinato-mallissa innovaatioekosysteemin analyysiprosessi toteutetaan kahdessa vaiheessa, datan kerääminen ja jalostaminen sekä verkoston luominen ja analyysi. Datan kerääminen ja jalostaminen toteutetaan neljällä askeleella: entiteetti-indeksin luominen, webin ja ohjelmointirajapintojen ryömiminen, datan raapiminen ja datan koostaminen. Verkoston luominen ja analyysi muodostuvat seitsemästä askeleesta: entiteettien valinta, solmujen ja yhteyksien luominen, tunnuslukujen laskenta, solmujen ja yhteyksien suodattaminen, entiteetti-indeksin täsmentäminen, asettelun prosessointi ja visuaalisten ominaisuuksien määrittely. Vuorottelu tutkimuksen ja automatisoinnin välillä leikkaa läpi prosessin vaiheiden.

Ostinato-mallin ohella määrittelemme joukon ohjenuoria innovaatioekosysteemien verkostomallinnuksen ja visuaalisen verkostoanalyysityön tueksi. Annamme väitöskirjassa panoksemme myös innovaatioekosysteemien empiiriseen tietämykseen mallintamalla joukon abstraktiotasoltaan ja kompleksisuudeltaan erilaisia innovaatioekosysteemejä. Tutkimuksen kohteena olleiden innovaatioekosysteemien tutkijat, politiikan päätöksentekijät, orkestroijat ja muut toimijat ovat omaksuneet esitetyn lähestymistavan. Työssä esittämämme ohjenuorat yhdessä ostinato-mallin kanssa antavat innovaatioekosysteemien tutkijoille ja toimijoille mahdollisuuden ottaa merkittäviä askelia visuaalisen verkostoanalytiikan soveltamisessa johtamisen ja orkestroinnin välineenä. Lähestymistavan kattava hyödyntäminen organisaatiorajat ylittävien innovaatiotoimien tutkimisessa, tukemisessa ja orkestroinnissa edellyttää lisää tutkimusta ja tuotekehitystä.

Abstract

More often than ever before, innovation activities are crossing organizational boundaries and taking place in the spaces between formal, organizational structures. This new context for innovation activities is increasingly referred to as an innovation ecosystem. Open innovation, co-creation, user-driven innovation, API and platform economies, and business ecosystems are key drivers of the transformation. Innovation ecosystems are open, dynamic systems that cross geographical as well as organizational boundaries and include financial, technological, and political dimensions. Talented humans have a crucial driving role in ecosystemic innovation activities. Innovation ecosystems set a new framework for analyzing, investigating, and therefore measuring innovation.

Measuring and visualizing innovation is difficult, particularly within innovation ecosystems where activities take very complex forms and even identifying all relevant actors and stakeholders is challenging. At the same time, ecosystem-level analyses of innovation ecosystem structures are imperative for three groups: innovation ecosystem scholars, policy and decision makers, and innovation ecosystem actors. Moreover, new sources of digital data on innovation activities have become available, introducing new opportunities to investigate innovation ecosystems at the ecosystem level.

In this dissertation, we seek to develop new means to utilize digital data in analyzing innovation ecosystems at the ecosystem level. We take an action design research approach to develop the means to investigate the structural properties of innovation ecosystems at the ecosystem level by using visual network analytics. We start from the realization that interconnectedness is a key property of innovation ecosystems. Addressing innovation ecosystems as networks, that is, as collections of pairs of interconnected innovation ecosystem actors, allows scholars and practitioners to gain insight into innovation ecosystem structures and the structural roles of individual ecosystem actors. To determine how innovation ecosystems should be modeled and analyzed as networks, we investigate several innovation ecosystems representing regional, metropolitan, national, and international contexts as well as investigating the context of programmatic activities that support innovation and growth. Our main objective in the dissertation is to develop a process model for data-driven visual network analytics of innovation ecosystems.

Visual network analytics is a valuable method for investigating and mapping the innovation ecosystem structure. In the proposed approach, transactional microdata on innovation ecosystem actors and their interconnections is collected from various digital sources. Innovation ecosystem actors are represented as network nodes that are connected through transactions, including investments and acquisitions and advisory, founder, and contributor affiliations. Network metrics are used to quantify actors' structural positions. Interactive visual analytics tools are used to support the visual exploration of the innovation ecosystem under investigation by using both top-down and bottom-up strategies.

This work makes several contributions to the art and science of data-driven visual network analytics of innovation ecosystems. Most importantly, the dissertation proposes the ostinato model, an iterative, user-centric, process-automated model for data-driven visual network analytics. The ostinato model simultaneously supports the automation of the process and enables interactive and transparent exploration. The model has two phases: data collection and refinement, and network creation and analysis. The data collection and refinement phase is further divided into entity index creation, Web/API crawling, scraping, and data aggregation. The network construction and analysis phase is composed of filtering in entities, node and edge creation, metrics calculation, node and edge filtering, entity index refinement, layout processing, and visual properties configuration. The cycle of exploration and automation characterizes the model and is embedded in each phase.

In addition to the ostinato model, we contribute a set of design guidelines for modeling and visualizing innovation ecosystems as networks. Finally, we contribute to the empirical body of knowledge on innovation ecosystems through a series of investigations of innovation ecosystems of different levels of abstraction and complexity. Innovation ecosystem scholars, policy makers, orchestrators, and other stakeholders in the innovation ecosystem under investigation in this dissertation have subscribed to the approach presented herein. The design guidelines, together with the ostinato model, allow innovation ecosystem investigators and actors an opportunity to significantly advance in utilizing visual network analytics in managing and orchestrating innovation ecosystems. Further research and development of supporting processes and tools are needed to take full advantage of the presented approach in analyzing, investigating, facilitating, and orchestrating inter-organizational innovation activities.

Preface

This study was carried out at the Intelligent Information Systems Laboratory (IISLab) at Tampere University of Technology (TUT). IISLab (2012 – 2016) was part of the Department of Mathematics and was previously known as the Hypermedia Laboratory.

There are many beginnings to the story of this dissertation. Let me start with the most important one, without which this dissertation would never have come to reality. I attended a conference in San Diego in 2009 to present a paper. To add a twist to the trip, I decided to visit Stanford University. I knew of Stanford's reputation, of course; however, the whole notion of Silicon Valley was not at all clear to me. Looking at the map, I saw that making the trip from San Diego to San Francisco was the perfect excuse to spend a weekend driving Highway 1.

My supervisor agreed to sponsor a few days at Stanford if I could arrange to meet some local researchers. Many request were kindly and politely turned down. Fortunately, Martha Russell agreed to meet me. Later I learned that she is in the business of having time to meet people, trusting that they have an important contribution to make, independent of their rank or status.

I was able to make my way to Stanford and Wallenberg Hall. Martha and I had an inspiring discussion. As it happened, the two pioneers of modern personal computing, Douglas Engelbart and Ted Nelson, were visiting mediaX the day of my visit and I got to meet them both. I began to understand how Silicon Valley operates and wanted to learn more. To make a long story short, I was sold!

Now, seven years later, following Martha's lead, the Innovation Ecosystems Network (IEN) has made a significant contribution to the art and science of innovation ecosystem investigation. The IEN team includes Kaisa Still, VTT Technical Research Centre of Finland; Neil Rubens, University of Electro-Communications, Tokyo; Rahul Basole, Georgia Tech; and Camilla Yu, VXPLO Innovation Lab. In addition, Neal Burns has made important contributions to IEN. The IEN team members are my co-authors in this dissertation and my dear friends. Thank you, team.

My contribution as a member of the IEN team was made possible by my working history at the Hypermedia Laboratory at Tampere University of Technology. Hypermedia Laboratory was an amazing platform for learning philosophical, theoretical, and practical skills and developing new methods and concrete tools in a truly multidisciplinary context. Ossi Nykänen has shown through his exemplary leadership how to conduct science and run teams in a respectful yet determined way. Anne-Maritta Tervakari, Kirsi Kuosa, Pekka Ranta, and Sami Hautakangas mentored me on the nuances of different disciplines. Jaakko Salonen and Juha Nurmi taught me to use computers and to develop code to solve problems. Through my years at the Hypermedia Laboratory, I had the pleasure to collaborate with tens of bright researchers and research assistants. Warm thanks to you all.

Professor Seppo Pohjolainen started the Hypermedia Laboratory and supported its operations through the years. Seppo also served as the supervisor of this dissertation. I am deeply grateful to Seppo for his trust and support in applying for funding, making new contacts, and traveling the world.

Professors Paavo Ritala and Elliot Bendoly served as the pre-examiners of this dissertation. Professor Ritala's feedback allowed me to take the precision and scientific quality of this dissertation to the next level. Professor Bendoly shared his knowledge on visual analytics and contributed to making the foundation of the dissertation stronger. Thank you, professors.

I am grateful for the network of academic professionals with whom I had the privilege to consult when needed. Samuli Kortelainen encouraged me to make the dissertation a reality, Kati Järvi supported me in identifying its core contribution, Professor Samuli Pekkola gave valuable comments on the methodology, Salla-Maaria Laaksonen pointed out a way to lay the philosophical foundations of the dissertation, Lysanne Lessard shared her expertise on using critical realism in information system research, Jari Jussila shared his insights on how to get the process over the finish line, and Karan Menon always had time for discussion.

The majority of research for this dissertation was conducted during the course of three projects: Social media supported innovation indicators (SINDI, 2010−2012), Relational capital for growth companies (REINO, 2012−2015), and Systemic paths of ecosystem evolution dynamics (SPEED, 2014−2016). All three projects were part of the Innovation research program at the Finnish Funding Agency for Innovation Tekes. My gratitude goes to Christopher Palmberg, Soile Ollila, Petri Räsänen, Antti Eskola, Marko Turpeinen, Ville Kairamo, and Ville Korpiluoto who provided continued support to our exploratory research venture.

I wrote this dissertation using Overleaf.com, a superb service for writing LaTeX. I used Zotero and Mendeley for managing the references. I owe warm thanks to the wonderful people serving customers at Kaffila, Wayne's Coffee in Tampere, and Philz Coffee in San Francisco for letting me sip my coffee for hours and hours, hands on the keyboard, headphones on. I am grateful to Terttu Etelämäki for her professional touch in finalizing the abstract in Finnish.

Exploration is where everything good starts, innovation included. My first explorations took place in Karjanlahti, a small village in Kauhava, Finland. Living on a small farm presented the perfect opportunity to learn the most important things. How to fix a mopeds and cars, how to drive a tractor sideways, and how to nurture fields and animals. My mother, Tuula, a librarian and informatics academic, maintained a constant flow of incoming knowledge. My father, Seppo, took us to everyday task at the farm and placed a lot of trust in the experimenting brothers, me and Janne, the trusted co-explorer. My sister Mari joined the adventures later and has always been there for me ever since. I am forever grateful to my family and friends for their support and for understanding that being a researcher means that I am not able to spend as much time with my loved ones as I want to.

On the day before leaving for the trip to San Diego in 2009, I was talking to the most beautiful girl in the world, that night dressed as the Finnish superhero Kari Grandi. We agreed that should she win the lottery that day, she would join me to cruise Highway 1.

Later that year, we bought an old wooden house and went to tour the world. Outi, your never-ending support has kept me sane throughout this rough process. I love you.

I would be happy to hear your feedback on the ostinato model and other aspects of this work that you find either interesting or in need of further development. You can reach me on Twitter: @jnkka.

Tampere, September 2016

Jukka Huhtamäki

This dissertation is dedicated to
all the brave researchers and research assistants
who served at the Hypermedia Laboratory and
the Intelligent Information Systems Laboratory
1996 – 2016

Contents

List of publications

This dissertation is a compilation of seven (7) publications, of which two (2) are journal articles, four (4) appear in conference proceedings and one (1) is published as a chapter of a peer-reviewed scientific book.

In this section, I will describe my role in each of the publications and add further information about them. Finnish Publication Forum[1] ratings are used to indicate the quality of the publication channels. Level 3 is the highest.

Publication I Jukka Huhtamäki, Ville Luotonen, Ville Kairamo, Kaisa Still, Martha G. Russell (2013) "Process for Measuring and Visualizing an Open Innovation Platform: Case Demola," *Academic MindTrek 2013, Tampere, Finland.* Publication Forum level 1.

Publication II Rahul C. Basole, Martha G. Russel, Jukka Huhtamäki, Neil Rubens (2012) "Understanding Mobile Ecosystem Dynamics: A Data-Driven Approach," *International Conference on Mobile Business (ICMB).* Publication Forum level 1.

Publication III Jukka Huhtamäki, Kaisa Still, Minna Isomursu, Martha G. Russell and Neil Rubens (2012) "Networks of Growth: The Case of Young Innovative Companies in Finland," *The 7th European Conference on Entrepreneurship and Innovation (ECIE 2012), Santarém, Portugal.* Publication Forum level 0.

Publication IV Jukka Huhtamäki, Martha G. Russell, Kaisa Still, Neil Rubens (2013) "A Network-Centric Snapshot of Value Co-Creation in Finnish Innovation Financing," *Value Co-Creation: Best of TIM Review.* Publication Forum level 0; TIM Review was promoted to Publication Forum level 1 in 2015.

Publication V Kaisa Still, Jukka Huhtamäki, Martha G. Russell, Rahul C. Basole, Jaakko Salonen, Neil Rubens (2013) "Networks of innovation relationships: multiscopic views on Finland," *International Society for Professional Innovation Management (ISPIM) 2013, Helsinki.* Publication Forum level 0.

Publication VI Kaisa Still, Jukka Huhtamäki, Martha G. Russell, Neil Rubens (2014) "Insights for orchestrating innovation ecosystems: the case of EIT ICT Labs and data-driven network visualisations," *International Journal of Technology Management.* Publication Forum level 1.

Publication VII Jukka Huhtamäki, Martha G. Russell, Neil Rubens, Kaisa Still (2015) "Ostinato: The exploration-automation cycle of user-centric, process-automated

[1] Finnish Publication Forum, `http://www.tsv.fi/julkaisufoorumi/english.php`

data-driven visual network analytics," in Sorin Adam Matei, Elisa Bertino, Russell, Martha G. (Eds.) *Transparency in Social Media: Tools, Methods and Algorithms for Mediating Online Interactions*, Springer. Publication Forum level 2.

Author's role and quality of publication channels

Let me start the description of my role in the research presented in this dissertation with an editorial comment. Although I have chosen to write this dissertation in first person plural form, that is, we rather I, I have selected the viewpoint of the dissertation to represent work in which I have played a leading role. The individual investigations that I have included in the dissertation are very much a group effort. Specifically, I have led the development of the design guidelines for investigating innovation ecosystems as networks as well as the ostinato model. I am, however, not willing to go as far as the guideline for selecting between I and we suggests: [2] "However, I am writing a thesis which means I am the only author and I even have to testify in writing that the work is my own and I did not receive any help other than from the indicated sources." Therefore, I will be referring to "we" throughout this dissertation.

The series of investigations that form the first part of the results of this dissertation are conducted by the Innovation Ecosystems Network (IEN) team. The majority of the work took place under the Tekes-sponsored innovation research projects SINDI (2010-2012), REINO (2012-2015) and SPEED (2014-2016). The author of this dissertation is one of the co-founders of IEN. Martha G. Russell, mediaX at Stanford University, leads the IEN team, which includes Kaisa Still, VTT Technical Research Centre of Finland Ltd; Neil Rubens, University of Electro-Communications, Tokyo; Rahul C. Basole, Georgia Tech; and Camilla Yu, VXPLO Innovation Lab.

The author of the dissertation led the design, development and operation of the data-driven visual analytics process in each of the investigations. Moreover, the author wrote the majority of the manuscript of **Publication I**, **Publication III**, **Publication IV** and **Publication VII**. In others, the author made significant contributions to the research design, manuscripts, analysis, discussion, and research project management.

Publication I was presented by the author of this dissertation at Academic MindTrek in Tampere, Finland. Academic MindTrek (MindTrek Conference in Publication Forum) is placed in Publication Forum level 1.

Publication II received the Best Paper Award at the International Conference on Mobile Business (ICMB). Rahul Basole gave the presentation. An extended version of the paper was later published in *ACM Transactions on Management Information Systems* (ACM TMIS). ICMB is in Publication Forum level 1.

Publication III was presented by the author of this dissertation at ECIE 2012 in Santarém, Portugal. ECIE is in Publication Forum level 0.

Publication IV was originally presented by the author of this dissertation at the 10th Annual EBRF conference in Nokia, Finland. The article was first invited to be published in *Technology Innovation Management Review* and was later selected to be included in the *Best of TIM Review* of co-creation studies. In 2015, TIM Review was promoted to Publication Forum level 1.

[2] `http://english.stackexchange.com/revisions/24629/1`

Publication V was co-presented by Kaisa Still and the author of this dissertation at ISPIM in Helsinki. Although ISPIM is in Publication Forum level 0, it is considered a key yearly meeting in innovation research domain.

Publication VI was invited to be published in a special issue of the *International Journal of Technology Management* (IJTM) on Business and Network Models for Innovation. IJTM is placed in Publication Forum level 1.

Publication VII was selected to be included in the second volume of *Roles, Trust, and Reputation in Social Media Knowledge Markets*, a book published by Springer. As a book publisher, Springer is placed in Publication Forum level 2.

1 Introduction

Innovation – what is it and why it is important? Innovation refers to the ability to create and capture economic value from invention (Hagel & Seely Brown, 2005). Innovation is a key driver of any sustainable business and, indeed, society. Innovation has become an imperative for business, academic and governmental organizations in response to environmental and technology-driven changes (Drucker, 2015; Gupta, Tesluk, & Taylor, 2007; Schumpeter, 1942; Van de Ven, 1986). Carlson and Wilmot (2006), among others, underline the imperative role of innovation: "Nothing is more important to business success than innovation." Over the past century, conceptual approaches for understanding and accelerating innovation have evolved.

What are innovation ecosystems and why are they relevant in 2016 and beyond? More often than ever before, innovation is taking place outside the boundaries of and in-between individual organizations. Moreover, customers now have an active role in innovation, that is, in value co-creation. Through open innovation and value co-creation, several steps have been taken from the manufacturing-driven, technology-push view (Schumpeter, 1934, 1942, 1950) to that of the ecosystemic view in both to business and innovation (Järvi, 2013; Moore, 1993; Russell, Still, Huhtamäki, Yu, & Rubens, 2011). Autio and Thomas (2013) define the innovation ecosystem as "a network of interconnected organizations, organized around a focal firm or a platform, and incorporating both production and use side participants, and focusing on the development of new value through innovation." Russell et al. (2011) take an even broader view to define innovation ecosystem as an "inter-organizational, political, economic, environmental and technological systems of innovation through which a milieu conducive to business growth is catalyzed, sustained and supported." At the ecosystem level, individual relationships take the shape of a network "through which information and talent flow through systems of sustained value co-creation" (Russell, Huhtamäki, Still, Rubens, & Basole, 2015).

The key influence in the adoption of the ecosystem concept in the context of innovation is the concept of business ecosystem that Moore (1993) introduced. Ecosystem is used in this context as a metaphor (e.g., Hwang & Horowitt, 2012), a strategy artifact (Moore, 1993), and to refer to ecosystem-level analyses (cf., Pentland, 2015). Innovation ecosystems are orchestrated rather than controlled or managed (Ritala, Agouridas, Assimakopoulos, & Gies, 2013; Ritala, Armila, & Blomqvist, 2009; Russell et al., 2011). When collective gains are sought at the ecosystem level, change agents seek to facilitate the emergence of networks, orchestrate existing networks, and manage their growth (Russell et al., 2011). To manage innovation ecosystems and to facilitate their emergence, the process of network orchestration is encouraged (Russell, Huhtamäki, et al., 2015; Russell et al., 2011).

In this dissertation, we follow the concept of innovation ecosystem as defined by Russell et al. (2011) for two key reasons. First, the research for this dissertation was conducted in close collaboration with Martha Russell and the Innovation Ecosystems Network team, and therefore their definition guides the research design used in the individual experiments described in Chapter 3. Second, Russell et al. (2011) included financial actors and highlighted the role of individual people and the importance of culture in their definition, and they considered them key enablers in innovation ecosystems that are conducive to the genesis of startup and growth companies. We do however acknowledge that, importantly, the role of customers, governmental institutions is not explicitly mentioned in their definition.

Why is visual network analysis valuable in exploring innovation ecosystems? Measuring and visualizing innovation is difficult. This is particularly true for innovation ecosystems where innovation activities take very complex forms, and even the identification of relevant actors and stakeholders is challenging. At the same time, ecosystem-level analyses of innovation ecosystems are imperative for three groups: innovation ecosystem scholars, policy and decision makers, and innovation ecosystem actors. Moreover, new sources of digital data on innovation activities have become available. In this dissertation, we seek to bridge the gap between the opportunities provided by the digital data available and the desire to investigate the innovation ecosystems at the ecosystem level. Specifically, our focus is on investigating the network structure in-between various ecosystem actors using visual network analytics.

> *Use a picture. It's worth a thousand words.* -Arthur Brisbaine (Bendoly, 2016)

> *...few people will appreciate the music if I just show them the notes. Most of us need to listen to the music to understand how beautiful it is. But often that's how we present statistics; we just show the notes, we don't play the music.* - Hans Rosling

From the beginning of social network analysis and its precursor, sociometry, visualization has been a key part of the analysis process (Moreno, 1953):

> *We have first to visualize [...] A process of charting has been devised by the sociometrists, the sociogram, which is more than merely a method of presentation. It is first of all a method of exploration. It makes possible the exploration of sociometric facts. The proper placement of every individual and of all interrelations of individuals can be shown on a sociogram. It is at present the only available scheme which makes structural analysis of a community possible.*

The exploratory visual analysis of network structure allows holistic investigations as well as sharing the findings to others (Freeman, 2000). The challenges and opportunities of visualization in general and visual network analytics in particular have received little attention in innovation ecosystem research. Basole (2014) states that while research on challenges and opportunities of visualization in corporate settings exists (Lam, Bertini, Isenberg, Plaisant, & Carpendale, 2012; Sedlmair, Isenberg, Baur, & Butz, 2011), "researchers haven't focused on visual business ecosystem intelligence tools." We maintain that the observation also holds in the context of innovation ecosystems.[1] Overall, we see that empirical research on innovation ecosystems continues to be relatively scarce and particularly so at the ecosystem-level (rather than actor or relationship level) (Järvi & Kortelainen, 2016).

Applying visual analytics in management uses carefully designed processes supporting the actors joining decision-making; moreover, the very design and development of these processes require a step-wise iterative and incremental approach (Bendoly, 2016). The use of visual network analytics in supporting ecosystem orchestration is essentially the application of enacted sensemaking (cf. Bendoly, 2016, 2017; Weick, Sutcliffe, & Obstfeld, 2005). The practices related to innovation ecosystem orchestration and the related enacted sensemaking are outside the scope of this dissertation. Sensemaking and orchestration do however serve as sources of requirements for the visual network analytics of innovation ecosystems. We will return to orchestration in discussing the potential implications of this dissertation as part of Chapter 8. Moreover, as we show in Chapter 5, the ostinato model draws from existing sensemaking models.

Why should these explorations and investigations be conducted using a data-driven, computational approach? New ways to measure innovation are needed (Still, Huhtamäki, Russell, & Rubens, 2012). Social media, socially constructed datasets, and other sources of digital data have created a plethora of new opportunities for measuring activities particularly toward the upstream of innovation. However, using social media as a source for measuring innovation is far from straightforward (Salonen, Huhtamäki, & Nykänen, 2013). The same goes for socially constructed data. In this dissertation, our objective is to develop a new research method for data-driven investigations of the structure of innovation ecosystems. The method falls under the field of computational social science (Lazer et al., 2009).

With the advent of theories and paradigms, such as the network-driven nature of the world (Barabási & Bonabeau, 2003; Watts, 1999) and, more recently, computational social science (Lazer et al., 2009) and social physics (Pentland, 2015), the anticipation of the possibility to develop theories for human behavior that are almost as specific as in physics or natural science is growing in a number of domains. The machinery to conduct the investigations, however, is largely in its infancy during the time of writing this dissertation. In the investigations we conducted for the dissertation, we focused on finding the kinds of datasets that would give us a ecosystem-level view into the network structure of innovation ecosystems (cf., Pan, Altshuler, & Pentland, 2012).

Based on these grounds, we began our venture in 2010 to investigate how social media and other publicly available sources of digital data could be used to map, analyze, and visualize the network structure of innovation ecosystems. In this dissertation, we take a data-driven approach to visual network analytics. Here, we refer to the data-driven meaning that the analysis process relies on data and is automated and conducted in a computational manner. Additional data can augment the dataset selected for analysis through an automated software process. Established analytical procedures can be automated, and new conditions for analysis-based insights can be introduced and refined incrementally with continuous computational iterations. We will revisit the development of these procedures in Chapter 5.

[1] Innovation Ecosystems Network is among the first to contribute to the field of visual innovation ecosystem analytics through Innovation Ecosystems Summit (2011), Visualizing Innovation Ecosystems workshop at Mindtrek 2011, and Analytics and Decision Support for Ecosystems Minitrack at HICSS 2016.

Drawing from our experience in running multiple experiments in the context of exploratory and descriptive innovation ecosystem investigations, we take a action design research (ADR) (Sein, Henfridsson, Purao, Rossi, & Lindgren, 2011) approach to describe a process model for use in data-driven visual network analytics.

1.1 Objectives of the dissertation

In this dissertation, we target the research gap in investigations of the structure of innovative ecosystems. First, as Järvi and Kortelainen (2016) point out, empirical studies on innovation ecosystem structure at the ecosystem level are scarce. Moreover, visual analytics tools and methods for studying innovation ecosystem structure are largely missing (cf., Basole, 2009). We answer Freeman's (2000) call to develop an integrated method to retrieve social network data, create networks, compute network metrics, and visualize networks. Finally and most importantly, we determine the requirements and limitations stemming from the development and use of visual network analytics tools for innovation ecosystem investigations that take place in their natural habitat, that is, in interdisciplinary teams that, according to Bendoly (2016), seek "not just simultaneous parallel use [of data and visualizations] but truly joint utilization and team-wise sensemaking for effective decisions." Although the enacted sensemaking process, an imperative step in utilizing visual analytics in management (Bendoly, 2016), is outside the scope of this dissertation, we take the iterative and incremental nature of the process into account when we develop the process model for use in visual network analytics.

The overall objective of this dissertation is to develop the means to investigate innovation ecosystems as networks through visual analytics. We claim that representing the structure of innovation ecosystems as networks is particularly intuitive and expressive in supporting their exploration and in allowing for a ecosystem-level view. Specifically, the individual objectives of this dissertation are summarized as follows:

Objective I To contribute to the empirical body of knowledge on innovation ecosystems through modeling and visualizing the network structure of a set of innovation ecosystems representing different grades of abstraction and complexity.

Objective II To identify patterns for modeling and analyzing the structure of innovation ecosystems as networks.

Objective III Finally and most importantly, to define a process model for the data-driven visual network analytics of innovation ecosystems.

In the next section, we will discuss the philosophical foundations of the dissertation and the research methodology applied to meet the objectives of the dissertation.

1.2 Research methods

This dissertation is in the domain of information systems (IS) research. IS researchers seek to develop means to create value in information for people, teams, and organizations (Nunamaker & Briggs, 2011). Due to the wide scope of knowledge from which IS draws, Nunamaker and Briggs (2011) urge IS researcher to "embrace multi-investigator, multidisciplinary, even multiuniversity research teams."

The IS field is an example of the sciences of the artificial (Simon, 1969) in which the objectives of investigations are man-made artifacts rather than phenomena in nature. Two main approaches to IS research exist. The first and more traditional is the behavioral approach where the usability, effectiveness, and overall quality of already existing IT artifacts is evaluated and measured through rigorous experiments. The second approach is to use the development process of an IT artifact to conduct research and create new knowledge. We adopt the latter approach.

To solve practical problems and to create new scientific knowledge in the domain of the sciences of the artificial, Hevner, March, Park, and Ram (2004) introduced design science research (DSR) as a research method that allows the creation of new knowledge through the design and implementation of new artifacts. Over the last 10 years, Vaishnavi and Kuechler (2007), Peffers, Tuunanen, Rothenberger, and Chatterjee (2007) and others have subscribed to DSR as a method of utility in scientific discovery. We claim that DSR has a perfect fit to meet the objectives of this dissertation, that is, to develop a) design guidelines for representing analyzing innovation ecosystems as networks (**Objective II**) and b) a process model for supporting data-driven investigations of innovation ecosystems (**Objective III**).

While DSR is often conducted without explicit "discussion about underlying philosophical assumptions in the IS design science research literature" (Carlsson, 2010), and a large part of engineering science research in general takes place without explicit reference to its philosophical underpinnings (Naukkarinen, 2015), we will next briefly state the ontological and epistemological stance taken in this dissertation. In short, we subscribe to Carlson's (2010) proposition that critical realism provides a workable philosophical basis for DSR. Bygstad and Munkvold (2011) note that critical realism is an "alternative to positivist and interpretive IS research" and provide a straightforward definition: "Critical realism combines a realist ontology with an interpretive epistemology." They go on and clarify that "although a real world exists, our knowledge of it is socially constructed and fallible" (Bygstad & Munkvold, 2011).

The critical realist wordview provides a firm foundation for innovation ecosystem investigations. Innovation ecosystems are open, complex, and adaptive systems (L. D. W. Thomas & Autio, 2012) where individual humans have a driving role. These properties yield analytic requirements that are very difficult to answer. The critical realist philosophy allows us to establish an epistemological and ontological platform for the investigations such that the complexity of the phenomenon under investigation can be considered.

We agree with Dobson (2001) in acknowledging the importance of taking a philosophical stance in establishing a common platform for investigative work. The combination of realist ontology with interpretive epistemology (Bygstad & Munkvold, 2011) means that the investigators assume that generalizable structures and mechanisms exist in a social phenomenon. In order to identify these mechanisms and structures, however, investigators need to move beyond superficial statistical measurements and case-specific qualitative observations to apply several complementing methods, both qualitative and quantitative, in the investigations.

More pragmatically, we point to Bygstad and Munkvold (2011) who introduce critical realism as an approach to data analytics and claim that the analytics process should serve the objective of identifying the structures and mechanisms that exist within a phenomenon and surface as observable events. They refer to Sayer (2010) for a layered ontology of critical realism and a related research strategy that is composed of three layers: 1) events that are caused by 2) mechanisms driven by the underlying 3) structure of a phenomenon under investigation. We further agree with Dobson (2001) and Archer (1995) that social structure should be considered dynamic, and therefore studying its evolution over time is important, that social structure is a key driver of social activity, and that social activity is a key driver of the evolution of social structure.

In the individual investigations included in this dissertation, we have applied different research approaches and methods—from the case study (Yin, 2003) to action research and straightforward computational network analysis (Lazer et al., 2009). In order to meet the objectives of the dissertation, however, the investigations serve first and foremost as experiments that provide a means to develop the approach used to model the network structure of innovation ecosystems in a data-driven manner. The experiments, their context, and the the results of the network analytics are described in detail in Chapter 3.

Furthermore, in terms of the different dimensions of the philosophy of science, the research in this dissertation is qualitative rather than quantitative (however, the investigations using the produced IT artifacts are computational), inductive rather than deductive, and subjective rather than objective (cf. Olsson, 2012).

Using DSR allows us to conduct research that is both 1) credible in terms of scientific theory and 2) has practical relevance to different innovation ecosystem stakeholders. DSR is a research method that allows “learning and investigation through artifact construction” (Vaishnavi & Kuechler, 2007). In contrast to the natural and social sciences where the objective is to understand reality, “design science attempts to create things that serve human purposes” (Simon, 1969). The rationale for DSR stems from the importance of the practical utility of research (Peffers et al., 2007). Design science research aims to support bridging IS research and its practical application by producing results that have real-life relevance. DSR “creates and evaluates IT artifacts intended to solve identified organizational problems” (Hevner et al., 2004).

Sein et al. (2011) argue that traditional DSR prioritizes technological rigor over organizational relevance and therefore might “fail to recognize that the artifact emerges from interaction with the organizational context even when its initial design is guided by the researchers’ intent.” To negotiate the issue, Sein et al. (2011) propose action design research (ADR) in which the research process is conducted in four stages: 1) problem formulation, 2) building, intervention, and evaluation, 3) reflection and learning, and 4) formalization of learning. Figure 1.1 presents the four stages of ADR.

Three types of designs are related to any IS initiative (Aken, 2004; Carlsson, 2010): 1) object design, 2) realization design, and 3) process design, “the professional’s own plan for the problem-solving cycle and includes the methods and techniques to be used in object and realization design” (Carlsson, 2010). Carlsson (2010) continues, “IS design science research should produce knowledge that can be used by the professionals in the three types of designs, including novel IT artifacts, methodologies, methods and techniques, and socio-technical implementation knowledge.”

ADR stresses the importance of iteration and interaction with the future users of the IT artifact. The building-intervention-evaluation (BIE) cycle enables the interaction, that is, an artifact is built and used to implement an intervention in an organization and to evaluate the artifact over several BIE cycles. To develop the guidelines for investigating innovation ecosystems as networks as well as the ostinato model, we repeat the BIE cycle over a series of investigations serving as ADR experiments described in detail in Chapter 3. The BIE cycle allows us to evaluate the ostinato model for the validity and added credibility of the presented results. Through guided emergence, we are able to

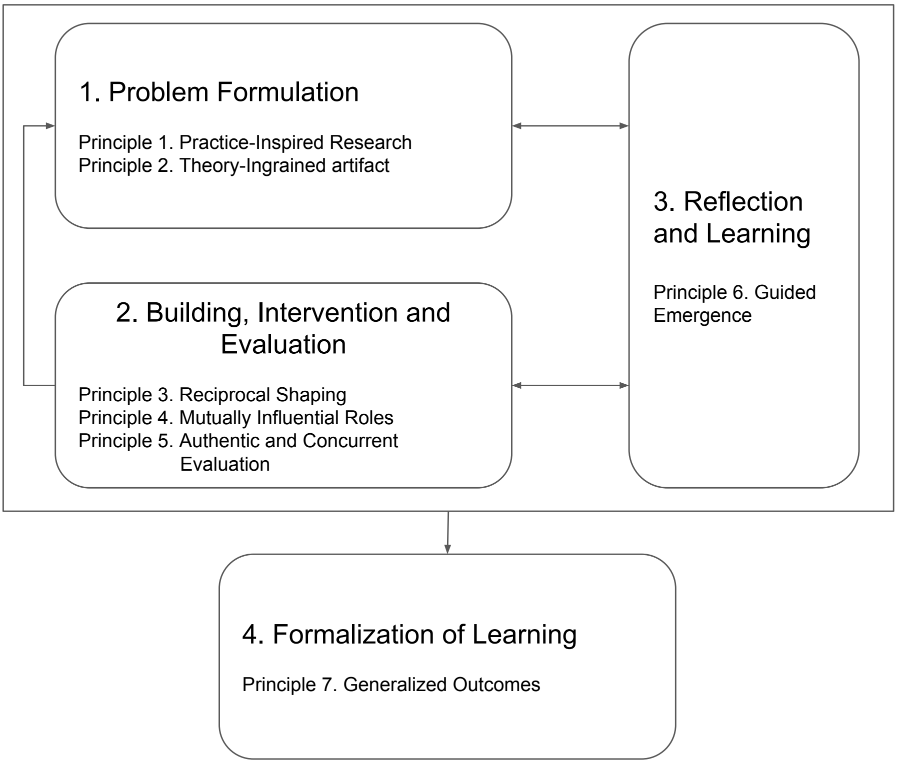


Figure 1.1: Action design research stages and principles (reproduced following Sein et al., 2011)

contribute to innovation ecosystem research with deep knowledge of the problem and the proposed solution.

Sein et al. (2011) note that while DSR separates artifact building and evaluation, ADR emphasizes the iterative nature of artifact development. ADR is, therefore, well in line with the principles of contemporary software development approaches. Readers with experience in software development will indeed notice a straightforward connection between the BIE cycle and agile software development (see e.g., Schwaber & Beedle, 2001). More recently, startups and other value-seeking communities within the software development industry have turned their attention to lean startup, another iteration-based method (Ries, 2011). Apart from the intention to publish the results, these three design and development processes move forward in an iterative and incremental fashion, and are guided by feedback collected from the users and other stakeholders of the developed software or other artifact, which in this dissertation is the ostinato process model.

DSR is sometimes criticized for its lack of scientific rigor and its bias in practice over theory. ADR responses to the critique with the principle of authentic and concurrent evaluation (Sein et al., 2011) that we apply through this dissertation. The utility and viability of artifacts created with the DSR process can be evaluated through proof-of-concept, proof-of-value, or proof-of-use (Nunamaker & Briggs, 2011), each of which adds to the requirements for the state of the artifact as well as the potential level of insight that the artifact has the ability to create. We will revisit the value and use of the artifacts developed in this dissertation in the final discussion in Chapter 7.

1.3 Scope and delimitations

Due to the holistic nature of this dissertation, we will next explicitly describe the scope of the research. Moreover, we will describe the delimitations we have set to guide the investigations. We will discuss separately the scope and delimitations of each of the three objectives.[2]

Objective I is to contribute to the empirical body of knowledge of innovation ecosystem network structure. We review the extant literature in detail in Section 2.2. Throughout the dissertation, our empirical work concentrates specifically on the ecosystem-level network structure of the selected innovation ecosystems. Importantly, we want to point out the difference between ecosystem-level analysis and the analysis of ecosystem-level network structure. The former is potentially a far more general objective, because it includes the study of mechanisms through which the ecosystem operates and evolves. The latter, our focus, is limited to investigating the network structures that emerge from individual connections between ecosystem actors including companies, individual people (C-level, advisors, and board members), and investors. These connections include both explicit connections between the actors (affiliations between companies and individual people) and implicit connections such as deals and alliances between companies, acquisitions, or investments from business angels, other companies, business angels, and financial actors. In short, we concentrate on mapping the ecosystem-level network structure of innovation ecosystem actors based on transactional microdata on these actors. Finally, we note that two important groups of innovation ecosystem actors are not included in any of the experiments reported in Chapter 3. First, apart from the Twitter followers of companies in the Tekes Young Innovative Companies Program, we do not include company customers as ecosystem actors. Second, we do not include institutional actors in any of the experiments.

Objective II is to develop general design guidelines for conducting empirical research on the network structure of innovation ecosystems. The identified design guidelines are specifically developed in the context of visual network analytics, which is an application of social network analysis. While we expect that the guidelines will serve investigations that take alternative approaches to visual (social) network analytics, we assume that new requirements will be introduced with other methods, including different variants of predictive statistical analysis, machine learning, and agent-based modeling. Finally, we want to stress that while the orchestration of innovation ecosystems is mentioned as a possible application of innovation ecosystem visual network analytics, the orchestration practices and related empirical work are excluded from this dissertation. We will however discuss orchestration as a potential implication of the results of this dissertation.

Objective III, the primary objective of this dissertation, is to develop a process model for investigating innovation ecosystems with data-driven visual network analytics. To reach the primary objective, we work through the series of experiments described in detail in Chapter 3. We use experiments involving a total of five (5) different innovation ecosystems to develop the process model. Despite the fact that we have interacted with a number of practitioners representing key stakeholder groups in a majority of the targeted innovation ecosystems, we realize that we have limited means to generalize the results beyond the contexts covered in the dissertation. We expect that new requirements will emerge when new types of innovation actors are introduced.

[2] We thank the pre-examiners for pointing out the importance of including this section.

To conclude, we accept that the selection of visual network analytics, an approach combining social network analysis and visual analytics, has limitations. In terms of the four categories of analytics – exploratory, descriptive, predictive, and prescriptive (cf., Bendoly, 2016; Davenport, 2013; Delen & Demirkan, 2013) – visual network analytics allows for exploratory and descriptive analytics. The possible use of visual network analytics together with supporting sensemaking and decision-making processes for predictive and prescriptive analytics is outside the scope of this dissertation. However, we want to point out that modeling the network structure of the innovation ecosystem under investigation introduces a means for quantifying the structure of the network of ecosystem actors at the ecosystem level as well as the structural positions of individual actors in the context of the overall network. Although using this type of quantification in statistical analysis is beyond the scope of the dissertation, we note that the exploratory work conducted in this dissertation is a significant step in using the types of data sources as well as the network modeling practices in future investigations that use a statistical analysis or agent-based modeling (cf., Huotari, Järvi, Kortelainen, & Huhtamäki, 2016).

1.4 Outline and contributions of the dissertation

This dissertation is composed of two complementary lines of investigation. First, a series of innovation ecosystem investigations using visual network analytics is presented. Second, the ostinato model, an exploration-automation cycle for a user-centric, process-automated, data-driven visual network analytics, is developed and described.

This dissertation is divided into eight chapters. The seven articles together with this summary and synthesis of results form the dissertation. The full versions of the articles are attached in this dissertation. Next, the contents of each chapter are described briefly.

Chapter 1 introduces the context of this dissertation, describes the applied research method and epistemological and ontological basis, and specifies the scope and delimitations, objectives, and key contributions of the dissertation.

Chapter 2 reviews the core literature on the key domains of the dissertation, namely innovation ecosystems, network analysis, and visual network analytics, and discusses the suitability of the network approach in investigating innovation ecosystems.

Chapter 3 discusses the results of the investigations from the point of view of data-driven visual network analytics. To ensure the brevity and tractability of the discussion, we concentrate on the specific ways in which the visual visual network analytics were applied in each of the investigations, and we discuss network-related insights.

Chapter 4 introduces the first part of the contributions of this dissertation, that is, the design principles for analyzing innovation ecosystems as networks, and it synthesizes representing and analyzing innovation ecosystems as networks based on the investigations serving as action design research experiments.

Chapter 5 reviews a set of existing process models and examines process model requirements stemming from the experiments to build a foundation for the ostinato model.

Chapter 6 describes the ostinato model and its suitability for data-driven visual investigations of innovation ecosystem structure.

Chapter 7 discusses the contributions of this dissertation, describes the process through which the contributions emerged, and provides a synthesis of the key results.

Chapter 8 concludes the dissertation and presents directions for further work.

We make contributions on two complementary levels. First and most importantly, the ostinato model for conducting data-driven visual network studies on innovation ecosystems is developed and described. Second, a series of investigations in the field of innovation ecosystem exploratory analysis is conducted. These investigations feed new insights into the emerging phenomena of innovation ecosystems and their research. A concrete outcome of the second contribution is a list of design guidelines for representing innovation ecosystems as networks.

In terms of increasing the transparency of the structures emerging from individual transactions and affiliations between innovation ecosystem actors, this dissertation contributes on three levels. First, network analysis is a key approach in supporting exploratory studies on the patterns and structures in-between actors founding, working for, advising, and investing in startups and growth companies; companies making deals and alliances, investing into, and acquiring other companies; and financial organizations investing in startups and growth companies, and in estimating the authority, role and structural positions of these actors, therefore allowing for increasing the transparency of the structure of innovation ecosystems that represent different grades of abstraction and complexity from platforms, programs and business sectors to national and international innovation ecosystems. These structures can be modeled, represented, analyzed, and visualized as networks to support investigation and exploration. Second, the presented data-driven approach allows extending these investigations of patterns and structures within and in-between groups of actors beyond the boundaries of individual datasets and over long periods. Third, actors with different sets of skills can all engage in the different phases of the investigative process.

Moreover, investigations of innovation ecosystems contribute to the body of knowledge on the domain of empirical innovation ecosystem studies. At the same time, these investigations serve as a vehicle for developing the process model in an incremental and iterative manner by following the action design research process.

Our main contributions are as follows:

Contribution I To meet **Objective I**, we contribute to the empirical body of knowledge by providing a visual description of innovation ecosystems from innovation platform, business domain and development program to national and international ecosystems through a series of investigations.

Contribution II To answer **Objective II**, we define a set of design guidelines for modeling and analyzing the structure of innovation ecosystems as networks to support their exploratory analysis.

Contribution III To satisfy **Objective III**, we define the ostinato model for data-driven visual network analytics in the context of innovation ecosystems.

The key result of this dissertation is the introduction of the process model for data-driven visual network analytics in the context of innovation ecosystem investigations. The model is described in detail in **Publication VII**.

2 Defining innovation ecosystems, network analysis and visual analytics

We operate across across three key domains in this dissertation, that is, innovation ecosystems, network analysis, and visual network analytics. In this chapter, we review the extant body of knowledge in these domains according to the approach taken in this dissertation.

2.1 Innovation ecosystems

The ecosystem concept is rooted in biology. According to Moran (1990), “ecosystem generally refers to the structural and functional interrelationships among living organism and the physical environment within which they exist.” Moore (1993) first introduced the concept in the business literature in his seminal article on business ecosystems. The article discusses new ways for a company to allow other companies to create value for them, that is, to the focal company of a business ecosystem. Moore notes that business ecosystems “condense out of the original swirl of capital, customer interest, and talent generated by a new innovation, just as successful species spring from the natural resources of sunlight, water, and soil nutrients” (Moore, 1993). In business and innovation literature, the term ecosystem is used as a metaphor (e.g., Huhtamäki, Russell, Still, & Rubens, 2011; Hwang & Horowitt, 2012), a business strategy artifact (Moore, 1993), and to refer to ecosystem-level analysis (cf., Pentland, 2015).

The innovation ecosystem is a relatively new concept, compared to business ecosystem. It is used in different ways in the literature and practical applications. Hwang and Horowitt (2012), for example, integrated the ecosystem metaphor in their rainforest framework for “building the next Silicon Valley.” To distinguish between ecosystems of business and innovation in the context of this dissertation, we point to their expected outputs. When the key objective in business ecosystems is to organize value creation and value appropriation in an ecosystemic setting, the main output of innovation ecosystems is the increase in information flow and collaboration and therefore the creation of new business-relevant knowledge, ideas and technologies that lead to new products, successful companies, and economic growth. Rephrasing Moore (1993), the innovation ecosystem is the “swirl” and its upstream. Innovation ecosystems survive through a constant idea flow, re-configuration, and evolution (cf., Pentland, 2015).

The key characteristics of ecosystems are interconnectedness, interdependency, co-evolution, value co-creation, and co-opetition (as summarized, e.g., in Huhtamäki et al., 2011; Järvi & Kortelainen, 2016). The actors in business ecosystems and innovation ecosystems are "loosely interconnected (Iansiti & Levien, 2004)."[1] The success of a given innovation often relies on the success the environment of the focal company, that is, companies are interdependent (Adner & Kapoor, 2010). L. D. W. Thomas and Autio (2012) state that co-evolving ecosystem actors "develop over time sympathetically with the other participants in order to maintain stability and health of the ecosystem in the face of change." Ramaswamy (2009) claims that value co-creation is an emerging business and innovation paradigm that leads to the need for "changing the very nature of engagement and relationship between the institution of management and its employees, and between them and co-creators of value - customers, stakeholders, partners and other employees." Finally, Ritala and Hurmelinna-Laukkanen (2009) note that co-opetition, that is, collaboration with competitors, is in some cases "an effective way of creating both incremental and radical innovations, especially in high-tech industries."

Russell et al. (2011) take an even broader scope and define innovation ecosystems as "inter-organizational, political, economic, environmental and technological systems of innovation through which a milieu conducive to business growth is catalyzed, sustained and supported." They note that at the ecosystem level, individual relationships take the shape of a network "through which information and talent flow through systems of sustained value co-creation." Importantly, Russell et al. (2011) include organizational investors and individual people−founders, advisors, and business angels−as innovation ecosystem actors and therefore potential units of analysis in innovation ecosystem investigations (cf., Huhtamäki et al., 2011).

To manage innovation ecosystems and to facilitate their emergence, the process of network orchestration is encouraged (Russell, Huhtamäki, et al., 2015). In general, innovation ecosystems are orchestrated rather than controlled or managed (Paquin & Howard-Grenville, 2013; Ritala et al., 2013, 2009; Russell et al., 2011). When collective gains are sought at the network level, change agents seek to facilitate the emergence of networks, orchestrate existing networks, and manage their growth (Russell et al., 2011). A dynamic innovation ecosystem is characterized by the continual realignment of synergistic relationships that promote the growth of the ecosystem (Russell, Huhtamäki, et al., 2015).

Innovation ecosystem research can take place at three different levels (Järvi & Kortelainen, 2016): "the individual actor, the relationship between actors and the ecosystem as a whole." We claim that the ecosystem-level is imperative in investigating, navigating and orchestrating innovation ecosystems. With Järvi and Kortelainen (2016) as evidence, we note that during the writing of this dissertation, empirical ecosystem-level research on the network structure of business and innovation ecosystems is scarce, which is very likely due to the mismatch between knowledge demand and the existing methods that are perhaps better suited to conduct analysis at the level of actors and relationships.

Discussion on the utility of the innovation ecosystem as a concept is ongoing. A recent critique of the concept (Oh, Phillips, Park, & Lee, 2016) assert that the eco-prefix in ecosystem may only add to the difficulties in communication related to research and its application in decision-making. The ecosystem as a metaphor is powerful and therefore prone to misconception and preposterous thinking (Oh et al., 2016). At the same time, scholars continue to explore the extent to which the biological ecosystem concept can eventually be applied to business and innovation studies. A study on the technospecies concept (Weber & Hine, 2015) is an example of the latter approach. Valkokari (2015)

[1] Iansiti and Levien (2004) stress that "like their biological counterparts, business ecosystems are characterized by a large number of loosely interconnected participants who depend on each other for their mutual effectiveness and survival."

takes a constructive stance to review how ecosystems as "meta-organizations" contribute to management studies.

In this dissertation, we are first and foremost interested in the properties that scholars and decision-makers attach to innovation (eco)systems and the ways the investigations of these properties can be conducted with a data-driven visual network analytics approach. In fact, we believe that empirical research can help to take the discussion to a more tangible level and provide means to test the validity of theoretical propositions related to innovation ecosystems.

In the investigations included in this dissertation, we subscribe to the definition of innovation ecosystem by Russell et al. (2011) and approach innovation ecosystems with a Silicon Valley-style mindset. This means that we do not limit the investigation into companies and their interconnections. Instead, we study the structures that emerge from activities taking place around startups and the individuals that found, advice, invest in, and work in key positions for the startups. We also investigate growth companies that have already crossed the death valley of company growth and continue to evolve toward a liquidity event, such as exiting through an acquisition or initial public offering (IPO). In order to provide extended context, we also study the deals and alliances that connect already established enterprises.

2.2 Network analysis

In this dissertation, innovation ecosystems are investigated using visual network analytics. In network analysis, phenomena under investigation are modeled as nodes and edges representing the entities and their interconnections. We fully subscribe to Yang and Leskovec (2014) that "Networks provide a powerful way to study complex systems of interacting objects." Network analysis is rooted in social network analysis (SNA) (Moreno, 1953; Wasserman & Faust, 1994). Although the first applications of network analysis dates back to 1950's and Milgram's famous experiment, which showed evidence of the small-world nature of the world, was reported in 1967 (Milgram, 1967), network analysis remains a relatively new method in a number of domains. The key steps in network science include the introduction of a model for generating small-world networks (Watts, 1999; Watts & Strogatz, 1998) and the discovery of scale-free networks (Barabási & Albert, 1999). The availability of interesting real-life social data and the development of computing capabilities have significantly increased the potential for applying network analysis in investigating various phenomenon (cf. Bastian, Heymann, & Jacomy, 2009; Cioffi-Revilla, 2010; Hansen, Shneiderman, & Smith, 2011; Lazer et al., 2009).

Social network analysis studies the social structures of actors. Sociogram is the core artifact in social network analysis. Wasserman and Faust (1994) define sociogram as "a picture in which people (or more generally, any social units) are represented as points in two-dimensional space, and relationships among pairs of people are represented by lines linking the corresponding points." The network structure is key in understanding the complex relationships that are latent in ecosystems (Barabási, 2003): "Small changes in the topology, affecting only a few of the nodes or links, can open up hidden doors, allowing new possibilities to emerge."

A phenomenon under investigation can be modeled either as a one-mode, two-mode, or multimodal network. In one-mode networks, all the nodes are of same type. Among company board members, for example, connections between the nodes can be formed based on friendship or company board co-membership. In two-mode networks, there are two types of nodes. A two-mode network of companies and investors show a company node connected to each investment firm from which it has received funding. The means used to visualize two-mode networks include hypergraphs and bipartite graphs (Freeman, 2000; Jesus, Schwartz, & Lehmann, 2009). The connections may be undirected or directed, the latter resulting in a digraph. Further, the connections between the actors can be either dichotomous or weighted, in which the strength of a connection can be expressed.

The analysis of overall network structure, the different characteristics of the network, the roles of the network actors, and the nuances of their interaction are of interest in many fields of research. The structure of a network may be characterized as random, small-world (Milgram, 1967; Watts & Strogatz, 1998), or scale-free (Barabási & Bonabeau, 2003). The actors in networks act as hubs or connectors as they diffuse information within the network (cf., Molka-Danielsen, Trier, Slykh, Bobrik, & Nurminen, 2007). Process phenomena such as preferential attachment (Barabási & Albert, 1999), homophily, reciprocity and transitivity (cf., Giuliani & Bell, 2008) shape the networks as they evolve. Precise SNA metrics can be calculated for all three units of analysis in innovation ecosystem investigations: network actors, connections between the actors, and the entire network. Node degree, that is, the number of connections that a node has is the simplest metric. The main categories of actor-level SNA metrics are centrality and prestige (Wasserman & Faust, 1994). The key metric for connections is their weight. Metrics such as density, diameter, and cohesion (Hansen et al., 2011) are used to describe networks quantitatively.

Network analysis has been applied to investigate companies and different company-related phenomena. Co-creator networks were one of the early applications of visual social network analysis. When reviewing the historical and theoretical foundations of SNA, Freeman (2009) refers to early work of Hobson, who in 1884 had already "produced a visual image of a social network that was not based on kinship." Unlike his predecessors, Hobson designed a two-mode network of corporations and their directors with, according to Freeman (2009), the rationale "that, to the degree that corporations shared directors, they could be expected to cooperate and work together." Levine's work on "corporate interlocks" presents another example of an early visual investigation of corporate networks as relationships through which social norms influence information flow for business intelligence (Levine, 1979). The notion of weak ties (Granovetter, 1973) is another important landmark in applying network analysis to sociological research in studying business and economy. Olson (2008) reports a more recent example of visualizing corporate interconnections implemented with socially constructed data. Basole (2009) pioneers the application of visual network analysis to analyze interfirm relations in the mobile ecosystem and conducts an extensive review of literature on visual network analysis in business studies.

Examples of innovation ecosystem investigations with a visual network analytics approach are rare outside the works of the author and his collaborators. In Table 2.1, we review seven (7) articles reporting empirical studies on ecosystem network structure that Järvi and Kortelainen (2016) identify. Even more recent examples of empirical studies of ecosystem structure exist. One of the research streams is the platform economy. Evans and Basole (2016) and Basole (2016) follow early work by Weiss and Gangadharan (2010) in investigating the network patterns latent in mashup ecosystems, an important manifestation of the digital economy that allows true self-organization in developing networks of platforms and complementary products. Parise, Whelan, and Todd (2015) present an additional recent example of how to utilize network analysis in combination with social network data in analyzing how network structure affects innovation activities.

Table 2.1: Empirical studies on ecosystem network structure (Järvi & Kortelainen, 2016)

Article	Ecosystem	Method
Iyer, Lee, and Venkatraman (2006)	Business ecosystem: "alliances and partnerships established by 509 firms in the packaged software industry"	Network analysis (actor and network-level metrics and small world coefficient), datasources include Thomson Reuters SDC (TR SDC)
Basole (2009)	Business (mobile) ecosystem: "complex structure and dynamics of nearly 7000 global companies and over 18,000 relationships in the converging mobile ecosystem"	Visual network analysis (actor and network-level metrics). Analysis in network and company level. Data sourced from TR SDC and Connexiti
Iyer and Henderson (2010)[a]	Cloud computing ecosystem	Identification of seven cloud capabilities "that that executives can use to formulate cloud-based strategies"
Weiss and Gangadharan (2010)	Mashup ecosystem: "422 new open APIs" ja "1,865 new mashups" created during the 600 days timeframe under investigation	Descriptive statistical analysis and networ analysis (one and two-mode networks). Data sourced mainly from ProgrammableWeb
Battistella, Colucci, De Toni, and Nonino (2013)	Business ecosystem: developing a method for modeling and analyzing business ecosystems as networks, case study using the method	Methodology of business ecosystem network analysis (MOBENA) is developed effectively with design science research approach, case study implemented with MOBENA
Clarysse, Wright, Bruneel, and Mahajan (2014)	Knowledge ecosystem and related business ecosystem utilizing the created knowledge: "hand-collected database of 138 innovative start-ups in the region of Flanders"	Visual network analytics. Descriptive statistics, including basic node and network-level metrics for the modeled ecosystems. Statistical analysis using the network metrics
Still, Huhtamäki, Russell, and Rubens (2014)	Innovation ecosystem: the six EIT ICT Labs co-location cities, companies operating in the cities, investors that have invested into the companies, individuals affiliated with the companies	Descriptive statistical analysis and visual network analysis (multimodal networks). In addition, a case scenario: what if Silicon Valley would be the seventh EIT ICT Labs co-location city?

[a] *We evaluated this article on basis of the abstract because we were not able to access the full article.*

In the large majority of these examples, however, the authors do not refer specifically to innovation ecosystems. A majority of empirical ecosystem studies on ecosystem structure shown in Table 2.1 target business ecosystems. Nevertheless, this research is relevant to the topic of this dissertation. Moreover, empirical work is presented using either vocabulary that does not meet the search criteria Järvi and Kortelainen (2016) use or published in outlets outside the list of journals that they targeted. Ritala and Hallikas (2011), for example, present a notable example of analyzing co-opetition patterns between industrial ecosystem actors. First, Ritala and Hallikas (2011) model a collaboration network with Thomson Reuters SDC data and, second, using "competitive relationships from the companies' public annual reports", they identify pairs of companies and investigate whether the network position of a company in competitive network affects the likelihood of company collaborating with its competitors. Their investigation is a prime example of the combined use of network modeling in introducing new levels of analysis to the data and traditional statistical analysis.

2.3 Visual network analytics

Because of the complexity of business ecosystems, the derivation of conceptual insights from business ecosystem data is challenging (Basole et al., 2015; Bizzi & Langley, 2012). These challenges also exist in innovation ecosystems, and the fact that innovation ecosystems are open dynamic systems further adds to these challenges. The visual revelation of patterns in complex ecosystem data allows for gaining important knowledge of the patterns and dynamics of [innovation] ecosystems (Basole, Clear, Hu, Mehrotra, & Stasko, 2013).

While a statistical analysis provides valuable insights into the structure and dynamics of ecosystems (see, e.g., Ritala & Hallikas, 2011), important knowledge can also be gained through the visual revelation of patterns in a complex business ecosystem data (Basole et al., 2013). Indeed, visualizations are more than merely artistic approaches to depicting structure in helping investigators to explore the data throughout the analysis process (Fox & Hendler, 2011). Visualizations have been used to explore, interpret, and communicate data in order to aid humans in overcoming their cognitive limitations, making structures, patterns, relationships, and themes visible, and providing a means to efficiently compare multiple representations of data in similar fields, such as medicine, dentistry, computer science and engineering. Tufte (1983) suggests that when applied properly, visualization is an extremely valuable tool for understanding and analyzing business issues, including strategy, scenario planning, and problem-solving.

We are excited to observe the existence of individual studies in the literature taking note of the development of analytics-based insights that go beyond quantitative (causal) relationships between individual measurements (Bendoly, 2016; Bygstad & Munkvold, 2011; Williams & Shepherd, 2015). In the context of information systems research, Bygstad and Munkvold (2011) present a critical realist-inspired analytical framework for identifying socio-technical mechanisms in a way that allows for "ontological depth, creative thinking and more precise explanations" that go beyond traditional statistical analysis. Williams and Shepherd (2015) recall the interdisciplinary origins of social network analysis and present a framework for processing data from multiple sources, including archival records for network analysis by using a mixed methods approach. Bendoly (2016) places visual analytics at the center of stage in data analytics, from data validation and curation through exploration and discovery to end-result communication and (re-)interpretation. Further, Bendoly points to the importance of "not just simultaneous parallel use [of data and visualizations] but truly joint utilization and team-wise sensemaking for effective decisions."

In using the term visual network analytics, we refer to taking a visual analytics (Heer & Shneiderman, 2012; J. J. Thomas & Cook, 2006) approach to network analysis. Visual analytics stresses the process-centric, interactive nature of using visualizations in supporting data-driven investigations (Heer & Shneiderman, 2012; Keim, Kohlhammer, Ellis, & Mansmann, 2010). Visual analytics is a particularly suitable approach for exploring new phenomenon using a data-driven approach (Shneiderman, 2014). Visual network analytics allows the emergence of insights into the structure and dynamics of business and innovation ecosystems (Basole, 2009), social media platforms (M. A. Smith, Himelboim, Rainie, & Shneiderman, 2015), and other networked phenomena. Transforming those insights into action, however, requires communicating the insights to the constituents of change (Russell et al., 2011; Still, Huhtamäki, Russell, & Rubens, 2014). Visual network analysis is a valuable method for investigating social configurations and for interactively communicating findings to others (Freeman, 2000).

The well-known mantra of visual information retrieval iterates three phases: "Overview first, zoom and filter, then details-on-demand" (Shneiderman, 1996). In the context of visual network analysis, however, users may prefer to follow the process of "start with what you know, then grow" (Heer & boyd, 2005). Visual analytics theory suggests that, at best, an investigator of a phenomenon is able to interact with all the phases of the process from view creation to exploration and refinement in an expressive manner (Heer & Shneiderman, 2012).

A data-driven process for understanding the roles of the different actors of an innovation ecosystem allows for interactive discovery that supports both investigation and orchestration of innovation ecosystems (Russell, Huhtamäki, et al., 2015). Multiple perspectives on ecosystem structure and the structural positions of individual actors can be invited and exchanged during the investigative process. In the subsequent automation of data updates and tracking analyses, the assumptions and contingencies that underlie decisions can be monitored for changes that would impact evidence-based conclusions, policies, and program directions.

While research issues related to innovation ecosystem orchestration and related management practices are beyond the scope of this dissertation, we will quickly review a topic of enacted sensemaking, which is imperative in utilizing visual analytics in management (Bendoly, 2016), or in this dissertation orchestration of innovation ecosystems. We are excited to take notice of Bendoly's (2016) recent article, which places enacted sensemaking into the center of the stage of using visual analytics in management. In order to support the design of visual analytical tools for management, Bendoly (2016) provides a three-part proposition. First, he applies Peirce's[2] "classic semiotic framework" (referring to Peirce (1958a, 1958b)) to develop a framework to support "the consideration of which visual representation of data is best suited to data analytic tasks" (Bendoly, 2016). Second, he proposes the standard convention, a taxonomy of visualization characteristics or a set of visual traits or idioms that serve as the building blocks for systems of idioms. In concrete terms, traits can be implemented as individual visualizations that can be compiled into dashboards or systems of traits. Third and perhaps most importantly, he points to the existing work on enacted sensemaking (Weick, 1988; Weiss & Gangadharan, 2010) in which an actor seeking to understand the dynamics of a phenomenon, here an orchestrator of an innovation ecosystem, iterates interventions and analytics of their impacts on the target phenomenon with the objective of gaining a holistic understanding of the causal mechanisms taking place in the phenomenon. We note that there seems to be good synergy between Bendoly (2016) and critical realist data analytics (Bygstad & Munkvold, 2011).

[2]Charles Sanders Peirce (1839–1914) was "an American philosopher, logician, mathematician, and scientist" who made significant contributions to the domain of semiotics, see `https://en.wikipedia.org/w/index.php?title=Charles_Sanders_Peirce&oldid=729438385`.

In the experiments included in this dissertation, we use network nodes to represent innovation ecosystem actors. The actors are connected to each other based on different kinds of transactions between them. These transactions include investments, acquisitions, deals, and alliances. Moreover, key individuals are connected to companies with which they are or have been affiliated. Network metrics are used to quantify the structural positions of these actors. Network representations are laid out in a way that allows their visual investigation. The visual properties of network nodes and edges are specified to support these investigations.

2.4 Cognitive fit of network analysis

The network representation of information is utilized extensively in information systems. Indeed, the network representation of information is at the core of hypertext. In presenting the notion of the hyperlink, Vannevar Bush highlights the similarity in between network representation and the way the human mind operates (Bush, 1945):

> *The human mind [...] operates by association. With one item in its grasp, it snaps instantly to the next that is suggested by the association of thoughts, in accordance with some intricate web of trails carried by the cells of the brain. It has other characteristics, of course; trails that are not frequently followed are prone to fade, items are not fully permanent, memory is transitory. Yet the speed of action, the intricacy of trails, the detail of mental pictures, is awe-inspiring beyond all else in nature.*

We claim that representing innovation ecosystems as networks allows for an intuitive way to investigate and revisit data representing innovation ecosystem actors and their activity. Indeed, Bush (1945) added that "[m]an cannot hope fully to duplicate this mental process [the intricacy of trails in human mind] artificially, but he certainly ought to be able to learn from it." In this dissertation, we take advantage of existing sources of digital data on innovation ecosystem actors and the transactions and affiliations between them to recreate traces that allow for ecosystem-level views of the currently scattered innovation ecosystem landscape. To reuse McGonigal's (2005) insightful metaphor, which was inspired by Lewis Carroll, we see visual network analytics as the means to craft "rabbit holes" through which innovation ecosystem explorers are drawn into new information landscapes beyond their imagination to find new, interesting ideas, companies, investors, communities, and similarly minded innovation champions (cf., Huhtamäki, 2007), as well as emerging patterns that may provide competitive intelligence (cf., Adner, 2012; Basole, 2014; Basole et al., 2015) to a company, ecosystem, program, region, or nation.

Cognitive fit theory suggests that in supporting problem-solving and decision-making, it is particularly important to find a fit between the problem-solving task and the problem representation and supporting tools; cognitive fit allows for faster and more accurate performance in decision-making (Vessey & Galletta, 1991). There is a delicate balance in developing tools with cognitive fit, however: empirical research shows that neither the time used to conduct a task nor the confidence that the users feel about their decisions are good predictors of the accuracy of their insights or decisions (cf., Dunn & Grabski, 2001). Cognitive fit theory further suggests that when the problem representation fits the problem-solving task, a preferable and more consistent mental representation of the problem will be realized, thereby facilitating the problem-solving process, and consequently resulting in preference for the representation, along with faster and more accurate performance in decision-making (Basole, Huhtamäki, Still, & Russell, 2016).

Using network analysis in the context of investigating and orchestrating innovation ecosystems is closely related to relational and social capital theory. In their seminal article on social capital,[3] Nahapiet and Ghoshal (1998) introduce a third dimension of social capital to complement the relational (relationship-level social capital) and structural dimension (ecosystem-level social capital) and name that the cognitive dimension to refer "to those resources providing shared representations, interpretations, and systems of meaning among parties." Nahapiet and Ghoshal (1998) "believe it represents an important set of assets not yet discussed in the mainstream literature on social capital but the significance of which is receiving substantial attention in the strategy domain." From the viewpoint of the research on relational capital, this dissertation seeks to use visual network analytics to augment the cognitive dimension of ecosystemic relational capital (Still, Huhtamäki, & Russell, 2014).

[3] Still, Huhtamäki, and Russell (2013) show that Nahapiet and Ghoshal (1998) is a key article in between social capital and relational capital research.

3 Experiments in investigating innovation ecosystems

In this chapter, we provide an overview of individual experiments on investigating innovation ecosystems with data-driven visual network analytics. In this dissertation, the experiments primarily serve as the means to develop an approach and methodology for investigating innovation ecosystems as networks, that is, to distill design principles and process requirements in interaction with innovation ecosystem actors, stakeholders, decision-makers, and investigators through the process of guided emergence (cf., Sein et al., 2011). Moreover, the results and insights contribute to **Objective I** of the dissertation.

The experiments are conducted on innovation ecosystems that represent different grades of abstraction and complexity from the Demola platform, a local ecosystem engager, to EIT ICT Labs (currently operating as EIT Digital), a large international "open innovation organisation."[1] In all the experiments, we take a data-driven approach to explore and describe the structure and in some cases the structural dynamics of the innovation ecosystems under investigation.

The innovation ecosystems in which the experiments take place represent a broad range of abstraction and complexity. For specificity, we present the following categorization of the innovation ecosystems: Innovation platform: Demola; Business domain: Mobile Ecosystem; Development program: Tekes Young Innovative Companies;[2] National ecosystem: Finnish Innovation Ecosystem; International ecosystem: EIT ICT Labs.

Here, we will review the experiments from local ecosystems to global ecosystems in that order. We want to note, however, that the order in which the experiments were conducted is different from the categorical order. The Finnish Innovation Ecosystem was the first experiment. Second, the investigation of the mobile ecosystem was conducted. Third, we took a network approach to investigate the ecosystem emerging through Tekes Young Innovative Companies program. Fourth, the means to use visualization to support orchestration of EIT ICT Labs were investigated. Fifth, we revisited the Finnish Innovation Ecosystem by using a multiscopic approach. Sixth and finally, the evolution of the innovation ecosystem operating on the Demola platform was investigated. Table 3.1 describes the individual experiments in more detail.

Before beginning the review of the experiments, we want to discuss the criteria used for the selection of innovation ecosystems the serve as the platform of our experiments. To reiterate the definition that we adopted for this dissertation (Russell et al., 2011), innovation ecosystems are "inter-organizational, political, economic, environmental and

[1] EIT Digital, `http://eit.europa.eu/eit-community/eit-digital`
[2] Tekes Young Innovative Company funding program, `http://www.tekes.fi/en/funding/yic/`

technological systems of innovation through which a milieu conducive to business growth is catalyzed, sustained and supported" where individual relationships in-between ecosystem actors, including companies, organizational investors and individual people−founders, advisors, and business angels−take the shape of a network "through which information and talent flow through systems of sustained value co-creation." Moreover, we restate the focal point of empirical innovation ecosystem investigations in this dissertation: the exploratory analysis of the ecosystem-level network structure of innovation ecosystems.

Demola is an "ecosystem engager," a platform that brings together companies and students and facilitates the process in which students seek ways to create new value for companies. Demola follows many of the practices of Five Disciplines of Innovation (Carlson & Wilmot, 2006), and seeks new ways to operate as an innovation ecosystem facilitator. Moreover, Demola Tampere operates as part of New Factory through which newly founded companies can explore ways to grow and get funding. The experiment on Demola is different from the other experiments included because we use Demola's internally collected project data and focus on making the Demola process and its immediate impact visible and tangible.

In the mobile ecosystem experiment, we seek ways to take a data-driven approach to investigate the "heterogeneous and continuously evolving set of firms that are interconnected through a complex, global network of relationships." Specifically, we investigate the ecosystem structure of two strategic relationships that were recently formed during the time of the investigation: the strategic partnership between Nokia and Microsoft and Google's acquisition of Motorola Mobility. The mobile ecosystem experiment gives us an opportunity to investigate the insights that different types of data, traditional deals and alliances data from Thomson Reuters SDC and socially constructed Innovation Ecosystems Network (IEN) Dataset, provide to investigate the interconnections between various types of innovation ecosystem actors. Although not all the deals and alliances are directly related to innovation, "alliance networks link the firm to a vast amount of knowledge, resources, and capabilities," and alliances can therefore serve as "access relationships" (Ritala & Hallikas, 2011). Moreover, the role of venture capital funding in different parts of the ecosystem is important for innovation strategies and policies.

The experiments related to the Finnish innovation ecosystem enable us to investigate the interconnections between various types of innovation ecosystem actors at the national level. In the first experiment, we focus on the funding of startups and growth companies, as well as the key individuals related to the companies. In the second experiment, we investigate both the growth companies as well as the deals and alliances between already established companies. Importantly, we also aggregate these separate views to answer the need that (Valkokari, 2015) points out: "relationships and interactions between ecosystems types need to be analyzed at several levels in order to understand how connections flow between different ecosystems in the real business world."

Finally, EIT ICT Labs represents yet another boundary specification for companies, venture capital investors, and individuals. As we note in **Publication VI**, "EIT ICT Labs (http://eit.ictlabs.eu) operates in a complex ecosystem of independent and interdependent actors, financing schemes and business models that create value for the European innovation landscape, and whose innovation strategy is positioned toward its mission of enhancing this ecosystem to synergize and accelerate innovations contributing to economic growth." A key objective in this experiment is to explore the ways in which international institutions, here EIT ICT Labs, are able to identify the existing structure of an innovation ecosystem to benchmark the structure before engaging with the ecosystem as well as to identify individual actors for tailored action.

For tractability and presentation brevity, we review the experiments specifically from the viewpoint of network analysis, and we address the findings that are related to the design decisions about the network analysis of innovation ecosystems.

Table 3.1: Overview of the investigations

Ecosystem	Data	Key outputs
Demola (**Publication I**)	Proprietary data on Demola projects, the companies that initiated the project and university affiliations of project members (university students)	The animation of the evolution of Demola project sphere including projects, companies, and the affiliations of project team members. Multimodal networks of 1) projects and affiliated actors and 2) projects and their key competences
Mobile Ecosystem (**Publication II**)	Thomson Reuters SDC and IEN Dataset	Dataset-specific visualizations of the innovation ecosystem surrounding pairs of ecosystem actors (Nokia and Microsoft, Google and Motorola Mobility)
Tekes Young Innovative Companies (**Publication III**)	IEN Dataset on growth companies, Twitter data on Tekes Young Innovative Companies (YIC) and their followers	One and two-step networks of the companies taking part in Tekes YIC program and their affiliations to investors and key individuals
Finnish Innovation Ecosystem (**Publication IV**)	IEN Dataset on growth companies	Network visualizations and metrics on Finnish companies and their first-step connections to other companies, investors, and key individuals
Finnish Innovation Ecosystem (**Publication V**)	Three separate datasets: 1) Thomson Reuters SDC for deals and alliances and IEN Dataset for 2) Executives and Finance and 3) Startups and Angels	Network visualizations and metrics on companies having their main office in Finland and their first-step connections to other companies, investors, and key individuals
EIT ICT Labs (**Publication VI**)	IEN Dataset for Executives and Finance	Network representation of EIT ICT Labs co-location cities and their first-step connections to investors, individuals, and other companies

3.1 Innovation platform: Demola

In this dissertation, Demola represents an innovation platform. Demola is an innovation ecosystem engager, an open innovation platform that takes real-life problems from companies and other organizations and puts together and facilitates projects where students from different universities collaborate to solve problems. The investigation focuses on the Demola Tampere site, the original location of the platform that by 2016 had spread to more than ten locations worldwide.[3]

[3]Demola – Building The World's Strongest Innovation Ecosystem, `http://www.demola.net/`

In **Publication I**, we (Huhtamäki, Luotonen, Kairamo, Still, & Russell, 2013) describe a set of network visualizations and animations that were developed in collaboration with the Demola operators with the objective of making Demola-initiated activity visible and tangible. Moreover, the development process that we used to design the visualizations and the technical process is described and discussed in the article. We claim that static network visualizations and, importantly, the animations of the evolution of an open innovation platform development are useful in presenting, describing, marketing, and selling the platform for existing and new stakeholders. Figure 3.1 shows a snapshot of the animation of the evolution of Demola project sphere. Our experience in the experiment suggests that in order to develop visualizations and animations that meet the requirements set by the different stakeholders, an iterative and incremental development process is needed. Moreover, we show that taking a data-driven approach to visualization development is a key enabler in supporting the development process.

3.1.1 Rationale for network analytics

Open innovation breaks the traditional patterns of innovation work. Consequently, this change introduces new requirements for measuring innovation (cf., Still et al., 2012). For an ecosystem engager such as Demola, many of the traditional innovation metrics, including change in company turnover, the number of patents, companies or scientific publications created, or the amount of new product launches, cannot be easily traced to individual projects or even to companies' engagement with Demola in general. In fact, many Demola stakeholders see these outputs as irrelevant to the core activity. At the same time, Demola needs to communicate its activities and its impact both internally and externally.

In this investigation, we joined with members of the Demola operating team to design ways to represent the structure and dynamics of the Demola platform. The investigative team made an inventory of the key challenges that Demola representatives face in measuring and communicating their innovation activities and their impact. The majority of visualization and animation development was conducted by a team of three: 1) a person with deep knowledge of Demola's vision, mission, and strategy; 2) a person with specific knowledge of the existing system used to manage project data; 3) a person with knowledge of applying visual network analytics for innovation ecosystem analysis and visualization.

The network approach allowed us to reuse and refine some of the existing processes developed for investigating other innovation ecosystems in the context of an individual open innovation platform.

3.1.2 Data sources

Demola runs a dedicated web-based platform for setting up new projects as well as for running existing ones. During the first planning sessions, it became evident that the data Demola already collected and produced provided a useful representation of the structure and evolution of the ecosystem.

For this investigation, we concentrated solely on the in-house project data. The data was exported from the Drupal-based Demola platform with a tailored batch script and serialized in comma-separated values (CSV) format for further processing. Demola's operating team curated the project data to meet the needs of the visualization process, particularly the naming of the project themes requiring harmonization.

3.1.3 Network modeling decisions

Projects, collaboration partners, project team members, and their universities were all intuitive candidates to be used as network nodes. In addition, we decided to use nodes to represent the project key areas. Figure 3.1 presents a screen capture of the key output of the experiment, which is an animation of the evolution of the network structure of Demola Tampere ecosystem.

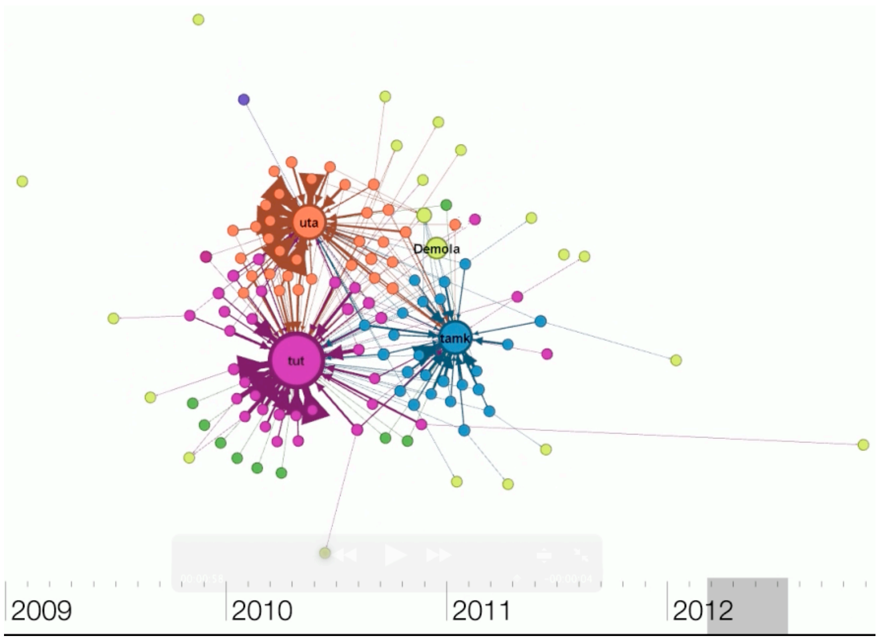


Figure 3.1: Innovation platform evolution: animating Demola evolution (Huhtamäki, Luotonen, et al., 2013)

The network analysis and visualization process model (Hansen et al., 2012) was used to frame the analysis process. The implementation of the visualization process is an interplay between tailored code and the use of pre-existing tools. Python and NetworkX are used to pre-process the data and the visualizations and the animation are created with Gephi.

In the network visualization shown in Figure 3.1, node size represents its betweenness value, and edge weight shows the number of students affiliated with a particular university. The node colors represent network clusters. A force-driven algorithm is used to lay out the nodes. To show the dynamics of the Demola platform, edges connecting companies to individual projects only exist during the project after which the layout algorithm starts pushing loose nodes away from the center. When the force-driven layout is run in real time, the animation shows the retention patterns of individual companies engaging with the platform. The capturing of the video was done with a screen-recording software and timeline is included during post-production.

3.1.4 Results and network-related insights

The key output of the Demola investigation was the animation of the evolution of Demola Tampere Ecosystem from its first day of operation to the beginning of 2013 when the investigation was conducted.

Based on the feedback received both during the development process as well as after the publishing the study, we claim that the dynamic animation of the platform evolution is particularly useful in presenting, describing, marketing, and selling the platform for existing and new stakeholders. As evidence, we offer the fact that the Demola operating team has continued to use the animated project network as a tool for communicating Demola activities and their evolution over time.

3.2 Business domain: mobile ecosystem

The mobile ecosystem consists of a heterogeneous and continuously evolving set of companies that are interconnected through a complex global network of relationships. However, there is very little theoretical understanding of how these networks emerge and evolve. Moreover, there is no well-established methodology to study these phenomena. Traditional approaches have primarily utilized the alliance data of relatively established firms; however, these approaches ignore the vast number of relevant ecosystem activities that occur at the personal, entrepreneurial, and university levels. In **Publication II**, we (Basole, Russell, Huhtamäki, & Rubens, 2012) argue and empirically illustrate that a data-driven approach can provide important complementary explanatory insights into the dynamics of the mobile ecosystem. We present our approach through two mobile ecosystem relationships that were formed just before the investigation−the strategic partnership between Nokia and Microsoft and Google's acquisition of Motorola Mobility. Our analysis is supported by network visualizations. We conclude with implications and future research opportunities, some of which we ourselves took up in an extended version of the study (Basole et al., 2015).

This investigation was conducted without external stakeholders.

3.2.1 Rationale for network analytics

The investigation stems from the fact that there is very little theoretical understanding of how ecosystems emerge and evolve (Ahuja, Soda, & Zaheer, 2012). Network modeling is an integral part of investigating ecosystem structure at the ecosystem level. Companies, investors, and individuals are represented as network nodes, and their interconnections are tracked over time to analyze the emergence of ecosystem patterns. Snapshots of the structure are used to analyze the evolution of the structure. We admit that network analysis alone is not enough to develop rich insight into the structures and mechanism underlying emergence and evolution. Hence, we propose further investigations based on agent-based modeling (cf., Huotari et al., 2016).

3.2.2 Data sources

The ecosystems surrounding the pairs of companies, Google and Motorola Mobility and Microsoft and Nokia, are investigated by using two complementary datasets. First, we use Thomson Reuters SDC Platinum, an institutionally curated dataset on deals and alliances between already established companies in order to gather transactional microdata on the direct and second-tier connections of the pair of companies. Second, to add to the insights into startups and growth companies, venture capital investors and even key individuals, we examine the IEN Dataset for complementary views of the ecosystem. The IEN Dataset is a collection of socially constructed datasets on executives, finance, business angels, and startups (Rubens, Still, Huhtamäki, & Russell, 2010).

3.2.3 Network modeling decisions

Figure 3.2 shows the ecosystem surrounding Google and Motorola Mobility based on Thomson Reuters SDC data. Figure 3.3 shows the startup and growth company centric view of the ecosystem surrounding Google and Motorola Mobility. Separate networks are created to represent the ecosystem structure emerging from Thomson Reuters SDC (Figure 3.2) and IEN Dataset (Figure 3.3). In the former, nodes represent companies and, in the latter, they represent companies, investors, and individuals. In the former, companies are connected by joint deals and alliances. In the latter, companies are connected to investors, key individuals, and other companies that have either acquired or invested in a company.

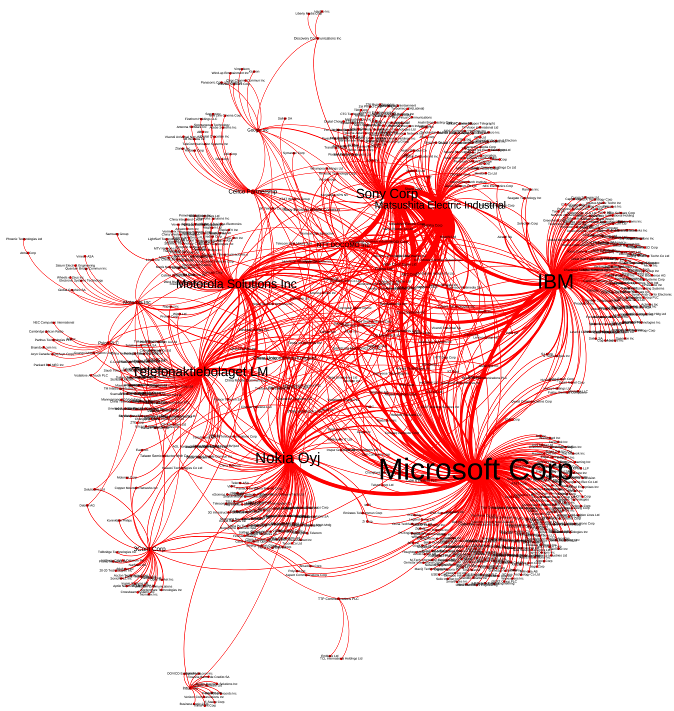


Figure 3.2: Cumulative network around Google and Motorola Mobility using Thomson Reuters SDC data for deals and alliances. Through second-step connections, Microsoft becomes the key node in the network. (Basole et al., 2012)

The boundary specification includes the selection of nodes, node types and relationship types, analysis timeframe, as well as the steps from the focal companies to include other companies and stakeholders (Basole et al., 2012). In this investigation, all companies and other actors within two steps from the focal companies are included in the network. A force-driven algorithm is used to lay out the network.

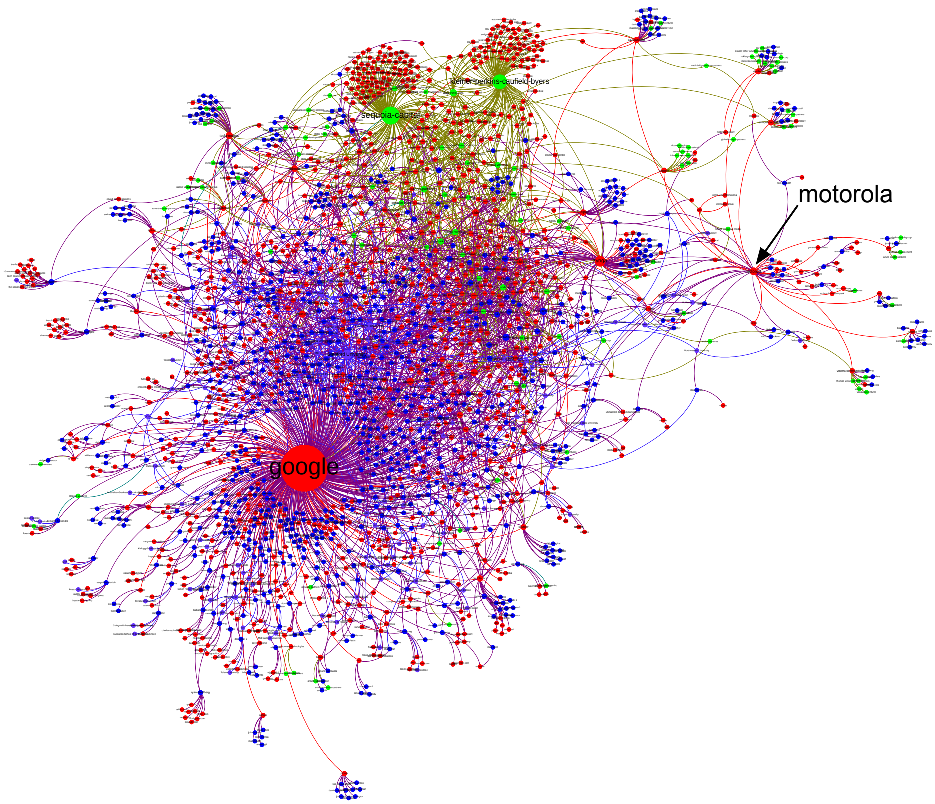


Figure 3.3: Cumulative network around Google and Motorola Mobility. Google's approach to operating through acquisitions rather than deals and alliances (cf. Figure 3.2) is revealed. Venture capital investors in green take a key role in the ecosystem. (Basole et al., 2012)

To investigate the evolution of the structural properties of the ego-centric networks, we created snapshots of the networks and calculated the development of a set of node and network metrics over time. These measurements are represented as small multiples in the article.

3.2.4 Results and network-related insights

Two illustrative investigations exemplify the use of the data-driven approach in order to understand the mobile ecosystem. In February 2011, Nokia and Microsoft announced a strategic alliance to work together in developing their mobile offering, in terms of devices as well as an ecosystem feeding applications and other content to add value to device users. Google acquired Motorola Mobility in August 2011 to strengthen its capabilities in the mobile domain.

Separate network representations are created for the two sets of data as well as the two cases. Figure 3.2 shows the two-step network of Google and Motorola mobility. Thomson Reuters SDC Platinum data represents the ways that traditional enterprises operate. Even though it is not a focal company in this representation, Microsoft emerges as the supernode in the network. The network representation of IEN Dataset in Figure 3.3, Google's true size accumulated particularly through a series of acquisitions, and the flow of talent is shown.

3.3 Development program: Tekes Young Innovative Companies

In **Publication III**, we (Huhtamäki, Still, Isomursu, Russell, & Rubens, 2012) explored a vital part of the Finnish innovation ecosystem: the startups that are selected to participate the Tekes Young Innovative Companies (YIC) program for support in fast international growth. Highlighting the importance of networks, we proceeded to analyze the existing relationships that these companies have with other companies, financing organizations, and the individuals taking part in their co-creation. Moreover, we investigated the network of Twitter followers of the companies taking part in the YIC program for creating insights into the volumes of perception that the startups have accumulated.

The investigation was conducted as part of a Tekes-funded innovation research project on using social media data to measure innovation. The investigative team designed the data collection and network modeling process, and they interacted with Tekes representatives to fine-tune the parameters related to boundary specification and network modeling.

3.3.1 Rationale for network analytics

The companies are selected into the YIC program based on their individual properties through evaluation panels. In taking a data-driven approach to investigating the innovation ecosystem around these companies, our objective is to gain insights into the underlying latent structure in-between companies and create an ecosystem-level view of the innovation ecosystem network structure.

We propose that these existing relationships in-between startups may be used to make sense of companies' role as resource integrators within an ecosystem, contributing to its growth and success. Overall, we claim that network analysis and resulting network visualizations provide novel insights into the understanding of possibilities for global growth and success and, importantly, of the impacts of the YIC program at the ecosystem-level.

3.3.2 Data sources

The list of startups participating in the Tekes YIC program was scraped from Tekes homepage.[4] The IEN Dataset (Rubens et al., 2010) was used to gather data on companies, investors, key individuals, and acquisitions.

Moreover, the Twitter usernames of the YIC companies were compiled in a spreadsheet in a semi-manual manner, and a tailored script was implemented to crawl Twitter REST API[5] to collect the list of followers of each YIC company with a Twitter account.

3.3.3 Network modeling decisions

For boundary specification, startups taking part in the Tekes YIC programs are used as focal points in the network shown in Figure 3.4. All the investors that have invested in the companies in the Tekes YIC program are included, as well as all the individuals

[4]Startups at the Tekes Young Innovative Company programme, `http://www.tekes.fi/en/funding/yic/companies/`

[5]Twitter Developers: REST APIs, `https://dev.twitter.com/rest/public`

that, according to IEN dataset, are affiliated with the companies. In addition, to explore the larger context of the companies in YIC, all the other companies that the individuals are or have been affiliated with are included in the network. Companies in the Tekes YIC program are represented as light blue nodes, other companies are red, individuals are blue, and investors are green. The node size represents its betweenness centrality to highlight the key nodes in bridging the different parts of the network. Finally, the network is laid out using a force-directed algorithm.

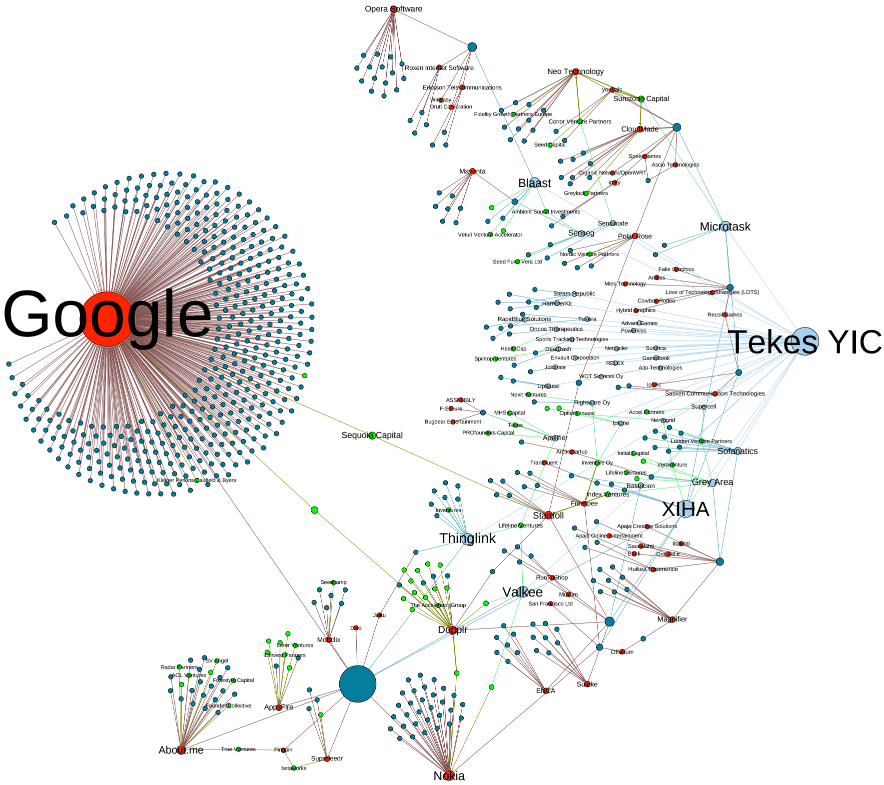


Figure 3.4: Boundary specification: 3-step network visualization of Tekes Young Innovative Companies, their direct connections and the companies, investors and individuals that can be reached through the direct connections (Huhtamäki et al., 2012)

In the Twitter network shown in Figure 3.5, all the followers of a company are connected to the company by a directed edge that points toward the company. The result is a two-mode or bipartite directed dichotomous network. The node size represents its indegree. A force-driven layout algorithm is used the lay out the Twitter network.

3.3.4 Results and network-related insights

The investigative team joined with Tekes YIC program representatives to identify the key insights. In the first steps of connecting investors and individuals, it became evident that the companies that were selected individually to participate in the Tekes YIC program are not without interconnections. When another step was added, the companies connected to the YIC program indirectly through investors or individual people were included

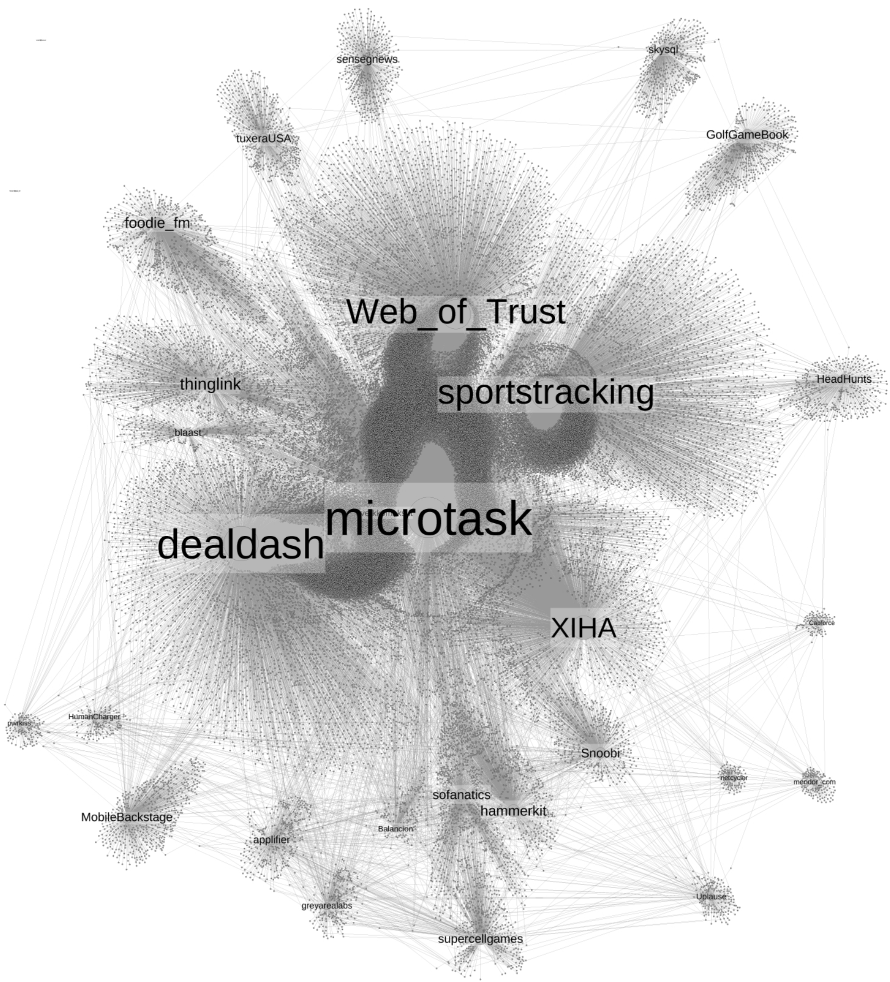


Figure 3.5: Social media analytics: 2-step network visualization of Tekes Young Innovative Companies and their followers on Twitter (Huhtamäki et al., 2012)

and additional connecting tissue was introduced. Both Google and Nokia were include through second-step connections. An individual becomes the sole bridge between Nokia and Google and therefore has the largest betweenness value.

Using background information, we know that the early social media platform Jaiku is one of the key individual developments in the Finnish Innovation Ecosystem. After Google acquired Jaiku, its founders, Jyri Engeström and Petteri Koponen, continued their ventures, Engeström as a serial entrepreneur in the San Francisco Bay Area and Petteri Koponen through Lifelines Ventures, which he co-founded. Interestingly, Lifeline Ventures is one of the early investors in Supercell, the most recent success story in the Finnish innovation ecosystem. More recently, Supercell CEO Ilkka Paananen joined the advisors of Lifeline Ventures to support the creation and growth of new startups.

Microtask, Web of Trust, Sportstracking, and DealDash form the core of the Tekes YIC startup follower network. Microtask leads the charts with more than 30,000 followers, followed by DealDash and Web of Trust. Supercell (@supercellgames) had less than 1000 followers during the time of the investigation. In May 2016, more than 280,000 Twitter users were following Supercell.

3.4 National ecosystem: Finnish Innovation Ecosystem

The Finnish Innovation Ecosystem is investigated in two consecutive experiments presented in **Publication IV** and **Publication V**. Of all the experiments included in this dissertation, **Publication IV** is the first we (Huhtamäki, Russell, Still, & Rubens, 2013) conducted.[6] Therefore, its main contribution is in serving as a stepping stone for further investigation.

In **Publication V**, we (Still, Huhtamäki, Russell, Basole, et al., 2013) take up a grand challenge to present a solution for modeling and visualizing the network structure of the innovation ecosystem of a single nation, in this case Finland. To create an ecosystem-level view of the Finnish Innovation Ecosystem, we utilize three complementary sources of data, each representing different aspects of the ecosystem.

We resolve the limitations of the separate datasets by building multiscopic views into networks of innovation relationships, by using separate datasets. Moreover, in order to create an ecosystem-level view of the Finnish Innovation Ecosystem, we create an aggregated dataset by combining the three individual sets of data. We proceed to support the interpretation of these visualizations by explaining the context of the network with network metrics as well as other descriptive metrics.

The two investigations of the Finnish Innovation Ecosystem were conducted in innovation research projects funded by Tekes innovation research. The investigative team conducted the investigations independently; however, project steering group members representing Finnish innovation ecosystem stakeholders had a role in formulating the research questions that guided the studies.

3.4.1 Rationale for network analytics

The investigation of the Finnish Innovation Ecosystem is an attempt to utilize the potential of the data-driven approach to network analytics at the level of a national ecosystem. It stems from our observation that traditional ways of measuring innovation inputs and outputs are often industry-level aggregates and therefore do not allow for ecosystem-level insights into the structure of an innovation ecosystem under investigation (Still et al., 2012).

The network representation of a national innovation ecosystem allows for observing several features, such as the role of individual investors and patterns of acquisition and workforce flow. In addition, the network representation enables the development of a context for measurements and tailored action. More importantly, the ecosystem-level view enables co-referencing and therefore more explicit support for sensemaking, and it attempts to form a shared vision between ecosystem stakeholders (Russell et al., 2011). To summarize, the presented multiscopic views of the Finnish Innovation Ecosystem allow for examining the relationships supporting value co-creation among various categories of ecosystem

[6] The paper on the investigation was first presented in EBRF conference in 2010 and invited to be published in TIM Review, see `http://timreview.ca/article/424`, and was eventually selected to book Value Co-Creation: Best of TIM Review.

actors as well as between those categories, thus providing novel possibilities for network orchestration and innovation management.

3.4.2 Data sources

Three complementary sets of data are used in the investigation. Thomson Reuters SDC includes connections between large, already established companies. In addition, to enable the investigation of the relationships around startups and growth companies, we use IEN Startup and IEN Growth to complement the analysis. Specifically, IEN Startup is used to create the microscopic view, IEN Growth the mesoscopic view, and Thomson Reuters SDC the macroscopic view.

In addition to using the three sources of data in parallel, an aggregate dataset is created. The three datasets in use in the investigation are complementary and partly overlapping, therefore necessitating a refinement and curation process similar to similar to that applied for example in data journalism (see Gray, Bounegru, & Chambers, 2012) in creating an aggregated dataset.

3.4.3 Network modeling decisions

Publication V is the first investigation in which we provide a multiscopic view of an innovation ecosystem, in this case, the Finnish innovation ecosystem. We create four different views of the Finnish Innovation Ecosystem. The macroscopic view shows the connections in-between already established enterprises. Finnish companies are included and connected to other companies, Finnish or foreign, through deals and alliances. The mesoscopic view is built around Finnish growth companies. All the organizational investors and key individuals affiliated with the companies are included and connected to the Finnish companies. The microscopic view includes Finnish startups as well as business angels and other seed-level investors and individual people affiliated with the companies. The multiscopic view, which is an aggregate of the three views, provides a holistic ecosystem-level view into the Finnish innovation ecosystem.

To complement the multiscopic views into the Finnish Innovation Ecosystem, several quantitative descriptions for the different network representations are included in the article. These include the number of nodes, number of connections, density, and diameter. Moreover, we list the top 10 actors based on betweenness centrality and degree for each of the networks.

3.4.4 Results and network-related insights

The main results of the investigation pertain to the multiscopic view of the Finnish Innovation Ecosystem, as shown in Figure 3.6. The aggregated network depicts an ecosystemic view of Finland in Figure 3.6, which combines the Finnish companies from the three separate datasets and shows their direct connections. Hence, for the first time, we can see a single network representation of an ecosystem of the founders and angels, executives and financing organizations, as well as companies from startups to established enterprises. Overall, the key actor in the ecosystem with the highest betweenness centrality is not surprising: Nokia is the super-node due to its connecting role in the Finnish ecosystem.

As the weight of micro and meso-level data is greater, the top-10 list of actors in the multiscopic level based on both betweenness and degree includes a significant number

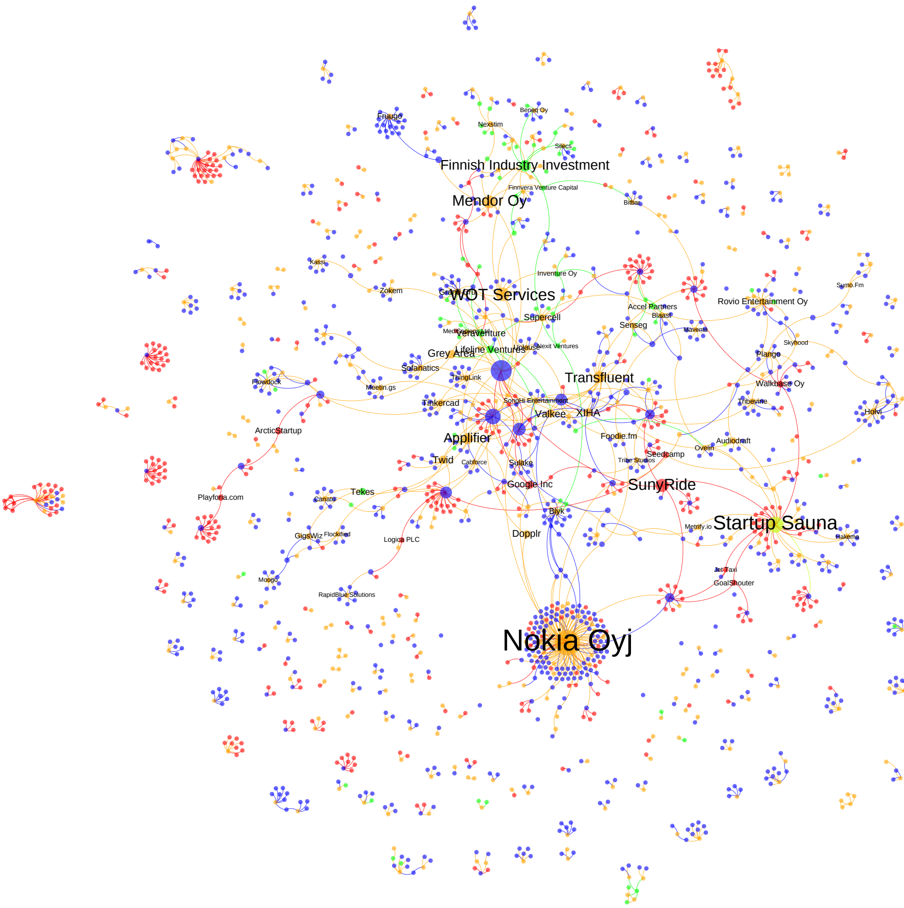


Figure 3.6: Data aggregation: Aggregate view of the Finnish Innovation Ecosystem using data on deals and alliances, executives and finance, and business angels and startups (Still, Huhtamäki, Russell, Basole, et al., 2013)

of individuals. There are seven shared nodes between micro and macro-level views; 184 between micro and meso-level views; 10 between meso and macro-level views. Four nodes appear in all three views: Rovio Entertainment, F-Secure, Mendor, and Nokia. Several foreign companies are included in the network through second-step connections, such as through acquisitions, investors, or individuals affiliated with Finnish companies.

3.5 International ecosystem: EIT ICT Labs

The network structure of the EIT ICT Labs innovation ecosystem is investigated in **Publication VI** (Still, Huhtamäki, Russell, & Rubens, 2014). EIT ICT Labs, rebranded as EIT Digital [7] in June 2015, is a “Leading European open innovation organisation.”

The investigation presented in **Publication VI** builds heavily on Still et al. (2012) with the objective of exploring the opportunities for supporting the orchestration of innovation

[7] EIT ICT Labs becomes EIT Digital, `https://www.eitdigital.eu/news-events/news/article/eit-ict-labs-becomes-eit-digital/`

ecosystems, hence contributing to a fundamental capability in the networked world. We use a data-driven, relationship-based and network centric approach to operationalize innovation ecosystems transformation framework (IETF) (Russell et al., 2011). We present an analysis, evaluation, and interpretation in order to provide decision support and insights into transforming innovation ecosystems.

The investigative team joined with representatives of EIT ICT Labs at the Helsinki co-location center to conduct the two investigations (Still et al., 2012; Still, Huhtamäki, Russell, & Rubens, 2014). In addition to mapping the network structure of EIT ICT Labs, the investigative team conducted a scenario planning experiment that emerged trough the interaction with EIT ICT Labs representatives: What if the San Francisco Bay Area was the seventh node of EIT ICT Labs? This is an example of a what-if question, which is a key feature in scenario planning (Schoemaker, 1995).

3.5.1 Rationale for network analytics

We investigate how data-driven network visualizations can be used to produce insights that support innovation ecosystem orchestration. The goal of network orchestration is the guided transformation of the ecosystem with continuous co-creation that allows the evolution of the processes needed to motivate and realize the transformation (Russell et al., 2011). This process evolution accommodates the complex influences on innovation in a networked world and energizes innovation processes and outcomes. Through the lens of the IETF, a shared vision of the transformational potential of a dynamic innovation ecosystem is created through changes in actors, the events that they enable, and the coalitions reflected in their relationships.

3.5.2 Data sources

After experimenting with in-house data on EIT ICT Labs activities, the investigative team, together with EIT ICT Labs representatives, decided to use the IEN Dataset instead. The use of an external data source enabled the exploration of existing pathways in between the EIT ICT Labs co-location cities previously unknown to the EIT ICT Labs operating team.

The IEN Dataset (Rubens et al., 2010) is used as the sole data source. The dataset for the EIT ICT Labs sample is drawn by selecting all the companies that have their primary office in one of the six co-location cities of EIT ICT Labs: Berlin, Eindhoven, Helsinki, Paris, Stockholm, and Trento. In addition, to collect data representing companies and related individuals and investors in the San Francisco Bay area, we list all the key cities located in the Silicon Valley area and complemented the list with San Francisco (including Berkeley).

3.5.3 Network modeling decisions

The six co-location cities, Paris, Berlin, Stockholm, Helsinki, Eindhoven, and Trento, form the core of the network. Each company in the sample is connected to the city in which its primary office is located. All key individuals (founders, board members, and C-level executives) in the dataset affiliated with one or more of the companies in the sample are connected to the companies. Next, financial organizations identified with funding events for those companies are added as nodes and connected to respective companies.

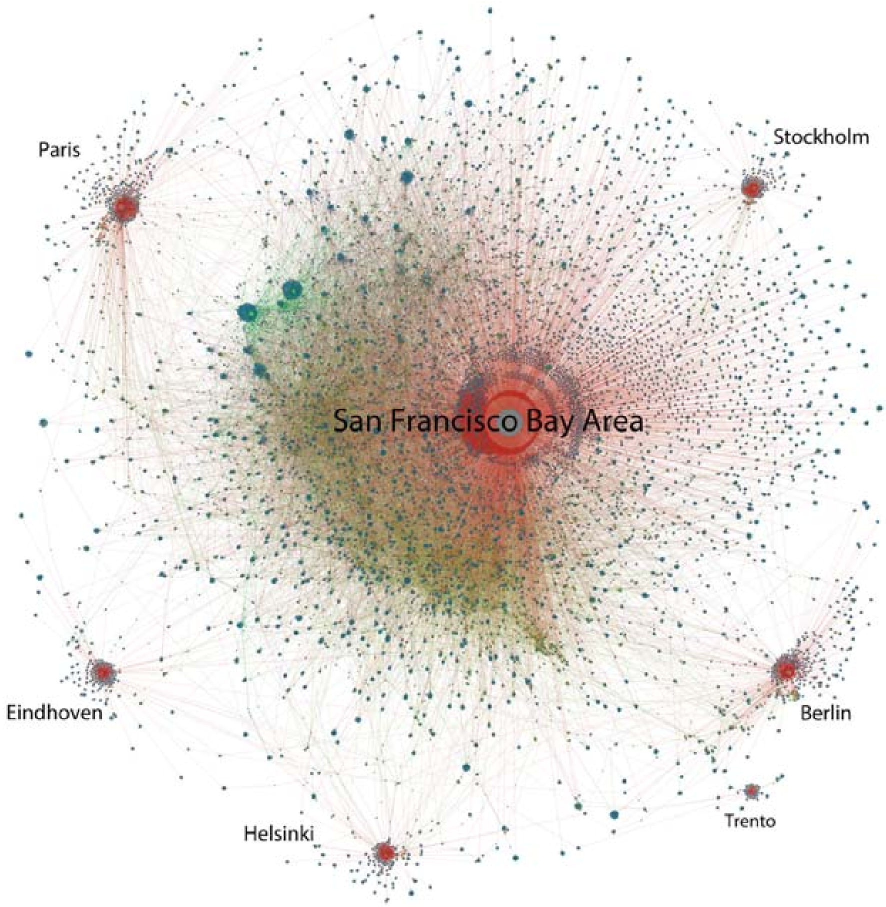


Figure 3.7: Scenario planning: What if the San Francisco Bay Area was the seventh EIT ICT Labs co-location city (Still, Huhtamäki, Russell, & Rubens, 2014)

The metrics used in the analysis include the number of each type of actor and the changes in the values over 2011, 2012, and 2013. Moreover, we measure the degree and betweenness of the six co-location cities and the change in the values between 2012 and 2013 to investigate the changes in the relative position of the co-location cities.

3.5.4 Results and network-related insights

The results of the investigation include changes in the numbers of actors connected to individual co-location cities, compared to our previous investigations, and a set of visualizations of the structure of the EIT ICT Labs ecosystems. Moreover, the role of individuals and investors in building interconnections between EIT ICT Labs co-location cities is shown.

The prominent role of investors as the connecting tissue between the individual co-location cities is a key result of the investigation. Moreover, the realization that many of these investors are, in fact, based in the Silicon Valley led us, in collaboration with EIT ICT Labs representatives, to ask a key what-if question: What if the San Francisco Bay

Area was the seventh EIT ICT Labs co-location city? Figure 3.7 shows that when the San Francisco Bay Area was added as the seventh co-location city and the network representation is laid out with force-driven algorithm, the San Francisco Bay Area became the focal point of the network through which most of the connections between the current co-location cities go through. With the addition of the San Francisco Bay Area, the number of nodes increased from 6,187 to 35,389, and the number of edges increased from 7,050 to 51,106.

3.6 Summarizing the investigations

We will conclude this chapter with a summary of the results and network-related insights into the investigated innovation ecosystems. Finally, we will discuss the utility and added value of the investigations.

For the investigation on Demola, we joined with the Demola operating team to explore ways to apply data-driven visual network analytics in representing the structure and dynamics of an ecosystem engager that is aiming to facilitate collaboration between (Tampere-based) universities and companies. In collaboration, through the process of guided emergence, we found that an animation of the evolution of the network structure of Demola platform was particularly useful in presenting, describing, promoting, and marketing the platform for existing and new stakeholders.

In the investigation on Tekes Young Innovative Companies, we explored the interconnections in-between companies taking part in the YIC program. Two sources of data were used to conduct the study: the IEN Dataset and Twitter. We showed that connections exist in-between the companies that were individually selected for participation in the YIC program. These interconnections were key in revealing the innovation ecosystem in which Tekes is interacting with through their support for individual companies. Moreover, the investigation makes a contribution with its use of social media data to give an ecosystem-level view into those interested in the companies. In the long term, should the Tekes YIC program be successful in selecting and supporting companies in growing, we would see a food chain of investors and acquirers emerge. Business angels, serial entrepreneurs−either active or successful−would also take a central role in this network representation of the innovation ecosystem around the YIC program.

In the investigation of the Finnish Innovation Ecosystem, we provided an ecosystem-level view of the structure of a national innovation ecosystem. The ecosystem representation includes established enterprises, growth companies, and startups as well as investors and key individuals affiliated with the companies. Specifically, we created four different representations of the ecosystem: microscopic, mesoscopic, macroscopic, and multiscopic. A handful of key individuals who entered the global startup ecosystem early and were successful in growing and selling a company maintain a prominent role in the Finnish Innovation Ecosystem. Nokia is visible through its role as a source of talent flowing into the ecosystem. Startup Sauna, a student-driven accelerator,[8] also has a notable role. The recently successful Rovio Entertainment and Supercell take a peripheral position; the presented approach does, however, enable monitoring the evolution of the ecosystem around them in the future. Our practical suggestions for startups include active communication and data sharing using a wide variety of media and, particularly for

[8]Startup Sauna accelerator is “Building a better startup ecosystem one company at a time”, `http://startupsauna.com/`

policy makers, the utilization of network views for targeted actions as well as for creating shared understanding and vision.

Lastly, in the investigation that we conducted in collaboration with EIT ICT Labs representatives, our results indicated that with the coordinated and continuously improved use of visual and quantitative social network analysis, special characteristics, significant actors, and connections in the innovation ecosystem can be revealed to develop novel insights. Creating a network representation of EIT ICT Labs, including the San Francisco Bay Area as the seventh node is an example of scenario planning that the approach proposed in this dissertation enables.

3.7 Utility and added value of the investigations

In summarizing the approach and key findings of our first investigation of EIT ICT Labs (Still, Russell, Huhtamäki, Turpeinen, & Rubens, 2011), Marko Turpeinen (2011), serving as the EIT ICT Labs Helsinki co-location center director at the time, pointed out the key role of mobility in EIT ICT Labs' attempt to turn Europe into a competitor equal to Silicon Valley. While EIT ICT Labs' key way of supporting mobility at the time was the European-level master school, the network representation of the interconnections between EIT ITC Labs co-location cities provided an interesting insight: the role of individual universities and particularly venture capital investors in bridging the co-location cities is very important. In revisiting the investigation on EIT ICT Labs (Still et al., 2012; Still, Huhtamäki, Russell, & Rubens, 2014), we joined with EIT ICT Labs representatives to ask a what-if question integral in scenario planning and, as a result, confirmed the bridging role of the San Francisco Bay Area.

In light of our investigations, we were excited to witness the news of, first, EIT ICT Labs opening up an office in San Francisco for "Building a bridge between the European ecosystem and the San Francisco Bay Area"[9] and, second, Marko Turpeinen's appointment as leader of the Silicon Valley Hub.[10] Even though the specifics of the decision-making process related to these two events remain unknown to us, we argue that the process related to the what if-scenario as well as the exploration of the structure of the existing ecosystem in general contributed and supported discussions and decision-making related to EIT ICT Labs' presence in Silicon Valley.

The Demola investigation was conducted during the time when knowledge of the Demola concept was only beginning to spread outside Tampere and Finland. Similar to the investigation of EIT ICT Labs, our collaborators in the Demola investigation knew the Demola operations by heart and therefore saw no value in exploring the structure of the Demola network for supporting their internal operations and decision-making. Instead, using network visualization, particularly the animation of the evolution of Demola platform as a network, was perceived very valuable in presenting, describing, promoting, and marketing the platform concept to existing and new stakeholders.

The investigation of the Finnish Innovation Ecosystem (Huhtamäki, Russell, Rubens, & Still, 2010) marks our first attempt to explore an innovation ecosystem with a data-driven visual network analytics approach. The alumni network study (Rubens et al., 2011) served

[9] EIT ICT Labs opens new Silicon Valley Hub, `http://eit.europa.eu/newsroom/eit-ict-labs-opens-new-silicon-valley-hub`

[10] Marko Turpeinen to lead EIT ICT Labs Silicon Valley Hub, `http://www.aalto.fi/en/current/news/2015-04-29-002/`

as an example of the approach in a related context. The follow-up study in **Publication V** is the first in which we used several sets of data both in parallel as well as in creating an aggregated dataset and a respective view into an innovation ecosystem. Overall, the two investigations of the Finnish innovation ecosystem supported introducing and experimenting with new features in the research process following the data-driven visual network analytics approach.

All the investigations included in this dissertation were conducted in Tekes-sponsored innovation research projects. Innovation ecosystem orchestrators and policy makers joined the projects to provide context for the investigations as well as to serve as project steering group members. We witnessed the applications of the presented approach being introduced into strategic foresight and impact assessment activities both within the Tekes programs and in the Council of Tampere Region.

4 Design principles for analyzing innovation ecosystems as networks

After reviewing the individual investigations conducted in this dissertation, we use this chapter to make explicit the first set of generalized outcomes of the research (Sein et al., 2011), that is, to synthesize the generalized outcomes of the experiments to develop guidelines for modeling and analyzing innovation ecosystems as networks, which contributes to **Objective II** of the dissertation.

To provide an overview of how the individual experiments contribute to the overall action design research process, we start the chapter by placing the experiments in a diagram of ADR stages and principles (Sein et al., 2011). Figure 4.1 summarizes the key principles of ADR in the context of this dissertation.

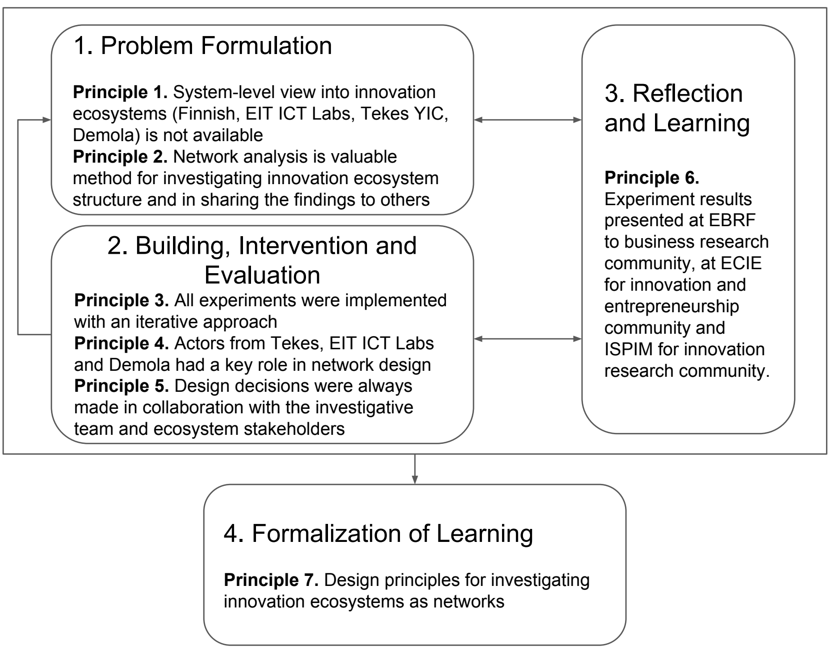


Figure 4.1: Action design research process for investigating innovation ecosystems as networks (following Sein et al., 2011)

4.1 Data for network analyses

The experiments included in this dissertation are based on two main sources of data: the Innovation Ecosystems Network Dataset (Rubens et al., 2010) and Thomson Reuters SDC. Thomson Reuters SDC is one of the most prominent sources of inter-firm relationships (Schilling, 2009). Thomson Reuters SDC is a prime example of a traditional company data sources from which start-ups and growth companies are often missing. The IEN Dataset provides socially curated (or crowd-sourced), rich data about startup and growth companies in almost real-time although with a public bias. In concert, the two sources of data provide means for creating several complementary views of innovation ecosystems. Following **Publication V**, the datasets represent different categories of actors in the innovation ecosystem and therefore enable the creation of both microscopic, mesoscopic, macroscopic, and multiscopic views of the innovation ecosystem (Still, Huhtamäki, Russell, Basole, et al., 2013).

Specifically, the IEN Dataset includes two subsets of data: Startups and Angels enable a microscopic view and Executives and Finance provide a mesoscopic view. In addition, Thomson Reuters SDC can be used to collect data on deals and alliances between the major enterprises that form the core of the existing business ecosystem, that is, the macroscopic view. The established enterprises acquire companies and act as sources of talent for startups and growth companies. Therefore, it is often important to include them in the network representations. While "alliance networks link the firm to a vast amount of knowledge, resources, and capabilities" and alliances can therefore serve as "access relationships" (Ritala & Hallikas, 2011) of potential importance for innovation, we want to stress that not all deals and alliances are related to innovation activities. Therefore, investigators should develop further the means to filter in only data that is relevant to innovation. We suggest that the use of machine learning methods should be investigated as a means to further develop automated data management. The fact that including a sufficiently rich context is imperative in order to provide a "wide lens" (Adner, 2012; Basole, 2014) to examine the ecosystem structure makes the exclusion of deals and alliance data a non-trivial task.

The investigation of the Finnish Innovation Ecosystem in **Publication V** is the first in which we create an aggregate set of data by combining the three different sets through actors that appear in more than one dataset. With the aggregated dataset, we provide an ecosystem-level view of the Finnish Innovation Ecosystem. In our investigations, we use a semi-manual process to aggregate the datasets and call for the means to fully automate the process by applying string matching (Navarro, 2001) and named entity recognition (Finkel, Grenager, & Manning, 2005).

In **Publication III**, we further use social media data from Twitter to investigate the structure and interconnections of the Twitter followers of the startups in the Tekes Young Innovative Companies program. The list of companies in the program is scraped from Tekes website, and the Twitter usernames were collected manually. In **Publication IV**, two complementary sources of data are used: the member list of the Finnish Venture Capital Association[1] and ArcticIndex,[2] a socially constructed dataset of Nordic and Baltic startups.

All of the aforementioned datasets originate in publicly available sources. Moreover, proprietary in-house data can be used to represent and analyze innovation ecosystems as

[1] Members of the FVCA, `http://www.fvca.fi/en/members`

[2] ArcticIndex is no longer active. It Was maintained by ArcticStartup, `http://arcticstartup.com/`

networks. The investigation of the Demola platform in **Publication I** is fully based on in-house data. The first attempt to create a network representation of EIT ICT Labs innovation ecosystem was based on their in-house data. We soon realized, however, that through their day-to-day operations, EIT ICT Labs actors know the structure emerging from the in-house data by heart. Therefore, we decided to use the IEN Dataset to investigate the already existing connections between the then six co-location cities to gain insight into the latent social structure of the ecosystem as well as mobility patterns within the co-location cities.

Table 4.1 presents a summary of the data sources used in each investigation.

Table 4.1: Details on collecting and processing data for the investigations

Experiment	Source	Collecting	Refinement
Demola	Project database internal to Demola	Tailored script for exporting the data from database in use in Drupal CMS	Unifying project category names
Mobile ecosystem	IEN Dataset, Thomson Reuters SDC (TR SDC)	Tailored script that traverses a NoSQL implementation of IEN Dataset proxy, TR SDC data imported as an Excel spreadsheet	Done as part of IEN Dataset and Thomson Reuters SDC curation
Tekes Young Innovative Companies	IEN Dataset, Twitter	Tailored scripts for traversing a NoSQL implementation of IEN Dataset proxy and accessing Twitter data through a REST API	Done as part of IEN Dataset curation
Finnish innovation ecosystem I	IEN Dataset	Tailored script that traverses a NoSQL implementation of IEN Dataset proxy	Done as part of IEN Dataset curation
Finnish Innovation Ecosystem II	IEN Dataset, TR SDC	Tailored script that traverses a NoSQL implementation of IEN Dataset proxy, TR SDC data imported as an Excel spreadsheet	Connecting companies across individual datasets through the creation of unified names
EIT ICT Labs	IEN Dataset	Tailored script that traverses a NoSQL implementation of IEN Dataset proxy	Done as part of IEN Dataset curation

4.2 Representing innovation ecosystems as networks

Visual analytics is a key source of requirements in designing the principles for investigating innovation ecosystems as networks. In the investigations in which we interacted with innovation ecosystem stakeholders and in the investigations that the research team conducted independently, we represented the innovation ecosystems as multimodal networks. Table 4.2 summarizes the network modeling decision in the individual investigations.

We do, however, realize that multimodal networks limit the possibilities for calculating node and network metrics to quantify the structural properties of the network and the structural roles of the individual actors to be utilized in statistical analysis. At the same time, we observed in the investigations that using a one-mode network representation of an ecosystem significantly reduced the complexity of the innovation ecosystem under investigation, therefore reducing transparency. Moreover, the one-mode network representation does not allow for truly ecosystem-level insights of the structural patterns of the innovation ecosystem. The use of one-mode and multimodal networks in concert to support processes that iterate exploration and specific measurement requires future development and experimentation.

Table 4.2: Network design details of experiments

Experiment	Boundary specification	Network modeling
Demola	Include the whole project database	Multimodal network of universities, projects, and companies. Connections are formed through project affiliation
Mobile ecosystem	Start from pairs of companies, include their first and second tier connections	One-mode networks for Thomson Reuters SDC, multimodal networks for IEND
Tekes Young Innovative Companies	Start from companies in Tekes YIC program, include their first and second tier connections	Multimodal network of companies, key individuals and investors
Finnish innovation ecosystem I	Start from companies in Finland, include their first tier connections	Multimodal network of companies, key individuals and investors
Finnish innovation ecosystem II	Start from companies in Finland, include their first tier connections	Microscopic, mesoscopic, macroscopic, and multiscopic view to Finnish Innovation ecosystem. Macroscopic view is non-directed one mode network, the others are non-directed multimodal networks
EIT ICT Labs	Start from the six EIT ICT Labs co-location cities, include companies having their primary office in one of the cities, include individuals and investors connected to the companies	Multimodal network of companies, key individuals, and investors

4.3 Analyzing innovation ecosystems as networks

Network analysis allows for measuring and analyzing the structural properties of innovation ecosystems at all three levels of analysis: ecosystem, relationship, and actor (Järvi & Kortelainen, 2016). At the ecosystem-level, the network metrics include density, diameter, and clustering coefficient. Edge weight is a key metric at the relationship level. In addition, detailed relationship properties can be included in network data and used in the analysis of dyads, that is, pairs of nodes. Node degree, indegree, outdegree, and betweenness are examples of actor-level metrics.

Table 4.3 summarizes the network and actor-level metrics that we applied in the different experiments. Because visual analytics are the focus of this dissertation, network metrics serve primarily to highlight actors in significant roles in an innovation ecosystem. Betweenness centrality is the most-often used actor-level metric. Edge weight is used throughout the investigations. Network metrics allow for comparing the structural properties of individual innovation ecosystems to each other. In this dissertation, network metrics describe the structure of innovation ecosystems. Moreover, network metrics are used in the ecosystem investigation to compare ecosystems with each other.

Table 4.3: Network design details for experiments

Experiment	Network metrics	Edge metrics	Node metrics
Demola	Not applied	Weight for the number of students representing a university	Betweenness for node's "connecting role in the network"
Mobile Ecosystem	Size, diameter, clustering, density	Weight for the number of deals and alliances between companies	Degree, betweenness, clustering coefficient
Tekes Young Innovative Companies	Size	Dichotomous weight	Betweenness to "to highlight the individuals, companies and investors that have an important connecting role in the network"
Finnish Innovation Ecosystem I	Size	Dichotomous weight	Degree, betweenness
Finnish Innovation Ecosystem II	Size, number of edges, density, diameter	Dichotomous weight	Betweenness
EIT ICT Labs	Size, number of edges	Dichotomous weight	Degree, betweenness "as a mobility factor to illuminate the potential of individual nodes to serve as bridges between the EIT ICT Labs co-locations"

4.4 Visualizing innovation ecosystems as networks

Visual network analysis is an organic part of the approach taken in this dissertation. To reiterate Freeman (2000), visual network analysis allows for two important tasks in investigating a phenomenon: first, it supports investigators in observing the social structures emergent in the empirical data representing a phenomenon under investigation and, second, it provides a tool for sharing the findings to others with representations that support co-referencing and the emergence of a shared understanding.

Table 4.4 summarizes the key visualization-related design decisions in the investigations. In all the investigations, we used a force-driven algorithm to lay out the nodes. Specifically, we applied the two variants of the force atlas algorithm implemented in Gephi (Bastian et al., 2009). A consistent color scheme for the nodes was established in the first investigation. The investigative team observed that using colors to identify node types−green for investors, red for companies, and blue for individuals−over time eases the visual investigation of networks as observers become familiar with the meaning of the different colors. We further note that the consistency of the network layout is an important objective in designing the analysis process when visual network analytics is used to investigate an innovation ecosystem over time. We will return to this topic in the context of the ostinato model in Chapter 6.

Table 4.4: Visualization details of the experiments

Experiment	Layout	Visual properties
Finnish Innovation Ecosystem I	Force-driven	Node color represents its type: red for companies, green for investors, and blue for individuals
Mobile Ecosystem	Force-driven	Node color represents its type: red for companies, green for investors, and blue for individuals
Tekes Young Innovative Companies	Force-driven	Node color represents its type: red for companies, green for investors, and blue for individuals
Finnish Innovation Ecosystem II	Force-driven	Node color represents its type: gold for Finnish companies, red for foreign companies, green for investors, and blue for individuals
Demola	Force-driven, dynamic when animating network evolution	For projects, node color represents its membership in a network cluster; company nodes are represented in light green
EIT ICT Labs	Force-driven	Node color represents its type: red for companies, green for investors, and blue for individuals

4.5 Design guidelines for network representation and analysis

Through the experiments reviewed in Chapter 3, we investigated several innovation ecosystems using the visual network analytics approach. Next, to generalize and make explicit the observations we made through the experiments and to support future efforts to model and represent innovation ecosystems as networks to support their investigation, we conclude the chapter with a set of design guidelines based on the results of the experiments to guide analysts and researchers in taking a data-driven visual network analytics approach to investigate innovation ecosystems.

4.5.1 Modeling innovation ecosystems as networks

The decisions made in network modeling represent the options for using different actor-level metrics in analysis. The directed one-mode network with weighted connections allows the use of the widest range of metrics. In all the investigations included in this dissertation, however, together with several ensembles of investigators, we decided to represent innovation ecosystems by using multimodal networks.

Guideline 1a: *Companies, key individuals, and investors are the basic entities in innovation ecosystem network representations.*

Guideline 1b: *Multimodal networks provide an intuitive starting point for representing innovation ecosystems as networks for an ecosystem-level view of its network structure.*

Guideline 1c: *One-mode networks enable detailed quantitative and statistical analysis with the expense of reducing the complexity of the ecosystem.*

Guideline 1d: *Edge direction and weight introduce additional means to utilize network metrics in supporting the analysis.*

Moreover, innovation ecosystem modelers should note that the innovation ecosystem literature also mentions customers and institutions as key actors in innovation ecosystems. In the context of innovation networks, Ritala and Huizingh (2014) list "customers, upstream supply chain partners, external knowledge providers, and competitors" as the key stakeholders. To generalize, modelers should conduct an inventory of relevant stakeholders when they begin a new innovation ecosystem investigation.

4.5.2 Analyzing the networks

In visual analytics, network metrics are used to highlight actors in different structural positions in the network. Node degree, indegree, and outdegree are the most simple metrics. Betweenness centrality is a useful metric in the context of multimodal networks. Hansen et al. (2011) refer to betweenness centrality as "Bridge Scores for Boundary Spanners." Moreover, additional network metrics become available when an innovation ecosystem is represented as a directed one-mode network.

Guideline 2a: *Node degree identifies actors with the largest amount of connections.*

Guideline 2b: *Outdegree identifies the most active actors.*

Guideline 2c: *Indegree is the simplest metric for prestige or authority.*

Guideline 2d: *Betweenness highlights nodes that have a bridging role in the network.*

4.5.3 Visualizing the networks

We applied a force-driven layout in all of the investigations included in this dissertation. A key reason is that the core investigative team remained the same throughout the investigations and continued to use the force-driven layout. Nevertheless, all the other stakeholders in the investigations also found the basic principle behind the force-driven layout to be intuitive. In our experience, the force-driven layout allows for the visual identification of network clusters and the actors bridging the clusters.

Visual consistency is important in supporting investigations where different network representation of the innovation ecosystem under investigation are created. We established a consistent color scheme for the innovation ecosystem actors. It is even more important to maintain the positions of the individual actors and clusters of actors in different representations, snapshots, and rounds of investigations. We discuss this in more detail when we describe the ostinato model in Chapter 6.

Filtering is a key approach used to reduce the complexity of the visual representations of innovation ecosystem networks. At best, an investigator analyzing the network representation of an innovation ecosystem is able to filter the network throughout the sensemaking process. To allow for expressive filtering, supporting data should be included into nodes and edges when creating the network.

Guideline 3a: *Force-driven network layout enables insights into the ecosystem-level network structure, key structural patterns as well as the structural roles of individual actors of the ecosystem.*

Guideline 3b: *Establishing a consistent and intuitive color scheme to differentiate node type is important. We use red for companies; green for finance organizations; and blue for key individuals (founders, C-level employees, board members, advisors).*

Guideline 3c: *Keeping network layout constant within and in between individual investigations supports investigators in establishing a mental model of the innovation ecosystem network representation.*

Guideline 3d: *Filtering enables revealing underlying patterns in the network under investigation.*

4.5.4 Investigating network evolution

There are two key approaches to investigating the development of a network structure, which is often referred to as network evolution (cf., Ahuja et al., 2012). First, the development of network metrics can be represented on a timeline. In **Publication II**, we used small multiple timelines, an application of small multiples (Heer & Shneiderman, 2012; Tufte, 1983), to represent the development of a number of networks to support their comparison and to gain insight into network evolution. In **Publication I**, we developed an animation of the evolution of Demola's innovation ecosystem as a network. We consider that supporting the investigations of network evolution through visual analytics provides a major venue for future research and development.

Guideline 4a: *Small multiple timelines provide insights on change in network and actor level metrics.*

Guideline 4b: *Network animation allows additional insights into the evolution of an innovation ecosystem.*

4.5.5 Interactive exploration

To reiterate, visual network analytics is a process (Heer & Shneiderman, 2012; Keim et al., 2010; Shneiderman, 2014). Therefore, supporting the processual nature of an investigation is imperative in selecting the tools for the analytics process. Gephi is the main tool used in all the investigations in this dissertation. In addition to network-centric tools, interactive tools from spreadsheet processors to business intelligence tools, including Tableau and others, can be used to explore network and node metrics. It is imperative to extend the ability to interact with data to upstream analysis, that is, to data transformations, boundary specification, and eventually data-collection routines. The ability to interact with the different steps of the data-driven visual network analytics process is at the core of the ostinato model.

Guideline 5a: *Both top-down and bottom-up analysis strategies are important. Support both "Start with what you know, then grow" (Heer & boyd, 2005) and "Overview first, details on demand" (Shneiderman, 1996).*

Guideline 5b: *Investigators should be able to experiment with the different metrics in defining visual properties of nodes.*

Guideline 5c: *Ability to filter nodes and edges on basis of parameters relevant to the investigation is imperative. Importantly, this calls for including node and edge properties that support filtering.*

Guideline 5d: *Investigators should be able to experiment with boundary specification.*

4.5.6 Sharing the findings

Interactivity is also a priority when selecting the tools for provisioning the visualizations to investigators and others interested in the findings. We used GEXF.js,[3] an interactive exploration tool running on a Web browser, for provisioning network visualizations.

Guideline 6: *Provisioning the outputs of the data-driven visual network analytics process with high-interaction visualization tools supports sensemaking and storytelling.*

[3] JavaScript GEXF Viewer for Gephi, `https://github.com/raphv/gexf-js`

5 Process model requirements

In the previous chapters of this dissertation, we focused on investigating innovation ecosystems as networks using a visual network analytics approach. In order to increase the level of architectural rigor of the analytics infrastructure and support automation in conducting the investigations, we will now discuss the process that is required to implement an investigation.

This chapter sets up the foundations for the development of a process model for data-driven visual network analytics. We draw from two complementing sources. First, we review existing work on data-driven visual network analytics and review existing related process models. Second, we identify a set of key requirements derived from the results of the investigations presented in Chapter 3.

5.1 Data-driven visual network analytics

Data-driven visual network analytics leverages computation to analyze potentially very large datasets in order to identify the structural patterns underlying a complex phenomenon. The investigations of such phenomenon are further complicated because data on the actors and their transactions often comes from multiple and diverse sources, some of which are not developed for computational use. Especially in cases involving data that is heterogeneous, an iterative, incremental analysis process is sometimes necessary (Telea, 2008). The analysis of complex phenomena often involves multiple pathways to conclusive insights, and actionable recommendations. Moreover, the assumptions underlying decisions may change over time.

We agree with Freeman (2000) that integrated tools that can be used to collect, manage and visualize the SNA data are key in supporting network investigations (cf., Huhtamäki, Salonen, Marttila, & Nykänen, 2010). The tradeoff between usability and automation sometimes creates a barrier for new entrants into data-driven visual network analysis (Hansen et al., 2012). However, a gap exists between the vision of easy-to-use integrated tools and the practice of data-driven visual network analytics. Data available for analysis is, for a number of reasons, notoriously difficult to process (Salonen et al., 2013). Individual investigators or small investigative teams often use manual processes or rely on ready-made tools that are operated through graphical user interfaces. Using these stand-alone tools is at best very straightforward. The available data sources and analysis and visualization functionalities are, however, somewhat limited. On the other end of the spectrum, the full-stack, programming-centric processes, in which massive sets of data are mined with tools that are developed and operated by experts, are generally run in complex cloud-based environments.

5.2 Review of existing process models

Several process models with different grades of abstraction exist to give structure to data-driven, visualization-centric investigations. In the next section, we will review a selection of the existing process models. The review follows **Publication VII** (Huhtamäki, Russell, Rubens, & Still, 2015).

Our approach to data-driven visual network analytics builds on several bodies of knowledge, including information visualization (Card, Mackinlay, & Shneiderman, 1999), data-driven visualization pipelines (Nykänen, Salonen, Haapaniemi, & Huhtamäki, 2008), interactive network analysis (Hansen et al., 2012), visual analytics (Keim et al., 2010; Wong & Thomas, 2004), sensemaking (Bendoly, 2016; Pirolli & Card, 2005; Weick et al., 2005), interactive visualization (Heer & Shneiderman, 2012), and scientific visualization (Telea, 2008). All these approaches introduce models and pose requirements that should be considered in developing next-generation tools and toolchains for visual network analytics. Moreover, the objective to conduct and publish research in a reproducible way (Ghosh, 2013; Peng, 2009, 2011) contributes to the overall quality of the analytics process and introduces further requirements.

To support the use of network analysis, Hansen et al. (2012) build on the sensemaking model (Pirolli & Card, 2005) to present the network analysis and visualization (NAV) model, a process model to support novices that enter network analysis. The NAV process starts by defining the goals of the analysis and continues through data collection and structuring, after which the data is interpreted through multiple loops of network visualization and SNA metrics calculation. Finally, the insights and conclusions are formatted, summarized, and then disseminated through a report. Seeking low-barrier entry, Hansen et al. (2011) introduce NodeXL, an Excel-based toolset for SNA, to conduct the analysis and define ways to use SNA in investigating phenomena in social media.

Card et al. (1999) present the information visualization reference model, a four-step process that can be used as a blueprint for implementing data-driven visualization processes. First, raw data is collected and then refined to data tables to allow straightforward processing. Third, the data tables are transformed into a portfolio of visual representations from which various concrete views are, fourth, provided to the visualization user for sensemaking. Imperatively, the reference model suggests that the best practice occurs when the user can interact with all steps of the process.

Component-based data-processing pipelines, a technical application of the information visualization reference model, introduce a viable approach for developing reusable pieces of software to support the automation of processes related to social network analysis across application domains (Huhtamäki, Salonen, et al., 2010; Nykänen et al., 2008). To support investigations of the social structure among wiki co-creators, for example, Huhtamäki, Salonen, et al. (2010) present a set of components and a process model for the orchestrated use of the components. A key benefit of the component-based approach (Nykänen et al., 2008) is that it is possible to integrate existing software tools implemented in different technologies into the data-processing pipeline, given that they can be operated from the command line. The main restriction of the approach is the need to implement the automation through scripting, that is, writing program code that describes rules for a particular functionality rather than operating a user interface.

The general sensemaking model (Pirolli & Card, 2005) divides the sensemaking process into two loops: the foraging loop and the sensemaking loop. To simplify, data is first collected and refined and then transformed into a selection of visualizations and other

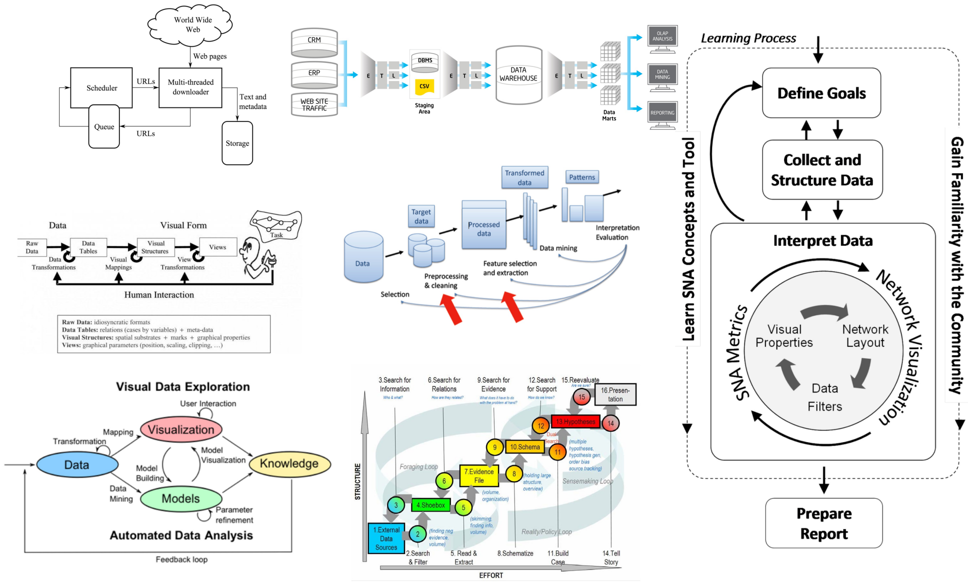


Figure 5.1: Process models related to data-driven visual network analytics. From the top left, the six small diagrams are Web crawling (Wikipedia.org, 2015), extract-transform-load (Intel, 2013), information visualization reference model (Card et al., 1999), knowledge extraction from databases (Indarto, 2013), visual analytics (Keim et al., 2010), sensemaking (Pirolli & Card, 2005). On the right: Network analysis and visualization (NAV) model (Hansen et al., 2012).

representations that support sensemaking. The process is reiterated as many times as required. Similarly, the process of visual analytics “typically progresses in an iterative process of view creation, exploration, and refinement” (Heer & Shneiderman, 2012).

The sensemaking part of the process can be implemented in different ways from purely manual processes where human investigators interact with various user interfaces to automated dashboard-centric information systems in which data are collected and processed in runtime. Sensemaking also includes the process of visual analytics (Keim et al., 2010; Wong & Thomas, 2004), which relies on the availability of software and tools supporting the users. Heer and Shneiderman (2012) provide an insightful overview of the specific function that users should be able to operate: 1) specify data and views, 2) manipulate views, and 3) process and provenance their findings. Sensemaking is indeed imperative when seeking true means to utilize visual analytics in management (Bendoly, 2016).

Peng (2009) gives three requirements for reproducibility: a piece of research is fully reproducible if both the data and code used are available and the code is executable by anyone. As Ghosh (2013) shows, reproducibility can be approached at many different levels from research policy to detailed technological solutions. Over the last few years, open research practices have become a priority in Finnish universities[1].

[1] Open Science and Research at TUT, `http://www.tut.fi/en/library/open-science-and-research`

To conclude the review, we note that many of the existing process models are either general or focus on particular parts of the visual network analytics process. For example, the use of parallel data sources is often not considered in the process models. A data-driven visual network analytics approach draws from a number of the presented process models. Moreover, network analytics introduces specific requirements in the process, importantly including the possibility of calculating node metrics to serve as additional data quantifying the different structural roles of the nodes.

5.3 Requirements from the series of investigations

Through the series of investigations included in this dissertation, we have shown that visual network analytics is a value-adding approach for exploring and investigating innovation ecosystems and sharing the findings to others. According to our experience, many of the process-related requirements stemming from the individual investigations are similar. At the same time, many of the investigation-specific analysis processes required that we tailor the process as we approached the final analysis.

By using the data-driven approach, the investigators of innovation ecosystems are able to move quickly at the beginning of the process. As the ways of visualizing and investigating a particular phenomenon mature, the investigators may wish to continue to follow the phenomenon with the support of close to real-time dashboards, thus adding transparency and supporting longitudinal investigations. The option of automating the process also supports developing these investigative tools toward end-user products to be used by avid innovation ecosystem actors, orchestrators, investigators, and policy makers.

The requirements that emerged through the investigations allow us to move toward the third and most important part of the results of this dissertation, namely the process model for data-driven visual network analytics. This will achieve **Objective III** of this dissertation. This section provides a synthesis of the requirements derived from the investigations presented in Chapter 3.

Developed through several rounds of iterations following the building-intervention-evaluation cycle (Sein et al., 2011), the core guidelines and requirements for the data-driven visual network analytics process model include the following: enabling manual steps; exploration; transparency; low entry barrier; interoperability; loose coupling; reproducibility; automation; and continuous data collection. In this dissertation, the requirements presented in this chapter and originally in **Publication VII** serve as a design rationale to support the definition of the process model for data-driven visual network analytics, that is, the ostinato model. Chapter 6 describes the model in detail. The next sections introduce each requirement.

5.3.1 Enabling manual steps

Although reproducibility is a key long-term objective, it is important to realize that automating some of the steps may not be feasible when an analysis is conducted the first time or requires intensive tailoring. Therefore, the process should support implementing any of the individual process steps manually. The use of file-based intermediary results is a practical solution that enables the manual steps of the analysis (cf., Huhtamäki, Russell, & Still, 2017).

-data/open-science-at-tut/index.htm

5.3.2 Exploration

The visual analytics (Heer & Shneiderman, 2012) approach is key in enabling users with varied technical skills to collaborate in exploring and making sense of a phenomenon. The ability to follow the visual analytics approach, however, requires flexible investigative tools and processes. That is, all the stakeholders in the analysis process should be able to conduct any of the individual steps by themselves even though the development of the overall process requires technical development skills.

In the investigations included in this dissertation, we relied extensively on using Gephi for exploring the networks. At best, however, the entire data processing pipeline from collection to refinement and transformation to visual representation also would allow the interaction of non-technical investigators.

5.3.3 Transparency

Developers with extensive technical skills may choose to manage the network analysis data throughout the analysis process by using a database. Graph databases in particular are appealing in conducting network investigations. To achieve transparency and flexibility in the process, however, other members of the investigative team will benefit from the option to access the data as files. The use of intermediary results is key in facilitating the transparency and flexibility of the process. Intermediary results refer to data in-between the individual steps of the analysis. This data should be available as files in widely used formats, including CSV and GEXF. In addition to the enhanced transparency, these intermediary results allow for speeding up the analysis process by using cached versions of source data and intermediary results when they have not changed.

5.3.4 Low entry barrier

The analysis of innovation ecosystems and other network-based investigations of complex phenomena require extensive domain knowledge, and hence require the active participation of domain experts (often without extensive technical expertise) throughout the analysis process. This requirement further underlines the need for transparency in the individual steps of the analysis process.

5.3.5 Interoperability

Despite the clear benefits of an integrated all-in-one tool for data-driven visual network analytics (cf., Freeman, 2000), there will always be individual tools that offer features not included in the all-in-one tool. Therefore, the investigative team should be able to use a number of existing analytics components as well as tools with high usability and rich interactivity, including Gephi, NodeXL, KNIME, and Tableau, for conducting the individual parts of the analysis. Moreover, provisioning the visualized networks and other outputs of the analysis should be possible through dashboard built with Web technologies such as D3.js, DC.js, and GEXF.js.

5.3.6 Loose coupling

At best, data-processing pipelines can be built with a range of tools and components implemented in different technologies. Loose coupling is key in enabling this kind of flexibility, which allows the introduction and use of new expressive tools from individual software components to full-featured applications as they become available to the investigative team. Many tools introduce new opportunities for advancing the analysis process, but generally it is not possible to integrate these tools in a data-processing framework at the program code level through the native application programming interfaces.

The data collection and pre-processing routines in the investigations described in Chapters 3 and 4 were implemented in Python using a collection of software modules for additional expressiveness. In addition, we used OpenRefine to clean the data, Gephi to lay out the networks and to calculate some of the network metrics, and NetworkX to calculate network metrics in Python for automation. For networks with 50,000+ nodes, we point to Snap.py.[2] The ability to use third-party routines to calculate different state-of-the-art network metrics is an example of the extendability from which the investigative team will benefit. Tableau allows visual analytics with a user interface and therefore serves as an example of a tool that many of the investigative teams will use due to loose coupling.

5.3.7 Reproducibility

In the data-driven visual network analytics approach, reproducibility is primarily a technical quality of the process: the investigative team should be able to repeat an investigation or one or more parts of the analytical process and reproduce the results. The reasons for the need to rerun the process include, among others, updates on the source data, development steps of the analysis process, the introduction of completely new processing steps, and new tools that insist on the use of a particular data format or a particular extension of the existing data. Moreover, dynamic sensemaking of complex phenomena necessitates the ability to refresh the data and derive new results with updated data. At the research collaboration level, reproducibility allows the investigative team to release the process, the data, and the results to other researchers interested in the phenomena under investigation in the spirit of open science.

5.3.8 Automation

The ability to develop automatically updating dashboards as needed gives the investigative team the opportunity to continue observing a particular phenomenon of interest over time. It is expected that production-ready analytical processes driving dashboards will operate without supervision; however, in the context of exploratory research, some requirements may be relaxed.

Automation is a key requirement in implementing a dashboard for an up-to-date ecosystem-level view of the structure of Demola, EIT Digital, the Tekes Young Innovative Companies program, or the Finnish innovation ecosystem.

5.3.9 Continuous data collection

Persistent processes for collecting data are often needed, particularly when the investigators wish to tap into social media to capture both the structure and structural dynamics of a phenomenon. Twitter, for example, currently provides only limited access to its historical data, but data on followers and friend connections between users do not include timestamps. The data collection sometimes takes weeks or "forever" to complete due to throttling or some other technical limitation (cf., Salonen et al., 2013) or the sheer size or the dynamic nature of the source data.

[2] Snap.py - SNAP for Python, `http://snap.stanford.edu/snappy/`

Although the collection process of the IEN Dataset (Rubens et al., 2010) falls outside the scope of this dissertation, we want to point out that IEN Dataset is an example of a data-collection process that has to run continuously in order to keep the views of innovation ecosystems up to date.

6 Ostinato model for data-driven visual network analytics

In this chapter, we describe the ostinato model, a process model for the data-driven visual network analytics of innovation ecosystems. The ostinato model achieves **Objective III** and represents the main contribution of this dissertation. The ostinato model was developed over the series of experiments presented in Chapter 3. The requirements identified through the investigations and existing process models (Chapter 5) were the key drivers in developing the ostinato model. The ostinato model was first presented in **Publication VII** in the context of social media studies. Here, we discuss the ostinato model in the context of investigating innovation ecosystems. This chapter is a revised version of a section of **Publication VII**.

In music, the word *ostinato* refers to both a repeating musical pattern as well as a composition that contains a repeating musical pattern. Similar to the repeating rhythms and melodies in Ravel's Boléro shown in Figure 6.1, or in post-rock,[1] small innovations are explored with each iteration, and some are incorporated into the melodic narrative. We apply the musical concept of ostinato to a cycle of user-centric exploration and automation that builds the transparency of authorship for evidence-based decision making.

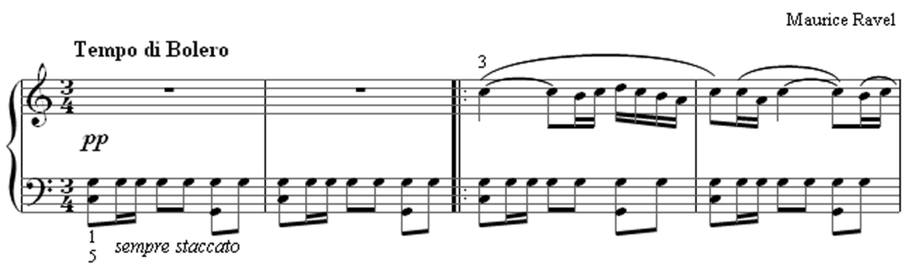


Figure 6.1: Ostinato patterns in Ravel's Boléro (Mawer, 2000)

In the ostinato model, the phenomena under investigation are modeled as a network, and interactive visualization tools are used to conduct the investigative process. Significantly, interaction is extended to all the different phases of the analytical process through transparency and the definition of explicit phases. Network analysis introduces a relationship approach to investigating the structure of many kinds of phenomena. Network analysis

[1]Categorizing music is debatable at best. However, music that falls under the post-rock category has been playing in the earphones of the author of this dissertation for hours and hours while conducting the investigations and writing this manuscript. For a sample, please refer to Magyar Posse. Paalanen (2015) presents inspiring work on supporting creativity by using repeating rhythmical structures.

allows for the exploratory analysis of the social roles of network actors and the complexity of relationships, as well as for the quantification of the structural properties of the network representation of the innovation ecosystem under investigation.

A key aspect of the ostinato model is the focal point of the user – in this dissertation, the investigator of an innovation ecosystem – in the investigative process. Putting the investigator at the center of the process answers the call for data scientists (Davenport, 2014), who are almost-mythical multi-skilled individuals capable of individually running the entire investigative process from the data collection and its analysis to deep sensemaking in the domain of interest, by allowing both experts of the domain under investigation, developers of the technical process as well as quantitative analysis specialists to have equal means to taking a proactive role in the investigative process. Moreover, the ostinato model defines the overall structure of the data-driven investigative process, which supports the coordination between the individual phases of the process and therefore allows all the members of the investigative team to contribute to the implementation of different phases of analysis and, importantly, to the sensemaking of the structures and mechanisms emergent in the empirical data representing an innovation ecosystem.

The ostinato model is developed over a series of experiments that investigated innovation ecosystems using an action design research approach. It is built on existing process models and the previous work presented in Figure 5.1, and it takes into account the process requirements presented in Chapter 5. Figure 6.2 shows a diagram of the ostinato model. Each step is described in the following sections.

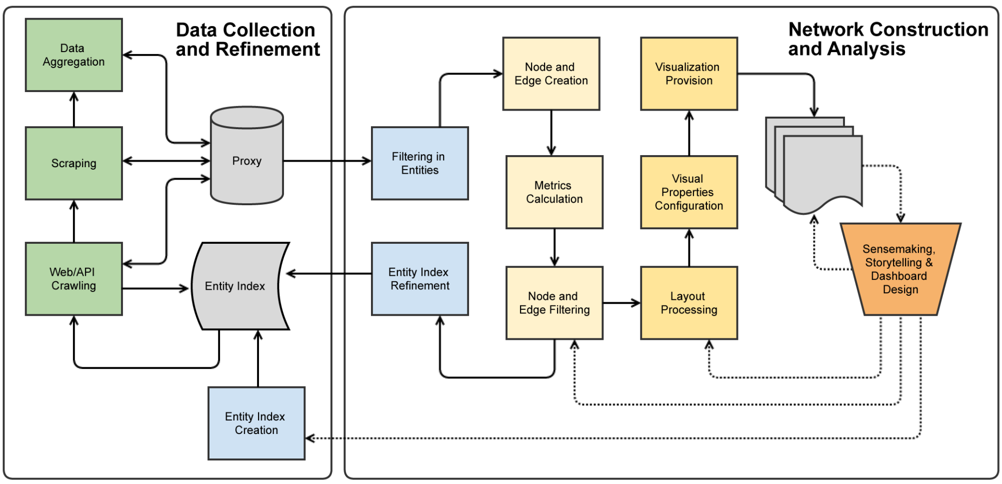


Figure 6.2: Ostinato model – a user-centric data-driven process model for visual network analytics

Phase 1: Data collection and refinement

1. Entity index creation
2. Web/API crawling
3. Scraping
4. Data aggregation

Phase 2: Network construction and visualization

1. Filtering in entities
2. Node and edge creation
3. Metrics calculation
4. Node and edge filtering
5. Entity index refinement
6. Layout processing
7. Visual properties configuration
8. Visualization provision
9. Sensemaking, storytelling, and dashboard design

6.1 Phase 1: Data collection and refinement

The general rules of data-driven analytics apply in implementing the ostinato model: collecting and cleaning the data will, in most investigations, consume most of the time and resources available for the investigation.

6.1.1 Entity index creation

In some investigations, the source data can be collected in full whereas in others only data on entities that are relevant for the analysis need to be collected. The entities for which data is collected are defined by boundary specification. In investigating the connections between companies taking part in the Tekes Young Innovative Companies program in **Publication III**, the list of companies defines the starting point of the analysis (Huhtamäki et al., 2012). In studying the structure emerging from individual deals and alliances around Google and Motorola Mobility, boundaries are set at two steps from the focal companies (Basole et al., 2012).

6.1.2 Web/API crawling

The data collection is the most heterogeneous step in the data-driven visual analytics process (cf., Salonen et al., 2013). Possible sources of data potentially include everything digital, from proprietary offline documents and document collections to spreadsheets, Web APIs, and Web sites that are designed primarily for human interaction.

Similarly, the functionality required to collect the source data can range from relatively simple reading of individual documents to functions often implemented in full-feature Web crawler. Compared to crawling random websites, Web APIs are, by default, more straightforward for data collection because they are often designed to support reuse (Vinoski, 2008). At best, source data is available as linked data (Bizer, Heath, & Berners-Lee, 2009), that is, data that has a clear structure and unique identifiers of individual facts. Linked data provides means to maintain referential integrity when data is integrated from complementary sources.

Thomson Reuters SDC provides a functionality for extracting data on alliances based on different search criteria, so crawling is not required. Crawling is, however, utilized extensively in collecting the IEN Dataset (see Rubens et al., 2010). Moreover, in **Publication III**, we use Twitter REST API to crawl the data of the Twitter followers of companies participating in Tekes YIC program.

At the end of the crawling phase, a set of web resources, or rather their representations in Hypertext Markup Language (HTML) or some other format, is made available in a local storage, a proxy that significantly speeds up the subsequent processing steps.

6.1.3 Scraping

When the raw source data is available locally, the next step is to filter, select, and distill the utility data relevant to the analysis process. Scraping refers to the process of distilling data from documents that are published on the Web for humans to use. This kind of data extraction and cleaning is sometimes referred to as data wrangling (Kandel et al., 2011). Scraping can further be considered a form of the extract, transform, and load (ETL) process, which is often applied in the context of data warehousing or other business intelligence processes to collect data from different sources to be refined, normalized, and finally loaded into a consistent database for later use (Petschulat, 2010; Vassiliadis, 2009).

The scraping function is required to distill the data from spreadsheets that are exported in Excel format from Thomson Reuters SDC. We use JSON to represent the data on individual deals and alliances. To support the access of non-technical investigators to the data, the use of CSV should be considered for representing intermediary results for added transparency and lowered entry barrier.

Using Wikipedia data in analyzing the structure of an innovation ecosystem is an example of scraping. When collecting data from Wikipedia on Finnish Young Innovative Companies, for example, the investigators were particularly interested in the facts presented in the Infobox section of the page (cf., Huhtamäki et al., 2012). To collect this data, the investigators took advantage of the HTML markup on the page to specify the semantics (meaning) of the different pieces of text. Each fact is represented as a table row including two cells, the first of which includes the label specifying the type of the fact and the second includes the actual value. Moreover, the value is also represented as a link to a separate page. These pages have to be included in the entity index for crawling and scraping additional facts relevant to the investigation.

6.1.4 Data aggregation

In contrast to data-driven social media studies in which data originates in an individual social media service, the complex context of innovation ecosystem investigations often requires on using several sets of data in parallel. This implies that in most investigations, linked data is not readily available and, therefore, links between individual sets of data have to be constructed through the creation of unique entity identifiers that allow referential integrity.

In innovation ecosystem investigations, the name of the company or another actor is sometimes the key data point that can be used to identify an entity. In **Publication V**, we used actors' names to find entities that appeared in more than one dataset. To take into account differences in the spelling of the names, we applied OpenRefine[2] to

[2]OpenRefine is an open source tool for working with messy data, `http://openrefine.org/`

harmonize the names through a semi-manual process. In determining whether co-author networks represent small-world properties, Newman (2001) uses author names with and without additional initials to create upper and lower bounds for network measurements. String matching (Navarro, 2001) and named entity recognition (Finkel et al., 2005) are examples of machine learning-based methods to support automation in creating unique identifiers for actors.

6.2 Phase 2: Network construction and analysis

When the data is available in a local proxy, the utility data is extracted from the source documents, and data from different sources is aggregated into a consistent set of linked data, the construction of the network representation of the innovation ecosystem under investigation can begin.

6.2.1 Filtering in entities

The network construction phase starts with a selection of the entities to be included in the network. The selection of nodes is guided by the boundary specification designed and defined by the investigative team. At least two approaches exist to implement the selection: starting from a list of entities and rule-based entity inclusion.

To continue the Finnish YIC example in **Publication III**, we started by compiling a list of companies participating in the program and continue to include all the individuals and investors directly connected with the company. Moreover, we included companies that are connected to the companies already in the sample through an investment or acquisition. For investigations on the Finnish innovation ecosystem and EIT ICT Labs in **Publication V** and **Publication VI**, respectively, companies were selected on basis of their location. In both investigations, directly connected companies, individuals, and investors were included in the sample.

The main reason for separating the selection of entities from node and edge construction is to support the transparency, reproducibility, and extensibility of the process. To create a shared understanding of the results of the analysis, it is vital that all the investigators taking part in a particular network investigation are able to access and understand the original raw data, in addition to any constructed variables, and the various analytics and metrics that represent the network, which means that investigation participants need access to all data, from raw to refined. According to our experience, answering specific questions raised by anyone interested in an investigation, drawing conclusions, generalizing the results, developing more specific and potentially more interesting questions all depend on the transparency of the data available and used in the analysis.

6.2.2 Node and edge creation

The creation of the network is a core part of the data-driven network analysis process. Network creation is based on the creation of nodes representing the actors and the creation of edges representing the connections between the actors. Several options are, however, available to specify details of the network creation process. First, the network can be either a one-node network or a two-node network. In one-mode networks, all the nodes are the same type: startup companies or investors, for example. Connections between the nodes are formed through relationships: investments, affiliations to individuals, acquisitions and transactions. In two-mode networks, there are two types of nodes, such as, startup companies and individuals related to them. Hypergraphs and bipartite graphs are examples of ways to visualize two-mode networks (Freeman, 2009; Jesus et al., 2009).

Further, the connections between network nodes can be either weighted or dichotomous. The strength of a connection can be expressed in weighted connections. In either case, the connections may be undirected or directed. Moreover, the temporal dimension can be included in networks if the data used to create the connections is time-stamped. As we showed in **Publication I** and **Publication II**, temporal data can yield insights into the evolution of the network.

In all the investigations included in this dissertation, we eventually decided to use multimodal networks for representing the innovation ecosystems under investigation. We assume that the main reason for this use was the exploratory, descriptive nature of the investigations, which places more importance on the ecosystem-level view than on the structural measurement of properties.

6.2.3 Metrics calculation

Network metrics enable the quantification of a variety of structural properties at both network and node levels. These range from simple metrics such as node degree (indegree, outdegree) (Freeman, 1978) and betweenness to PageRank (Page, Brin, Motwani, & Winograd, 1999), hub and authority values with HITS (Kleinberg, 1999), and other sophisticated measures. Whereas in principle, every metric can be calculated for all of the networks and their nodes, in practice this is not always feasible for reasons of efficiency. Moreover, new metrics for networks are being developed continually, and the investigative team is likely to find−or develop−new metrics that fulfill specific investigative purposes. From an implementation viewpoint, it is unlikely that one tool would be found to support all the metrics the team wishes to use. Therefore, a combination of tools may be required to calculate the metrics, which loose coupling enables.

Network metrics for the network representation should be stored for later usage. For transparency, a list of exported network nodes and edges should include the various metrics used. In practice, node and network metrics must be recalculated after each change in the network structure; however, reference to previous calculations is often needed, such as to analyze the change in the structural position of nodes.

The selected network structure dictates the metrics that can be calculated for the network and individual nodes. A one-mode network that has directed and weighted edges allows using the widest range of node and network metrics. In a multimodal network of companies, investors, and individuals, for example, metrics such as density or authority are not fully relevant because, for example, investors can never be connected to other investors if the connections are based on investments.

6.2.4 Nodes and edge filtering

A key limitation of visual network analysis is the amount of space available both on screen and particularly on paper, to present the visualization. Depending on the level of detail required in the analysis, hundreds or thousands of nodes can be presented in one visualization view. For networks of tens of thousands of nodes and more, only general structures and patterns can be observed in a visualization.

Two means exist to address this limitation: the best option is to allow the visualization users to filter in and out nodes and edges. If the end-user tools used to present the visualizations do not allow filtering, it can be done as one part of the automated process. Reducing the size of the visualized network is often accomplished by using a combination of filtering out edges that have the least amount of weight as well as filtering out nodes that 1) are left without edges; 2) have a value of the degree or some other a network analysis metric under a specified threshold; or 3) are (not) of particular type (even though this can already be taken into account when filtering in the entities used to construct the network in the first place).

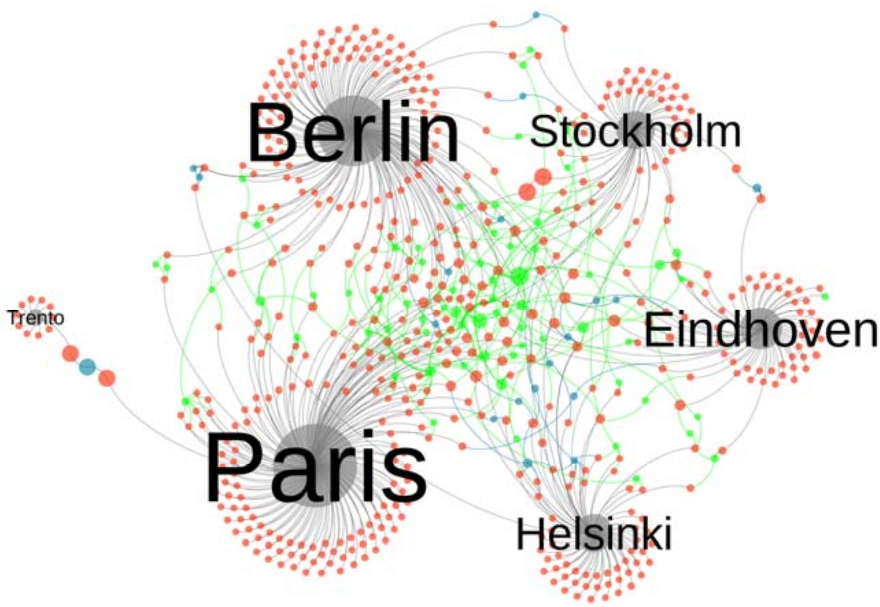


Figure 6.3: Top 10 percent of individual, companies and investors connecting EIT ICT Labs co-location cities according to their betweenness. The role of venture capital investors as enablers for mobility becomes evident.

In Figure 6.3 in **Publication VI**, we provide a filtered version of the network structure of EIT ICT Labs to highlight the importance of venture capital investors in connecting the different co-locations centers and therefore in enabling mobility in Europe. The nodes are filtered in according to their betweenness centrality.

6.2.5 Entity index refinement

At this stage of the process, the network representation of the innovation ecosystem under investigation is constructed, and the required metrics are calculated for each of the nodes. Depending on the boundary specification applied in the investigation, the network is either ready to be visualized or, alternatively, additional data can be collected to complement the network. In the Finnish Young Innovative Companies case examined in **Publication III**, the boundary specification is designed to include all the individuals involved in one or more of the companies in YIC program as well as all the other companies with which individuals are or have been affiliated. Moreover, the data includes all the investors that have invested in any of the companies as well as all the companies that have acquired any of the YIC companies.

Entity index refinement has a particularly important role when multiple sources of data are used. If, for example, the boundaries of the innovation ecosystem under investigation are set to two steps from the focal companies, the actors coming in through the first step should be taken into account across datasets.

6.2.6 Layout processing

The principle of processing the network layout is simple. Nodes are given a position in two-dimensional space such that the network structure is revealed in an expressive, intuitive way. Despite their simplicity, novel layout algorithms have been developed over several decades.

In the investigations included in this dissertation, various stakeholders found a specific implementation of force driven layout, force atlas, to be particularly suitable for laying out networks representing innovation ecosystems of different grades of abstraction and complexity. In fact, force atlas was used in all the investigations included in this dissertation. Force atlas is implemented in Gephi (Bastian et al., 2009) and can be used as a batch process with the help of the Gephi Toolkit.[3]

In practice, the parameters of the layout algorithm must be adjusted manually for a particular kind of a network before fully automating the layout processing. Alternatively, the layout can be processed with the user interface version of Gephi and the resulting network, including the X and Y coordinates for each node, can be exported as part of the network representation in GEXF or another suitable format.

Storing the network layout data is particularly important for improving the efficiency of the analytical process, as well as for reducing investigators' cognitive load and promoting transparency. In particular, it is important that after the data is refreshed, the investigators are able to find the pre-existing nodes in the area of the network where the nodes were previously located. This stability can be achieved by inserting the existing positions into the network data before re-running the force driven layout algorithm. In most cases, investigators will find the pre-existing nodes close to the initial area of the network.

Future work is needed to determine how features, such as layout algorithms implemented into NodeXL, could be used as a component of data-driven visual network analysis pipelines.

6.2.7 Visual properties configuration

There is a limited set of possibilities for defining the visual appearance of a network. Nodes have size, color, and perhaps a border and shape as elected visual features. Edges have color and width.

Both node and edge properties that originate in the source data as well as node metrics can be used to define the visual properties of nodes and edges. In our investigations, node size in most cases represented its betweenness centrality, and node color represented the type of the actor. At best, visual properties are defined as part of the automated pre-processing routines instead of selected manually in Gephi.

Allowing the user of the visualization to select and change the visual properties according to node metrics and other node properties is perhaps the easiest way to allow end user

[3] Gephi Toolkit, `http://gephi.github.io/toolkit/`

interactivity in network analysis. Depending on the tools used by the investigators to conduct the analysis, the visual properties of nodes and edges can continue to be tweaked as part of the interactive analysis process.

6.2.8 Visualization provision

In this stage, a network has all the required information available, and it therefore can be visualized. The means used to finalize this step depend greatly on the tools that have been selected for use by the investigative team. In most cases, however, the created network is serialized into a file following a selected vocabulary or format for representing a network. These vocabularies and formats range from different CSV-based applications to XML-based languages designed for representing networks.

A minimum approach to provisioning the network visualizations is to export network data in GEXF or other suitable format and place the resulting file into a folder that a software component such as Gexf.js can access. Generally, viewer composition scenarios can include the following:

Scenario 1. Network viewer component with fixed functionality, that is, following a fully descriptive approach. Visual properties, such as node size and color, need to be defined in the data during its processing. Gexf.js is an example of a component that we have found useful in adding value to a fully static PDF-based approach in disseminating network visualizations.

Scenario 2. Implementing a dashboard with Web technologies, specifically frameworks such as Highcharts, D3.js, Crossfilter.js, DC.js and others. In this case, tailored interactive features for data exploration can be provided to the user, thus adding options for representing network data.

Scenario 3. Using full-feature explorative analytics tools, such as Gephi, NodeXL and Tableau, which can be used to process the data further and to connect source data to visual properties of the visualization. The key here is to produce visualizations that are sufficiently rich in data to enable the analyst to utilize the critical properties of the chosen analytics tool for investigation and exploration. In Gephi, for example, it is useful to include attribute data for nodes to assist network filtering according to the intention of the investigator.

6.2.9 Sensemaking, storytelling and dashboard design

Although information visualization includes data transformation, representation, and interaction, it is ultimately about harnessing human visual perception capabilities to help identify trends, patterns, and outliers. Sensemaking is rooted in cognitive psychology, and many different models have been developed. Sensemaking procedures are cyclic and interactive, involving both discovery and creation (North, 2006). During the data collection and refinement phase, an investigator searches for representations. In the network generation phase, these representations are instantiated, and based on these insights the representation may be shifted, to begin the process again. Sensemaking is closely linked to the insight objectives (Konno, Nonaka, & Ogilvy, 2014), and the ostinato cycle of exploration−automation is key in supporting sensemaking practices that are required in achieving actionable insights that innovation ecosystem investigators, analysts, and orchestrators can utilize (cf., Bendoly, 2016).

When the sensemaking requirements of investigators and other users are satisfied, the steps in the ostinato process can be formalized with automated procedures for iteration over time. Key actors, relationships and events of the network can be incorporated into dashboards that will track changes in critical assumptions and into stories that will share visions of actionable change.

6.3 Utility and added value of the Ostinato Model

In this dissertation, the ostinato model, a new process of data-driven visual network analytics was developed and described. The ostinato model contributes to the call for more expressive means for supporting innovation ecosystem investigations in two ways. First, it can be applied to support the data-driven investigations of innovation ecosystem structure and dynamics. Second, to ensure the validity and reliability of these investigations, it is vital to increase the transparency of the processes behind the data originating in various digital sources.

Moreover, the ostinato model contributes to the data-driven network investigations of innovation ecosystems in three different ways. First, the network approach has great strength in supporting the exploratory investigations of the patterns in-between actors of innovation ecosystems. Second, with specific reference to the first phase of the ostinato model, the data-driven approach allows tracking processes over the boundaries of individual sources of innovation ecosystem data. Third, the user-centricity of the data-driven process adds to the transparency of the process itself, therefore providing a means to triangulate different phases of data refinement and transformation and allowing different stakeholders in investigations to take as proactive role as they wish in moving a particular investigative process forward.

Because of the continued and rising interest in big data analysis, new tools are continually introduced to support investigative work. Despite the development of all-in-one tools, a combination of tools is likely to continue to provide more flexibility in accessing and aggregating data and in processing and analyzing such data. Finding a balance between user interface-operated low barrier tools and expressive computational strategies that require technical knowledge is key in making the investigative process as productive as possible while maintaining transparency and process flexibility.

The proposed ostinato model for user-centric, process-automated, data-driven visual network analytics meets many of the requirements outlined in Section 5.3 for the exploration−automation cycle recommended for developing shared understanding.

Using files rather than databases for representing intermediary results supports both loose coupling and transparency of the process. It also allows for implementing some of the steps manually, if seen feasible, and the flexibility of the process in general is increased.

Allowing exploration is based on the selection of the end user tools for investigators to visualize and explore the data. If a rather static tool, such as Gexf.js, is used, the user is limited to browsing and searching the data. If importing the data into an exploration platform, such as Gephi or NodeXL, is permitted, it is possible to provide the users with rich node and edge data, enabling them to continue their explorations with more independence. The availability of expressive visual analytics tools, such as Tableau, adds to investigation options of analyzing network data, either as a network or using node and edge level data to provide new inspirations for other kinds of data analyses.

The low-entry barrier is enabled through making the intermediary results available to all the members of the investigative team. Because the process is repeatable and its individual steps are automated, new projections of the data can be implemented in an iterative and incremental manner. Implementing completely new steps of analysis becomes possible even without technical skills. Automating the steps, however, requires the developers' attention. The ostinato model requires a multidisciplinary data science team, or a multi-skilled data scientist (cf. Davenport, 2014), to conduct the investigation.

Interoperability can be built into the computational approach. This requires that the technical architecture is flexible enough to permit different software components and tools, which may be implemented with different technologies, to be introduced into the process. When an analytical pipeline is built completely from scratch, it is recognizably important to minimize the number of technologies used. However, moving fast and in an agile manner is an objective that we claim can be achieved when existing tools can be integrated into the pipeline to implement the individual steps of the analysis process and to provide the visualizations to investigators and other end users.

Reproducibility is both a technical and a policy requirement. For an investigative team revisiting or extending an existing investigation, the availability of runnable code, source data, and intermediary results provides a fruitful starting point. Moreover, the results of reproducible studies can be published in a way that both data and runnable code are available, providing a solid foundation for others to add their contributions. A reasonable proposition is that such knowledge attracts the attention of other researchers and therefore has increased potential for impact. Automation is a key requirement in reproducibility, as well as in creating dashboards that continues to update visualizations of the phenomenon under investigation, sometimes close to real time.

Setting up persistent data-collecting routines requires, in general, a programmatic implementation and must be designed and implemented case by case. To maintain the transparency of the process, it is important that the investigators are able to access both the raw data as well as to track down the individual steps used to derive the data that is eventually used for the analysis and visualizations.

Generally, implementation of the ostinato model can serve as the core engine of an investigation. It can also be used to develop a pre-processing pipeline that collects and refines the data, creates a network representation, and serializes the outputs to be analyzed and processed with expressive tools that, standing alone, allow the full visual analytics cycle for users.

A key challenge of the presented approach concerns the number of options for investigators and other end users to interact with the data in real-time while conducting the analysis, particularly the non-technical investigators on a multi-disciplinary team. The action design research approach favors an iterative approach to both data-driven explorations and evidence-based decision making. However, investigators with limited programming skills or technical know-how are restricted in their participation, even though they may possess vital domain intelligence. Through access to data, documentation of changes in the analytical approach, flexible means to produce network representations in various formats, and exposition of intermediary results, barriers to participation are lowered. The cycle of exploratory visual analytics, confirmation of data selection rules, and analytical results made accessible through high interactivity visual analytics, allows the investigative team to confirm assumptions and investigative procedures, identify aspects of the analysis that can be automated, and establish a transparent, reproducible process.

To conclude, we claim that the ostinato model is a necessary prerequisite to the use of visual analytics in innovation ecosystem analytics and orchestration following enacted sensemaking Bendoly (2016). We look forward to further research on the use of enacted sensemaking in innovation ecosystem orchestration.

7 Discussion

The objective of this dissertation was to explore the ways network analytics should be applied to investigate the structural properties of innovation ecosystems at the ecosystem level. Using a data-driven approach to conduct the investigations was set as the key design criteria. That is, it should be possible to collect and aggregate data from various heterogeneous sources in an automated fashion to allow reproducible analysis. Specifically, three objectives were set. First, **Objective I** was to contribute to the empirical body of knowledge by running a series of investigations on the ecosystem-level network structure of innovation ecosystems representing different grades of abstraction and complexity. Second, **Objective II** was to develop design guidelines on how to model innovation ecosystems as networks for visual analytics. Third and most importantly, **Objective III** was to design a general process model for the data-driven visual analytics of innovation ecosystems.

To reach these objectives, an action design research approach was taken to conduct research in two complementary streams. First, a series of investigations of innovation ecosystems was conducted to gain knowledge about the ways investigators, innovation ecosystem actors, and stakeholders prefer to model innovation ecosystems as networks. Second, a set of requirements was derived from the experiments to support the design of a general process model for the data-driven visual analytics of innovation ecosystems as networks. Third and most importantly, the ostinato model was developed through aggregating and extending existing process models in a way that the requirements specific to data-driven visual network analytics of innovation ecosystems could be met.

Gregor and Hevner (2013) suggest, “with socio-technical artifacts in IS, when the design is complex in terms of the size of the artifact and the number of components (social and technical), then explicit extraction of design principles” may be included in the discussion section of a scientific publication. We have already introduced two key sets of design principles. First, in Chapter 4, we enumerated several design principles for modeling innovation ecosystems as networks for their visual investigation. Second, in Chapter 6, we described the ostinato model for the data-driven visual analytics of the network structure of innovation ecosystems.

In the following sections, we will discuss and review the key contributions of this dissertation to give evidence that we have successfully bridged the identified the research gap that was the starting point for this dissertation work.

7.1 Empirical investigation on innovation ecosystems

The investigations contributing to **Objective I** are described in detail in Chapter 3. The research in this dissertation was conducted by a multidisciplinary, global team of researchers with backgrounds in business, policy making, and academia. The members of the research team posses expertise and real-life experience in knowledge management, innovation ecosystem orchestration, innovation research, information system design, machine learning, and network science among other domains. The investigations included in this dissertation explore and describe the innovation ecosystems in a novel way. Through the individual investigations, we provided new insights into the ecosystem-level network structure innovation ecosystems representing different grades of abstraction and complexity.

Specifically, investigations of innovation ecosystems took place in five different contexts of varying grades of abstraction and complexity. The contexts include the innovation platform, business domain, innovation program, national ecosystem, and international ecosystem.

Demola is an innovation platform. We collaborated with Demola operating team to explore ways to use data-driven visual network analytics in representing the structure and dynamics of an ecosystem engager that is aiming at facilitating the collaboration between (Tampere-based) universities and companies. Through guided emergence, we found that animating the evolution of the network structure of the Demola platform was a particularly useful approach to present, describe, market, and support sales work of the platform to existing and new stakeholders.

The mobile ecosystem represents a business domain. We investigated the key pairs of actors in the mobile ecosystem. During the time of the investigation, Nokia and Microsoft had just recently announced a strategic alliance to work together in developing their mobile offering. Google acquired Motorola Mobility in August 2011 to strengthen its capabilities in the mobile domain. The investigation highlights the importance of data triangulation for covering the different aspects of innovation ecosystems. Using Thomson Reuters SDC Platinum, the standard data source in strategy research, Microsoft was the supernode even in the network representation of actor network surrounding Google and Motorola Mobility. Only when IEN Dataset – a compilation of socially constructed set of data on innovation activities – was used did Google's true size that the company has accumulated through a series of acquisitions, and the flow of talented individuals became visible.

Tekes Young Innovative Companies is an example of an innovation program. Here, we investigated the interconnections between startups taking part in Tekes YIC program. Two sources of data were used to conduct the study: IEN Dataset and Twitter. We showed that connections existed in-between the companies that were individually selected to participate in the YIC program. This finding highlights the importance of ecosystem-level analysis in giving context and therefore in supporting decision-making. In the long term, should the YIC program be successful in selecting and supporting companies in their growth, we would see the emergence of a "food chain" of investors and acquirers. Business angels, serial entrepreneurs – either active or successful – would also take a key role in such a network representation of the innovation ecosystem around YIC. Moreover, we used social media data to derive an ecosystem-level view of those interested in the companies.

The Finnish innovation ecosystem is an example of a national innovation ecosystem. The ecosystem-level view of the Finnish innovation ecosystem includes established enterprises, growth companies, and startups as well as investors and key individuals affiliated with the companies. Specifically, we created four different representations of the ecosystem: microscopic, mesoscopic, macroscopic, and multiscopic. The results showed that a handful of key individuals who entered the global startup ecosystem early and were successful in growing and selling a company have a prominent role in the Finnish innovation ecosystem.

Nokia is visible through its current role as a source of talent flowing into the ecosystem. Startup Sauna, a student-based initiative for supporting startup creation, also has a notable role. The recent successes of Rovio Entertainment and Supercell take a peripheral position; the presented approach will enable monitoring the evolution of the ecosystem around them in the future. Our practical suggestions for national ecosystem actors include active communication, data sharing using a wide variety of media, and utilizing network views for targeted actions as well as for creating shared understanding and vision.

Finally, EIT ICT Labs is an example of a large-scale international innovation ecosystem. In this investigation, we explored ways of using visual network analytics to gain insights into existing network structure across EIT ICT Labs co-location cities. Our results indicate that with the coordinated and continuously improved use of visual and quantitative social network analysis, special characteristics, significant actors and connections in the innovation ecosystem can be revealed to develop new insights. Creating a network representation of EIT ICT Labs, including San Francisco Bay Area as the hypothetical seventh node, is an example of scenario planning enabled by the approach developed in this dissertation. Our results showed the value of socially constructed data in gaining insights on the structure of a broad-based international innovation ecosystem.

A key part of our contribution to the empirical literature is the introduction of novel sources of data on innovation ecosystems. Several datasets were used in the investigations, including social media, socially constructed data available online, and proprietary sets of data represented as spreadsheets and other formats. In most of the investigations, we used data sources that were external to the focal organization of the innovation ecosystem under investigation. However, for Demola, the least abstract and least complex of the investigations, we decided to use the project data set that the Demola team maintains internally. For the other investigations, we used two main sources of data: Innovation Ecosystems Network Dataset and Thomson Reuters SDC Platinum. In addition, Twitter data was used in one of the investigations. Moreover, in the first investigation of the Finnish Innovation Ecosystem in **Publication IV**, we aggregated data collected from several additional sources. The data sources are covered in detail in Section 4.1.

In all the investigations included in this dissertation, we took advantage of new sources of data from socially constructed to institutionally constructed to data accumulated through day-to-day operations, which Williams and Shepherd (2015) refer to as archival data. The experiments therefore contribute to organizational research in showing evidence of the usefulness of archival records that are observed to be novel in organizational research literature (Williams & Shepherd, 2015): “despite the possible benefits, secondary data are rarely used in organizational social network studies and are almost never considered from a qualitative perspective.”

To conclude, we claim that in-house data is particularly useful in supporting an organization in its operations to others in the context of marketing and public relations. In contrast, data that originate in external sources allows for new viewpoints and novel insights into the structure of innovation ecosystems and the roles of individual actors in them.

7.2 Investigating innovation ecosystems as networks

Investigations of innovation ecosystems can take place in three different levels of analysis: actor, relationship, and ecosystem (Järvi & Kortelainen, 2016). The majority of existing investigations on innovation ecosystems focus on either individual firms or pairs of firms and their relationship (dyads). Our review of the extant literature literature on network structure of innovation ecosystems in Section 2.2 showed that only a limited number of ecosystem-level investigations exist. The specific viewpoint that we take in this section, the approach for investigating innovation ecosystems as networks, is in most empirical studies outside the scope of the publication and instead appears in research design and the exploratory phases of the investigative process.

The investigations included in this dissertation provide evidence that representing and analyzing innovation ecosystems as networks adds value to two interrelated fields: the scholarly investigation of innovation ecosystems and innovation ecosystem analytics. In Chapter 4, we presented a set of design guidelines to support network investigations of the innovation ecosystem structure and its evolution and the structural roles of individual actors in the ecosystem. These guidelines provide a basis for the consistent analysis of innovation ecosystems in the aforementioned fields.

In the investigations, the main finding was that both innovation ecosystem stakeholders and academic co-investigators preferred to model the innovation ecosystems as multimodal networks that included key organizational investors and key individuals in addition to companies. In a visual analysis, this yields insights into the overall ecosystem-level network structure of an innovation ecosystem because all the actors appearing in the data are also present in the visualizations. Multimodal networks are, however, not an optimal starting point for the quantitative analyses. The key reason is the fact that often−in all the investigations included in this dissertation, for example−the possible connections between actors are limited: in our investigations, venture capital investors were only connected to companies, not to each other. The same restriction applies to individuals. This means that network metrics, such as density and node metrics, which take into account the larger network structure, such as Page Rank, HITS, and eigenvector centrality, provide only limited value for analysis compared to situation where all the actors represented as nodes in the network can be connected with each other through directed nodes. Because of the limited connectivity between actors, a network can never be fully connected. Therefore, density value has to be interpreted with particular care. Similarly, metrics such as Page Rank and HITS, which consider both the direction of connections as well as their context−for example, in Page Rank the authority of a node referring to another node is considered in calculating the authority value for the referred node−are of limited utility in partially connected multimodal networks.

Betweenness centrality was perceived to be a particularly useful metric in the investigations. There are several reasons for its utility. To begin, betweenness can be calculated in undirected networks as well as in networks with limited connectivity in-between modes of nodes. Moreover, betweenness does take into account the relative position of a node in connecting different parts of the ecosystem. Perhaps most importantly, the principle for calculating betweenness centrality value is relatively straightforward to comprehend, yet it takes into account the network-level structural position of an actor. It should be noted, however, that betweenness centrality is particularly prone to errors in the data; therefore, investigators should be able to use a set of network metrics for comparison and context as well as to interact with network construction, filter parameters, and boundary specification.

From the viewpoint of using visual analytics as means to supporting sensemaking for management (Bendoly, 2016), we contribute to the development of interactive visualization tools that are developed to investigate the ecosystem-level network structure of innovation ecosystems. This process encourages the voicing of different perspectives that can “give rise to subsequent visual consideration, more targeted evaluations, additional requests to redevelop and enhance the visual platforms that the organization leverage" (Bendoly, 2017) and facilitates rich dialogue on those perspectives. This is imperative in applying the results of this dissertation in sensemaking. Further, we make a significant contribution to the development in applying network as a visual and modeling idiom, an instance of the construction trait, for empirical innovation ecosystem investigations (Bendoly, 2016).

Finally, we would like to point out the steps we have begun to take toward the formal validation of the utility and usefulness of visual network analytics of innovation ecosystems. Basole et al. (2016) presents the results of an experiment in which three different representation of network data were used to support a collection of decision-making task. Moreover, Russell, Still, and Huhtamäki (2015) used a more qualitative, descriptive, and reflective approach to describe the ways the network approach contributes to decision-making in the context of innovation ecosystems.

7.3 Main result: Ostinato model

*This section is based on **Publication VII**.*

The ostinato model was developed and validated in multiple investigations serving as experiments that followed action design research. The ostinato model has two main phases, data collection and refinement, and network creation and analysis. The data collection and refinement step is divided into entity index creation, Web/API crawling, scraping, and data aggregation. The network creation and analysis step is composed of filtering in entities, node and edge creation, metrics calculation, node and edge filtering, entity index refinement, layout processing, and visual properties configuration. In the final step, the visualizations are provided to investigators and other end users. The results of sensemaking and feedback activates an iteration of the process. A cycle of exploration and automation is embedded in each phase of the model.

The ostinato model allows both the exploratory approach during the early phases of the investigation and the automation of the data collection and analysis process when the investigative routines gain maturity. The iteration cycle is especially beneficial in working with multi-source datasets, complex phenomena, and changing externalities that may impact assumptions for decisions, and in establishing a dashboard for continued observation of the phenomenon, perhaps in real time.

The ostinato model has several operational implications for investigative teams that adopt the data-driven visual network analytics approach. These implications are described in the following paragraphs.

First, the facilitation and documentation of the investigative process are required. Low-barrier entry in exploration and analysis poses risks that increase when transparency is not present. In other words, through transparency, intermediate results, and easy access, the risk of false conclusions is decreased. Co-ordinated discussion on raw data and its journey to the finalized visualizations and other results is imperative. Documentation of assumptions and rationale for changing data selection or analytical procedures enables transparency. Facilitation also helps in creating literacy of the processes and its outputs within the investigative team. When the intermediate results are available, all the members of the investigative team are able to maintain control of the process and continue to introduce new, novel ways of analyzing the data according to their skills and methodological expertise.

Second, the cycle of exploration-automation introduces new requirements for the governance of both the data and the analysis process. Intermediary results require transparent authorship in their provenance. The transparent authorship of new datasets, constructed variables, and analytical iterations must be ensured.

Third, starting with exploration and moving toward automation is supported with the help of the ostinato model. The investigative team is able to move quickly in the beginning of the process while maintaining control over the process as its complexity increases. With appropriate technology selections, the process can eventually be relegated to the background to collect, process, analyze, and visualize data in an automated manner to support a longitudinal study of a particular innovation ecosystem. Significantly, a mature procedure−or one or more of its components−can be reused to investigate other innovation ecosystems of interest.

Fourth, increased reproducibility is an asset for future investigations, but it requires explicit governance. Technical reproducibility of the process allows revisiting the analytical results of an investigation even after a long time period. Refreshing and collecting new data or, alternatively, adding new dimensions to existing data is straightforward when the process or its individual parts can be run computationally. Rules must be developed for data curation, and the access to code and data has to be designed at both the technical and policy levels. Governance of the data from rawness to intermediate results to outputs as well as the components and software process must be articulated clearly.

The ostinato model provides blueprints for designing analytical processes with technologies ranging from Python to R and even Javascript. At best, the process is able to support the inclusion of several different technologies in a similar manner than the Wille Visualisation System (Nykänen et al., 2008).

7.4 Support for analytics

After discussing the outcomes of the dissertation in terms of the three key objectives, we will next continue to discuss broader issues related to innovation ecosystem analytics. One important aspect is the support for analytics that the approach proposed in this dissertation provides. In both research and analytics, several approaches are used to investigate a phenomenon. One taxonomy categorizes the approaches into exploratory, diagnostic, descriptive, predictive, and prescriptive (Bendoly, 2016; Davenport, 2013). In the individual investigations, our approach was exploratory and descriptive rather than predictive or prescriptive. It is imperative to develop ways to facilitate balanced discussion between members of the investigative team with a multidisciplinary members (cf. Bendoly, 2016; Nunamaker & Briggs, 2011; Pentland, 2015). The data-driven visual network analytics approach presented in this dissertation is, we claim, a major step in that direction. At the same time, we point to the research gap in using visual analytics in innovation ecosystem orchestration.

Through the transparent creation of visualizations of the innovation ecosystems representing different grades of abstraction and complexity from platforms to development programs to national and international ecosystems, our research does “map the terrain of a specific phenomenon”[1] and therefore takes steps from exploratory to descriptive

[1] This nicely formulated phrase originates from an online note “Research Methods: Some Notes to Orient You,” see `http://isites.harvard.edu/fs/docs/icb.topic851950.files/Research%20Methods_Some%20Notes.pdf`

research. We realize that with the capability to predict, results could be achieved to help in developing new knowledge on how the world works. At the same time, we see that, as an object of research, innovation ecosystems are either complex or chaotic rather than known or knowable (cf., Kurtz & Snowden, 2003). Therefore, we subscribe to the argument that Kurtz and Snowden (2003) use to select their approach for developing Cynefin, a sensemaking framework for supporting real-life decision-making and policy-making:

> *We consider Cynefin a sense-making framework, which means that its value is not so much in logical arguments or empirical verifications as in its effect on the sense-making and decision-making capabilities of those who use it. We have found that it gives decision makers powerful new constructs that they can use to make sense of a wide range of unspecified problems. It also helps people to break out of old ways of thinking and to consider intractable problems in new ways.* (Kurtz & Snowden, 2003)

In line with the methodological and philosophical foundations of this dissertation, which are described in Section 1.2, we consider that the ostinato model supports innovation ecosystem investigations that are conducted by those with a critical realist worldview. Specifically, data-driven visual network analytics support the early phases of the investigative process when research questions that may eventually lead to the identification of the structure and mechanisms driving a particular phenomenon surfacing as empirical data are only being derived and specified. In addition, the ostinato model will in general support the creation, analysis, and validation of network representations created to represent an innovation ecosystem in a transparent and structured manner.

In terms of supporting visual network analysis of innovation ecosystems with the type of a holistic and exploratory approach discussed in this section, we see that key pieces of related literature are, first, the MOBENA methodology for business ecosystem network analysis (Battistella et al., 2013) and the NAV model (Hansen et al., 2012). MOBENA was developed to be used in the context of business ecosystems, however, and it does not focus on visual analytics. We consider that the MOBENA model has much value. Therefore, we will return to this model in Section 8.3. The NAV model was developed for conducting social media investigations with low-barrier tools that require running parts of the investigations manually, whereas the ostinato model is designed with the potential to automate the process at hand.

7.5 Evaluation of results in action design research and beyond

A myriad of approaches exists to evaluate the reliability and validity of research results. In quantitative studies, internal validity, external validity (generalizability), and reliability are the three evaluation criteria. Shenton (2004) propose that in qualitative studies, the trustworthiness of research can be evaluated through credibility, transferability, dependability, and confirmability.

However, we would like to remind both ourselves and the reader that we subscribe to the critical realist worldview in this dissertation. This stance will have implications for discussions on validity and the overall quality of this research. Moreover, we reiterate that the investigations included in this dissertation are exploratory or descriptive rather than predictive.

In critical realist data analysis and realism research in general, the core objective is the identification of generative mechanisms that surface as empirical data (Bygstad & Munkvold, 2011; Healy & Perry, 2000). In discussing the validity of the explanatory power of the identified mechanism, Bygstad and Munkvold (2011) refer to Sayer (2010) by stating that instead of identifying as many mechanisms as possible, the investigators should focus on finding the key mechanisms that cause the behavior of a system: "the mechanism with the strongest explanatory power related to the empirical evidence, that is, the causal structure that explains best the events observed." Although these mechanisms may not be identified in the empirical data per se but rather through analytical generalization (Healy & Perry, 2000), the existence of identified mechanisms can be validated through theory-testing studies (S. P. Smith & Johnston, 2014).

Moreover, because we decided to include the individual empirical investigations in the dissertation, we are aware of our responsibility to discuss the quality of these results as well as the results of action design research. Because the individual investigations are exploratory, it is too early to determine individual key mechanisms. We have, however, shown evidence that the companies participating in Finnish Young Innovative Companies program are interconnected through individual people, that a handful of individuals have an important role in the Finnish innovation ecosystem, that the acquisition of venture-backed startups is key part of Google's strategy and differs from that of deals-and-alliances centric Microsoft, that venture capital investors are mobile across the European innovation ecosystem, and that, effectively, the San Francisco Bay Area is by far the most important bridge in the European innovation ecosystem. We admit, however, that these insights are views of these innovation ecosystems rather than generative mechanisms. However, we want to stress the importance of the exploratory phase of the investigative process: revealing the structure supports the development of specific research questions and hypotheses on generative mechanisms.

Construct validity, that is, the extent to which the collected empirical data represents the phenomenon, is as important in critical realist research as it is in positivist investigations (Healy & Perry, 2000). As we will note in our discussion of the limitations of the present research, important actor groups including customers and institutions are missing from our network representations. Moreover, visual network analytics primarily affords the study of the innovation ecosystem structure, not its dynamics. Only in **Publication II** do we investigate the structural evolution of the ecosystem around the focal companies in the study.

We want to raise one more point of discussion regarding validity, that is, the fact that we have used alternative datasets to investigate both the mobile ecosystem as well as the Finnish innovation ecosystem. We showed that the views derived from socially constructed data are very different from those constructed using deals and alliances data sourced from Thomson Reuters SDC. In the context of qualitative studies that are conducted from a realist perspective, Healy and Perry (2000) note that "realism relies on multiple perceptions about a single reality," and point both to multiple sources of data and "several peer researcher's interpretations of those triangulations" as ways to increase the quality of research.

Reliability refers to the extent to which researchers repeating the investigation would arrive at the same results. We claim that at the mechanical level, the computational approach we used in all the investigations contributes to the reliability of the research: reproducibility allows for repeating the data collection and analyses in exactly the same way. However, it is important to note that Bárabasi's (2003) observation applies in the presented case: “Small changes in the topology, affecting only a few of the nodes or links, can open up hidden doors, allowing new possibilities to emerge.” This means that minor changes in boundary specification may lead to major changes in network representations of innovation ecosystems, particularly in actor-level structural metrics. For this reason, we stress the iterative and incremental research process and point to enacted sensemaking (Bendoly, 2016; Weick et al., 2005) as a key topic of further research.

In action design research, there are two axiomatic principles for evaluating research quality: guided emergence, and authentic and concurrent evaluation. Sein et al. (2011) state that ADR emphasizes the organizational relevance of the artifact over its technological rigor and the emergence of the artifact through interaction between the ADR researchers and the organizational context.

To reiterate, this dissertation seeks to satisfy two key objectives that are related to action design research:

- **Objective II** Develop design principles for modeling, representing, and analyzing innovation ecosystems as networks to support their visual investigation.
- **Objective III** Develop a process model to support taking a computational approach to the visual investigation of innovation ecosystems in interdisciplinary teams.

Next, we discuss the building-intervention-evaluation cycles (Sein et al., 2011) that led to the guided emergence of the two key artifacts proposed in this dissertation.

7.5.1 BIE cycles for design principles for modeling innovation ecosystems as networks

The first artifact designed in this dissertation was the set of design principles for modeling innovation ecosystems as networks to support their visual investigation. The design principles are the results of guided emergence that took place in the different investigations that served as experiments in modeling innovation ecosystems as networks. The design principles are described in detail in Chapter 4.

The experiments that investigated the Finnish innovation ecosystem were the starting point of this dissertation research. The investigative team conducted the investigations independently and with limited interaction with the organizational context, in this case the Finnish innovation policy makers at Tekes and Ministry of Employment and Economy. The individual visualizations of the innovation ecosystem were, however, presented to innovation policy actors through project steering groups and a series of round table discussions. To explicate, the experiments on the Finnish innovation ecosystem follow the IT-dominant BIE (cf., Sein et al., 2011).

The experiment on mapping the innovation ecosystem relevant to EIT ICT Labs was conducted using an organization-dominant BIE cycle, in close interaction with representatives of EIT ICT Labs. The premise of the experiment was to use data collected by EIT ICT Labs to represent the innovation ecosystem structure for EIT ICT Labs actors and stakeholders. Moreover, a key objective of the experiment was to investigate the latent structure and existing connections within and in-between the EIT ICT Labs co-location cities. After constructing an early version of the network representation of EIT ICT Labs by using their internal, proprietary data on EIT ICT Labs activities, the investigative team together with EIT ICT Labs representatives, concluded that the insights provided by the visualization were already known to EIT ICT Labs actors.

This realization led to the construction of the second prototype of the innovation ecosystem network representation, which used socially constructed data on companies, their founders, advisors, business angels, and other key individuals as well as organizational investors as the data source. This approach was particularly useful for EIT ICT Labs actors because they were able to observe existing, previously latent connections in-between the co-location cities. Moreover, they gained new evidence of the very limited mobility taking place in-between the co-locations. Most importantly, however, new insights on the imperative role of venture capital investors, both European as well as US-based in general and Silicon Valley-based in particular, were revealed. Further investigation of Silicon Valley's role led to the design of the most important individual visualization artifact in this dissertation, namely the representation of the ecosystem of EIT ICT Labs including Silicon Valley as the hypothetical seventh co-location city of EIT ICT Labs.

The experiment conducted on visualizing the innovation ecosystem of Demola also followed the organization-dominant BIE. The author of this dissertation worked in collaboration with Demola operators to design a way to represent the Demola community as a network. This experiment provided an example that countered EIT ICT Labs with regard to the selection of data source for representing the innovation ecosystem. Guided by Demola operators, we chose to start the network modeling using data on Demola projects and the new connections that project members introduce in between universities and companies that propose the ideas for Demola projects to solve. Moreover, instead of a static visualization, we chose to develop an animation that revealed the evolution and dynamic nature of the Demola plaform in engaging the actors in its extended ecosystem.

Guided emergence through a series of BIE cycles led to the identification of the key difference between Demola and EIT ICT Labs investigations in their usage of network visualizations. In EIT ICT Labs, the network visualizations were primarily used by the EIT ICT Labs operators to investigate and make sense of the existing structure between the co-location cities. Moreover, insights into Silicon Valley's imperative role led to the formulation of a new question: what if Silicon Valley were the seventh co-location center of EIT ICT Labs? In Demola, the key usage of the developed visualizations and animations was to make the Demola process visible to different stakeholders.

7.5.2 BIE cycles for the ostinato model

The second and most important artifact developed in this dissertation is the ostinato model. The model is described in detail in Chapter 6. The ostinato model is a second-level generalized outcome (cf., Sein et al., 2011) of this dissertation research, that is, emerged through conducting several rounds of innovation ecosystem investigations following the design principles described in Chapter 3.

In the following paragraphs, we will describe what we mean by referring to second-level generalized outcome.

Following ADR principles, the ostinato model originated in a practical need, and it builds on existing theory. To use ADR vocabulary (Sein et al., 2011), the ostinato model is equally a result of practice-inspired research as well as an theory-ingrained artifact. The field problem that provided the knowledge-creation opportunity that, through guided emergence, eventually led to the definition of the ostinato model stems from individual experiments where we engaged with innovation ecosystem scholars (i.e., researchers), innovation ecosystem operators (i.e., practitioners) and−towards the end of the dissertation research process and beyond−with actors and stakeholders of the investigated innovation ecosystems (i.e., end-users).

During the experiments, we observed the existence of several repeating phases in implementing the data processing functionalities to support the innovation ecosystem investigations. However, the experimentation-specific requirements often required tailoring these phases in a major way from one experiment to another. This variation suggested that implementing a general-purpose software was not possible in the short term.

Therefore, instead of building a software, we researched the existing literature and theory of process models related to data-driven visual investigations covered in detail in Chapter 5. This brings in the ADR principle of the theory-ingrained artifact (Sein et al., 2011). Key leads in the process model literature were NAV process model (Hansen et al., 2012) and Derek Hansen's talk on infrastructure for supporting computational social science (Hansen, 2013). The final push that inspired us to externalize our accumulated knowledge on the data-driven visual analytics process of innovation ecosystems came from the Kredible.net community.[2] We participated in the Kredible.net Reputation, Trust and Authority Workshop at Stanford University[3] to present our ideas on infrastructure for data-driven visual network analytics of innovation ecosystems. The Kredible.net community took up the idea and accepted our proposal for a book chapter on the topic (Huhtamäki et al., 2015).

The final version of the ostinato model is the result of a number of iteration rounds in which individual investigations serving as ADR experiments were analyzed for processual steps and their interconnections. The validity of the ostinato model is further evaluated and described in Huhtamäki et al. (2017) where a number of the investigations included in this dissertation are analyzed through the ostinato model lens.

[2]Kredible.net: Understanding roles and authority in knowledge markets – An NSF Funded Project. Award No. 1244708, `http://kredible.net/in/`

[3]Kredible.net workshop at Stanford on October 2013, `http://kredible.net/in/second-kredible-net-workshop-stanford-university/`

8 Conclusions

The objective of this dissertation was to develop a methodology for applying network analytics in investigating innovation ecosystems with a data-driven approach. Because the approach is data driven, it should be possible to collect and aggregate data from various heterogeneous sources in a computational fashion. Moreover, it should be possible to automate the overall analysis process. In short, the process should allow reproducible analysis. Specifically, three individual objectives were set for the dissertation. **Objective I** was to contribute to the empirical body of knowledge on innovation ecosystems by running a series of investigations on the structural properties of innovation ecosystems representing different grades of abstraction and complexity. **Objective II** was to develop guidelines for modeling, representing, and measuring innovation ecosystems as networks for their visual analytics. Most importantly, **Objective III** was to design a generic process model for data-driven visual analytics of innovation ecosystems as networks.

To reach these objectives, research was conducted in two complementary streams. First, a series of investigations of innovation ecosystems was conducted to gain knowledge of the structural properties of innovation ecosystems and, even more importantly, of the ways that investigators as well as innovation ecosystem actors and stakeholders prefer to model the innovation ecosystems as networks. Second, a set of requirements were distilled from the investigations, which serve as action design research (Sein et al., 2011) experiment to support the design of the generic process model for data-driven visual analytics of innovation ecosystems as networks.

The main conclusion we draw from the investigations is that conducting data-driven investigations on innovation ecosystems with a network-centric mindset is a valid approach for mapping, exploring, and describing the structure of these ecosystems. Through the series of investigations, we showed that innovation ecosystem investigators as well as innovation ecosystem actors and other stakeholders find value in using network analytics to investigate and explore the structure of their innovation ecosystems of interest.

Addressing innovation ecosystems as networks enables scholars and practitioners to study ecosystem complexity and provides means for mapping and monitoring the individual actors of an innovation ecosystem under investigation, as well as identifying actors for tailored action. In this dissertation, we have presented experiments in taking a data-driven network analytics approach to investigating innovation ecosystems from innovation platform, business domain and development program to national and and international ecosystems. In all these contexts, the main objective of the investigations was to support innovation and growth. Following the action design research approach, a design science variant that is based on iterative and incremental construction of artifacts, in this case network visualizations and supporting processes, allowed us to collect constant feedback from the innovation ecosystem stakeholders through the process of guided emergence.

This dissertation contributes to the emerging field of data-driven innovation ecosystem research in three major ways. First, we contribute to the field of innovation ecosystem research by conducting a series of empirical investigations on the structure and structural evolution of innovation ecosystems representing different grades of abstraction and complexity. These investigations are described in Chapter 3. Second, through the individual investigations serving as action design research experiments, we developed and described design guidelines for modeling, representing and analyzing innovation ecosystems as networks. The guidelines are presented in Chapter 4. Third and most importantly, we developed and began the validation of the ostinato model, a process model for data-driven visual network analytics of innovation ecosystems. The ostinato model, described in detail in Chapter 6, is the key contribution and result of this dissertation.

The ostinato model has two main phases, data collection and refinement, and network creation and analysis. The data collection and refinement phase is divided into entity index creation, Web/API crawling, scraping, and data aggregation. The network creation and analysis phase is composed of filtering in entities, node and edge creation, metrics calculation, node and edge filtering, entity index refinement, layout processing, and visual properties configuration. Finally, the visualizations are provisioned to investigators and other end users with interactive exploration tools and discussion, and their feedback activates an iteration of the process. A cycle of exploration and automation characterizes the model and is embedded in each step.

In addition to the ostinato model, we contribute a set of design principles for investigating innovation ecosystems as networks. The design principles we identified support modeling, analyzing, and visualizing innovation ecosystems as networks, investigating network evolution, allowing for interactive network exploration, and sharing findings with others.

Both the ostinato model and the design principles to be used to investigate innovation ecosystems as networks were developed over multiple experiments following action design research. Importantly, the action design research approach defines inbuilt mechanisms supporting the validation of the created artifacts. The driving principle of validation is guided emergence through which an artifact is created in constant collaboration with actors and organizations for which the artifact is developed. Moreover, perpetual collaboration allows for authentic and concurrent evaluation. This means that evaluation is not a separate step but takes places throughout the artifact creation process. As discussed in detail in Section 7.5, we applied authentic and concurrent evaluation in developing both of the key artifacts of the dissertation.

Three primary target groups will find utility and value in the results of this dissertation.

First, the capability of deriving structural insights into the ecosystem and actor levels based on the data analysis is imperative in pushing the academic research forward. Although there is a substantial body of extant empirical research on innovation ecosystems, the majority of innovation ecosystem research takes place in either individual organizations or pairs (dyads) of these organizations (Järvi & Kortelainen, 2016). As presented in Section 2.2, there is a limited amount of empirical research on the ecosystem-level network structure of innovation ecosystems outside this dissertation. Key means for breaking the limitations of providing ecosystem-level insights into the network structure of innovation ecosystems implicit in name-generating surveys, which is the traditional way to collect data on connections between ecosystem actors, is to utilize archival records in sourcing data for computational analysis of innovation ecosystems (cf. Williams & Shepherd, 2015). The ostinato model supports the utilization of new sources of digital data in conducting innovation ecosystem research. Importantly, the digital data enabling the analysis of the network structure of innovation ecosystems often is transactional microdata, that is, it represents timestamped, actor-level interactions.

Second, the domain of computational innovation ecosystem analytics, an extension of business ecosystem analytics, benefits equally from the ecosystem-level views of the innovation ecosystem network structure created with the ostinato model. Even more importantly, the ostinato model can be used to design and implement the architecture of analytical tools and dashboards that render views into innovation ecosystem structure eventually in the self-service mode. The target users of computational innovation ecosystem analytics include policy, business, and investment decision makers that seek to facilitate the emergence of innovation ecosystems and orchestrate their evolution.

Third, views on the structure and dynamics of an innovation ecosystem have a focal role in innovation ecosystem orchestration. Data-driven visualizations are an organic part of the innovation ecosystem transformation framework (Russell et al., 2011), in which visualizations are used to support innovation ecosystem actors in arriving at a shared vision of a joint future toward which they seek to make their way following individual paths. The empirical work on innovation ecosystem orchestration is outside the scope of this dissertation. However, we derived requirements from the orchestration process.

We used several datasets in these investigations, including social media, socially constructed data available online, and proprietary sets of data represented as spreadsheets and other formats. The investigations included in this dissertation were conducted using two main sources of data. The socially constructed and curated IEN Dataset, specifically IEN Executives and Finance and IEN Angels and Startups, has served as the main source of data. The IEN Dataset was used to gather data on startups and growth companies and their transactions with individuals, investors, and other innovation ecosystem actors. Moreover, Thomson Reuters SDC data was used to gather data on deals and alliances between already established enterprises. In addition, we used Twitter data to map the customer landscape around Tekes Young Innovative Companies and Demola in-house data for data on innovation projects and their actors over time. Moreover, the ostinato model was designed to include the process for collecting new sets of digital data relevant to a particular investigation.

We claim that that the ostinato model is general enough to be used in domains outside innovation ecosystem studies. In fact, we have already taken steps toward generalizing and validating the ostinato model in investigations outside the innovation ecosystem sphere. The author of the dissertation has joined with other investigators to apply the ostinato model in case studies in addition to those included in this dissertation. This work includes innovation ecosystem analysis (Russell, Huhtamäki, et al., 2015), investigations of the use of Twitter in emerging communities (Aramo-Immonen, Jussila, & Huhtamäki, 2015; Aramo-Immonen, Kärkkäinen, Jussila, Joel-Edgar, & Huhtamäki, 2016; Jussila, Huhtamäki, Henttonen, Kärkkäinen, & Still, 2014), and communication between political and journalistic elites (Ruoho, Kuusipalo, Vainikka, & Huhtamäki, 2016; Vainikka & Huhtamäki, 2015). However, further research is needed by investigators other than the present author to evaluate the true value and generalization potential of the ostinato model.

8.1 Implications for innovation ecosystem actors

The research for this dissertation was conducted in close interaction with several innovation ecosystem actors and stakeholders. The innovation ecosystem investigators, that is, the co-authors of the articles included in this dissertation, form the core group of co-creators. Moreover, through the investigations, we interacted with innovation ecosystem orchestrators (EIT ICT Labs), policy makers (Tekes, Ministry of Employment and the Economy, Council of Tampere Region) as well as startup entrepreneurs, investors, and other innovation ecosystem actors. In the following sections, we will discuss the implications of the results of this research.

8.1.1 Innovation ecosystem investigators

Innovation ecosystems are a new domain on which there is a limited amount of empirical research. Several novel streams of digital data are, however, available to be analyzed in order to extend the empirical body of knowledge on innovation ecosystems. The network-based approach comes with new metrics and the objective for creating an ecosystem-level view instead of a more atomistic, sample-through-survey based way for conducting research. Developing the investigative process in an iterative and incremental manner is imperative in managing its complexity. An analytics process that is exploratory, interactive, and transparent supports maintaining balanced communication between the members of the investigative team and therefore supports interdisciplinary collaboration.

The most significant result of this dissertation is the ostinato model for data-driven visual network analytics. This model will help innovation ecosystem investigators in designing data-processing pipelines and architectures for visual network analytics of innovation ecosystems. In order to be able to develop independent software components or integrate existing API-based services that excel in individual parts of the process, there has to be a clear distinction between the individual steps of the process. Moreover, a consistent structure used in individual investigations allows for the reuse of analytics components. We look forward to observing how innovation ecosystem investigators will make use of the ostinato model in seeking interoperability in developing components and toolchains for the data-driven visual network analytics process.

In addition to visual analytics, innovation ecosystem scholars will benefit from applying the ostinato model and design principles for analyzing innovation ecosystems as networks when creating network models of innovation ecosystems for quantitative and statistical analysis. As we have shown in this dissertation, a rich variety of options exists in data sources, network modeling decisions, selection of node and network metrics, and, decisively, boundary specification. All these options will make a major difference in the metrics that are fed into statistical and machine learning models. Therefore, the creation process of the network model of an innovation ecosystem should be made as transparent and tractable as possible.

Through a process following the ostinato model, innovation ecosystem investigators gain access to empirical data that can be fed to agent-based and other models of innovation ecosystems. This enables investigations of the impacts of the structure and conduct (Afuah, 2013; Ahuja et al., 2012) of the innovation ecosystem of interest; networks represent the pathways that show where recent activity has taken place.

In all, quantitative and statistical analysts will find the ability to maintain low entry barrier and high transparency as well as automation and overall reproducibility as useful as those who investigate innovation ecosystems using the visual analytics approach.

8.1.2 Innovation ecosystem orchestrators and policy makers

In this section, for brevity, we use (innovation ecosystem) orchestrators to refer to policy makers, program developers, university and technology park actors, and others with a similar role in facilitating innovation activities that cross the boundaries of individual organizations.

From the viewpoints of policy making and innovation ecosystem orchestration, the key contribution of this dissertation is the provision of an ecosystem-level view for innovation ecosystems of interest in the investigative context. With the help of the design guidelines described in Chapter 4 and the ostinato model, similar ecosystem-level views can be created to individual ecosystems representing different grades of abstraction and complexity. Moreover, the innovation ecosystem transformation framework (Russell, Huhtamäki, et al., 2015; Russell et al., 2011) provides a holistic way to organize data-driven, evidence-based orchestration efforts in a rigorous manner.

Figure 8.1: New technology allows more interactive ways to explore data for sensemaking. Prototype visualization explored using Bluescape on MultiTaction.

We do not propose replacing traditional statistical indicators with the ecosystem-level view. Instead, we claim that the ecosystem-level view does provide holistic insights into the innovation ecosystem structure and evolution. Moreover, they provide a context for individual indicators and measurements. There is, however, a mismatch between the existing statistics available to orchestrators and the objective of creating an ecosystem-level view. Statistics are often aggregated into categories, such as according to the classification of economic activities.[1] Transactional microdata is needed to create an ecosystem-level view. Such data is long term, representing individual actor-level transactions over time.

Data on scientific articles is a good example of transactional micro-level relational data. Article authors are enumerated and publishing dates are available. Moreover, the authors whose work has provided the knowledge baseline for the article are explicitly mentioned

[1] Toimialaluokitus in Finnish.

Figure 8.2: Visual analytics supporting decision-making. Experimental setting at Stanford in 2013.

through citations. The availability of bibliographic data as well as its importance to researchers has led to a significant body of literature on bibliometrics and scientometrics. However, accessing and particularly aggregating bibliometrical data from multiple sources continue to be far from trivial tasks (Huhtamäki, 2016). Data formats and access mechanisms vary, and global identifiers of actors do not exist in the data.

We urge orchestrators not to settle for data that is aggregated through a static, out-dated taxonomy. These taxonomies are useful in maintaining long-term, consistent, comparable statistics and statistical analysis. For action-oriented, real-time policy making, learning what is working and therefore providing a feedback mechanism for ecosystemic innovation activities is a key priority. Therefore, orchestrators should make sure that they have access to multi-level temporal data on companies, their creators, and enablers−including venture-capital investors−in a way that supports the creation of multiscopic views of the ecosystem of interest. Moreover, orchestrators should at least consider providing access to this data to scholars and other stakeholders. Steps toward this direction are being taken at the moment: datasets on projects funded by both Academy of Finland, Tekes as well as European Commission's Horizon 2020 are already available online. However, work remains to be done even among the aforementioned examples, particularly to make sure that the properties that are expected from open data are met.[2] Organizations leading the digital transformation, such as Cap Digital,[3] extensively apply the data-driven approach in developing and orchestrating their innovation ecosystem.

[2]According to https://okfn.org/opendata/, "'Open knowledge' is any content, information or data that people are free to use, re-use and redistribute — without any legal, technological or social restriction."

[3]Cap Digital: the French business cluster for digital transformation, `http://www.capdigital.com/`

Finally, we encourage the orchestrators to use the new interactive tools in making sense of innovation ecosystems of their interest. Figure 8.1 gives an example of a setup based on a large touchscreen allowing for co-exploration and co-referencing. A bolder vision for supporting evidence-based decision making for innovation ecosystem orchestration would be to build a situation room, physical or virtual, with a visual representation supporting situation awareness[4]. Figure 8.2 shows an example of a possible setup. However, before building the situation room, the orchestrators must make an effort to support the organizational use of visualization (Bendoly, 2016): "Creating dynamic interactive interfaces for multiple stakeholders, accessible at their desktops, focusing not on how many superfluous frills can be added but rather on practical storytelling... This is what will tend to support the organizational use of visualization."

Finally, those who take up visual analytics should bear in mind that while visualizations support users in making faster decisions, the confidence that actors have about their decisions does not predict the accuracy of the decisions. Therefore, it is important to apply and develop further the practices of enacted sensemaking and to validate the usefulness of the visual tools developed to support decision-making trough user experiments.

8.1.3 Other ecosystem actors: startups, enterprises, investors

We acknowledge that in socially constructed data, there is very likely a significant bias in favor of startups, incubators, and programs where actors invest into communicating their activities, the funding rounds they receive, the advisers that support the creation and development of the companies, and perhaps even the versions of products being built. However, we want to note that such a bias may very likely exist in the field of innovation in general, particularly in the context of consumer products and services. It is fair to assume that those who invest in communicating their results will improve their chances to be successful and make a more significant impact. In scientific publishing, for example, those that communicate their results and publications will gain more attention, and are likely to receive more citations, and therefore get better marks in citation-based measurements, including the h-index (cf., Terras, 2012).

In order to enable visibility and attract attention, a startup should make sure that they have a presence in different social media platforms as well as community-curated databases, including Wikipedia, CrunchBase, and others. The startups should have a website with the latest information on the team, advisors, investors, references, and other connections of significance.

Using navigation in the physical world as a metaphor, the approach presented in this dissertation was used to draw maps of the topology of innovation ecosystems. When the approach and results of the investigations included in this dissertation were presented to startup ecosystem actors, they noted the value of real-time maps of innovation ecosystems in keeping up with the evolution of the ecosystem−their competitors, customers, and developers of complementary products and services. Network maps contextualize information on individual companies, their products and services to show how the companies and their offering are positioned in the ecosystem.

[4]Tilannekuva in Finnish.

8.2 Limitations

We are fully aware that this type of multidisciplinary, exploratory investigation, in which the work is conducted at three complementary levels, is necessarily limited. In this section, we explicate the limitations in four categories. First, we point to the limitations related to the volume of data used in the investigations. Second, we acknowledge the limits of big data and related computational methods in terms of the extent to which a phenomenon can indeed be investigated by using big data. Third, we note the limitations of the exceptionally broad boundary specification we used in our empirical investigations. Fourth and last, we discuss the limitations of the methodology selected for this dissertation, particularly social network analysis and visual network analytics.

First, we consider that key limitation in the presented ostinato model is the volume of data in use in the investigation. Specifically, although the ostinato model per se can be used to structure a process that is built using big data-scale technology such as Apache Spark[5] or Hadoop, the use of static (tabular) files in representing data adding to the transparency of the process to non-technical investigators becomes impossible once the volume of the data increases to a certain limit. However, while the source data may be voluminous, entity index creation, node and edge filtering, and boundary specification all provide means to manage the amount of data in the investigative process.

Second, big data, computational social science, and visual analytics are all fields that receive justified criticism (cf., boyd & Crawford, 2012). We acknowledge that a series of investigations in both laboratories and in the "wild" must be conducted to validate and refine the ways data-driven visual analytics best supports cross-organizational innovation activities and their analyses.

Third, this dissertation was conducted in the context of innovation ecosystem investigations. The definition (Russell et al., 2011) we adopted to guide the investigations and the selection of innovation ecosystems we engaged with to run the experiments was very broad; therefore, setting the boundaries of the innovation ecosystems under investigation was challenging. Although these challenges have, in fact, worked in our favor in developing the ostinato model, we admit that we are limited in contributing to the extant literature on innovation ecosystems in which the boundaries are often defined specifically, and therefore the investigations have led to specific conclusions. We consider that more focused investigations will allow empirical contributions to the identification of different mechanisms latent in innovation ecosystems. We will return to this topic in the final section on future work.

Fourth and finally, our experience showed that visual network analytics is best suited for exploratory analysis, an important step in any empirical investigation. Moving forward to descriptive, predictive, and prescriptive analysis however requires the application of alternative analytical approaches.

8.3 Future work

This dissertation opens up several avenues for further research. In this concluding section of the dissertation, we focus on four of these opportunities. First, we discuss future work related to the technological aspects of the ostinato model. Second, we urge innovation ecosystem scholars to seek ways to conduct focused predictive investigations at the

[5] Apache Spark - Lightning-fast cluster computing, `http://spark.apache.org/`

ecosystem level. Third, we point to research on closing the gap between the ostinato model and the enacted sensemaking process for management. Fourth, we encourage research on utilizing visual network analytics-backed enacted sensemaking as means of conducting orchestration in innovation ecosystems.

Future work includes, first, additional rounds of refining the ostinato model based on the feedback collected from researchers and practitioners working with the exploration−automation cycle of data-driven visual network analytics and applying the model in different contexts.

On a related note, to lower the barrier to applying the ostinato model in supporting data-driven investigations and orchestrations of innovation ecosystems, we believe that a software framework following the ostinato model should be implemented. At best, the framework would allow easy access (Web-based), real-time operations, and stream-based processing. Moreover, the practices of both contemporary software development and interactive computing should be taken into account when developing such a framework. Practices related to existing tools, such as Grunt[6], a popular Javascript-based task runner as well as Python-based automation tools, including PyBuilder,[7,8] and Luigi,[9] should be investigated for infrastructure and best practices.

As an ecosystem of tools and components is developed and further requirements for interoperability are articulated, we see the possibility of forming a community of people that move the field forward. They will need a package management framework, system components, and means to collaborate. We look forward to contributing to this work.

Second, we grant that the empirical work on ecosystem-level investigations is still a very early stage. Therefore, we suggest conducting carefully focused investigations on the individual ecosystem-level mechanisms that drive and take place in innovation ecosystems, including interconnectedness, interdependency, co-evolution, value co-creation, and coopetition. The use of statistical analysis, agent-based modeling, and other modeling methods are needed to complement network modeling and visual network analysis. We further believe that the critical realist worldview provides these investigations with a firm foundation.

In this dissertation, we focused on supporting the development of technological architecture and related process models that will add to the automation, reproducibility, and overall manageability of the analytics process used in the data-driven visual network analytics of innovation ecosystems. This focus forced us to delimit two particularly important domains of research out of scope of this dissertation: visual analytics-driven enacted sensemaking and innovation ecosystem orchestration. Both of these domains are imperative in applying the proposed approach in orchestrating innovation ecosystems and related management activities

Therefore, third, in order to close the gap between the data-driven process of the visual network analytics of innovation ecosystems and the application of this process and effectively move innovation ecosystem analytics to the prescriptive stage, we propose conducting a series of continued action design research experiments complemented by experiments following the human-computer interaction tradition of evaluating the developed information system artifacts for their quality and efficiency. To ensure their relevance,

[6]Grunt: The JavaScript Task Runner, `http://gruntjs.com/`
[7]PyBuilder: Build automation for Python, `http://pybuilder.github.io/`
[8]Cf. related discussion on Twitter, `https://twitter.com/jsalonen/status/612644821052878848`
[9]Luigi: build complex pipelines of batch jobs in Python, `https://github.com/spotify/luigi`

these experiments should be conducted in the "wild", that is, in a real context where the practices of visual network analytics-driven enacted sensemaking are applied and developed. We suggest that in conducting such investigations, the MOBENA methodology (Battistella et al., 2013) may provide useful guidelines, and it therefore deserves attention.

Fourth and lastly, we urge the researchers and practitioners of visual network analytics, enacted sensemaking, and innovation orchestration to combine their efforts in researching ways to utilize visual network analytics-backed enacted sensemaking as means to satisfy the requirements stemming from the contemporary approach to innovation ecosystem orchestration.

We are confident that this dissertation provides a solid basis for these future endeavours.

Bibliography

Adner, R. (2012). *The Wide Lens: A New Strategy for Innovation.* New York: Penguin Publishing.

Adner, R., & Kapoor, R. (2010). Value creation in innovation ecosystems: how the structure of technological interdependence affects firm performance in new technology generations. *Strategic Management Journal*, *31*(3), 306–333. doi: 10.1002/smj.821

Afuah, A. (2013). Are network effects really all about size? The role of structure and conduct. *Strategic Management Journal*, *34*(3), 257–273. doi: 10.1002/smj.2013

Ahuja, G., Soda, G., & Zaheer, A. (2012). The genesis and dynamics of organizational networks. *Organization Science*, *23*(2), 434–448.

Aken, J. E. v. (2004). Management Research Based on the Paradigm of the Design Sciences: The Quest for Field-Tested and Grounded Technological Rules. *Journal of Management Studies*, *41*(2), 219–246. doi: 10.1111/j.1467-6486.2004.00430.x

Aramo-Immonen, H., Jussila, J., & Huhtamäki, J. (2015). Exploring co-learning behavior of conference participants with visual network analysis of Twitter data. *Computers in Human Behavior*, *51*, 1154–1162. doi: 10.1016/j.chb.2015.02.033

Aramo-Immonen, H., Kärkkäinen, H., Jussila, J. J., Joel-Edgar, S., & Huhtamäki, J. (2016). Visualizing informal learning behavior from conference participants' Twitter data with the Ostinato Model. *Computers in Human Behavior*, *55, Part A*, 584–595. doi: 10.1016/j.chb.2015.09.043

Archer, M. S. (1995). *Realist social theory: the morphogenetic approach.* Cambridge: Cambridge University Press.

Autio, E., & Thomas, L. D. W. (2013). Innovation Ecosystems: Implications for Innovation Management. In M. Dodgson, N. Philips, & D. M. Gann (Eds.), *The Oxford Handbook of Innovation Management.* Oxford University Press.

Barabási, A.-L. (2003). *Linked: How Everything Is Connected to Everything Else and What It Means.* USA: Plume.

Barabási, A.-L., & Albert, R. (1999). Emergence of scaling in random networks. *Science*, *286*(5439), 509–512.

Barabási, A.-L., & Bonabeau, E. (2003). Scale-Free Networks. *Scientific American*, *288*(5), 50–59.

Basole, R. C. (2009). Visualization of interfirm relations in a converging mobile ecosystem. *Journal of Information Technology*, *24*(2), 144–159. doi: 10.1057/jit.2008.34

Basole, R. C. (2014). Visual Business Ecosystem Intelligence: Lessons from the Field. *IEEE Computer Graphics and Applications*, *34*(5), 26–34. doi: 10.1109/MCG.2014.104

Basole, R. C. (2016). Accelerating Digital Transformation: Visual Insights from the API Ecosystem. *IT Professional Magazine*, Forthcoming.

Basole, R. C., Clear, T., Hu, M., Mehrotra, H., & Stasko, J. (2013). Understanding Interfirm Relationships in Business Ecosystems with Interactive Visualization. *IEEE*

Transactions on Visualization and Computer Graphics, *19*(12), 2526–2535. doi: 10.1109/TVCG.2013.209

Basole, R. C., Huhtamäki, J., Still, K., & Russell, M. G. (2016). Visual Decision Support for Business Ecosystem Analysis. *Expert Systems with Applications*, *65*, 271–282. doi: 10.1016/j.eswa.2016.08.041

Basole, R. C., Russell, M. G., Huhtamäki, J., & Rubens, N. (2012). Understanding Mobile Ecosystem Dynamics: A Data-Driven Approach. In *Proceedings of 2012 International Conference on Mobile Business, Delft, The Netherlands.* Retrieved from `http://aisel.aisnet.org/icmb2012/15/`

Basole, R. C., Russell, M. G., Huhtamäki, J., Rubens, N., Still, K., & Park, H. (2015). Understanding Business Ecosystem Dynamics: A Data-Driven Approach. *ACM Transactions on Management Information Systems (TMIS)*, *6*(2), 6. Retrieved from `http://dx.doi.org/10.1145/2724730`

Bastian, M., Heymann, S., & Jacomy, M. (2009). Gephi: An Open Source Software for Exploring and Manipulating Networks. In *Proceedings of the Third International AAAI Conference on Weblogs and Social Media, May 17-20, 2009, San Jose, California, USA.*

Battistella, C., Colucci, K., De Toni, A. F., & Nonino, F. (2013). Methodology of business ecosystems network analysis: A case study in Telecom Italia Future Centre. *Technological Forecasting and Social Change*, *80*(6), 1194–1210. doi: 10.1016/j.techfore.2012.11.002

Bendoly, E. (2016). Fit, Bias and Enacted Sensemaking in Data visualization: Frameworks for Continuous Development in operations and Supply Chain Management Analytics. *Journal of Business Logistics.*

Bendoly, E. (2017). Best Practices in Visual Design. In E. Bendoly & S. Clark (Eds.), *Visual analytics for management: Translational science and applications in practice.* Routledge, Forthcoming.

Bizer, C., Heath, T., & Berners-Lee, T. (2009). Linked Data - The Story So Far. *International Journal on Semantic Web and Information Systems*, *5*(3), 1–22.

Bizzi, L., & Langley, A. (2012). Studying processes in and around networks. *Industrial Marketing Management*, *41*(2), 224–234. doi: 10.1016/j.indmarman.2012.01.007

boyd, d., & Crawford, K. (2012). Critical Questions for Big Data. *Information, Communication & Society*, *15*(5), 662–679. doi: 10.1080/1369118X.2012.678878

Bush, V. (1945). As We May Think. *The Atlantic Monthly*. Retrieved from `http://www.theatlantic.com/magazine/archive/1945/07/as-we-may-think/303881/` doi: 10.1145/227181.227186

Bygstad, B., & Munkvold, B. E. (2011). In Search of Mechanisms. Conducting a Critical Realist Data Analysis. In *Thirty Second International Conference on Information Systems, Shanghai 2011.*

Card, S. K., Mackinlay, J. D., & Shneiderman, B. (1999). *Readings in information visualization: using vision to think.* San Francisco, California: Morgan Kaufmann.

Carlson, C. R., & Wilmot, W. W. (2006). *Innovation: The Five Disciplines for Creating What Customers Want.* Crown Business.

Carlsson, S. A. (2010). Design Science Research in Information Systems: A Critical Realist Approach. In (pp. 209–233). Springer US.

Cioffi-Revilla, C. (2010). Computational social science. *Wiley Interdisciplinary Reviews: Computational Statistics*, *2*(3), 259–271. doi: 10.1002/wics.95

Clarysse, B., Wright, M., Bruneel, J., & Mahajan, A. (2014). Creating value in ecosystems: Crossing the chasm between knowledge and business ecosystems. *Research Policy*, *43*(7), 1164–1176. doi: 10.1016/j.respol.2014.04.014

Davenport, T. H. (2013). Analytics 3.0. *Harvard Business Review*, *91*(12), 64–72.

Davenport, T. H. (2014). *Big Data at Work: Dispelling the Myths, Uncovering the Opportunities.* Boston, Massachusetts, USA: Harvard Business Press Books.

Delen, D., & Demirkan, H. (2013). Data, information and analytics as services. *Decision Support Systems*, *55*(1), 359–363. doi: 10.1016/j.dss.2012.05.044

Dobson, P. J. (2001). The Philosophy of Critical Realism—An Opportunity for Information Systems Research. *Information Systems Frontiers*, *3*(2), 199–210. doi: 10.1023/A:1011495424958

Drucker, P. (2015). *Innovation and Entrepreneurship: Practice and Principles* (Routledge ed.). New York, New York, USA: Routledge.

Dunn, C., & Grabski, S. (2001). An Investigation of Localization as an Element of Cognitive Fit in Accounting Model Representations. *Decision Sciences*, *32*(1), 55–94. doi: 10.1111/j.1540-5915.2001.tb00953.x

Evans, P. C., & Basole, R. C. (2016). Revealing the API Ecosystem and Enterprise Strategy via Visual Analytics. *Communications of the ACM*, *59*(2), 26–28. doi: 10.1145/2856447

Finkel, J. R., Grenager, T., & Manning, C. (2005). Incorporating Non-local Information into Information Extraction Systems by Gibbs Sampling. In *ACL '05 Proceedings of the 43rd Annual Meeting on Association for Computational Linguistics* (p. 363–370). Association for Computational Linguistics. doi: 10.3115/1219840.1219885

Fox, P., & Hendler, J. (2011). Changing the equation on scientific data visualization. *Science*, *331*(6018), 705–708. doi: 10.1126/science.1197654

Freeman, L. C. (1978). Centrality in social networks: Conceptual clarification. *Social Networks*, *1*(3), 215–239. doi: 10.1016/0378-8733(78)90021-7

Freeman, L. C. (2000). Visualizing Social Networks. *Journal of Social Structure*, *1*(1), [np]. Retrieved from `http://www.cmu.edu/joss/content/articles/volume1/Freeman.html`

Freeman, L. C. (2009). Methods of Social Network Visualization. In R. A. Meyers (Ed.), *Encyclopedia of complexity and systems science* (p. 145–156). Berlin: Springer.

Ghosh, S. (2013). *Python Tools for Reproducible Research in Brain Imaging.* Retrieved from `https://speakerdeck.com/satra/pydata-2013-python-tools-for-reproducible-research-in-brain-imaging`

Giuliani, E., & Bell, M. (2008). *Industrial clusters and the evolution of their knowledge networks: revisiting a Chilean case* (Tech. Rep.). Falmer, Brighton, UK: SPRU.

Granovetter, M. (1973). The Strength of Weak Ties. *American journal of sociology*, *78*(6), 1360–1380.

Gray, J., Bounegru, L., & Chambers, L. (Eds.). (2012). *Data Journalism Handbook* (1.0 Beta ed.). Retrieved from `http://datajournalismhandbook.org/1.0/en/index.html`

Gregor, S., & Hevner, A. R. (2013). Positioning and Presenting Design Science Research for Maximum Impact. *MIS Quarterly*, *37*(2), 337-A6.

Gupta, A. K., Tesluk, P. E., & Taylor, M. S. (2007). Innovation At and Across Multiple Levels of Analysis. *Organization Science*, *18*(6), 885–897. doi: 10.1287/orsc.1070.0337

Hagel, J., & Seely Brown, J. S. (2005). *The Only Sustainable Edge: Why Business Strategy Depends on Productive Friction and Dynamic Specialization.* Boston, Massachusetts, USA: Harvard Business Press.

Hansen, D. L. (2013). *Infrastructure for Supporting Computational Social Science.* Purdue University, West Lafayette, Indiana, USA. Retrieved from `http://kredible.net/in/infrastructure_for_computational_social_science/`

Hansen, D. L., Rotman, D., Bonsignore, E., Milic-Frayling, N., Rodrigues, E. M., Smith, M., & Shneiderman, B. (2012). Do You Know the Way to SNA?: A Process Model for Analyzing and Visualizing Social Media Network Data. In *Proceedings of the 2012 International Conference on Social Informatics* (pp. 304–313). IEEE. doi: 10.1109/SocialInformatics.2012.26

Hansen, D. L., Shneiderman, B., & Smith, M. A. (2011). *Analyzing Social Media Networks with NodeXL: Insights from a Connected World.* Burlington, MA, USA: Morgan Kaufmann.

Healy, M., & Perry, C. (2000). Comprehensive criteria to judge validity and reliability of qualitative research within the realism paradigm. *Qualitative Market Research: An International Journal*, *3*(3), 118–126. doi: 10.1108/13522750010333861

Heer, J., & boyd, d. (2005). Vizster: Visualizing online social networks. In *Proceedings of the 2005 IEEE Symposium on Information Visualization (INFOVIS '05)* (p. 32–39). IEEE.

Heer, J., & Shneiderman, B. (2012). Interactive Dynamics for Visual Analysis. *Communications of the ACM*, *55*(4), 45–54.

Hevner, A. R., March, S. T., Park, J., & Ram, S. (2004). Design Science in Information Systems Research. *MIS Quarterly*, *28*(1), 75–105.

Huhtamäki, J. (2007). Community visualisations in Open Knowledge Space: Uncovering rabbit holes in a digital ecosystem. In *Proceedings of the 1st OPAALS conference, November 26-27, 2007, Rome, Italy.* opaals.eu.

Huhtamäki, J. (2016). Visualizing Co-authorship Networks for Actionable Insights: Action Design Research Experiment. In *Proceedings of 20th International Academic Mindtrek Conference, October 17- 19, 2016, Tampere, Finland.* Forthcoming.

Huhtamäki, J., Luotonen, V., Kairamo, V., Still, K., & Russell, M. G. (2013). Process for Measuring and Visualizing an Open Innovation Platform: Case Demola. In *17th International Academic MindTrek Conference 2013: "Making Sense of Converging Media", October 1-3, Tampere, Finland.* ACM. Retrieved from http://urn.fi/URN:NBN:fi:tty-201312201533

Huhtamäki, J., Russell, M. G., Rubens, N., & Still, K. (2010). A Network-Centric Snapshot of Value Co-Creation in Finnish Innovation Financing. In *Proceedings of EBRF 2010 - Co-creation as the way forward, Nokia, Finland.*

Huhtamäki, J., Russell, M. G., Rubens, N., & Still, K. (2015). Ostinato: The exploration-automation cycle of user-centric, process-automated data-driven visual network analytics. In S. A. Matei, M. G. Russell, & E. Bertino (Eds.), *Transparency in social media: Tools, methods and algorithms for mediating online interactions* (p. 197–222). Springer International Publishing Switzerland. Retrieved from http://link.springer.com/chapter/10.1007/978-3-319-18552-1_11 doi: 10.1007/978-3-319-18552-1_11

Huhtamäki, J., Russell, M. G., & Still, K. (2017). Processing data for visual network analytics: innovation ecosystem experiences. In E. Bendoly & S. Clark (Eds.), *Visual analytics for management translational science and applications in practice.* New York: Taylor & Francis / Routledge.

Huhtamäki, J., Russell, M. G., Still, K., & Rubens, N. (2011). A Network-Centric Snapshot of Value Co-Creation in Finnish Innovation Financing. *Open Source Business Resource*, 13–21. Retrieved from http://timreview.ca/article/424

Huhtamäki, J., Russell, M. G., Still, K., & Rubens, N. (2013). A Network-Centric Snapshot of Value Co-Creation in Finnish Innovation Financing. In S. Tanev, M. Seppä, & A. Chowaniec (Eds.), *Value Co-Creation: Best of TIM Review.* Talent First Network.

Huhtamäki, J., Salonen, J., Marttila, J., & Nykänen, O. (2010). Context-Driven Social Network Visualisation: Case Wiki Co-Creation. In D. Karabeg & J. Park (Eds.), *Proceedings of the Second International Workshop on Knowledge Federation: Self-Organizing Collective Mind, October 3-6, 2010, Dubrovnik, Croatia.* Dubrovnik, Croatia: CEUR-WS.org. Retrieved from `http://urn.fi/URN:NBN:fi:tty-201201161008`

Huhtamäki, J., Still, K., Isomursu, M., Russell, M. G., & Rubens, N. (2012). Networks of Growth: Case Young Innovative Companies in Finland. In *Proceedings of the 7th European Conference on Innovation and Entrepreneurship (ECIE), September 20-21, 2012, Santarém, Portugal.* Santerém, Portugal.

Huotari, P., Järvi, K., Kortelainen, S., & Huhtamäki, J. (2016). Winner Does Not Take All: Selective Attention and Local Bias in Platform-based Markets. *Technological Forecasting and Social Change*, Forthcoming. Retrieved from `http://dx.doi.org/10.1016/j.techfore.2016.08.028` doi: 10.1016/j.techfore.2016.08.028

Hwang, V. W., & Horowitt, G. (2012). *The Rainforest: The Secret to Building the Next Silicon Valley.* Los Altos Hills, California: Regenwald.

Iansiti, M., & Levien, R. (2004). *The Keystone Advantage: What the New Dynamics of Business Ecosystems Mean for Strategy, Innovation, and Sustainability.* Boston, Mass: Harvard Business Review Press.

Indarto, E. (2013). *Data Mining.* Retrieved from `http://recommender-systems.readthedocs.org/en/latest/datamining.html`

Intel. (2013). *Extract, Transform, and Load Big Data with Apache Hadoop* (Tech. Rep.). Intel White Paper. Retrieved from `https://software.intel.com/en-us/articles/extract-transform-and-load-big-data-with-apache-hadoop`

Iyer, B., & Henderson, J. C. (2010). Preparing for the Future: Understanding the Seven Capabilities of Cloud Computing. *MIS Quarterly Executive*, *9*(2).

Iyer, B., Lee, C.-H., & Venkatraman, N. (2006). Managing in a "Small World Ecosystem": Lessons from the Software Sector. *California Management Review*, *48*(3), 28–47.

Järvi, K. (2013). *Ecosystem architecture design: endogenous and exogenous structural properties* (Doctoral Dissertation, Lappeenranta University of Technology). Retrieved from `http://urn.fi/URN:ISBN:978-952-265-465-6`

Järvi, K., & Kortelainen, S. (2016). Taking stock of empirical research on business ecosystems: a literature review. *International Journal of Business and Systems Research*, Forthcoming.

Jesus, R., Schwartz, M., & Lehmann, S. (2009). Bipartite networks of Wikipedia's articles and authors: a meso-level approach. In *Proceedings of the 5th International Symposium on Wikis and Open Collaboration (WikiSym '09)* (pp. Article 5, 10 pages). Orlando, Florida: ACM. doi: 10.1145/1641309.1641318

Jussila, J., Huhtamäki, J., Henttonen, K., Kärkkäinen, H., & Still, K. (2014). Visual network analysis of Twitter data for co-organizing conferences: case CMAD 2013. In *Proceedings of the 47th annual hawaii international conference on system sciences, january 6-9, 2014* (pp. 1474–1483). Computer Society Press. Retrieved from `http://urn.fi/URN:NBN:fi:tty-201401221053`

Kandel, S., Heer, J., Plaisant, C., Kennedy, J., van Ham, F., Riche, N. H., ... Buono, P. (2011). Research Directions in Data Wrangling: Visualizations and Transformations for Usable and Credible Data. *Information Visualization*, *10*(4), 271–288. doi: 10.1177/1473871611415994

Keim, D., Kohlhammer, J., Ellis, G., & Mansmann, F. (2010). *Mastering the Information Age - Solving Problems with Visual Analytics.* Eurographics Association.

Kleinberg, J. M. (1999). Authoritative sources in a hyperlinked environment. *Journal of the ACM*, *46*(5), 604–632. doi: 10.1145/324133.324140

Konno, N., Nonaka, I., & Ogilvy, J. (2014). Scenario Planning: The Basics. *World Futures*, *70*(1), 28–43. doi: 10.1080/02604027.2014.875720

Kurtz, C. F., & Snowden, D. J. (2003). The new dynamics of strategy: Sense-making in a complex and complicated world. *IBM Systems Journal*, *42*(3), 462 - 483. doi: 10.1147/sj.423.0462

Lam, H., Bertini, E., Isenberg, P., Plaisant, C., & Carpendale, S. (2012). Empirical Studies in Information Visualization: Seven Scenarios. *IEEE Transactions on Visualization and Computer Graphics*, *18*(9), 1520–1536. doi: 10.1109/TVCG.2011.279

Lazer, D., Pentland, A., Adamic, L., Aral, S., Barabási, A.-L., Brewer, D., ... Alstyne, M. V. (2009). Computational Social Science. *Science*, *323*(5915), 721–723. doi: 10.1126/science.1167742

Levine, J. H. (1979). Joint-space analysis of "pick-any" data: Analysis of choices from an unconstrained set of alternatives. *Psychometrika*, *44*(1), 85–92. doi: 10.1007/BF02293787

Mawer, D. (2000). *The Cambridge Companion to Ravel.* Cambridge, United Kingdom: Cambridge University Press.

McGonigal, J. (2005). *Down the rabbit hole.* Berkeley, CA, USA. Retrieved from `http://www.avantgame.com/McGonigal_down%20the%20rabbit%20hole_050505.pdf`

Milgram, S. (1967). The Small-World Problem. *Psychology Today*, *1*(1), 61–67.

Molka-Danielsen, J., Trier, M., Slykh, V., Bobrik, A., & Nurminen, M. I. (2007, 8). IRIS (1978-2006) Historical Reflection through Visual Analysis. In *Proceedings of IRIS (Information Systems Research in Scandinavia) Conference.* Tampere, Finland. Retrieved from `http://www.commetrix.de/iris/iris.html`

Moore, J. F. (1993). Predators and prey: a new ecology of competition. *Harvard Business Review*, *71*(3), 75–86.

Moran, E. F. (1990). *The Ecosystem Approach in Anthropology: From Concept to Practice.* University of Michigan Press.

Moreno, J. L. (1953). *Who Shall Survive?: Foundations of Sociometry, Group Psychotherapy and Sociodrama.* Beacon, NY, USA: Beacon House Inc. Retrieved from `http://www.asgpp.org/docs/WSS/WSS.html`

Nahapiet, J., & Ghoshal, S. (1998). Social Capital, Intellectual Capital, and the Organizational Advantage. *The Academy of Management Review*, *23*(2), 242–266. doi: 10.2307/259373

Naukkarinen, J. K. (2015). *What Engineering Scientists Know and How They Know It. Towards Understanding the Philosophy of Engineering Science in Finland.* Retrieved from `http://URN.fi/URN:ISBN:978-952-15-3641-0`

Navarro, G. (2001). A Guided Tour to Approximate String Matching. *ACM Computing Surveys*, *33*(1), 31–88. doi: 10.1145/375360.375365

Newman, M. E. J. (2001). The structure of scientific collaboration networks. *Proceedings of the National Academy of Sciences*, *98*(2), 404–409. doi: 10.1073/pnas.98.2.404

North, C. (2006, 5). Toward measuring visualization insight. *IEEE Computer Graphics and Applications*, *26*(3), 6–9. doi: 10.1109/MCG.2006.70

Nunamaker, J. F. J., & Briggs, R. O. (2011). Toward a Broader Vision for Information Systems. *ACM Transactions on Management Information Systems*, *2*(4), 20:1–20:12. doi: 10.1145/2070710.2070711

Nykänen, O., Salonen, J., Haapaniemi, M., & Huhtamäki, J. (2008). A Visualisation System for a Peer-to-Peer Information Space. In *Proceedings of the 2nd International OPAALS Conference on Digital Ecosystems: OPAALS 2008, October 7-8, Tampere, Finland* (pp. 76–85). Tampere, Finland: Tampere University of Technology. Retrieved from `http://urn.fi/URN:NBN:fi:tty-201104153766`

Oh, D.-S., Phillips, F., Park, S., & Lee, E. (2016). Innovation ecosystems: A critical examination. *Technovation*, *54*, 1–6. doi: 10.1016/j.technovation.2016.02.004

Olson, R. (2008, 7). Great Apps Using The CrunchBase API. *TechCrunch*. Retrieved from `http://techcrunch.com/2008/07/27/great-apps-using-the-crunchbase-api/`

Olsson, T. (2012). *User Expectations and Experiences of Mobile Augmented Reality Services* (Doctoral dissertation, Tampere University of Technology). Retrieved from `http://urn.fi/URN:ISBN:978-952-15-2953-5`

Paalanen, A. (2015). *Palkeen kieli: Vaihtoäänisen haitarin paljerytmiikka sävellystyössä* (Doctoral Dissertation). Sibelius Academy of the University of the Arts Helsinki.

Page, L., Brin, S., Motwani, R., & Winograd, T. (1999). *The PageRank Citation Ranking: Bringing Order to the Web* (Tech. Rep.).

Pan, W., Altshuler, Y., & Pentland, A. (2012). Decoding Social Influence and the Wisdom of the Crowd in Financial Trading Network. In *Proceedings of 2012 International Conference on Social Computing* (pp. 203–209).

Paquin, R. L., & Howard-Grenville, J. (2013). Blind Dates and Arranged Marriages: Longitudinal Processes of Network Orchestration. *Organization Studies*, *34*(11), 1623–1653. doi: 10.1177/0170840612470230

Parise, S., Whelan, E., & Todd, S. (2015). How Twitter Users Can Generate Better Ideas. *MIT Sloan Management Review*, *56*(4).

Peffers, K., Tuunanen, T., Rothenberger, M., & Chatterjee, S. (2007). A Design Science Research Methodology for Information Systems Research. *Journal of Management Information Systems*, *24*(3), 45–77. doi: 10.2753/MIS0742-1222240302

Peirce, C. S. (1958a). *Collected Papers, Vol. 1–6* (C. Hartshorne & P. Weiss, Eds.). Cambridge, MA: Harvard University Press.

Peirce, C. S. (1958b). *Collected Papers, Vol. 7–8* (A. Burks, Ed.). Cambridge, MA: Harvard University Press.

Peng, R. D. (2009). Reproducible research and Biostatistics. *Biostatistics*, *10*(3), 405–408. doi: 10.1093/biostatistics/kxp014

Peng, R. D. (2011). Reproducible Research in Computational Science. *Science*, *334*(6060), 1226–1227. doi: 10.1126/science.1213847

Pentland, A. S. (2015). *Social Physics: How Social Networks Can Make Us Smarter*. New York, New York, USA: Penguin Books.

Petschulat, S. (2010). Other people's data. *Communications of the ACM*, *53*(1), 53. doi: 10.1145/1629175.1629196

Pirolli, P., & Card, S. (2005). The sensemaking process and leverage points for analyst technology as identified through cognitive task analysis. In *Proceedings of international conference on intelligence analysis.*

Ramaswamy, V. (2009, 3). Leading the transformation to co-creation of value. *Strategy & Leadership*, *37*(2), 32–37. doi: 10.1108/10878570910941208

Ries, E. (2011). *The Lean Startup*. USA: Crown Business.

Ritala, P., Agouridas, V., Assimakopoulos, D., & Gies, O. (2013). Value creation and capture mechanisms in innovation ecosystems: a comparative case study. *International Journal of Technology Management*, *63*(3/4), 244–267.

Ritala, P., Armila, L., & Blomqvist, K. (2009). Innovation orchestration capability — Defining the organizational and individual level determinants. *International Journal of Innovation Management*, *13*(04), 569–591. doi: 10.1142/S136391960900242X

Ritala, P., & Hallikas, J. (2011). Network position of a firm and the tendency to collaborate with competitors – a structural embeddedness perspective. *International Journal of Strategic Business Alliances*, *2*(4), 307–328. doi: 10.1504/IJSBA.2011.044859

Ritala, P., & Huizingh, E. (2014). Business and network models for innovation: strategic logic and the role of network position. *International Journal of Technology Management*, *66*(2), 109–119. doi: 10.1504/IJTM.2014.064608

Ritala, P., & Hurmelinna-Laukkanen, P. (2009). What's in it for me? Creating and appropriating value in innovation-related coopetition. *Technovation*, *29*(12), 819–828. doi: 10.1016/j.technovation.2009.07.002

Rubens, N., Russell, M., Perez, R., Huhtamäki, J., Still, K., Kaplan, D., & Okamoto, T. (2011). Alumni network analysis. In *Proceedings of 2011 IEEE Global Engineering Education Conference, EDUCON 2011, Amman, Jordan* (pp. 606–611). IEEE. Retrieved from `http://dx.doi.org/10.1109/EDUCON.2011.5773200` doi: 10.1109/EDUCON.2011.5773200

Rubens, N., Still, K., Huhtamäki, J., & Russell, M. G. (2010, 2). *Leveraging Social Media for Analysis of Innovation Players and Their Moves* (Tech. Rep.). Media X at Stanford University.

Ruoho, I., Kuusipalo, J., Vainikka, E., & Huhtamäki, J. (2016). Politics on Twitter: Twitter Networks of Political Elite and Political Journalists in Finland. *European Journal of Communication*, *Submitted*.

Russell, M. G., Huhtamäki, J., Still, K., Rubens, N., & Basole, R. C. (2015). Relational Capital for Shared Vision in Innovation Ecosystems. *Triple Helix: A Journal of University-Industry-Government Innovation and Entrepreneurship (THJI)*. doi: 10.1186/s40604-015-0017-2

Russell, M. G., Still, K., & Huhtamäki, J. (2015). Visual tools to support innovation development: user experiences from the Parisian ecosystem. In *Proceedings of International Forum on Knowledge Asset Dynamics, June 10-12, 2015, Bari, Italy.*

Russell, M. G., Still, K., Huhtamäki, J., Yu, C., & Rubens, N. (2011). Transforming Innovation Ecosystems through Shared Vision and Network Orchestration. In *Proceedings of Triple Helix IX International Conference: "Silicon Valley: Global Model or Unique Anomaly?", July 2011, Stanford, California, USA* (p. 17).

Salonen, J., Huhtamäki, J., & Nykänen, O. (2013). Challenges in Heterogeneous Web Data Analytics - Case Finnish Growth Companies in Social Media. In *17th International Academic MindTrek Conference 2013: "Making Sense of Converging Media", October 1-3, Tampere, Finland* (pp. 131–138). ACM. Retrieved from `http://urn.fi/URN:NBN:fi:tty-201312191527` doi: 10.1145/2523429.2523481

Sayer, A. (2010). *Method in Social Science: A realist approach.* Routledge.

Schilling, M. A. (2009). Understanding the alliance data. *Strategic Management Journal*, *30*(3), 233–260. doi: 10.1002/smj.731

Schoemaker, P. J. H. (1995). Scenario Planning: A Tool for Strategic Thinking. *Sloan Management Review*, *36*(2), 25–40.

Schumpeter, J. A. (1934). *The Theory of Economic Development: An Inquiry Into Profits, Capital, Credit, Interest, and the Business Cycle.* Transaction Publishers.

Schumpeter, J. A. (1942). *Capitalism, Socialism and Democracy.*

Schumpeter, J. A. (1950). The March Into Socialism. *The American Economic Review*, *40*(2), 446–456.

Schwaber, K., & Beedle, M. (2001). *Agile Software Development with Scrum.* New Jersey: Prentice Hall.

Sedlmair, M., Isenberg, P., Baur, D., & Butz, A. (2011). Information visualization evaluation in large companies: Challenges, experiences and recommendations. *Information Visualization*, *10*(3), 248-266. doi: 10.1177/1473871611413099

Sein, M., Henfridsson, O., Purao, S., Rossi, M., & Lindgren, R. (2011). Action Design Research. *MIS Quarterly*, *35*(1), 37–56.

Shenton, A. K. (2004). Strategies for ensuring trustworthiness in qualitative research projects. *Education for Information*, *22*, 63–75.

Shneiderman, B. (1996). The Eyes Have It: A Task by Data Type Taxonomy for Information Visualizations. In *Proceedings of IEEE Symposium on Visual Languages, Boulder, Colorado, USA* (pp. 336–343). IEEE Computer Society.

Shneiderman, B. (2014). The Big Picture for Big Data: Visualization. *Science*, *343*(6172), 730–730. doi: 10.1126/science.343.6172.730-a

Simon, H. A. (1969). *The Sciences of the Artificial.* Cambridge, MA: MIT Press.

Smith, M. A., Himelboim, I., Rainie, L., & Shneiderman, B. (2015). The Structures of Twitter Crowds and Conversations. In S. A. Matei, M. G. Russell, & E. Bertino (Eds.), (pp. 67–108). Springer International Publishing.

Smith, S. P., & Johnston, R. B. (2014). How Critical Realism Clarifies Validity Issues in Information Systems Theory-Testing Research. *Scandinavian Journal of Information Systems*, *26*(1), 5–28.

Still, K., Huhtamäki, J., & Russell, M. G. (2013). Relational Capital and Social Capital: One or two Fields of Research? In *Proceedings of the 10th International Conference on Intellectual Capital, Knowledge Management and Organisational Learning, 24-25 October 2013, The George Washington University, Washington, DC, USA* (pp. 420–428).

Still, K., Huhtamäki, J., & Russell, M. G. (2014). Ecosystemic relational capital: framework and process for measuring it. In *Proceedings of the XII Triple Helix International Conference 2014 (THC 2014), September 11-13, Tomsk, Russia.*

Still, K., Huhtamäki, J., Russell, M. G., Basole, R. C., Salonen, J., & Rubens, N. (2013). Networks of innovation relationships: multiscopic views on Finland. In *Proceedings of the XXIV ISPIM Conference – Innovating in Global Markets: Challenges for Sustainable Growth, June 16-19, 2013, Helsinki, Finland* (p. 15).

Still, K., Huhtamäki, J., Russell, M. G., & Rubens, N. (2012). Paradigm shift in innovation indicators: from analog to digital. In *Proceedings of the 5th ISPIM Innovation Forum, 9-12 December, 2012, Seoul, Korea.* Seoul, Korea.

Still, K., Huhtamäki, J., Russell, M. G., & Rubens, N. (2014). Insights for orchestrating innovation ecosystems: the case of EIT ICT Labs and data-driven network visualisations. *International Journal of Technology Management*, *66*(2-3), 243–265. Retrieved from `http://www.inderscienceonline.com/doi/abs/10.1504/IJTM.2014.064606` doi: 10.1504/IJTM.2014.064606

Still, K., Russell, M. G., Huhtamäki, J., Turpeinen, M., & Rubens, N. (2011). Explaining Innovation with Indicators of Mobility and Networks: Insights into Central Innovation Nodes in Europe. In *Proceedings of Triple Helix IX International Conference: "Silicon Valley: Global Model or Unique Anomaly?", July 2011, Stanford, California, USA.*

Telea, A. C. (2008). *Data visualization: principles and practice.* Wellesley, Mass.: A K Peters.

Terras, M. (2012). The verdict: is blogging or tweeting about research papers worth it? *Impact of Social Sciences blog, London School of Economics*. Retrieved from `http://blogs.lse.ac.uk/impactofsocialsciences/2012/04/19/blog-tweeting-papers-worth-it/`

Thomas, J. J., & Cook, K. A. (2006). A visual analytics agenda. *IEEE Computer Graphics and Applications*, *26*(1), 10–13. doi: 10.1109/MCG.2006.5

Thomas, L. D. W., & Autio, E. (2012). Modeling the ecosystem: A meta-synthesis of ecosystem and related literatures. In *DRUID 2012, June 19-21, CBS, Copenhagen, Denmark.* (p. 28).

Tufte, E. (1983). *Visual Display of Quantitative Information.* Graphics Press.

Turpeinen, M. (2011). What Kind of Questions in Innovation Ecosystems Insist Visualization? Mobility Analysis in the Context of European Innovation Ecosystem. In *Snapshots, Movies and Interactive Tools: Analyzing and Communicating the Power of Relationships in Innovation Ecosystems (of Digital Media), Mindtrek 2011.* Tampere, Finland. Retrieved from `http://www.mindtrek.org/2011/visualizing-innovation-ecosystems`

Vainikka, E., & Huhtamäki, J. (2015). Tviittien politiikkaa – poliittisen viestinnän sisäpiirit Twitterissä. *Media & viestintä*, *38*(3), 165–183. Retrieved from `http://www.mediaviestinta.fi/arkisto/index.php/mv/article/view/299/282`

Vaishnavi, V. K., & Kuechler, W. (2007). *Design Science Research Methods and Patterns: Innovating Information and Communication Technology.* Boca Raton: Auerbach Publications.

Valkokari, K. (2015). Business, Innovation, and Knowledge Ecosystems: How They Differ and How to Survive and Thrive within Them. *Technology Innovation Management Review*, *5*(8), 17–24.

Van de Ven, A. H. (1986). Central Problems in the Management of Innovation. *Management Science*, *32*(5), 590–607. doi: 10.1287/mnsc.32.5.590

Vassiliadis, P. (2009). A Survey of Extract–Transform–Load Technology. *International Journal of Data Warehousing and Mining*, *5*(3), 1–27. doi: 10.4018/jdwm.2009070101

Vessey, I., & Galletta, D. (1991). Cognitive fit: An empirical study of information acquisition. *Information Systems Research*, *2*(1), 63–84. doi: 10.1287/isre.2.1.63

Vinoski, S. (2008). Serendipitous Reuse. *Internet Computing*, *12*(1), 84–87.

Wasserman, S., & Faust, K. (1994). *Social network analysis: Methods and applications* (Vol. 8). Cambridge university press.

Watts, D. J. (1999). Networks, Dynamics, and the Small-World Phenomenon. *American Journal of Sociology*, *105*(2), 493–527. doi: 10.1086/210318

Watts, D. J., & Strogatz, S. H. (1998). Collective dynamics of 'small-world' networks. *Nature*, *393*(6684), 440–442. doi: 10.1038/30918

Weber, M., & Hine, M. (2015). Who Inhabits a Business Ecosystem? The Technospecies as a Unifying Concept. *Technology Innovation Management Review*, *5*(5), 31–44. Retrieved from `http://timreview.ca/article/896`

Weick, K. E. (1988, jul). Enacted Sensemaking in Crisis Situations. *Journal of Management Studies*, *25*(4), 305–317. doi: 10.1111/j.1467-6486.1988.tb00039.x

Weick, K. E., Sutcliffe, K. M., & Obstfeld, D. (2005). Organizing and the Process of Sensemaking. *Organization Science*, *16*(4), 409–421. doi: 10.1287/orsc.1050.0133

Weiss, M., & Gangadharan, G. R. (2010). Modeling the mashup ecosystem: structure and growth. *R&D Management*, *40*(1), 40–49. doi: 10.1111/j.1467-9310.2009.00582.x

Wikipedia.org. (2015). *Web crawler.* Retrieved from `https://en.wikipedia.org/w/index.php?title=Web_crawler&oldid=670025903`

Williams, T. A., & Shepherd, D. A. (2015). Mixed Method Social Network Analysis Combining Inductive Concept Development, Content Analysis, and Secondary Data for Quantitative Analysis. *Organizational Research Methods*, 31. doi: 10.1177/1094428115610807

Wong, P. C., & Thomas, J. (2004). Visual Analytics. *IEEE Computer Graphics and Applications*, *24*(5), 20 - 21. doi: 10.1109/MCG.2004.39

Yang, J., & Leskovec, J. (2014). Overlapping Communities Explain Core-Periphery Organization of Networks. *Proceedings of the IEEE*, *102*(12), 1892–1902. doi: 10.1109/JPROC.2014.2364018

Yin, R. K. (2003). *Case Study Research: Design and Methods* (3rd ed.). Thousand Oaks, California: SAGE Publications, Inc.

Publications

Publication I

Jukka Huhtamäki, Ville Luotonen, Ville Kairamo, Kaisa Still, Martha G. Russell (2013) "Process for Measuring and Visualizing an Open Innovation Platform: Case Demola," *Academic MindTrek 2013, Tampere, Finland.*

Available at http://dx.doi.org/10.1145/2523429.2523478

Process for Measuring and Visualizing an Open Innovation Platform: Case Demola

Jukka Huhtamäki
Tampere University of Technology
Korkeakoulunkatu 3
33720 Tampere, Finland
+358 40 5854771
jukka.huhtamaki@tut.fi

Ville Luotonen
Hermia Oy, New Factory
Väinö Linnan aukio 15
33210 Tampere, Finland
+358 44 5741937
ville.luotonen@hermia.fi

Ville Kairamo
Hermia Oy, New Factory
Väinö Linnan aukio 15
33210 Tampere, Finland
+358405667182
ville.kairamo@hermia.fi

Kaisa Still
VTT Technical Research Centre of Finland
Kaitoväylä 1
90571 Oulu, Finland
+358 40 1508741
kaisa.still@vtt.fi

Martha G. Russell
mediaX at Stanford
Cordura Hall
210 Panama Street
Stanford, CA 94305-4115
+1 650 7231616
martha.russell@stanford.edu

ABSTRACT
Open innovation breaks the traditional pattern for developing new innovation leading to new business and the activities toward it. Consequently, new requirements are posed to innovation measurement. Demola is an open innovation platform that takes real-life problems from companies and other organizations and puts together and facilitates projects where students from different universities come together to solve the problems. This paper describes a set of network visualizations and animations that were developed in co-creation with the Demola operators to make visible the activity that Demola has initiated. Moreover, the development process used to design the visualizations and the technical process that was applied are described and discussed. We claim that static network visualizations and animations of an open innovation platform development are useful in presenting, describing, marketing and selling the platform for existing and new stakeholders. Our experience shows that in order to develop visualizations and animations that meet the requirements set by the different stakeholders, an iterative and incremental development process is needed. Moreover, we claim that taking a data-driven approach to visualization development is a key enabler in supporting the development.

Categories and Subject Descriptors
H.5.1 [**Information Interfaces and Presentations**]: Multimedia Information Systems – *animations.*

General Terms
Management, Measurement, Documentation, Performance, Design, Economics, Experimentation, Human Factors.

Keywords
Innovation ecosystems, co-creation, open innovation, information visualization, dynamic network analysis, animation, data-driven analytics.

1. INTRODUCTION
Carlson and Wilmot, among others, underline the critical nature of innovation: "Nothing is more important to business success than innovation" [6]. New approaches to innovation break the traditional patterns of in-house R&D. Innovation is nowadays seen to focus around customers, services and business models rather than solely around technology, in combining existing technologies and solutions with human knowledge resources from multiple sources, oftentimes outside of the boundaries of established companies. The multitude of theoretical approaches to innovation and the consequent paradigm shift place new demands for measuring innovation activities and their impact [14]. However, little research addresses the practical implications of creating and using these novel measures.

This paper describes the process of creating novel means of measuring open innovation developed for the context of an innovation ecosystem called Demola[1], an open innovation platform in Tampere, Finland. A Network Analysis and Visualization (NAV) process model [7] was applied and evaluated in a co-creative manner with the Demola team, resulting in network visualizations and animations that demonstrate the innovation activities and their impact.

1.1 Case Demola
Demola is an open innovation platform established in 2008 in Tampere, Finland. It puts together and facilitates innovation projects in which students from different universities, with backgrounds in different fields and cultures, come together with company representatives to solve company-initiated real-life

[1] Demola Tampere: http://tampere.demola.fi/

problems, challenges and new openings for the company portfolio [9][13]. By the beginning of 2013, 86 companies and 3 universities, with a total of about 1200 students, have participated in more than 250 projects. These projects are seen to energize the surrounding, larger ecosystem as they encourage the stakeholders to take alternative approaches to innovation work.

Nowadays, Demola is a part of Uusi Tehdas/New Factory[2], a platform for several initiatives supporting startups, innovation and business creation in different phases. Recently, new Demola sites have been opened in Oulu, Finland; Budapest, Hungary; Vilnius, Lithuania as well as in East and South Sweden. This study concentrates on Demola activities in Tampere, Finland. Demola projects from the other sites are excluded here as are all other connections Uusi Tehdas/New Factor has had a role in facilitating that are not directly related to Demola projects.

In Demola many of the traditional innovation metrics (changes in company turnover, the number of patents, companies or scientific publications created, or the amount of new product launches) cannot be easily tracked down to individual projects or even to the organizational level. In fact, many Demola stakeholders see these to be less relevant to the core activity. Still, Demola needs to communicate about its activities and impact internally as well as externally.

1.2 Network visualization process

As network visualization and animation tools have developed, interest in visual analysis of dynamic networks has increased [3][10]. More recently, the developers of an open source network visualization and exploration platform Gephi [2] have implemented functionalities that support dynamic network analysis and animation of network evolution over time. One of the more recent examples of visualizing network evolution, more specifically network construction, is the retweeter network of the Egyptian revolution that a developer was able to capture by incident [11]. The video has raised interest certainly among network analysis enthusiasts.

Whereas we draw from existing work on visual analytics and component-based data processing pipelines for visualization [11], we found the Network Analysis and Visualization (NAV) process model [7] to be a suitable framework to structure the process of this exploration. We see the NAV process model includes the steps of general information visualization reference model [5] that provides a framework for the technical process needed to make the process data-driven and reproducible in an automated manner.

According to the NAV model, the network analysis process starts from defining the goals of the analysis, after which data can be collected and structured. Next, an interpretation of the collected data is done by defining rules to transform the data into a network format. Producing an insightful network representation requires iterating over the steps of 1) laying out the nodes of the network, 2) possibly filtering the data and 3) adjusting the visual properties of the nodes and edges of the network. Different SNA metrics such as node degree can, for example, be used to define node size or color. Importantly, reaching a result that meets the requirements set by the different stakeholders of the visualization process insists on following iterative and incremental development process.

[2] Uusi Tehdas/New Factory: http://newfactory.fi/

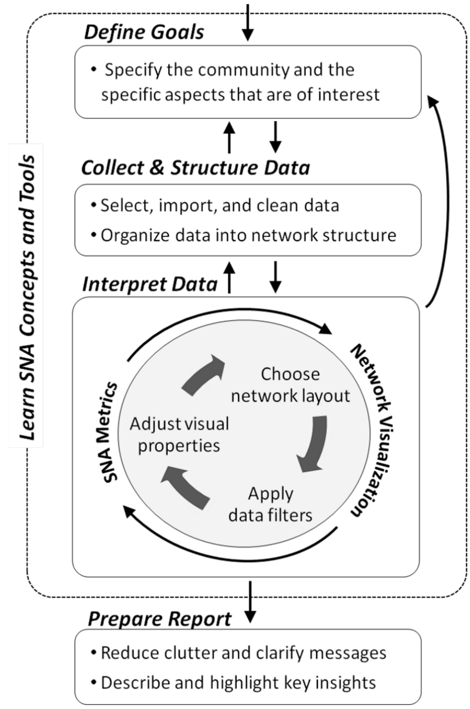


Figure 1. Network Analysis and Visualization Process Model [7].

2. Research method

As we have applied the visual network analytics paradigm for providing insights on innovation ecosystems in national [8] and European level [15][16] as well as within e.g. mobile domain [1], the network approach allowed us to reuse and refine our existing processes to the context of an individual open innovation platform.

In this research, an action research approach was followed to make an inventory of the key challenges that the members of the Demola team face in measuring and communicating about their innovation activities and the impact of those. The most part of visualization and animation development was conducted by a team of three including 1) a person with deep knowledge on Demola vision, mission and strategy, 2) a person with specific knowledge of the existing system used to manage project data, and, 3) a person with knowledge on applying visual network analytics for innovation ecosystem analysis and visualization.

2.1 Defining goals

Due to the interconnected nature of the Demola platform as a co-creation space between companies and teams students from different universities, taking a network approach for visualizing Demola activities was found to resonate with the operators and decision makers at Demola. Hence, in this study, we used NAV to develop new instruments for making the activities and impact more tangible and transparent to its different stakeholders, existing and new in the context of Demola.

Whereas the approach taken was seen promising also in providing value for operative tasks within the team running Demola, this study focuses in measurements and related visualizations that are targeted for stakeholders external to Demola. The target audience includes students, company representatives, university representatives, policy makers, without excluding general public.

Use cases for the visualizations include:

- Demola team member giving a presentation on Demola demonstrating existing Demola partners and processes. Audiences are heterogeneous and the specific needs vary accordingly. General introduction, marketing, sales. Requires “tailored storytelling” at best, for example pulling up and focusing into specific actors in the overall network, thus e.g. a fixed video is not an optimal solution.
- A student wishing to know more about Demola visits Demola website and plays a video showing the Demola process with real data on projects.
- A company representative is planning the first engagement with Demola. Browsing through the website, the person is interested in knowing more about the previous projects, the companies involved, the types of students participated and about the topics tackled.
- A policy maker is interested in the impact that Demola has had to the surrounding ecosystem. He or she visits the website, plays a video and makes an appointment with a Demola representative to discuss the specifics of the insights of the dynamics that the video provides. During the meeting, the dynamics are investigated in detail.

As the use cases demonstrate, several requirements are posed to the visualizations and animations going beyond individual static snapshots of the actor networks and even ready-made animations available in video format.

2.2 Collect & Structure Data: Demola Projects

Demola runs a dedicated Drupal-based web-based platform for setting up new projects as well as for running existing ones. During the first planning sessions, it became evident that Demola already collects and produces a useful data trail on projects.

Table 1 shows the structure of the data. As is often the case, new usage scenario for any set of data poses additional requirements. The data schema remained in practice the same from the beginning of the process but particularly the enumerated values for project key areas had to be harmonized over the course of the development project.

Table 1. Project data example

Project Detail	Example
Project Id	Project 115
Name	Koukkuniemi 2020
Started	2010-05-04
Ended	2010-10-31
Status	Completed
Collaboration Partner	City of Tampere
Type of Partner	Public
Project Domain	Non-profit
Location	Tampere
Key Areas	well-being, knowledge management, regional studies
Project Team Members	uta, uta, tut, tut

The start and end times of the project enable temporal analysis. Project status allows filtering in only project that are completed without loosing information on projects that e.g. were proposed but never started. Key areas field includes a comma-separated list of the areas that Demola operators have assigned for a project. The areas are selected from a curated list of domains for more specific semantics. Project team members are anonymized but their university affiliations are kept through listing the each university as many times as there are team members affiliated with a given university.

In this case, project data was the sole raw data used. The data was exported from the Drupal-based Demola platform with a tailored batch script and serialized in CSV (Comma Separated Values) format for further processing in a harmonization process. While less error-prone and more expressive formats for representing the data exist, CSV allows the use of general spreadsheet processors and other analysis software for managing and refining the data. A team of two Demola operators that were familiar with the origins of Demola as well as its evolution over the years conducted the harmonization process with simply using their collective recall and a spreadsheet processor as the refinement tool. Missing timestamps as well as some other inconsistencies were also fixed.

2.3 Interpret Data: Project Networks

While we realize that the project data available allows various kinds of analysis, the network approach was selected as the sole approach for this particular study. Projects, collaboration partners, project team members and their universities are all intuitive candidates to be used as network nodes. In addition, we decided to use nodes for representing the project key areas.

Whereas the NAV process model leads to the creation of a report of the network analysis results, our aim was set to developing static, interactive and animated network visualizations for a set of audiences with particular requirements. More specifically, our main objective was to develop views that allow insights on the immediate impact that Demola has had through its projects.

The technical implementation of the visualizations is an interplay between tailored code and the use of pre-existing tools. Whereas the technical process for creating the visualizations and animations is simplified, it does follow the logical steps of the information visualization reference model [5]. The model defines four steps for information process: First, Raw Data is collected. Second, Raw Data is refined into Data Tables that allow straighforward processing. Third, Data Tables are transformed into Visual Structures from which, finally, Views are created for representing the data. Importantly, the model states that at best, a visualization user should be able to interact with all the four steps of the process.

To create the Data Tables, data was exported from the spreadsheet processor used to refine the data in CSV format and a simple Python script was implemented to parse the data for further processing. After the refinement process, data was ready for the creation of Visual Structures, here networks. The interpretation rules were implemented in Python and NetworkX[3], an expressive Python library for analysis of complex networks, was used to help in constructing the network representations. The Python script is set to serialize the constructed network into files following Graph Exchange XML Format, in short GEXF, allowing e.g. the representation of dynamic networks.

For network visualization, i.e. the final part of the information visualization process, View Creation, we used Gephi. Gephi is an open source platform for explorative network analysis and visualization [2]. Gephi developers' original object to develop "the Photoshop for networks" has extended to include dynamic network analysis as one of its key features. Network visualization is an interactive process in which, as the NAV process model suggests, network layout, data filtering and the adjustment of visual properties are applied iteratively to create insightful views into the network. As its rapidly increasing popularity[4] indicates, Gephi provides all of the key functionalities required for network visualization.

2.4 Preparing Report: Static and Dynamic Visualizations

As result of the iterative and incremental co-creation process, a set of static network visualizations and a dynamic animation of network evolution were created.

2.4.1 Static Network Visualizations

Instead of creating one network including all the possible combinations between different types of nodes, two separate networks were constructed.

Firstly, the Project Network (Figure 2) is composed of companies, project and universities. It includes the three universities at its core. Each project is represented as a node and connected to all the universities from where project team members are from. Companies are connected to each project that they are involved in.

Network metrics are used to highlight features of the network. Node sizes indicates its connecting role for the whole network: the size is defined on basis of node betweenness value, i.e. the number of times a shortest path from all network nodes to all others goes through a particular node. Edge width shows the weight of the connection, i.e. the number of students that participated in a project from a given university.

A force-driven layout algorithm is used to define the position of each node in the network. The basic principle of force-driven layout is simple: nodes are programmed to repel each other and connections between nodes act as springs pulling them towards each other. In the resulting view, nodes that have the most interconnections are often placed close each other, thus revealing the overall structure of the network. The overall structure of the network is further highlighted through node colors showing the cluster of nodes in the network that particular node is included in with the exception that company nodes are always light green. Gephi's implementation of the modularity algorithm [4] is used for cluster analysis.

Secondly, the Domain Network (Figure 3) also starts from the three universities. Again, universities are connected to project nodes through project team member affiliations. Key areas are represented as nodes and connected to each project that has mentioned them. Finally, collaboration partners are connected to projects that they have been affiliated with.

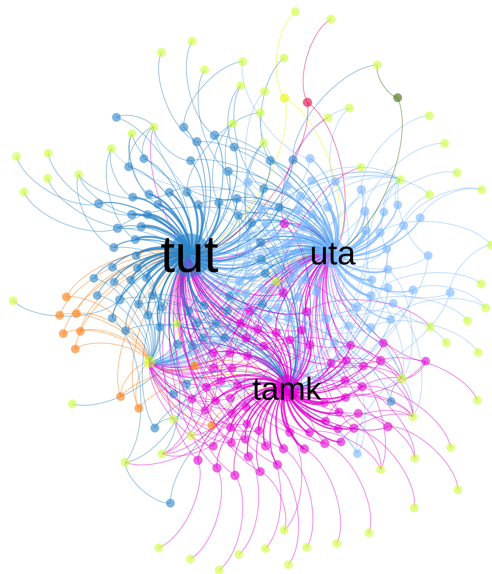


Figure 2. Demola Project Network.

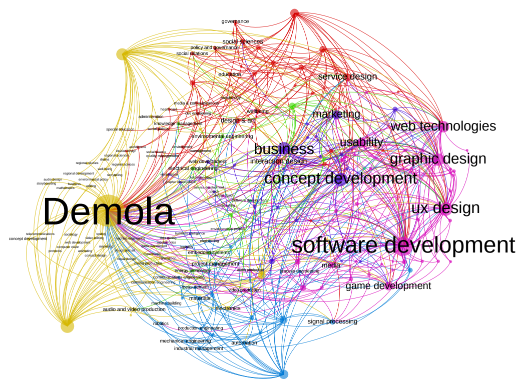


Figure 3. Demola project Domain Network.

2.4.2 Animation

An animation showing the current situation of the project network was also created. The project network is partly cumulative, partly showing a situation view. Connections between projects and universities only include the time when the project starts, thus projects are cumulated around the universities over time. The

---

[3] NetworkX is available at http://networkx.github.io/

[4] Google Trends for Gephi: http://www.google.com/trends/explore#q=gephi

connections between companies and projects use both the start and end dates. This means that, together with force driven layout algorithm, company nodes are pulled towards the center of the ecosystem during the project and start drifting away once a project engagement is finished.

A snapshot of the resulting animation is shown in Figure 4. Gephi allows network animation through two key features: First, it implements a timeline component with play functionality. Second, it allows graph layout algorithms to be run simultaneously while playing the timeline. Capturing the video was done with screen recording software. Post-production was required to include a timeline component into the video in an elegant manner.

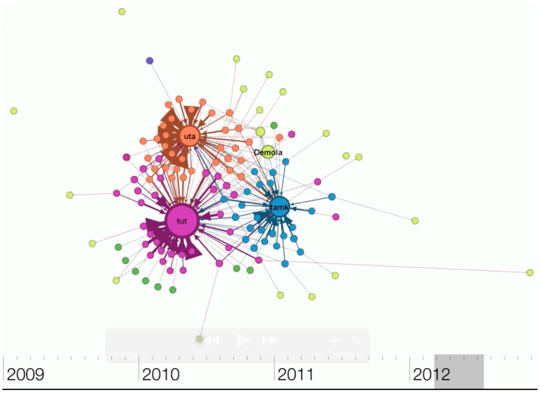


Figure 4. Snapshot of the animated project network.

3. DISCUSSION

In this paper, we used the Network Analysis and Visualization (NAV) process model for supporting the measurement of an innovation ecosystem. The resulting artifacts, i.e. the static visualizations as well as the animation, were all created in a co-creative manner with members of the case context, Demola.

Based on the feedback received during the co-creation process, we claim that static network visualizations and dynamic animations of an open innovation platform structure and evolution are useful in presenting, describing, marketing and selling the platform for existing and new stakeholders. As evidence, we offer the fact that a decision has already being made to start using the animated project network as a tool for communicating Demola activities and their evolution over time. Also the international collaborators of Demola have indicated their interest in using the tool to facilitate the discussions with their stakeholders.

From a technical viewpoint, the study allowed the following observations:

- Moving between tabular and structured format insists custom code development. Data-driven information visualization allows automation of the process but during the prototyping phase, we found interactive computing to be a more suitable paradigm to frame the development.
- The current implementation of dynamic network analysis in Gephi is rudimentary. To reach the level that e.g. software control management visualization tool Gource[5] allows for animation developers, additional work is required for developing Gephi further. Plans exist, already: https://gephi.org/2013/rebuilding-gephis-core-for-the-0-9-version/

We acknowledge the fact that more thorough user studies are required to evaluate the utility of the developed visualizations and to define specific steps to develop them further. More work is also needed to be able to show the change in the way that companies change their thinking over several Demola engagements, something that Demola operators have first-hand experience in. While the role of software development, for example, appears to be central in the Domain Network, the operators' experience shows that many companies start with software development (on prototype level) but continue to propose projects with a more cross-disciplinary framing. A way to take steps towards showing the shift in companies' thinking includes, for example, measuring and showing how the key areas of the projects change over time for a company. In addition, future possibilities include:

- Constructing steps from the Demola ecosystem through New Factory to national, European and global levels of the innovation ecosystem.
- Creating specific visualizations for different purposes / target groups: e.g. marketing, internal CRM, reporting of results.
- Creating visualizations from different perspectives, such as segments identified by industrial domains, university-oriented, expertise and skills –oriented.
- Interactive storytelling with data-driven, yet narrative-based views allowing real data driven, animated network view to the Demola ecosystem.
- Developing a real-time situation view implementing a fully automated, data-driven operation of the system to allow daily use of both the animatied and static cumulative views. For example, this requires solving the current requirement to create the layout through an interactive process manually with Gephi.

The NAV process model, coupled with the iterative and incremental approach taken in the process, was found to provide a useful task structure for the process of measuring and visualizing activities and the impact of an ecosystem through explicating the key steps required in the analysis. We found out that visualizations are useful in validating the source data. Disconnected or redundant nodes e.g. indicate errors in the data. Furthermore, the data-driven process and especially the flows of feedback built in the NAV process model did support the development of insightful visualizations and animations. In those discussions, we observed how the iterative detailed specification of visualizations and animations required the availability of the prototypes. From the first iteration, the visualizations that were used to validate the data catalyzed discussions on further requirements for the visualizations and animations.

More generally, the relationships identified with network connections allowed a preview of potential alliances for collaboration that could be created through participation in Demola. These relationship resources extended the value of the business ideas, talented employees and captured markets. They

[5] Gource homepage: https://code.google.com/p/gource/

described the character of the current Demola impact as well as the potential of the continuing impact, as individuals in these relationships collaborate on the current and future projects.

The selected set of tools was found to be fit for prototyping, creating versions of the animated network that are of high-quality and engaging. Appealing to the more hedonistic qualities of the observers that are attached to the concept of user experience (in an arcade-like mode) was found to be difficult task with Gephi. Thus, in the future, we seek to find an approach stemming from game development to develop an animation player allowing interactive storytelling with high user experience.

4. SUMMARY

In this study, static and dynamic visualizations of the network representation of Demola, an open innovation platform and an ecosystem engager, were developed in co-creation between Demola representatives and researchers developing new approaches to use visual network analysis as a tool for measuring innovation. Particularly, an animation representing the Demola key activities was found to be useful and of interest to many of the stakeholders. While this kind of an approach to measure innovation is very different from the more traditional approaches, we see that it has potential in allowing shared insights on the dynamics of innovation activities long before their impact surfaces as new product releases, patents filed, publications accepted, startups created or venture capital funding collected.

5. ACKNOWLEDGMENTS

This research is sponsored by Tekes – the Finnish Funding Agency for Technology and Innovation through project Reino (Relational Capital for Innovative Growth Companies) and is done in collaboration with the Innovation Ecosystems Network (http://www.innovation-ecosystems.org/). The authors wish to extend their gratitude to Ville Ilkkala from Meanfish Ltd for his insightful feedback and support in developing the animations.

6. REFERENCES

[1] Basole, R.C., Russel, M.G., Huhtamäki, J. and Rubens, N. 2012. Understanding Mobile Ecosystem Dynamics: A Data-Driven Approach. *Proceedings of the 2012 International Conference on Mobile Business (ICMB 2012)* (Delft, Netherlands, Jun. 2012), 17–28.

[2] Bastian, M., Heymann, S. and Jacomy, M. 2009. Gephi: An Open Source Software for Exploring and Manipulating Networks. *Proceedings of the Third International AAAI Conference on Weblogs and Social Media* (San Jose, California, USA, May. 2009).

[3] Bender-deMoll, S. and McFarland, D.A. 2006. The Art and Science of Dynamic Network Visualization. *Journal of Social Structure*. 7, 2 (2006).

[4] Blondel, V.D., Guillaume, J.-L., Lambiotte, R. and Lefebvre, E. 2008. Fast unfolding of communities in large networks. *Journal of Statistical Mechanics: Theory and Experiment*. 2008, 10 (Oct. 2008), P10008.

[5] Card, S.K., Mackinlay, J. and Shneiderman, B. 1999. *Readings in Information Visualization: Using Vision to Think*. Morgan Kaufmann.

[6] Carlson, C.R. and Wilmot, W.W. 2006. *Innovation: The Five Disciplines for Creating What Customers Want*. Crown Business.

[7] Hansen, D., Rotman, D., Bonsignore, E., Milic-Frayling, N., Rodrigues, E., Smith, M. and Shneiderman, B. 2009. *Do You Know the Way to SNA?: A Process Model for Analyzing and Visualizing Social Media Data*. University of Maryland Tech Report: HCIL-2009-17.

[8] Huhtamäki, J., Russell, M.G., Still, K. and Rubens, N. 2011. A Network-Centric Snapshot of Value Co-Creation in Finnish Innovation Financing. *Open Source Business Resource*. (Mar. 2011), 13–21.

[9] Kilamo, T., Hammouda, I., Kairamo, V., Räsänen, P. and Saarinen, J.P. 2011. Applying Open Source Practices and Priciples in Open Innovation. *Open Source Systems: Grounding Research*. S.A. Hissam, B. Russo, M.G.M. Neto, and F. Kon, eds. Springer Berlin Heidelberg. 1–10.

[10] Molka-Danielsen, J., Trier, M., Slykh, V., Bobrik, A. and Nurminen, M.I. 2007. IRIS (1978-2006) Historical Reflection through Visual Analysis. (Tampere, Finland, Aug. 2007).

[11] Nykänen, O., Salonen, J., Haapaniemi, M. and Huhtamäki, J. 2008. A Visualisation System for a Peer-to-Peer Information Space. *Proceedings of OPAALS 2008: The 2nd International OPAALS Conference on Digital Ecosystems* (Tampere, Finland, Oct. 2008), 76–85.

[12] Panisson, A. 2011. The Egyptian Revolution on Twitter. *https://gephi.org/2011/the-egyptian-revolution-on-twitter/*. Accessed: 2013-05-08.

[13] Pippola, T., Poranen, T., Vuori, M., Kairamo, V. and Tuominiemi, J. 2012. Teaching Innovation Projects in Universities at Tampere. *Proceedings of the International Conference on Engineering and Education* (Turku, Finland, Jul.-Aug. 2012), 785–792.

[14] Still, K., Huhtamäki, J., Russell, M.G. and Rubens, N. 2012. Paradigm shift in innovation indicators—from analog to digital. *Proceedings of the 5th ISPIM Innovation Forum* (Seoul, Korea, Dec, 2012).

[15] Still, K., Huhtamäki, J., Russell, M.G. and Rubens, N. 2012. Transforming Innovation Ecosystems Through Network Orchestration: Case EIT ICT Labs. *Proceedings of the XXIII ISPIM Conference – Action for Innovation: Innovating from Experience* (Barcelona, Spain, Jun. 2012).

[16] Still, K., Russell, M.G., Huhtamäki, J., Turpeinen, M. and Rubens, N. 2011. Explaining Innovation with Indicators of Mobility and Networks: Insights into Central Innovation Nodes in Europe. *Proceedings of Triple Helix IX International Conference: "Silicon Valley: Global Model or Unique Anomaly?"* (Stanford, California, USA, Jul. 2011).

Publication II

Rahul C. Basole, Martha G. Russel, Jukka Huhtamäki, Neil Rubens (2012) "Understanding Mobile Ecosystem Dynamics: A Data-Driven Approach," *International Conference on Mobile Business (ICMB)*.

Available at http://aisel.aisnet.org/icmb2012/15/

UNDERSTANDING MOBILE ECOSYSTEM DYNAMICS: A DATA-DRIVEN APPROACH

Rahul C. Basole, Georgia Institute of Technology, Tennenbaum Institute, 760 Spring Street NW, Atlanta, GA, USA, rahul.basole@ti.gatech.edu

Martha G. Russell, Media X at Stanford University, 119 Cordura Hall, Stanford, CA, USA, martha.russell@stanford.edu

Jukka Huhtamäki, Tampere University of Technology, Hypermedia Laboratory, Tampere, Finland, jukka.huhtamaki@tut.fi

Neil Rubens, University of Electro-Communications, Knowledge Systems Laboratory, Tokyo, Japan, rubens@activeintelligence.org

Abstract

The mobile ecosystem consists of a heterogeneous and continuously evolving set of firms that are interconnected through a complex, global network of relationships. However, there is very little theoretical understanding how these networks emerge and evolve and no well-established methodology to study these phenomena. Traditional approaches have primarily utilized alliance data of relatively established firms; however, these approaches ignore the vast number of relevant ecosystem activities that occur at the personal, entrepreneurial, and university level. We argue and empirically illustrate that a data-driven approach, using both alliance and socially-curated datasets, can provide important complementary explanatory insights into the dynamics of the mobile ecosystem. We present our approach through two recently formed mobile ecosystem relationships – the strategic partnership between Nokia and Microsoft and Google's acquisition of Motorola Mobility. Our analysis is complemented using network visualization techniques. The paper concludes with implications and future research opportunities.

Keywords: mobile ecosystem, transformation, strategic alliances, socially-constructed data, data and knowledge engineering, visualization.

1 Introduction

The mobile ecosystem consists of a heterogeneous and continuously evolving set of firms that are interconnected through a complex, global network of relationships. These firms come from a variety of market segments, each providing unique value propositions (Basole, 2009). It is quite unlikely for a single market segment to deliver all products or services to end-consumers. In fact, value creation and delivery requires a careful orchestration between firms across segments (Basole and Karla, 2012; Dhanaraj and Parkhe, 2006). For instance, the massive rollouts and upgrades of cellular networks by mobile network operators are useless without devices that can fully leverage them. Similarly, smartphones would just be boxes with little or no value without a platform and platform-enabled applications (Basole and Park, 2012). App stores provide third-party developers ways to offer content and reach consumers. Co-creation is hence an essential ecosystem characteristic, because a continual realignment of synergistic relationships of talent, knowledge and resources is required for growth of the system and responsiveness to changing internal and external forces (Rubens et al., 2011).

However, there is very little theoretical understanding how ecosystems emerge and evolve (Ahuja et al., 2011). Methodological approaches to quantitatively study these transformation phenomena have usually focused on event sequences at single levels in the biotechnology sector (Owen-Smith and Powell, 2004), national innovation ecosystem (Huhtamäki et al., 2011), mobile applications (Basole and Karla, 2012), and knowledge-intensive industries (Iansiti and Richards, 2006). Research that would answer how ecosystem emerge and evolve depends on data. The collection of primary data for business network research is time-consuming and costly. There are several potentially complementary data sources – some proprietary, others publicly available and still emerging data, like social media - containing relevant stakeholder activity information. Often these data sources are disconnected and reflect different units and periodicity; they are rarely interoperable. In some instances there is overlap, in others they are complementary, and in others they provide different insights and even conflicting insights. How can researchers leverage the wealth of data available to make new insights into how the ecosystem emerges and evolves? Historically, data acquisition was a resource-intensive step in data-driven research; it was a scarce resource. Open access to online data has made data widely available. A key challenge has now become the qualification and choice of data for analysis.

Our paper contributes in several ways. From a practical perspective, it provides competitive intelligence and insights into the systemic behavior and outcomes of firms. Small companies want to identify opportunities. Large companies want to know what's around the corner - technology and innovation, about which competitors they should worry and with which collaborators they should partner. Together they want to learn who has succeeded, why and how long it took. Theoretically, our paper contributes to the understanding of what elements and processes shape the evolution and transformation of the mobile ecosystem. It also contributes to our understanding how large, disconnected, potentially complementary structured and unstructured datasets can best be handled for insight, exploration, and discovery and how ecosystem evolution can be visually represented.

2 Related Work

Our paper draws on three distinct, but interrelated literature streams: interfirm networks and ecosystems, socially-constructed and curated data, and visualization and visual analytics.

2.1 Interfirm Networks and Ecosystem

An ecosystem consists of interdependent firms that form symbiotic relationships to create and deliver products and services (Basole and Rouse, 2008; Dougherty and Dunne, 2011). The conceptualization of markets as ecosystems is a result of theoretical extensions of work in inventor networks (Powell and Giannella, 2009) and of interfirm networks, alliances, and innovation (Gulati, 1998; Moore, 1993;

Oliver, 1990). With the complexity of product and service development and markets becoming increasingly disintegrated vertically and horizontally, there have been both a need and opportunity for the creation of interfirm relations (Iansiti and Levien, 2004). The formation of networks and alliances has been found particularly beneficial in technology industries as it has allowed firms to share risks in development and have access to synergistic knowledge (Eisenhardt and Schoonhoven, 1996). Studies have shown that interfirm networks are an effective organizational form to improve firm performance, speed of innovation, and organizational learning (Ahuja, 2000; Gulati et al., 2000).

More recently, studies have adopted a complex networked systems perspective to examine why, when, and how interfirm networks and alliances form and change (Gulati et al., 2000). This view combines both the resource-dependency and embeddedness perspective and suggests that interfirm networks are complex systems characterized by co-evolving actors engaged in collaboration, coopetition (Iansiti and Levien, 2004) and collective invention (Powell and Giannella, 2009). The complex networked systems approach has also been used to study value network and ecosystems in a variety of industries (Basole and Rouse, 2008; Rosenkopf and Schilling, 2008).

2.2 Provenance in Socially-Constructed and Curated Data

In the current tsunami of data, the provenance of data is of critical importance. Curated data is perceived to have the advantages of consistent ontologies, predictable data gathering methods and consistently applied data-cleaning rules. With the standardized data practices and policies of curated data, analytical methods can become standardized, and interpretation of analytical results benefits from consistent comparisons and a shared understanding of metrics. These very advantages, however, bring with them some disadvantages. Bias becomes baked into the data policies. Categories and classifications sometimes persist in data practice long after real world semantics have shifted to new classifications or reformulated categories. The time required for the curation processes may introduce significant delays into the timeliness of even the most recently available curated data. Additionally, many curated databases have limited availability and access may be exclusive and/or very expensive.

While some have argued that data in and of itself has little meaning and that the knowledge (Borgman, 2007) and meaning of data (Smagorinsky, 1995) are inherently socially constructed, the social nature of the Internet has added a new data frontier – in socially constructed data. Extensive data about businesses is now openly available through company websites, published announcements and filings, blogposts and microblogging, and community-built information resources. These sources provide unprecedented access to data, updated in real-time. One of the firsts of its kind, Wikipedia established itself as the most reliable source of accurate information (Giles, 2005) because it invited additions and tracked the provenance of changes; a data source that is socially constructed has observable patterns of governance (Leskovec et al., 2010). Advantages include its open access and availability, potentially large coverage, timeliness, and community verification of data quality. Some of the disadvantages are the potential of incompleteness and inconsistencies, lack of established perspective, and the issue (although slightly different from that of curated data) of incompleteness and inconsistencies.

2.3 Visualization and Visual Analytics

While an analytical approach provides valuable insights to the structure and dynamics of ecosystems, important knowledge can also be gained through the visualization of complex ecosystem data. Contrary to the perception that visualizations are merely artistic approaches to depicting structure, they have been used to explore, interpret, and communicate data in order to aid humans in overcoming their cognitive limitations, making structure, patterns, relationships, and themes visible, and providing a means to efficiently comparing multiple representations of the same data in fields such as medicine, dentistry, computer science and engineering. It has been suggested that visualization approaches can be extremely valuable for understanding and analyzing business issues, including strategy, scenario planning, and problem-solving (Tufte, 1983).

One explanation for the relatively slow uptake of visualization technologies in organizational and management sciences may be that visualization of complex systems is not only a very challenging and difficult task and but also, if not developed, implemented or applied correctly, may lead to non-conclusive results. Particularly in visualizing temporal changes of business ecosystems, node-link configurations are not necessarily unique and results may be misleading. The boundary-setting problem, or inclusion of nodes, is often artificial. Conclusions based on these models must thus be carefully scrutinized for the possibility of alternative explanations. Along the same lines, the amount of information that is captured and presented can often be overwhelming to the end-user. In many instances, what and how ecosystem data is visualized depends not only on the nature of the data but also on the question that is being asked and ultimately the cognitive abilities of the user. In order to overcome the aforementioned challenges, researchers must therefore ensure a balance between detail, abstraction, accuracy, efficiency, perceptual tension, and aesthetics in their complex network visualizations (Segel and Heer, 2010). These observations highlight the importance of setting the context and defining the elements in an ecosystem visualization study very carefully.

3 Data

We explore the dynamics of the mobile ecosystem using two complementary types of data sources – SDC Platinum and the IEN Dataset. Because the validity of our results and insights depends heavily on the nature and quality of the datasets, we first describe those datasets and then explain our conceptual approach and present our empirical results.

3.1 SDC Platinum

The SDC database is one of the most prominent, comprehensive, and accurate commercial databases used in the study of global interfirm relationships across multiple sectors (Schilling, 2009). It has been used extensively in strategic management and the management and organization sciences (e.g. Hsu (2006); Sampson (2004); Schilling and Phelps (2007)). Alliances and inter-organizational relationships are thus only one aspect of this broad database. The SDC database contains information on joint ventures, strategic alliances, R&D agreements, sales and marketing agreements, supply and manufacturing agreements, and licensing and distribution data, curated from SEC filings, trade publications, wires and news sources. In addition, it provides access to 200+ additional data elements, including names, SIC codes and nationality of participants, and relationship terms and synopsis.

3.2 IEN Dataset

The Innovation Ecosystems (IEN) Dataset (Rubens et al., 2010) is a quarterly updated collection of socially constructed data about technology-oriented companies in the ICT fields and the service companies (legal, accounting, advertising) that support them. Drawn from press release type information on multiple websites that permit comment and correction, it includes data about more than 68,000 companies (including accounting, legal and marketing services firms, and includes a high proportion of startup companies), their executives and board personnel, investment organizations, and financial transactions. People included in the dataset are key individuals in their respective companies (e.g. founders, executives, lead engineers, etc.), members of boards of advisors, or investors. The dataset further includes background data of individuals (e.g. degrees and institutions).

3.3 The Complementarity of the Two Datasets

The utilization of both datasets promises enormous complementary value for the analysis of ecosystem dynamics. While the SDC Platinum database contains validated alliance information for primarily large, global, and public companies, the IEN dataset contains information about small, private companies and startups. As many innovation activities occur in entrepreneurial settings or at the people level, the IEN dataset thus fills in the "blanks" between major ecosystem events. In contrast to high-quality and validated SDC data, however, the IEN dataset also inherits both the advantages and disadvantages of socially constructed data. Some of the advantages are availability, large coverage, timeliness, and community verification of data quality. Some of the disadvantages are potentially erroneous data and public bias (vs. the editorial bias often extant in traditional data settings). A comparative summary of the two datasets is provided in Table 1.

Table 1. Comparison of Datasets

	SDC Platinum 4.0	**IEN Dataset**
Source	Proprietary (Thomson Reuters Financial) based on U.S. SEC data	Open-Source based on socially-curated data from news, press releases, and social media
Type of Data	Alliance data (strategic, R&D, marketing, manufacturing, licensing, and supply) and status (active, terminated, pending) of public and private firms (37 SIC Codes, 4-digit)	Relationship Data of Public and Private Firms, Financial Organizations, Educational Institutions, Funding Rounds, Acquisitions, Investments by Individuals and Companies
Years covered	1/1/1990 - 12/31/2011	1/1/1994 - 01/31/2012

4 Approach

We use a three-stage process for analyzing the dynamics of the mobile ecosystem, consisting of boundary specification, metrics identification and computation, and analysis and visualization.

4.1 Step 1: Boundary Specification

Boundary specification involves determining the primitives of the network architecture (Ahuja et al., 2011), including nodes, node types (e.g. firms, people, universities, etc.), and relationship types (e.g. R&D, supply chain, marketing, licensing, etc.) and specification of the desired analysis timeframe (e.g. start/end-date). The choice of these parameters is driven by the nature and intent of the problem.

The specification of nodes, however, is not a trivial task, as firms continuously enter and leave the ecosystem. If the analytical focus is on the evolution of a particular market segment, one may begin by considering all companies that operate in that market sector and the second level companies to which the selected first-level companies connect. This leads to a related decision concerning the number of third, fourth and subsequent levels of companies to include in the selected data. Which other companies should be included in the analysis (only those directly connected companies outside of the first-level market sector or companies connected *k*-steps from companies in the focal first-level market sector)? The larger *k* is (upper bound limit defined by the maximum k-steps of the graph), the more companies will be included. However, this expansion carries risks of diluting the analysis with potentially irrelevant companies. The smaller *k* is, however, the greater the risk of ignoring important companies that may be a few steps removed.

The specification of the appropriate timeframe is an equally challenging task. How far back in time does the data need to go in order to capture the events and activities that led to the alliance? In many instances, researchers either choose the largest timeframe available (e.g. the first activity for any of the companies involved in the alliance) or a particularly important or relevant point in time (e.g. announcement, product launch, policy decision). It is quite foreseeable that a singular event/activity did not necessarily cause the activity the researcher is trying to explain. It may have been a result of multiple events/activities that occurred in a particular order.

4.2 Step 2: Metrics Identification and Computation

There are many social network and graph theoretic metrics that can be useful for understanding the dynamics of an ecosystem. Broadly, these can be categorized at two levels of analysis – the whole network (ecosystem) and the node level (firm/individual). This differentiation is important because network dynamics at each level, although related, are also distinct (Zaheer et al., 2010). A description of representative metrics (e.g. (Ahuja et al., 2011) is provided in Table 2.

Table 2. Node and Network-Level Ecosystem Dynamics Metrics

Level	Metric	Description
Network	• Size	Change in the size of the network is reflective of the overall growth of the relevant ecosystem.
	• Degree Distribution	Change in the degree distribution is reflective of changes in the status hierarchy of the observed system.
	• Diameter	Change in the diameter is reflective of the connectivity or “small worldness” of the network.
	• Clustering	Change in clustering represents the reconfiguration of clusters or constellations of firms that may be competing against each other as alliance networks.
	• Density	Change in density (the proportion of ties that are realized in the network relative to the hypothetical maximum possible) represents how tightly the network is connected.
	• Degree Assortavity	Change in degree assortavity is reflective of the degree to which nodes with similar degrees connect to each other.
Node	• Degree	Change in the degree is reflective of the number of new connections a firm has gained or established.
	• Betweenness Centrality	Change in betweeness centrality measure is reflective of the positional prominence of a firm (node) in a network.
	• Cluster Coefficient	Change in the cluster coefficient is reflective of the level of connectivity between a firm’s directly connected partners.

4.3 Step 3: Analysis and Visualization

There are a number of ways analyzing and visualizing temporal data. One approach includes a tabular description of key metrics; another includes a timeline representation of changes in key network metrics. If multiple metrics want to be compared simultaneously and structural patterns matter more than specific metric levels, sparklines or small-multiples are a frequent choice. Ideally, an interactive, animated approach is required. Due to page constraints, we utilize a tabular representation and cumulative network visualization to depict the dynamics of the mobile ecosystem in the paper and provide an interactive representation online.

5 Illustrative Examples

We illustrate our data-driven approach to understanding mobile ecosystem dynamics with two recent examples. The visualizations represent a 2-step network using two layout algorithms: OpenORD to create clusters; ForceAtlas2 to aesthetically space nodes. Node and relationship types are differentiated by color (e.g. red=firms; green=investment firms; blue=people; purple=educational institutions).

5.1 Nokia and Microsoft

The alliance between Nokia and Microsoft in February 2011 was considered by many pundits to be an inevitable move given the recent struggles of both companies in the mobile ecosystem. Once a leader in the global handset market, Nokia has been falling behind other device manufacturers in the lucrative smartphone segment. Microsoft, a perennial leader in the desktop market, never really achieved any traction in the mobile market despite its Windows Mobile platform. Many attributed Microsoft’s shortcoming to a lack of an appropriate hardware partner. A collaboration became increasingly realistic when Nokia appointed Stephen Elop, a former Microsoft executive, as its next CEO in 2010.

Figure 1 shows the SDC alliance network of Microsoft and its partners, Nokia and its partners, and alliances between the partners. Microsoft and Nokia have direct relationships with 275 and 123 companies, respectively. Both firms have many second order relationships. The strength of ties between Microsoft and Nokia can be observed in their proximity to each other and in the thickness of the edges connecting the two major nodes.

Figure 2 shows the IEN network, which adds company leadership, investment firms, and educational institutions. The patterns of relationships among these constituents show multiple connections, with key individuals as critical nodes in the network of relationships. The importance of the personal network in creating relationship pathways between Microsoft and Nokia is visible. Stephen Elop, shown as the individual at the top of Figure 2, with direct connections to both Nokia and Microsoft, is not the sole relationship connection. The links to investment firms from Microsoft's second order companies creates a venture-influenced mega cluster. The cluster of companies around Nokia is less influenced by different investment firms. A few investment firms and their key people link Microsoft to Nokia. The multiple relationship pathways through which information, resources and talent can flow between Microsoft and Nokia reflect a multidimensional form of collaboration.

5.2 Google and Motorola Mobility

Google's proposed acquisition of Motorola Mobility in August 2011 received significant attention by players in the mobile ecosystem. Motorola Mobility had been struggling to (re)gain market share in the lucrative smartphone segment. Through various business transformations in recent years it had tried to reposition itself, but still failed to deliver on its past innovative pedigree. Contrary, Google - not a traditional mobile player - was speculated to enter the ecosystem full-force on many occasions. For instance, Google was a key bidder on wireless spectrum a few years back. More recently, Google was a key investor and creator of the Android mobile platform. However, there were no signs that Google would offer its own hardware.

Figure 3 shows the SDC alliance network of Google and its partners, Motorola Mobility and its partners, and alliance between the partners (both firms are both shown in the upper left area). Google and Motorola Mobility are directly connected to 16 and 22 companies, respectively. There is no direct alliance between the two firms. Interestingly, Microsoft dominates this network, due to the first-degree connections between Microsoft and both Motorola and Google and the second degree connections to the other strongly connected firms.

Figure 4 shows the IEN network. This network is characterized by a dense web of companies and investment firms – a venture network. Google shows two connections to investment firms – Sequoia and Google Ventures. Two observations are of particular interest: the lack of a node connecting Google with Motorola and the relative independence from investment firms for both. Motorola Mobility hangs on a connection to Motorola Solutions and links to this ecosystem with a connection to Vivotech, which has a connection to Draper Fisher Jurvetson, and a connection to a123systems, which has investment from Fisker and Sequoia Capital, which is connected to Google.

The relative isolation of Motorola from the Google and venture subnetworks is apparent in this graph. The sole link visible here is one individual who is connected to both Motorola and Motorola Mobility, and is also connected to another company that received investment from Sequoia Capital. Indirect pathways between Google and Motorola are created by the relationships of several individuals, but these appear to be relatively few, especially in contrast to the Microsoft-Nokia network.

The strong presence of Google in the IEN Dataset is highlighted by the fact that Google's absolute degree value is larger for IEN data even though the total network for IEN data is significantly smaller than for the SDC network, see Table 3. As indicated by data drawn from the IEN Dataset one could argue that this is due to Google's strategy of growing through acquiring small startups rather than forming alliances. Seen from a network perspective, the acquisition of Motorola Mobility by Google is more likely to be an event in which Motorola Mobility and its relationships are consumed by Google.

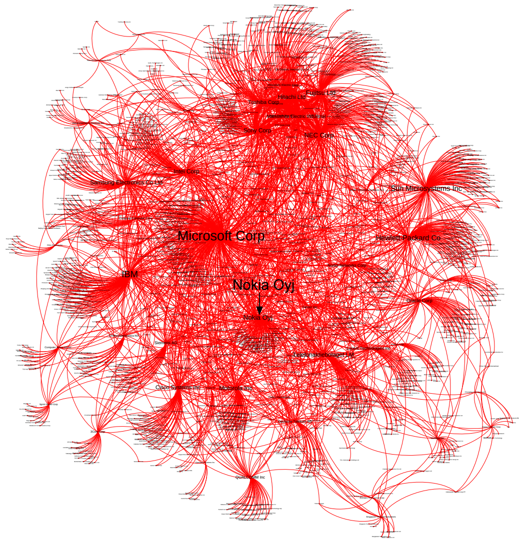


Figure 1. *Nokia & Microsoft -- Cumulative Network using SDC Alliance Data*

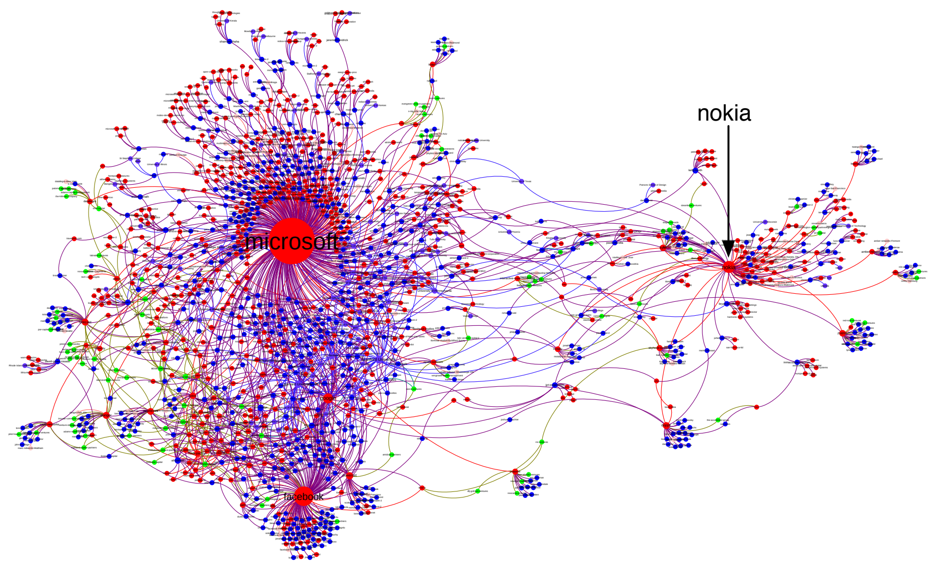


Figure 2. *Nokia & Microsoft -- Cumulative Network using IEN Data*

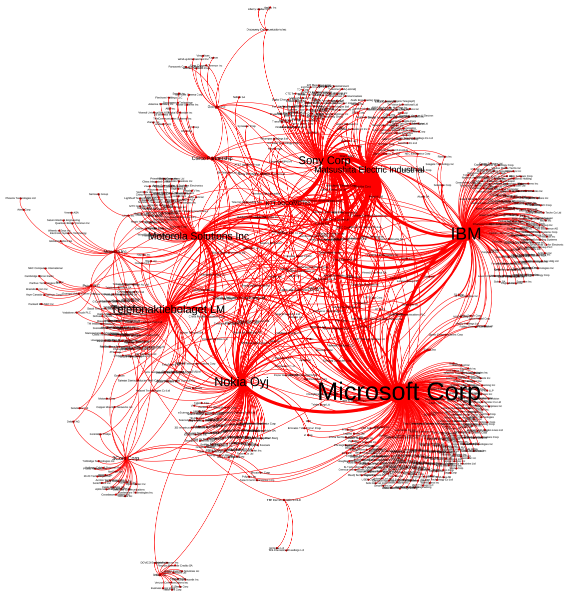


Figure 3. *Google & Motorola Mobility -- Cumulative Network using SDC Alliance Data*

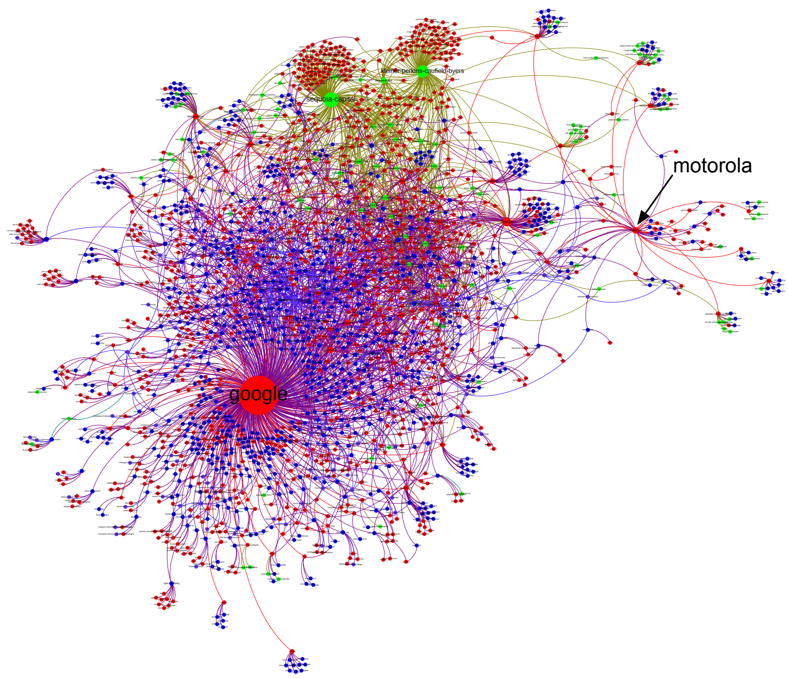


Figure 4. *Google & Motorola Mobility -- Cumulative Network using IEN Data*

Table 3. Nokia and Microsoft -- Comparison of Representative Ecosystem Dynamics Metrics

		SDC Data						IEN Data					
		6/2009	**12/2009**	**6/2010**	**12/2010**	**6/2011**	**12/2011**	**6/2009**	**12/2009**	**6/2010**	**12/2010**	**6/2011**	**12/2011**
Network	• Size	1,621	1,643	1,646	1,652	1,659	1,666	125	142	146	155	156	157
	• Diameter	4	4	4	4	4	4	4	4	4	4	4	4
	• Clustering	1	1	1	1	1	1	2	2	2	2	2	2
	• Density	0.0023	0.0023	0.0023	0.0023	0.0023	0.0023	0.016	0.014	0.014	0.13	0.013	0.013
Node	**Nokia Oyj**												
	• Degree	197	120	120	122	122	123	14	17	19	20	20	20
	• Betweeness Centrality	135193	135220	135197	138627	139011	140970	104	205	292	342	342	342
	• Clustering Coefficient	0.035	0.036	0.036	0.035	0.035	0.035	0	0	0	0	0	0
	Microsoft Inc.												
	• Degree	266	272	273	274	274	275	69	72	72	74	75	76
	• Betweeness Centrality	405609	416156	417728	419434	421013	424517	5624	6846	6846	7649	7775	7902
	• Clustering Coefficient	0.014	0.014	0.014	0.014	0.014	0.014	0	0	0	0	0	0

Table 4. Google and Motorola Mobility -- Comparison of Representative Ecosystem Dynamics Metrics

		SDC Data						IEN Data					
		6/2009	**12/2009**	**6/2010**	**12/2010**	**6/2011**	**12/2011**	**6/2009**	**12/2009**	**6/2010**	**12/2010**	**6/2011**	**12/2011**
Network	• Size	719	788	789	792	794	797	60	73	91	118	129	137
	• Diameter	6	6	6	6	6	6	4	4	4	4	4	4
	• Clustering	1	1	1	1	1	1	2	2	2	2	2	1
	• Density	0.0038	0.0036	0.0036	0.0036	0.0036	0.0036	0.033	0.028	0.022	0.017	0.016	0.015
Node	**Google Inc.**												
	• Degree	14	16	16	16	16	16	45	51	65	78	87	98
	• Betweeness Centrality	14200	16224	14725	9450	9691	9720	1416	2097	3432	5587	6812	9067
	• Clustering Coefficient	0.077	0.1	0.108	0.117	0.117	0.117	0	0	0	0	0	0
	Motorola Mobility												
	• Degree	22	22	22	22	22	22	3	4	4	6	6	7
	• Betweeness Centrality	8363	8881	8886	8953	8977	9011	5.0	9.0	9.0	33.0	33.0	795
	• Clustering Coefficient	0.087	0.087	0.087	0.087	0.087	0.087	0	0	0	0	0	0

Note: Due to page constraints, we did not include a detailed description of the degree distribution metric.

6 Concluding Remarks

This paper advocates a data-driven approach for understanding the dynamics of the mobile ecosystem. We illustrate our approach with an exploratory analysis of two recently formed relationships – Microsoft/Nokia and Google/Motorola Mobility – using two data sources. Our initial results show that each dataset has its advantages and disadvantages, but used jointly can reveal consistent patterns and create synergistic insights. The SDC dataset emphasizes deal-based relationships and does not include data about key individuals in the companies; the IEN dataset includes individuals and emphasizes the relationships formed among companies through those individuals' leadership activities. We think that the data-driven approach can provide important insights into patterns of event sequences between nodes for a particular type of event (e.g. R&D alliance) and the average duration it takes.

Many challenges and opportunities remain. Arguably the most foundational task is the careful integration of datasets. Datasets use different unique identifiers or naming conventions. Consequently, matching names and labels of firms or individuals across datasets is not a trivial task. Firm names may be inconsistent and use different enterprise labelling. As a result of mergers, acquisitions, or corporate restructuring, firms may also change names over time. Appropriate identification and matching algorithms to ensure consistency across datasets must therefore be developed. Another challenge is the selection and assignment of companies to market segments. Various industry classifications exist, but datasets often use different classification schemes. The identification of primary and secondary market segments is particularly challenging for large firms that operate in multiple and equally important segments. Intelligent market segment identification and assignment methods must therefore be developed. As firms transform or enter and exit the ecosystem it is critical to devise appropriate data persistency protocols by identifying events by time and actors involved.

Our study also provides the foundation to explore many interesting ecosystem issues including what relationship configurations characterize growth, how the position and role of firms in the ecosystem influences their access to talent, information, resources, what event windows and types are relevant for observing ecosystem dynamics and what sequences matter.

There are also many opportunities for creating appropriate representations of mobile ecosystem dynamics. This may include the development of an interactive visualization system using multiple views. The alignment and representation of time units at potentially different scales is an important representational aspect. While established datasets may capture large, less frequent events, socially-curated data may capture activities that occur in closer time intervals. Enabling a user-driven selection of time units will enable greater insight and discovery of the temporal nature of ecosystem activities.

References

Ahuja, G. (2000). Collaboration networks, structural holes, and innovation: A longitudinal study. Administrative Science Quarterly, 45 (3), 425-455.

Ahuja, G., Soda, G., and Zaheer, A. (2011). The genesis and dynamics of organizational networks. Organization Science (Article in Advance).

Basole, R.C. (2009). Visualization of interfirm relations in a converging mobile ecosystem. Journal of Information Technology, 24 (2), 144-159.

Basole, R.C., and Karla, J. (2012). Value transformation in the mobile service ecosystem: A study of app store emergence and growth. INFORMS Service Science, 4 (1), 1-18.

Basole, R.C., and Park, H. (2012). The evolution of smartphones and platform type preference. Georgia Tech, pp. 1-10.

Basole, R.C., and Rouse, W.B. (2008). Complexity of service value networks: Conceptualization and empirical investigation. IBM Systems Journal, 47 (1), 53-70.

Borgman, C.L. (2007). Scholarship in the digital age. MIT Press, Cambridge, MA.

Dhanaraj, C., and Parkhe, A. (2006). Orchestrating innovation networks. Academy of Management Review, 31 (3), 659-669.

Dougherty, D., and Dunne, D.D. (2011). Organizing ecologies of complex innovation. Organization Science, 22 (5), 1214-1223.

Eisenhardt, K.M., and Schoonhoven, C.B. (1996). Resource-based view of strategic alliance formation: Strategic and social effects in entrepreneurial firms. Organization Science, 7 (2), 136-150.

Giles, J. (2005). Internet encyclopaedias go head to head. Nature, 438 (1), 900-901.

Gulati, R. (1998). Alliances and networks. Strategic Management Journal, 19 (4), 293-317.

Gulati, R., Nohria, N., and Zaheer, A. (2000). Strategic networks. Strategic Management Journal, 21 (3), 203-215.

Hsu, D. (2006). Venture capitalists and cooperative start-up commercialization strategy. Management Science, 52 (2), 204-219.

Huhtamäki, J., Russell, M.G., Still, K., and Rubens, N. (2011). A network-centric snapshot of value co-creation in finnish innovation financing. Open Source Business Resource (March), 13-21.

Iansiti, M., and Levien, R. (2004). The keystone advantage: What new dynamics of business ecosystems mean for strategy, innovation, and sustainability. Harvard Business School Press, Boston, MA.

Iansiti, M., and Richards, G.L. (2006). The information technology ecosystem: Structure, health and performance. The Antitrust Bulletin, 51 (1), 77-110.

Leskovec, J., Huttenlocker, D., and Kleinberg, J. (2010). Governance in social media: A case study of the wikipedia promotion process. In Proceedings of the Fourth International AAAI Conference on Weblogs and Social Media, Washington, DC.

Moore, J.F. (1993). Predators and prey: A new ecology of competition. Harvard Business Review, 71 (3), 75-86.

Oliver, C. (1990). Determinants of interorganizational relationships: Integration and future directions. Academy of Management Review, 15 (2), 241-265.

Owen-Smith, J., and Powell, W.W. (2004). Knowledge networks as channels and conduits: The effects of spillovers in the boston biotechnology community. Organization Science, 15 (1), 5-21.

Powell, W.W., and Giannella, E. (2009). Collective invention and inventor networks, in Handbook of economics of invention, B.H. Hall and N. Rosenberg (eds.). Elsevier, Amsterdam, Netherlands, 2009.

Rosenkopf, L., and Schilling, M.A. (2008). Comparing alliance network structure across industries: Observations and explanations. Strategic Entrepreneurship Journal, 1 (3-4), 191-209.

Rubens, N., Still, K., Huhtamäki, J., and Russell, M.G. (2010). Leveraging social media for analysis of innovation players and their moves. Stanford University.

Rubens, N., Still, K., Huhtamäki, J., and Russell, M.G. (2011). A network analysis of investment firms as resource routers in the chinese innovation ecosystem. Journal of Software, 6 (9), 1737-1745.

Sampson, R.C. (2004). Organizational choice in r&d alliances: Knowledge-based and transaction cost perspectives. Managerial and Decision Economics, 25 (6-7), 421-436.

Schilling, M.A. (2009). Understanding the alliance data. Strategic Management Journal, 30 (3), 233-260.

Schilling, M.A., and Phelps, C.C. (2007). Interfirm collaboration networks: The impact of large-scale network structure on firm innovation. Management Science, 53 (7), 1113-1126.

Segel, E., and Heer, J. (2010). Narrative visualization: Telling stories with data. IEEE Transactions on Information Visualization & Computer Graphics, 16 (6), 1139-1148.

Smagorinsky, P. (1995). The social construction of data: Methodological problems of investigating learning in the zone of proximal development. Review of Educational Research, 65 (3), 191-212.

Tufte, E. (1983). The visual display of quantitative information. Graphics Press, Cheshire, CT.

Zaheer, A., Gözübüyük, R., and Milanov, H. (2010). It's the connections: The network perspective in interorganizational research. Academy of Management Perspectives, 24 (1), 62-77.

Publication III

Jukka Huhtamäki, Kaisa Still, Minna Isomursu, Martha G. Russell and Neil Rubens (2012) "Networks of Growth: The Case of Young Innovative Companies in Finland," *The 7th European Conference on Entrepreneurship and Innovation (ECIE 2012), Santarém, Portugal.*

Networks of Growth: Case Young Innovative Companies in Finland

Jukka Huhtamäki[1], Kaisa Still[2], Minna Isomursu[2], Martha G. Russell[3], and Neil Rubens[4]
[1] Hypermedia Laboratory, Tampere University of Technology, Tampere, Finland
[2] VTT Technical Research Centre of Finland, Oulu, Finland
[3] Media X, Stanford University, Palo Alto, CA, USA
[4] Knowledge Systems Laboratory, University of Electro-Communications, Tokyo, Japan
jukka.huhtamaki@tut.fi, kaisa.still@vtt.fi, minna.isomursu@vtt.fi, martha.russell@stanford.edu, neil@hrstc.org

Abstract: In this paper, we explore a vital part of the Finnish innovation ecosystem: young, innovative companies that are supported for fast international growth. Highlighting the importance of networks, we proceed to analyze the perceived existing relationships these companies have with other companies, financing organizations as well as with individuals taking part in their co-creation. We propose that these existing relationships, or connections, may be used to explain the firm as a resource integrator within a network, contributing to its growth and success. Overall, we propose that network analysis and resulting network visualizations can provide novel insights into the understanding of possibilities for global growth and success.

Key words: innovation, growth, networks, network analysis, young companies

1 Introduction

Networks (whether you call them innovation networks or business networks) address the notion that value creation, and especially innovation activities, are rarely carried out within a single organization in the flat-world economy (Friedman, 2005). Rather, resources are extracted from multiple sources; see e.g. Triple Helix Model highlighting the university-business-government interactions (Etzkowitz and Leydesdorff 2000); open innovation referring to the idea that ideas could come from both inside or outside of the company (Chesbrough 2003), and co-creation emphasizing collaboration with customers, suppliers and other stakeholders (Ramaswamy and Gouillart 2010). Service-dominant logic takes this even further as it states that all economic and social actors are resource integrators, implying that the context of value creation is in networks (Vargo and Lusch 2004, Vargo 2009). These networks rely on the infrastructure of relationships of people and organizations, and relationships are recognized as channels through which resources flow (Wind and Fung 2008), in forms such as linkages between executives (Ibarra and Hansen 2011). Therefore, in the background of this study is the understanding that innovation takes place in the context of relationships that form a network via the linkages between firms and their human and financial resources (Russell et al. 2011), manifesting between people, in teams, organizations as well as between organisations.

In this study, we explore approaches for understanding the participation of a company in the networked world. We propose that looking at existing relationships of firms as resource integrators according to Vargo provides glimpses of the participation of the company, and can show the channels, or access, to needed resources, or other resource integrators in the network. Flows of knowledge are also recognized as important resources in networks: sharing, acquiring and deploying knowledge is integral in networks (Dhanaraj and Parkhe 2006). On individual level good communication and social skills, and interactions are seen important; on organizational level formal and informal communication channels and knowledge exchange forums are highlighted for creating a strong reputation that enhances operating in networked environments (Ritala et al. 2009).

As network actors are clearly social actors, we propose using social network analysis (SNA, which has been used for several decades to study the sociological relationships of people and organization) to gain insights into the social configurations of the network (Wellman and Berkowitz 1988; Wasserman and Faust 1994). For example, network analysis has been used to study the interdependence of industries and nations (Yim and Kang 2008) as well as the dependence of innovation networks on knowledge flows (Owen-Smith and Powell 2004) and to explore the investment flows into an innovation ecosystem (Huhtamäki et al. 2011). One of the benefits of network analysis is in enabling investigators of networks to gain insight in the social configurations of the networks and in supporting them in communicating their findings to others (Freeman 2009).

2 Research methodology: Case young innovative companies in Finland

We will be using a case-study methodology into exploring the possibilities of growth and success of SME companies, and the role of networks in it. The explanatory nature of a case study approach, and its applicability to social studies (Yin 1994) guided our selection. In addition, case-study methodology is seen to provide detailed and analysed information about real world environments which can be seen as examples of phenomena under research, allowing the researcher to answer “how” and “why” questions (Benbasat 1987), which are seen relevant for this study.

We will be using an integral part of the Finnish innovation ecosystem called “Program for funding young innovative companies” as our case environment. The case environment is presented as an example of an approach to take in order to gain insight on the people- and investor-based networks surrounding and possibly interconnecting companies.

2.1 Program for young innovative companies in Finland

According to Statistics Finland, “Of all enterprises, 99.1 per cent were small enterprises, that is, employing under 50 persons. They employed 48 per cent of all personnel and accounted for 35 per cent of total turnover (http://www.stat.fi/til/syr/2010/syr_2010_2011-11-25_tie_001_en.html). Overall, the contribution of small companies to Finnish society and its wealth creation as well as employment creation is enormous. Consequently, the Finnish innovation ecosystem and its guiding policies have recently been emphasizing the role of these SMEs, start-ups and growth companies, also due to the changes attributed to restructuring of Nokia Corporation.

The Finnish Funding Agency for technology and development Tekes has a major role in building and sustaining the Finnish innovation ecosystem, through funding and other services that it provides for individual companies as well as clusters of organizations. In 2011 Tekes made funding decisions regarding 1,928 projects, which resulted in total investment of €610 million, of which 58 per cent was targeted at SMEs (Tekes annual review 2011, http://www.tekes.fi/u/Annual_Review_2011.pdf). Tekes new strategy reflects the new emphasis in Finland: it states that it gives priority to growth-seeking, innovative SMEs (http://www.tekes.fi/en/community/Structure/557/Structure/1428).

One example of the Tekes strategy is program for “funding for young innovative companies”, supporting young companies for international growth. Through it, Tekes not only provides funding resources, but also other resources, such as expertise and experience of its personnel, access to accelerator environments, and as well as its connections, for the selected companies so that they can grow and succeed in global markets. This program was initiated in 2011. It is intended for a company that (1) has a capacity and willingness to strive for fast international growth, (2) has products or services that can generate considerable business, (3) has a credible growth plan, and a committed and skilled management team, (4) has been in operation for less than 6 years and is small, and (5) invests strongly in innovation activities (http://www.tekes.fi/en/community/Young_innovative_growth_enterprises/1155/Young_innovative_growth_enterprises/2528).

2.2 Using social media data

Our approach to explaining the possibilities for growth and success for young innovative companies is to show how the companies participate in the world, seen through the lense of social media. Towards this goal, we will be using data-driven, network centric methods and two sets of data.

First, this study takes and extends the approach of Ecosystem Network Analysis that has been applied e.g. when looking into the co-creator configuration of Finnish Innovation Ecosystem (Huhtamäki et al. 2011). Accordingly, as data source, the IEN dataset (Rubens et al. 2010) will be used: a socially constructed dataset, which is built by crawling the Internet for socially curated information on press-worthy technology-based companies, their executives and board level personnel, and investment organizations as well as transaction flows. It is socially constructed, like Wikipedia, referring to the fact that individuals can add data to it when they want (they can also verify and correct its data), therefore contributing to its availability and timeliness, but also to its potentially erroneous data and public bias. Therefore, it basically has the power to combine the interesting activities happening in technology-based companies, and can show how different actors are connected. The dataset is a rather large one: in April 2012 it includes over 100,000 people, 80,000 companies and 7,000 financial organizations, and is based on sources in English.

We see that when something interesting and newsworthy happens in the company, it wants to share its news and communicates through its web site, press releases, or through its social media activities which also allows for engagement with the surrounding ecosystem resulting in impacting the perception of the company. This interesting information can prompt individuals within the company or outside the company to add its information that then ends to IEN dataset and eventually to network analysis.

Our assumption is that at least some of the Young Innovative Companies are included in the IEN dataset. The resulting visualizations of their relationships to other individuals, organizations and investors may provide insights into understanding ways to act as resource integrators for growth.

The second method of analysis looks at social media presence of Young Innovative Companies from a different perspective. Therefore, it is proposed to provide findings for bringing forth at least some of their activities in attracting resources and using their resources for engaging the innovation ecosystem around them, impacting the perception of the company. This provides an alternative and, at the same time, broader view of social networks surrounding the companies. For example, it allows for bringing forth the activities of users and customers in relation to the individual company. Hence, we reached for Twitter as a social media allowing relatively straightforward data-collection in real-time.

Our assumption is that especially using social media the companies want to interact with their environment. By looking at their Twitter activities, we can see the networks of their resource integration interactions, including those with users and customers, within this social media platform.

3 Findings

Our findings present the two separate, yet interrelated sets of analysis and resulting visualizations that explain how young innovative companies that are part of the Tekes program are seen to participate in the world. For visualization, we used Gephi, an open interactive visualization and exploration platform for networks (Bastian, Heymann and Jacomy, 2009) for graph metrics, visualization and layout. Traversing and other network-creation procedures are implemented as Python-based batch processes. MongoDB, an open source document-oriented NoSQL database system, was used for managing the data.

3.1 Networks based on IEN dataset

In their public website, Tekes provides a list of companies that were included in its Young Innovative companies program by 31.12.2011. In all, 94 companies are listed. As only the name of the company is mentioned, we applied a fuzzy text-matching algorithm Levenshtein to bring up the potential company instances in IEN Dataset. To ensure that we do not include any false positive matches or miss false negatives, we set the match threshold to 0.7 and double-checked the matches manually.

A total of 33 (contributing to 35 percent) of the Young Innovative Companies were found in IEN dataset. For these companies, we could proceed with the network analysis.

To present the individuals and investors co-creating companies within the Finnish innovation ecosystem, we processed the network layout in two stages: (1) cluster-based stage, (2) relation-based compacting stage. In the cluster-based stage we use OpenOrd layout algorithm (Martin, Brown, Klavans and Boyack 2011) since it produces a layout that allows us to better distinguishing clusters based on the interconnections between the nodes. We then apply the Force Atlas (Bastian, Heymann and Jacomy 2009) to compact the graph (nodes that are connected to each other are pulled closer together) and to make the representation more easy to read and aesthetically pleasing. The network visualizations are embedded in the document by using vector graphics so it is possible to look at network details by zooming in.

Modeling of the network is an important part of the visualization process. Here, the resulting network is a directed one with connections pointing from individuals and investors towards companies. Instead of using node outdegree for sizing the nodes (cf. Huhtamäki et al. 2011), we chose to size the nodes proportional to their betweenness centrality, i.e. the amount of times a node is included in the shortest path between any two nodes in the network. Betweenness was select to highlight the individuals, companies and investors that have an important connecting role in the network as a whole instead of solely having a large amount of direct connections. For easier viewing, we used different colors to characterize the three kinds of nodes in the network: light blue nodes represent Young Innovative Companies, green nodes represent venture capital investors – individual people, companies and financial organizations that have invested to at least one Young Innovative Company, and blue nodes represent people that have a press worthy relationship to a company. The positions of individuals vary from CEO and board membership to positions on research and development activities. However, for this visualization we have removed the names of individuals, as we recognize the limitations of the dataset as well as want to emphasize the patterns instead of particular individuals. Then, for easier storytelling purposes, we split the visualization process into two steps. Both present the networks and relationships of Young Innovative Companies.

The first visualization (Figure 1) presents the direct network of Young Innovative Companies, showing the people that are working or have been working in the companies, and the investors (both individuals as well as organizations) that have invested in these companies. Tekes Young Innovative Companies (YIC) program is in the center, as it is the connecting entity through its financing activities (all of the Young Innovative Companies are connected to it). The network visualization also shows the connections between these actors: (a) some individuals are connected with more than one company, and (b) some investors are connected with more than one company. It hence introduces a network of 119 actors and 130 connections. We can see that several companies have clusters of actors around them, indicating the number of their direct connections. The participation of investors in this network, marked by green nodes, is also clear.

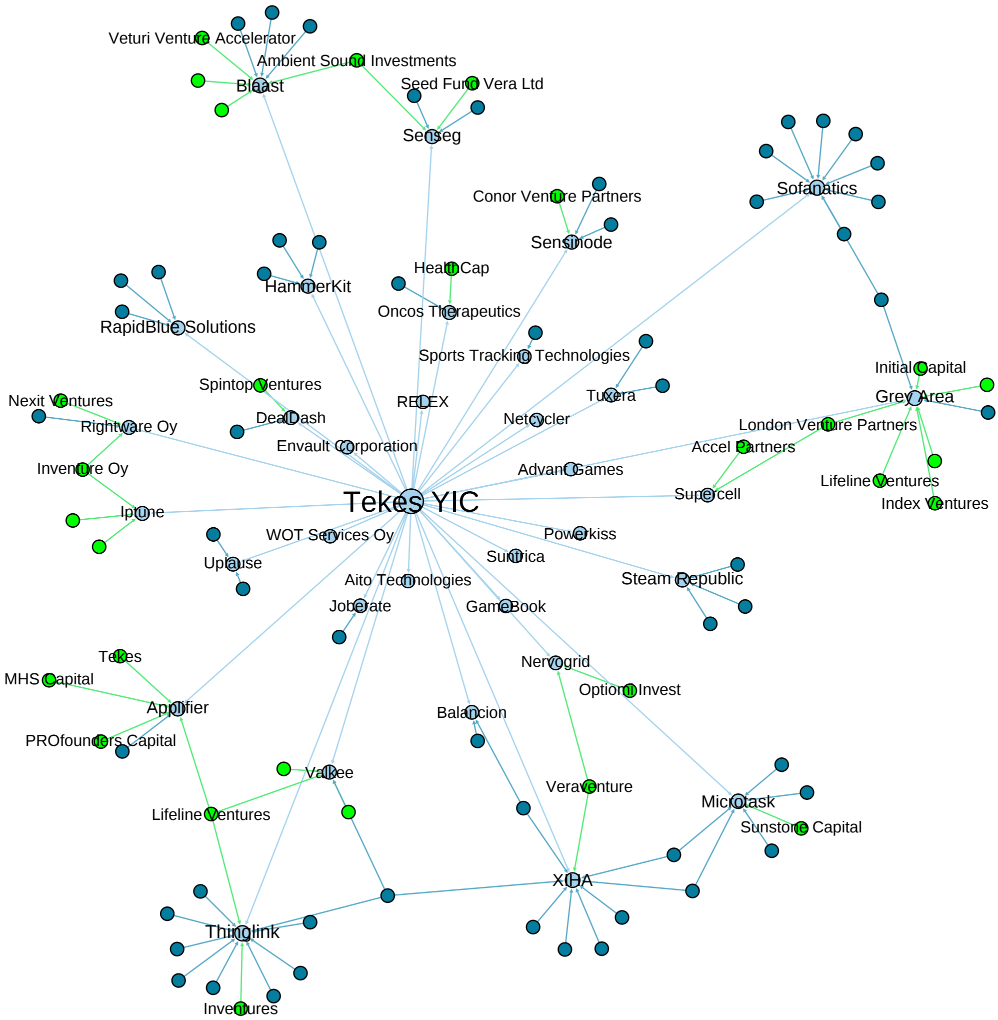


Figure 1. 1-step network visualization of the Tekes Young Innovative Companies and their direct connections

The second visualization (Figure 2) provides a view toward the horizon behind the immediate connections. We traverse through each individual in the first level network to include all companies, represented in red, that the people are connected into but not included in Tekes Young Innovative Companies program. Moreover, people and investors related to the newly added companies are included. The visualization hence adds new actors and connections beyond the direct connections, introducing a network total of 786 actors and 815 connections—a network that could be within reach for Young Innovative Companies in their resource integrating activities. The clusters of individuals and investors around Young Innovative Companies as well as their related companies become evident in the visualization. Due to the visualization technique (node size indicating the betweenness of the actor), one node representing an individual stands out: according to the data, this person is the only individual directly connecting Nokia with Google (the latter with a large number of individuals connected with it), which of course is a significant role. In all, this individual is connected to 12 companies, out of which 3 are included in Tekes Young Innovate Companies program. Further, the presence of Google becomes dominant due to the number of individuals connected. While this may not be truly representative of the Finnish innovation ecosystem, it reflects the global nature of networks as well as Google's key role in the global innovation ecosystem.

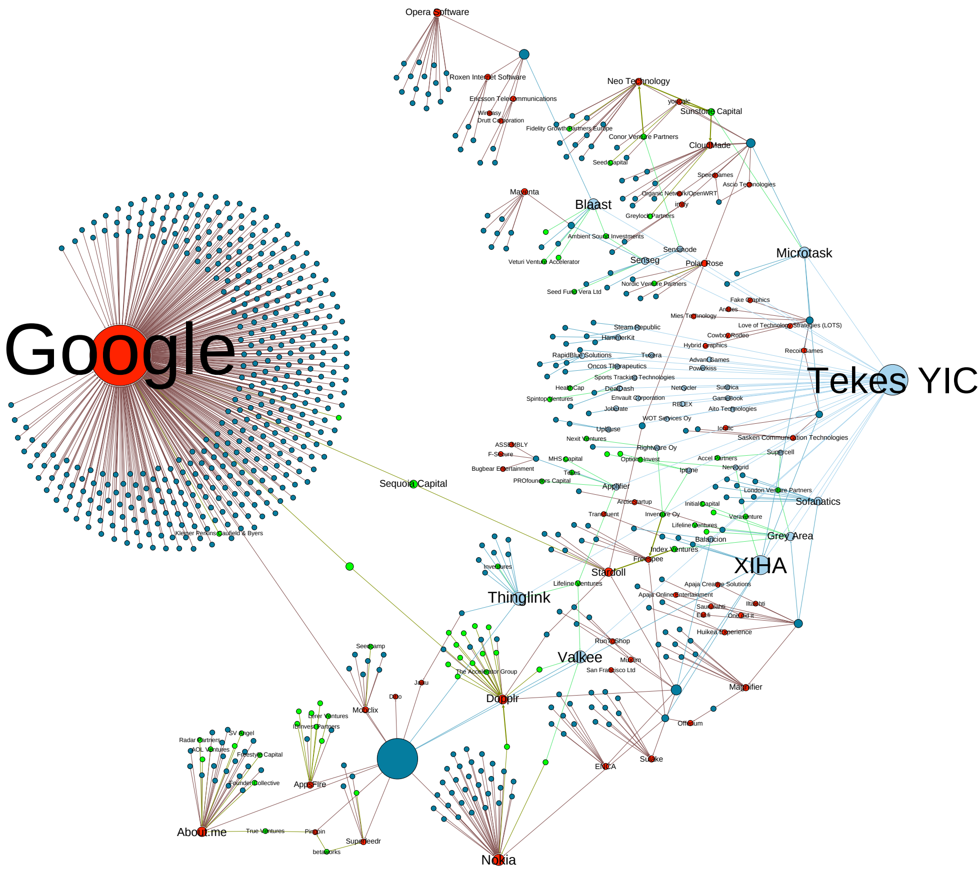


Figure 2. 3-step network visualization of Tekes Young Innovative Companies, their direct connections and the companies, investors and individuals that can be reached through the direct connections

3.2. Networks based on Twitter

For a list of Twitter accounts of Young Innovative companies, we first queried the IEN Dataset and complemented the list by manually adding the missing account information. In all, 46 Twitter accounts were found for the 94 Tekes Young Innovative Companies (49 %). Through a tailor-made batch script, we collected followers for each company through the Twitter API. A total of more than 70 000 followers were found for the companies.

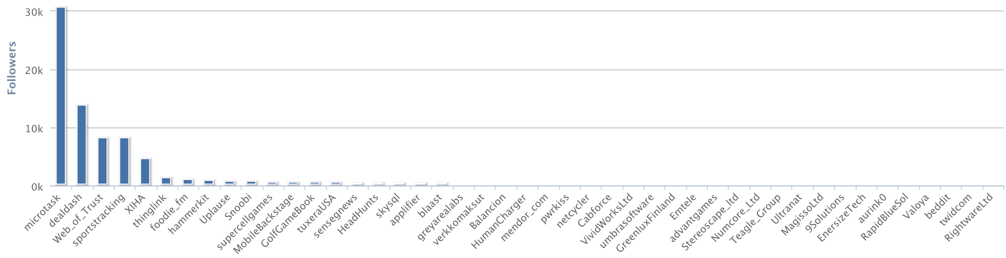


Figure 3. The distribution of Twitter follower count for Tekes Young Innovative Companies

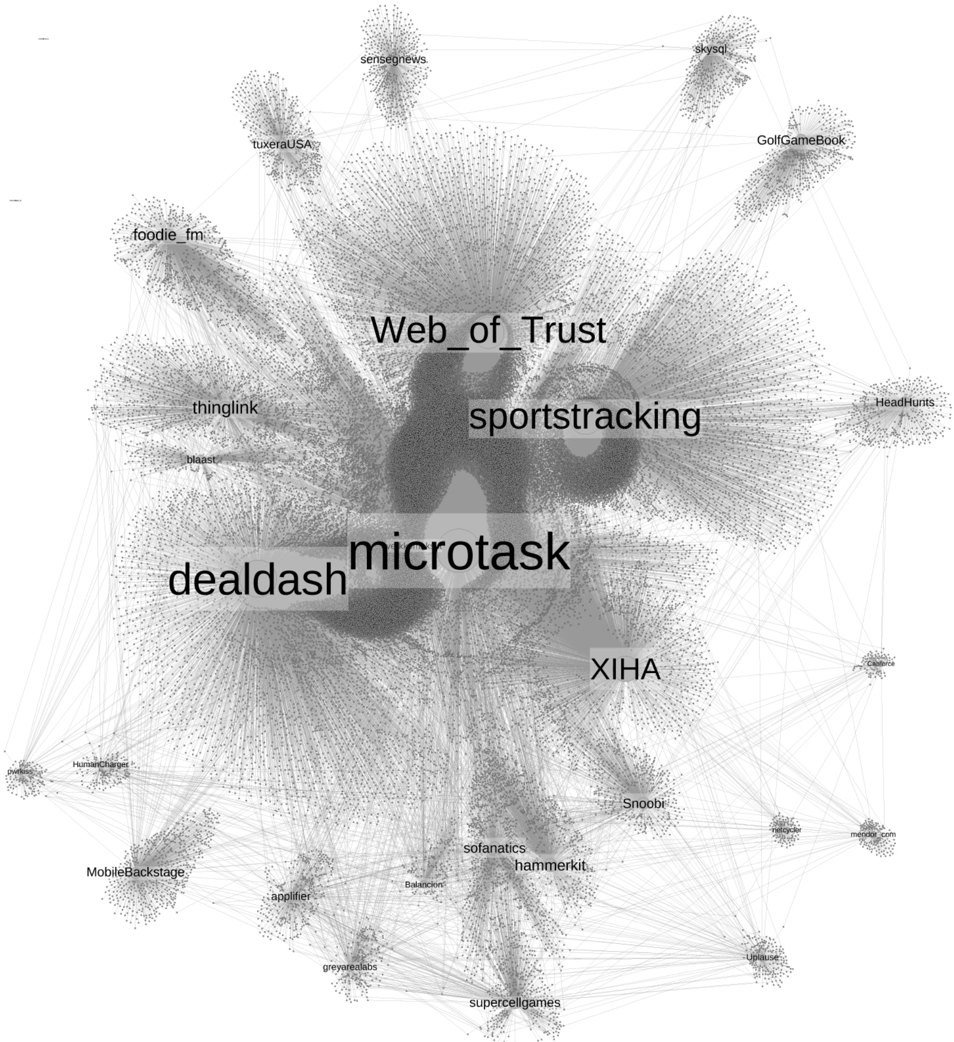


Figure 4. Network of Tekes Young Innovative Companies and their followers

Figure 3 shows the distribution of Twitter follower count for the different companies. As we can see, Microtask (http://www.microtask.com/), a company providing solutions to human powered document processing, has attracted over 30,000 followers and a few other companies have some thousands of followers. However, for most of the companies the follower count is small. This power law like distribution is commonly found in networks that are scale free (Barabási and Bonabeau 2003): this means that as a result of the process of preferential attachment, one actor dominates the connections in a network. We see that in this context, direct conclusions should not be drawn from the values for individual companies. Rather, a more valuable approach for providing insights on follower data is to create an overall or ecosystemic view to the companies.

For showing the Twitter-mediated interlinkages between the companies, we created a union of 1-step egocentric networks. In other words, the followers of a company were connected to a company with a directed connection but the connections between the followers were not included. Again, a combination of clustering and relation-based compacting stages were used to layout the network. The resulting visualization in Figure 4 shows all the 72,880 nodes and 75,755 connections in the network and can be used to pinpoint e.g. patterns in follower-based connections between the companies. We chose to present a view to the network where node size is representational to its indegree, i.e. the number of Twitter followers of a company. We acknowledge that the modeling and visualization design decisions depend on the questions that one seeks to discuss and answer with the help of the network: if one would be interested in finding the most active followers in the network, node size should be proportional to its outdegree. To find the companies that are most strongly connected to each other through shared followers, a one mode network of companies could be created with connection weight proportional to the number of shared followers. For example, Microtask shares the most amount of followers with other companies: 444 followers with XIHA; 188 with Web of Trust and 155 with Hammerkit.

4. Discussion

Our goal was to explore the networks of growth of Tekes Young Innovative Companies using social network analysis, social media data and network visualizations. We wanted to highlight the role of companies as resource integrators within networks, acting with other resource integrators in order to succeed and grow. Through the process from social media data through analysis we produced concrete visualizations of the networks. Our findings made visible the existing relationships that young Innovative Companies in Finland have: (1) directly with individuals and financing organizations, (2) with individuals, other organizations and financing organizations within their reach, and (3) with people and organizations interacting with them. They show that Young Innovative Companies already have connections (both direct and indirect ones) with a number of individuals and organizations through which they can access resources and interact with resources needed for their growth and success, and due to the nature of data from social media, these visualizations are rather timely snapshots. Hence, we want to highlight that the results are often not generalizable or conclusive. This should not come as a surprise when taking into account the complex, paradoxical and context-sensitive nature of innovation.

First, we evaluated the process and results internally. Initially, we explored the fact that not all Young Innovative Companies were found in the IEN dataset. Through the analysis of their websites, we found out that only a few of the companies did not have their web site in English and that most of them actively issue press releases and news to communicate about them (also in English)—therefore they seem to be actively gearing for international communication and presence, and have the potential for being added into the IEN dataset, furthermore demonstrating the applicability of using the IEN dataset for analysing this sample of the Finnish innovation ecosystem. However, to be included into the IEN dataset requires activities of individuals for recognizing and adding entities to it (either within or outside of the companies), which means that we acknowledge that there are individuals, companies, and financing organizations missing from it, which can be explained with the concept "public bias". In addition, we addressed the limited connections that are visible in the IEN dataset: for example, we know that certain individuals are married, have gone to school together, and might be neighbors—these connections are not visible in the IEN dataset nor in Twitter data, altogether highlighting the fact that some actions take place in social media or can be traced through social media and some remain outside of it.

As our proposition is that the network visualizations can provide insights for Young Innovative Companies as well as for the ecosystem trying to support them, we then presented the visualizations to experts of the Finnish innovation ecosystem at Tekes for their evaluation, and discussed the value of the process and the results with them. We noted that as network analysis and visualizations are not traditional ways of exploring innovation, the first reaction was "interesting, fascinating", followed by questions (1) about the methods of producing the visualization: Why is this individual node so big?

What do these different colors of line mean? and (2) about the meaning of the overall results: Is this about gatekeepers in networks? About weak ties? Is this a system? What is the bigger phenomenon that can be explained with these visualizations? Only after arriving in the shared understanding about the visualizations and what they represent, the experts had comments about insights of them: for example, one of them commented that “I would have expected company X to have more connections”. One novelty offered by the visualizations was clearly that it showed the investors within the network— this was something that the experts had not seen before and something that the data they have routinely access to does not show. They then proceeded to provide suggestions for allowing for better and clearer insights: indicating by color the Finnish and international actors of the networks, indicating the specific market segments of companies (for example mobile, gaming, pharma etc.), and bringing in timelines, for example for showing what connections have formed after the company has become participant of the Tekes Young Innovative Companies program.

The kind of modeling used in both parts of this study can be utilized to create network views for further insights on several complementary aspects. First, due to the fact that the networks are directed, node indegree and outdegree values can be used to highlight actors in different roles. In Twitter analysis, for example, the companies that have the most followers have a large indegree and the Twitter users that follow many companies a large outdegree. Second, as the first part of the study shows, betweenness value provides an easy and intuitive way to find actors that have a particularly significant role as connectors in the network. Finally, the network can be filtered or split to smaller pieces e.g. in Twitter case to find individuals following a specific set of companies, for example, operating in the same domain.

We are tempted to suggest the process of navigation as a metaphor or analogy for the kind of cartography we provide here: while being a long way from a modern proactive car navigator, the visualizations shown here make the topology of parts of innovation ecosystem explicit. Indeed, visual network analysis affords investigators insights on the (often latent) (social) configurations of the networks and allows sharing the insights to others (cf. Freeman 2009). The results presented in this article represent the first evolution of the analysis. In order to validate the approach, we plan to engage with different stakeholders to see both what is missing from the networks and, even more interestingly, what are the new insights that the networks afford the stakeholders.

Acknowledgements

The authors would like to extend their gratitude to Tekes experts involved in the study for their comments and feedback and further to Teemo Tebest for technical assistance in collecting Twitter data. This study is partly funded by Tekes.

References

Barabási, A.-L. and Bonabeau, E. (2003). Scale-Free Networks. *Scientific American*, 288(5), 50–59.

Bastian, M., Heymann, S., & Jacomy, M. (2009). Gephi: An Open Source Software for Exploring and Manipulating Networks. Presented at the International AAAI Conference on Weblogs and Social Media, San Jose, California,USA.

Benbasat, I., Goldstein, D.K.and Mead, M. (1987). The Case Strategy in Studies of Information Systems. MIS Quarterly, 11 (3): 369-386.

Chesbrough, H. (2003). Open Innovation: The New Perspective for Creating and Profiting from Technology. Boston: Harvard Business School Press.

Dhanaraj, C. and Parkhe, A. (2006). Orchestrating innovation networks. Academy of Management Review, Vol. 31, No. 3, pp. 659-669.

Etzkowitz, H. and Leydesdorff, L. (2000). The Dynamics of Innovation: From National Systems and ‘Mode2’ to Triple Helix of University-Industry-Government Relations. Research Policy, 29 (2): 109-123.

Freeman, L.C. (2009). Methods of social network visualization. Encyclopedia of Complexity and System Science. Springer: Berlin.

Friedman, T. (2005). The World is Flat: A Brief History of the Twenty-First Century. Farrar, Straus and Giroux: New York.

Huhtamäki, J., Russell, M.G., Still, K. and Rubens, N. (2011). A Network-Centric Snapshot of Value Co-Creation in Finnish Innovation Financing. Open Source Business Resource, March 2011.

Ibarra, H. and Hansen, M.T. (2011). Are you a Collaborative Leader? Harvard Business Review, July-August, 69-74.

Martin, S., Brown, W. M., Klavans, R., & Boyack, K. W. (2011). OpenOrd: an open-source toolbox for large graph layout. Presented at the Visualization and Data Analysis 2011, San Francisco Airport, California, USA. doi:10.1117/12.871402

Owen-Smith, J. and Powell, W.W. (2004). Knowledge Networks as Channels and Conduits: The Effects of Spillovers in the Boston Biotechnology Community. Organization Science, 15: 5-12.

Ramaswamy, V. & Gouillart, F. (2010). Building the Co-Creative Enterprise. Harvard Business Review, 88(10).

Ritala, P., Armila, L. and Blomqvist, K. (2009). Innovation Orchestration Capability—Defining the Organizational and Individual Level Determinants. International Journal of Innovation Management, Vol. 13, No. 4, (569-591).

Rubens, N., Still, K., Huhtamäki, J. & Russell, M.G. (2010). Leveraging Social Media for Analysis of Innovation Players and Their Moves. Innovation Ecosystems Network, Media X, at Stanford University.

Russell, M.G., Still, K., Huhtamäki, J., Yu, C. and Rubens, N. (2011). Transforming Innovation Ecosystems through Shared Vision and Network Orchestration. Proceedings of Triple Helix Conference.

Tekes annual review 2011, http://www.tekes.fi/u/Annual_Review_2011.pdf

Vargo, S. and Lusch, R.F. (2004). Evolving to a New Dominant Logic for Marketing. Journal of Marketing, Vol. 68 (January 2004), 1-17.

Vargo, S. (2009). Toward a transcending conceptualization of relationship: a service-dominant logic perspective. Journal of Business and Industrial Marketing, 24/5/6 (2009) 373-379.

Wasserman,S. and Faust, K. (1994). Social Network Analysis: Methods and Applications. 1st Edition. New York, NY: Cambridge University Press.

Wellman, B. (1988). Structural Analysis: From Method and Metaphor to Theory and Substance. In Wellman, B. and Berkowitz, S.D. (Eds.) (1988) Social Structures: A Network Approach. Cambridge University Press: New York, NY, 19-61.

Wind, J., Fung, V.K.K. and Fung, W (2008). Competing in a Flat World: Building Enterprises for a Borderless World. Wharton University Publishing: Upper Saddle River, NJ.

Yim, D.S. and Kang, B. (2008). Policy Options of Establishing Effective Subnational Innovation Systems and Technological Capacity-building. Asia-Pacific Trade and Investment Review, 4, 115-137.

Yin, R. (1994). Case study research: Design and methods. 2nd edition. Beverly Hills, CA: Sage Publishing.

Publication IV

Jukka Huhtamäki, Martha G. Russell, Kaisa Still, Neil Rubens (2013) "A Network-Centric Snapshot of Value Co-Creation in Finnish Innovation Financing," *Value Co-Creation: Best of TIM Review*.

Available at http://timreview.ca/article/424

A Network-Centric Snapshot of Value Co-Creation in Finnish Innovation Financing

Jukka Huhtamäki, Martha G. Russell, Kaisa Still, and Neil Rubens

"In co-creation, strategy formulation involves imagining a new value chain that benefits all players in the ecosystem."
Venkat Ramaswamy and Francis Gouillart (2010)

In this article, we apply the concept of value co-creation to the analysis of linkages between organizations and their human and financial resources to observe the emergence of cooperative activities in a specific innovation system. Through visual network analysis of a federated and socially constructed dataset of organizations and their related actors, we show how co-creation occurs through financial linkages.

We use the ecosystem concept as a metaphoric reference to value co-creation with a network-centric mindset. Business financing linkages reveal convergence and co-creation in the innovation ecosystem, and network analysis is used to visualize the relationships between firms. Through the lens of relationship-based synergy, we provide a snapshot of innovation funding, which highlights the collaboration of venture capital and government agencies in co-creating the emerging Finnish innovation ecosystem.

Introduction

The term co-creation was coined to explain emerging relationships between customers and the companies though which they were jointly creating value. Recently, the frame of reference has been extended to an emerging business and innovation paradigm that leads to the need of “changing the very nature of engagement and relationship between the institution of management and its employees, and between them and co-creators of value - customers, stakeholders, partners and other employees” (Ramaswamy, 2009; http://tinyurl.com/47c9ook).

Strategic value creation networks can be observed through network analysis of small, medium, and large enterprises, and they are important examples of co-creation. A leading idea in open innovation is that, because valuable knowledge exists outside of an individual organization, companies purposively co-create value networks through vendor-supplier relationships and collaborative service offerings that are specific to market segments. Inter-firm relationships created by the participation of executives and board members in two or more enterprises with related missions, markets, products, or social initiatives are additionally a potentially powerful force for value co-creation. In a similar way, enterprises receiving investment resources from the same financial source may share complementary visions of the future, complementary benefits from new technologies, and synergistic market development. Business ecosystems are comprised of the aggregate of these relationships among individuals and groups of individuals in clusters of companies. The competitive advantage of clusters accrues from the linkages and the synergy between activities (Porter, 2000; http://tinyurl.com/4csuj9u).

Co-creation is an essential force in a dynamic innovation ecosystem because a continual realignment of synergistic relationships of people, knowledge, and resources is required for growth of the system and responsiveness to changing internal and external forces (Rubens, et al., 2011; http://tinyurl.com/4rnup6h). On one hand, venture capital is the "independent, professionally managed, dedicated pools of capital that focus on equity or equity-linked investments in privately held, high growth companies" (Gompers and Lerner, 2001; http://tinyurl.com/4vd5r2z), has specific termination objectives that drive investments. On the other hand, government development agencies are often framed around capacity building missions – building markets, standards, supply chains, and technical and managerial talent. The investment strategies of development agencies vary in outcome objectives, as well as in time frame and financial objectives. For examples, differences in the "cultivation vs. harvesting" strategies evidenced by investments into and out of China have been described (Rubens et al., 2011).

Jungman and Seppä (2004; http://tinyurl.com/4cpwxm5) differentiate the role of angel investors, incubators, advisors, and corporate investments in bridging the gap between seed funding of prospective companies and capital infusion into investable companies. While all these types of financial resources may be available for business investment in a region, the role and proportion may vary. Investors' ultimate objective is for a new company to undergo the major liquidity event that allows it to become listed on a stock exchange. An ecosystem including both experiential and financial resources is needed to co-create successful journeys across the gap from a prospective to a listable company.

In this article, we use data-driven social network visualization to present a network analysis of venture funding in the Finnish innovation ecosystem. A socially constructed dataset is used to study the nature of business co-creation through syndicated venture capital investments. We show that the dataset can be explored to provide value to researchers as well as ecosystem facilitators and other agents of change. The snapshot of innovation funding in Finland is examined by means of network analysis to visualize inter-firm relationships, following the ecosystem as metaphoric reference for value co-creation in a network-centric mindset. The analysis concentrates on investments of venture capital, which in Finland have been oriented to early equity-phase financing of high-tech startups. A total, all-inclusive analysis of the Finnish system is outside of the scope of this article, but the visualization snapshot of venture funding will serve as a starting point to stimulate the development of insights relevant to innovation experts, analysts, and decision makers within the context of the Finnish innovation ecosystem.

Venture Funding for the Finnish Innovation Ecosystem

The Finnish national innovation system has been described as a network of various actors, with education, research, product development, and knowledge-intensive business and industry at its core. Regarding the flows of investments into this system, it has been noted that "because of the importance of the public venture capital/private equity organizations, the Finnish venture capital system can be described as dual one in which some private venture capital funds have been initiated by public intervention" (Luukkonen, 2006; http://tinyurl.com/5v4tota). Furthermore, special characteristics have been noted: i) due to the small markets in Finland, the growth expectations oftentimes have been limited, which has impacted non-Finnish investors' perceptions of the attractiveness of investment in Finnish companies; ii) these existing public investors many times have been passive; and iii) that there are very few corporate venture capitalists in Finland (Luukkonen, 2006).

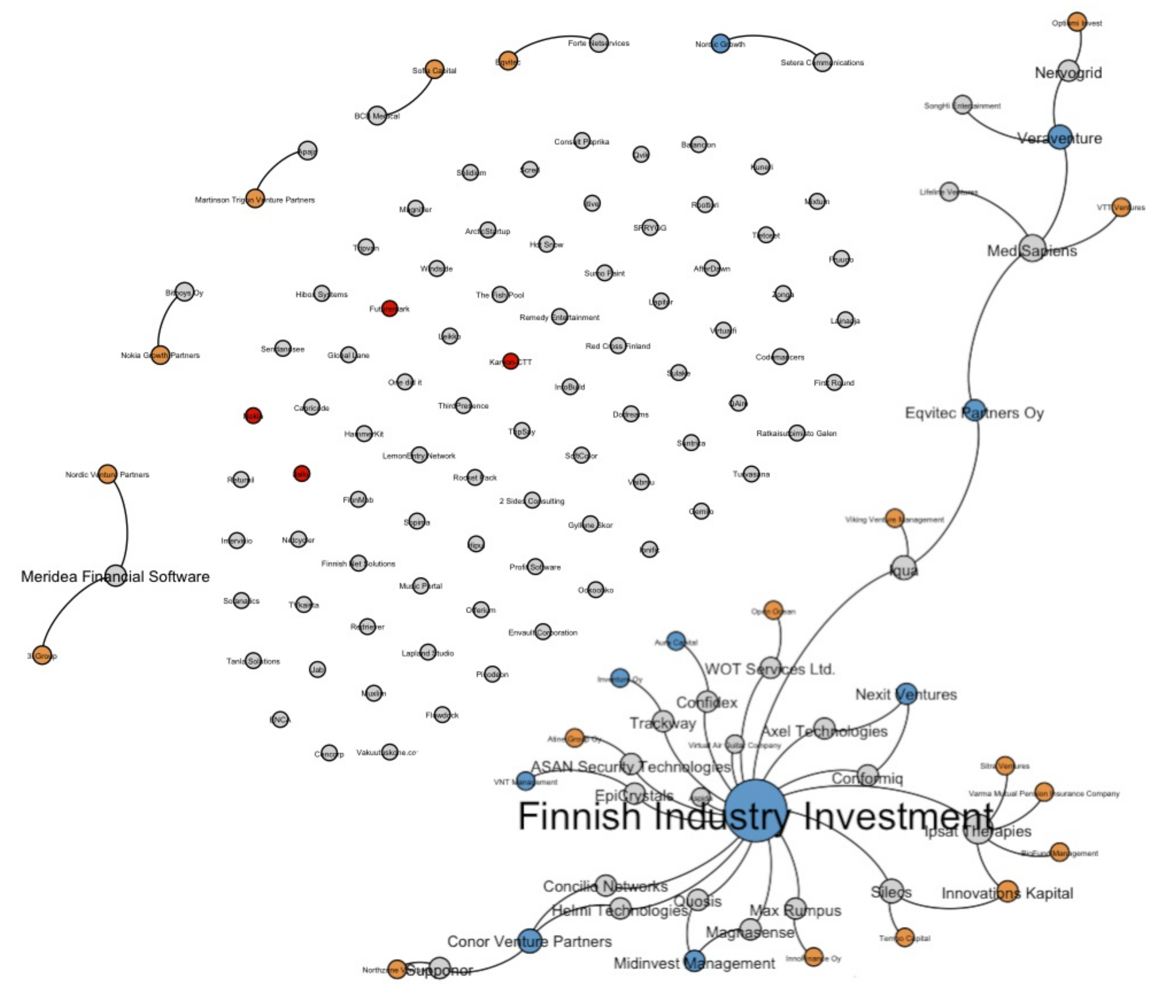
Value Co-Creation in Finnish Innovation Financing

Jukka Huhtamäki, Martha G. Russell, Kaisa Still, and Neil Rubens

In this sample of 108 high-tech companies, 53 investments were announced from 28 institutional investors, made in 29 rounds between 2005 and 2010. An examination of the social networks and other structures produced from this data is much like a walkabout in the Finnish innovation funding ecosystem. Visual analysis shows the patterning of connections between company actors as well as those of financial resources flowing to Finnish technology-based companies, implying co-creation from innovation funding. For example, the walkabout reveals a landscape of four companies that have come of age – sold or issued an initial public offering (IPO), amidst many independent firms – and a few with international connections. One actor dominates the investment landscape.

Figure 1 shows all 136 actors in our sample, which consisted of 108 technology-based companies with a home office in Finland and 28 investment organizations. Companies and their funding organizations are interconnected with edges. The actors are colour-coded: companies

Figure 1. Network of Finnish Technology Companies and their Investment Organizations



are gray, unless they were sold or have issued IPO, in which case they are red. Investors with their home office in Finland are blue; investors whose whereabouts are international or unknown in the dataset are orange. The nodes are inflated according to their degree (i.e., the number of connections that they have to other nodes): the bigger the node, the more connections it has.

Among the notable relationships in the sample, Figure 1 shows:

1. Ipsat Therapies, Medisapiens, Iqua, and Silecs have the largest number of connections to investors.

2. Finnish investment organizations represent roughly half of the investors for these Finnish companies.

3. Conor Venture Partners, Veraventure, Eqvitec Partners, Innovations Kapital, Midinvest Management, and Nexit Ventures are linked to more than one company by their investments.

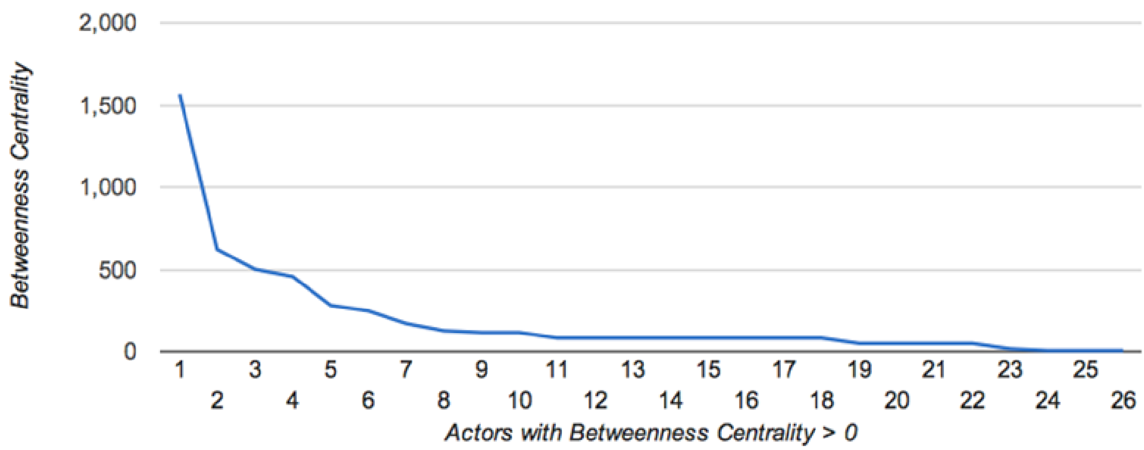
4. Biofund Management, Sitra Ventures, Varma Mutual Pension Insurance Company (Varma), and Finnish Industry Investment invested in Ipsat Therapies. This was the first investment in the sample and occurred in April 2005.

5. Medisapiens received investment from VTT Ventures, Eqvitec Partners, Veraventure, and Lifeline Ventures. This was the most recent investment and occurred in June 2010.

6. Most of the companies (75%) in this sample are not receiving funding from an investment organization. Although some companies have investments from individuals, angel investors are not included in this analysis.

In our sample, 56 of the companies and investment organizations (41%) are connected to one or more actors. Figure 2 shows the betweenness centrality values for the 26 actors that have a value larger than zero. Betweenness centrality is one of the key metrics in social network analysis (http://wikipedia.org/wiki/Centrality#Betweenness_centrality). It is based on counting the number of times that a given node is included in the shortest path between two nodes. Of the companies, Iqua has the largest betweenness centrality value: 610. Of the investment organizations, government-owned Finnish Industry In-

Figure 2. Distribution of Betweenness Centrality



vestment is connected to the largest number of companies, with a betweenness centrality value of 1557. For the whole sample, including the actors with no connections, betweenness centrality values of the lowest, low-medium, and upper medium quartiles are zero, making the average value 36.

The value distribution of betweenness centrality roughly follows a power law. Node degree value, the number of connections per actor, has a similar kind of distribution. This suggests that the network is scale free – characterized by a very small number of nodes that are highly connected and many nodes with little connection (Barabási and Bonabeau, 2003; http://tinyurl.com/4e3oxof). In scale-free networks, growth patterns that show preferences for attaching to highly connected nodes are typical and generally lead to the development of hubs (i.e., nodes with an enormous number of links) in a rich-get-richer manner. Scale-free networks tend to be "robust against accidental failures but vulnerable to coordinated attacks" (Barabási and Bonabeau, 2003).

Through the companies they co-fund, relationships between investment organizations are of strategic interest for co-creation. Sunburst diagrams were applied to visualize patterns in the Finnish innovation ecosystem. Figure 3 shows the co-investments of 22 investment organizations into 19 Finnish companies. Each investor that co-invested with another investor in this sample is shown in the inner circle. Their co-investors are placed in the outer circle adjacent to each investor, without specification of the time of investment. In this design, each investor appears as co-investor at least two times in the diagram. Investment organizations identified as Finnish are shown in blue. The Finnish Industry Investment co-invested with 15 other funding organizations; some co-investors were Finnish, while the location of others was not available in the data. (It should be noted that some of the investors are known by the authors to be Finnish, but their Finnish locations were not identifiable programmatically. The locations of these investors were therefore classified as unknown and are shown in orange in Figure 3. These organizations include, among others, Varma, Sitra Ventures, and VTT Ventures.)

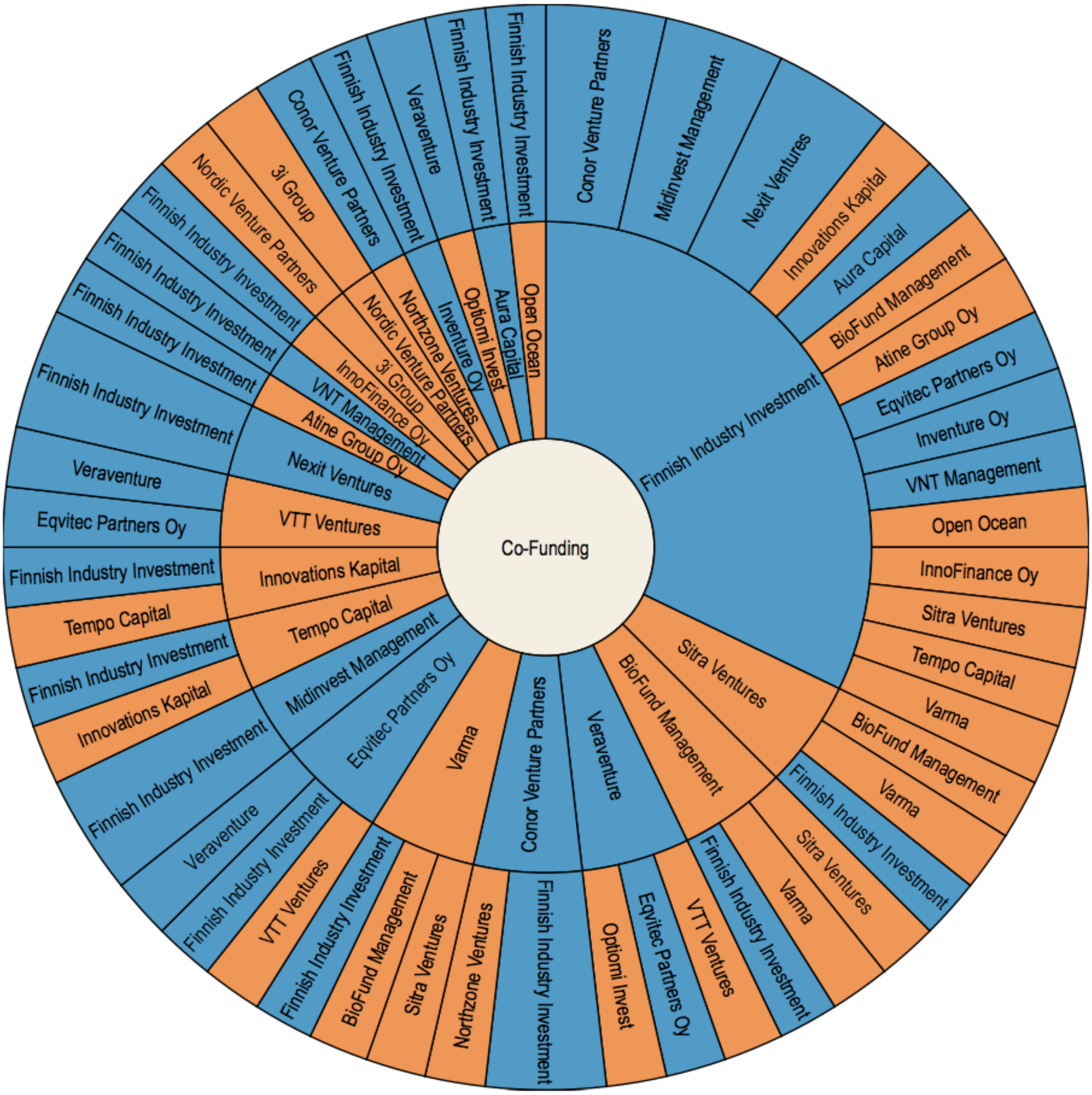
Figure 4 reveals funding paths or bursts for companies that have received two rounds of funding; no companies in this dataset were reported to have received a third-round investment. Second-round investors are shown on the outer circle adjacent to the investors of the first round for the same company. Finnish Industry Investment, for example, has been both a first-round investor and a second-round investor. When Sitra Ventures and Varma are regarded as being Finnish, we can see that a small majority (57%) of funding organizations participating in multiple funding rounds are Finnish organizations.

Discussion

The approach for visual co-creation analysis presented here is a synthesis of visual social network analysis and data-driven information visualization. Visualization and measurement are claimed to be the two main factors enabling the explosive development of modern science. Visualization has been a key element of social network analysis - and its precursor, sociometry - in supporting the exploration, presentation, and analysis of the structure of communities. The general objective of information visualization is to amplify the cognition of a user through an expressive, often interactive view that gives insight on a given phenomena represented by the data.

Data-driven visual storytelling allows insights on the structure and dynamics of a network to be shared with the help of visualizations. "[S]torytelling allows visualization to reveal information as effectively and intuitively as if the viewer were watching a movie" (Gershon & Page, 2001; http://tinyurl.com/6k8nb3t). Hans Rosling gives particularly inspiring examples of such storytelling; his presentations are some-

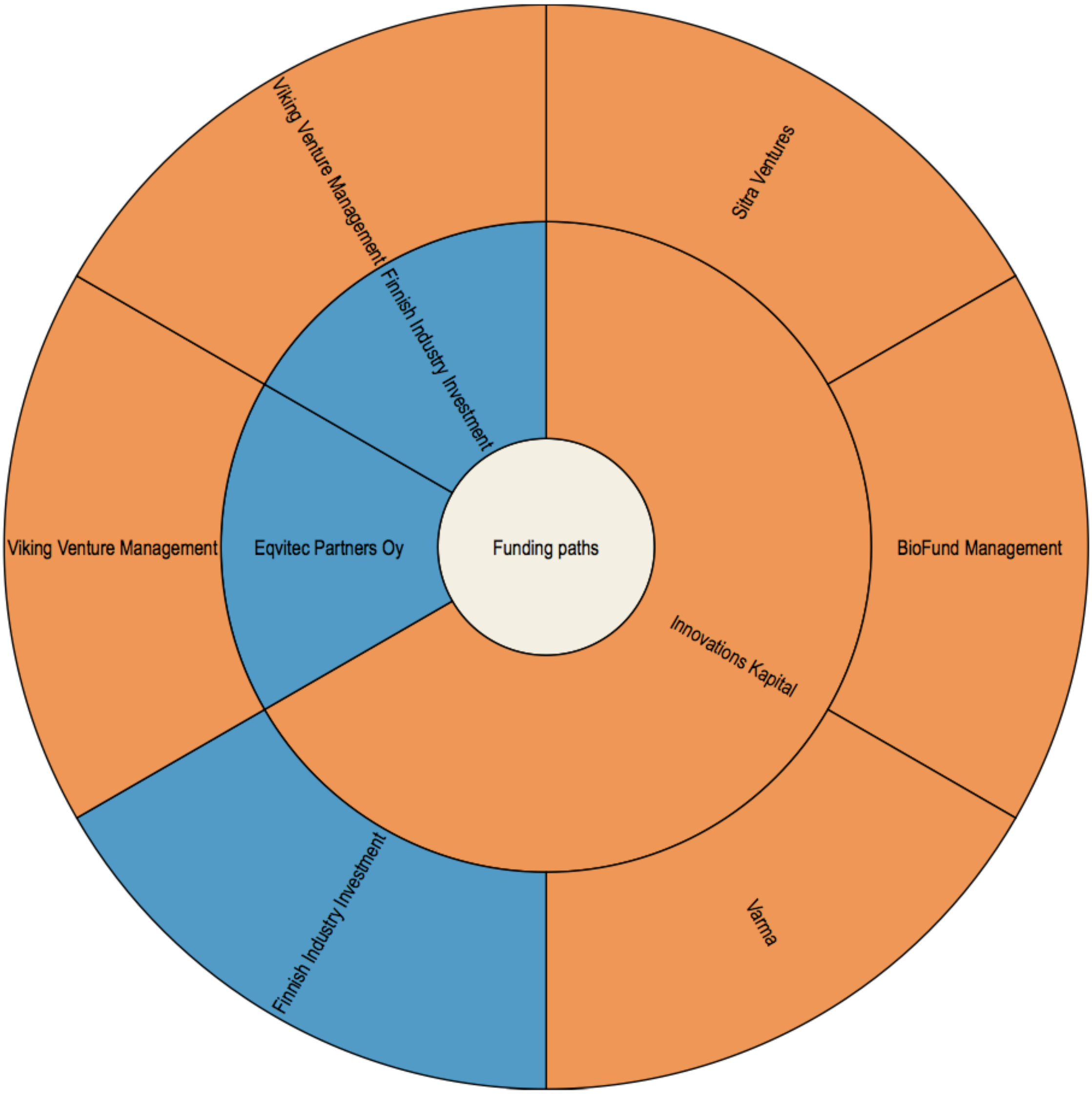
Figure 3. Patterns of Co-Investing in Finnish Technology Companies



times referred as "the best stats you've ever seen" (TED Talks, 2006; http://tinyurl.com/99rnmm)

This study's visual social network analysis revealed structural connections between Finnish technology-based companies and their investment organizations. A significant proportion of Finnish companies in the high-tech sector have not received funding from investment organizations since 2005. For those Finnish companies that have received funding, 63% of have received either first or second-round funding from Finnish Industry Investment. A handful of investment organizations (some Finnish and some not) provide modest diversification to the

Figure 4. First and Second-Round Investment Paths in Finnish Technology Companies



Finnish funding landscape, which shows a scale-free pattern.

Further, this analysis has generated preliminary insights about the general patterns of co-creator networks supporting the Finnish innovation ecosystem in the high-tech sector. The sunburst visualizations display funding pathways and highlight the flexibility of Finnish government investment organizations to co-create in both first-round and second-round funding. The co-creation role of these organizations is visualized through both concurrent and sequential cooperative investments. At the same time, the visualizations also reveal a dependency on Finnish Industry Investment and an opportunity to further diversify institutional investments in Finnish companies.

These initial patterns suggest avenues for future study. Investment relationships reflect an intentional alignment of business resources and goals that may be based on technologies, markets, or globalization strategies. A resource-based relationship implies that the partners share objectives, share risks, and share rewards as they co-create value through investments. In co-creation, both the risks and rewards are shared; however they may not be equal. The roles of first and second-round investors may be specialized with respect to the amount of risk, the financial and temporal objectives for exit, and the value of the network itself. Across public and private Finnish organizations making investments in technology-based companies with headquarters in Finland, this study showed that Finnish Industry Investment is unique in both leading and following the investments made by other entities.

This study lacks two very important investment players for a full view of the Finnish innovation ecosystem. Since firms were used as the unit of analysis, individuals serving as angel investors were not included. In a subsequent study, we seek to gain further insight on the business angels' vital role in seed financing for new technology-based companies – an act of co-creation in this sense. An interesting, though difficult, task for future work is visualizing the role of incubators and business angels in closing the gap between venture and capital.

Further studies could include the utilization of temporal data, which often yields insights about the evolution of a network. Network visualization tool-development initiatives such as Gource (http://code.google.com/p/gource/) and Gephi (http://gephi.org) are clear indicators of the interest that the open source community has in temporal network visualization. These tools are of high value when the dynamics of innovation ecosystems are studied for insights on trends, the roles of different actors, diffusion of information and innovations et cetera, but they insist on the availability of rich data sources.

Conclusion

Applying information visualization and visual social network analysis has huge potential for revealing the social structures and network dynamics within innovation ecosystems, from individual organizations to the whole world. Despite recent rapid development of visual tools for social network analysis, one major issue that hinders data-driven visual analysis of co-creator networks in innovation ecosystems is the lack of accessible, timely data about the global ecosystem of high-tech companies. We anticipate development in this area in the near future with the advent of (open) linked data (see http://linkeddata.org), which is currently endorsed with respect to opening up public administration. The authors are contributing to this opportunity by creating a dataset representing high-tech companies and building up research methods for this dataset.

The scale-free patterning of the Finnish venture capital network is similar to the findings of Barabási (2010; http://brsts.com) who claims that such patterning can be found in nearly all kinds of human activities. Adding the temporal dimension to data enables the analysis of the evolution of the network. This opens up a new level of insights into changes in the network that, at best, supports the formulation of future scenarios for agents of change in different innovation ecosystems. Two important opportunities for innovation policy analysts concern identifying incentives to effectively encourage the reinvestment of exit resources and orchestrating mechanisms to strategically encourage global participation in a manner that provides a return on investment back to its origin.

Value Co-Creation in Finnish Innovation Financing

Jukka Huhtamäki, Martha G. Russell, Kaisa Still, and Neil Rubens

This article is based on a paper presented in EBRF 2010: Co-Creation as a Way Forward. The authors express their gratitude to the Venture Capital Industry, Business Angels, and Knowledge Investors session chairs Prof. Markku Maula, Prof. Marko Seppä, and Dr. Jennifer Walske, as well as the other participants of EBRF 2010 for their co-creation efforts contributing to the article. Camilla Yu provided valuable feedback for revising the article. For additional versions of these visualizations, please refer to http://bit.ly/fininnofin.

Jukka Huhtamäki (M.Sc, Hypermedia) is a researcher, a post-graduate student, and a teacher working for the Hypermedia Laboratory (HLab) at Tampere University of Technology. Jukka also collaborates with the Innovation Ecosystems Network, lead by Martha G. Russell. His interests include visual social media analytics, methods of streamlining social network visualisation and information visualisation, user and information modeling and the development methods, and implementation technologies of social, adaptive, and distributed hypermedia. Currently, Jukka is working to develop data-driven visual analysis processes for insights on, for example, social media usage and innovation diffusion.

Martha G. Russell is Senior Research Scholar at Human Sciences Technology Advanced Research Institute and Associate Director of Media X at Stanford University, a membership-based, interdisciplinary research catalyst focused on people, media, technology, and innovation. Dr. Russell's background spans a range of business development, innovation, and technology-transfer initiatives in information sciences, communications, and microelectronics at the University of Minnesota, The University of Texas at Austin, and Stanford University. She collaborates with Stanford's Innovation Ecosystems Network and serves on the editorial boards of the Journal of Interactive Advertising, the Journal of Electronics, and Technology Forecasting and Social Change.

Kaisa Still studies innovation, technology transfer, and the role of technology, with a focus on information and knowledge creation, sharing, and management. Her studies focus on supporting collaboration and cooperation in organizations as well as in community settings. Recent studies include mobile, online and social networking communities, innovation ecosystems, and innovation indicators. Dr. Still has over 10 years of cross-sector business and academic experience in Finland, USA and China. She currently works for VTT Technical Research Centre of Finland and collaborates with Stanford's Innovation Ecosystems Network.

Dr. Neil Rubens is an Assistant Professor at the Knowledge Systems Laboratory, University of Electro-Communications, Japan. He is the Director of Active Intelligence Research Group and is a member of the Innovation Ecosystems Network at Stanford University. He holds an M.Sc. degree from the University of Massachusetts and a Ph.D. degree from the Tokyo Institute of Technology - both in Computer Science. His research focuses on developing Active Intelligence systems, which are systems Artificial Intelligence systems that are self-adaptable utilizing unsupervised and semi-supervised learning, and active communication and data acquisition. He collaborates with Stanford's Innovation Ecosystems Network.

Publication V

Kaisa Still, Jukka Huhtamäki, Martha G. Russell, Rahul C. Basole, Jaakko Salonen, Neil Rubens (2013) "Networks of innovation relationships: multiscopic views on Finland," *International Society for Professional Innovation Management (ISPIM) 2013, Helsinki.*

This paper was presented at The XXIV ISPIM Conference – Innovating in Global Markets: Challenges for Sustainable Growth in Helsinki, Finland on 16-19 June 2013. The publication is available to ISPIM members at www.ispim.org.

Networks of innovation relationships: multiscopic views on Finland

Kaisa Still*

VTT Technical Research Centre of Finland, Kaitoväylä 1, FI-90571 Oulu, Finland.
E-mail: kaisa.still@vtt.fi

Jukka Huhtamäki

Tampere University of Technology, IISLab
Korkeakoulunkatu 3, 33720 Tampere, Finland.
E-mail: jukka.huhtamaki@tut.fi

Martha G. Russell

Human Sciences Technology Advanced Research Institute, Stanford University, Cordura Hall, 200 Panama Street, Stanford CA, 94305.
E-mail: martha.russell@stanford.edu

Rahul C. Basole

School of Interactive Computing & Tennenbaum Institute, Georgia Institute of Technology, 85 Fifth Street NW, Atlanta GA 30332, USA.
E-mail: basole@gatech.edu

Jaakko Salonen

Tampere University of Technology, IISLab
Korkeakoulunkatu 3, 33720 Tampere, Finland.
E-mail: jaakko.salonen@tut.fi

Neil Rubens

Active Intelligence Group, Knowledge Systems Laboratory, 1-5-1 Chofugaoka, Chofu, Tokyo 182-8585, Japan.
E-mail: neil@hrstc.org
* Corresponding author

Abstract: In this study, we present a solution for describing and visualizing networks of innovation relationships in the context of a single nation, in this case Finland. We resolve the limitations of separate datasets by building multiscopic views into networks of innovation relationships, using separate datasets as well as an aggregated dataset that federates them. We proceed to support the interpretation of these visualizations explaining context with network metrics as well as other descriptions. Our approach allows examining the relationships needed for value co-creation at various levels of the ecosystem as well as between those levels, providing novel possibilities for network orchestration and innovation management. Our practical suggestions include active communication and data sharing using a wide variety of media, and utilizing network views for targeted actions as well as for creating shared understanding and vision.

Keywords: Innovation; networks; relationships; multiscopic; Finland; metrics; data-driven; visualization; transformation; orchestration

1 Background

This research addresses the challenge of managing innovation in an increasingly global business environment. While there is a growing recognition that networks, relationships, and ecosystems are essential in understanding innovation today, very little is known about the character, role and impact of multilevel relationships (i.e. relationships between organizations, between organizations and individuals, and between individuals) on accelerating company growth and enabling a local to global context transformation.

Our research aims to fill this important scholarly gap by providing a "multiscopic" view of innovation ecosystems that allows for actionable insights into the complex, evolving relationship structure within and across multiple levels. These views are accomplished using separate, though complementary data about the various actors of the ecosystem, and especially with aggregating data representing them to explore the ecosystem. Through resulting network analytics and visualizations, we support the different stakeholders of the innovation ecosystem with their innovation management and network orchestration activities.

1.1 Networks of innovation relationships

The shift of innovation from a single firm toward an increasingly network-centric activity (Chesbrough 2003) has added significant complexities to innovation management. The importance of collaboration and value co-creation (Ramaswamy and Goullart 2010) and resulting networks of relationships (Kogut and Zander 1996, Vargo 2009) between individual and organizational entities (i.e. policy makers; educational institutions; venture capitalists, business angels and other investors; serial entrepreneurs; employees, managers and board-members of start-ups, growth companies and in established companies as well as the entities surrounding them) have consequently led to the study of innovation ecosystems (Iansiti and Levien 2004, Russell et al. 2011, Basole et al. 2012, Hwang and Horowitt 2012, Mars et al. 2012).

It is generally acknowledged that (networks of) relationships are at the core of innovation ecosystems shaping the behaviour and outcome of all stakeholders as well as the system-level effects (Hwang and Horowitt 2012). This perspective is further corroborated by Burke (2011) who argued that "Innovation is about people. Once you remove the obstacles to entrepreneurship, the most important ingredient is the network." The ability to connect and manage competencies across a broad network of relationships is considered as one of the most important meta-capabilities for a networked world (Wind et al. 2008) and is commonly referred to as network orchestration (Russell et al. 2011, Nambisan and Sawhney 2011).

1.2 Toward multiscopic views

The introduction of the network perspective, and especially that of social structures (Wasserman and Faust, 1994) as the defining characteristic of innovation ecosystems, allows for utilizing visual analysis of social networks for exploring innovation ecosystems and their clusters of unique actors and unique reciprocal links among them (Chandler and Vargo 2011). Visualizations enable researchers and other stakeholders to 'see' the structural context and the scalable influence of the context within market structures (Freeman 2009, Chandler and Vargo 2011), showing the connections of individual nodes, organizations or the network at large (Basole et al. 2011). Furthermore, 'seeing' with multiple layers of views, outlooks or perspectives offers advantages in addressing the inherent complexities of innovation.

There is very little theoretical understanding on how ecosystems emerge and evolve, or how to address innovation in multiple levels (Ahuja et al. 2011). Methodological approaches to quantitatively study these transformation phenomena have focused on event sequences at single levels in the biotechnology sector (Owen-Smith and Powell 2004), local innovation ecosystems (Hwang and Horowitt 2012), national innovation ecosystem (Huhtamäki et al. 2011), and knowledge-intensive industries (Iansiti and Richards 2006). Still, theoretical concepts of addressing multiple levels of innovation and their structures are available. For example Nahapiet and Ghosnal (1998) when talking about the dimensions of social capital, have introduced three distinctive levels, micro-meso-macro, addressing first individual contacts and personal relations, then social networks, and finally institutions. A similar naming convention can be found in research addressing resource integration and structurization of service ecosystems, in which levels are contexts that influence each other (Chandler and Vargo 2011), hence adding dimensionality to the networks and their visualizations. The micro-context is seen to frame exchanges among actors as dyads; the meso-context as triads (which are based on the dyad of micro-context); and the macro-context as complex networks (based on triads of the meso-context); with service ecosystems as meta-layers of context.

As the methods to explore ecosystems have developed, so have the computational capabilities that allow for managing vast amounts of data continuously generated by actors and their activities in innovation ecosystems (McKinsey 2011, Kohlhammer et al. 2012). This data can be accessible through company reports and other company filings (such as patent filings) contributing to official government data about companies, as well as in data shared or contributed to social media—all providing data that can link the entities of ecosystems together (for example by the alliances or other deals signed by companies, by linking individuals to companies where they are employed, etc.), and allowing for network and data-driven approaches.

2 Research methodology and findings: case Finland

Our earlier data-driven studies have revealed insights about ecosystemic innovation and its actors on multiple levels, for example about EIT ICT Labs at local and European level (Still et al. 2011, Still et al. 2012) and about the converging mobile ecosystem at the firm and individual level (Basole et al. 2012, 2013). However, showing the interactions between the different levels or perspectives has been historically constrained by the limitations of separate datasets.

In this study, we proceed to bridge the limitations of separate datasets by building multiscopic views into networks of innovation relationships, using separate datasets as well as an aggregated dataset that federates them. Hence, we are addressing validity, which is one of the key challenges of data-driven research (Barnes and Vidgen 2006). It can be managed with data-triangulation for building a richer, more complete picture of the phenomena under investigation and for validating and cross-checking findings, in particular when data from different sources point to congruent insights (Kaplan and Duchon 1988).

We apply a four-stage process for analysing a business ecosystem (Basole et al. 2013). It consists of (1) boundary specification for determining the primitives (nodes, relationships) of the networks as well as the analysis timeframe (2) metrics identification for selecting the appropriate social network and graph theoretic metrics for understanding the dynamics of an ecosystem, (3) computation, analysis and visualization toward analysing and visualizing temporal, relational ecosystem data, and (4) sense-making and storytelling, describing the processes from data to understanding and visual narratives for telling the story.

2.1 Boundary specification

This study concentrates on Finland. Though small, Finland is generally considered a vibrant innovation ecosystem which has been achieving high results in global rankings such as the Global Innovation Index (#4 in 2012) and the Global Competiveness Index (#3 in 2012-2013), with some very successful start-ups and growth companies (such as Supercell and Rovio) and some established companies with global presence (Nokia and Nokia Siemens Networks). In addition, in our previous research (Huhtamäki et al. 2011, Huhtamäki et al 2012) we have explored the Finnish ecosystem, and are familiar with it. Hence, we used Finnish companies and the ecosystem around them as our case for exploring data-driven network analytics at multiple levels between organizations and individuals. In this study, we focused on the recent five years of data (from Jan 1, 2008 to Dec 31, 2012) to provide timely insights and possibilities for comparisons of temporal changes.

Traditional company data sources tend to have data about the established, larger companies; start-ups and growth companies are oftentimes missing from that data. Therefore, we complemented Thomson Reuters' SDC dataset—one of the most prominent sources of inter-firm relationships (Schilling 2009)—with two datasets that reveal the relationships of start-ups and growth companies, IEN Startup and IEN Growth. These two data sources provide socially curated (or crowd-sourced) rich data about companies at the meso- and micro-levels, as well as individuals and investors related to them in almost real-time, though with "public bias".

For bringing out the variety of actors of the Finnish ecosystem and showing the specific types of relationship highlighted in each of the datasets, we use the micro-meso-macro naming convention. Founders and angels and their relationships with start-ups drive the microscopic view; executives and financing relationships with growth companies drive the mesoscopic view; and the macroscopic view is generated by the deals and alliances at the enterprise level.

Table 1 Data sources enabling multiscopic views

For	*Microscopic view*	*Mesoscopic view*	*Macroscopic view*
Source of data	IEN Start-up dataset: socially curated English language data from news, press releases, and social media; data on more than 100,000 companies and individuals; updated quarterly.	IEN Growth dataset: socially curated English language data from news, press releases, and social media; data on more than 100,00 companies and individuals; updated quarterly	SDC Platinum 4.0: proprietary (Thomson Reuters Financial) based on U.S. SEC data; more than 1.9 million financial transactions, updated monthly.
Ecosystem entities	Firms, investors, individuals: Time stamps on individuals; strong emphasis on datadated from 2010	Firms, investors, individuals	Firms
Types of relationships	Founders and angels: *prominent individuals and companies and their relationships with location "Finland"*	Executives and financing: *Finnish companies and their relationships*	Deals and alliances: *Finnish companies and their relationships to any company*

Going beyond the snapshots of relationship networks for innovation, provided by the lenses of these datasets, we then combined the three datasets toward an aggregate dataset. In the aggregated dataset, the three datasets in use are complementary but, at the same time, partly overlapping, necessitating a refinement and curation process similar to what is being applied e.g. in data journalism (Gray, Bounegru and Chambers 2012).

2.2 Metrics identification

The metrics for understanding the dynamics of an ecosystem are categorized based on the distinct but related levels of analysis: the network as the whole (ecosystem) and the node level (firm/individual) (Basole et al. 2013). As these metrics reveal insights about the types of links in the ecosystem as well as the structure of the ecosystem, we use standard metrics (density, diameter, components) to describe the whole network and network clusters of the Finnish ecosystem.

For understanding the roles of individual nodes (actors in the ecosystem), we use node degree and betweenness centrality. Node degree values show the number of connections for a given node, indicating its immediate connectivity and importance in the networks. The betweenness centrality value equals the number of times a given node appears in the shortest path from all nodes in the network to all others. Hence, betweenness centrality shows the importance of a node in bridging the different parts or components of the network together.

2.3 Computation, analysis & visualization

First, a projection was created for each dataset including the Finnish companies, their directly connected actors and interconnections between the actors, using time-span of 5 years. The result is a cumulative 1-step networks include all of the relationships formed during the timeframe. Next, an aggregate dataset was created from the three different datasets and duplicate entities for companies, individuals and other actors were merged.

As can be seen from the descriptions of the multiscopes (Table 2) with different views, the datasets included significant numbers of Finnish companies and their relationships with other companies, individuals and financing organizations. Using the metric of betweenness and degree as defining factors, the top 10 actors from each dataset were identified (note: following the practices and guidelines related to privacy, in this research we do not provide the names of individuals). Some actors were found to have positions as key connecting nodes in more than one view, suggesting interlocking relationships between the different levels. However, their roles were different.

Table 2 Descriptions of the multiscope of cumulative networks of case Finland

View	*Microscopic*	*Mesoscopic*	*Macroscopic*	*Multiscope*
Nodes	844	821	231	1698
Connections	883	824	186	1664
Network description	Directed relationships, multimode network	Directed relationships, multimode network	Undirected relationships, 1-mode network	Directed and undirected relationships, multimode network
Network metrics	density: 0.002 diameter: 15	density: 0.002 diameter: 18	density: 0.007 diameter: 4	density: 0.001 diameter: 16
Top 10 actors based on betweenness centrality	Startup Sauna Nokia Ind-PK SunyRide Ind-KB Ind-TT Transfluent Ind-JE Ind-AK Ind-VM	Nokia Ind-PK Ind-TT Ind-JE Applifier Mendor Finnish Industry Investment WOT Services XIHA Tinkercad	Nokia Nokia Siemens Networks Wartsila Metso Kemira Finnair Microsoft Outokumpu Stora Enso Ilmarinen	Nokia Ind-PK Startup Sauna Ind-TT WOT Services SunyRide Ind-JE Mendor Ind-AK Ind-KB
Top 10 actors based on degree	Nokia Startup Sauna Ind-AK Ind-TL Ind-TT Holvi Ind-PK Ind-JE Ind-KL Ind-VM	Nokia Blyk Fruugo Grand Cru Rovio Entertainment XIHA Flowdock Finnish Industry Investment Sofanatics Sulake	Nokia Nokia Siemens Networks Wartsila Finnair Tabuk Cement Alcatel Lucent Telefonaktie–bolaget LM NEC NTT Docomo Zecon Bhd Elematic Fujitsu	Nokia Startup Sauna Ind-TL Nokia Siemens Networks Ind-AK Ind-TT Grand Cru Holvi Blyk Rovio Entertainment

In the resulting visualizations (created using Gephi), the nodes (points) represent the various actors, with lines between them indicating relationships. The size of the node signals its role based on betweenness centrality. Node color shows its type: blue is for individuals, red for companies, green for investors, and light green for incubators. Finnish companies are highlighted in orange.

The macroscopic view highlighting enterprise level relationships (Figure 2) depicts a landscape of a rather loose network with many dyadic company relationships. However, its comparably higher density can be explained with its composition as a 1-mode network, where all nodes can be connected. Only a few Finnish companies are connected to more than one company. Both Nokia and Nokia Siemens Networks are shown as the most prominent nodes that have each collected a cluster of companies around them, emphasizing their role of connecting the Finnish ecosystem to the world. The cluster including players from more traditional industries – Wartsila, Metso and Kemira – indicates their connecting role both within Finland as well as globally. Interestingly, Rovio Entertainment, which by many is still regarded more of a growth company than an established company, is present in this view, due to its enterprise level relationships. Due to the nature of the data, all of the actors in the top 10 based on betweenness are companies, including Microsoft which is there due the its strategic alliance with Nokia.

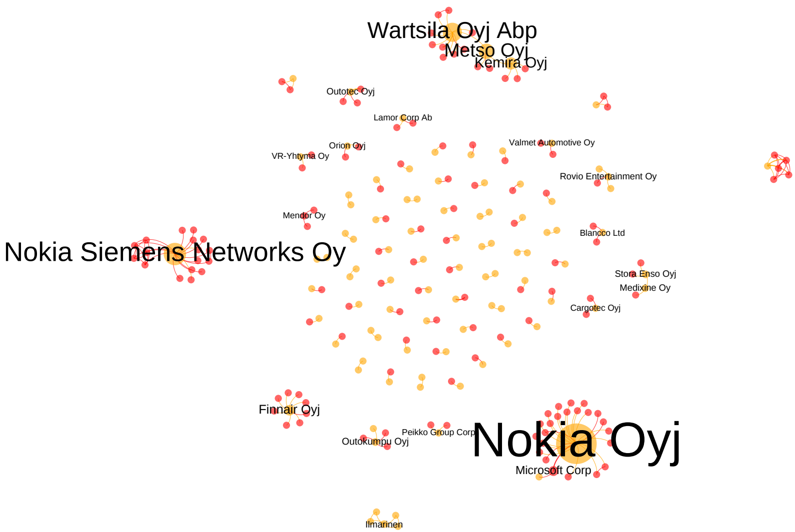


Figure 2 Macroscopic view highlighting enterprise level relationships

The mesoscopic view highlighting growth companies (Figure 3) shows many Finnish companies with relationships to 1-3 actors; it also introduces more complex, networked relationships, showing a chain of nodes connecting key nodes. The key nodes that act as bridges between various network actors are not only companies (such as Nokia and WOT Services), but include also prominent individuals—in their roles as company executives, advisors and investors—as well as financing organizations (Finnish Industry Investment).

For Rovio Entertainment, this view indicates the connected individuals as well as investors. Accordingly, three out of the top 10 actors based on betweenness are individuals. However, as degree measures the number of connections, all top 10 actors based on it are companies, including one financing organization.

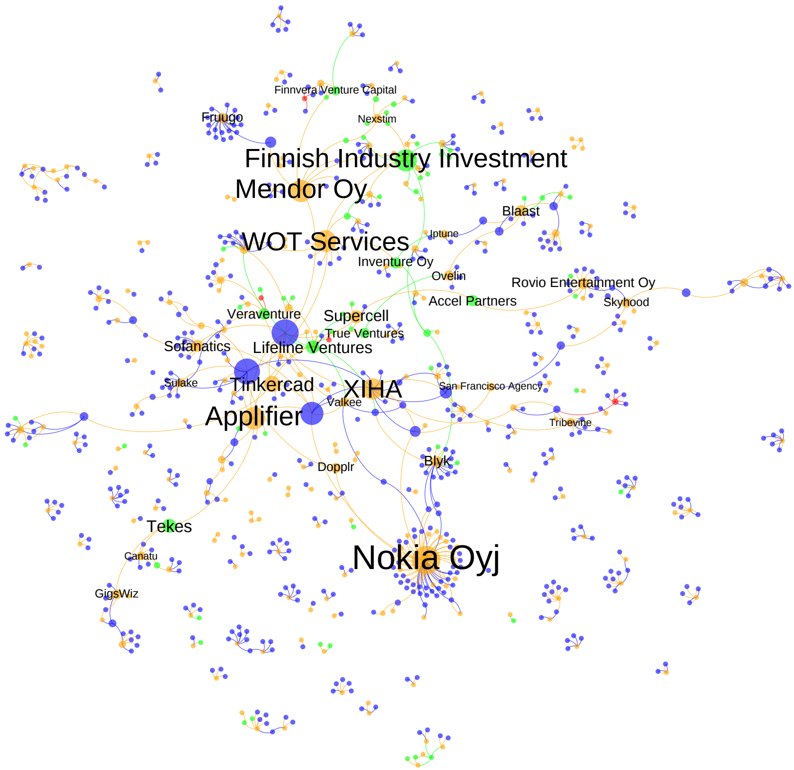


Figure 3 Mesoscopic view of growth companies

The microscopic view highlighting startup companies (Figure 4) illustrates an intricate web of connections within the Finnish ecosystem. In addition to start-ups, the key nodes now include prominent individuals (in roles of founders, advisors and angels) as well as a business incubator, Startup Sauna, reflecting this particular incubator's role as active advocate of start-up culture as well as home for start-ups, a place for building relationships. Six out of the top 10 actors based on betweenness are individuals. Highlighting the emphasis on individual connections, most of the top 10 actors based on degree are individuals. The role of Nokia is again important, as individuals with Nokia background are connected to other companies and thus interconnecting the Finnish ecosystem. As this view is drawn from data centered on individuals and their relationships to startups, a number of non-Finnish actors are introduced.

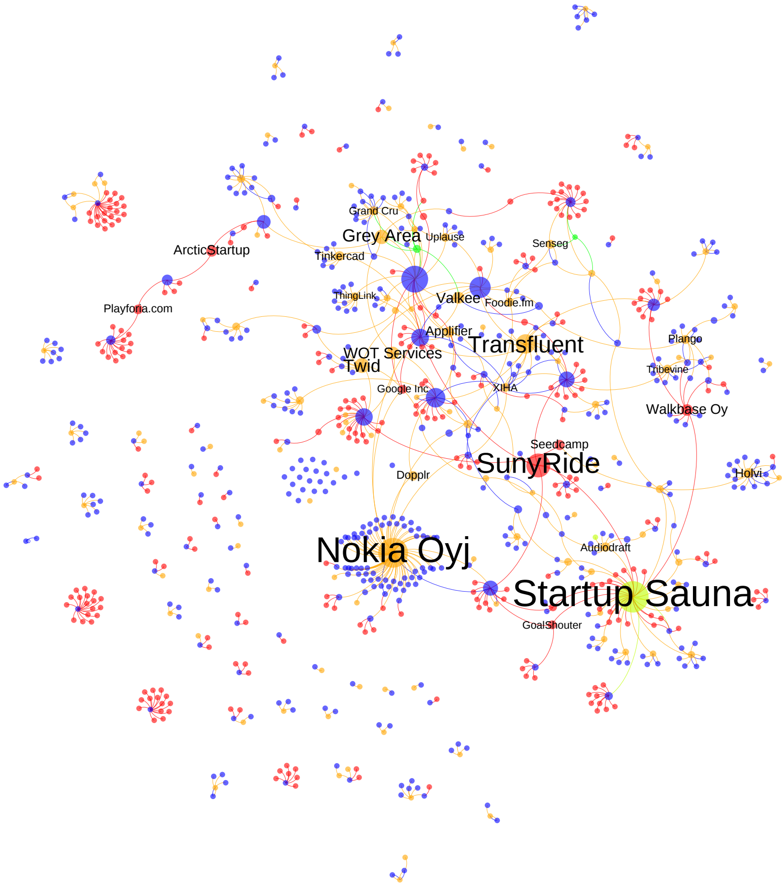


Figure 4 Microscopic view of start-up companies

The aggregated network depicts an ecosystemic view of Finland (Figure 5) as it combines the Finnish companies from the three separate datasets, and shows their direct connections. Hence, for the first time, we can see in a single visualization the founders and angels, executives and financing organizations, as well as companies from start-ups to established enterprises. Overall, key actor of the ecosystem with the highest betweenness centrality is not surprising: Nokia is the super-node underscoring its connective role in the Finnish ecosystem. Accordingly, the same companies, financing organizations and individuals that have been prominent in previous lists and visualizations are highly visible in this ecosystemic view. As the weight of data from the micro and meso levels is greater, the top 10 of actors in the ecosystem based on their betweenness as well as degree includes a significant number of individuals. There are 7 shared nodes between micro and macro views; 184 between micro and meso views; 10

between meso and macro views; four nodes appear in all three views: Rovio Entertainment, F-Secure, Mendor and Nokia.

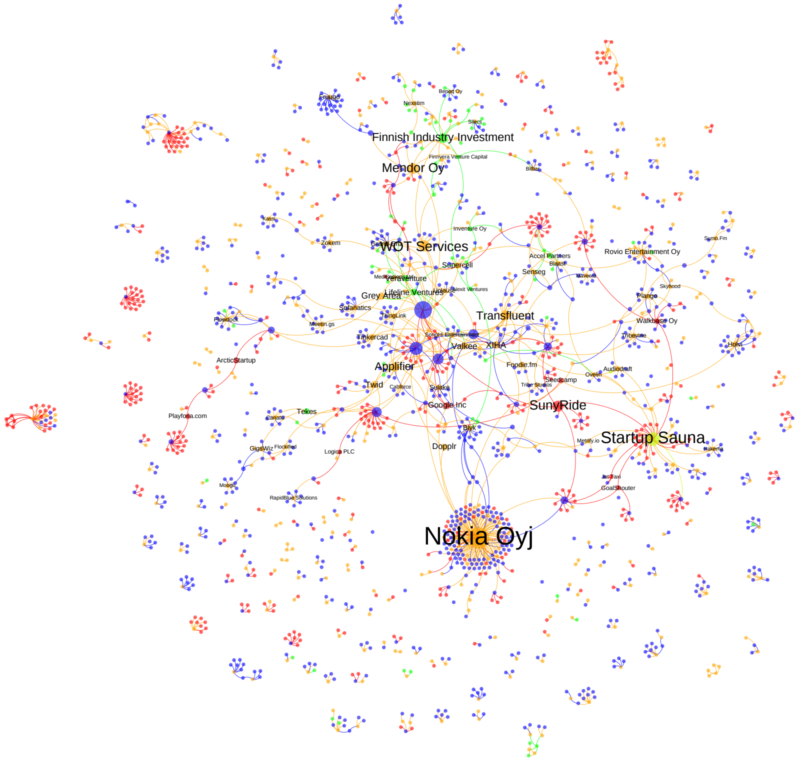


Figure 5 Aggregate view to the ecosystem

2.4 Sense-making & storytelling

Our visualizations of metrics and networks can be seen to model the skeleton of an ecosystem. However, they rely on human insights for emerging “sense-making” as well as for forming narratives and telling the stories that help stakeholders view and interpret the images. The visualizations are highly contextual and, for most stakeholders, are interpreted in the context of the user by the user’s actions such as inspecting, ranking, comparing, categorizing, inferring, associating and correlating (Xu et al. 2009). They make knowledge about existing networks explicit, however, the interpretation and understanding of the visualizations is built on tacit knowledge and ground truths of individuals investigating them. Hence, providing information to support users in sense-making is essential. Additional multiscopic context with explanatory insights about the data sources as well as processes used for curation and visualization can be used for improving the sense making processes. For example:

- Understanding the different sources of data (Table 1)—official, curated data vs. socially constructed data with public bias—and their impact on richness as well as timeliness of the data can build insight for interpreting the multiscopic views drawn from the aggregate dataset. It is also important to point out that this data is based on English language sources.
- The metrics presented (Table 2) about the primitives of the network (number of nodes and connections) indicate that each of the three separate datasets provide only a limited view into the multitude of relationships of Finnish companies, both in Finland as well as globally. At the same time, by showing the numbers of shared nodes—entities that are present in more than one view—it makes visible the existing interconnections between the different views.

Overall, the visualizations highlight the small scale of the Finnish innovation ecosystem. Furthermore, the blue color dominating the overall ecosystem indicates that a small group of individuals form the interconnecting core of the ecosystem. Nokia continues to be the focal point of the network. With some exceptions, investment organizations and venture firms, shown in green, are scarce in the ecosystem. These insights are not new as such but our results show, for the first time, a complete structure of the Finnish innovation ecosystem with a federation of three different data sources.

3 Discussion

This study presents a solution for describing and visualizing an innovation network on multiple levels. The benefit of the visualization approach is that tacit knowledge about the intangible nature of networks becomes visible and shared. Our data-driven visualizations using separate and aggregated data provided multiscopic views on one country, in this study Finland. With additional contextual information, the analytical process was communicated to the observer for supporting the subsequent intepretations processes of sense making and storytelling. This allowed for understanding the scope and limitations of the explicit mapping of relationships, and tracking the changes with the context of Finland, based on this particular data using these particular processes toward data-driven visualizations.

We believe that the benefit of the multiscopic views approach lies in the dimensionality of the overall network that can be understood by considering various levels of the network but also by considering the holistic network. The combined results of the visualizations and interpretations can be used to examine the relationships on various levels with expanding network connections toward understanding the ways relationships have begun to converge. Hence, they provide competitive intelligence and insights into a coherent ecosystem. Using separate and aggregated data, the visualizations of micro, meso and macro levels span from startups to growth companies to established companies for perspectives on company maturity, and traverse toward systemic behavior and outcomes on the ecosystemic innovation.

3.1 Implications for data sharing and data management

We have entered the era of both big data and open data. The success of individual startups, incubators, investors and other innovation ecosystem stakeholders is increasingly dependent on their visibility. We encourage the different stakeholders to make sure that the data representing their key relationships is present in public sources, such as Wikipedia, Angel List, Crunchbase and services alike. Moreover, we harnessing these sources of data for ecosystem analysis in micro, meso and macro level to keep the knowledge of the surrounding ecosystem up to date. For example, our visualizations highlight the role of Startup Sauna in the Finnish innovation ecosystem—however, there are rather successful business incubators also in cities such as Tampere and Oulu that have been noted press-worthy both nationally and internationally but yet do not appear regularly in socially constructed data.

Theoretically, this study contributes to our understanding of how large, disconnected, potentially complementary structured and semi-structured datasets can be leveraged for insight, exploration, and discovery, and how ecosystem complexity can be analyzed and results visually communicated. For the scholar interested in innovation ecosystems, this approach to holistic multiscopic ecosystem analysis invites the exploration of dynamic multiple networks and forces of transposition and refunctionality (Padgett and Powell 2012), such as those elaborated by Padgett (2012) in his analysis of Renaissance Italy. It begs the fundamental questions of emergence and transformation.

For the practitioner, we emphasize that (a) the dataset, (b) the filters for creating the projections, (c) and the rules for creating the connections from the data and for including and excluding nodes, all have an impact on the network metric values and the resulting views. For this reason, we argue that the data acquisition and analytical process should be transparent so that the observer can not only react to the static snap-shot of the network, but can interact with the views created and also with the processes used to create them. With these insights, a social media savvy company can for example easily connect more individuals to their company by publicly sharing company information, and this may make the company more visible in network visualizations, with an impact on increasing the betweenness value of the company.

3.2 Implications for innovation management and policy-making

With the data-driven visualizations, descriptions of the current relationship-based links of the network are revealed, allowing observers to see visual indicators of the broad systems of value co-creation. As each of the resulting visualizations shows a different aspect of the ecosystem, according to the Finnish ecosystem stakeholders, the insights that come from visualizing the "invisible" provide concrete possibilities for improving network orchestration. These activities also provide an opportunity for various stakeholders to come together to discuss their interpretations of the visualizations. In addition to contributing to the understanding of elements and processes shaping the transformation of innovation ecosystems, the process can enhance the discussion about global relationships by company stage (startup, growth, establishment) as well as contribute to the national level discussion of the local-to-global relationships within the Finnish innovation ecosystem.

As influential and connecting actors of the ecosystem are revealed, they can be contacted and included in discussions and other tailored actions. Furthermore, they can be targeted: used for benchmarking activities or for learning about “best practices”, for regional as well as national levels of development. Hence, visualizations and supportive contextual information provide practical tools for innovation ecosystems stakeholders, as well as methods toward the controllability, manageability and orchestrability of the network.

References and Notes

Ahuja, G., Soda, G., and Zaheer, A., 2011. The genesis and dynamics of organizational networks, Organization Science, 23(2), 443-448.

Barnes, S.J., and Vidgen, R.T., 2006. Data triangulation and web quality metrics: A case study in e-government, Information & Management, 43(6), 767-777.

Basole, R.C. and Karla, J, 2011. On the evolution of mobile platform ecosystem structure and strategy, Business & Information Systems Engineering, 3 (5): 313-322.

Basole, R.C., Russell, M., Rubens, N., and Huhtamäki, J., 2012. Understanding Mobile Ecosystem Dynamics: A Data-Driven Approach. 13th International Conference on Mobile Business, Delft, Netherlands.

Basole, R.C., Russell, M., Huhtamäki, J., Rubens N. and Still, K., 2013. Understanding Mobile Ecosystem Dynamics: A Data-Driven Approach, Journal of Information Technology (JIT), Special Issue on Mobile Platforms and Ecosystems. Submitted.

Burke, A., 2011, How to build an innovation ecosystem, The New York Academy of Sciences, http://www.nyas.org/publications/Detail.aspx?cid=da1b8e1d-ed2d-4da4-826d-00c987f63c82 (Accessed April 15, 2013)

Chandler, J.D., and Vargo, S.L., 2011. Contextualization and value-in-context: How context frames exchange. Marketing Theory, 11:35.

Chesbrough, H., 2003. Open innovation: the new perspective for creating and profiting from technology. Boston: Harvard Business School Press.

Freeman, L. C., 2009. Methods of Social Network Visualization, in R. A. Meyers, ed. *Encyclopedia of Complexity and Systems Science*, Berlin: Springer.

Gray, J., Chambers, L., and Bounegru, L., 2012. The Data Journalism Handbook, O´Reilly Media. http://datajournalismhandbook.org

Huhtamäki, J., Russell, M.G., Still, K., and Rubens, N., 2011. A network-centric snapshot of value co-creation in Finnish innovation financing, Open Source Business Resource, March, 13-21.

Huhtamäki, J., Still, K., Isomursu, M., Russell, M.G., and Rubens, N., 2012. Networks of growth: Case young innovative companies in Finland. Proceedings of the 7th European Conference on Innovation and Entrepreneurship, Santarém, Portugal, September 20-21, 2012.

Hwang, V.W. and Horowitt, G., 2012. *The Rainforest: The Secret to Building the Next Silicon Valley*. Los Altos Hills, CA: Regenwald.

Iansiti, M., and Levien, R., 2004. The keystone advantage: *What new dynamics of business ecosystems mean for strategy, innovation, and sustainability*. Boston: Harvard Business School Press.

Iansiti, M., and Richards, G.L., 2006. The information technology ecosystem: Structure, health and performance. The Antitrust Bulletin, 51 (1), 77-110.

Kaplan, B., and Duchon, D., 1988. Combining qualitative and quantitative methods in information systems research: A case study, MIS Quarterly, 12(4): 571-586.

Kohlhammer, J., Nazemi, K., Ruppert, T. and Burkhardt, D., 2012, Toward Visualization in Policy Making, IEEE Computer Graphics and Applications, 32 (5): 84-89.

Kogut, B., and Zander, U., 1996. What firms do? Coordination, Identity and Learning. Organization Science, 7 (5), 502-518.

Mars, M.M., Bronstein, J.L. and Lusch, R.F., 2012. The Value of a metaphor: Organizations and Ecosystems, Organizational Dynamics, 41,271-280.

McKinsey, 2011. Big Data: The next frontier for innovation, competition and productivity. McKinsey Global Institute, Overall Presentation, October. http://www.jegi.com/sites/default/files/McKinsey_Presentation.pdf (Accessed Feb 14, 2013).

Nahapiet, J. and Ghoshal, S. 1998. Social capital, intellectual capital, and the organizational advantage, Academy of Management Review, 23, 242-266.

Nambisan, S. and Sawhney, M., 2011. Orchestration Processes in Network-Centric Innovation: Evidence from the Field, Academy of Management Perspectives, 25(3).

Owen-Smith, J., and Powell, W.W., 2004. Knowledge networks as channels and conduits: The effects of spillovers in the boston biotechnology community, Organization Science, 15(1), 5-21.

Padgett, J.F. and Powell, W.W., 2012. The problem of emergence, in J.F. Padgett and W.W. Powell, eds., *The emergence of organizations and markets*, Princeton, NJ: Princeton University Press.

Padgett, J.F., 2012. Transposition and refunctionality: The birth of partnership systems in Renaissance Florence, in J.F. Padgett and W.W. Powell, eds., *The emergence of organizations and markets*, Princeton, NJ: Princeton University Press.

Ramaswamy, V., and Gouillart, F., 2010. Building the Co-Creative Enterprise. Harvard Business Review, 88 (10).

Russell, M.G., Still, K., Huhtamäki, J., Yu, J., and Rubens, N., 2011. Transforming Innovation Ecosystems through Shared Vision and Network Orchestration, Proceedings of Triple Helix IX Conference, July 2011, Stanford University.

Schilling, M.A., 2009. Understanding the alliance data. Strategic Management Journal, 30 (3), 233-260.

Still, K., Russell, M.G., Huhtamäki, J. and Turpeinen, M., 2011, Explaining innovation with indicators of mobility and networks: insights into central innovation nodes in Europe, Proceedings of Triple Helix IX Conference, July 2011, Stanford University.

Still, K., Huhtamäki, J., Russell, M.G. and Rubens, N., 2012, Transforming Innovation Ecosystems through Network Orchestration: Case EIT ICT Labs, Proceedings of the XXIII ISPIM Conference—Action for Innovation: Innovating from Experience, June 17-20, Barcelona, Spain.

Vargo, S., 2009. Toward a transcending conceptualization of relationship: a service dominant logic perspective. Journal of Business and Industrial Marketing, 24, (5/6), 373-379.

Wasserman,S., and Faust, K., 1994. *Social Network Analysis: Methods and Applications*. 1st Edition. New York, NY: Cambridge University Press.

Wind, J., Fung, V. K. K., and Fung, W., 2008. Network Orchestration: Core Competency Borderless World, in J. Wind, V.K.K. Fung, and W. Fung. *Competing in a Flat World: Building Enterprises for a Borderless World*. Upper Saddle River, NJ: Wharton University Publishing.

Xu, S., Chen, X. and Liu, D, 2009. Classifying software visualization tools using the Bloom’s taxonomy of cognitive domain, Proceedings of the Electrical and Computer Engineering, CCECE´09, Canadian Conference on, IEEE, 13-18.

Publication VI

Kaisa Still, Jukka Huhtamäki, Martha G. Russell, Neil Rubens (2014) "Insights for orchestrating innovation ecosystems: the case of EIT ICT Labs and data-driven network visualisations," *International Journal of Technology Management.*

Available at `http://urn.fi/URN:NBN:fi:tty-201501201000`

Insights for orchestrating innovation ecosystems: the case of EIT ICT Labs and data-driven network visualisations

Kaisa Still*

VTT Technical Research Centre of Finland,
Kaitoväylä 1, 90571 Oulu, Finland
E-mail: kaisa.still@vtt.fi
*Corresponding author

Jukka Huhtamäki

Tampere University of Technology,
Korkeakoulunkatu 3, 33720 Tampere, Finland
E-mail: jukka.huhtamaki@tut.fi

Martha G. Russell

mediaX at Stanford University,
Cordura Hall, 210 Panama Street,
Stanford, CA 94305-4115, USA
E-mail: martha.russell@stanford.edu

Neil Rubens

Graduate School of Information Systems,
University of Electro-Communications,
182-8585 Tokyo, Japan
E-mail: rubens@ai.is.uec.ac.jp

Abstract: This paper explores opportunities for supporting the orchestration of innovation ecosystems, hence contributing to a fundamental capability in the networked world. We present analysis, evaluation and interpretation toward the objective of decision support and insights for transforming innovation ecosystems with a case study of EIT ICT Labs, a major initiative intended to turn Europe into a global leader in ICT innovation. Towards this, we use a data-driven, relationship-based and network centric approach to operationalise the 'innovation ecosystems transformation framework'. Our results indicate that with coordinated and continuously improved use of visual and quantitative social network analysis, special characteristics, significant actors and connections in the innovation ecosystem can be revealed to develop new insights. We conclude that the IETF transformation framework can be used to develop shared vision and to support the orchestration of innovation ecosystem transformations.

Keywords: innovation ecosystem; network visualisation; social network analysis; SNA; network orchestration; data-driven; transformation; technology management; EIT ICT Labs.

Reference to this paper should be made as follows: Still, K., Huhtamäki, J., Russell, M.G. and Rubens, N. (2014) 'Insights for orchestrating innovation ecosystems: the case of EIT ICT Labs and data-driven network visualisations', *Int. J. Technology Management*, Vol. 66, Nos. 2/3, pp.243–265.

Biographical notes: Kaisa Still studies innovation, technology transfer and the role of technology with a focus on information and knowledge creation, sharing and management. Her studies focus on supporting collaboration and cooperation in organisations as well as in community settings. Her recent studies include mobile, online and social networking communities, innovation ecosystems and innovation indicators. She has over ten years of cross-sector business and academic experience in Finland, USA and China. She currently works for VTT Technical Research Centre of Finland and is a founding member of Stanford's Innovation Ecosystems Network.

Jukka Huhtamäki is Researcher and Teacher at the Intelligent Information Systems Laboratory (IISLab) at Tampere University of Technology and founding member of the Innovation Ecosystems Network. His interests include the methods, processes and tools of data-driven visual analytics, social network analysis and the development of modern hypermedia. Currently, he is working to develop means and methods of data-driven visual network analytics for insights on the structure and dynamics of innovation ecosystems. He is also preparing a PhD on the topic.

Martha G. Russell is senior research scholar at Human Sciences Technology Advanced Research Institute and Executive Director of mediaX at Stanford University, a membership-based, interdisciplinary research catalyst focused on people, media, technology and innovation. Her background spans a range of business development, innovation and technology-transfer initiatives in information sciences, communications and microelectronics at the University of Minnesota, The University of Texas at Austin and Stanford University. She is a founding member of Stanford's Innovation Ecosystems Network and serves on the editorial board of the journal *Technology Forecasting and Social Change*.

Neil Rubens is an Assistant Professor at the Knowledge Systems Laboratory, University of Electro-Communications, Japan. He is the Director of Active Intelligence Research Group and is a founding member of the Innovation Ecosystems Network at Stanford University. He holds an MSc degree from the University of Massachusetts and a PhD degree from the Tokyo Institute of Technology – both in Computer Science. His research focuses on developing active intelligence systems, which are artificial intelligence systems that are self-adaptable utilising unsupervised and semi-supervised learning and active communication and data acquisition.

This paper is a revised and expanded version of a paper entitled 'Transforming innovation ecosystems through network orchestration: case EIT ICT labs' presented at the XXIII ISPIM Conference – Action for Innovation: Innovating from Experience, Barcelona, Spain, 17–20 June 2012.

1 Introduction

This study is motivated by the pursuit of network-based findings to reveal new insights on how interventions can be orchestrated to facilitate transformation of an innovation ecosystem. Our study is based on the understanding that firms are embedded in networks of relationships that remarkably affect their potential success in the markets (Ritala et al. 2009). These complexities related to innovation have increasingly been addressed with the term ecosystem (Durst and Poutanen, 2013). Orchestration, or network orchestration, refers to capability to purposefully build and manage inter-firm innovation networks (Dhanaraj and Parkhe, 2006); when network-level, collective gains are sought, organisations seek to assemble or orchestrate networks and manage their growth (Paquin and Howard-Grenville, 2013) which we explore in the context of innovation ecosystems.

In this study, we show how data-driven network visualisations can be used to produce insights for orchestrating innovation ecosystems. Our data-driven approach stems from the potential of the vast sea of available data which can be referred to as information overload or as big data; it is also touted as the next frontier for innovation, competition and productivity (McKinsey, 2011). Big data is seen to provide possibilities for promoting better measurement, better management and better decisions (McAfee and Brynjolfsson, 2012). For this study, we see that openly available data about innovation, coupled with its analysis and presentation, provides possibilities for insights that can promote better measurement, better management and better decisions in the context of innovation ecosystems. The network visualisations demonstrate how connections at the level of the individual nodes and links can have complex effects that ripple through the ecosystem as a whole (Easley and Kleinberg, 2010).

Network orchestration is an understudied process with mainly conceptual studies addressing it (Ritala et al., 2009). A better understanding of it is considered of both scholarly and practical importance (Paquin and Howard-Grenville, 2013) for "an integrated understanding of the mechanisms for value creation and capture in the innovation ecosystem context" [Ritala et al., (20130m p.246]. Hence, in this study we attempt to provide empirical qualitative and quantitative evidence for supporting network orchestration in the form of data-driven network visualisations. In addition, we demonstrate how these visualisations can be used to produce insights for orchestration for innovation ecosystems. We explore the possibilities for supporting understanding, monitoring and managing innovation ecosystems and their transformations using innovation ecosystem transformation framework (IETF) which has been previously and successfully used to create insights on network orchestration (Russell et al., 2011).

The structure of the paper is as follows: we begin with an overview on previous studies, from which our approach is derived, then we describe the data sample and its analysis, which allows us to discuss the insights based on the findings, as well as the opportunities they provide for orchestrating transformation. Finally, we present recommendations for replication and extension of this approach and we describe limitations of our study. Overall, with this research we invite researchers, programme managers and policy makers to embrace the value of understanding and measuring complex relationships underlying innovation in a networked world.

2 Theoretical background

2.1 Innovation ecosystems

Sustainable innovation activities are rarely carried out by a single individual or within a single organisation; they are sometimes addressed with the ecosystem approach. Innovation ecosystems, generally seen as entities consisting of organisations and connections between them, have been defined as human networks that generate extraordinary creativity and output on a sustainable basis (Hwang and Horowitt, 2012) and also as consisting of interdependent firms that form symbiotic relationships to create and deliver products and services (Basole and Rouse, 2008). A broader definition sees innovation ecosystems as a network of relationships through which information, talent and financial resources flow through systems, creating sustained value co-creation (Russell et al., 2011), including human networks and firm-level networks as well as the "inter-organizational, political, economic, environmental and technological systems of innovation through which a milieu conducive to business growth is catalysed, sustained and supported" [Russell et al., (2011), p.3].

Networks are described by connections or social links (Krackhardt and Hanson, 1993) and as nested structures of individuals, firms and their relationships (Halinen et al., 2012). Addressing ecosystems as networks allows studying their complex relationships, providing means for mapping the ecosystem structure to support its monitoring and management, sometimes addressed as orchestration. Also from the policy side, networks have been at the centre of attention. The significance of actors and the relationships between them have become targets for innovation policy, under "the rationale for network formation and for their support is the assumption that the whole (the network) is greater than the sum of its individual parts (the network members) in terms of the activities performed" (Cunningham and Ramlogan, 2012).

The utility of network modelling for studying innovation ecosystems comes from the revelation of patterns of connections and interactions within an ecosystem that are captured (Green and Sadedin, 2005) and revealed as structures. Social network analysis (SNA) studies the structure of networks of social actors (Wellman, 1988). SNA has been used to study the sociological relationships of people and organisations (Wasserman and Faust, 1994; Welser et al., 2007). For example, node degree is the simplest metric for centrality and connectivity. Degree value shows the number of direct connections of a node (Wasserman and Faust, 1994). Betweenness is another centrality metric useful for measuring the importance of a node's bridging role in a network; betweenness value represents the number of times a particular node is in the shortest path for any node-pair in the network (Wasserman and Faust, 1994).

With the rise of consumer-generated content, SNA has been deployed to analyse communication structures, content and virality in social media (Welser et al., 2007) and promises to do so also for other sources of big data. Recently, Liu et al. (2011) have shown that understanding the structure of a network is a key factor in the controllability of both engineered and real complex networks.

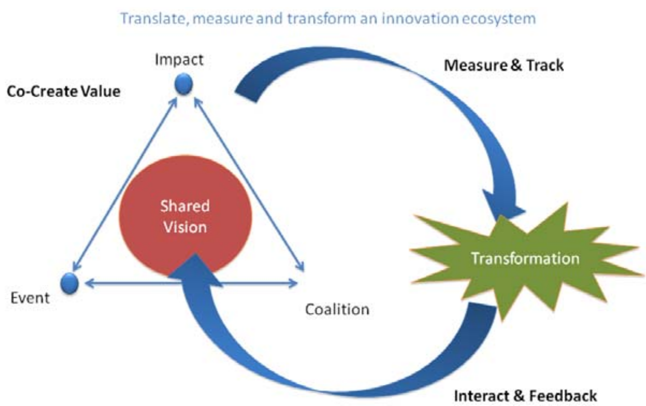
2.2 Orchestrating transformation

The concept of network orchestration goes beyond both knowledge management and innovation management, to include 'discrete influence' that addresses the interdependencies and flexibility of actors in the network (Rizova, 2006). This perspective enables coordination of the innovation network and signals for the innovation output (Dhanaraj and Parkhe, 2006). Increasingly, networks are intentionally 'orchestrated' or 'engineered' by an organisational actor who recruits network members and shapes their interactions, corresponding to phases of innovation ecosystem building and management (Ritala et al., 2013); the impacts of such orchestration have been shown to be pervasive, robust and long-lived (Paquin and Howard-Grenville, 2013). The ability to connect and manage competences across a broad network of relationships has also been seen as one of the most important meta-capabilities for a networked world (Wind et al., 2008; Ritala et al., 2009). In accordance, there have been many programmes of government interventions to create and support networks (Cunningham and Ramlogan, 2012).

The capacity to continually co-create and maintain value is essential (Christensen, 1997). To fully explore the processes in innovation ecosystems that are enacted through time and how nested network structure shapes the process, Halinen et al. (2012) recommend examining relationships and interactions. Furthermore, network orchestrators are urged to engage in sense-making for external audiences who have little or no prior understanding of the transformation activity and its 'rightness' (Möller and Rajala, 2007).

The goal of network orchestration is guided transformation of the ecosystem with continuous co-creation that allows the evolution of the processes needed to motivate and realise the transformation (Russell et al., 2011). This process evolution accommodates the complex influences on innovation in a networked world and energises innovation processes and outcomes. Through the lens of the IETF, a shared vision of the transformational potential of a dynamic innovation ecosystem is created through changes in actors, the events that they enable and the coalitions reflected in their relationships. The infrastructure of the network evolves through their coalitions, accommodating and stimulating innovation in line with their objectives and the collective shared vision. Actors perform roles as arbiters, catalysts and gatekeepers in open and closed-elite dynamics across time (Powell and Owen-Smith, 2013).

Figure 1 Innovation ecosystems transformation framework (IETF) (see online version for colours)



The IEFT (Figure 1) is based on the premise that shared vision for transformation in an innovation ecosystem is created and continually updated through relationships that motivate and guide decisions to realise that vision. Hence, it simultaneously calls for and allows for action research, which "seeks to bring together action and reflection, theory and practice, in participation with others, in the pursuit of practical solutions for pressing concern to people, and more generally the flourishing of individual persons and their communities" [Reason and Bradbury, (2001) p.1]. It extends the process studies of change in organisations that "conceptualize change as a succession of events, stages, cycles or states in the development or growth of an organization" [van de Ven and Poole, (2005), p.1389)] to transformation as a process that is continually updated and collectively realised, as suggested by Hagel and Seely Brown (2005). It further recognises that shared vision is a significant resource (Hagel and Seely Brown, 2005) for innovation ecosystems. Every decision point for effective change cannot be discussed and approved in committee or agreed before implementation. Across the constituents of change, many critical decisions must be made individually and independently. It is the shared vision of these decision makers that allows their independent decisions to synergise change and transform the present into to a shared future.

People and other resources referred to as actors participate in events that over time effect changes in the initial conditions; one such change is the emergence of new coalitions through which joint participation reveals their relationships, shown as links between the actors. Changes in the actors and changes in their links document modifications in the network and the coalitions that provide its structure. Over time these shifts result in new actors, new events, new impacts and new coalitions that continuously evolve into the shared vision of the future.

Previously the IETF has been used to measure, track and visualise snapshots of regional innovation ecosystems. For example, mapping the local events and participants of projects supported by the Southeastern Minnesota Initiative Fund's regional development revealed the emergence of a regional perspective as community organisations began to include newsletter coverage of events sponsored by related organisations in the region (Southeastern Minnesota Initiative Fund, 1995). The programmatic and financial support networks for afterschool programmes in Dallas County, Texas revealed the accomplishment of shared vision in the programmatic strength of afterschool programmes that relied on multiple sources of support which in turn provided similar types of programmatic services in addition to their financial supports (Russell and Smith, 2011). Relationships between CapDigital companies jointly applying for and receiving government funding awards highlighted the ecosystem growth of a network of Parisian companies pursuing new opportunities as well as programmatic opportunities for further accelerating regional transformation (Russell et al., 2011). Insights about these successes and opportunities when shared in interactive visual format with the CapDigital board of advisors, stimulated ideation and actions to mobilise support for new initiatives (personal conversation with Patrick Cocquet, Cap Digital, Paris, France, 9 October 2013).

2.3 Data-driven network visualisation

Calculating network metrics and tracing their changes over time are methods that have been used to study the longitudinal processes of network orchestration (Paquin and

Howard-Grenwille, 2013). As access to and availability of unprecedented amounts of data about the complex innovation ecosystem and its parts now exist, network visualisations have evolved to a data-driven process (Nykänen et al., 2008) with phases of data collection, refinement, analysis and visualisation (Card et al., 1999). Visual network analysis affords insight into the social configurations of the networks and assists in communicating the findings to others (Freeman, 2009). Hence, visualisations can help us 'see through the forest of data'. They are more than pretty pictures as they allow for real-time exploration of complex, interacting variables (Hadhazy, 2011) and can provide evidence about ecosystem transformation and opportunities for orchestrating this transformation.

The data-driven process starts with data which can exist: in official company data; compiled through surveys; and as organisational data about collaborations and activities within the company and outside the company. Much of this data is proprietary and not easily available. However, information created and shared in social media also exists. This socially constructed data is created as innovation actors such as company founders, entrepreneurs, knowledge and financial investors, journalists, policy makers and customers use social media to share information, discuss events and communicate about their needs, experiences and opinions related to innovation (Still et al., 2012). For example, companies issue press releases and blog about their activities, results of their funding rounds and new personnel. Their information is picked up and added to publicly available sources such as Wikipedia, TechCrunch, CrunchBase, Arctic Startup and AngelList. Socially constructed data has the characteristics of open access and availability, potentially large coverage, timeliness and community verification of data quality. Some of the disadvantages are the potential of incompleteness and inconsistency, lack of established perspective and the issue (although slightly different from that of officially curated data) of incompleteness and inconsistencies. This data, sometimes referred to as 'big data' is by default in digital format which combined with computational power available today, provides potentially revolutionary business intelligence for business advantage and performance improvement and for management decisions based on evidence rather than intuition (McAfee and Brynjolfsson, 2012).

This kind of data can be arranged as relational data to define the relationships within an ecosystem and can be used to create network representation of the ecosystem. The relational context of this data allows the use of SNA metrics. To present the data as a network and its metrics in a visual form, we compiled a set of tailored batch-processing tools in Python for network creation. Moreover, we used Gephi for calculating network metrics as well as for network visualisation and layout. The network layout was created using a force-driven algorithm in which nodes repel each other and the links pull the connected nodes together (Noack, 2009). The resulting network layout reveals the clusters in the network as well as the key nodes and pathways that build bridges among the clusters.

Hence, data-driven network visualisations can be seen to offer a powerful approach to providing evidence-based information when talking about ecosystems, their structures, actors and interactions. The visualisations can reflect the structure of an innovation ecosystem at a single point in time and they can also show the evolution of an ecosystem's actors and their relationships over time (Basole et al., 2012).

3 Methodology

The objective of this research is to explore how data-driven network visualisations can be used to produce insights for orchestrating an innovation ecosystem. The IETF is used to translate or understand the ecosystem to empower change agents to measure and transform it with network orchestration, extending the process studies of change (van de Ven and Poole, 2005) to include relationships of co-creation. Our operationalisation of IETF includes measuring and tracking through socially constructed data and using network analysis metrics and visualisations to implement the sense-making and feedback mechanism. These correspond to network orchestration actions that are seen to shape network structures and outcomes, which in turn create shifts in orchestration actions (Paquin and Howard-Grenville, 2013). We see value in this activity, even though;

1 we are aware of the confines of the IETF due to the inherent characteristics of all frameworks as simplified models

2 we acknowledge that data-driven visualisations rely on data and that our data is not complete and hence the resulting network visualisations cannot show all of the connections nor nodes and might be seen incomplete for the purposes of network orchestration.

In this study, we employ a case study of EIT ICT Labs to demonstrate the use of IETF for addressing how data-driven network visualisations can be used for orchestrating innovation ecosystems. The case study method has been found to be a legitimate way of adding to the body of knowledge by providing detailed and analysed information about real world environments which can be seen as examples of phenomena under research (Benbasat, 1987). Owing to the call for practical solutions for the EIT ICT Labs community, through employing the IETF, we follow action research practices towards combining the expertise of evidence-based research with local, contextual knowledge (Brydon-Miller et al., 2003).

Accordingly, we actively communicated and collaborated with EIT ICT Labs representatives of two senior managers and one business developer. These representatives are knowledgeable and have a holistic understanding of the context, thus using them as informants was seen applicable, especially as they are interested and open to collaboration. These informal, unstructured and iterative discussions took place through face-to-face meetings and online meetings, were short (maximum of 1 hour) and were documented in notes. The discussions were conducted in the summer of 2011 for validity checks after initial data analysis, in the summer of 2012 for refinement of visualisations and in the fall of 2013 for collaborative sense-making after the release of the final versions of the visualisations.

Within EIT ICT Labs it is recognised that relationships between key individuals open channels through which talent, information and financial resources can flow across Europe. This flow of resources – as relational capital – through relationships is a key premise of IETF; in this study we refer to this flow as mobility. Mobility is widely recognised as an important aspect of knowledge creation and sharing within innovation networks (Saxenian, 2007) and its use as the indicator for the exchange and innovation potential in the economy is recognised (Graversen, 2003). We focus on changes in relationships during a three-year period as an indicator of mobility with which we measure and track the process of transformation. Therefore, we use SNA metrics of degree and betweenness as a mobility factor to illuminate the potential of individual nodes to serve as bridges between the EIT ICT Labs co-locations. The resulting metrics and visualisations (geospatial network representations) provide evidence on ecosystemic actors and linkages that create the context of mobility for the EIT ICT Lab as well as reveal the operational impact of its activities act as events whose impact can create new coalitions that will serve as conduits for the mobility of information, talent and financial resources. Insights connecting events to coalitions can be used for creating shared understanding and programme planning for the management of networks in this emerging innovation ecosystem.

3.1 Case EIT ICT Labs

EIT ICT Labs (http://eit.ictlabs.eu) operates in a complex ecosystem of independent and interdependent actors, financing schemes and business models that create value for the European innovation landscape and whose innovation strategy is positioned toward its mission of enhancing this ecosystem to synergise and accelerate innovations contributing to economic growth (http://www.eitictlabs.eu/ict-labs/about-eit-ict-labs/our-approach/). For the purposes of our research, we view EIT ICT Labs as an innovation ecosystem and apply IETF to its transformation:

- with the shared vision of turning Europe into the global leader in ICT innovation
- with coalitions of sub-networks of actors ‘nodes’ – and the various actors around them
- participating in events (collaborative activities and interactions shown by ‘links’) that result in impact (changes) in the flow of company information, of talent and of financial and other innovation resources
- looking at relationship links as indicators of the potential to increase the mobility of information, talent and resources.

For validation of the ground truth of our data-source and for feedback on opportunities for transformation, we collaborated with representatives of EIT ICT Labs. We presented our early results and initial sense-making to them, allowing them as context-experts to engage in their own sense-making of the findings and to derive the insights needed to support decisions regarding network orchestration.

3.2 Data collection and sample characteristics

The analysis of the ecosystem of EIT ICT Labs uses an annual sampling of Innovation Ecosystems Network (IEN) Dataset (Rubens et al., 2010) which is a quarterly updated collection of socially-constructed and curated data. It is data scraped from sources such as Crunchbase, TechCrunch, Arctic StartUp, Wikipedia etc. that has been cleaned and organised so that it can be used for further analysis. The dataset for this study describes executive and funding relationships which then allows for the network visualisations: it includes data on companies (including enterprises and start-up companies), their key individuals (with data about the educational institutions they have been associated with) and their financing firms (investment organisations and venture capital investors). Individuals in the dataset are key individuals in their respective companies (e.g., founders, executives, lead engineers, members of boards of advisors and investors). As shown in Table 1, the full dataset from which the EIT ICT Labs sample is drawn includes more than 100,000 companies.

Table 1 Full dataset over selected time periods

Network actor	*2011*	*2012*	*2013*	*2011–2012*	*2012–2013*
Individual	76,000	100,000	150,000	32%	50%
Company	65,000	80,000	100,000	23%	25%
Financial	5300	7,000	10,000	32%	43%

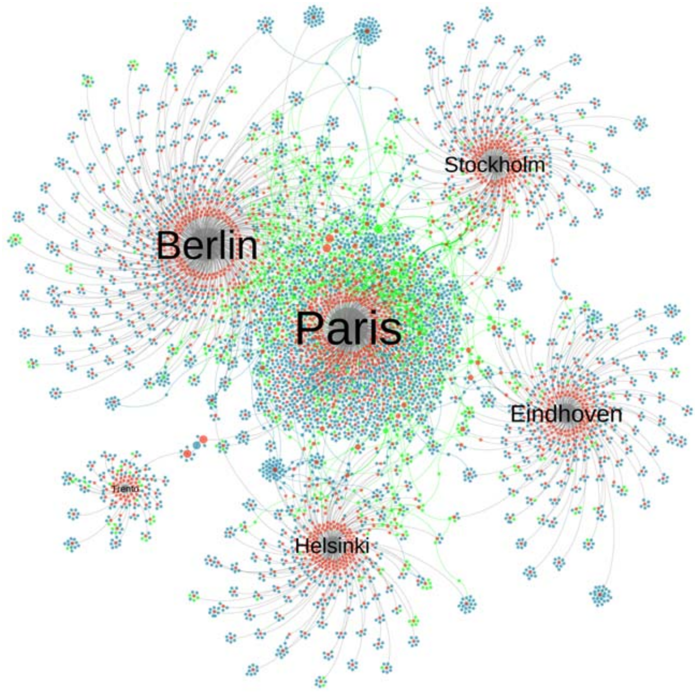
The dataset for the EIT ICT Labs sample is drawn by selecting all the companies that have their primary office in one of the six co-location cities of EIT ICT Labs: Berlin, Eindhoven, Helsinki, Paris, Stockholm and Trento. Each company is connected to the city of its primary office with a link. Then, we select all the key individuals (founders, board members and C-level executives) in the dataset who are identified either with a previous or a current connection to one or more of the companies in the sample and showed their relationship to those companies with a link. Next, financial organisations identified with funding events for those companies are added as nodes and the relationships between financial organisations and companies are shown with additional links. These procedures were used to create a sample of IEN data for EIT ICT Labs, first in 2011, the year the EIT ICT Labs programme was initiated. Same procedure was used in 2012 and 2013 to recreate those samples of EIT ICT Labs data from the updated IEN dataset.

Intrigued with how relationships with international companies, key individuals and especially financial organisations might reveal new insights about the EIT ICT Labs innovation ecosystem, we expand the year 2013 sample to include the ICT companies in the San Francisco Bay Area of California, then also include the individuals and financial organisations that have relationships with those companies. Using the same data selection procedures as for the initial sample, links to actors in the EIT ICT Labs sample are established based on relationships such as board or advisory roles (individuals) or investments (financial). This expanded sample introduces more than 6,800 companies, almost 20,000 key individuals and some 1,800 financial firms from SF Bay Area to the overall ecosystem. The expanded sample adds the relationships of the San Francisco Bay Area to the EIT ICT Labs network, revealing its potential to enhance the mobility of information, talent and financial resources.

3.3 Network metrics and visualisations

To present evidence of the ecosystemic transformations taking place in the European innovation ecosystem and measuring and tracking the EIT ICT Labs ecosystem, we utilise network metrics of degree and betweenness that are also basis for the data-driven network visualisations. Especially betweenness supports the understanding of existing connections between innovation nodes and also provides potential insights for targeted actions based on the innovation actors with highest betweenness values.

Figure 2 Key individuals, companies and financial firms as resources for mobility in 2013: highlighting the roles of Paris and Berlin in the ecosystem and also the role of financial firms (green) as enablers for mobility (see online version for colours)



Overall representation of the EIT ICT Labs network across co-location cities is created (Figure 2). In this, we include and examine key individuals, companies and investors. The lines between the nodes show links-relationship connections (total number of nodes is 6,187, total number of links is 7,050). A company is linked to a co-location city if its primary office is located in the city. In all of the visualisations, key individuals are in blue, companies in red and financial firms in green. The names of individuals, companies and investors are not shown. Link colour follows the colour of the source node. Gray links point from EIT ICT Labs co-locations to companies, green links from investors to companies and blue links from individuals to companies. To more clearly reveal key patterns in the structure of the EIT ICT Labs innovation ecosystem, actors in each co-location having the top 10% betweenness values are selected across all node categories, i.e., individuals, companies and investors. The top 10% network includes 29 individuals, 51 investors and 513 companies that form the key pathways in between the six EIT ICT Labs co-location cities. Node size is again proportional to its betweenness value.

Companies are initially clustered close to their respective EIT ICT Labs co-location cities. However, the force-directed layout algorithm pulls the company toward other locations based on links to locations established through individuals or investment firms who have relationships with multiple companies in different locations. Similarly, this algorithm pulls the location of financial firms to a place in the network that reflects all the relationships with companies held by that firm. In this way, nodes representing companies, financial firms and key individuals are positioned relative to each other, in the context of the companies' connection to the EIT ICT Lab co-location cities. In the resulting network visualisation, many actor nodes are clustered around each of the co-location cities. Paris and Berlin show the largest clusters because they have the greatest number of actors – key individuals, companies and financial firms – connected to them. They share connections, which pulls them close to each other and Paris with its multiple and powerful connections takes its central place in the EIT ICT Labs ecosystem.

4 Results

This research analyses mobility in the context of relationships between companies, individuals and financial organisations in the innovation ecosystem based on the co-location cities in which the EIT ICT Labs initiative is operating.

4.1 Network metrics

The size of the EIT ICT Labs ecosystem increased from 2011 to 2013 as presented in Table 2. Compared to the analysis conducted in 2011, the number of companies in the EIT ICT Labs sample dataset in 2012 and 2013 increased at approximately the same pace (23%) as the number of companies in the IEN Dataset as a whole (25%). The changes in number of key individuals as well as investors are different in the full dataset compared to the sub-set for EIT ICT Labs. This interesting trend deserves more analysis and could be a result of changes in the online availability of data rather than changes in activities of actors. However, for the purposes of this research, we concentrate on changes within EIT ICT Labs. For example, a total of 55 new investors emerge in its network between 2012 and 2013.

Table 2 Growth over time of EIT ICT Labs ecosystem sample dataset

Network actors	*2011*	*2012*	*2013*	*2011–2012 change*	*2012–2013 change*
Individual	1,634	2,817	3,660	72%	30%
Companies	1,056	1,665	2,041	58%	23%
Financial	280	425	480	52%	13%

Table 3 shows the changes in degree and betweenness metrics across the EIT ICT Labs co-location cities during the 2011 to 2013 period. Berlin and Paris have the greatest connectivity; in this innovation ecosystem these co-location cities have the largest number of actors connected to them. In 2013, the sample reveals that Paris had 589 companies in the ICT sector; Berlin had 507. This is roughly twice the number of Eindhoven, Stockholm and Helsinki and more than five times more than Trento. Paris and Berlin exhibit the greatest betweenness values in 2012 and continue to have the greatest betweenness values in 2013 (9,762,717 and 8,159,381s respectively).

Table 3 Change over time in relationship metrics for co-location cities

Co-location cities	*Betweenness*			*Degree*		
	2012	*2013*	*Change*	*2012*	*2013*	*Change*
Paris	6,950,362	9,762,717	40%	505	589	17%
Berlin	5,284,815	8,159,381	54%	389	507	30%
Eindhoven	2,802,845	4,445,841	59%	202	257	27%
Stockholm	2,695,012	3,978,408	48%	230	273	19%
Helsinki	2,741,119	3,914,762	43%	230	264	15%
Trento	820,993	1,246,415	52%	56	71	27%

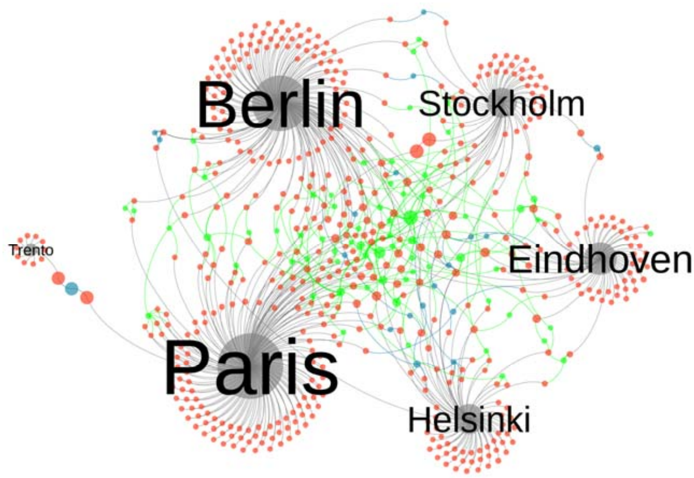
The largest change in betweenness values is observed in Eindhoven and Berlin, (59% and 54% respectively). These co-location cities show the greatest positive changes in betweenness and the largest increases in degree value since 2011. Trento follows closely with an increase of 52% in its betweenness and an increase of 27% in its degree of connectivity.

4.2 *Visualisations*

In the network visualisation shown in Figure 2, Paris and Berlin occupy key roles in the ecosystem; financial organisations (green) occupy central positions between co-location cities and thus are revealed as key enablers for mobility.

The size and complexity of the network visualising the 2,000+ companies, their key individuals and financial organisations was confirmed by our EIT ICT Lab collaborators but proved too complex for visual exploration of meaning in the network of relationships.

Figure 3 Top 10% of individual, companies and investors connecting EIT ICT labs co-location cities according to their betweenness in 2013: highlighting the role of financial firms (green) as enablers for mobility (see online version for colours)



The simplified network of co-locations and their most connected actors in Figure 3 (across all node categories, i.e., individuals, companies and investors) shows the top 10% of nodes according to their betweenness value.

Figure 4 San Francisco Bay Area as a 7th EIT ICT Labs co-location city according to year 2013 sample: highlighting the possibilities of extended network for mobility (see online version for colours)

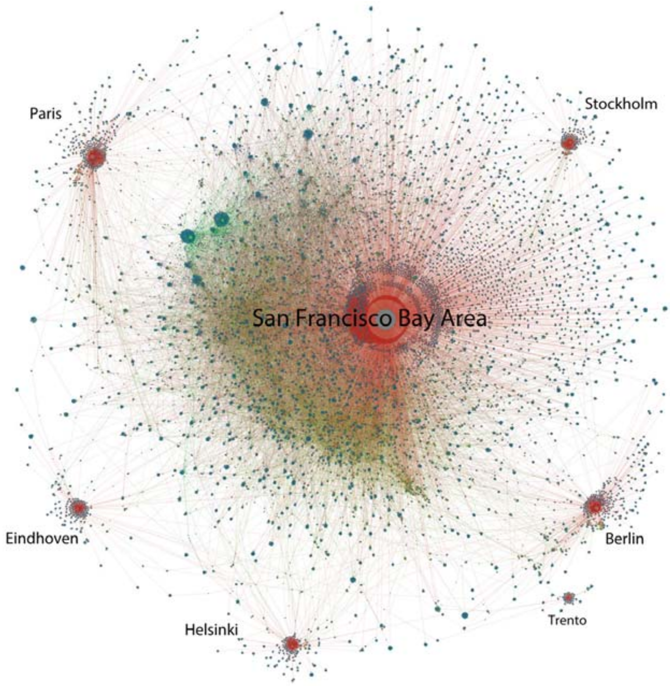


Figure 4 shows the network visualisation of the full sample that includes the presence of companies, key individuals and financial firms located in the San Francisco Bay Area as well as within of EIT ICT Labs ecosystem. This expanded sample shows a vastly larger ecosystem with a significantly larger number of nodes and links (number of nodes is 35,389 and total number of links is 51,106). The potential expansion of relational capital through which information, talent and financial resources could flow to and from companies in the EIT ICT Labs co-location cities is illustrated.

4.3 Sense-making with interaction and feedback

Sense-making discussions note ecosystem events – the existence of companies, individuals and financing firms, as well as their relationships to each other. The discussions also address impact – the growth in size of the ecosystem and its different actors over time and the differences in mobility factors among the co-location cities. During this study, revealing the structure of the EIT ICT Labs innovation ecosystem enabled sense-making conversations, in which both researchers and EIT ICT Labs representatives participated.

The review of Figure 3 resulted in the insight that only a small proportion of the individuals and financial firms in the EIT ICT Labs overall network exhibit relationship capital that spans more than one EIT ICT Labs co-location city. The review of Figure 3 also prompted the observation that financial organisations occupy key connector roles between the companies in various co-location cities, leading to sense-making discussions about additional opportunities to include financial organisations in the EIT ICT Labs' programmes and activities. Initially, EIT ICT Labs had not emphasised the role of financial organisations in its ecosystem: for example, none of its core or affiliate partners were financial organisations. However, with the data-driven approach and the data that inherently included connections through financing organisations, the role of financing organisations became evident, introducing more collaboration between EIT ICT Labs and financial organisations.

Visual examination of the network including SF Bay Area as the hypothetical 7th EIT ICT Labs co-location city provided an example of a 'what-if' question for policy makers and decision makers. This analysis (Figure 4) validated the strong role of SF Bay Area in the European venture-backed innovation ecosystem, also indicated by the high degree values for US corporate investment entities and US-based venture capital investors. The force-directed layout algorithm positions nodes with greatest connectivity in the centre of the network and thus we see the SF Bay Area at the centre of the network and EIT ICT Labs co-location cities on the periphery. The notable green belt accentuates the presence of financial actors as mobility enablers in connecting the companies and key individuals of the EIT ICT Nodes to the SF Bay Area. Sense-making discussions about Figure 4 led to conversations about ways in which EIT ICT Labs programmes could establish a presence in the bay area to enable and accelerate mobility factors – events whose impact would lead to new coalitions and the development of relationships through which information, talent and financial resources could flow.

As summarised in Table 4, sense-making and interpreting the visualisations support the understanding of the ecosystem, provide the contextual view of the larger innovation ecosystem in which EIT ICT Labs operates and allow for interaction and feedback using the IETF framework. Our investigation of actors and relationships in this specific innovation landscape focuses on revealing relationships of individuals in the network (the mobility of knowledge and talent) as well as financing firms (the mobility of financial resources), as measures of the transformation potential of an innovation ecosystem. These measures of mobility are considered key in deriving insights that can contribute toward decisions to support the network orchestration of relationships among companies, individuals and financial firms. Hence, the findings are interpreted with that specific context in mind. The growth of the ecosystem indicates mobility of individuals, companies and financial firms entering into the ecosystem; the key metrics indicate increasing potential of relationships as resources for mobility; and the role of financial firms as well as key individuals is seen as a key mobility enabler. The potential for mobility within Europe, as well as between Europe and the bay area reveals opportunities for greater access to resources for growth.

Table 4 Sense-making of research findings (evidence of ecosystemic linkages and actors)

*Findings**	*Sensemaking*
Growth over time of EIT ICT Labs Ecosystem sample (Table 2)	Describes growth of the overall EIT ICT Labs ecosystem: the increased numbers of individuals, companies and financial firms; mobility of entities entering into the ecosystem.
Change over time in key metrics for co-location cities (Table 3)	Reveals changes in the EIT ICT Labs ecosystem over time.
	The increased value of degree reflects new connections to the cities.
	The increased value of betweenness reflects new connections between the companies, financial firms and individuals with locational relationships to these cities.
	Paris and Berlin are the largest and best-connected clusters.
	New connections are resources for mobility of individuals (knowledge) and financing.
Key individuals, companies and financial firms as resources for mobility (Figure 2)	Shows the central role of financial actors as resource-brokers and key mobility enablers.
	The relationship connections, shown by green lines, linking the financial actors to companies and individuals, highlight this. These are created by their relationships with entities in multiple cities.
	Fewer blue nodes and lines, indicating limited roles of individuals in connecting the ecosystem.
Top 10% of individual, companies and investors connecting EIT ICT Labs co-location cities according to their betweenness (Figure 3)	Indicates the importance of financial actors, companies and a limited number of key individuals in connecting the ecosystem.
	The potential mobility of information, talent and financial resources through financial actors', companies' and individuals' relationships, shown by links.
	The numbers of key individuals with multi-node connections remains low, though has increased from 2011 to 2013.
San Francisco Bay Area as a 7th EIT ICT Labs co-location city (Figure 4)	Adds the 7th node expanded the network.
	Showed the sheer size as well as the central role of San Francisco Bay Area.
	Potential mobility of financial resources and key individuals in an extended network.

Note: *A new city Trento was added in 2012 and the geographical areas of each city were re-defined in 2012, hence the exact comparisons are not possible.

5 Discussion

This study proposes that data-driven network visualisations play a role in generating insights for orchestrating innovation ecosystems. At the core are:

1 the component of visualisations and the data-driven processes toward those

2 the component of understanding them in the context of innovation ecosystem transformation, which are conducted using the IETF construct.

The two components are tightly linked. As uncertainty around the network activity and its value is inherently high (Ritvala and Salmi, 2010), one benefit of using IETF for EIT ICT Labs is measuring and tracking its networked nature (structure, key actors, patterns of interest and flows of interaction) as empirical evidence for interactions and feedback, which allows for shared vision and understanding about the entity being orchestrated.

5.1 Data-driven network visualisations

Our six-location modelling of the EIT ICT Labs co-location cities resulted in network metrics and geospatial social network visualisations, describing the growth and evolution of the EIT ICT Labs ecosystem. Highlighting the relationships of companies, key individuals and financial organisations – and their potential contributions to the mobility objective – reveals the existing network of relationship capital into which the activities of the EIT ICT Labs can be integrated. Consistent with the understanding of emergent structures that are only visible after they have emerged (Padgett and Powell, 2013), it is assumed that additional nodes and links are in emergent states within the existing innovation ecosystem that described the EIT ICT Labs.

Our ecosystemic, evidence-based method uses relationships to reflect the mobility factor of ecosystem development; these and the overall values of this study are validated through interactions and feedback with leaders in the EIT ICT Labs, through informal, in-depth discussions. In these discussions, the observed changes in degree and betweenness over time were attributed to the programmatic activities of EIT ICT Labs, such as research activities along mobility objectives, action lines and meetings as catalysts. Programme leaders interpreted the increased relationships and networks for mobility to the process of creating the EIT ICT Lab programme and used their interpretations also for targeting some identified key individuals for potential collaboration as well as for planning for international outreach enhancing mobility, especially to Silicon Valley/San Francisco Bay Area.

EIT ICT Labs representatives affirmed the value of geospatial representations, though the insights from the network metrics and their interpretation were considered challenging. The use of analysis and network level snapshots over a three-year period was seen to provide a valuable baseline for updating measurements with the emphasis on the issue of mobility. It was seen to contribute a shared understanding of the complex issue of evolution and changes of an ecosystem. Future use of such indicators for understanding, monitoring, managing and evaluating the impact of the activities in the context of EIT ICT Labs was encouraged in these discussions. Especially as the co-location cities reflect a newly established network, one of the contributions of this research can be seen to be toward establishing a visible description for EIT ICT Labs' mobility objective, creating shared vision and communicating to broad audiences about its organisational objectives and accomplishments.

5.2 Operationalising IETF for network orchestration

The network visualisations allowed for observations of the network as the whole and for the identification of individual actors such as those in financing and it allowed for targeted orchestration. Hence, our research supports the understanding of orchestration as a set of evolving actions (Paquin and Howard-Grenville, 2013). For the interact and feedback elements of IETF, data-driven network visualisations can reflect the structure of an innovation ecosystem at a single point of time; they can also show the evolution of an ecosystem's actors and their relationships over time, corresponding to the opportunity for examining relationships and interactions (Halinen et al., 2012) and showing how network participants build valuable positions through their activities (Powell et al., 1996). Using IETF can contribute to the establishing the legitimacy of the EIT ICT Labs and communicating widely about it (Möller and Rajala, 2007). For example, two EIT ICT Labs representatives have used the network visualisations created in this study in management meetings as well as in conference presentations (Turpeinen, 2011; Jonker, 2013).

In dynamic ecosystems, networks compete against networks; and leading actors, as network orchestrators, must help entities in the ecosystem understand their roles in the network and collaborate for integrated synergy with common vision that creates value for the entire network (Ritala et al., 2013). Skilful network orchestration enables shared vision about the whole ecosystem, as well as for individuals' egocentric networks, corresponding to the value flows from targeted and directed connections arranged by an orchestrator (Paquin and Howard-Glenville, 2013). Our operationalisation of IETF in this study highlights the importance of shared vision that is collectively realised and continually updated (Hagel and Seely Brown, 2005), extending the process studies of change in organisations that "conceptualize change as a succession of events, stages, cycles or states in the development or growth of an organization" (van de Ven and Poole, (2005), p.1389] to innovation ecosystems.

We recommend the development of improved methods for managing the volume, velocity and variety of data (McAfee and Brynjolfsson, 2012). To transform an innovation ecosystem, we suggest that the network orchestrator should simultaneously:

1 facilitate the network

2 enable the individual actors and their activities.

We invite researchers, programme managers and policy makers to explore and embrace the possibilities of combining a variety of data sources, for data-driven network visualisations as well as for other evidence about ecosystems and their actors and interactions. In addition, we see that to allow for faster and deeper insights, metrics and their representations must move beyond static snapshots. Interactive visualisations introduce means to apply the methods and tools of visual analytics (Keim et al., 2010) for decision-making support.

5.3 Limitations

During this research, we collaborated actively with the representatives of EIT ICT Labs in order to seek fundamental insights on:

1 the most representative data sources available to support the analysis

2 selecting the most appropriate metrics and network visualisation parameters

3 ways to use the visualisations in a meaningful way in EIT ICT Labs management processes.

These discussions led to deeper understanding about the limitations related to the evidence-based approach presented here, as well as potential improvements for analysis processes (for example, for data sampling) and suggested questions for guiding the process.

Naturally, the limitations of this research as well as the applicability of the results and subsequent recommendations are amplified with the use of the case study approach and its inherent challenge of generalisation. In addition, the use of action research components further limits generalisation with the context and time specificity of interpretations conducted throughout the research. For example, the participants' voluntary participation and openness to collaboration with research partners can be seen to have impacted the results, but the extent and the nature of the impact remain imprecise.

Importantly, the value of the data-driven insights on how an ecosystem emerges and evolves depends strongly on both the quantity and quality of data. The novelty of our results comes from the use of socially constructed data that is (almost) real-time, providing information that we find to be difficult to obtain through other sources. With data openly available on the internet and curated through social media practices, important contextual insights can be provided to augment programme-specific and internal data collecting and reporting practices. It is possible that some relevant data was not included. Some of the potential biases in the data were counter-balanced by its large quantity. Further, the applied data curation processes optimised the quality and accessibility of this data. For example, the increase in the number of companies related to EIT ICT Labs from 2011 to 2012 may have partially been due to the changes in the sampling procedure used for those two years as the initial sample dataset was expanded to include Trento. It is also worth noting that English language bias of the dataset might have been partially responsible for the extremely strong representation of SF Bay Area actors in the expanded sample.

Additionally, it is not clear at this time whether growth in the number of companies from 2012 to 2013 reflects growth in the availability of data that includes the companies or growth in actual activities of companies represented by this sample of data. Also why the growth of the subset used for this research differs from the growth of the overall dataset is not clear. The details of such biases, which may be inherent in socially constructed data or might correlate with larger societal and/or business trends, are not well documented to date and they present an opportunity for further study. Still, the patterns that emerge from large quantities of data can be seen produce insights about the character of phenomena represented by the data.

Though we agree with Kohlhammer et al. (2012) that visualisation and visual analytics are vital for informed decision-making and policy modelling in a highly complex information environment overloaded with data and information, we do not advocate using network visualisations as the only evidence for decision-making or policy setting. The literacy of decision makers in visual analytics and network metrics is just beginning to emerge. Most managers are not accustomed to reading network visualisations and the metrics behind them are not yet common knowledge. Our experience emphasised the important responsibility of action researchers to educate about the methodologies at the same time as presenting the results and to communicate the data and the analytical processes in a way that makes sense for the decision makers.

6 Conclusions

In this study, we utilised the IEFT to investigate and explain the complexities of innovation and to create a shared understanding and insights toward possibilities for network orchestration within the case environment of EIT ICT Labs. We demonstrate that data-driven network visualisations offer a powerful approach for providing evidence-based information when talking about ecosystems, their structures, key actors and interactions, revealing their context and the potential for novel structures and relationships, especially in this case of operationalising mobility as relational capital. For EIT ICT Labs as an example, the network visualisations and the network metrics show key roles of financial organisations as well as the impact and size of the Bay Area/Silicon Valley ecosystem.

Accordingly, resulting network metrics and visualisations provide a significant step forward in addressing the call for better understanding of network orchestration (Paquin and Howard-Grenville, 2013), simultaneously highlighting the possibilities of data-driven decisions (McAfee and Brynjolfsson, 2012) for addressing the complexities of ecosystems and the use of multiple methods for understanding the processes over time (Bizzi and Langley, 2012). For EIT ICT Labs, the visualisations provided a description of the complexities of its ecosystem; the metrics describe some changes in the dynamics of its ecosystem.

Data-driven visualisations can support the development of insights needed to orchestrate transformations of ecosystems, recognising that activities orchestrated through individual actors of a network impact the whole network with the potential to leverage the relationship complexities of innovation. We claim that our approach not only describes and visualises the innovation network, but also provides insights for enhancing methods for the development, controllability and manageability of innovation networks, as it shows the individual and influential actors through which transformation can take place.

Acknowledgements

The authors express their gratitude to Rahul C. Basole for suggestions on the visualisations and to Marko Turpeinen of EIT ICT Labs for feedback and continued support. The authors would also like to thank the reviewers for their constructive comments on the manuscript. This research has been supported in part by the Finnish Funding Agency for Innovation, Tekes.

References

Basole, R.C. and Rouse, W.B. (2008) 'Complexity of service value networks: conceptualization and empirical investigation', *IBM Systems Journal*, Vol. 47, No. 1, pp.53–70.

Basole, R.C., Russell, M.G., Huhtamäki, J. and Rubens, N. (2012) 'Understanding mobile ecosystem dynamics: a data-driven approach', *Proceedings of 2012 International Conference on Mobile Business (ICMB 2012)*, Delft, Netherlands, pp.17–28 [online] http://aisel.aisnet.org/icmb2012/15/ (accessed 25 September 2013).

Benbasat, I., Goldstein, D.K. and Mead, M. (1987) 'The case research strategy in studies of information systems', *MIS Quarterly*, Vol. 11, No. 3, pp.369–386.

Bizzi, L. and Langley, A. (2012) 'Studying processes in and around networks', *Industrial Marketing Management*, Vol. 41, No. 2, pp.224–234.

Brydon-Miller, M., Greenwood, D. and Maguire, P. (2003) 'Why action research?', *Action Research*, Vol. 1, No. 1, pp.9–28 [online] http://www.arj.sagepub.com/content/1/1/9 (accessed 5 February 2014).

Card, S.K., Mackinlay, J.D. and Shneiderman, B. (1999) *Readings in Information Visualization: Using Vision to Think*, Morgan Kaufmann Publishers, San Francisco, CA.

Christensen, C. (1997) *The Innovator's Dilemma*, Harvard Business Press, Boston, MA.

Cunningham, P. and Ramlogan, R. (2012) 'The effects of innovation network policies', *Compendium of Evidence on the Effectiveness of Innovation Policy Intervention*, Manchester Institute of Innovation Research, Manchester Business School, University of Manchester [online] http://www.innovation-policy.net/compendium/section/Default.aspx?topicid=23 (accessed 5 February 2014).

Dhanaraj, C. and Parkhe, A. (2006) 'Orchestrating innovation networks', *Academy of Management Review*, Vol. 31, No. 3, pp.659–669.

Durst, S. and Poutanen, P. (2013), 'Success factors of innovation ecosystems – initial insights from a literature review', *Proceedings of Co-Create 2013: The Boundary-Crossing Conference on Co-Design in Innovation*, 16–19 June, 2013, Aalto University, Espoo, Finland.

Easley, D. and Kleinberg, J. (2010) *Networks, Crowds and Markets: Reasoning About a Highly Connected World*, Cambridge University Press, New York, NY, USA.

Freeman, L.C. (2009) 'Methods of Social network visualization', in Meyers, R.A. (Ed.): *Encyclopedia of Complexity and System Science*, Springer, Berlin.

Graversen, K.E. (2003) *Human Capital Mobility – A Comparable Knowledge Indicator*, Science and Technology Indicators in the Nordic Countries, OECD Report.

Green, D.G. and Sadedin, S. (2005) 'Interactions matter – complexity in landscapes and ecosystems', *Ecological Complexity*, Vol. 2, No.2, pp.117–130.

Hadhazy, A. (2011) 'Visualization tools can 'see' through a forest of data', *TechNewsDaily* [online] http://www.technewsdaily.com/2053-visualization-tools-can-see-through-a-forest-of-data.html (accessed 10 February).

Hagel, J. and Seely Brown, J. (2005) *The Only Sustainable Edge*, Harvard Business School Press, Boston.

Halinen, A., Medlin, C.J. and Törnroos, J-Å. (2012) 'Time and process in business network research', *Industrial Marketing Management*, Vol. 41, No. 2, pp.215–223.

Hwang, V.W. and Horowitt, G. (2012) *The Rainforest: The Secret to Building the Next Silicon Valley*, Regenwald, Los Altos Hills, CA, USA.

Jonker, W. (2013) 'EIT ICT labs: create for value', *Presented at the iMinds The Conference 2013*, Brussels, December 5 [online] http://www.slideshare.net/iMinds/eit-ict-labs-external-presentation-dec-2013 (accessed 6 March 2014).

Keim, D., Kohlhammer, J. and Ellis, G. (Eds.) (2010) *Mastering the Information Age – Solving Problems with Visual Analytics*, Eurographics Association [online] http://www.vismaster.eu/book/ (accessed 31 March 2014).

Kohlhammer, J., Nazemi, K., Ruppert, T. and Burkhardt, D. (2012) 'Toward visualization in policy making', *IEEE Computer Graphics and Applications*, Vol. 32, No. 5, September–October, pp.84–89.

Krackhardt, D. and Hanson, J.R. (1993) 'Informal networks: the company behind the charts', *Harvard Business Review*, Vol. 71, No. 4, pp.104–111.

Liu, Y-Y., Slotine, J-J. and Barabasi, A-L. (2011) 'Controllability of complex networks', *Nature*, Vol. 473, No. 7346, pp.167–173.

McAfee, A. and Brynjolfsson, E. (2012) 'Big data: the management revolution', *Harvard Business Review*, October, Vol. 90, No. 10, pp.61–68.

McKinsey (2011) *Big Data: The NextFfrontier for Innovation, Competition and Productivity* [online] McKinsey Global Institute, Overall Presentation, October [online] http://www.jegi.com/sites/default/files/McKinsey_Presentation.pdf.

Möller, K. and Rajala, A. (2007) 'Rise of strategic nets: new modes of value creation', *Industrial Marketing Management*, Vol. 36, No. 7, pp.895–908.

Noack, A. (2009) 'Modularity clustering is force-directed layout', *Physical Review E*, Vol. 79, No. 2, p.8, doi:10.1103/PhysRevE.79.026102.

Nykänen, O., Salonen, J., Haapaniemi, M. and Huhtamäki, J. (2008) 'A visualisation system for a peer-to-peer information space', *Proceedings of OPAALS 2008: The 2nd International OPAALS Conference on Digital Ecosystems*, 7–8 October 2008, Tampere, Finlandpp.76–85 [online] http://matriisi.ee.tut.fi/hypermedia/events/opaals2008/articlelist.html#opaals2008-article14 (accessed 1 October 2013).

Padgett, J. and Powell, W.W. (2013) 'The problem of emergence', in Padgett, J. and Powell, W.W. (Eds.): *The Emergence of Organizations and Markets*, Princeton University Press, Princeton, NJ, USA.

Paquin, R. and Howard-Grenville, J. (2013) 'Blind dates and arranged marriages: longitudinal processes of network orchestration', *Organization Studies*, Vol. 34, No. 11, pp.1623–1653.

Powell, W.W. and Owen-Smith, J. (2013) 'An open elite', in Padgett. J. and Powell, W.W. (Eds.): *The Emergence of Organizations and Markets*, Princeton University Press, Princeton, NJ, USA.

Powell, W.W., Koput, K.W. and Smith-Doerr, L. (1996) 'Interorganizational collaboration and the locus of innovation: networks of learning in biotechnology', *Administrative Science Quarterly*, Vol. 41, No. 1, pp.116–145.

Reason, P. and Bradbury, H. (Eds.) (2001) *Handbook of Action Research: Participative Inquiry and Practice*, Sage Publications, London.

Ritala, P., Agouridas, V., Assimakopoulos, D. and Gies, O. (2013) 'Value creation and capture mechanisms in innovation ecosystems: a somparative case study', *International Journal of Technology Management*, Vol. 63, Nos. 3/4, pp.244–267.

Ritala, P., Armila, L. and Blomqvist, K. (2009) 'Innovation orchestration capability – defining the organization and individual determinants', *International Journal of Innovation Management*, Vol. 13, No. 4, pp.569–591.

Ritvala, T. and Salmi, A. (2010) 'Value-based network mobilization: a case study of modern environmental networkers', *Industrial Marketing Management*, Vol. 39, No. 6, pp.898–907.

Rizova, P. (2006) 'Are you networked for successful innovation?', *MIT Sloan Management Review*, Vol. 47, No. 3, pp.49–55.

Rubens, N., Still, K., Huhtamäki, J. and Russell, M.G. (2010) *Leveraging Social Media for Analysis of Innovation Players and Their Moves*, mediaX White Paper, Stanford University.

Russell, M.G. and Smith, M.A. (2011) 'Network analysis of regional ecosystem of afterschool programs', *Afterschool Matters*, Spring, No. 13, pp.1–11.

Russell, M.G., Still, K., Huhtamäki, J., Yu, C. and Rubens, N. (2011) 'Transforming innovation ecosystems through shared vision and network orchestration', *Proceedings of Triple Helix IX Conference*, July 2011, Stanford University.

Saxenian, A. (2007) *The International Mobility of Entrepreneurs*, Berkeley University Papers [online] http://people.ischool.berkeley.edu/~anno/Papers/International_Mobility_of_Entrepreneurs.pdf (accessed 25 April 2014).

Southeastern Minnesota Initiative Fund (1995) *Regional Development Planning and Evaluation Report*, Connect Consultants International, Inc.

Still, K., Huhtamäki, J., Russell, M.G. and Rubens, N. (2012) 'Paradigm shift in innovation indicators – from analog to digital', *Proceedings of 5th ISPIM Innovation Forum*, 9–12 December, 2012, Seoul, South Korea.

Turpeinen, M. (2011) 'What kind of questions in innovation ecosystems insist visualization? Mobility analysis in the context of European innovation ecosystem', Presented at *Visualizing Innovation Ecosystems Workshop*, MindTrek 2011, Tampere, Finland [online] http://www.mindtrek.org/2011/visualizing-innovation-ecosystems.

van de Ven, A.H. and Poole, M.S. (2005) 'Alternative approaches for studying organizational change', *Organization Studies*, Vol. 26, No. 9, pp.1377–1404.

Wasserman, S. and Faust, K. (1994) *Social Network Analysis: Methods and Applications*, 1st ed., Cambridge University Press, New York, NY.

Wellman, B. (1988) 'Structural analysis: from method and metaphor to theory and substance', in Wellman, B. and Berkowitz, S.D. (Eds.): *Social Structures: A Network Approach*, pp.19–61, Cambridge University Press, New York, NY.

Welser, H.T., Gleave, E., Fisher, D. and Smith, M. (2007) 'Visualizing the signatures of social roles on online discussion group', *The Journal of Social Structure*, Vol. 8, No. 2 [online]. http://www.cmu.edu/joss/content/articles/volume8/Welser/ (accessed 25 April 2014).

Wind, J., Fung, V.K.K. and Fung, W. (2008) 'Network orchestration: core competency in a borderless world', in Wind, J., Fung, V.K.K. and Fung, W. (Eds.): *Competing in a Flat World: Building Enterprises for a Borderless World*, Wharton University Publishing, Upper Saddle River, NJ.

Publication VII

Jukka Huhtamäki, Martha G. Russell, Neil Rubens, Kaisa Still (2015) "Ostinato: The exploration-automation cycle of user-centric, process-automated data-driven visual network analytics," in Sorin Adam Matei, Elisa Bertino, Russell, Martha G. (Eds.) *Transparency in Social Media: Tools, Methods and Algorithms for Mediating Online Interactions.*

Available at http://dx.doi.org/10.1007/978-3-319-18552-1_11

Ostinato: The Exploration-Automation Cycle of User-Centric, Process-Automated Data-Driven Visual Network Analytics

Jukka Huhtamäki, Martha G. Russell, Neil Rubens, and Kaisa Still

1 Introduction

This chapter introduces the Ostinato Model, an exploration-automation cycle for a user-centric, process-automated, data-driven visual network analytics.

In terms of increasing the transparency of editorial processes on social media, this chapter contributes to the general theme of the book and particularly its second volume at hand in three levels. First, network analysis is a key approach in supporting explorative studies on the patterns and structures in between actors creating, curating, refining, and distributing social media content and in estimating the authority and trust these actors have, therefore allowing for increasing the transparency of the editorial structure of Wikipedia co-authors, discussion and dissemination structures on Twitter and other social media. These structures can be modeled, represented, analyzed and visualized as networks to support the investigations and exploration. Second, the presented data-driven approach allows extending these investigations beyond the boundaries of individual social media and over long periods of time. Third, actors with different sets of skills from means to crawl online sources for data to domain knowledge allowing deep sensemaking can all fully engage into the different phases of the investigative process.

J. Huhtamäki (✉)
Tampere University of Technology, Tampere, Finland
e-mail: jukka.huhtamaki@tut.fi

M.G. Russell
mediaX at Stanford University, Stanford, CA, USA

N. Rubens
University of Electro-Communications, Tokyo, Japan

K. Still
VTT Technical Research Centre of Finland, Espoo, Finland

S.A. Matei et al. (eds.), *Transparency in Social Media*,
Computational Social Sciences, DOI 10.1007/978-3-319-18552-1_11

These contributions allow the use of visual representations of the structures behind various social media phenomena to improve social interaction, estimations of trust and credibility on social media. With the data-driven approach, the investigators of social media phenomena and patterns of social interaction, trust and credibility are able to move fast in the beginning of the process. As the ways of visualizing and investigating a particular phenomena matures, the investigators may continue to follow the phenomena with the support of close to real-time dashboards adding transparency and supporting e.g. longitudinal investigations. The option for automating the process also supports developing these investigative tools toward end-user products for avid social media content authors and users.

In music, the word "ostinato" refers to both a repeating musical pattern as well as a composition that contains a repeating musical pattern. Like the repeated rhythms and melodies in Ravel's *Bolero* (Fig. 1)—small innovations are explored with each iteration, and some are incorporated into the melodic narrative—we apply the musical concept of "ostinato" to a cycle of user-centric exploration and automation that builds transparency of authorship for evidence-based decision making.

Here, data-driven means that the analysis process relies on data, is automated and conducted in a computational manner, and visual network analytics refers to taking a visual analytics (Heer & Shneiderman, 2012; Thomas & Cook, 2006) approach to network analysis. Additional data can augment the dataset selected for analysis through an automated software process. Established analytical procedures can be automated, yet new conditions for analysis-based insights can be introduced and refined incrementally with continuous computational iterations.

In this implementation of the Ostinato Model, the phenomena under investigation are modeled as a network, and highly interactive visualization tools are used to conduct the investigative process. Network analysis introduces a relationship approach to investigating the structure of many kinds of phenomena. Network analysis allows for exploratory analysis of the social roles of network actors and the phenomena of relationships, as well as for quantifying the structural properties of networks.

A key aspect of the Ostinato Model is the focal point of the user—here, the investigator of particular network-driven phenomena—in the investigative process.

Fig. 1 Ostinato patterns from Bolero's *Ravel* (Mawer, 2000)

This answers to the call for data scientists,[1] somewhat mythical multi-skilled individuals that are capable of individually running the whole investigative process from collecting data to analysis to deep sensemaking in domain of interest, by allowing both experts of the domain under investigation, developers of the technical process as well as e.g. quantitative analysis specialists to possess equal means to take a proactive role in the investigative process. Moreover, the Ostinato Model defines an overall structure for the data-driven investigative process that supports the coordination between the individual phases of the process and therefore allows all the members of the investigative team to contribute to the implementation of different phases of analysis.

Visual network analytics allows the emergence of insights on the structure and dynamics of innovation ecosystems, social media platforms and other networked phenomena. Existing research on networks shows that network analysis has a good fit for explorative analysis of (eco)systems: much is already known about structure in networks (Barabási & Bonabeau, 2003; Granovetter, 1973), the roles of individual actors in the network (Hansen, Shneiderman, & Smith, 2011), the drivers of network evolution (Giuliani & Bell, 2008) as well as the latent structures and dynamics behind the diffusion of information through networks (Leskovec, Backstrom, & Kleinberg, 2009), network control (Liu, Slotine, & Barabási, 2011) and virality (Shakarian, Eyre, & Paulo, 2013; Weng, Menczer, & Ahn, 2013). Transforming those insights into action requires communicating the insights to constituents of change (Russell et al., 2011; Still et al., 2014). Visual network analysis is a promising method for investigating social configurations and for interactively communicating their findings to others (cf. Freeman, 2009).

Data-driven visual network analytics leverages computation to analyze potentially very large datasets in order to identify the patterns driving complex phenomena. Moreno (1953), Freeman (2000, 2009), Hansen et al. (2009, 2011), Russell et al. (2011), Still et al. (2014), Basole et al. (2012), Ritala and Hallikas (2011), and Ritala and Huizingh (2014) give examples of using a network approach to investigate complex phenomena that are driven by sets of interconnected actors. The investigations of such phenomena are further complicated because data about these actors frequently come from multiple and diverse data sources, some of which are not developed for computational use. Especially in cases involving data that are heterogeneous by nature, an iterative, incremental analysis process is sometimes necessary (Telea, 2008). Analysis of complex phenomena often involves multiple pathways to actionable recommendations, and assumptions underlying decisions may change over time.

We agree with Freeman (2000) that integrated tools that can be used to collect, manage and visualize the SNA data are key in supporting network investigations. The tradeoff between usability and automation sometimes creates a barrier for new entrants into data-driven visual network analysis (Hansen et al., 2009). In order to

[1] Ideally, a data scientist is a hacker, scientist, quantitative analyst, trusted adviser and business (domain) expert, all in one person (cf. Davenport, 2014).

provide a low barrier approach to using network analysis to study complex phenomena, we prioritize usability over process automation when possible.

However, a gap exists between the vision and the practice. Manually operated processes used by individual investigators or small investigative teams rely on ready-made tools that are operated through graphical user interfaces. Using these stand-alone tools is very straightforward. The available data sources and analysis and visualization functionalities are, however, somewhat limited. The full-stack, programming-centric processes, in which massive sets of data are mined with tools that are developed and operated by experts, are generally run in complex cloud-based environments. We are aware that several process models, with different levels of abstraction, exist to structure data-driven, visualization-centric investigations; a selection of these models will be covered as part of the description of previous work in the next section.

Many of the existing models are either very general or focus on particular parts of the process. A data-driven visual network analytics approach requires drawing from a number of process models. Using parallel data sources is often not considered in the process models. Moreover, network analysis introduces specific requirements to the process, importantly including the possibility to calculate node metrics as additional data quantifying the different structural roles of the nodes.

Drawing from our experience in running multiple case studies in the context of explorative innovation ecosystem analysis, we take a design science research (Hevner et al., 2004) approach to describe a process model for data-driven visual network analytics. In this book, our chapter contributes to the body of knowledge on computational frameworks, tools and algorithms for supporting transparent authorship in social media knowledge markets by defining an interactive and iterative process model for data-driven visual network analytics to explore relationships in ecosystems. Our process model takes into account requirements stemming from a call for transparent authorship in social media knowledge markets and builds on existing models for data-driven analytics and sensemaking. It is designed to support iterative and incremental investigative processes, as well as to automatically update a visualization dashboard revealing the dynamics and evolving network structure of a phenomenon under investigation.

The rest of the chapter is organized as follows. In second section, we review previous work on which this Ostinato process model is based. The third section introduces the research methodology and a selection of cases we have used to develop the model. The fourth section describes the requirements for the process as well as the different steps that constitute the Ostinato process model (Fig. 2). In the fifth section, we discuss how this model satisfies these criteria and adapts to the exploration–automation cycle. The sixth section concludes the chapter and describes key implications and ideas for future work.

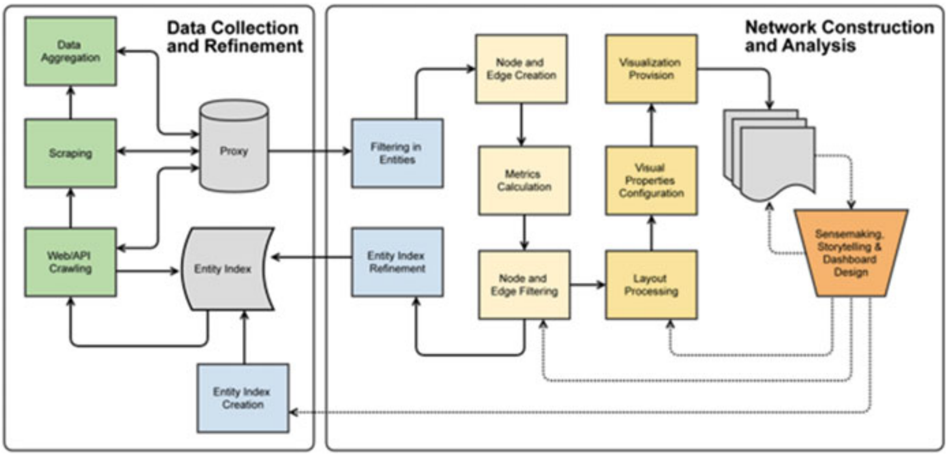


Fig. 2 Ostinato model—user-centric data-driven process model for visual network analytics

2 Previous Work

Our approach into data-driven visual network analytics builds on a number of bodies of knowledge (Fig. 3), including traditional SNA (Wasserman & Faust, 1994), information visualization (Card, Mackinlay, & Shneiderman, 1999), data-driven visualization pipelines (Nykänen et al., 2008), interactive network analysis (Hansen et al., 2009), visual analytics (Thomas & Cook, 2006), sensemaking (Pirolli & Card, 2005), interactive visualization (Heer & Shneiderman, 2012) and scientific visualization (Telea, 2008). All these fields offer models and approaches, and additionally they pose key requirements to be considered when developing next-generation analytics tools for very large networks. The objective to conduct (and publish) research in a reproducible way (Ghosh, 2013; Peng, 2009) contributes to the quality of the process and also introduces additional requirements.

Traditional SNA (Wasserman & Faust, 1994) introduces a set of node and network level metrics that can be used to describe the structural properties of networks and to quantify the various social roles of network actors. To support the use of network analysis for novices, Hansen et al. (2009) introduce the Network Analysis and Visualization (NAV) process model that builds on top of the general sensemaking model. The NAV process starts with defining the goals for the analysis and continues through data collection and structuring, after which data are interpreted through multiple loops of network visualization and SNA metrics calculation. Finally, the insights and conclusions are formatted and summarized, then disseminated through a report. Seeking low-barrier entry, the authors introduce NodeXL, an Excel-based toolset for SNA, to conduct the analysis. Among others, Hansen et al. (2011) define ways to apply these metrics in investigating phenomena taking place in social media.

The information visualization reference model (Card et al., 1999) presents a four-step process that can be used as a blueprint for implementing data-driven visualization processes. Raw data is (1) first collected and then (2) refined into data tables to

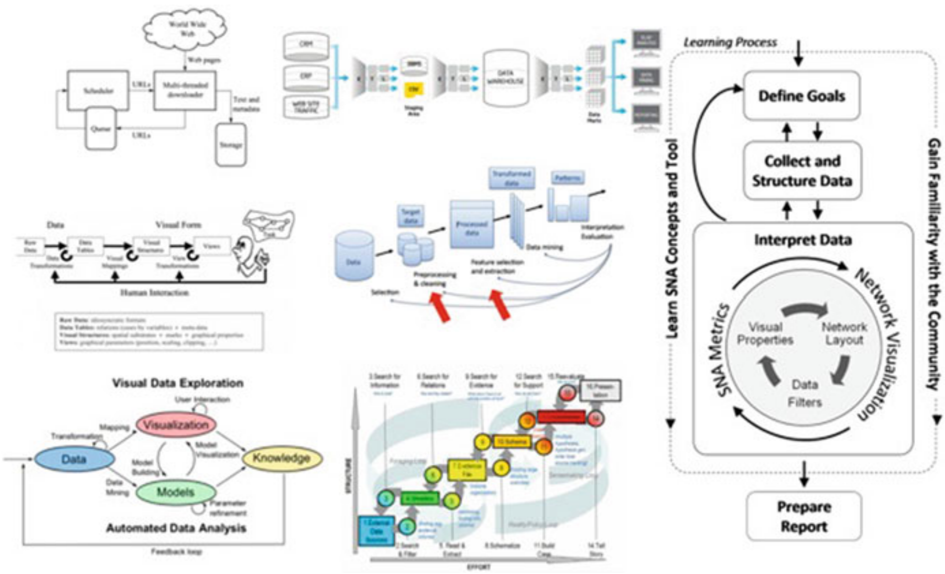


Fig. 3 Process models related to data-driven visual network analytics. The six small diagrams, from *top-left*: Web Crawling (Wikipedia, 2014), extract-transform-load (Intel, 2013), information visualization reference model (Card et al., 1999), knowledge extraction from databases (Indarto, 2013), visual analytics (Keim, Kohlhammer, & Ellis, 2010), sensemaking (Pirolli & Card, 2005). On the *right*: Network Analysis and Visualization (NAV) model (Hansen et al., 2009)

allow straightforward processing. Data tables are then (3) transformed into a portfolio of visual representations from which various concrete views are (4) served to the visualization user for sensemaking. Importantly, the reference model suggests that best practice is when the user can interact with all steps of the process (Fig. 3).

Component-based data-processing pipelines, a technical application of the information visualization reference model, introduce a viable approach for developing reusable pieces of software to support the automation of processes related to social network analysis across application domains (Huhtamäki et al., 2010; Nykänen et al., 2008). To support investigating the social structure among wiki co-creators, Huhtamäki et al. (2010) present a set of components and a process model to orchestrate the use of the components. A key benefit of the component-based approach presented by Nykänen et al. (2008) is the possibility to integrate existing software tools implemented in different technologies into the data-processing pipeline, given that they can be operated from the command line. The main restriction of the approach is the need to implement the automation through scripting, i.e. writing program code that describes rules for a particular functionality.

The general sensemaking model by Pirolli and Card (2005) divides the sensemaking process into two loops, the foraging loop and the sensemaking loop. To simplify, data is first collected and refined and then transformed into various visualizations and other representations that support the sensemaking. The process is iterated as many times as required. Similarly, the process of visual analytics "typically progresses in an iterative process of view creation, exploration, and refinement" (Heer & Shneiderman, 2012).

The sensemaking step can be applied in different ways—from purely manual processes in which humans interact with various user interfaces to conduct the analysis to automated information systems in which data are collected and processed in runtime. Sensemaking also includes the process of visual analytics (Thomas & Cook, 2006) that, by default, relies on the availability of software and tools supporting the users. Heer and Shneiderman (2012) give an insightful overview to the specific functionalities that users should be able to operate: (1) specify data and views; (2) manipulate views; and (3) process and provenance their findings.

Peng (2009) builds his definition of reproducible research on three categories: a piece of research is fully reproducible when both the data and code used to are available and, moreover, the code is executable by anyone. As Ghosh (2013) shows, reproducibility can be approached at many different levels, from policy to detailed technological solutions.

3 Methodology

In this section, we briefly describe the context in which the data-driven network analytics takes place. This illustrates the explanatory power and novelty introduced by the Ostinato Model to network analysis workflows. Further, we discuss the use of Design Science Research (Vaishnavi & Kuechler, 2007) as a method that we apply in our venture to develop the Ostinato Model in a way that is both credible in terms of scientific theory as well as practical utility. We also refer to a selection of case studies we have used to develop and validate the process model presented in this chapter. This shows that the Ostinato Model is a general approach to conducting data-driven network analysis investigations that has already been applied extensively in a series of real life experiments and investigations.

Addressing innovation ecosystems as networks allows scholars and practitioners to study their complexity, providing a means for mapping, monitoring and managing the ecosystem components. To do this, we have taken a data-driven network analysis approach to study innovation ecosystems in regional, metropolitan, national and international level as well as e.g. in the context of programmatic activities supporting innovation and growth. We have followed a design science research approach that is based on iteration through construction of network visualizations as artifacts. We have used a number of different datasets in these studies, including social media, socially constructed data available online, and proprietary sets of data represented as spreadsheets and other formats.

3.1 Context

The research that led to the development of this process model for data-driven network analysis began in the context of studying complex networks of relationships in innovation ecosystems. Russell et al. (2011) use the concept of innovation ecosystem to refer to the inter-organizational, political, economic, environmental, and technological systems through which a milieu conducive to business growth is catalyzed, sustained, and supported. A dynamic innovation ecosystem is characterized by a continual realignment of synergistic relationships that promote growth of the system (Russell et al., 2015).

Ecosystems are a complex phenomenon, with multiple entities connected through multiple level relationships, as well as multiple stakeholder perspectives into those relationships. Ecosystems that promote innovation have become a quest for companies, cities, regions and countries. It is agreed that “relationships shape the behavior and outcome of all stakeholders as well as the system-level effects” (Hwang & Horowitt, 2012), and that it is through the relationships of individuals within and across organizations in an ecosystem that knowledge transfer, technology dissemination and organizational change are accomplished (Russell et al., 2015). Program managers and policy analysts in charge of transforming innovation ecosystems seek to define and describe innovation ecosystems in order to set goals, determine interventions and evaluate change, and visualizing the innovation ecosystem has proven instrumental to strategy setting and decision-making (Still et al., 2014). By making the roles and relationships explicit, both numbers and visualizations can be used to support the creation and management of innovation ecosystems. By tracking the provenance of data and authorship of analytical refinements, the collaborative exploration gains transparency.

To manage as well as to create innovation ecosystems, network orchestration has been encouraged (Paquin & Howard-Grenville, 2013; Ritala, Armila, & Blomqvist, 2009; Still et al., 2014). A data-driven process for understanding roles allows for interactive discovery of the innovation ecosystem. Multiple perspectives can be invited and exchanged in the process of developing and orchestrating transformation programs. With subsequent automation of data updates and tracking analyses, the assumptions and contingencies underlying decisions can be monitored for changes that would impact policy and program directions.

3.2 Design Science Research

In this research, we take a Design Science Research (DSR) approach to describe a process model for data-driven visual network analytics applicable in replicable investigations of innovation ecosystems as well as other domains in which network structures over time are of interest. DSR is a research method that allows “learning and investigation through artifact construction” (Vaishnavi & Kuechler, 2007, p. 187). “Whereas natural sciences and social sciences try to understand reality, design science attempts to create things that serve human purposes” (Simon, 1969). The rationale for DSR hails from the importance of the practical utility of research (Peffers et al., 2007). Design science research aims to build a bridge between information system (IS) research and its practical application by producing results that have real-life relevance. "Design science [. . .] creates and evaluates IT artifacts intended to solve identified organizational problems" (Hevner et al., 2004).

The General Design Cycle (GDC) is a key part of any DSR process (Vaishnavi & Kuechler, 2007). The process begins from the awareness of the problem and continues to one or more suggestions for solution. Next, an implementation of the plan is developed and evaluated, and finally, the process is concluded and the results shared; in the case of a scientific process, they are published. In each of these steps, new knowledge is both created and fed back to previous phases. The phases are repeated in an iterative fashion until a satisfactory end result (one that has practical utility) is achieved.

Readers with experience in software development will notice a straightforward connection between general design cycle and agile software development (see e.g. Schwaber & Beedle, 2001). Apart from the intent to publish the results, both design and development processes move forward in an iterative and incremental fashion and are guided by feedback collected from the users and other stakeholders of the developed software or other artifact, here the process model.

To develop the Ostinato Model for data-driven visual analytics presented in this chapter, we effectively applied and repeated the General Design Cycle. To evaluate the process model for added credibility of the presented results, we applied the Experimentation Pattern defined by Vaishnavi and Kuechler (2007), more specifically the case-based prototype development pattern on which the prototype is developed in an incremental, iterative manner over a number of cases, leading to deep knowledge of the problem and the proposed solution.

3.3 Experimental Cases

The Ostinato Model has been developed over a number of cases in which a variety of innovation ecosystems have been investigated in collaboration with their stakeholders using various sets of data sources (Basole et al., 2012; Jussila et al., 2014; Rubens et al., 2011; Russell et al., 2015). Table 1 describes core cases in which the automation-exploration cycle was implemented, using structured and semi-structured data sources, involving stakeholders in the exploration process as well as the sensemaking of key visualizations and other outputs for each case.

4 Ostinato Model

This section presents summary of the results of our research. First, we describe the requirements for the data-driven visual network analytics process; these requirements stem from existing process models and are augmented through results that emerged in case studies on which we applied the method. Second, as the core contribution of this chapter, we describe the process model for the exploration–automation cycle of data-driven visual analytics, the Ostinato Model.

Table 1 Illustrative cases for developing the exploration-automation cycle and the process model for data-driven visual network analytics

Case	Data	Co-creators/case shareholders	Visualizations/outputs
Demola (Huhtamäki et al., 2013)	Proprietary data on Demola projects, the companies that initiated the project and university affiliations of project members (university students)	Demola leaders and operators and the investigative team	The animation of the evolution of Demola project sphere including projects, the affiliations of project team members and companies. Multimode networks on (1) projects and affiliated actors and (2) projects and their key competences
Tekes Young Innovative Companies (Huhtamäki et al., 2012)	Innovation Ecosystem Network Dataset on growth companies, Twitter data on Tekes Young Innovative Companies (YIC) and their followers	Policy makers at Tekes—the Finnish Funding Agency for Innovation and the investigative team	One and two-step networks of the companies part of Tekes YIC program and their affiliations to investors and key individuals
Finnish Innovation Ecosystem (Still et al., 2013)	Three separate datasets: (1) Thomson Reuters SDC for deals and alliances and IEN Dataset for (2) Executives and Finance and (3) Startups and Angels	Finnish national-level policy makers and the investigative team	Network visualizations and metrics about companies having their main office in Finland and their first-step connections to other companies, investors and key individuals
Network orchestration for EIT ICT Labs (Still et al., 2014)	IEN Dataset for Executives and Finance	EIT ICT Labs representatives and the investigative team	Network visualizations of companies having their main office in one of the EIT ICT Labs co-location centers and their first-step connections to investors and individuals as well as to other companies through investments and acquisitions

4.1 Process Requirements

Developed through several rounds of iterations following the General Design Cycle, the core guidelines and requirements for the data-driven visual network analytics process model include the following: continuous data collection; exploration; transparency; loose coupling; reproducibility; automation; enabling manual steps; low entry barrier; and interoperability. Each is described.

Continuous data collection. When collecting data from social media, persistent processes are often needed, particularly when the investigators want to capture both the structure and dynamics of a phenomenon. Twitter, for example, currently provides only limited access to its historical data, and even then data on followers and friend connections between users do not include timestamps. At times, collecting the data takes days or weeks or "forever" to complete, due to throttling or other technical limitation or the sheer size or the dynamic nature of source data.

Exploration. A visual analytics approach is key to enable users with varied technical skills to collaboratively explore and make sense of a phenomenon. Being able to follow the visual analytics approach requires process flexibility. That is, all the stakeholders of the analysis process should be able to conduct any of the individual steps by themselves even though development of the overall process requires technical development skills.

Transparency. Developers with technical skills may select to manage the network analysis data, in its different phases, with a database. To accomplish transparency and flexibility in the process, other members of the investigative team may, however, need less technical means to access the data. The use of intermediary results is key in facilitating the transparency and flexibility of the process. Intermediary results refer to data in between the individual steps of the analysis. These data should be available as files in widely used formats, such as CSV and GEXF. In addition to the enhanced transparency, these intermediary results allow for speeding up the analysis process by using cached versions of source data and intermediary results when they have not changed.

Loose coupling. At best, data-processing pipelines can be built with a range of tools and components that have been implemented with different technologies. This kind of flexibility allows the introduction and use of new expressive tools from individual software components to full-featured applications as they become available to the investigative team. Many of them introduce new opportunities for advancing the analysis process but generally it is not possible to integrate these tools to a data-processing framework in program code (API) level.

Reproducibility. In the data-driven visual network analytics approach, reproducibility is first and foremost a technical quality of the process: the investigative team should be able to repeat the study or one or more steps of the analysis process and reproduce the results. Reasons for the need to rerun the process include, among others, updates on the source data, development steps of the analysis process, and the introduction of completely new processing steps and tools that insist on the use of a particular data format or extending the existing data. Moreover, dynamic sensemaking for complex phenomena mandates being able to refresh the data and derive new results with updated data. Reproducibility at this technical level also allows the investigative team to release the process, data and results to other researchers interested in the phenomena under investigation.

Enabling manual steps. While reproducibility is important, at the same time it is important to realize that automating some of the steps may not be feasible when an analysis is conducted the first time or requires intensive tailoring. Therefore, the process should support implementing any of the process steps manually. The use of file-based intermediary results is a practical approach in enabling manual analysis steps.

Automation. Allowing the development of automatically updating dashboards as needed gives the investigative team the opportunity to continue observing particular phenomena over time. It is expected that production-ready analysis processes for dashboards will operate without supervision; however, in the context of exploratory research, some requirements may be relaxed.

Low entry barrier. Analysis of innovation ecosystems and other network-based investigations of complex phenomena require extensive domain knowledge, and hence insist on active participation from domain experts (often without extensive technical expertise) throughout the analysis process. This requirement further underlines the need for transparency of the analysis process and the individual analysis steps.

Interoperability. The investigative team should be able to use a number of existing analytics tools with high usability and rich interactivity such as Gephi, NodeXL, KNIME and Tableau for conducting the analysis. Moreover, provisioning the visualized networks and other outputs of the analysis should be possible through dashboard built with Web technologies such as D3.js, DC.js, GEXF.js and the like.

In terms of the General Development Cycle, these requirements can be used to describe the Definition of the problem that serves as the starting point of artifact development (cf. Vaishnavi & Kuechler, 2007). These requirements form a design rationale for the Ostinato exploration–automation cycles of the process model for data-driven network analysis.

4.2 Process Model

The Ostinato process model that is presented in this section is developed over multiple case studies with a design research approach. It is built on existing models and previous work, and it takes into account the process requirements presented in Sect. 4.1. Each step is described. Figure 2 shows a diagram of the process model.

Phase 1: Data Collection and Refinement

1. Entity Index Creation
2. Web/API Crawling
3. Scraping
4. Data Aggregation

Phase 2: Network construction and visualization

5. Filtering in Entities
6. Node and Edge creation
7. Metrics Calculation
8. Node and Edge Filtering
9. Entity Index refinement
10. Layout Processing
11. Visual Properties Configuration
12. Visualization Provision
13. Sensemaking, Storytelling & Dashboard Design

4.2.1 Phase 1: Data Collection and Refinement

The general rules of data-driven analytics apply here: collecting and cleaning the data will in most cases consume most of the time and resources available for the investigation.

Entity Index Creation

In some cases, the source data can be collected in full; whereas, in other cases only data on entities that are relevant for the analysis need to be collected. In one use case, we were interested in the Twitter discussions taking place in relation to a conference, #cmadfi. We collected all the Tweets sent by conference participants before, during and after the event in order to create a network representing the social structure of the conversation. For this, we created an entity index including the Twitter handles of conference participants, as well as those mentioned in the discussion (Jussila et al., 2014).

In the context of innovation ecosystem studies, the entities for which we collected data were defined by boundary specification (Basole et al., 2012). For example, in investigating the connections between companies taking part in Young Innovative Companies program[2] run by the Finnish Funding Agency for Innovation Tekes, the list of companies defined the starting point of the analysis (Huhtamäki et al., 2012).

Web/API Crawling

Collecting the data is the most heterogeneous step in the data-driven visual analytics process. Possible source data potentially includes everything digital, from proprietary offline documents and document collections to spreadsheets to Web

[2] Funding for young innovative companies, http://www.tekes.fi/en/funding/companies/funding-for-young-innovative-growth-companies/

APIs (Application Programming Interface) to Web sites that are designed primarily for human interaction.

Similarly, the functionality required to collect the source data can range from relatively simple reading of individual documents to functions similar to a fully featured Web crawler. Compared to crawling random websites, Web APIs are, by default, more straightforward for data collection as they are often designed to support reuse (Vinoski, 2008). At best, source data is available as linked data (Bizer, Heath, & Berners-Lee, 2009), i.e. data that has a clear structure with individual facts that can be interconnected with the help of unique identifiers. This is key in ensuring referential integrity.

At the end of the crawling phase, a set of web resources, or rather their representations in Hypertext Markup Language (HTML) or some other format, is made available in a local database or other storage, a proxy that significantly speeds up the subsequent processing steps.

Scraping

Once the raw source data is available locally, the next step is to filter, select and distill the utility data relevant to the analysis process. Scraping refers to the process of distilling data from documents that are published to the Web for humans to use. Scraping can be seen as a form of the Extract, Transform, Load (ETL) process that is often applied in the context of data warehousing or other business intelligence processes to collect data from different sources to be refined and normalized and finally loaded into a consistent database for later use (Petschulat, 2010; Vassiliadis, 2009).

When collecting data from Wikipedia on Finnish Young Innovative Companies (YIC), for example, we were particularly interested in the facts presented in the Infobox section[3] of the page. To collect this data, we took advantage of the HTML markup on the page to specify the semantics (meaning) of the different pieces of text.[4] Each of the facts is represented as a table row including two cells, the first of which includes the label specifying the type of the fact and the second includes the actual value. Moreover, the value is also represented as a link to a separate page, a fact that we included in the crawl.

[3] Help:Infobox, http://en.wikipedia.org/wiki/Help:Infobox

[4] The Terms of Service for a Web page must also be considered. When using Wikipedia as a data source, for example, one has to take into account the Terms of Service that specifically deny crawling Wikipedia for large amount of files. Instead of crawling the live website, users of the data are advised to download a copy of Wikipedia’s contents and set up a proxy for serving further processing.

Data Aggregation

Social media studies often take place within the boundaries of an individual social media service; and therefore, ways of accessing data and identifying individual entities can be straightforward when one source of data is used. The complex context of innovation ecosystem studies, however, led us to use several sets of data in parallel. This meant that in many, if not most, of the cases, linked data was not readily available; and therefore, links between individual sets of data had to be created through finding unique entity identifiers that allow referential integrity. In innovation ecosystem studies, the name of the company or another actor is sometimes the key data point that can be used to identify an entity; in other cases, more advanced entity recognition procedures can be applied.[5] This kind of data cleaning is sometimes referred to as data wrangling (Kandel et al., 2011). Applying the methods of entity recognition provides a potentially more general solution to creating unique identifiers for entities in the data.

4.2.2 Phase 2: Network Construction and Analysis

Once the data is available on a local proxy, the utility data has been extracted from the source documents and data from different sources has been aggregated into a consistent set of linked data, the construction of the network representation of the phenomena under investigation can begin.

Filtering in Entities

The network construction phase starts by selecting the entities that will be included in the network. The selection of nodes is guided by the boundary specification designed and defined by the investigative team. At least two approaches exist to implement the selection: starting from a list of entities and rule-based entity inclusion. To continue the Finnish YIC example, we started from the list of companies participating in the program. We scraped Wikipedia data on the connections between the YIC companies and key individuals running them. If data on the individuals was not available in a clean format, we followed the crawling pattern by including the individuals in the list of web resources to be crawled. We continued to complement the dataset with data from the Innovation Ecosystems Network Dataset (IEN Startups and Angels, IEN Executives and Growth) and other sources of data about investments, acquisitions and affiliations.

[5] When using names as identifiers, one can apply fuzzy string matching and semi-automated tools such as OpenRefine (http://openrefine.org/) or DataWrangler (http://vis.stanford.edu/wrangler/) to assist in the aggregation process.

A key reason to separate the selection of entities from node and edge construction is to support the transparency, reproducibility and extensibility of the process. To create a shared understanding of the analytical results, it is absolutely vital that all the investigators taking part in a particular network study are able to understand the original raw data, in addition to any constructed variables, and the various analytics and metrics that represent the network; this means that investigation participants need access to the analysis process as a whole, including access to the raw data. In our experience, we found that answering specific questions raised by anyone interested in the study, drawing conclusions, generalizing the results, developing more specific and potentially more interesting questions all depend on transparency of the data available and used for the analysis.

Node and Edge Creation

A key part of the data-driven network analysis process is, of course, the actual creation of the network. Network creation boils down to the creation of nodes representing the actors and the creation of edges representing the connections between the actors. Several options are available, however, when specifying details of the network creation process. First, the network can be either one-mode or two-mode. In one-mode networks all the nodes are of same type: startup companies, for example. Connections between the nodes are formed through relationships: investments, affiliations to individuals, acquisitions and transactions. In two-mode networks, there are two types of nodes, for example, startup companies and individuals related to them. Hypergraphs and bipartite graphs are examples of means to visualize two-mode networks (Freeman, 2009; Jesus, Schwartz, & Lehmann, 2009).

Further, the connections between network nodes can be either valued or dichotomous. With valued connections, the strength of a connection can be expressed. In either case, the connections may be undirected or directed. Finally, the temporal dimension can be included in networks if the data used to create the connections is time-stamped. With temporal data, insights about the evolution of the network can be gained.

Metrics Calculation

Network metrics enable quantifying a variety of structural properties, both in network and node level. These range from simple metrics such as node degree (indegree, outdegree) and betweenness to hub and authority values with HITS and other more sophisticated measures. Whereas in principle, every metric can be calculated for all of the networks and their nodes, in practice this is not feasible due to reasons of efficiency. Moreover, new metrics are being developed continually, and the investigative team is likely to find—or develop—new metrics that fulfill specific investigative purposes. From an implementation viewpoint, it is unlikely to find one tool that supports all the metrics the team wishes to use. Therefore, a combination of tools may be required to calculate the metrics.

As part of this step, network metrics for the network representation should be archived for later usage. For transparency, a list of exported network nodes and edges should include the various metrics used. In practice, node and network metrics must be recalculated after each change in the network structure; however, reference to previous calculations is often needed.

Nodes and Edge Filtering

A key limitation in visual network analysis is the amount of space available, both on screen and particularly on paper, to present the visualization. Depending on the level of detail required in the analysis, hundreds or thousands of nodes can be presented in one visualization view. For networks of tens of thousands of nodes and more, only more general structures and patterns can be observed from the visualization. Two means exist to address this limitation: the best option is to allow the visualization users to filter in and out nodes and edges. If the end-user tools used to present the visualizations do not allow filtering, it can be done as one part of the automated process. Often, reducing the size of the visualized network is accomplished with a combination of filtering out edges that have the least amount of weight as well as filtering out nodes that: (1) are left without edges; (2) have a value of the degree or some other a network analysis metric under a specified threshold; or (3) are (not) of particular type (even though this can already be taken into account when filtering in the entities used to construct the network in the first place).

Entity Index Refinement

At this stage, the network is constructed and the required metrics are calculated for each of the nodes. Depending on the boundary specification applied in a particular investigation, the network is either ready to be visualized or, alternatively, additional data can be collected to complement the network. Revisiting the Finnish Young Innovative Companies case, the boundary specification was designed to include all the individuals involved in one or more of the companies in YIC program as well as all the other companies the individuals are or have been affiliated with. Moreover, the data included all the investors that had invested into any of the companies as well as all the companies that had acquired any of the YIC companies.

Layout Processing

The principle of processing network layout is simple. Nodes are given a position in two-dimensional space in a way that network structure is revealed in an intuitive way. Despite the simplicity, novel layout algorithms have continued to be developed over several decades. In our research cases, various stakeholders found a specific implementation of force driven layout, Force Atlas, to be particularly suitable for laying out networks representing innovation ecosystems at different levels. Force Atlas is implemented in Gephi and can be used as a batch process with the help of Gephi Toolkit.[6] In practice, the parameters of the layout algorithm must be adjusted manually for a particular kind of a network before fully automating layout processing. Alternatively, the layout can be processed with the UI version of Gephi and the resulting network, including the XY-coordinates for each node, can be exported, e.g. in GEXF.

Storing the network layout data is particularly important for improving the efficiency of the layout process, as well as for reducing investigators' cognitive load and promoting transparency. In particular, it is important that after the data is refreshed, the investigators are able to find the pre-existing nodes in an area of the network where the nodes were previously located. This stability can be achieved by inserting the existing positions into the network data before re-running the force driven layout algorithm. In most cases, investigators will find the pre-existing nodes close to the initial area of the network.

Future work is needed to determine how features such as layout algorithms, e.g., those implemented into NodeXL, could be used as a component of data-driven visual network analysis pipelines.

Visual Properties Configuration

In networks, there are limited selection possibilities when defining the visual appearance of nodes and edges. Nodes have size, color and perhaps a border and shape as elected visual features. Edges have color and width. Allowing the user to select and change the visual properties according to node metrics and other node properties is perhaps the easiest way to allow end user interactivity in network analysis. Depending on the tools used by the investigators to conduct the analysis, the visual properties of nodes and edges can continue to be tweaked as part of the interactive analysis process.

Visualization Provision

At this stage, a network has all the required information available and therefore can be visualized. The means to finalize this step depend greatly on the tools that have been selected for use by the investigative team. In most cases, however, the created network is serialized into a file following a selected vocabulary or format for

[6] Gephi Toolkit, http://gephi.github.io/toolkit/

representing a network. These vocabularies and formats range from different CSV based applications to XML-based languages designed for representing networks.

A minimum approach to provision the network visualizations is to export network data in GEXF or other suitable format and place the resulting file into a folder from where a library such as Gexf.js can access it. More generally, viewer composition scenarios can include the following:

Scenario 1. Network viewer component with fixed functionality, i.e. following a fully descriptive approach. Visual properties such as node size and color need to be defined into the data during its processing. Gexf.js is an example of such a component that we have found useful in adding value to a fully static PDF-based approach in disseminating network visualizations.

Scenario 2. Implementing a dashboard with Web technologies, more specifically frameworks such as Highcharts, D3.js, Crossfilter.js, DC.js and others. In this case, tailored interactive features for data exploration can be provided to the user, adding options for representing network data.

Scenario 3. Using full-feature explorative analytics tools such as Gephi, NodeXL and Tableau, which can be used to further process the data and to connect source data to visual properties of the visualization. The key here is to produce visualizations rich-enough in data that the analyst can fully utilize the critical properties of the chosen analytics tool for investigation and exploration. In Gephi, for example, it is useful to include attribute data for nodes to assist network filtering in a way the investigator desires to do.

Sensemaking, Storytelling and Dashboard Design

While information visualization includes data transformation, representation, and interaction, it is ultimately about harnessing human visual perception capabilities to help identify trends, patterns, and outliers. Sensemaking has its roots in cognitive psychology and many different models have been developed. Sensemaking procedures are cyclic and interactive, involving both discovery and creation (North, 2006). During the data collection and refinement phase, an individual searches for representations. In the network generation phase these representations are instantiated, and based in these insights the representation may be shifted, to begin the process again. Sensemaking is closely linked to the insight objectives (Konno, Nonaka, & Ogilvy, 2014), and the Ostinato cycle of exploration–automation is key in achieving actionable insights that network orchestrators can utilize.

When sensemaking requirements are satisfied for investigators and users, steps of the Ostinato process can be formalized with automated procedures for iteration over time. Key actors, relationships and events of the network can be incorporated into dashboards that will track changes in critical assumptions and into stories that will share vision for actionable change.

5 Discussion

The present chapter adds a new perspective on the heuristic and application development process that may lead to new tools, applications, services, and algorithms dedicated to understanding how social media content is created, curated and disseminated and how the authority and trust of social media content creators accrues and how this matters in terms of trust and credibility. The Ostinato Model contributes to this call in two levels. First, it can be applied to support the data-driven investigations of innovation ecosystem structure and dynamics. Moreover, in the context of our investigations, social media serves first and foremost as a source of data that is fed into the investigations of innovation ecosystems and the structure between their actors. Therefore, second, for validity and reliability of these investigations, it is key to be able to increase the transparency of the processes behind these data originating from social media.

The Ostinato Model contributes to the data-driven network investigations of social media, innovation ecosystems and other network-driven phenomena in several ways. First, the network approach has great strength in supporting the explorative studies of the patterns in between actors creating, curating and disseminating social media content. Second, referring specifically to the first phase of the Ostinato Model, data-driven approach allows tracking down processes over the boundaries of individual social media platforms and services. Third, user-centricity of the data-driven process adds to the transparency of the process itself, therefore providing means to triangulate different phases of data refinement and transformation and allowing different stakeholders of investigations to take as proactive role as they wish in moving forward a particular investigative process.

Due to the continued and rising interest in social media analytics and general big data analysis, new tools are continually introduced to support investigative work. Despite the tool development, a combination of tools is likely to continue to provide more flexibility in accessing and aggregating data and in processing and analyzing it. Finding a balance between user interface-operated low barrier tools and expressive computational strategies that require technical knowledge is key in making the investigative process as productive as possible while maintaining transparency and process flexibility.

This Ostinato Model for user-centric, process-automated, data-driven visual network analytics meets many of the requirements outlined earlier in this chapter for the exploration–automation cycle recommended for developing shared understanding.

Setting up persistent data-collecting routines requires, in general, a programmatic implementation and must be designed and implemented case by case. To maintain the transparency of the process, it is important that the investigators are able to access both the raw data as well as to track down the various steps used to derive the data that is eventually used for the analysis and visualizations.

Allowing exploration boils down to the selection of the end user tools available for investigators to visualize and explore the data. If a rather static tool such as Gexf.js, for example, is used, the user is limited to browsing and searching the data.

If importing the data into an exploration platform such as Gephi or NodeXL is permitted, it is possible to provide the user with node and edge data, enabling them to continue their explorations with more technical independence. The availability of particularly expressive visual analytics tools, such as Tableau, adds to investigation options of analyzing network data, either as a network or using node and edge level data to provide new inspirations for other kinds of data analyses.

Using files rather than databases for representing intermediary results supports both loose coupling and transparency of the process. It also allows for implementing some of the steps manually, if seen feasible, and the flexibility of the process in general is increased.

Reproducibility is both a technical and a policy requirement. For an investigative team revisiting or extending an existing case, the availability of runnable code, source data and intermediary results provides a fruitful starting point. Moreover, results of reproducible studies can be published in a way that both data and runnable code are available, allowing a solid foundation for others to add their contributions as well. A reasonable proposition is that such a piece of knowledge draws attention from other researches and therefore has true potential for impact.

Automation is a key requirement for reproducibility, as well as for creating a dashboard that continues to update visualizations of the phenomena under investigation, sometimes in close to real time.[7]

Low entry barrier is enabled through making intermediary results available to all the members of the investigative team. As the process is repeatable and its individual steps are automated, new projections of the data can be implemented in an iterative and incremental manner. Implementing completely new steps of analysis becomes possible even without technical skills. Automating the steps, however, requires developers' attention. The Ostinato process model requires a multidisciplinary data science team or the somewhat mystical multi-skilled data scientist (cf. Davenport, 2014) to conduct the investigation.

Interoperability can be built into a computational approach. This requires that the technical architecture is flexible enough to permit different software components and tools—that may be implemented with different technologies—to be introduced into the process. When an analysis pipeline is built completely from scratch, it is recognizably important to minimize the number of technologies used. However, moving fast and in an agile manner is an objective we claim can be achieved when existing tools can be integrated to implement the individual steps of the analysis process and to provide the visualizations to investigators and other end users.

An implementation of the Ostinato user-centric, process-automated model for data-driven visual network analytics can serve as the core engine of an investigation. It can also be used to develop a pre-processing pipeline that collects and

[7] Using a full stack programming language such as Python gives the developers more opportunities to turn the scripts developed for analysis into processes that run in the cloud, intermittently collecting and preprocessing the data and feeding results into dashboards implemented in Web technologies.

refines the data, creates a network representation and serializes the outputs to be analyzed and processed with expressive tools that, standing alone, allow the full visual analytics cycle for users.

6 Summary

In this chapter, we have presented the Ostinato Model of the exploration—automation cycle user-centric for data-driven visual network analytics. This model has two main phases, data collection and network analysis; they iterate through a cycle of exploration and automation. The Data Collection and Refinement step is divided into Entity Index Creation, Web/API Crawling, Scraping, and Data Aggregation. The Network Creation and Analysis step is composed of Filtering in Entities, Node and Edge Creation, Metrics Calculation, Node and Edge Filtering, Entity Index Refinement, Layout Processing, and Visual Properties Configuration. As a final step, the visualizations are provisioned to investigators and other end users with interactive exploration tools and discussion, and their feedback activates an iteration of the process. This Ostinato process model allows both an exploratory approach during the early phases of the investigation as well as the automation of the data collection and analysis process. The iteration cycle is especially beneficial in working with multi-source datasets, complex phenomena, changing externalities that may impact assumptions for decisions, and establishing a dashboard for continued observation of the phenomena over time, perhaps in real time.

A key challenge of this approach concerns the number of options for investigators and other end users to interact with the data in real-time while conducting the analysis, particularly the non-technical investigators on a multi-disciplinary team. The design research approach favors an iterative approach for both data-driven explorations and evidence-based decision making. However, investigators with limited programming skills or related technical know-how are limited in their participation, even though they may possess vital domain intelligence. Through access to data, documentation of changes in the analytical approach, flexible means to produce network representations in various formats, and exposition of intermediary results, barriers to participation can be lowered. The cycle of exploratory visual analytics, confirmation of data selection rules and analytical results made accessible through high interactivity visual analytics, allows the investigative team to confirm assumptions and investigative procedures, identify aspects of the analysis that can be automated and establish a transparent, replicable process.

The Ostinato process model has several implications for investigative teams taking the data-driven visual network analytics approach.

First, facilitation and documentation of the investigative process are required. Low barrier for entry in exploration and analysis poses risks that increase without transparency. Put another way, with added transparency and through intermediate results and easy access, the risk of false conclusions is lowered. Co-ordinated discussion on raw data and its journey to the finalized visualizations and other results is imperative; documentation of assumptions and rationale for changing data selection or analytical procedures enables transparency. Facilitation also helps in creating literacy of the processes and its outputs within the investigative team. Having the intermediate results available, all the members of the investigative team are able to maintain more of the control of the process and continue to introduce new, novel ways of analyzing the data as their skills and methodological know-how allows.

Second, the cycle of exploration–automation introduces new requirements for governance. Intermediary results require transparent authorship in their provenance. The transparent authorship of new datasets, constructed variables and analytical iterations must be ensured.

Third, starting from exploration and moving toward automation is straightforward with the help the process model. The investigative team is able to move fast in the beginning of the process while, at the same time, maintaining control over the process as its complexity increases. With appropriate technology selection, the process can eventually be relegated to the background to collect, process, analyze and visualize data in an automated manner to support a longitudinal study of a particular phenomena. And, more importantly, a mature procedure—or one or more of its components—can be reused to investigate other phenomena of interest.

Fourth, increased reproducibility is an asset for future studies but requires explicit governance. Technical reproducibility of the process allows revisiting analytical results of a case even after a long time period. Refreshing (collecting new) data or, alternatively, adding new dimensions into existing data is straightforward when the process or its individual parts can be run computationally. Curational rules must be developed, and access to code and data has to be designed at both the technical and policy levels. Governance of the data from raw to intermediate results to outputs as well as the components and software process must be articulated.

Within the constraints imposed by the level of abstraction in this article, this Ostinato process model provides blueprints for designing analytical processes with technologies ranging from Python to R to Javascript. At best, the process is able to support the inclusion of several different technologies, as implemented e.g. by the Wille Visualisation System (Nykänen et al., 2008).

Future work includes, first, the refinement of this model on basis of the feedback collected from researchers and practitioners working with the exploration–automation cycle of data-driven visual network analytics and applying the model and, second, the implementation of a software framework—perhaps similar to Grunt (http://gruntjs.com/), a popular Javascript-based task runner—to support the development of processes of data-driven visual network analytics on very large datasets.

As an ecosystem of tools and components develops and requirements for interoperability are articulated, we see the possibility of developing a community of people moving the field forward. They will need a package management framework, system components and a supportive community.

The Kredible.net initiative is an important step toward establishing a community like this.

Acknowledgement The research reported in this chapter was funded through resources provided by Tekes—the Finnish Funding Agency for Innovation and mediaX at Stanford University.

References

Barabási, A.-L., & Bonabeau, E. (2003). Scale-free networks. *Scientific American, 288*(5), 50–59.

Basole, R. C., Russel, M. G., Huhtamäki, J., & Rubens, N. (2012). Understanding mobile ecosystem dynamics: A data-driven approach. In *Proceedings of the 2012 International Conference on Mobile Business* (*ICMB 2012*) (pp. 17–28). Delft, The Netherlands. Retrieved from http://aisel.aisnet.org/icmb2012/15/

Bizer, C., Heath, T., & Berners-Lee, T. (2009). Linked data: The story so far. *International Journal on Semantic Web and Information Systems, 5*(3), 1–22.

Card, S. K., Mackinlay, J. D., & Shneiderman, B. (1999). *Readings in information visualization: Using vision to think*. San Francisco: Morgan Kaufmann. Retrieved from http://www.amazon.com/Readings-Information-Visualization-Interactive-Technologies/dp/1558605339

Davenport, T. H. (2014). *Big data at work: Dispelling the myths, uncovering the opportunities*. Boston: Harvard Business Press Books.

Freeman, L. C. (2000). Visualizing social networks. *Journal of Social Structure, 1*(1). Retrieved from http://www.cmu.edu/joss/content/articles/volume1/Freeman.html

Freeman, L. C. (2009). *Methods of social network visualization*. Berlin, Germany: Springer.

Ghosh, S. (2013). *Python tools for reproducible research in brain imaging*. Retrieved from https://speakerdeck.com/satra/pydata-2013-python-tools-for-reproducible-research-in-brain-imaging

Giuliani, E., & Bell, M. (2008). *Industrial clusters and the evolution of their knowledge networks: Revisiting a Chilean case*. Brighton, England. Retrieved from http://www.sussex.ac.uk/spru/documents/sewp171

Granovetter, M. (1973). The strength of weak ties. *American Journal of Sociology, 78*(6), 1360–1380. Retrieved from http://www.jstor.org/discover/10.2307/2776392?uid=3737976&uid=2&uid=4&sid=21104852601921

Hansen, D. L., Rotman, D., Bonsignore, E., Milic-Frayling, N., Rodrigues, E. M., Smith, M., & Shneiderman, B. (2009). *Do you know the way to SNA?: A process model for analyzing and visualizing social media data*. University of Maryland Technical Report: HCIL-2009-17. Retrieved from http://www.smrfoundation.org/wp-content/uploads/2010/05/2009-UMD-TechReport-Do-you-know-the-way-to-SNA.pdf

Hansen, D., Shneiderman, B., & Smith, M. A. (2011). *Analyzing social media networks with NodeXL: Insights from a connected world*. Burlington, MA: Morgan Kaufmann. Retrieved from http://www.amazon.com/dp/0123822297

Heer, J., & Shneiderman, B. (2012). Interactive dynamics for visual analysis. *Communications of the ACM, 55*(4), 45–54. Retrieved January 31, 2013, from http://dl.acm.org/citation.cfm?id=2133821

Hevner, A. R., et al. (2004). Design science in information systems research. *MIS Quarterly, 28*(1), 75–105.

Huhtamäki, J., Salonen, J., Marttila, J., & Nykänen, O. (2010, October 3–6). Context-driven social network visualisation: Case wiki co-creation. In D. Karabeg & J. Park (Eds.), *Proceedings of the Second International Workshop on Knowledge Federation: Self-Organizing Collective Mind*, Dubrovnik, Croatia. Dubrovnik, Croatia: CEUR-WS.org. Retrieved from http://urn.fi/URN:NBN:fi:tty-201201161008

Huhtamäki, J., Still, K., Isomursu, M., Russell, M. G., & Rubens, N. (2012, September 20–21). Networks of growth: Case young innovative companies in Finland. In *Proceedings of the 7th European Conference on Innovation and Entrepreneurship (ECIE)*, Santarém, Portugal

Huhtamäki, J., Luotonen, V., Kairamo, V., Still, K., & Russell, M. G. (2013, October 1–3). Process for measuring and visualizing an open innovation platform: Case Demola. In *17th International Academic MindTrek Conference 2013: "Making Sense of Converging Media*". Tampere, Finland: ACM. Retrieved from http://urn.fi/URN:NBN:fi:tty-201312201533.

Hwang, V. W., & Horowitt, G. (2012). *The rainforest: The secret to building the next silicon valley* 1.02 Edition. Los Altos Hills, CA: Regenwald. Retrieved from http://www.amazon.com/The-Rainforest-Secret-Building-Silicon/dp/0615586724

Indarto, E. (2013). *Data mining*. Retrieved December 13, 2014, from http://recommender-systems.readthedocs.org/en/latest/datamining.html

Intel. (2013). *Extract, transform, and load big data with Apache Hadoop*. Retrieved from https://software.intel.com/en-us/articles/extract-transform-and-load-big-data-with-apache-hadoop

Jesus, R., Schwartz, M., & Lehmann, S. (2009). *Bipartite networks of Wikipedia's articles and authors: A meso-level approach* (p. Article 5, 10 pages). Orlando, FL: ACM. Retrieved September 2, 2010, from http://portal.acm.org/citation.cfm?id=1641309.1641318

Jussila, J., Huhtamäki, J., Henttonen, K., Kärkkäinen, H., & Still, K. (2014, January 6–9). Visual network analysis of Twitter data for co-organizing conferences: Case CMAD 2013. In *Proceedings of the 47th Annual Hawaii International Conference on System Sciences* (pp. 1474–1483). Computer Society Press. Retrieved from http://urn.fi/URN:NBN:fi:tty-201401221053.

Kandel, S., Heer, J., Plaisant, C., Kennedy, J., van Ham, F., Riche, N. H., ... Buono, P. (2011). Research directions in data wrangling: Visualizations and transformations for usable and credible data. *Information Visualization, 10*(4), 271–288. doi:10.1177/1473871611415994

Keim, D., Kohlhammer, J., & Ellis, G. (Eds.). (2010). *Mastering the information age—Solving problems with visual analytics*. Geneva, Switzerland: Eurographics Association. Retrieved from http://www.vismaster.eu/book/

Konno, N., Nonaka, I., & Ogilvy, J. (2014). Scenario planning: The basics. *World Futures, 70*(1), 28–43. Retrieved December 14, 2014, from http://www.tandfonline.com/doi/abs/10.1080/02604027.2014.875720

Leskovec, J., Backstrom, L., & Kleinberg, J. (2009). Meme-tracking and the dynamics of the news cycle. In *KDD '09* (pp. 497–506). New York: ACM. Retrieved September 14, 2012, from http://doi.acm.org/10.1145/1557019.1557077

Liu, Y.-Y., Slotine, J.-J., & Barabási, A.-L. (2011). Controllability of complex networks. *Nature, 473*(7346), 167–173. Retrieved June 15, 2011, from http://dx.doi.org/10.1038/nature10011

Mawer, D. (2000). Ballet and the apotheosis of the dance. In D. Mawer (Ed.), *The Cambridge companion to ravel* (p. 157). Cambridge, England: Cambridge University Press.

Moreno, J. L. (1953). *Who shall survive?: Foundations of sociometry, group psychotherapy and sociodrama*. Beacon, NY: Beacon House Inc. Retrieved August 23, 2010, from http://www.asgpp.org/docs/WSS/WSS.html

North, C. (2006). Toward measuring visualization insight. *IEEE Computer Graphics and Applications, 26*(3), 6–9. Retrieved from http://ieeexplore.ieee.org/xpl/login.jsp?tp=&arnumber=1626178&url=http%3A%2F%2Fieeexplore.ieee.org%2Fxpls%2Fabs_all.jsp%3Farnumber%3D1626178

Nykänen, O., Salonen, J., Haapaniemi, M., Huhtamäki, J., Huhtamäki, J., & Huhtamäki, J. (2008). *A visualisation system for a peer-to-peer information space* (pp. 76–85). Tampere, Finland: Tampere University of Technology. Retrieved from http://matriisi.ee.tut.fi/hypermedia/events/opaals2008/articlelist.html#opaals2008-article14.

Paquin, R. L., & Howard-Grenville, J. (2013). Blind dates and arranged marriages: Longitudinal processes of network orchestration. *Organization Studies, 34*(11), 1623–1653. Retrieved March 11, 2014, from http://oss.sagepub.com/content/34/11/1623

Peffers, K., Tuunanen, T., Rothenberger, M., & Chatterjee, S. (2007). A design science research methodology for information systems research. *Journal of Management Information Systems, 24*(3), 45–77. doi:10.2753/MIS0742-1222240302.

Peng, R. D. (2009). Reproducible research and biostatistics. *Biostatistics, 10*(3), 405–408. Retrieved December 14, 2014, from http://biostatistics.oxfordjournals.org/content/10/3/405

Petschulat, S. (2010). Other people's data. *Communications of the ACM, 53*(1), 53. Retrieved December 14, 2014, from http://cacm.acm.org/magazines/2010/1/55742-other-peoples-data/fulltext

Pirolli, P., & Card, S. (2005). The sensemaking process and leverage points for analyst technology as identified through cognitive task analysis. In *Proceedings of International Conference on Intelligence Analysis*.

Ritala, P., Armila, L., & Blomqvist, K. (2009). Innovation orchestration capability—Defining the organizational and individual level determinants. *International Journal of Innovation*

Management, 13(04), 569–591. Retrieved March 6, 2014, from http://www.worldscientific.com/doi/abs/10.1142/S136391960900242X

Ritala, P., & Hallikas, J. (2011). Network position of a firm and the tendency to collaborate with competitors—A structural embeddedness perspective. *International Journal of Strategic Business Alliances, 2*(4), 307–328. Retrieved December 19, 2014, from http://dx.doi.org/10.1504/IJSBA.2011.044859

Ritala, P., & Huizingh, E. (2014). Business and network models for innovation: Strategic logic and the role of network position. *International Journal of Technology Management, 66*(2), 109–119. Retrieved December 19, 2014, from http://dx.doi.org/10.1504/IJTM.2014.064608

Rubens, N., Russell, M., Perez, R., Huhtamaki, J., Still, K., Kaplan, D., & Okamoto, T. (2011). Alumni network analysis. In *Proceedings of 2011 IEEE Global Engineering Education Conference, EDUCON* 2011 (pp. 606–611).

Russell, M. G., Still, K., Huhtamäki, J., Yu, C., & Rubens, N. (2011, July). Transforming innovation ecosystems through shared vision and network orchestration. In *Proceedings of Triple Helix IX International Conference: "Silicon Valley: Global Model or Unique Anomaly?"*, Stanford, CA.

Russell, M. G., Huhtamäki, J., Still, K., Rubens, N., & Basole, R. C. (2015). Relational capital for shared vision in innovation ecosystems. *Triple Helix: A Journal of University-Industry-Government Innovation and Entrepreneurship*.

Schwaber, K., & Beedle, M. (2001). *Agile software development with Scrum*. Upper Saddle River, NJ: Prentice Hall.

Shakarian, P., Eyre, S., & Paulo, D. (2013). *A scalable heuristic for viral marketing under the tipping model*. Retrieved September 23, 2013, from http://arxiv.org/abs/1309.2963

Simon, H. A. (1969). *The sciences of the artificial*. Cambridge, MA: MIT Press.

Still, K., Huhtamäki, J., Russell, M. G., Basole, R. C., Salonen, J., & Rubens, N. (2013, June 16–19). Networks of innovation relationships: Multiscopic views on Finland. In *Proceedings of the XXIV ISPIM Conference—Innovating in Global Markets: Challenges for Sustainable Growth*, Helsinki, Finland (p. 15).

Still, K., et al. (2014). Insights for orchestrating innovation ecosystems: The case of EIT ICT Labs and data-driven network visualisations. *International Journal of Technology Management, 66*(2/3), 243–265.

Telea, A. C. (2008). *Data visualization: Principles and practice*. Wellesley, MA: A K Peters. Retrieved from http://www.amazon.com/Data-Visualization-Principles-Alexandru-Telea/dp/1568813066

Thomas, J. J., & Cook, K. A. (2006). A visual analytics agenda. *IEEE Computer Graphics and Applications, 26*(1), 10–13.

Vaishnavi, V. K., & Kuechler, W., Jr. (2007). *Design science research methods and patterns: Innovating information and communication technology*. Boca Raton, FL: Auerbach.

Vassiliadis, P. (2009). A survey of extract–transform–load technology. *International Journal of Data Warehousing and Mining, 5*(3), 1–27.

Vinoski, S. (2008). Serendipitous reuse. *Internet Computing, 12*(1), 84–87. Retrieved February 27, 2009, from http://steve.vinoski.net/pdf/IEEE-Serendipitous_Reuse.pdf

Wasserman, S., & Faust, K. (1994). *Social network analysis: Methods and applications* (1st ed.). New York: Cambridge University Press. Retrieved from http://www.amazon.com/dp/0521387078

Weng, L., Menczer, F., & Ahn, Y.-Y. (2013). Virality prediction and community structure in social networks. *Scientific Reports, 3*, 2522. Retrieved March 20, 2014, from http://www.nature.com/srep/2013/130828/srep02522/full/srep02522.html

Wikipedia. (2014). Web crawler. *Wikipedia*. Retrieved December 14, 2014, from http://en.wikipedia.org/w/index.php?title=Web_crawler&oldid=635502147

Tampereen teknillinen yliopisto
PL 527
33101 Tampere

Tampere University of Technology
P.O.B. 527
FI-33101 Tampere, Finland

ISBN 978-952-15-3833-9
ISSN 1459-2045