Visualizing Co-authorship Networks for Actionable Insights: Action Design Research Experiment

Jukka Huhtamäki
Tampere University of Technology
Korkeakoulunkatu 3
Tampere, Finland
jukka.huhtamaki@tut.fi

ABSTRACT

Increasing interest has been expressed in lowering the barrier for research access. Several approaches exist, including more active communication on research, the use of social computing-oriented networking tools for researchers, parallel publishing of research publications, and the use of research management systems for collecting, managing, and publishing bibliographical data. In this paper, we target the first step of research access, namely the use of publication metadata available in current research information systems. More specifically, we will take an action design research approach to experiment how visual network analytics could be used to create additional value for bibliographical data. We will tap into the current research information system of a selected university to develop a prototype of a self-service co-authorship network visualization and engage with four researchers to identify the key requirements for taking such an approach and to explore the potential value that could be created with visual analytics of bibliographical data. We contribute a set of design guidelines to support the development of computational visual network analytics tools for research collaboration analyses using bibliographical data.

CCS Concepts

•**Information systems → Data analytics;** *Reputation systems; Digital libraries and archives; Search interfaces;* •**Human-centered computing → Empirical studies in visualization;**

Keywords

Social network analysis; visual network analytics; ostinato model; scientometrics; action design research; human-in-the-loop analytics

1. INTRODUCTION

Despite the extensive discussion on digitalization, data-driven society, and digital infrastructures, the digital ecosystem for academic work is in its infancy. While the reasons for this slow development are beyond the scope of this paper, we suggest that the current transformation in the business models of the academic publishing industry is one of the main reasons for the slow uptake of digitalization.

The role of measurement and research data analytics is gaining importance in academic work. Therefore, academic institutions are developing or acquiring tools for managing research data. On one hand, this can be seen as the first step toward more open research practices. On the other hand, from the publishing industry's viewpoint, the need for bibliographic data management opens up new business opportunities around bibliographic data.

In this paper, we report an action design research (ADR) experiment in which data sourced from an institutional bibliographic data repository is used to conduct visual network analytics on the co-authorship network of a university. The objective of this experiment is, first, to identify key requirements for developing analytical tools for academic work in self-service mode, that is, without support from academic institutions. Second, we engage with four researchers to explore the potential for creating additional value for open bibliographic data through analytics.

This paper reports the very first findings of an experiment that will continue over several rounds of iteration. The overall objective of the experiment is to design tools and supporting processes for utilizing visual network analytics as a means for the researcher and other university actors to investigate the social structures latent in bibliographic data available in current research information systems (CRIS) currently deployed in universities in Finland[1] and around the world. In this experiment, we conduct the first round of the building-intervention-evaluation (BIE) cycle [38].

The overall objective of the experiment is to progress toward a system-level view of a university for actionable insights for individual researchers. In order to achieve this objective, we begin with the first draft of a specification for a holistic information system (IS) artifact, that is, an artifact that includes the information artifact, technology artifact, and social artifact. We will define the context more specifically as part of the methodology description in Section 3.

We begin this paper by laying the groundwork for self-service visual network analytics of university collaboration networks. We continue with a description of the methodology and related worldview we take in the experiment. Next, in Section 4, we review the key results from this first round

[1]Current research information systems (CRIS) in Finland, see http://scienceport.tut.fi/automation/cris

AcademicMindtrek'16, October 17-18, 2016, Tampere, Finland
© 2016 ACM. ISBN 978-1-4503-4367-1/16/10...$15.00
DOI: http://dx.doi.org/10.1145/2994310.2994340

of the experiment using the lenses of social, information, and technology artifacts. In Section 5, we discuss the results and define a set of design guidelines for metadata management for universities. Finally, we conclude with some general remarks.

2. BACKGROUND

This experiment draws from a number of existing bodies of knowledge and literature. In this section, we will cover the core parts of background knowledge. These include open linked data, linked data, application programming interfaces (APIs), resource-centric architecture, and data-driven visual network analytics.

2.1 Open linked data

Overall, in order to move forward with the quality and value-creation potential of bibliographical data, we point to two core sets of principles that should be applied when further developing the data management practices. The first set of principles stem from open data and the second from linked data.

Open science is an initiative to reintroduce openness and transparency to academic research. A number of dimensions are related to open science. From this investigation's viewpoint, the access to academic literature is the key dimension. While open access publications and parallel publishing introduce a means for providing free access to scientific publications, the availability of the bibliographic data, that is, metadata on publications, including their authors, title, abstract, publication venue and other details, is imperative to managing large bodies of literature. However, despite extensive discussion on open science and digitalization, in practice "even publications are not described in a machine-understandable way" and "are not openly accessible" [20].

In recent years, governmental organizations have been adopting open data practices to add transparency and introduce opportunities for businesses, individual developers, and others to create value to government data. Essentially, open data has to meet three criteria [8]. First, the data must be available and accessible in a straightforward manner, at best freely downloadable from the Internet. Second, terms for using the data must be defined explicitly and in a way that allows reuse, redistribution, and remixing with other data. Third, universal participation to data use, reuse, and redistribution should be enabled. In addition to the three core criteria, the data should be distributed in open machine-processable format.[2] Relatedly, the FAIR guiding principles for scientific data management and stewardship include four dimensions: data must be findable, accessible, interoperable, and reusable [45].

The second set of principles are drawn from linked data. Linked data refers to the use of the Web to create typed links between data from different sources [4]. Berners-Lee [3] defines four principles for linked data: use Uniform Resource Identifiers (URIs) as names for things; use Hypertext Transfer Protocol (HTTP) URIs so that people can look up those names; when someone looks up a URI, provide useful information, using the standards, including Resource Description Framework (RDF) and SPARQL; and include links to other URIs, so that they can discover more things. In contemporary Web development, these principles are often implemented following the RESTful approach for developing Web APIs [32].

[2]Open Definition 2.0, see http://opendefinition.org/od/2.0/en/

Linked science is an initiative to develop methods and practices that enable the interconnection of all scientific assets from data to methods and final publications [20]. Linked science goes beyond reproducible research where the objective is to publish data and analytical processes in a way that others can reproduce the results and develop them further [30, 31] to allow for aggregating and interconnecting sets—large or small—of scientific assets in an automated manner. Linked science draws from linked data and the practices of semantic web, web standards, open source software, web-based online environments, and cloud computing for "the legal and technical infrastructure by Creative Commons, and naturally also a joint effort by the scientific community and academic publishers" [21].

Open data is one of the cornerstones of open science, and moreover, bibliographic data is one of the core parts of open data for research. Open bibliographic data allows for the computational analysis of large quantities of scientific literature. Scientific practices that benefit from open bibliographic data include computational literature reviews, bibliometrics, and other means of scientometrics [24].

2.2 APIs and resource-centric architecture

In addition to open science, the presented experiment also draws from contemporary Web development. The architecture of contemporary Web applications routinely follows the RESTful approach, an architectural approach that is based on Roy Fielding's [10] dissertation work on representational state transfer (REST). Importantly, REST defines a way to apply the HTTP protocol to interact with resources exposed through Web APIs [11]. Moreover, APIs are a key concept in Web 2.0, a compilation of practices that transformed the Web from a page-centric collection of documents to an "architecture of participation" [28] in which users create and share data with the help of a network of Web applications and communication among each other through APIs. For Web developers, Web APIs and the RESTful approach have introduced the means and scale to both take advantage of and to engineer to support serendipitous reuse, effectively making the Web "an expansive application framework" [44].

Recently, APIs have entered the stage related to discussion on platform economy [9]. Simply put, in platform economy, companies and organizations seek ways to allow other actors to co-create value to the services and products they offer. In the context of business ecosystems, an organization with a platform economy strategy is able to enable the genesis of an ecosystem that "condense[s] 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" [25]. In the context of digital academic ecosystems, however, the virtues rather include increased information flow, more intense collaboration, and a growing shared pool of data, tools, and other resources conducive to science.

2.3 Visual network analytics

In the network analysis approach, a phenomenon is modeled as a set of nodes and edges connecting the nodes. With a background in social network analysis (SNA), network analysis is a general method that provides "a powerful way to study complex systems of interacting objects" [46]. Visualization is an organic part of SNA [26] and network analysis in general [1]. Visual network analysis is a valuable method for investigating social configurations and for interactively communicating their findings to others [12].

In this paper, visual network analytics refers to taking a visual analytics [16, 42] approach to network analysis (cf. [18]). Visual analytics stresses the process-centric, interactive, and iterative nature of using visualizations in supporting data-driven investigations [16, 22]. Visual network analytics allows the emergence of insights on the structure and dynamics of business and innovation ecosystems [1], social media platforms [41], and other networked phenomena through the process of sensemaking.

The well-known mantra of visual information retrieval iterates over three phases: "Overview first, zoom and filter, then details-on-demand" [39]. In the context of visual network analytics, however, users may prefer to follow the process of "start with what you know, then grow" [15]. 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 [16].

Visual analytics is a particularly suitable approach for exploring new phenomena with a data-driven approach. 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 [34]. Multiple perspectives on ecosystem structure and the structural positions of individual actors can be created and exchanged during the investigative process. With subsequent automation of data updates and tracking analyses, the assumptions and contingencies underlying decisions can be monitored for changes that would affect evidence-based conclusions, policy, and program directions.

Visual network analytics is an equally important method in linked science [20], bibliometrics, and its science-technology-innovation counterpart scientometrics [24] as well as science of science [37], the data-enabled approach to investigating the structure and dynamics of scientific processes. In this experiment, we investigate the ways individual researchers and university actors prefer to utilize visual network analytics in exploring the collaboration patterns between authors of scientific literature.

3. METHODOLOGY

This paper falls under the domain of IS research. IS researchers seek to develop a means to create value out of information for people, teams, and organizations [27]. Specifically, an ADR approach is taken in this paper. ADR is a recent variant of design science research [17, 29], the science of the artificial [40], where the focus in knowledge creation is on the design of human-made artifacts. Holistic IS research is interested in three interconnected dimensions or types of artifacts: social artifacts, information artifacts, and technology artifacts [23, 43].

On the epistemological and ontological level, this paper subscribes to a critical realist worldview that, according to Carlsson [7], provides a workable philosophical basis for design science research. Moreover, Bygstad and Munkvold [6] 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 to clarify that "although a real world exists, our knowledge of it is socially constructed and fallible" [6].

In more pragmatic terms, Bygstad and Munkvold [6] 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. Further, they refer to Sayer [36] 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 3) the underlying structure of a phenomenon under investigation.

This paper reports the very first finding of an experiment that will continue over several rounds of iteration. The overall objective of the experiment is to design tools and supporting processes for utilizing visual network analytics as a means for the researcher and other university actors to investigate the social structures latent in bibliographic data available in current research information systems. In this experiment, we conduct the first BIE cycle [38].

We claim that ADR in general and the process of guided emergence [38] in particular are key to answering to the challenges of the epistemological stance of the socially constructed and fallible nature of the way we can create knowledge of the real world. In ADR, "the artifact emerges from interaction with the organizational context even when its initial design is guided by the researchers' intent [38]."

The experiment presented in this paper is based on the interactive version of the network visualization shown in Figure 1. The visualization was used as a prototype, an early version of the self-service application that could serve as a minimum viable product (MVP) [33] that would allow collecting data on the ways researchers prefer to explore bibliographic data.

Four interviews were conducted during the BIE cycle. All the informants, one female and three male, work at the university for which the data for the visualization was collected. The one-hour interviews were conducted following the same general script. The interview started with a quick introduction to the dataset that was used to construct the network. Moreover, the network modeling principles that were applied in constructing the network were introduced. However, topics that the informants brought up during the interview were discussed to the extent that the delimitations of the experiment allowed.

In this experiment, the interviews were used as a low-barrier tool to identify the insights that the visual network analytics approach introduce as well as the requirements for data and the visualization process. In ADR terms, the interviews serve as a means to enable guided emergence.

4. RESULTS

In this section, we will cover the key insights from the experiment following the three dimensions of information system artifacts: social, information, and technology. The visualization in Figure 1 serves as the starting point of the investigation. The ultimate objective of this investigation is to develop a means for self-service visual network analytics in university collaboration. Therefore, we will first describe the social dimension, the key dimension in terms of the functional requirements. Second, we will discuss the information

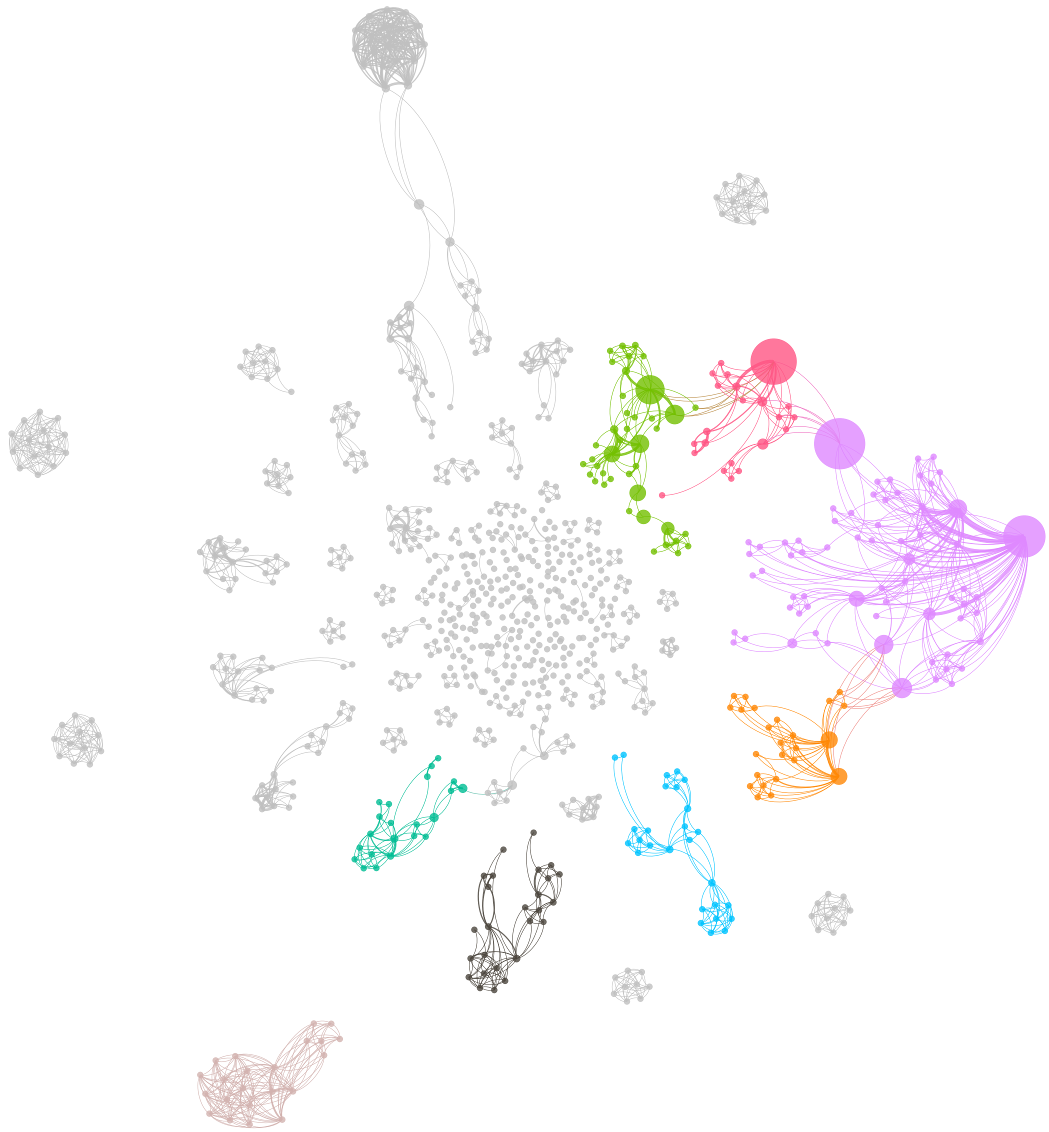

Figure 1: Co-authorship network used in the experiment. Nodes represent authors, and edges represent co-authorship affiliations. Nodes are sized according to their betweenness value and colored to indicate their membership in associative clusters. Edge weight shows the number of articles a pair of authors have co-authored.

artifacts. Last, we will describe the technology artifacts that will enable the actual implementation.

4.1 Social artifact

The four informants subscribed to the value-creation potential of visual network analytics in exploring university co-authorship. A set of key observations that are related to visual co-authorship network analytics as social artifact are described next.

Before moving to investigate the contents of the networks, the informants expressed an interest in learning the specifics of the data used to create the network visualization—what is and what is not included in the data. We will discuss the mechanisms of collecting the data as part of the section on technological artifacts.

First, most informants were interested in locating themselves in the visualization. This is in line with the “start with what you know, then grow” principle that Heer and boyd [15] first introduced. We suggest that the presented approach serves as a measurement tool that gives researchers the means to quantify both their work and the larger context in which it takes place, in this experiment the university.

Second, the informants noted the value of the visualization in revealing latent social structures in university co-authorship. They were interested in investigating what defines the existing structure: organizational structure, existing or future research groups, shared research interests, or something else. One informant gave an example of a related actionable insight that the network view provides: A shared connection between the informant and a researcher whose work the informant is interested in can be used as a reason to approach the researcher and make an introduction.

Third, the network view inspired the informants to explore and to benchmark the ways other researchers operate and to investigate the impact of those ways of operation. Several potential questions were raised. Which kind of strategy leads into a large number of publications or high degrees of research impact, measured, for example, as the number of citations? Is the number of co-authors important? Does collaboration with international researchers make a difference?

Fourth, the network approach was seen as a valuable means to measure research activity at universities. Some informants suggested that measuring researcher performance on the basis of the amount of publications does not encourage collaboration between researchers representing different departments or other institutional units. If actors’ position in bridging the co-authorship or more general collaboration network was seen favorable, they might be more inclined to seek ways to collaborate. Existing research, indeed, does suggest that bridging structural holes is a good strategy for increasing creativity [5].

4.2 Information artifact

Two sets of information artifacts are at the core of university co-authorship network analytics. First, article metadata was the axiomatic information artifact used in this experiment. Second, collections of articles were compiled with the help of the search functionality of the bibliographic repository.

In the experiment reported in this paper, several issues were raised when processing article metadata. First, despite the fact that the CRIS in use has unique profiles for article authors, we noted that the author names exported from the repository are not unique. Second, organizational information on the authors, including their department and faculty, is not available. It is in fact not possible to differentiate those with affiliation to the university under investigation from those co-authors external to the university, national or international. Third, there is a cap of 1000 articles when exporting the articles from the CRIS. Moreover, the license for using the bibliographic data in applications is not given. All of these issues came up during the interviews with the informants.

4.3 Technology artifact

The technology artifact in this experiment was a quick-and-dirty implementation of data collection and transformation batch script for getting bibliographic data into network format. The batch script was implemented in Python using Pandas[3] for processing tabular data and NetworkX [4] for constructing the network representation. The process roughly follows the ostinato model, a process model for data-driven visual network analytics [18]; however, due to the quick-and-dirty nature, the implementation as a whole is lightweight and indeed a simple minimum viable product.

The interactive network exploration platform Gephi [2] was used to conduct the sensemaking process. To move from quick-and-dirty to full self-service mode, we recommend using a Web-based visualization tool. A simple example of such a tool is GEXF.js that, however, insists on implementing the network layout algorithm as part of the pre-processing pipeline. Ecoxight is an example of an advanced Web-based tool that can be used to explore network representations.

For a relatively lightweight and yet novel approach for utilizing visual network analytics to make sense of organizational bibliographic repositories, we propose exploring the options to implement a Web-based visualization service that can be fired up in a context-sensitive manner [35, 19]. We will next briefly describe such an implementation. In order to select the articles to be analyzed, the user starts in a regular manner with the search service of the repository. A preferred set of articles is filtered in with a keyword search as well as using the advanced search options, including filters using publication data, the availability of full-text publications, and other features available in article metadata. After a set of articles of interest to the user is selected and appear as search results, the user clicks a bookmarklet, that is, an intelligent bookmark capable of collecting contextual data, that launches the visualization. Behind the scenes, the visualization tool connects with the bibliographic repository to collect the data in a machine-processable format or, alternatively, performs a set of crawling and scraping steps required to collect the data (cf. [18]). A visualization of the co-authorship network opens in a browser and, importantly, the user is able to download the data in formats that self-service network analytics tools, including NodeXL [14] and Gephi, are able to import.

Moreover and importantly, in order to conduct self-service analytics in a straightforward manner, several requirements must be met. Accessing sets of articles computationally is imperative. Currently, the bibliographic repository supports exporting a maximum of 1000 articles in a limited number

[3]Python Data Analysis Library, see http://pandas.pydata.org/

[4]NetworkX, see https://networkx.github.io/

of formats. For computational analytics, RIS and BibTeX are very likely the most suitable formats. From software development viewpoint, we next give a brief description of a more suitable setup for article access: performing queries and fetching article metadata through a well-documented API following the REST principles. Article metadata representations should be available in JSON and XML.

5. DISCUSSION AND FUTURE WORK

Gregor and Hevner [13] suggest that "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" a may be included in the discussion section of a scientific publication. We will follow this instruction in this section and for reasons of brevity and tractability concentrate on design principle extraction. The following is the first draft of a set of design principles, which describes the development of university article metadata repositories in a way that supports the development of visual network analytics tools in a self-service mode, that is, without the need to contact, negotiate, or in any way work with university representatives in order to create value for metadata:

1. Release the bibliographical data with a specific open data license. Creative Commons 4.0 is the recommended to be used, for example, in Finland.[5]
2. Provide a REST interface for accessing metadata on individual articles and authors.
3. Provide a REST interface for the search. Support a rich set of search parameters for selecting sets of data.
4. Provide complete data on articles and other entities in JSON and XML in addition to traditional bibliographic formats including BibTeX and RIS.
5. Provide unique identifiers for articles, authors, and organizational and other entities to ensure referential integrity.
6. Apply linked data practices to support traversing the metadata. Unique identifiers and entity-specific access through systematic URI schema are key here.
7. Do not limit the number of articles that can be fetched from the repository.
8. To enable the launch of external, third-party visual analytics tools, equip search result representations with machine-readable means to access results data in a computational manner.

In order to ensure that the access mechanism for bibliographic data allows for computation access in self-service mode, work in collaboration with early adopters. You may consider starting with a minimum viable API and implement new features following user feedback.

6. CONCLUDING REMARKS

In this paper, we reported the first round of action design research investigating how researchers and other academic actors prefer to explore bibliographical data available in current research information systems.

In order to add to the value-creation potential from the bibliographical and other science, technology, and innovation data collected into current research information systems, we developed an early prototype of a visual network analytics tool for exploring co-authorship networks on the basis of bibliographical data and engaged with four experienced researchers to investigate their preferences as well as the requirements that stem from both the implementation of the data-processing pipeline and the requirements the researchers expressed.

On the basis of the first experiment, we claim that value-creation potential exists in the presented approach. Researchers find visual network analytics useful for finding new potential collaborators, observing collaboration patterns beyond the institutional structure latent in the data as well as for analyzing and, perhaps, measuring universities at the ecosystem level.

In order to enable visual network analytics, however, the existing CRIS functionality has to be developed further. Imperatively, universities must adopt linked open data practices in order to enable the use, reuse and redistribution of bibliographical data. This is the minimal starting point. Moreover, in order to enable the development of computational tools that assist researchers and other university actors in gaining insights on, for example, collaboration patterns within universities, contemporary Web development practices, including the RESTful approach, must be applied in serving bibliographic data.

Knowing that the role of data and analytics is growing in importance in most streams of academic research including computational social science and digital humanities, we have no doubt that there is a growing demand in bibliographical data access for conducting computational literature reviews, analyzing academic ecosystems for new knowledge and potential collaborators, and applying the practices of bibliometrics and linked open science. Therefore, we conclude with the strongest possible recommendation for university policy makers to step into the open science sphere. Providing open linked CRIS data is an intuitive first step.

7. ACKNOWLEDGMENTS

The work reported in this paper is part of the Academy of Finland-sponsored project Enhancing Knowledge Work and Co-creation by Big Social Data Weak Ties in Web-based Online Services (COBWEB, grant number 295894). The author extends his gratitude to the informants that dedicated their time to support the experiment and provided invaluable insights into supporting research activities with self-service visual network analytics. Overleaf was used to author this document in LaTeX.

8. REFERENCES

[1] R. C. Basole. Visualization of interfirm relations in a converging mobile ecosystem. *Journal of Information Technology*, 24(2):144–159, June 2009.

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

[3] T. Berners-Lee. Linked Data - Design Issues, July 2006.

[5]Recommendation for open data licensing (in Finnish): http://www.jhs-suositukset.fi/suomi/jhs189

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

[5] R. S. Burt. Structural Holes and Good Ideas. *American Journal of Sociology*, 110(2):349–399, 2004.

[6] B. Bygstad and B. Munkvold. In Search of Mechanisms. Conducting a Critical Realist Data Analysis. In *Proceedings of International Conference on Information Systems*, 2011.

[7] S. A. Carlsson. Design Science Research in Information Systems: A Critical Realist Approach. In A. Hevner and S. Chatterjee, editors, *Design Research in Information Systems*, number 22 in Integrated Series in Information Systems, pages 209–233. Springer US, 2010. DOI: 10.1007/978-1-4419-5653-8_15.

[8] D. Dietrich, J. Gray, T. McNamara, A. Poikola, R. Pollock, J. Tait, and T. Zijlstra. *Open Data Handbook*. Open Knowledge International, 2009.

[9] P. C. Evans and R. C. Basole. Revealing the API Ecosystem and Enterprise Strategy via Visual Analytics. *Commun. ACM*, 59(2):26–28, Jan. 2016.

[10] R. Fielding. *Architectural Styles and the Design of Network-based Software Architectures*. Doctoral Dissertation, University of California, Irvine, 2000.

[11] R. T. Fielding and R. N. Taylor. Principled Design of the Modern Web Architecture. *ACM Trans. Internet Technol.*, 2(2):115–150, May 2002.

[12] L. C. Freeman. Visualizing Social Networks. *Journal of Social Structure*, 1(1):[np], 2000.

[13] S. Gregor and A. R. Hevner. Positioning and Presenting Design Science Research for Maximum Impact. *MIS Quarterly*, 37(2):337–A6, June 2013.

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

[15] J. Heer and d. boyd. Vizster: Visualizing Online Social Networks. In *Proceedings of the 2005 IEEE Symposium on Information Visualization (INFOVIS'05), October 23-25, 2005, Minneapolis, MN, USA*, pages 33–40. IEEE Computer Society, 2005.

[16] J. Heer and B. Shneiderman. Interactive dynamics for visual analysis. *Commun. ACM*, 55(4):45–54, 2012.

[17] A. R. Hevner, S. T. March, J. Park, and S. Ram. Design science in information systems research. *MIS Q.*, 28(1):75–105, Mar. 2004.

[18] J. Huhtamäki, M. G. Russell, N. Rubens, and K. Still. Ostinato: The Exploration-Automation Cycle of User-Centric, Process-Automated Data-Driven Visual Network Analytics. In S. A. Matei, M. G. Russell, and E. Bertino, editors, *Transparency in Social Media*, Computational Social Sciences, pages 197–222. Springer International Publishing, 2015. DOI: 10.1007/978-3-319-18552-1_11.

[19] J. Huhtamäki, J. Salonen, J. Marttila, and O. Nykänen. Context-Driven Social Network Visualisation: Case Wiki Co-Creation. In *Proceedings of the Knowledge Federation 2010: Self-Organizing Collective Mind, October 3-6, 2010, Dubrovnik, Croatia*.

[20] T. Kauppinen, A. Baglatzi, and C. Keßler. Linked science: interconnecting scientific assets. In *Data Intensive Science*. CRC Press, USA, 2012.

[21] T. Kauppinen and G. M. d. Espindola. Linked Open Science-Communicating, Sharing and Evaluating Data, Methods and Results for Executable Papers. *Procedia Computer Science*, 4:726–731, Jan. 2011.

[22] D. A. Keim, F. Mansmann, and J. Thomas. Visual Analytics: How Much Visualization and How Much Analytics? *SIGKDD Explor. Newsl.*, 11(2):5–8, May 2010.

[23] A. S. Lee, M. Thomas, and R. L. Baskerville. Going back to basics in design science: from the information technology artifact to the information systems artifact. *Information Systems Journal*, 25(1):5–21, Jan. 2015.

[24] P. Mayr and A. Scharnhorst. Scientometrics and information retrieval: weak-links revitalized. *Scientometrics*, 102(3):2193–2199, Nov. 2014.

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

[26] J. L. Moreno. *Who Shall Survive?: Foundations of Sociometry, Group Psychotherapy and Sociodrama*. Beacon House Inc., Beacon, NY, USA, 1953.

[27] J. F. Nunamaker, Jr. and R. O. Briggs. Toward a Broader Vision for Information Systems. *ACM Trans. Manage. Inf. Syst.*, 2(4):20:1–20:12, Dec. 2011.

[28] T. O'Reilly. What is Web 2.0: Design Patterns and Business Models for the Next Generation of Software. *Communications & Strategies*, (1):17, Aug. 2007.

[29] K. Peffers, T. Tuunanen, M. Rothenberger, and S. Chatterjee. A Design Science Research Methodology for Information Systems Research. *Journal of Management Information Systems*, 24(3):45–77, Dec. 2007.

[30] R. D. Peng. Reproducible research and Biostatistics. *Biostatistics*, 10(3):405–408, July 2009.

[31] R. D. Peng. Reproducible Research in Computational Science. *Science*, 334(6060):1226–1227, 2011.

[32] L. Richardson, M. Amundsen, and S. Ruby. *RESTful Web APIs*. O'Reilly Media, 2013.

[33] E. Ries. *The Lean Startup*. Crown Business, USA, 2011.

[34] M. G. Russell, K. Still, J. Huhtamäki, C. Yu, and N. Rubens. Transforming Innovation Ecosystems through Shared Vision and Network Orchestration. In *Proceedings of Triple Helix IX International Conference: "Silicon Valley: Global Model or Unique Anomaly?", 11-14 July 2011, Stanford, California, USA*, 2011.

[35] J. Salonen and J. Huhtamäki. Launching Context-Aware Visualisations. In F. A. B. Colugnati, L. C. R. Lopes, and S. F. A. Barretto, editors, *Digital Ecosystems*, number 67 in Lecture Notes of the Institute for Computer Sciences, Social Informatics and Telecommunications Engineering, pages 146–160. Springer Berlin Heidelberg, Jan. 2010.

[36] A. Sayer. *Method in Social Science: A Realist Approach*. Routledge, 2010.

[37] F. Schweitzer. Scientific networks and success in science. *EPJ Data Science*, 3(1), Dec. 2014.

[38] M. Sein, O. Henfridsson, S. Purao, M. Rossi, and R. Lindgren. Action Design Research. *MIS Quarterly*,

35(1):37 – 56, 2011.

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

[40] H. A. Simon. *The Sciences of the Artificial*. The MIT Press, Cambridge, Mass, 3rd edition, 1996.

[41] M. A. Smith, I. Himelboim, L. Rainie, and B. Shneiderman. The Structures of Twitter Crowds and Conversations. In S. A. Matei, M. G. Russell, and E. Bertino, editors, *Transparency in Social Media*, Computational Social Sciences, pages 67–108. Springer International Publishing, 2015. DOI: 10.1007/978-3-319-18552-1_5.

[42] J. Thomas and K. Cook. A visual analytics agenda. *IEEE Computer Graphics and Applications*, 26(1):10–13, 2006.

[43] T. Vartiainen and T. Tuunanen. Value Co-Creation and Co-Destruction in an IS Artifact: Contradictions of Geocaching. In *2016 49th Hawaii International Conference on System Sciences (HICSS)*, pages 1266–1275, 2016.

[44] S. Vinoski. Serendipitous Reuse. *Internet Computing*, (January/February 2008):84–87, 2008.

[45] M. D. Wilkinson, M. Dumontier, I. J. Aalbersberg, G. Appleton, M. Axton, A. Baak, N. Blomberg, J.-W. Boiten, L. B. da Silva Santos, P. E. Bourne, J. Bouwman, A. J. Brookes, T. Clark, M. Crosas, I. Dillo, O. Dumon, S. Edmunds, C. T. Evelo, R. Finkers, A. Gonzalez-Beltran, A. J. Gray, P. Groth, C. Goble, J. S. Grethe, J. Heringa, P. A. 't Hoen, R. Hooft, T. Kuhn, R. Kok, J. Kok, S. J. Lusher, M. E. Martone, A. Mons, A. L. Packer, B. Persson, P. Rocca-Serra, M. Roos, R. van Schaik, S.-A. Sansone, E. Schultes, T. Sengstag, T. Slater, G. Strawn, M. A. Swertz, M. Thompson, J. van der Lei, E. van Mulligen, J. Velterop, A. Waagmeester, P. Wittenburg, K. Wolstencroft, J. Zhao, and B. Mons. The FAIR Guiding Principles for scientific data management and stewardship. *Scientific Data*, 3, 2016.

[46] J. Yang and J. Leskovec. Overlapping Communities Explain Core - Periphery Organization of Networks. *Proceedings of the IEEE*, 102(12):1892–1902, 2014.