

KOLMEN POHJOISMAALAISEN KIRJASTO- JA IN-  
FORMAATIOTIETEIDEN YKSIKÖN BIBLIOMETRI-  
NEN ANALYYSI

Janne Karin-Oka

Tampereen yliopisto

Viestintätieteiden tiedekunta

Informaatiotutkimus ja interak-  
tiivinen media

Pro gradu -tutkielma

Huhtikuu 2018

TAMPEREEN YLIOPISTO, Viestintätieteiden tiedekunta

Informaatiotutkimus ja interaktiivinen media

KARIN-OKA, JANNE: Kolmen pohjoismaalaisen kirjasto- ja informaatiotieteiden yksikön bibliometrinen analyysi

Pro gradu -tutkielma, 61 s.

Huhtikuu 2018

---

Pro gradu -tutkielmassa perehdyn kolmen pohjoismaalaisen yliopiston Boråsin, Kööpenhaminan ja Tampereen yliopiston kirjasto- ja informaatiotieteiden julkaisuaktiivisuuteen vuosien 2005–2015 välillä ja tarkastelen miten näinä vuosina julkaistut teokset ovat keränneet viittauksia. Tutkin myös miten viitatuimmat tutkijat ovat jakautuneet kolmen yliopiston välille ja miten julkaisut ja viittaukset ovat jakautuneet eri julkaisukanaviin.

Kolmen yliopiston julkaisuaktiivisuus valitulla aikajaksolla on kehitykseltään erilainen. Kööpenhaminan yliopiston aktiivisuus on ollut koko aikajakson laskussa, kun taas Tampereen yliopistolla aktiivisuus on kokonaisuutena ollut laskusuhdanteinen, mutta kuitenkin välillä kokien pieniä nousuja. Boråsin yliopiston kirjasto- ja informaatiotieteiden julkaisuja tarkastelujakson alussa oli hyvin vähäisesti, mutta muutos tapahtui vuonna 2011, jolloin kasvua edelliseen vuoteen oli hurja edellisvuoden kahdeksasta julkaisusta 72 julkaisuun. Kööpenhaminan yliopisto oli selkeästi aktiivisin julkaisija ja Tampereen yliopisto seuraavaksi aktiivisin. Jos otetaan huomioon ainoastaan vuosien 2011–2015 julkaisut, niin Boråsin yliopisto nousee Tampereen yliopiston ohitse aktiivisuudessa.

Julkaisuaktiivisuus yliopistoissa heijastaa viittausmääriä eli, enemmän julkaissut yliopisto on kerännyt enemmän viittauksia. Viittaukset suhteutettuna kirjoittajien määrään, Tampereen yliopisto oli kerännyt enemmän viittauksia. Tarkoittaen, että pienempi joukko kirjoittajia oli suhteessa kerännyt julkaisuillaan enemmän viittauksia.

Koko aineiston viitatuin tutkija oli Tampereen yliopiston Reijo Savolainen ja seuraavaksi viitatuin oli Kööpenhaminan yliopiston Birger Hjørland. Viitatuimmat kymmenen tutkijaa koko aineistossa jakautuivat tasan Tampereen yliopiston ja Kööpenhaminan yliopiston kesken. Kun huomioitiin vain vuosien 2005–2015 julkaisut myös Boråsin yliopiston tutkijat pääsivät kärkikymmenikköön.

Viittaustietokantojen kattavuudessa tämän aineiston perusteella on valtava ero. Scopus kattoi kolmen yliopiston vuosien 2005–2015 julkaisuista ainoastaan 21,4 % kun Google Scholar sen sijaan kattoi 86,1 %. Erot olivat pienemmät tarkasteltaessa eri julkaisutyyppejä. Scopus oli melko kattava konferenssijulkaisuissa jääden kuitenkin vain 34,7 %.

Avainsanat: Viittausanalyysi, Google Scholar, Scopus, Bibliometriset menetelmät

# Sisällysluettelo

1	JOHDANTO.....	6
2	BIBLIOMETRINEN TUTKIMUS.....	8
2.1	Bibliometrisen tutkimuksen käsitteitä ja menetelmiä.....	8
2.1.1	Bibliometrisen tutkimuksen tarkasteltavia kohteita .....	8
2.1.2	Indikaattoreita .....	9
2.2	Bibliometrisen tutkimuksen menetelmiä .....	12
2.2.1	Viittausanalyysi .....	12
2.2.2	Matemaattisia malleja.....	14
2.2.3	Muita bibliometrisia menetelmiä.....	16
3	BIBLIOMETRIKKA OSANA TIEDEKENTTÄÄ JA TIETEEN ARVIOINTIA	18
3.1	Bibliometrisen tutkimuksen asema tieteessä .....	18
3.1.1	Bibliometriikkaan kohdistunut kritiikki .....	21
3.1.2	Viittaus- ja julkaisukäytännöt.....	21
3.2	Bibliometrinen tutkimus osana tieteen arviointia ja tiedepolitiikkaa .....	22
3.2.1	Tieteen arviointi suomessa .....	23
4	TUTKIMUSKYSYMYKSET .....	26
5	AINEISTO JA VIITTAUSTIETOKANNAT .....	27
5.1	Viittaustietokannat.....	29
5.2	Viittausaineistonkeruu ja käsittely .....	33
6	TULOKSET .....	36
6.1	Julkaisutoiminta vuosien 2005 ja 2015 välillä .....	36
6.2	Scopusin ja Google Scholarin kattavuus kirjasto- ja informaatiotieteissä... ..	38
6.3	Viittausten kerääntyminen vuosien 2005-2015 välillä .....	42
6.4	Viitatuimpien tutkijoiden jakautuminen yliopistojen välille.....	48
6.5	Viittausten ja julkaisujen jakautuminen julkaisukanavien välillä .....	53
7	YHTEENVETO .....	55
	LÄHTEET .....	59
	LIITTEET	

# 1 JOHDANTO

Tutkin tässä pro gradu -tutkielmassa kolmea pohjoismaalaista kirjasto- ja informaatiotieteisiin keskittynyttä yliopistoyksikköä. Tutkimus on bibliometrinen analyysi näiden kolmen yliopistoyksikön kirjasto- ja informaatiotieteiden julkaisujen perusteella. Vertailtavat yliopistot valitsin pohjoismaisten yliopistojen kirjasto- informaatiotieteiden tutkijakoulutuksen verkoston Nordic Research School in Library and Information Sciencen yliopistoista. Halusin erityisesti vertailla miten Tampereen yliopiston entisen informaatiotieteiden yksikön kirjasto- ja informaatiotieteiden julkaisut suhtautuvat kahteen muuhun kyseiseen tieteenalaan keskittyneeseen yliopistoon. Vertailuun halusin ottaa Tampereen kanssa samaa kokoluokkaa olevan Boråsin yliopiston The Swedish School of Library and Information Sciencen sekä verkoston suurimman Kööpenhaminan yliopiston Royal School of Library and Information Sciencen. Koska halusin keskittyä nimenomaan kirjasto- ja informaatiotieteen julkaisuihin, rajasin aineistosta pelitutkimuksen julkaisut pois, jotka Tampereen yliopistolla kuuluvat informaatiotieteiden yksikön julkaisuihin.

Bibliometrinen analyysi kertoo yliopistojen julkaisuaktiivisuudesta sekä siitä, miten yliopistojen julkaisut on otettu vastaan tieteenalalla. Viittaukset kertovat osaltaan siitä, miten julkaisulla on ollut merkitystä tieteenalalla. Julkaisumääriä ja julkaisukanavia käytetään myös yliopistojen rahoituksen mittareina Suomessa. Yliopistojen sisäisessä taloudessa julkaisuaktiivisuus sekä bibliometriset mittarit ovat myös tarkastelun kohteena, kun pohditaan yksiköiden rahoitusta.

Toinen tämän tutkimuksen pääkysymys on miten ilmainen Google Scholar viittaustietokanta vertautuu kaupallisen toimijan Scopusin tietokantaan. Tarkastelen näiden tietokantojen osalta kattavuutta kirjasto- ja informaatiotieteissä keräämälläni kolmen yliopiston julkaisuaineistolla.

Luvussa kaksi kerron mitä bibliometrinen tutkimus on ja mitä tarkasteltavia käsitteitä, tarkastelun kohteita ja mittareita bibliometrisessä tutkimuksessa seurataan. Samassa luvussa kerron menetelmistä näiden kohteiden ja mittarien tarkasteluun. Luvussa kolme perehdyn siihen miksi bibliometrinen tutkimus tehdään ja mikä on sen merkitys tiedepolitiikassa, yliopistojen vertailussa ja tieteenalojen tutkimuksessa. Luvuissa neljä ja viisi esittelen tutkimuskysymykseni ja aineistoni. Erittelen tarkemmin luvussa viisi, miten aineisto on kerätty ja miten sitä on käsitelty jatkoanalyysia varten. Luvussa kuusi esittelen

tulokset julkaisuaktiivisuudesta ja viittausteitokantojen eroista sekä siitä, miten yliopistot ja tutkijat ovat keränneet viittauksia vuosien 2005 ja 2015 välillä julkaistuille teoksille.

## **2 BIBLIOMETRINEN TUTKIMUS**

Bibliometrinen tutkimus voidaan tehdä hyvin monella tavalla. Pohjimmiltaan bibliometrisessä analyysissä on kyse määrällisestä tutkimuksesta, jossa lasketaan julkaisujen saamia viittauksia. Käsittelen luvussa 2.1 bibliometrisen analyysin indikaattoreita ja määreitä sekä luvussa 2.2 avaan bibliometrisiä analyysimenetelmiä.

### **2.1 Bibliometrisen tutkimuksen käsitteitä ja menetelmiä**

#### **2.1.1 Bibliometrisen tutkimuksen tarkasteltavia kohteita**

Julkaisu, lähde ja viite käsittelevät kaikki samaa keskeistä määrettä tieteellisessä keskustelussa ja tieteen mittaamisessa. Kyse on vain siitä, mistä suunnasta kulloinkin asiaa tarkastellaan. Julkaisu voi olla esimerkiksi tieteellinen artikkeli, kirja, tutkimusraportti, dokumentti tai patentti. Ei myöskään ole merkitystä sillä, onko kyseessä fyysinen julkaisu vai onko julkaisu ilmestynyt sähköisesti. Laajasti käsitettynä julkaisu voi olla myös pienemmälle yleisölle saataville tarjottu teos. Lähde ja viite ovat julkaisussa olevia osia. Kirjoittaja viittaa aiemmin julkaistuihin teoksiin omassa teoksessaan samalla linkittäen oman tutkimuksensa osaksi aiempaa tietämystä. Tämä on siis viittaus kyseisen aiemman julkaisun näkökulmasta ja kirjoittajan lisäämä lähde omaan teokseensa. Julkaisussa on siis lähteitä, kun taas julkaisu voi saada viittauksia. Suomeksi näiden molempien kvantitatiivisesta tarkastelusta keskustellaan viiteanalyysinä, joka haarautuu useisiin erilaisiin viittauksia ja lähteitä käsitteleviin tutkimuksen muotoihin. (Kärki, Kortelainen, 1996, 8–14.)

Bibliometrinen tutkimus on syntynyt viittauksien ja lähteiden ympärille. Toisena bibliometrisen tutkimuksen tarkasteltavana kohteena voi toimia julkaisun kirjoittaja tai kirjoittajat. Kirjoittajia voidaan tarkastella heidän julkaisujensa keräämien viittauksien kautta esimerkiksi h-indeksin avulla. Kirjoittajia voidaan tutkia myös tutkimalla heidän muodostamia verkostoja yhteisviittausanalyysin avulla tai tutkimalla kahden ryhmän välisiä viittauksia.

Tieteenaloilla on hyvin erilaisia tapoja viitata aiempiin tutkimuksiin ja toisaalta missä, kuinka usein ja millä kielellä tutkimusta julkaistaan, vaihtelee tieteenaloittain. Tämän lisäksi tieteenalojen välillä on suuria eroja ja erilaisia käytäntöjä siinä, miten julkaisulle



määritellään kirjoittaja. Monesti yhden yksittäisen tutkijan tarkastelu ei ole kovin mielekästä vaan tutkimus kohdennetaan tutkimusyhteisöihin kuten yliopistoihin, tutkimusyksiköihin ja kokonaiseen maahan. (Kärki, Kortelainen, 1996, 11–12; Forsman, 2016, 30–32).

Kolmantena tarkasteltavana kohteena toimii julkaisukanavat. Julkaisukanavia ovat tieteelliset aikakauslehdet, avoimen tieteen julkaisukanavat kuten Plos ja ArXiv ja konferenssit. Julkaisukanavia tutkitaan, muiden menetelmien ohella, pohjimmiltaan myös viittausten ja lähteiden kautta. Analyysia lehdistä voidaan tehdä myös käyttämällä matemaattisia malleja. Tällaisia matemaattisia malleja ovat muun muassa Bradfordin laki, joka kuvaa julkaisukanavien läheisyyttä tieteenalan keskeisimpiin julkaisijoihin (Kärki, Kortelainen, 1996, 17 ja 21). Matemaattisia malleja käsitellään tarkemmin luvussa 2.2.2. Eugen Garfieldin kehittämä viittauskerroin, impact factor, luotiin alun perin mittaamaan tieteellisten lehtien vaikuttavuutta tai merkittävyyttä omalla alallaan, mutta nykyään sitä käytetään myös muihin tarkoituksiin. Bibliometrisessä tutkimuksessa voidaan lehtien vaikutavuuskertoimia käyttämällä mitata muun muassa julkaisun tai kirjoittajan suhteellista vaikuttavuutta. (Kärki, Kortelainen, 1996, 25–26; Garfield, 2006)

Julkaisujen sisällään pitämien lähteiden, julkaisujen keräämien viittausten sekä kirjoittajien ja lehtien ohella bibliometrisen tutkimuksen kohteena voivat olla ideoiden leviäminen, tutkijoiden sosiaaliset verkostot tai esimerkiksi tieteenalan tilaa kuvaavat mallinnukset. Näissä kaikissa tarkastelun kohteissa keskeisenä osana ovat julkaisut ja niiden lähteet sekä niiden keräämät viittaukset. Nämä eivät välttämättä ole itse tarkastelun kohteita, mutta monesti ne toimivat mitattavina määreinä tarkastelun kohteille tai niistä johdetuilla indikaattoreilla tutkitaan kohdetta.

### **2.1.2 Indikaattoreita**

Tässä luvussa esittelen yleisesti käytettyjä bibliometrisia indikaattoreita. Indikaattorit ovat julkaisujen viittausmääristä laskettavia määreitä, joilla pyritään mittaamaan julkaisukanavien tai kirjoittajien tuotteliaisuutta tai vaikuttavuutta. Indikaattoreita saatetaan käyttää myös laadun mittaamiseen, vaikeivat ne siihen ole tarkoitettukaan.

## Viittauskerroin ja muut julkaisukanavien indikaattorit

Lehtien laatua on pyritty kvantitatiivisesti mittaamaan esimerkiksi Eugene Garfieldin alun perin kehittämän viittauskertoimen (impact factor), viittauskertoimen, avulla. Garfield mainitsi viittauskertoimen ensimmäisen kerran 1955 artikkelissaan Science-lehdessä. Artikkelin käsitteli tieteen viittausindeksiä, josta voisi löytää tieteenalan tärkeimmät ja keskeisimmät julkaisut ja niiden merkityksen alalla. Garfield ehdottikin artikkelissaan, että viittauskerroin on parempi määre kuin absoluuttinen luku kertomaan lehden merkityksestä, sillä se poistaa epäsuhtaan pienten ja suurten julkaisumäärien lehdiltä. (Garfield, 2006.) Nykyisin Clarivate Analytics julkaisee julkaisukanavien tärkeimpiä mittareita. Alun perin Thomson&Reutersin alaisuudessa toiminut Institute of Scientific Information määritteli viittauskertoimen kyseisenä vuonna lehden kahden edellisen vuoden julkaisujen saamien viittausten keskiarvoksi (The Clarivate Analytics Impact Factor). Vuosittain ilmestyvässä Journal Citation Reportissa julkaistaan nykyään myös viiden vuoden viittauskerroin, joka lasketaan muutoin samalla tavalla, mutta aikayksiköksi otetaan viisi vuotta.

Clarivate analyticsin kilpailija Elsevier on alkanut julkaista vuodesta 2016 myös omaa viittauskerrointa Scopus CiteScorea. Tämä muodostuu julkaisun edellisen kolmen vuoden viittausten keskiarvosta verrattuna Clarivate Analyticsin Impact Factoryn kahden vuoden keskiarvoon. Suurin ero näiden kahden viittausindeksin välillä on viittauksia kerrottävät artikkelityypit. Journal Citation Reportissa viitattavissa oleviksi julkaisuihin lasketaan ainoastaan artikkelit ja arvostelut, kun taas CiteScoreen otetaan mukaan näiden lisäksi myös kaikki muut julkaisukanavissa julkaistavat tekstit. (Scopus: Journal metrics; InCites Journal Citation Reports Help.)

Journal Citation Reportin muita julkaisukanavan mittareita on viittausten välittömyyttä kuvaava immediacy index. Välittömyyttä tarkastellaan katsomalla julkaisukanavan saamat viittaukset vuonna x ja jakamalla sen samana vuonna viitattavien artikkelien määrällä. (InCites Journal Citation Reports Help). Välittömyys kuvaa sitä kuinka nopeasti keskimäärin kyseisessä julkaisukanavassa julkaistu artikkeli alkaa saada viittauksia

## H-indeksi ja kirjoittajien indikaattorit

H-indeksi on Jorge E. Hirschin (2005) kehittämä tutkijan julkaisuaktiivisuutta mittaava indeksi. Indeksillä määritellään seuraavasti ”tutkijan h-indeksi on kokonaisluku, joka saa arvon  $h$ , mikäli kaikista hänen  $N$  julkaisustaan  $h$  julkaisuun on viitattu muissa julkaisuissa  $h$  kertaa tai useammin ja loppuihin ( $N-h$ ) julkaisuun  $h$  kertaa tai vähemmän” (Hirsch 2005, käänös Forsman, 2016, 59). H-indeksi voidaan laskea myös tutkimusryhmille tai joukolle tutkijoita heidän henkilökohtaisen julkaisuhistorian pohjalta. Kyse ei ole h-indeksien summa, vaan samaan tapaan laskettu indeksiluku kuin yksittäiselle tutkijalle laskettu indeksiluku. H-indeksi on pätevä vain oman tieteenalan ja jopa tarkemmin vain oman erikoistumisalan sisällä vertailukelpoisena lukuna, sillä viittaus-, kirjoittajuus- ja julkaisukäytännöt vaihtelevat aloittain. Indeksillä ei myöskään ota huomioon julkaisun kirjoittajien määrää, josta johtuen se ei kuvasta kaikilla aloilla yksittäisen tutkijan vaikutusta tieteenalalle. Tutkijan ikä vaikuttaa myös positiivisesti h-indeksiin, sillä iäkkäämpi tutkija on ennättänyt keräämään viittauksia pidemmän aikaa kuin juuri uransa aloittava tutkija. H-indeksiä on myös kiiteltu sen yksinkertaisesta laskentakaavasta ja siitä, että se tuottaa vain yhden luvun, jolla voidaan arvioida tutkijan aktiivisuutta sekä vaikutusta alallaan. H-indeksi myös kiittää muiden mittareiden ongelmia kuten yksittäisten julkaisujen saamien tai ei saamien viittausten määrän tarkasteltaessa tutkijan julkaisujen saamien viittausmäärien keskiarvoa. (Hirsch, 2005; Hirsch, Buéla-Casal, 2014; Forsman, 2016, 59–61.)

Hirsch on ehdottanut korjauksia h-indeksin määritelmään ja tulkintaan. Tutkijan ikä voidaan ottaa huomioon niin sanotulla  $m$ -indeksillä, jossa  $h$ -indeksi jaetaan tutkijan tutkijajäällä, eli vuosilla ensimmäisestä julkaisusta. (Hirsch, Buéla-Casal, 2014.)

Hirsch on ehdottanut myös  $h_{bar}$ -indeksiä täydennyksenä  $h$ -indeksille.  $h_{bar}$ -indeksi ottaa huomioon yhteiskirjoittajuuden laskennassaan.  $h_{bar}$ -indeksi muodostuu tutkijan oman  $h$ -indeksin sekä hänen kanssaan yhdessä julkaisseiden tutkijoiden  $h$ -indeksien kautta.  $h_{bar}$ -indeksiin lasketaan mukaan kaikki sellaiset tutkijan omat julkaisut, jotka on julkaistu joko yksin tai sellaisten tutkijoiden kanssa, joiden oma  $h$ -indeksi on pienempi kuin tarkasteltavan tutkijan. Lisäksi tarkasteltavan tutkijan  $h_{bar}$ -indeksiin lasketaan julkaisut, jotka ovat osa yhteisjulkaisuun osallistuneiden tutkijoiden  $h$ -indeksiä. Esimerkiksi jos tarkasteltavan tutkijan  $h$ -indeksi on 20 ja hän on julkaissut yhdessä tutkijan kanssa, jonka oma  $h$ -indeksi on 40, lasketaan julkaisu osaksi tarkasteltavan tutkijan  $h_{bar}$ -indeksiä, jos

julkaisulla on 43 viittausta. Jos julkaisulla on 35 viittausta, sitä ei lasketa tarkasteltavan tutkijan hbar-indeksiin, koska se ei ole osa yhteisjulkaisun toisen tutkijan h-indeksiä. (Hirsch, 2010.)

Vaihtoehtona h-indeksille on myös ehdotettu kirjoittajan viittauskerrointa, author impact factoria. Kirjoittajan viittauskerroin voidaan laskea samankaltaisesti kuin lehden viittauskerroin. Kirjoittajan kahden vuoden viittauskerroin vuodelle x on kahden edellisen vuoden artikkelien keräämien viittausten keskiarvo. Kirjoittajan viittauskerroin mittaa paikkaa ajassa, kun taas h-indeksi mittaa tutkijan uraa kokonaisuutena. (Pan, Fortunato, 2014.) Myös verrattuna m-indeksiin, kirjoittajan viittauskerroin on vahvasti kirjoittajan tämän hetkiseen aktiivisuuteen sidottu ja ei ota huomioon tutkijan uran pituutta. Toisaalta, pitkän uran tehneellä tutkijalla on kenties paremmat mahdollisuudet kokemuksensa johdosta päästä arvostettuihin ja paljon luettuihin julkaisuihin ja täten saada enemmän viittauksia myös uusiin julkaisuihin.

## **2.2 Bibliometrisen tutkimuksen menetelmiä**

Bibliometristä tutkimusta on jäsennetty sen sovellusalueen ja toisaalta sen kohteen mukaan. Menetelmät tutkimukselle valitaan sen perusteella mitä halutaan tutkia. Suurin osa tutkimuksesta pohjautuu julkaisujen lähteisiin ja viittauksiin, vaikka itse tutkimuksen kohde olisikin esimerkiksi tutkijoiden välisten verkostojen tutkiminen tai uusien tieteenalojen välisten yhteyksien havaitseminen. Hyödyksi tutkimuksessa käytetään myös johdettuja bibliometrisia indikaattoreita kuten h-indeksejä tai viittauskertoimia. Käsittelen luvussa 2.2.1 viittausanalyysia, luvussa 2.2.2 yhteisviittausanalyysia ja luvussa 2.2.3 matemaattisia menetelmiä Lotkan lakia ja Bradfordin lakia. Nämä eivät ole suinkaan ainoita bibliometrisiä menetelmiä vaan menetelmiä on useita muitakin kuten lähdeanalyysi, yhteissana-analyysi ja verkostanalyysi (Kärki, Kortelainen, 1996, 14–24; Forsman, 2016, 47–69.)

### **2.2.1 Viittausanalyysi**

Viittausanalyysi on keskeisimpiä bibliometrisen tutkimuksen menetelmiä. Viittausanalyysissa tarkastellaan miten julkaisut, tutkijat, instituutiot tai jopa maat kerryttävät viittauksia. Yhtä hyvin tutkimus voi arvioida julkaisukanavan, julkaisun tai kirjoittajan saa-

maa huomiota tietyn aikajakson aikana tai kokonaisuudessaan vuosien saatossa. (Forsman, 2016, 54–55; Kärki, Kortelainen, 1996, 16–17). Yleisesti voidaan todeta, että viittausanalyysillä pyritään tutkimaan kvantitatiivisesti julkaisukanavan, julkaisun tai tutkijan vaikutusta, huomiota, tuotteliaisuutta ja joskus myös laatua omalla tieteenalallaan. Yksi viittausanalyysin suosiota selittävä tekijä on sen yksinkertaisuus ja helppo ymmärrettävyys. Lasketaan viittaukset ja korkeampi arvo on parempi. Bibliometriikassa huomiota mitataan julkaisujen viittauksista ja toisinaan huomiota käytetään myös suorana mittarina laadulle. Viittauksista laskettavat indikaattorit eivät kuitenkaan ole täysin suoraviivaisia ja tieteenaloittain verrattavissa. Etenkään laadun mittarina niitä ei yksistään voida käyttää. (Andrés, 2009, 56–57.) Viittaajat eivät välttämättä viittaa kaikkiin julkaisuihin, joista he ovat saaneet vaikutteita ja toisaalta voidaan myös viitata julkaisuihin negatiivisessa mielessä tai siten, ettei niitä ole täysin luettu. Osa viittauksista tehdään myös tieteellisen konvention vuoksi, eli viitataan julkaisuun, koska siihen tulee viitata yleisen tavan vuoksi. (Forsman, 2016, 56–57.) Konvention lisäksi myös viittauksiin pätee niin sanottu Matteus-efekti. Termin teki tunnetuksi bibliometriikan yhteydessä Merton ja hänen oppilaansa Zuckerman myös testasi ilmiön paikkansapitävyyttä Nobel palkinnon voittajilla. Matteus-efektillä tarkoitetaan bibliometriikassa, että jo viitattuihin viitataan jatkossakin todennäköisemmin ja mitä enemmän viittauksia, niin sitä todennäköisempää saada niitä lisää. (Merton, 1968; De Bellis, 2009, 56.) Syitä viittaamiselle on monia ja osa viittausanalyysistä siirtyykin hieman enemmän sisällönanalyysin suuntaan tutkittaessa tarkemmin konteksteja missä viittaus on tapahtunut. (Kärki, Kortelainen, 1996, 40–41 ja 51.)

Viittausanalyysia tehdessään, ja etenkin mitatessaan tutkijoiden tuotteliaisuutta, tutkijat joutuvat myös ratkaisemaan, miten käsitellään yhteisjulkaisut tutkimuksessa. Menetelmiä on kolme: lasketaan jokaiselle yhteisjulkaisun tekijälle yksi julkaisu, lasketaan vain ensimmäiselle kirjoittajalle yksi julkaisu tai ositetaan yksi julkaisu kaikille sen kirjoittajien kesken. (Andrés, 2009, 24–25.) Eri tavat laskea yhteisjulkaisuja saattavat tuottaa merkittävästi erilaisen näkymän siihen, miltä esimerkiksi tutkijan julkaisuaktiivisuus näyttää. Suosituin tapa kuitenkin laskea viittauksia tuntuu olevan laskea kokonainen julkaisu jokaiselle kirjoittajalle (Andrés, 2009, 24).

Viittausanalyysi lähtee olettamuksesta, että viittaus on aina myönteinen tunnustus toista kirjoittajaa kohtaan. Mitä enemmän julkaisu tai kirjoittaja on saanut viittauksia, sitä enemmän hän on myös omalla alallaan saanut huomiota. (Kärki, Kortelainen, 1996, 17.)

Julkaisukanavista mitataan viittausanalyysillä muun muassa julkaisukanavan saamia viittauksia, joista muodostetaan myös viittauskerroin (Andrés, 2009, 81–82).

Viittausanalyysin suosio on kasvanut viitetietokantojen suosion myötä. Suurimmat tällaiset viittaustietokannat ovat Web of Science, Scopus ja Google Scholar. Selkeästi vanhin näistä kolmesta tietokannasta on Web of Science, joka on rakennettu 1963 perustetun Science citation index -viittaustietokannan (SCI) päälle. Rinnalla toimivat vielä myös Social science citation index, joka myöskin on SCIn kehittäneen Institute for Scientific Informationin perustama. ISI on nykyään Clarivate analyticsin omistuksessa ja yhtenä heidän tuotteenaan on Clarivate analytics Web of Knowledge. Scopus ja Google Scholar ovat sen sijaan perustettu 2000-luvulla ja niiden arkisto ei näistä syistä ole kovin pitkälle historiaan kattava. (Andrés, 2009, 56; Forsman, 2016, 54.)

### 2.2.2 Matemaattisia malleja

Alfred J. Lotka oli yhdysvaltalainen matemaatikko, joka vaikutti myös bibliometriikan saralla. Hän loi niin sanotun Lotkan lain, joka on keskeinen matemaattinen laki bibliometriikassa. Lotkan laki kuvaa kirjoittajien julkaisuaktiivisuuden jakautumista tieteenalan sisällä. Yleisesti Lotkan laki toteaa, että noin 61% kirjoittajista tuottaa vain yhden julkaisun kukin. Julkaisuaktiivisuus laskee  $1/n^2$  jolloin yhden kaksi julkaisua kirjoittavien määrä on  $1/4$  yhden julkaisun kirjoittajien määrästä. Lotkan laki voidaan ilmaista myös

$$y_x = c \times 1/x^2$$

Kaavassa  $y_x$  on kirjoittajien määrä, joilla on  $x$  julkaisua,  $c$  on sellaisten kirjoittajien määrä, joilla on vain yksi julkaisu ja  $x$  on julkaisujen määrä. Jos siis halutaan ennustaa kolme julkaisua kirjoittaneiden määrä tieteenalan sisällä, jonka otoksessa on yhden julkaisun kirjoittajia 100 tulee kaavasta  $y_3 = 100/32 = 11$ . (1926 tässä Andrés, 2009, 23–24; Egghe, 2005, 16; Voos, 1974) Lotka laati eksponentin  $\exp=2$  käyttäen fysiikan tieteenalan julkaisuja 200 vuoden ajalta, mutta rajaten julkaisut vain kirjoittajiin joiden sukunimi alkoi A ja B -kirjaimella. Käytetty eksponentti saattaa vaihdella valitun julkaisujen perusjoukon mukaan. Jos kyseessä on joukko, jossa on paljon julkaisevia kirjoittajia, voi eksponentti lähennellä pikemminkin kolmea kuin kahta. Näissä tapauksissa siis paljon julkaisevien ja vähän julkaisevien välillä oleva kuilu on suurempi. Tieteenalojen välillä

vaihtelee myös kirjoittajien määrä julkaisua kohden. Tutkittaessa tieteen alaa ja julkaisuaktiivisuuden jakautumista tieteenalan sisällä on suositeltavaa käyttää yhteisjulkaisujen tapauksessa kokonaislaskentamenetelmää. Kokonaislaskennassa jokaiselle julkaisuun osallistuneelle kirjoittajalle lasketaan yksi julkaisu. (Andrés, 2009, 24–25.)

Voos (1974) teki laskelman informaatiotieteen julkaisuihin käyttäen Lotkan menetelmää ja päätyi eksponenttiin  $\exp=3,5$ . Hän tarkasteli informaatiotieteen julkaisuja vuosilta 1966–1970, mutta kuitenkin laskelmat hän teki yksittäisiin vuosiin. Näiden vuosien pohjalta hän totesi, että informaatiotieteissä 88% julkaisee vain yhden julkaisun. Coile (1975) osoitti laskelmat kuitenkin vääräksi. Vastauksessaan Coilen kritiikkiin Voos totesi, että  $\exp=3,5$  on vain yleistys. Lotka on alkuperäisessä fyysikoita ja kemistejä koskeneessa julkaisuaktiivisuuden tutkimuksessaan ottanut mukaan vain ensimmäisen kirjoittajan. Myöskään Voos ei omassa laskelmassaan käsitellyt kuin ensimmäistä kirjoittajaa eli monen kirjoittajan yhteisjulkaisuja käsiteltiin samoin kuin yhden kirjoittajan julkaisuja.

Samuel C. Bradford oli englantilainen matemaatikko ja kirjastonhoitaja, joka kehitti vuonna 1934 niin sanotun Bradfordin lain. Bradfordin laki tai hajaantumista kuvaava laki kuvaa tieteenalan tai aihealueen sisällä julkaisujen jakautumista eri julkaisukanaviin. Monissa yhteyksissä Bradfordin lakia visualisoidaan sisäkkäisillä ympyröillä, joista sisin ympyrä edustaa tieteenalan tai aihealueen keskeisiä julkaisuja. Nämä keskeiset julkaisut kattavat  $1/n$  -osan tieteenalan julkaisuista, jossa  $n$  on kuvattujen ympyröiden määrä. Monesti näitä sisäkkäisiä ympyröitä kutsutaan myös Bradfordin alueiksi. Kaikilla näillä alueilla julkaisujen määrä on siis sama, mutta julkaisukanavien määrä nousee siirryttäessä sisemmistä ympyröistä ulospäin. Yleisesti on havaittu, että julkaisut jakautuvat julkaisukanavien väleille suhteessa  $1:n:n^2$ . Tunnistamalla tai laskemalla tieteenalan keskeiset julkaisukanavat voidaan estimoida muiden alueiden koko. (Andrés, 2009, 31–35.) Bradfordin lakia voi pitää vertauksena sille minkä julkaisukanavien seurantaan tutkijan kannattaa pääasiassa resursoida aikaansa. Jos tutkija seuraa oman alansa keskeisiä julkaisukanavia, Bradfordin lain sisimpää aluetta, hänen ei tarvitse seurata kovin montaa julkaisukanavaa nähdäkseen kuitenkin ison osan alansa julkaisuista.

Kolmas matemaattinen malli, jonka kuvaaja muistuttaa kahta edellistä on Zipfin laki. Zipfin laki kuvaa tekstissä ja kielessä sanojen tiheyttä. Jos missä tahansa tekstissä oleva sanat järjestetään niiden frekvenssin mukaan laskevaan järjestykseen, niin huomataan sanojen tiheyden laskevan jokaisen sijan mukaisesti. Yleisin sana esiintyy tekstissä kaksi kertaa

niin usein kuin toiseksi yleisin sana. Yksinkertaisimmillaan Zipfin laki voidaan määrittellä niin, että sanan frekvenssi on käänteisesti verrannollinen sen sijaan, jos sanat järjestetään määrän mukaisesti. Zipf määritteli sen matemaattisesti  $ab^2 = k$ . (Kärki, Kortelainen, 1996,22; Forsman, 2016, 51–52.)

Kaikkia näitä kolmea matemaattista mallia tarkastellessa kuvaajan avulla, muistuttavat ne hyvin paljon toisiaan. Kaikissa näissä on vinouma vasemmalla ja nopeasti laskeva hyvin pitkä häntä. Tämänkaltaista jakaumaa ja matemaattista mallia voidaan kuvata yläkäsitteellä potenssilaki. Potenssilakeja esiintyy hyvin monen ilmiön yhteydessä kuten maanjäristysten tiheydessä taikka varallisuuden jakautumisessa yksilöiden välille. (Newman, 2005, 327.) Potenssilakeja siis esiintyy myös muualla kuin bibliometriikassa ja kielitieteissä ja ei täten ole vain tähän tieteenalaan sidoksissa olevia. Kuitenkin Lotkan ja Bradfordin lait molemmat kuvaavat nimenomaan julkaisutoimintaa. Lotkan laki selittää kirjoittajien ja julkaisujen lukumäärien suhdetta ja Bradfordin laki selittää julkaisujen jakautumista julkaisukanaviin. Lotkan ja Bradfordin lakien avulla voidaan havaita eroja eri tieteenalojen välillä tai tarkastella eroaako jokin yliopisto tai maa toisesta julkaisutoimintansa perusteella. Zipfin laki ei ole pelkästään julkaisuihin taikka yhteen kieleen sidottu vaan se on myös muilla kielillä ja hyvin monenlaisessa tekstissä osoitettu pitävän paikkansa.

### **2.2.3 Muita bibliometrisia menetelmiä**

Yhteisviittausanalyysillä pyritään tunnistamaan tieteenalalla vallitsevia suhteita ideoiden, tutkimuskohteiden tai tekijöiden välillä. Sillä voidaan tutkia ideoiden leviämistä ja tieteen rakennetta. Keskeistä yhteisviittausanalyysissä on tunnistaa erilaisia klustereita tutkittavien kohteiden välillä. Yhteys syntyy, kun samassa julkaisussa viitataan kahteen eri tutkijaan tai dokumenttiin. Ajatellaan, että samassa dokumentissa viitatuilla henkilöillä tai asioilla on jokin temaattinen yhteys keskenään. Mitä useammassa yhteydessä nämä kaksi tutkijaa tai dokumenttia mainitaan, sitä vahvempi temaattinen yhteys heillä on. (Small, 1973; Kärki, Kortelainen, 1996,18–20; Andrés, 2009, 75.) Yhteisviittaus muistuttaa jonkin verran bibliografista kytkentää, jossa tarkastellaan kahden tai useamman julkaisun lähdeluetteloita. Bibliografinen kytkentä muodostuu, kun kahdella tai useammalla julkaisulla on yhteisiä lähteitä lähdeluettelossaan. (De Bellis, 2009, 155–157.) Bibliografinen kytkentä on staattinen, kun taas yhteisviittaus muuttuu uusien julkaisujen myötä (Kärki, Kortelainen, 1996, 19). Kuten De Bellis sanoo (2009, 157), katsotaan yhteisviittauksia



tarkastellessa eteenpäin, kun bibliografisessa kytkennässä tarkastellaan mennyttä. Yhteisviittausanalyysi muistuttaa myös verkostanalyysia siinä suhteessa, että molemmissa tutkitaan yhteyksiä ja suhteita. Yhteisviittausanalyysista muodostuu sosiokognitiivinen kartta tieteenalan rakenteesta. Kartalla olevat tutkijat eivät välttämättä ole tehneet toistensa kanssa yhteistyötä, mutta heidän tutkimuskohteensa ovat lähellä toisiaan. Voidaan sanoa, että tällaiset kartat paljastavat näkymättömät yliopistot (invisible colleges) omalla tutkimusalueella (Andrés, 2009, 51–53). Verkostanalyysi taas mittaa nimenomaan yhteistyötä ja selkeää yhteyttä kirjoittajien välillä.

Pohjimmiltaan yhteisviittausanalyysi on hyvin samankaltainen kuin viittausanalyysikin. Sen tavoitteena on tutkia yhden yksittäisen kohteen sijasta kohdeparien kokonaisuuksia ja niiden muodostamia rakenteita. Forsmanin (2016) mukaan yhteisviittausanalyysi on joko tekijöitä tai dokumentteja käsittelevää, mutta yhtä lailla yhteisviittausanalyysia voidaan käyttää julkaisukanavien tarkasteluun (Andrés, 2009, 118–119). Näistä kuitenkin yleisempi yhteisviittausanalyysin kohde ovat tekijät ja dokumentit.

### **3 BIBLIOMETRIKKA OSANA TIEDEKENTTÄÄ JA TIE- TEEN ARVIOINTIA**

Bibliometriikka on mitannut sen kehittämisestä asti tiedettä julkaisujen kautta. Sitä on käytetty arvioidessa julkaisujen vaikuttavuutta tieteessä, etsittäessä keskeisiä tieteellisiä julkaisuja tai tutkijoita tietyllä tieteenalalla, mutta yhtä hyvin arvioidessa tieteen tilaa maittain, korkeakouluittain tai jopa yksilön tasolla hänen vaikuttavuuttaan tieteenalalla tai hänen tehokkuuttaan kirjoittajana. Menetelmiä on monia, joista viittausanalyysin ja yhteisviittausanalyysin esittelin toisessa luvussa, ja osa näistä on tieteellisen yhteisössä hyvinkin kritisoituja ja moniin näistä on ehdotettu vuosien saatossa parannuksia.

Käsittelen tässä luvussa bibliometriikan asemaa tieteessä, siihen kohdistuvaa kritiikkiä ja sitä miten bibliometriikkaa käytetään osana tieteen ja yliopistojen kansallista ja kansainvälistä arviointia.

#### **3.1 Bibliometrisen tutkimuksen asema tieteessä**

Bibliometrinen tutkimus voidaan jakaa kahteen selkeään kenttään niiden sovellusalueen mukaan: evaluoivaan ja suhteelliseen. Evaluoiva bibliometrinen tutkimus tarkastelee tieteen vaikuttavuutta vertailemalla eri kokoisten yksiköiden suhteellista julkaisuaktiiviteettia tai julkaisujen saamia viittauksia toisiinsa. Yksiköiden koot voivat vaihdella yksittäisten tutkijoiden, tutkimusryhmien, tiedekuntien, yliopistojen, tieteenalojen tai jopa maiden väliseen tarkasteluun. Suhteellinen tutkimus tarkastelee pikemminkin tieteenalan rakennetta tutkimalla yhteisviittaus- tai yhteissana-analyysin avulla tutkijoiden ja ideoiden yhteyksiä toisiinsa tai heidän luomien ideoiden tai julkaisujen samankaltaisuutta. (Kärki, Kortelainen, 1996, 14–24; Thelwall, 2008, 606–607; Forsman, 2016, 45–46.) Evaluoivaa bibliometrista tutkimusta käytetään vahvasti tieteen arvioinnissa myös kansallisella tasolla. Suomessa opetus- ja kulttuuriministeriö pohjaa 13 % yliopistojen ja 2 % ammattikorkeakoulujen rahoituksesta julkaisutoimintaan (Ehdotus ammattikorkeakoulujen rahoitusmalliksi 2017 alkaen, 2015, 42; Ehdotus yliopistojen rahoitusmalliksi 2017 alkaen, 2015, 36). Arviointia tehdään Suomen lisäksi muun muassa Norjassa ja Tanskassa. Suomen julkaisufoorumiluokitusjärjestelmä on Norjan luokitusjärjestelmän kaltainen (Forsman, 2016, 72).

Wolfgang Gänzel (2003, 37) määrittelee bibliometriselle tutkimukselle kolme tutkimuskohteiden tasoa, joille aggregointi tapahtuu: kirjoittajien tai tutkimusryhmien, julkaisukanavien tai instituutioiden sekä maiden taso. Nämä ovat tutkimustasoina esillä niin suhteita tutkivassa kuin evaluoivassa bibliometriikassa. Mielenkiintoisempi lajittelu Glänzellillä (2003, 9–10) on kuitenkin tutkimuksen tarkoituksen lajittelussa, siis siinä minkä takia tutkimusta tehdään. Tämäkin on kolmiportainen:

- Bibliometriikka bibliometriikan tutkijoille – matemaattista tutkimusta bibliometriikan luonteesta, sen sovellusalueista ja metodeista
- Bibliometrinen tutkimus tieteenalojen tutkimuksessa – tutkimusta jonka fokuksessa on tieteenalan kehitys ja rakenne
- Bibliometriikka korkeakoulu- ja tiedepolitiikan johtamisessa – vertailevaa tutkimusta maan, alueen tai instituution rakenteesta tai sen aktiivisuudesta.

Bibliometrisessä tutkimuksessa, joka pyrkii arvioimaan tieteellisen tutkimuksen vaikuttavuutta, lähes kaikki tutkimus pohjautuu tavalla tai toisella julkaisujen lähteisiin tai viitteisiin. Robert K. Merton (1973, tässä Thelwall, 2008, 606) kirjoittaessaan tieteen sosiologiasta on sanonut, että viittaaminen on kirjoittajien tapa hyväksyä tai kunnioittaa aiempia tutkijoita ja tutkimusta. Tämä ei kuitenkaan ole koko totuus, sillä viittaamiselle on useita syitä. Merton itse aiemmassa artikkelissaan (1968) on puhunut Matteus-efektin vaikutuksesta viittaamiseen, joka tarkoittaa viittausten kertymistä jo viitatuille. (De Bellis, 2009, 55–56.) Thed van Leeuwen (2004, 374–375) erottaa evaluoivan ja kuvailevan tutkimuksen toisistaan. Hänen mukaansa kuvaileva tutkimus tarkastelee korkeamman tason, esimerkiksi yliopiston tai koko maan, bibliometrisiä indikaattoreita, kun taas evaluoiva tutkimus perehtyy yksilön, tutkimusyksikön tai muun instituution sisäisen yksikön tarkasteluun. (van Leeuwen, 2004, 374–375.)

Evaluoivaa eli vaikuttavuuden tutkimusta tehdään jatkuvasti ja siitä on tullut myös osa yritysten liiketoimintaa. Vaikuttavuutta, etenkin julkaisukanavien osalta, tutkii Web of Sciencen takana oleva Clarivate Analytics julkaisemalla vuosittain Journal Citation Indexin. Vastaavia tilastoja keräävät ja laskevat myös Web of Sciencen suurimmat kilpailijat Scopus ja Google Scholar. Näitä lukuja käytetäänkin muun tutkimuksen lähdeaineistona. Suomessa vaikuttavuutta on väitöskirjojen ja pro gradu tutkielmien lisäksi pääosin tehty Suomen Akatemian, yliopistojen ja opetus- ja kulttuuriministeriön johdolla. Suomen Akatemia on vertaillut Suomen tieteen tilaa 25 muun EU- ja OECD- maan välillä (Lehvo,

Nuutinen, 2006). Suomen Akatemia käyttää myös Tieteen tila -tutkimuksessa bibliometrista analyysiä yhtenä menetelmänä mittaamaan tieteen tilaa (Tieteen tila). Tieteellinen tuotteliaisuus ja viittauskertoimet ovatkin yhtenä osana tieteen arvioinnissa niin Suomessa kuin kansainvälisestikin.

Suhteiden bibliometriikka tutkii tieteen rakennetta, sitä miten tieteentekijät tekevät yhteistyötä, miten ideat ja uudet tutkimusalueet muodostuvat ja keräävät kannatusta ja mistä asioista tai tutkijoista koostuu tieteenalojen keskeisimmät vaikuttajat ja ideat. (Thelwall, 2008, 607; Andrés, 2009, 41; Forsman, 2016, 67–69.) Suhteiden bibliometriikassa lähestytään sisällönanalyttisiä keinoja yhteissana-analyysin ja viittausten kontekstien tutkimuksen kautta, mutta myös tutkimalla artikkeleissa ja muissa julkaisuissa esitettyjä kiitoksia tai kirja-arvioita. Nämä paljastavat näkymättömiä tekijöitä, jotka kuitenkin ovat voineet olla merkittävässä roolissa tutkijan työn kannalta tarkastajina, keskustelukumppaneina tai ideoita tarkennettaessa. (Kärki, Kortelainen, 1996, 51–52.) Perinteisempää bibliometrista näkökulmaa suhteiden tutkintaan edustaa bibliografinen kytkentä, yhteisviittausanalyysi ja verkostanalyysi (Forsman, 2016, 54–59; Kortelainen, Kärki, 1996, 15–16; sama, 23–24.) Näistä yhteisviittausanalyysia on käsitelty tarkemmin luvussa 2.2.2. Suhteita tutkivassa bibliometriikassa keskeinen käsite on tekijöiden, julkaisukanavien tai aiheiden muodostamat klusterit. Näiden klustereiden välisiä yhteyksiä tutkimalla löydetään linkkejä tieteenalojen tai aiheiden välillä. Lisäämällä tutkimukseen aikadimensio, voidaan perehtyä ideoiden tai vaikutusten leviämiseen yli ajan. Visualisoinnissa klustereiden etäisyydet, muoto ja värit kertovat tieteen rakenteesta ja klustereiden läheisyydestä toisiinsa. (Forsman, 2016, 59; sama, 82; Andrés, 2009, 49–51; Gmür, 2003, 31–32). Thelwall nostaa (2008) nimenomaan visualisoinnin yhdeksi nousevaksi kentäksi bibliometriikan alalla. Forsman (2016) nosti sen myös yhdeksi alaluvukseen bibliometriikan oppikirjassaan. Tutkittaessa Web of Sciencesta viittausten ja julkaisujen määrää otsikolla vizualisation ja bibliometrics, nähdään, että viimeisen 10 vuoden aikana artikkeleiden sekä viittausten määrä on kasvanut lähes tasaisesti. Van Eck ja Waltman julkaisivat 2010 ilmaisen työkalun VOSviewerin, jolla voidaan helposti visualisoida tutkimusverkostoja ja bibliometrista dataa, joka osaltaan varmasti on tehnyt visualisointien laatumisen suosiota (van Eck, Waltman, 2010).

Bibliometriikka on keskeinen osa tieteen tutkimusta. Sitä käytetään avuksi korkeakoulu- ja tiedepoliittisissa päätöksissä ja yhtä lailla tieteen rakenteen hahmottamiseen ja kuvailuun. Juurikin näistä syistä on siis tärkeää myös ymmärtää bibliometrisia indikaattoreita;

mitkä asiat vaikuttavat indikaattoreiden laskentaan ja miten indikaattoreita voidaan soveltaa päätöksiä ja arviointia tehtäessä, jotta vältetään väärin päätelmien teolta.

### **3.1.1 Bibliometriikkaan kohdistunut kritiikki**

Yksikään bibliometrinen mittari ei ole riittävä yksistään ja käytännössä kaikkia mitta-  
reista on kritisoitu liian kapea-alaiseksi tai sellaiseksi ettei ne ota tarpeeksi huomioon tie-  
teellisen kommunikaation piirteitä eri aloilla. Yksi keskeinen rajoite viittauksia tutkitta-  
essa on lähdejärjestelmien suppeus koko tieteen kentällä. Keskeisimmät tietokannat eivät  
kata kaikkia julkaisukanavia tai tieteenaloja. Pienten kielialueiden ja perifeeristen maiden  
edustus näissä tietokannoissa on myös rajallinen. (Forsman, 2016, 77; Puuska, 2014, 38–  
42.)

Viittauskerrointa on kritisoitu paljon siitä, että se on hyvin angloamerikkalainen ja suosii  
aloja, joissa julkaisutahti ja julkaisuihin reagointi on nopeaa. Viittauskerroin voi olla  
myös hyvin vino jakaumaltaan lehden artikkeleiden kesken ja lehtien viittauskerroimen  
vuosittainen vaihtelu voi olla hyvinkin suurta. Korkean viittauskerroimen lehdissä kes-  
kiarvo voi olla myös hämäävä määre, jos lehdessä on julkaistu muutama hyvin paljon  
viitattu teos ja suuri määrä vähän tai ei lainkaan viitattuja artikkeleita. Viittauskerroin ei  
välttämättä siis kuvasta lehden laatua. Tieteellisesti merkittävä löydös voidaan yhtä hyvin  
julkaista matalamman viittauskerroimen julkaisussa. (Goldfinch, Yamamoto, 2012, 79-80;  
Kärki, Kortelainen, 1996, 26.) Lariviere ynnä muut (2016) testasivat lehtien viittausker-  
rointa ja sen kykyä ennustaa artikkeleiden mahdollisia viittausmääriä. Jopa 75 % tutki-  
muksessa mukana olleiden 11 lehden artikkeleista oli saanut vähemmän viittauksia kuin  
lehtien sille vuodelle laskettu viittauskerroin olisi voinut antaa ymmärtää. (Lariviere et  
al., 2016.)

### **3.1.2 Viittaus- ja julkaisukäytännöt**

Tutkijat pyrkivät julkaisemaan tutkimuksensa mahdollisimman arvostetussa julkaisussa,  
joka yleensä takaa julkaisulle näkyvyyttä. Eri tieteenaloilla julkaisukäytännöt ovat erilai-  
set ja julkaisukanavatkin eroavat tieteenaloittain. Monesti informaatio- ja kirjastotieteiden  
julkaisut osuvat luontevasti yhteiskuntatieteiden alle, kun taas joissakin tapauksissa  
ne ovat hyvinkin vahvasti tietojenkäsittelytieteisiin kallellaan.

Informaatio- ja kirjastotieteet lasketaan Tieteen tila tutkimuksessa yhteiskuntatieteiden alle (Bibliometrinen tieteenalaluokitus 2016). Web of Sciencesta löytyy informaatio- ja kirjastotieteiden aihealue ja myös tietojenkäsittelytieteiden alta cybernetics, informaatijärjestelmät ja tieteidenväliset sovellusalueet. Informaatio- ja kirjastotieteen lehdet löytyvät Web of Sciencen Social Sciences Citation Indeksistä, joka viittaa siihen, että se on yhteiskuntatieteiden alle luokiteltu. Julkaisukanava voi Web of Sciencessä saada 1-6 aihekategoriaa. Scopuksessa informaatio- ja kirjastotieteet ovat luokiteltu yhteiskunta- ja humanistisiin tieteisiin.

Hanna-Mari Puuska (2014) on tutkinut väitöskirjassaan tieteenalojen julkaisukäytäntöjä, mutta hän ei mene aihekategoriatasolle tutkimuksessaan, vaan keskittyy päätieteenaloihin. Julkaisukäytännöt ovat tietojenkäsittelytieteissä todella erilaiset verrattuna yhteiskuntatieteisiin. Yhteiskuntatieteissä suositaan kansallista julkaisemista sekä julkaisemista muissakin kuin tieteellisissä julkaisukanavissa. Tietojenkäsittelytieteissä, joka on luokiteltu bibliometrisen tieteenalaluokituksen mukaan luonnontieteisiin, taasen julkaisukanavat ovat yleensä kansainvälisiä ja todennäköisesti artikkeleita tieteellisissä lehdissä tai konferensseissa. (Puuska, 2014, 52 ja 65.)

### **3.2 Bibliometrinen tutkimus osana tieteen arviointia ja tiedepolitiikkaa**

Tiedettä arvioidaan tiedeyhteisön sisältä ja ulkoa käsin. Sisäistä arviointia tapahtuu osana perustutkimusta arvioitaessa yksiköitä ja tutkijoita sekä heidän meritoitumistaan. Tutkijoiden arviointi on sekä itse tutkijalle hyödyllistä todistettaessa omaa osaamistaan, mutta myös tiedeyhteisön validoinnin takia. Monelle tutkijalle nimenomaan muiden tutkijoiden arvostus ja huomio ovat kannustavia tekijöitä. Yliopistot, tutkimusryhmät sekä yliopistoyksiköt arvioivat myös omaa suoriutumistaan suhteessa muihin yliopistoihin ja muuhun tieteenkenttään. Tällaista arviointia saatetaan teettää yliopistoyksiköiden ja tutkimusryhmien ulkopuolella yliopistojen kirjastoissa tai kansainvälisillä bibliometriseen vaikuttavuuteen keskittyneillä tutkimusorganisaatioilla kuten Leidenin yliopiston tieteen ja teknologian tutkimuskeskuksella. (Forsman, 2016, 115–118). Valtion eri viranomaiset ja valtion tutkimuskeskukset kuten Valtion tieteen tutkimuslaitos tai Suomen Akatemia ovat kansallisella tasolla tiedettä arvioivia instansseja. Tieteen rahoittajina he osallistuvat tieteelliseen keskusteluun ja vaikuttavat siihen, millaista tutkimusta tehdään.

Tiedeyhteisön ulkopuolelta tulevaa tutkimusta tekevät kansainväliset tutkimusryhmät tehdessään yliopistojen ranking-listoja 2000-luvulla valtiollisten toimijoiden kuten Suomessa opetus- ja kulttuuriministeriön tai Norjassa Nordforskin rooli on kasvanut tieteen arviointien teettäjänä. Ministeriöt käyttävät myös bibliometrista tutkimustietoa korkeakoulu- ja tiedepolitiikan suunnittelun tukena. (Forsman, 2016, 100–106.)

Ranking-listoja on ensimmäisenä alkanut tehdä Shanghain Jiao Tong -yliopisto ja heti seuraavana vuonna perässä tuli Times Higher Education omalla ranking-listallaan. Suomalaisien yliopistojen sijoitukset näillä listoilla kiinnostavat mediaa kovasti. Sijoitukset vaihtelevatkin vuosittain jonkin verran. Listojen järjestykseen vaikuttavat eri ranking-listoilla eri mittarit. Shanghain listalla Nobel-palkinnot, julkaisutoiminta, top-1 listalle sijoittuneet tutkijat kuin myös Nature ja Science -lehdissä julkaistut artikkelit saavat arvostusta. Times Higher Education taas painottaa opetuksen ja tutkimuksen mainetta, julkaisujen määrää ja viittauksia sekä kansainvälisyyttä. Erilaiset painotukset listoilla aiheuttavat sen, että joihinkin tieteenaloihin kuten luonnon- tai lääketieteisiin painottuneet yliopistot voivat menestyä paremmin niiden nopean julkaisutahdin tai julkaisujen kansainvälisyyden johdosta. Lisäksi taustalla käytettävät indikaattorit ja tietokannat saattavat vaihtua, josta johtuen vuosien välinen vertailu on ongelmallista. (Forsman, 2016, 119–122.) Eri tavalla lasketut listat ja niiden takana vaikuttavien indikaattorien määrä voi olla vaikeasti hahmotettava ja tästä syystä ranking-listat palvelevatkin parhaiten yliopistoja silloin kun nämä määreet otetaan huomioon.

### **3.2.1 Tieteen arviointi suomessa**

Suomessa tieteen arviointia tapahtuu sekä rahoitusta arvioitaessa, mutta myös Suomen Akatemian Tieteen tila -arvioinnissa. Korkeakoulu- ja tiedepolitiikassa bibliometriikka näkyy siinä, miten bibliometrista tietoa käytetään osana korkeakoulujen rahoitusmekanismia. Yliopistojen ja ammattikorkeakoulujen rahoitukseen vaikuttaa julkaisuaktiivisuus sekä julkaisukanavien laatu niin Suomessa kuin kansainvälisestikin. Ehdotuksessa yliopistojen rahoitusmalliksi 2017 alkaen (2015) todetaan ”Yliopiston tuottamat julkaisut kuvaavat yliopiston tutkimustoiminnan tuotosta eli tuotettua tieteellistä ja muuta julkaisutoimintaa sekä sen siirtämistä yhteiskunnan ja tieteellisen toiminnan käyttöön.” Julkaisujen laatua pyritään mittaamaan käyttämällä tutkimuksen arvioinnissa julkaisufoorumin luokitusta julkaisujen julkaisukanavasta. Vuodesta 2017 alkaen tasoluokkien pisteytystä muutettiin entistä enemmän korostamaan tutkimuksen laatua painottamalla kertoimia

kohti tasoja kaksi ja kolme. Vertaisarvioituja julkaisuja painotetaan huomattavasti korkeammalle kuin vertaisarvioimattomia julkaisuja. Tutkimuksen osuus rahoituksesta on yliopistoille 13 % ja ammattikorkeakouluille 2 %. (Ehdotus yliopistojen rahoitusmalliksi 2017 alkaen, 2015.) Julkaisufoorumin käyttöä tieteen arvioinnissa tukee havainnot kansainvälisten viitetietokantojen kattavuuden rajoituksista (Julkaisufoorumi, Tausta).

Julkaisufoorumi toimii hieman samaan tapaan kuin Journal Citation Reports ja Scopus CiteScore siinä, että julkaisufoorumissa luokitellaan myös tieteellisiä julkaisukanavia eri tasoille niiden ominaisuuksien perusteella. Julkaisufoorumi toimii Tieteellisten seurain valtuuskunnan alaisuudessa ja sen ohjausryhmään kuuluu jäseniä yliopistoista, opetus- ja kulttuuriministeriöstä, CSC -tieteen tietotekniikkakeskukselta, kansalliskirjastosta, Suomen tieteellisten kirjastojen neuvostosta, Suomen Akatemiasta sekä Tieteellisten seurain valtuuskunnasta. Julkaisukanavia arvioivat kyseisten alojen tutkijoiden ja asiantuntijoiden paneelit.

Tasolle 1 julkaisufoorumissa voivat päästä sellaiset suomalaisen tutkimuksen kannalta keskeiset kotimaiset ja ulkomaiset kanavat, joka voidaan katoa tieteellisiksi julkaisukanaviksi. Tämä tarkoittaa, että julkaisukanavalla tulee olla asiantuntijoista koostuva toimintuskunta ja sen julkaisut ovat tieteellisiä sekä vertaisarvioituja. Tasolle 2 ja 3 pääsevät pääasiassa kansainväliset julkaisut, mutta ihmistieteissä hyväksytään myös suomen- ja ruotsinkielisiä julkaisukanavia. Tasolle 3 yltävät vain alojen korkeatasoisimmat julkaisukanavat. Tasolle 0 sijoitetaan sellaiset julkaisukanavat, jotka eivät täytä joitakin tason 1 vaatimuksia. Tasoilla 2 ja 3 on julkaisuvolyymeihin perustuvat kiintiöt eri tieteenaloilla. Panelisteilla on arviointityön tukena erilaisia vaikuttavuusindikaattoreita ja Tanskan ja Norjan vastaavat luokitukset. Jufo-luokituksen saavat kaikki julkaisut, joissa on julkaistu suomalaista tutkimusta. Lisäksi Jufo-luokituksen saavat myös kirjakustantajat. (Julkaisufoorumi, luokitteluperusteet.) Rahoituksessa vertaisarvioitujen julkaisukanavien painotukset vaihtelevat tasoilla 1–3 yhdestä kuuteentoista kertoimeen, kun taas tason 0 ja vertaisarvioimattomat ovat kertoimilla 0,1 ja 0,4. Vastaavia kansallisesti julkaisukanavia luokittelevia elimiä löytyy myös muista maista kuten Norjasta ja Tanskasta.



## **Julkaisufoorumia kohtaan esitetty kritiikki**

Suomessa korkeakoulujen julkaisutoiminnan arviointiin käytetään myös julkaisufoorumin luokituksia. Luokitusta on arvosteltu kuten viittauskerrointa ja muita bibliometrisia indikaattoreita. Jufo-luokituksiin voi ehdottaa korjauksia sekä lisäyksiä luokiteltaviin julkaisukanavien valikoimaan. Luokituksia arvioidaan uudelleen neljän vuoden välein ja lisäksi tehdään jatkuvasti täydennysarviointia. Kritiikkiä ja huolta on aiheuttanut lähinnä korkeakoulusektorilla opetus- ja kulttuuriministeriön korkeakoulutuksen rahoitusmalli, jossa 13% korkeakoulun rahoituksesta riippuu julkaisuista. Nämä julkaisut arvotetaan julkaisufoorumin luokituksen mukaisesti niin, että esimerkiksi tason kolme sarjajulkaisussa julkaistu artikkeli saa korkeamman kertoimen kuin tason kaksi tai yksi sarjajulkaisussa ilmestynyt artikkeli. Etenkin pienille tutkimusaloille tai jopa yksittäiselle tutkijalle voi muodostua paineita julkaista tutkimustaan korkeamman tasoluokituksen omaavassa lehdessä, vaikka tutkimus saattaisi saavuttaa paremmin alan tutkijat alemman tasoluokituksen julkaisussa. Forsman (2016, 117) sekä Sari Kivistö ja Sami Philström (2015) esittävät huolen tästä yliopistotasolta alas yksittäiselle tutkijalle kantautuvan viestin vaikutuksista pienille tieteenaloille ja yksittäisille tutkijoille. Tason kolme julkaisut ovat luonteeltaan helposti tieteen valtavirtaa ja tästä syystä niissä tieteenalaa mullistavan tai luovasti tieteenaloja yhdistelevän tutkimuksen julkaiseminen voi olla haasteellista. Huolta on aiheuttanut myös yleistajuisten kirjojen ja ammatillisten julkaisujen painoarvo julkaisufoorumissa. (Poropudas, Pölönen, 2015; Kivistö, Philström, 2015.) Toisenlainen huoli on julkaisufoorumiluokkien käyttö osana rekrytointiprosessia. Tutkijoiden julkaisuhistorian tarkastelu eri tasoissa julkaisukanavissa ei ole hedelmällistä kaikilla tieteenaloilla niiden erojen johdosta. (Forsman, 2016, 117)

## 4 TUTKIMUSKYSYMYS

Tutkin tässä työssä pohjoismaisia informaatiotutkimuksen yksiköitä käyttämällä bibliografisia menetelmiä. Erityisesti haluan vertailla Tampereen yliopiston informaatiotutkimuksen julkaisuja suhteessa Boråsin ja Kööpenhaminan yliopiston informaatiotieteiden julkaisujen saamiin viittauksiin. Siihen kuinka paljon viittauksia julkaisut ovat keränneet, kuinka jakautuvat parhaiten viitatus julkaisut ja tutkijat yliopistojen väleille, mutta toisaalta myös, missä julkaisuissa parhaiten viitatus julkaisut ovat ilmestyneet. Tarkastelen julkaisujen saamien viittausten kehitystä vuosien 2005 ja 2015 välillä. Vertaamalla yliopistojen julkaisutietoja Scopukseen ja Google Scholariin voin myös tutkia näiden tietokantojen kattavuutta tutkimusalalla.

Tutkimuskysymykseni muodostuvat siis kahdesta pääkysymyksestä:

- Kuinka paljon yliopistojen kirjasto- ja informaatiotieteiden julkaisut ovat keränneet viittauksia vuosien 2005 ja 2015 välillä?
  - Miten viitatuimmat tutkijat ovat jakautuneet yliopistojen välille
  - Miten Tampereen yliopisto vertautuu muihin yliopistoihin julkaisuaktiivisuudeltaan ja viittausmäärältään osalta?
- Miten julkaisut ovat jakaantuneet eri julkaisukanaviin ja tieteellisiin aikakauslehtiin?
- Miten maksullinen ja ilmainen viittaustietokanta eroaa toisistaan kattavuudella verrattuna kirjasto- ja informaatiotieteissä?

## 5 AINEISTO JA VIITTAUSTIETOKANNAT

Otin aineistoni lähtökohdaksi pohjoismaisten yliopistojen kirjasto- informaatiotieteiden tutkijakoulutuksen verkoston Nordic Research School in Library and Information Science (jatkossa NORSLIS). Mukana tässä verkostossa on Suomesta Tampereen yliopisto, Oulun yliopisto sekä Åbo Akademi. Ruotsalaisia yliopistoja verkostossa on viisi: Boråsin, Lundin, Umeån, Uppsalan ja Linnaeuksen. Norjalaisia korkeakouluja verkostossa on Tromsøen yliopisto sekä Oslon ja Akershusin ammattikorkeakoulu. Muista NORSLIS-verkoston kuuluvista maista kustakin on yksi korkeakoulu mukana verkostossa. Virosta Tallinnan yliopisto, Latviasta Latvian yliopisto, Liettuasta Vilnuihin yliopisto, Islannista Islannin yliopisto ja Tanskasta Kööpenhaminan yliopisto. Tutkimuksen kohteeksi valitsin kolme suurinta julkaisijaa kirjasto- ja informaatiotieteiden alalla NORSLIS-verkoston jäsenistä: Tampereen yliopiston Informaatiotieteiden yksikön (nykyisin osa viestintätieteiden tiedekuntaa), tanskalaisen Kööpenhaminan yliopiston Royal School of Library and Information Sciences -yksikön sekä ruotsalaisen Boråsin yliopiston Swedish School of Library and Information Science -yksikön. Viittaan näihin yliopistoihin ja heidän tutkimusaloihinsa tai yksiköihinsä jatkossa yliopiston nimellä helpottaakseni luettavuutta.

Bibliometrasta tutkimusta voi tehdä käyttämällä tutkijoiden nimiä tai julkaisujen tietoja. Julkaisut ovat menneiden vuosien osalta staattinen lista tutkijoiden kirjoituksia. Verrattuna julkaisuihin yliopistojen henkilöstö on vaihtuva ja samalla tutkijalla voi olla työsuhte usean yliopiston kanssa valitulla ajalla. Tästä johtuen haasteeksi muodostuu tämän tutkimuksen osalta löytää vuosien 2005 ja 2015 välillä yliopistoilla työskennelleiden tutkijoiden tiedot. Yliopiston julkaisuissa työskennelleet tutkijat voisi julkaisuarkistoista etsiä, mutta riippuen yliopiston tietokannasta ja merkitsemiskäytännöstä, joko kaikkia kirjoittajia ei ole tallennettu tai julkaisulle on tallennettu kaikki kirjoittajat, muttei heidän affiliaatioyliopistoa. Yliopistojen julkaisuarkistot ovat henkilökuntaan verrattuna staattisempia ja niiden perusteella on tärkeiden tietojen saaminen nopeampaa kuin henkilökunnan nimillä hakuja tehtäessä. Julkaisujen ongelma on kuitenkin yliopistoyksikköjen määrittelyt, jotka voivat kattaa suuremman tieteenalan alueen kuin mitä tässä työssä tutkin. Kolmen valitsemani yliopiston osalta pystyin kuitenkin etsimään tutkimusalueiden (Borås) ja tutkimusyksiköiden perusteella julkaisut joiden viittauksia tutkin. Rajasin pelituttimukseen julkaisut Tampereen yliopiston osalta tutkimusalueen ulkopuolelle, sillä halu-

sin nimenomaan tarkastella kirjasto- ja informaatiotieteiden julkaisuja. Joissakin tapauksissa julkaisun luokittelu kirjasto- ja informaatiotutkimukseen tai sen ulkopuolelle, oli hie-man hankalaa. Sama ongelma on varmasti ollut myös muissa yliopistoissa tieteenaloja ja aiheita luokitellessa. Toinen ongelma julkaisuarkistoja ja julkaisulistoja käytettäessä on se, etten voi olla täysin varma siitä, edustaako löytämäni julkaisut kaikkia kyseisten yliopistojen julkaisuja kirjasto- ja informaatiotutkimuksen alalta.

Tampereen yliopiston aineisto muodostuu Research Group on Information and Media Practicesin ja Finnish Information Retrieval Experts Groupin julkaisuarkistosta. Näiden kahden julkaisuarkiston julkaisuja täydensin TRIM-tutkimuskeskuksen ja viestintätieteiden yksiköiden julkaisuarkistoista, sillä kaikkia tutkijoita taikka julkaisuja ei löytynyt suoraan edellä mainittujen tutkimusryhmien omista julkaisuarkistoista. Lisäksi täydensin aineistoa neljän tutkijan Google Scholarista löytämilläni julkaisuilla. Kööpenhaminan Royal School of Library and Information Sciencen julkaisutiedot pyydettiin suoraan yliopiston yksiköltä. Boråsin yliopistossa toimivan The Swedish School of Library and Information Sciencen julkaisut poimittiin yliopiston julkaisuarkistosta suodattamalla sellaiset tutkimusartikkelit, joiden aihe on library and information science ja kuuluvat Boråsin yliopiston organisaatorakenteessa faculty of librarianship, information, education and IT tiedekuntaan. Jokaisesta julkaisusta on tiedossa tekijät, vuosi ja julkaisun nimi. Muut julkaisua koskevat tiedot vaihtelevat yliopistoittain. Tampereen yliopiston julkaisuarkistoista poimituista julkaisuista löytyi lähes kaikista myös julkaisukanava, kun taas Kööpenhaminan julkaisuista 46,1 % ja Boråsin julkaisuista 46,8% oli tallennettu julkaisukanava. Kööpenhaminan yliopiston julkaisutiedoista valitettavasti osasta (16,2 %) julkaisuja puuttui myös kirjoittajan nimi.

Boråsin ja Kööpenhaminan yliopistot ovat molemmat omissa tietokannoissaan luokitelleet julkaisut tyypeittäin kuten tieteellinen artikkeli tai artikkeli kirjassa. Tampereen yliopiston julkaisut luokittelin käyttämällä Tampereen yliopiston SoleCRIS-tietojärjestelmää, johon julkaisut on luokiteltu valtakunnallisen julkaisutyypiluokittelun mukaisesti. Kaikkia ennen 2011 julkaistuja teoksia ei löytynyt SoleCRIS:stä. Näissä tapauksissa luokittelin teokset niiden julkaisukanavan mukaan. Pieniä eroja näiden kolmen yliopiston luokittelussa oli. Boråsin yliopiston luokittelu oli jaottelultaan selkeästi yleismaallisempi ja on toiminut oman luokitteluni pohjana. Kööpenhaminalla erilaisia tyyppikategorioita oli 32. Tampereen yliopistolla ja siitä jatkamassa omassa luokittelussani kategorioita oli 14.

## 5.1 Viittaustietokannat

Viittaustietokannan valintaan vaikuttaa tehtävän bibliometrisen analyysiin kohde sekä tieteenala. Kolme merkittävää viittaustietokantaa, jotka kaikki kattavat useita tieteenaloja ovat Web of Science, Scopus ja Google Scholar. Nämä ovat tieteenalojen, julkaisukanavien, julkaisuvuosien, julkaisukielten sekä maantieteellisen kattavuutensa osalta erilaisia. Monitieteisiin tietokantoihin voidaan katsoa myös Microsoftin Academic, jonka toimintaperiaatteet ovat Google Scholaria vastaavat. Se on pääasiassa hakukone tieteellisille julkaisuille, mutta sen avulla voidaan tarkastella myös viittausmääriä eri aggregointitasoilla, yhteiskirjoittajuutta ja yhteistyöverkostoja. Academic ei kuitenkaan laske automaattisesti bibliometrisia indikaattoreita vaan tyytyy näyttämään viittausten ja julkaisujen määrät. Academic julkaistiin uudelleen kesällä 2017, jolloin siihen lisättiin uusia ominaisuuksia ja sen kattavuutta parannettiin (Microsoft Academic, 2017). Harzing ja Alakangas ovat tutkineet Academicin kattavuutta ja käyttöä viittausanalyysin suorittamiseen, mutta heidän tutkimuksensa on tehty ennen heinäkuun 2017 uudistusta. He totesivat, että Academic vertautuu hyvin Scopukseen ja Web of Scienceen kattavuudeltaan. Google Scholar oli kuitenkin viittausmäärällä mitattuna kattavampi kuitenkin vielä tammikuussa 2017. Harzing ja Alakangas totesivat, että Microsoft Academic tarjoaa hyvän vaihtoehdon viittausanalyysin tekemiseen ja se tulee entisestään parantamaan kattavuuttaan. (Harzing & Alakangas, 2017/1 ja 2017/2.)

Seuraavassa viittaustietokannoista esitellään tarkemmin Scopus ja Google Scholar, koska ne valittiin käytettäväksi viittaustietokannoiksi tässä työssä. Kirjasto- ja informaatiotieteiden alalla konferenssijulkaiseminen on yleistä, kuten näkyy keräämästäni aineistosta, ja tästä syystä päädyin valitsemaan Scopuksen Web of Sciencen sijasta. Tarkoitus oli myös vertailla maksullista tietokantaa ilmaisen tietokannan kattavuuteen ja tästä syystä toiseksi tietokannaksi valikoitui Google Scholar. Microsoft Academic oli uudistunut vasta kesällä 2017 ja sen käytöstä bibliometrisessä arvioinnissa ei ollut vielä kovinkaan paljon tutkimustietoa.

### Scopus

Scopus on Elsevierin vuonna 2004 perustama viittaustietokanta ja ensimmäinen merkittävä kilapilija Web of Sciencelle. Scopus on ainoa Yhdysvaltojen ulkopuolella toimiva

monitieteinen viittaustietokanta. Sillä onkin suurin valikoima ei-englanninkielisiä indeksoitavia julkaisukanavia. (Falagas et al., 2007). Scopus kattaa noin 22 800 julkaisukanavaa eri tieteenaloilta ja maantieteellisiltä alueilta (Scopus, 2017).

Falagas ja kumppanit (2007) vertailivat vuonna 2007 keskeisiä viittaustietokantoja ja niiden ominaisuuksia. He ottivat vertailuun mukaan Web of Sciencen, Scopusen, Google Scholarin sekä lääketieteeseen keskittyneen PubMedin. Heidän vertailussaan Scopus osoittautui kattavimmaksi niin indeksoitavien julkaisukanavien kuin artikkelien määrässä mitattuna. Vertailu tehtiin lääketieteestä käsin, eli tarkastelussa oli viittaustietokantojen erot ja ominaisuudet lääkärin ja lääketieteentutkijoiden käyttöön. (Falagas et al., 2007). Mongeon ja Paul-Hus tarkastelivat Scopusen ja Web of Sciencen kattavuutta indeksoitavien julkaisukanavien osalta. Scopus on huomattavasti kattavampi mitattuna indeksoitavien julkaisukanavien määrällä. He kuitenkin huomasivat samanlaisia vinoumia mitä Web of Sciencessä on aiemmin löydetty. Molemmat painottuvat lääke- ja luonnontieteisiin sekä englanninkielisiin julkaisukanaviin. (Mongeon & Paul-Hus, 2016.)

Meho ja Yang (2007) vertailivat Indianan yliopiston kirjasto- ja informaatiotieteiden tutkijoiden näkyvyyttä Web of Sciencessä verrattuna Scopusen ja Google Scholariin. He valitsivat 15 tutkijaa yliopistolta ja tutkivat heidän julkaisujen saamia viittaussuureita näissä kolmessa viittaustietokannassa. Scopus oli Web of Scienceen verrattuna lähes jokaisen tutkijan kohdalla kattavampi. Etenkin konferenssijulkaisuissa Scopusen kattavuus oli merkittävästi parempi. He huomasivat myös, että kirjasto- ja informaatiotieteiden alan sisällä erilaiset tutkimusalueet näyttäytyvät eri lailla. Scopusen kattavuus oli kuitenkin kaikilla tutkimusaloilla parempi kuin Web of Sciencen. Heidän päätelmänsä oli verrattaessa Web of Sciencea ja Scopusta, että tietokannat pikemminkin täydentävät toisiaan kuin pystyvät korvaamaan toinen toistaan. (Meho & Yang, 2007, 2112–2115 ja 2123.)

Gunnar Sivertsen (2016) tutki Web of Sciencen ja Scopusen kattavuutta yhteiskunta- ja humanistisissa tieteissä. Hän vertasi Norjan kansallisen julkaisutietokannan julkaisuja Web of Sciencesta ja Scopusesta löytyviin julkaisuihin. Tarkemmassa alavertailussa kirjasto- ja informaatiotieteiden osalta Web of Science löysi 58% kun taas Scopus 83% aikakauslehtien julkaisuista. Kirjojen kattavuus molemmissa tietokannoissa oli hyvin heikkoa norjalaisten kirjasto- ja informaatiotieteiden kirjojen osalta. Web of Sciencen Book Citation Indexissä 14% ja Scopusessa vain 5%. Sivertsen epäilee tämän johtuvan siitä,

että molemmat tietokannat kattavat pääasiassa englanninkieliset kustantajat. (Sivertsen, 2014.)

## **Google Scholar**

Google Scholar on Googlen tarjoama hakukone akateemisen kirjallisuuden hakuun ja viittausanalyysiin. Google julkaisi Scholarin vuonna 2004. Google Scholar kattaa tieteelliset julkaisut lehdissä ja konferensseissa, opinnäytteet ja väitökset, monografiat, vasta julkaisuvaiheessa olevat teokset, abstraktit, tekniset raportit ja Googlen määritelmän mukaan muut tieteelliset julkaisut. Viimeistä kategoriaa Google ei aukaise tarkemmin. Lisäksi Google indeksoi oikeuden päätökset ja patentit. Google Scholarin yksi merkittävä ero kaupallisiin viitetietokantoihin verrattuna on sen indeksointitavassa. Lehtien taikka konferenssien indeksoinnin sijasta Google Scholar indeksoi itse artikkeleita. (Google Scholar, 2017.) Tämä kertoo Googlen toimintaperiaatteen eron Scopukseen ja Web of Scienceen, sillä niiden pääasiallinen fokus on nimenomaan julkaisukanavien seuraaminen ja sitä kautta artikkelien, kirjoittajien ja viittausten seuranta. Google lähtee siis liikkeelle artikkelien indeksoinnista ja sitä kautta aggregoi viittaukset muille tarkasteltaville tasoille kuten julkaisukanavan tai kirjoittajan tasolle.

Google Scholarin julkaisun jälkeen sen ominaisuuksia, kattavuutta ja indeksoinnin laatua on pyritty arvioimaan useissa tutkimuksissa. Anne-Wil Harzing on yksi tunnetuista Googlen kattavuuden ja laadun arviointiin perehtyneistä tutkijoista. Vuonna 2006 hän julkaisi Publish or Perish -sovelluksen, jolla voi tehdä hakuja Google Scholariin. Sovellus tarjoaa työkaluja monipuolisiin hakuihin ja viittaustietojen tallentamiseen. Alun perin ohjelma kehitettiin vain Google Scholariin tehtäviä hakuja varten, mutta uusimmassa kuudennessa versiossa hakuja voi spesifioida tehtäväksi Google Scholariin, Web of Scienceen ja Microsoft Academiciin. Scopukseen hakujen tekeminen ei ole mahdollista tällä hetkellä Publish or Perishissä. Anne-Wil Harzing julkaisi yhdessä van der Walin kanssa vuonna 2008 tutkimuksen, jossa vertailtiin Google Scholaria Scopukseen ja Web of Scienceen. Tutkijat huomasivat muun muassa, että Google kattoi heidän aineistossaan huomattavasti paremmin myös ei-englanninkieliset artikkelit kuin Scopus taikka Web of Science. Harzing jatkoi tutkimustaan Google Scholarin kattavuudesta ja arvioi sitä väliinään viittausanalyysin tekoon. Hän julkaisi vuonna 2010 Publish or Perish kirjan, jossa opastetaan samannimisen sovelluksen käyttämiseen viittausanalyysin tekemisessä. (Harzing, van der Wal, 2008; Harzing, 2010, Preface.)

Meho ja Yang (2007) ovat myös vertailleet Google Scholaria, Scopus ja Web of Scienceen tuottamiin viittaustietoihin. Tarkoitus heillä oli nimenomaan tarkastella uusia viittaustietokantoja suhteessa vakiintuneempaan Web of Scienceen. He käyttivät vertailuunsa Indianan yliopiston Bloomingtonissa kirjasto- ja informaatiotieteiden tutkijoiden julkaisutietoja. Vertailussaan he havaitsivat, että Google Scholar löytää viittauksia huomattavasti enemmän kuin Scopus ja Web of Science. Mehon ja Yangin havainto oli, että Google Scholar löysi noin 53% enemmän viittauksia kuin Scopus ja Web of Science yhteensä. He huomasivat myös, että johtuen Google Scholarin toimintaperiaatteesta, tutkimusalat joilla verkkonäkyvyys ei ole kovin merkittävää, julkaisujen saamat viittaussuureet eivät kasvaneet merkittävästi käytettäessä Google Scholaria. Konferenssijulkaisut ja niiden viitteet taas löytyivät Google Scholarin avulla erittäin hyvin verrattuna molempiin Web of Scienceen ja Scopusiin. Tutkijat epäilevät, että suuri osa konferenssijulkaisuista julkaistaan tutkijoiden omilla verkkosivuilla tai muulla tavoin saatetaan yleisön nähtäväksi verkossa, jonka takia ne myös indeksoidaan Google Scholarissa. He kuitenkin huomauttavat, että Google Scholarin tuottamat viittaukset eivät ole painoarvoltaan yhtä vahvoja kuin Web of Scienceen taikka Scopusin. (Meho & Yang, 2007, 2115–2118.)

Googlen sanotaan olevan huomattavasti kattavampi kuin Scopusin taikka Web of Scienceen, mutta tämän varmistaminen on hieman hankalaa. Google ei itse kerro tarkkaan mitä sivuja, lehtiä tai tietokantoja sen robotit käyvät lukemassa (Harzing, van der Wal, 2008; Google Scholar.) Kuten aiemmin on mainittu, niin Google kuitenkin avaa indeksoitavia dokumenttityyppejä ja että indeksoitavat dokumentit ovat luotettavista lähteistä peräisin. Vaikka dokumenttityypit ovat luotettavista lähteistä peräisin, niin niiden merkittävyys ei välttämättä ole sama kuin viittausten, jotka Web of Science tai Scopus rekisteröi. Kuten Meho ja Yang (2007) totesivat, se että Google Scholar on kattavampi ei välttämättä tarkoita sitä, että viittaukset ovat kaikki merkityksellisiä. Harzing ja Alakangas varoittavat samasta asiasta 9 vuotta myöhemmin. He kuitenkin huomauttavat, että aiemmissa tutkimuksissa on todettu suurimman osan viittauksista pitävän sisällään kuitenkin intellektuellia sisältöä, vaikka ne olisivatkin peräisin esimerkiksi blogeista taikka sanomalehdistä. (Harzing & Alakangas, 2017/1) Toisin sanoen siis Google Scholarin rekisteröimät tiedeyhteisön ulkopuolelta tulevat viittaukset eivät täysin turhia kuitenkaan ole. Etenkin mietittäessä yhteiskunnallista vaikuttavuutta, Google Scholarin eduksi tulee mai-



nita poliittisten asiakirjojen indeksointi. Viittaukset tällaisissa asiakirjoissa voivat olla yhteiskunnan kannalta merkittäviä, vaikka ne eivät vertaisarvioituja tieteellisiä artikkeleita olekaan.

Toinen merkittävä haaste Google Scholaria käytettäessä viittausten keräämiseen on sen paikottainen epätarkkuus. Google Scholarin tapa indeksoida altistaa sen myös eri versioiden indeksoinnille. Samasta artikkelista saattaa olla indeksoituna tutkijoiden itse julkaisemat ennakkoversiot sekä vertaisarvioitu versio. Artikkelien indeksoinnissa on myös jonkin verran virheitä. Artikkelien bibliografisista tiedoista kuten julkaisuvuodesta, -kanavasta, nimestä taikka kirjoittajista osa voi olla monella tavalla indeksoitu ja tuottavat näin ollen väärin rekisteröityjä viittauksia. Harzin ja Alakangas huomauttavat, että virheellisiä viittauksia löytyy myös muista viittaustietokannoista ja ei siis ole vain Googlen ongelma. (Meho & Yang, 2007, 2111; Harzing & Alakangas, 2017/1.)

## **5.2 Viittausaineistonkeruu ja käsittely**

Aineisto kerättiin syksyn 2017 aikana Google Scholarista ja Scopuksesta. Pohdittaessa valittavia viittaustietokantoja käytettäväksi tässä tutkimuksessa oli huomioitava tietokantojen kattavuus julkaisukanavissa, maantieteellisesti sekä artikkelityypeittäin. Huomiotavaa oli myös käytettävien tietokantojen mahdollistavat hakutavat tietojen tallentamiseen liittyvät kysymykset. Molemmat tietokannat mahdollistavat hakujen tekemisen suoraan artikkelin nimellä, niin että voidaan koneellisesti tallentaa vain yhden artikkelin saamat viittausmäärät ja kerätä ne yhdeksi kokoavaksi yliopistokohtaiseksi listaksi.

Haut Google Scholariin tehtiin artikkelin nimellä siinä muodossa kuin se oli löytynyt julkaisuarkistoista. Joidenkin hyvin lyhyiden julkaisujen nimien kohdalla jouduttiin määrittelemään erikseen vielä kirjoittajatietoja, jotta oikean julkaisun löytäminen helpottuisi. Jos Google Scholar tunnisti usean artikkelin samalla nimellä, tutkittiin näistä kirjoittajatiedot, julkaisuvuosi sekä julkaisukanava ja jos kyseessä oli sama julkaisu, yhdistettiin viittausmäärät. Publish or Perish -ohjelmalla haettaessa artikkelin nimellä teoksia, tulosjoukko on haun ehdot täyttävä lista artikkeleita. Yhdistäminen tässä yhteydessä tarkoittaa, että valitsen yhden artikkelin sijasta kaksi tai useamman artikkelin tulosjoukosta. Ohjelma yhdistää tällöin näiden kahden julkaisun saamat viittaukset kyseisen haun alle. Löydettyt artikkelit viittaustietoineen tallennettiin yliopistokohtaiseen hakujen listaan. Listan

pystyi viemään Publish or Perishistä ulos CSV-tiedostona jatkokäsiteltäväksi ja analysoitavaksi. Publish or Perish -ohjelmasta saa metriikkatietoja kahdella tavalla tallennettua ja molemmista on tämän työn kannalta hyötyä. Ensimmäinen tapa on tallentaa löydetty artikkelit, jossa näkyy artikkelitasolla artikkelien saamat viittaukset. Toinen tapa on tallentaa hakujen metriikka, jossa on yhdistetyt viittausmäärät jokaisen haun osalta.

Haut Scopukseen tehtiin käyttämällä Scopuksen omaa käyttöliittymää ja sen edistynyttä hakutoimintoa käyttäen. Julkaisujen tiedot olin saanut Kööpenhaminasta Excel-tiedostona ja Boråsin julkaisut poimin heidän julkaisutietokannastaan CSV-tiedostona. Tampereen yliopiston julkaisutiedot keräsin suoraan Excel-taulukkoon. Ennen hakujen suorittamista Scopukseen, muotoiltiin Excelissä julkaisun nimi -kenttä muotoon `OR TITLE("[julkaisun nimi]")`. Näin voitiin kopioida useampi rivi kerrallaan Scopuksen edistyneeseen hakuun ja poistaa ensimmäinen OR-ehto hausta. Yksi hakulauseke piti sisällään kymmenen julkaisua muodossa `TITLE("[julkaisun 1 nimi]") OR TITLE("[julkaisun 2 nimi]")` ja niin edelleen. Hakuja tehtiin tällä tavalla noin 10 julkaisun ryppäissä. Jokaisen tuloksen kohdalla tarkastettiin, että kirjoittajatiedot vastasivat yliopistoilta kerättyjä tietoja ennen tallennusta yliopistokohtaiseen listaan. Hakulauseke hakee käytännössä kaikki julkaisut, joiden nimessä esiintyy suluissa ja lainausmerkeissä oleva teksti. Tästä johtuen julkaisuja, joiden nimet olivat lyhyitä tai muutoin monessa julkaisun nimessä esiintyviä lauseita, kuten kielikuva, koskevat haut jouduttiin suorittamaan yksittäin ja suodattamaan hakutuloksia kirjoittajien nimien perusteella. Julkaisut poimittiin tuloslistasta omaksi yliopistokohtaiseksi listakseen. Kun kaikilla yliopiston julkaisuilla oli tällä tavalla tehty haut, tallennettiin ne jatkokäsittelyä ja analysointia varten Excel-muodossa.

Haettuani molemmista tietokannoista julkaisut tarkastin kaikkein viitatuimmat tutkimukset molempien tietokantojen tuottamista tuloksista ja vertasin toisiinsa. Jos toisen tietokannan hakutuloksissa oli viittauksia saaneissa julkaisuissa sellaisia julkaisuja joita toisessa ei ollut, tehtiin kyseisen julkaisun osalta haku uudestaan. Näissä tapauksissa muokattiin alkuperäistä hakua niin, että siihen yhdistettiin julkaisun kirjoittajien tiedot sekä muutettiin julkaisun nimi helpommin löydettävään lyhennettyyn muotoon.

Molempien tietokantojen tiedoissa on tallennettu kirjoittajien nimet, julkaisun nimi, julkaisuvuosi ja viittausten määrä. Scopuksesta kerätyissä tiedoissa on myös jokaisen jul-

kaisun kohdalla tieto julkaisukanavasta sekä julkaisun yksilöivä tunnus Scopusin tietokannassa. Yksilöivää tunnusta taikka julkaisukanavaa ei saanut Google Scholarista tallennettua. Google Scholarista löydetyille julkaisuille luotiin yksilöivä tunnus jälkikäteen jokaiselle julkaisulle. Indeksien luonnin jälkeen Google Scholarin julkaisut käytiin läpi ja yhdistettiin yhden indeksitunnuksen alle kaikki kyseistä julkaisua koskevat rivit. Tämän jälkeen Google Scholarin julkaisut käytiin vielä kerran läpi ja luokiteltiin löydetyt julkaisut alkuperäisen luokittelun mukaisesti. Scopusista löydettyjä julkaisuja ei luokiteltu enää uudelleen, sillä Scopusista sai tallennettua myös luokituksen.

Kirjoittajien nimet tuottavat bibliometrisessä analyysissä monesti myös ongelman, sillä nimet ovat voitu kirjoittaa väärin viittauksen yhteydessä tai Google Scholarin tapauksessa ne on voitu algoritmin toimesta tulkita väärin ja väärässä järjestyksessä. Scopusissa ja Google Scholarissa on käytössä kirjoittajan yksilöivä tunnus, mutta se ei ole tietokantojen rajat ylittävä. Se ei myöskään ole virheetön, sillä Google Scholarissa, jos tutkija ei ole perustanut profiilia ja kertonut Googlelle mitkä teoksista ovat hänen, ei Google Scholar tätä pysty tietämään varmasti. Samoin kirjoitusvirheitä voi olla myös Scopusin tietokannassa, jolloin pohjoismaisista nimistä saattavat ä ja ö -kirjaimet korvaantua a ja o -kirjaimilla. Kävin molemmista viittaustietokannoista keräämiäni tietojen osalta myös kirjoittajien nimet läpi ja korjasin väärin kirjoitetut nimet.

## 6 TULOKSET

Tulokset jakautuvat kolmeen osaan aineiston eri kysymysten perusteella. Ensimmäisessä luvussa kerron kolmen yliopiston välisestä julkaisutoiminnasta 2005 ja 2015 vuosien välillä. Luvussa 6.2. käsitelen viittaustietokantojen eroja kattavuudessa ja luvussa 6.3. viittausten kerääntymistä. Luvuissa 6.4. ja 6.5 käsitelen viitatuimpien kirjoittajien jakautumista yliopistojen välille sekä julkaisujen jakautumista julkaisukanaville.

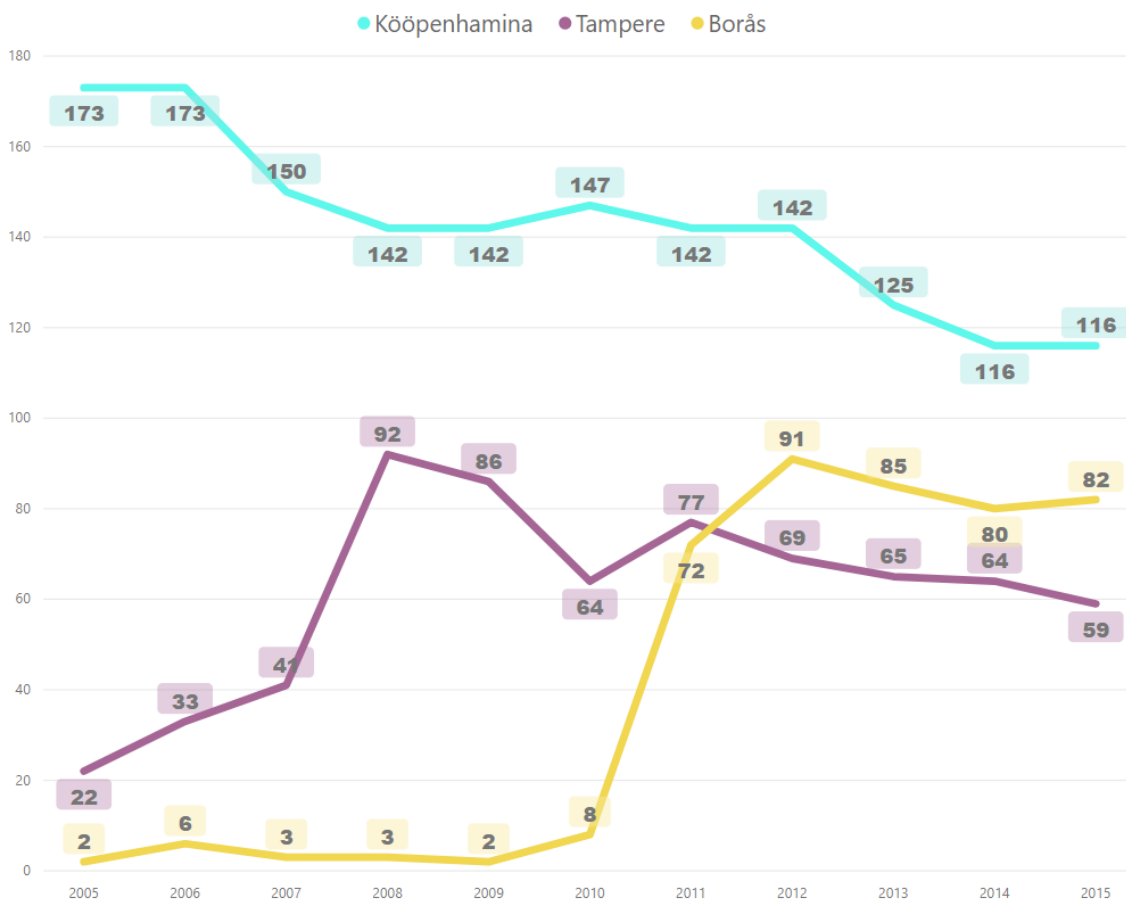
### 6.1 Julkaisutoiminta vuosien 2005 ja 2015 välillä

Julkaisutoimintaa tarkastellessa huomaa yliopistojen kokoeron sekä julkaisutoiminnan painottumisen erilaisuudet. Kööpenhaminan Royal School of Library and Information Sciencen on selkeästi suurin kolmesta tutkimastani yliopiston yksiköstä. Tampereen kirjasto- ja informaatiotutkimuksen julkaisumäärään verrattuna Kööpenhamina on julkaisut vuosien 2005-2015 aika yli kaksi kertaa sen mitä Tampereen yliopisto on julkaissut. Yhteensä julkaisuja vuosina 2005–2015 Tampereen yliopistolla oli 672, Kööpenhaminan yliopistolla 1568 ja Boråsin yliopistolla 440.

Taulukko 1. Kööpenhaminan, Boråsin ja Tampereen julkaisut tyypeittäin

	Borås	Kööpenhamina	Tampere
Julkaisu lehdessä	206 (46,8 %)	705 (44,9 %)	244 (36,3 %)
Julkaisu kirjassa tai raportissa	50 (11,4 %)	367 (23,4 %)	81 (12,1 %)
monografia tai raportti	35 (8 %)	138 (8,8 %)	43 (6,4%)
Väitöskirja	14 (3,2 %)	23 (1,5 %)	21 (3,1 %)
Konferenssi abstrakti, paperi tai juliste	124 (28,2 %)	270 (17,2 %)	203 (30,3 %)
Muut	11 (2,5 %)	65 (4,2 %)	79 (11,8 %)
<b>Yhteensä</b>	<b>440</b>	<b>1568</b>	<b>672</b>

Taulukosta 1 voi nähdä, että Kööpenhaminan yliopiston Royal School of Library and Information Sciencen julkaisuista suurin osa on joko julkaisuja lehdissä, kirjoissa tai raporteissa. Konferenssijulkaisut ovat seuraavaksi suurin kategoria lähes viidenneksellä julkaisuista. Painotus julkaisuihin lehdissä on kaikissa yliopistoissa, mutta Tampereella kuitenkin osuudeltaan pienin. Mielenkiintoista on huomata, että Tampereen yliopiston ja Boråsin yliopiston kirjasto- ja informaatiotieteen julkaisuista noin 30% on konferenssijulkaisuja, kun taas Kööpenhaminan julkaisuista ainoastaan 17 % kuului tähän kategoriaan. Kööpenhaminan yliopistolla painottuvat hieman erilaiset julkaisukanavat suhteessa Tampereen ja Boråsin yliopistojen julkaisuihin. Muiden julkaisujen osuus Tampereen julkaisuista on myös huomattavan erilainen verrattuna Boråsin ja Kööpenhaminan julkaisuihin tässä kategoriassa. Tällaisia julkaisuihin on luokiteltu Tampereen yliopiston julkaisuissa mukaan muun muassa artikkelit yhdistysten lehdissä artikkelit verkossa. Kööpenhaminan lehtien julkaisuista 109 ovat arvosteluja, kommentteja, johdantoja tai kommentteja.



Kaavio 1. Vuosittainen julkaisumäärän kehitys.

Vuosittainen julkaisuaktiivisuus on Kööpenhaminan yliopistolla ollut lievässä laskussa koko ajanjaksolla. Laskua Kööpenhaminan yliopiston julkaisuissa on kokonaisuudessaan

33 %. Tampereen yliopistolla lasku on alkanut vuodesta 2008 hieman piristyneen vuonna 2011 ja jatkaneen laskua siitä. Prosentuaalisesti laskua Tampereen yliopistolla on vuodesta 2008 tapahtunut 36 %. Boråsin yliopisto osoittautui mielenkiintoiseksi tapaukseksi julkaisujen osalta. Kokonaisuudessaan julkaisumäärä on uskottava vuosien 2005-2015 ajalta, mutta jotain merkittävää yliopistolla on tapahtunut vuonna 2011 sillä sen jälkeen julkaisumäärä nousee yhdeksänkertaiseksi 2010 vuoteen verrattuna. Huomattavaa on myös, että huolimatta ensimmäisen viiden vuoden vähäisestä julkaisutoiminnasta, Boråsin yliopisto näyttää julkaisujakaumaltaan hyvin samanlaiselta kuin Tampereen yliopisto. Se on julkaissut 206 lehtiartikkelia ja 124 konferenssipaperia kun Tampereella vastaavat luvut ovat 246 ja 200. Prosentuaalinen muutos Boråsin julkaisuissa vuodesta 2011, ensimmäisestä vuodesta kun julkaisumäärä vastaa Tampereen yliopiston määrää ja kuulos-tavat realistisilta, on 12,3 %. Kasvu näyttää kuitenkin olevan normaalia heilahtelua vii-meisen viiden vuoden perusteella. Kun otetaan vertailtaviksi vuosiksi jokaisen yliopiston kohdalla vuosi 2011 ja vuosi 2015, niin julkaisuaktiivisuus Kööpenhaminan on laskenut 18,3 % ja Tampereen yliopistossa 23,4 %.

## 6.2 Scopuksen ja Google Scholarin kattavuus kirjasto- ja informaatio-tieteissä

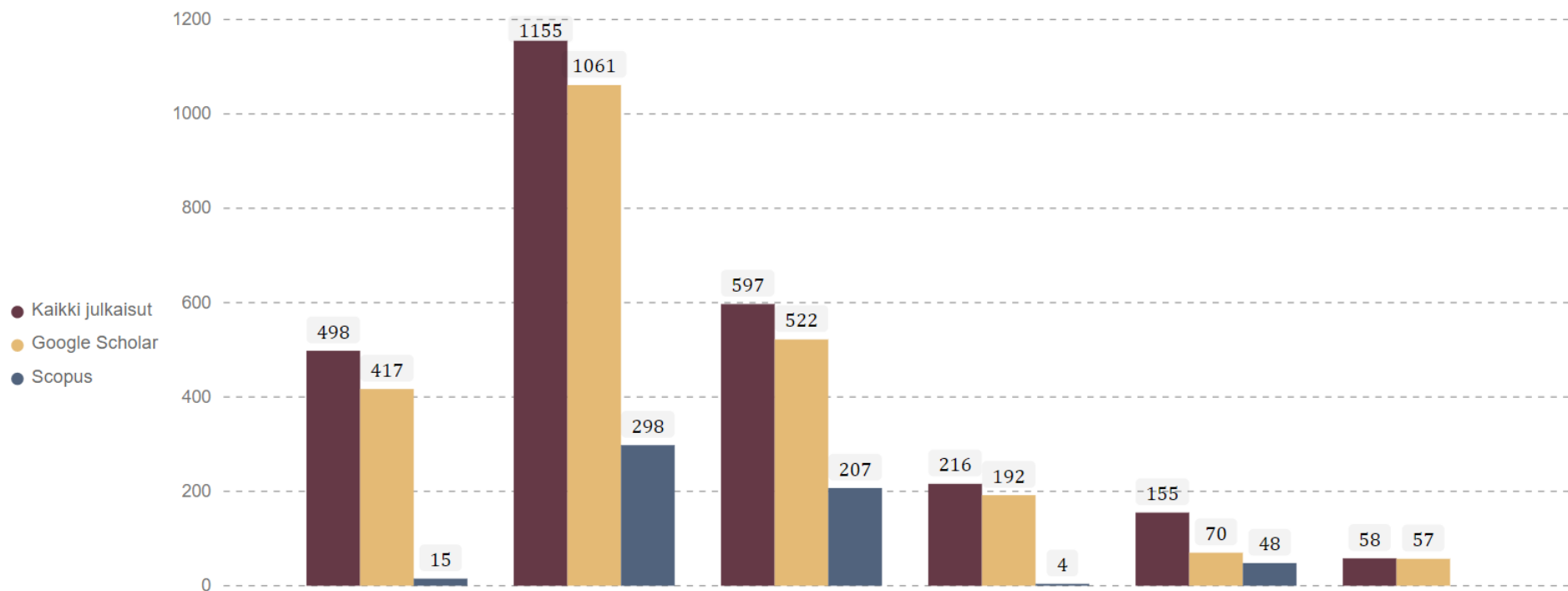
Scopuksen ja Google Scholarin tapa kerätä julkaisu- ja viittaustietoja eroavat hyvin paljon toisistaan. Luvussa 5.1. olen tarkemmin esitellyt viittaustietokantojen eroja, mutta katta-vuutta tutkiessa ero keräystavoissa tulee selkeästi esille. Scopus kerää julkaisutietoja in-deksoimistaan lehdistä, joiden kautta julkaisut päätyvät heidän tietokantaansa ja viittauk-set kertyvät indeksoitujen lehtien julkaisujen lähdeluetteloista. Googlen tapa toimia on erilainen, sillä he indeksoivat suoraan verkosta löytyneet julkaisut ja näiden julkaisujen lähdeluettelot.

Taulukko 2. Viittaustietokantojen kattavuus kirjasto- ja informaatiotieteissä.

	Borås	Kööpenhamina	Tampere
Kaikki julkaisut	440	1568	671
Google Scholar	416 (94.5%)	1403 (89.5%)	489 (72.9%)
Scopus	80 (18.2%)	260 (16.6%)	236 (35.2%)

Taulukoista kaksi voi nähdä, että Google Scholarin kattavuus on julkaisujen löydettävyydessä toista luokkaa verrattuna Scopukseen. Etenkin Boråsin ja Kööpenhaminan yliopistojen kohdalla Google Scholarin ylivertaisuuteen näyttäisi vaikuttavan molempien yliopistojen kattava julkaisuarkisto, jota hakukone indeksoi. Suurella osalla löydetyistä julkaisuista ei ollut viittauksia juurikaan, mutta ne löytyivät julkaisuarkiston sivuilta, joten ne rekisteröityivät Google Scholariin. Tampereen julkaisut ovat listattu hajallaan yksikön ja tutkimusryhmien sivuilla, mutta niissä ei ole linkkiä itse julkaisuun. Ihan kaikkea tämäkään ei selitä, sillä osassa Boråsin ja Kööpenhaminan julkaisuista ei myöskään ollut linkkiä itse julkaisuun, vaan ainoastaan viitetiedot.

Kaaviossa kaksi kattavuus on jaettu kuuteen julkaisutyypikategoriaan. Tässäkin tarkastelussa Google Scholar on kokonaisuudessaan huomattavasti kattavampi kaikissa kategorioissa. Erot eivät kuitenkaan ole niin valtavia osassa kategorioista. Scopuksen kattavuus konferenssien abstrakteissa ja lehtiartikkeleissa on huomattava. Etenkin Tampereen julkaisuista molemmat lehtiartikkeleista ja konferenssi abstrakteista on löytynyt lähes puolet.



Luokittelu	Julkaisu kirjassa tai raportissa		Julkaisu lehdessä		Konferenssi abstrakti, paperi tai juliste		Monografia tai raportti		Muut		Väitöskirja	
	GS %	SCOPUS %	GS %	SCOPUS %	GS %	SCOPUS %	GS %	SCOPUS %	GS %	SCOPUS %	GS %	SCOPUS %
Yliopisto												
Borås	94.0%	4.0%	94.7%	24.3%	96.0%	18.5%	94.3%	5.7%	100.0%	27.3%	100.0%	
Kööpenhamina	86.4%	2.2%	91.3%	19.6%	93.7%	32.6%	90.6%		73.8%	40.0%	100.0%	
Tampere	65.4%	6.2%	91.0%	46.3%	73.9%	47.8%	79.1%	4.7%	13.9%	24.1%	95.2%	
<b>Total</b>	<b>83.7%</b>	<b>3.0%</b>	<b>91.9%</b>	<b>25.8%</b>	<b>87.4%</b>	<b>34.7%</b>	<b>88.9%</b>	<b>1.9%</b>	<b>45.2%</b>	<b>31.0%</b>	<b>98.3%</b>	

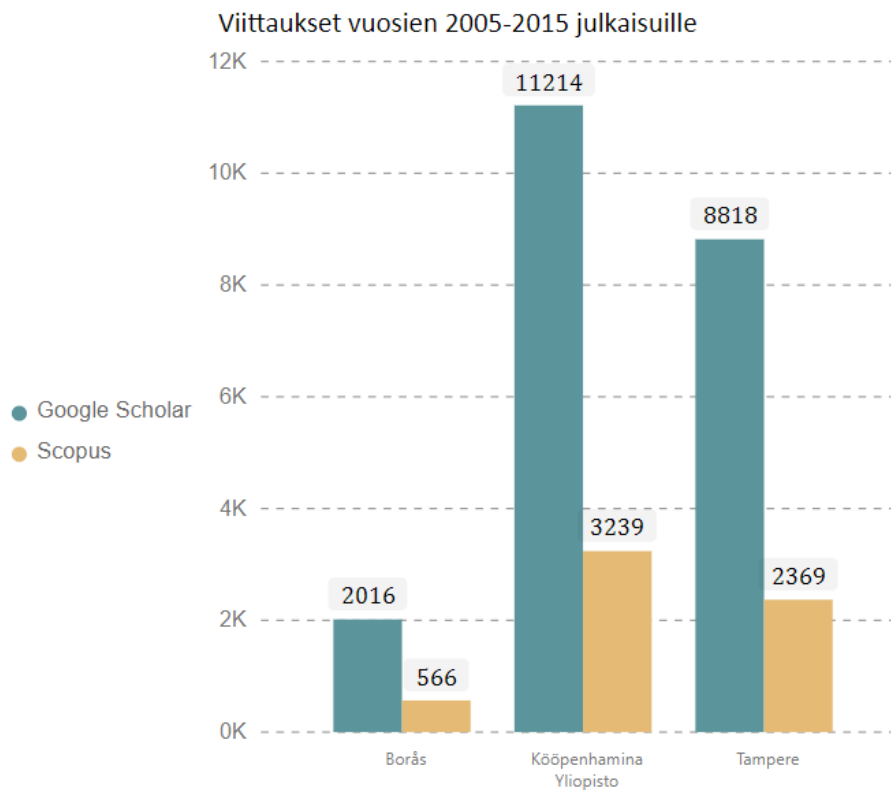
Kaavio 2. Viittaustietokantojen kattavuus julkaisutyypeittäin.



Kirjat, raportit sekä näissä esiintyneet artikkelit ovat Scopuksessa heikosti edustettuna. Samoin väitöskirjoja ei Scopuksesta löydy lainkaan. Tämä on tietoinen valinta Scopukselta, se on päättänyt jättää viittaustietokannan ulkopuolelle väitöskirjat, populaaritieteen kirjat, ohjekirjat, elämäkerrat, vuosikirjat, atlakset sekä perustutkinto-opiskelijoille suunnatut kirjat. Kirjojen osalta Scopus seuraa myös samaa periaatetta kuin tieteellisissä lehdissä, eli se indeksoi tietyn julkaisijan valikoiman. (Scopus, 2018.) Indeksoitavien julkaisijoiden ja kirjatyyppien valinnalla on vaikutusta kattavuuden vertailuun etenkin tieteenalojen välillä, mutta myös tämän tutkimuksen tapauksessa. Scopus mainitsee myös, että julkaisijan maine, julkaisijan kirjojen julkaisumäärä sekä tieteenalojen kattavuus vaikuttavat siihen otetaanko julkaisija osaksi indeksoitavia julkaisijoita (Scopus, 2018). Pienet yliopistot, kuten tässä työssä mukana olleiden Kööpenhaminan, Boråsin ja Tampereen yliopistot, eivät pysty kilpailemaan julkaisumäärässä suurten kustantamoiden tai suurten yliopistokustantamoiden kuten esimerkiksi Harvard University Press tai Cambridge University Press kanssa. toiminta kirjojen julkaisijana on julkaisumäärällisesti pientä. Google Scholar loistaa näissä kategorioissa verrattuna Scopukseen. Toimintaperiaatteensa vuoksi, se ei erottele teosta sen perusteella kuka sen on julkaissut taikka onko kyseessä raportti, populaaritieteen julkaisu vai kenties väitöskirja. Kirjojen, raporttien, väitöskirjojen ja näissä olevien artikkeleiden osalta julkaisijoina Google Scholarin tiedoissa esiintyy etenkin Tanskan ja ruotsin kansallinen julkaisujen viittaustietokanta ja Google Books. Julkaisija on tässä yhteydessä hieman väärä termi, sillä kyse on monessa näistä julkaisuista pikemminkin Googlen indeksoinnin osoite kuin varsinainen julkaisija, vaikka niitäkin listalta löytyy kuten Springer.

### 6.3 Viittausten kerääntyminen vuosien 2005-2015 välillä

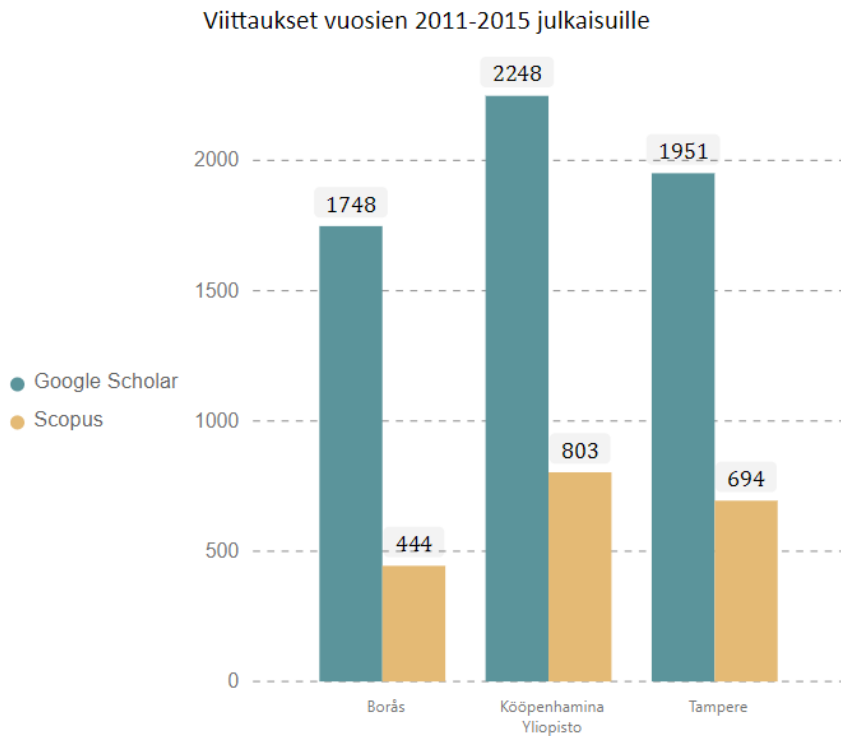
Viittauksia kolmen yliopiston julkaisut ovat kerryttäneet parhaillaan 10 vuotta ja viimeisenäkin vuonna julkaistujen osalta kahden vuoden ajan. Kokonaisuudessaan Google Scholarista löydetyt julkaisut ovat kerryttäneet 22 048 viittausta ja Scopuksesta löydetyt julkaisut 6088. Kaaviosta kolme näkee miten viittaukset ovat jakautuneet yliopistojen välille. Kattavuudeltaan Google Scholar oli merkittävästi Scopusta parempi löytäen noin 86 % kaikista yliopistojen julkaisuista, kun taas Scopus löysi noin 21 % kaikista julkaisuista. Tampereen yliopisto oli tässä poikkeus, sillä sen julkaisuista Google Scholar löysi 72,9 % ja Scopus 35 %. Tietokantojen kattavuus näkyy myös niistä kerätyissä viittausmäärissä, sillä ero kunkin yliopiston kohdalla on hieman yli 70 %.



Kaavio 3. Viittausten jakautuminen yliopistojen välillä julkaisuille vuosina 2005–2015.

Julkaisuaktiivisuus näinä vuosina kolmella yliopistolla on erilainen, mutta etenkin Tampereen yliopiston ja Kööpenhaminan yliopiston välinen ero Scopuksen viittausmäärissä ei kerro suuresta erosta julkaisumäärässä. Boråsin yliopiston viittausmäärää selittää osin se, ettei ensimmäiseltä viideltä vuodelta ole juurikaan julkaisuja joihin viitata ja näin ollen

Tampereen yliopistolla ja Kööpenhaminan yliopistolla viittausmäärää nostavat näinä viitenä vuonna julkaistut teokset. Jos vuosien 2005–2010 julkaisut jätetään huomiotta yliopistojen jakauma näyttää tasaisemmalta kuten kaaviosta neljä voi nähdä.



Kaavio 4. Yliopistojen viittaukset julkaisuille vuosina 2011–2015.

Taulukosta 3 voi nähdä, että suurin osa kaikkien yliopistojen viittauksista tulee lehtiartikkeleista. Kööpenhaminan yliopistolla viittauksista suurempi osa tulee julkaisuista kirjoissa ja raporteissa, kun taas Tampereen yliopistolla yhtä suuri osa viittauksista tulee monografioista ja raporteista. Boråsin yliopiston muut kuin lehtijulkaisut jakaantuvat tasaisemmin eri kategorioihin kuin verrokeillaan. Tampereen ja Kööpenhaminan yliopistolla myös konferenssijulkaisujen osuus viittauksista on huomattava. Scopusen lehdissä julkaistut kirjeet, arvostelut, pääkirjoitukset ja muut kirjoitukset näkyvät myös Kööpenhaminan yliopiston Scopusesta kerätyissä viittauksissa.

Taulukko 3. Viittausten jakautuminen julkaisutyypeittäin.

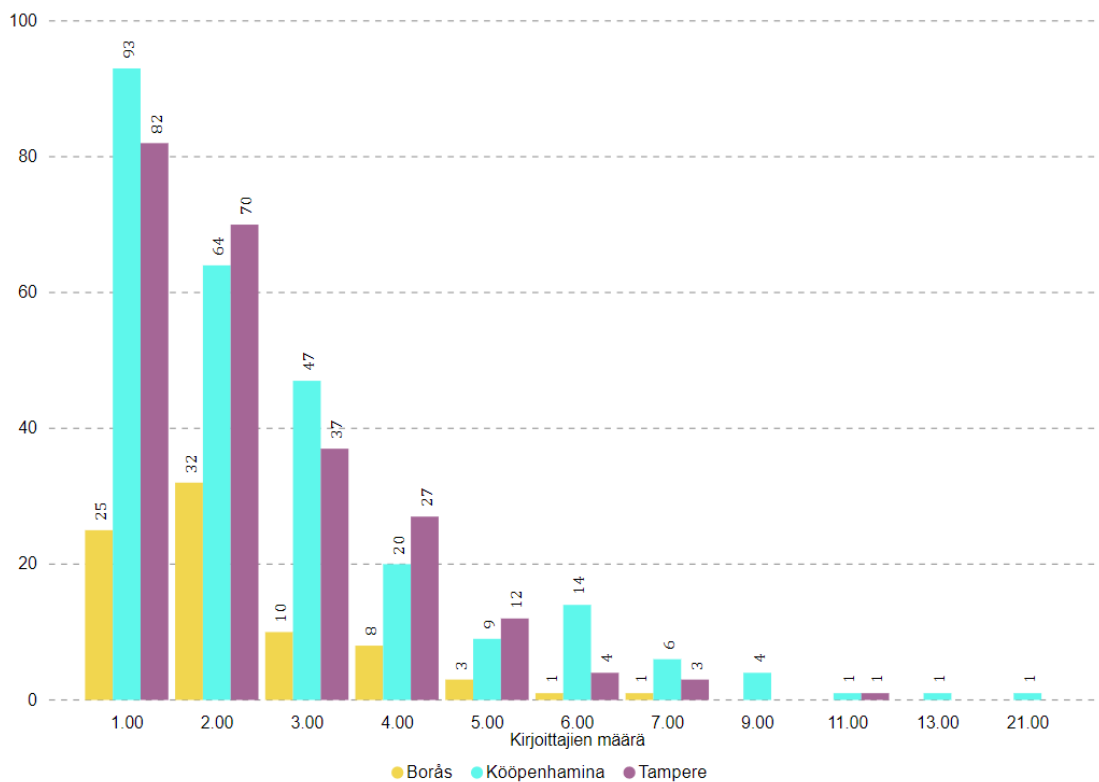
Viittausten jakautuminen julkaisutyypeittäin vuosien 2005-2015 julkaisuille

Yliopisto	Google Scholar julkaisumäärä	Google Scholar viittaukset	Scopus julkaisumäärä	Scopus viittaukset
<b>Borås</b>				
Julkaistu kirjassa tai raportissa	47	4.37%	2	0.18%
Julkaistu lehdessä	195	67.06%	50	72.79%
Konferenssi abstrakti, paperi tai juliste	119	9.08%	23	7.95%
Monografia tai raportti	33	13.19%	2	9.54%
Muut	11	1.24%	3	9.54%
Väitöskirja	14	5.06%		
<b>Kööpenhamina</b>				
Julkaistu kirjassa tai raportissa	317	19.20%	8	0.90%
Julkaistu lehdessä	644	53.71%	138	63.57%
Konferenssi abstrakti, paperi tai juliste	253	11.79%	88	19.60%
Monografia tai raportti	125	14.00%		
Muut	48	0.05%	26	15.93%
Väitöskirja	23	1.25%		
<b>Tampere</b>				
Julkaistu kirjassa tai raportissa	53	4.97%	5	0.76%
Julkaistu lehdessä	222	61.86%	113	65.01%
Konferenssi abstrakti, paperi tai juliste	150	12.69%	97	24.15%
Monografia tai raportti	34	19.06%	2	0.21%
Muut	11	0.07%	19	9.88%
Väitöskirja	20	1.35%		

### Viittaukset ositettuna kirjoittajille

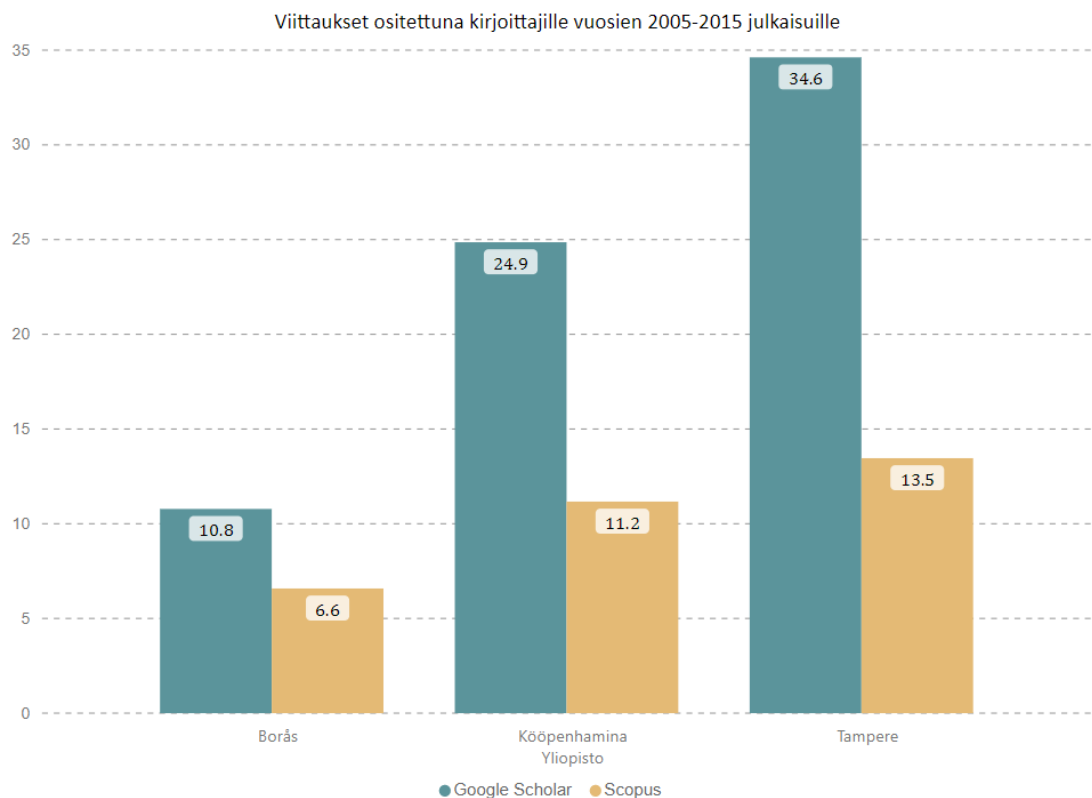
Jotta huomioitaisiin kolmen yliopiston erot julkaisumäärissä sekä kirjoittajien määrissä, on mielekästä tarkastella viittauksia ositettuna julkaisuille sekä kirjoittajille. Kirjoittajien määrä seuraa hyvin julkaisujen määrää, sillä yhteensä eri kirjoittajia Google Scholarissa Boråsin yliopistolla oli 188, Kööpenhaminan yliopistolla 452 ja Tampereen yliopistolla 256. Scopusessa kirjoittajien määrä oli pienempi; Boråsin yliopistolla 86, Kööpenhaminan yliopistolla 290 ja Tampereen yliopistolla 176. Ero kirjoittajien määrässä selittynee pienemmällä määrällä löydettyjä julkaisuja ja myös Scopusen indeksoimalla aineistolla. Muun muassa väitöskirjan kirjoittajista kaikki eivät julkaise muita julkaisuja ja eivät siksi näy kirjoittajien määrässä Scopusessa. Kirjoittajien määrä on laskettu kaikissa yliopiston julkaisuihin osallistuneista yksittäisistä kirjoittajista. Mielekkäämpää tässä kohtaa olisi käyttää alkuperäistä julkaistua aineistoa lähtökohtana kirjoittajien määrän laskennalle. Tämä kuitenkin vääristäisi valittujen kolme yliopiston osalta tilannetta, sillä Kööpenhaminan yliopiston julkaisulistassa osasta julkaisuista puuttui kirjoittaja (16,2 %) ja suuressa osassa kirjoittajaksi oli merkitty vain tosiasiallisesti ensimmäinen kirjoittaja.

Kirjoittajien määrä jokaisen julkaisun kohdalla vaihtelee luonnollisesti, mutta kolmen yliopiston välillä keskiarvot näyttävät hyvin samanlaisilta. Keskimäärin kirjoittajia Tampereen yliopistolla on 2,37, Kööpenhaminan yliopistolla 2,69 ja Boråsin yliopistolla 2,30. Scopuksesta löydettyt julkaisut ovat tässä kohtaa parempi lähdemateriaali kirjoittajien laskennalle. Google Scholar ei näytä kuin maksimissaan viisi ensimmäistä kirjoittajaa julkaisua kohden, joten se ei anna oikeaa kuvaa kirjoittajien määrästä. Yhden kirjoittajan julkaisuja on selkeästi enemmän Tampereen (34,7 %) ja Kööpenhaminan yliopistoissa (35,8 %) kun taas Boråsin yliopistolla kahden kirjoittajan julkaisuja (41,2 %) on enemmän kuin yhden kirjoittajan julkaisuja. Kuten kaaviosta viisi näkee, niin Kööpenhaminan ja Tampereen yliopistojen julkaisumäärä laskee hyvin tasaisesti kirjoittajamäärän kasvaessa. Boråsin yliopiston kahden kirjoittajan julkaisut tekevät tähän selkeän poikkeuksen.



Kaavio 5. Julkaisujen kirjoittajamäärän jakautuminen yliopistoissa.

Tampereen yliopiston ositettu viittausmäärä on molempien tietokantojen osalta korkeampi verrattuna Boråsin ja Kööpenhaminan yliopiston julkaisuihin, kun tarkastellaan viittausmäärää ositettuna kirjoittajille. Kaaviosta kuusi käy ilmi, että Tampereen yliopiston pienempi joukko kirjoittajia on saanut suhteellisesti enemmän viittauksia kuin kahden muun yliopiston kirjoittajat.



Kaavio 6. Viittausmäärä ositettuna kirjoittajille

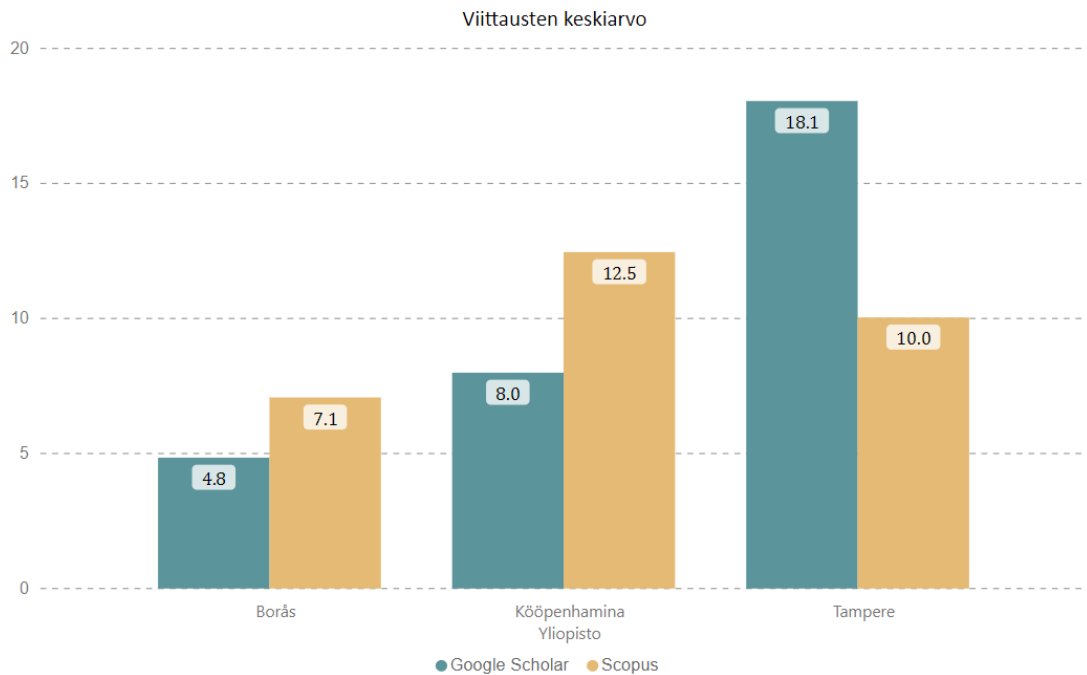
Jos huomioidaan vain vuodet 2011–2015, niin etenkin Boråsin yliopiston ja Kööpenhaminan yliopiston tilanne tasaantuu huomattavasti. Tampereen yliopiston pienempi joukko kirjoittajia on vieläkin kerryttänyt suhteellisesti enemmän viittauksia kuin kaksi muuta yliopistoa. Mielenkiintoista on huomata miten tilanne Boråsin ja Kööpenhaminan yliopiston välillä tasaantuu, kun ei huomioida alkupään vuosia, jossa Boråsilla ei juurikaan ollut julkaisuutoimintaa. Boråsin yliopiston kirjoittajille ositettu lukema on marginaalisesti parempi kuin Kööpenhaminan yliopiston sen ollessa Google Scholarin aineistossa 9,8 ja Scopusen aineistossa 5,3 kun taas Kööpenhaminan yliopiston vastaavat luvut ovat 9,6 ja 5,1. Vertailun vuoksi Tampereen yliopistolla vuosien 2011-2015 kirjoittajille ositettu viittausmäärä on Google Scholarissa 11,8 ja Scopusessa 6,0.

Jaettuna julkaisutasolla julkaisun saamat viittaukset tasaisesti kaikille kirjoittajille ja tarkastelemalla näiden keskiarvoa voi nähdä miten paljon yksi yksittäinen kirjoittaja keskimäärin saa ositettuja viittauksia. Keskimäärin kirjoittajat ovat saaneet viittauksia Tampereen yliopiston julkaisuissa Google Scholarin aineistossa 11,0 ja Scopusen aineistossa 6,9. Kööpenhaminan yliopistolla kirjoittajien keskimääräinen viittausmäärä Google

Scholarin aineistossa on 5,3 ja Scopusen aineistossa 9,8. Boråsin yliopiston keskimääräiset viittausmäärät kirjoittajaa kohden ovat Google Scholarin aineistossa 3,28 ja Scopusen aineistossa 6,0.

### Viittausten keskiarvo

Yliopistojen julkaisumäärä on erilainen ja syytä on myös kirjoittajakohtaisen tarkastelun lisäksi tarkastella viittausten keskiarvoa.



Kaavio 7. Viittauksia keskimäärin per julkaisu

Tampereen yliopisto erottuu Google Scholarin aineistossa ainoana, joka on kerännyt Scopuskeen verrattuna keskimäärin enemmän viittauksia julkaisuilleen. Julkaisuista löytyi kuitenkin pienempi osa Google Scholarista Tampereen yliopiston kohdalla kuin Kööpenhaminan tai Boråsin yliopistojen osalta. Scopusen kattavuus taasen Tampereen julkaisuissa oli korkeampi. Eroon keskimääräisissä viittausmäärissä ei vaikuta myöskään julkaisutyypit, sillä valitsemalla minkä tahansa yksittäisen julkaisutyypikategorian, pysyy jakauma samanlaisena. Molemmissa aineistoissa julkaisut lehdissä ovat myös keskiarvoisilta viittausmääriltä suuremmat. Tampereen Google Scholarissa kerätyssä aineistossa saattaa keskiarvoa nostaa muutama hyvin paljon viitattu monografia, jotka eivät näy Scopusen aineistossa. Nämä ovat Järvelinin ja Ingerwsenin kirjoittama *The turn: integration of information seeking and retrieval in context* (1178 viittausta) sekä Savolaisen kirjoittama *everyday information practices: a social phenomenological*

perspective (304 viittausta). The turn on kirjoitettu yhdessä Kööpenhaminan yliopiston tutkijan Peter Ingwersenin kanssa ja sama kirja toki näkyy myös Kööpenhaminan julkaisuissa piikkinä, mutta koska julkaisuja on enemmän kokonaisuudessaan, niin tämä ei nosta keskiarvoa niin paljoa.

#### **6.4 Viitatuimpien tutkijoiden jakautuminen yliopistojen välille**

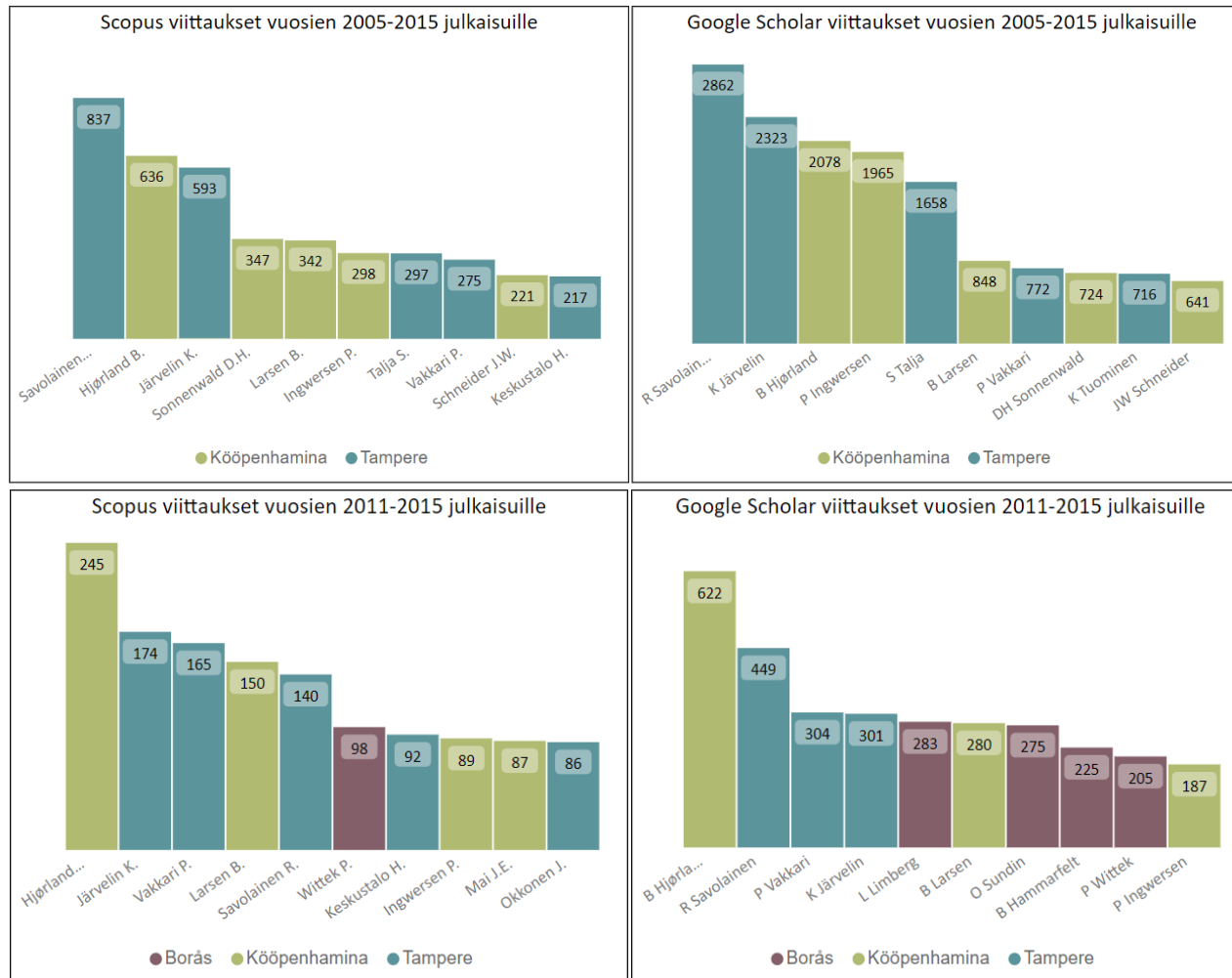
Toinen tapa vertailla yliopistoja toisiinsa on tarkastelemalla yksittäisten paljon viitattujen tutkijoiden jakautumista yliopistojen välillä sekä sitä, miten yliopistot tekevät yhteistyötä. Tarkastelen tässä kymmentä eniten viitattua tutkijaa molempien aineistojen osalta. Viittausmääriä verratessa on järkevää tarkastella sekä osittamattomia että ositettuja viittausmääriä. Ositettu viittausluku tasapainottaa yhden kirjoittajan julkaisun saamat viittausmäärät useamman kirjoittajan julkaisujen saamien viittausten kanssa, siten että ne ovat paremmin vertailtavissa yhden kirjoittajan lukuina. Osittamatonta viittauslukua mielekästä tarkastella, sillä yhden julkaisun kerryttämät viittaukset ovat kuitenkin kyseisen teoksen meriitti, jolloin siihen julkaisuun osallistuneet ansaitsevat siitä tunnustuksen. Tarkastelu tässäkin yhteydessä on järkevää tehdä koko aineistolle sekä vuosien 2011–2015 julkaisuille Boråsin yliopiston julkaisuaktiivisuuden kehityksen johdosta. Google Scholarin aineiston osalta tulee huomioida molemmissa, ositetuissa ja osittamattomissa viittausmäärissä se, että Google Scholarin kautta kerätyssä aineistossa näytetään vain maksimissaan viisi kirjoittajaa teokselle, kun taas scopusuksessa tätä määrää ei ole rajattu.

Keräämäni aineiston viitatuin teos oli Kalervo Järvelinin ja Peter Ingwersenin *The Turn*-kirja, jonka ensimmäinen painos oli julkaistu juuri valitsemani tarkasteltavan aikajakson alussa. Viittauksia *The Turn* oli aineiston keräämishetkellä syksyllä 2017 kerryttänyt 1178, mutta tätä kirjoittaessa viittauksia on 1188. Muitakin yhteistyöteoksia kolmen yliopiston välillä on.

Erot viitatuimpien tutkijoiden välillä on hurjat osittamattomissa viittausmäärissä kahden aineiston välillä. Aineistojen välillä on eroja kattavuudessa kuten luvussa 6.1 käy ilmi. Muutamat erittäin viitatu kirjat ovat jääneet Scopusin aineiston ulkopuolelle, koska Scopus ei näiden kirjojen kustantajien teoksia indeksoi. Kaaviosta kahdeksan näkee selvästi, miten viittausmäärät ovat aineistojen välillä erilaiset. Vuosien 2005–2015 julkaisu-



jen osalta kymmenen viitatuimpien tutkijoiden paikat menevät tasan Tampereen yliopiston ja Kööpenhaminan yliopiston kesken. Järjestyksessä tosin on hieman eroja kahden aineiston osalta, mutta Reijo Savolainen on kuitenkin viitatuin tutkija osittamattomien viittausmäärien osalta koko aineistossa. Savolaisella on Scopusin aineistossa 37 julkaisua ja Google Scholarin aineistossa 57 julkaisua. Hänen viitatuin teoksensa on molemmissa aineistossa ”Isms” in information science: Constructivism, collectivism and constructionism. Hän on kirjoittanut sen yhteistyössä Sanna Taljan ja Kimmo Tuomisen kanssa. Kyseinen teos on saanut Google Scholarissa aineistossa 366 viittausta ja Scopusin aineistossa 126 aineiston keruuhetkellä syksyllä 2017.

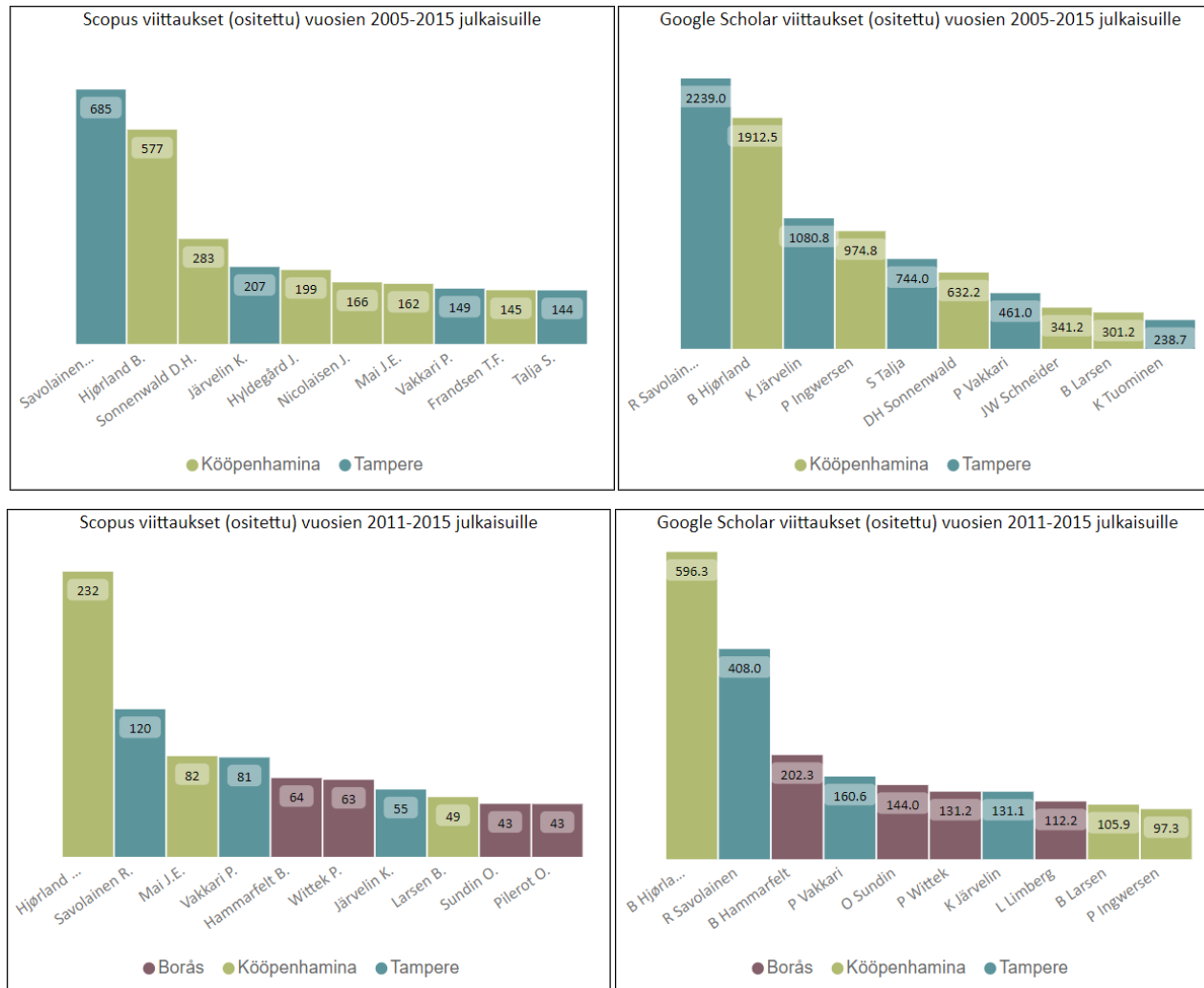


Kaavio 8. Viitatuimpien tutkijoiden jakautuminen yliopistojen välille osittamattomien viittausmäärien perusteella

Kun rajataan viittaukset ainoastaan vuosien 2011–2015 julkaisuille, niin myös Boråsin yliopiston tutkijat löytyvät viitatuimpien kymmenen tutkijan joukosta. Google Scholarin aineistossa Boråsin tutkijoita on neljä, mutta Scopusin aineistossa ainoastaan yksi. Viitatuin tutkija rajatussa aineistossa on Birger Hjørland Kööpenhaminan yliopistosta. Tampereen yliopiston Kalervo Järvelin, Pertti Vakkari ja Reijo Savolainen ovat viiden kärjessä hieman paikkaa vaihtaen. Birger Hjørlandin viitatuin teos vuosien 2011–2015 julkaisuissa on molemmissa aineistoissa Evidence-based practice: An analysis based on the philosophy of science. Yhteensä julkaisuja hänellä on Google Scholarin aineistossa 35 ja Scopusin aineistossa 17.

Osa kirjoittajista on julkaisumäärältään hyvin tuottoisia, mutta julkaisevat pääasiassa yhteistyössä muiden kirjoittajien kanssa. Mielekästä on siis tutkia viitatuimpien kirjoittajien jakautumista yliopistojen välille myös osittettujen viittausmäärien osalta. Osittaminen tapahtuu jakamalla julkaisun viittausmäärä tasan kaikkien kirjoittajien kesken. Osittettujen viittausmäärien perusteella viitatuimpien tutkijoiden kuva muuttuu hieman. Koko aineistossa Reijo Savolainen on edelleen viitatuin tutkija, hänen viittausmäärät ositettuna ovat hyvin lähellä itseasiassa osittamattomia lukuja. Hänen julkaisuistaan suuri osa on yksin kirjoitettuja. Samoin Birger Hjørlandin viittausmäärät ovat hyvin samankaltaisia kuin osittamattomien viittausmäärä. Hänen viitatuimmista teoksistaan molemmissa aineistoissa neljä viitatuinta ovat yksin tehtyjä. Viitatuimman teoksen viittausmäärän jakautuminen puoliksi Järvelinille ja Ingwersenille näkyy Järvelinin sijan laskuna. Kahden viitatuimman ero seuraavaan sijaan on myös merkittävä ositetuissa viittausmäärissä ja näytetty huipun terävyytenä.

Vuosiin 2011–2015 rajatussa aineistossa Boråsin yliopiston tutkijat pääsevät viitatuimman kymmenen joukkoon. He eivät yllä kuitenkaan tässäkään tapauksessa listan kärkeen. Boråsin yliopiston viitatuin kirjoittaja ositettulla viittausmäärällä on Björn Hammarfelt. Hänen viitatuin teos on yksin kirjoitettu Using altmetrics for assessing research impact in the humanities molemmissa aineistoissa. Google Scholarissa se on kerännyt 182 viittausta ja Scopusissa 56. Tämä on hyvä esimerkki ajankohtaisesta aiheesta, joka on kerännyt nopeasti viittauksia, sillä se on julkaistu vasta 2014, eli tarkastelujakson loppuvaiheilla. Julkaisuja Hammarfeltillä on tässä rajatussa aineistossa ainoastaan seitsemän Google Scholarissa ja kolme Scopusissa.



Kaavio 9. Viitatuimpien tutkijoiden jakautuminen yliopistojen välille ositettujen viittausmäärien perusteella

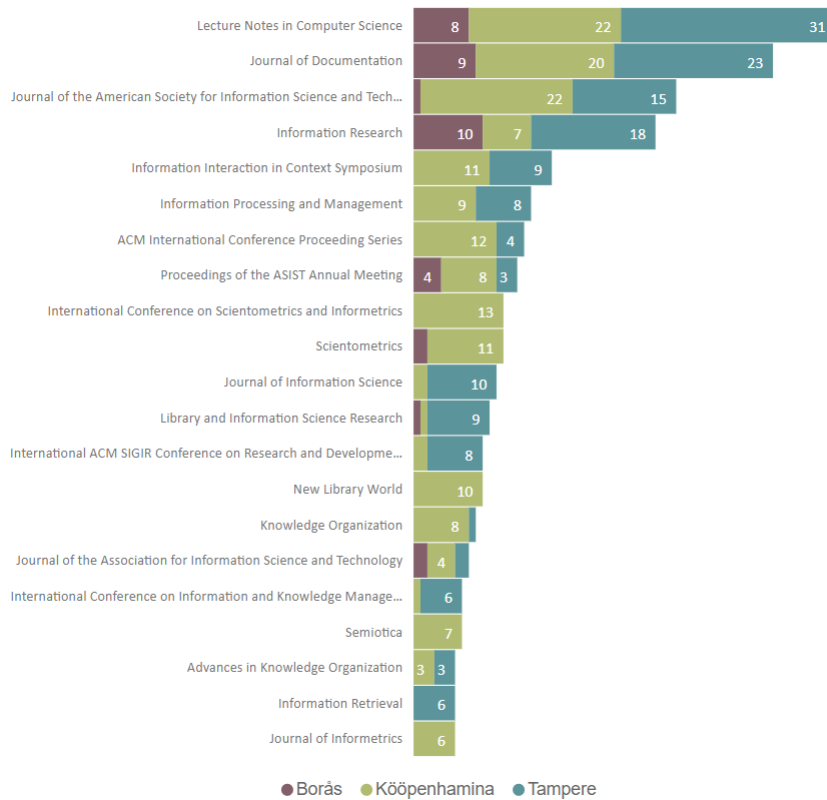
## 6.5 Viittausten ja julkaisujen jakautuminen julkaisukanavien välillä

Olin myös kiinnostunut, miten yliopistojen julkaisut jakautuvat eri julkaisukanaviin. Halusin tietää mikä on tämän aineiston perusteella keskeisimmät julkaisukanavat ja miten näissä julkaistut teokset ovat kerryttäneet viittauksia. Viittausten lähteiden tarkasteleminen olisi toinen näkökulma siihen, miten näkyvyys jakautuu eri lehtiin, mutta se ei ollut tämän työn yhteydessä mahdollista. Julkaisujen sekä viittausten jakautumisen tarkastelun teen Scopuksen aineistolla. Google Scholarin aineistossa hyvin suurelta osalta julkaisuja puuttuu julkaisukanava tai se on vain osittain tunnistettavissa. Julkaisujen jakautumista ei myöskään pysty tekemään alkuperäisellä yliopistoilta kerätyllä julkaisuaineistolla, sillä Boråsin ja Kööpenhaminan yliopiston aineistossa suurelta osalta julkaisuja puuttuu julkaisukanava.

Scopuksen aineistossa on yhteensä 36 eri julkaisukanavaa, joista kaaviossa kymmenen esitetään 20 eniten julkaisuja kerännyttä. Julkaisujen kasautuminen muutamalle suosituille julkaisukanavalle on tämän aineiston perusteella selkeää, vaikkakin huippu ei olekaan kovin terävä. Seitsemän suosituimmasta julkaisukanavasta ovat konferenssijulkaisusarjoja. Eniten julkaisuja on konferenssisarjajulkaisu Lecture Notes in Computer Science (61) ja perässä tulevat Journal of Documentation (52), Journal of the American Society for Information Science and Technology (38) ja Information Research (35) -lehdet. Viidentenä on keskeinen konferenssi Information Interaction in Context Symposium (20). Julkaisujen määrä seuraavissa lehdissä alkaa vähenemään nopeasti.

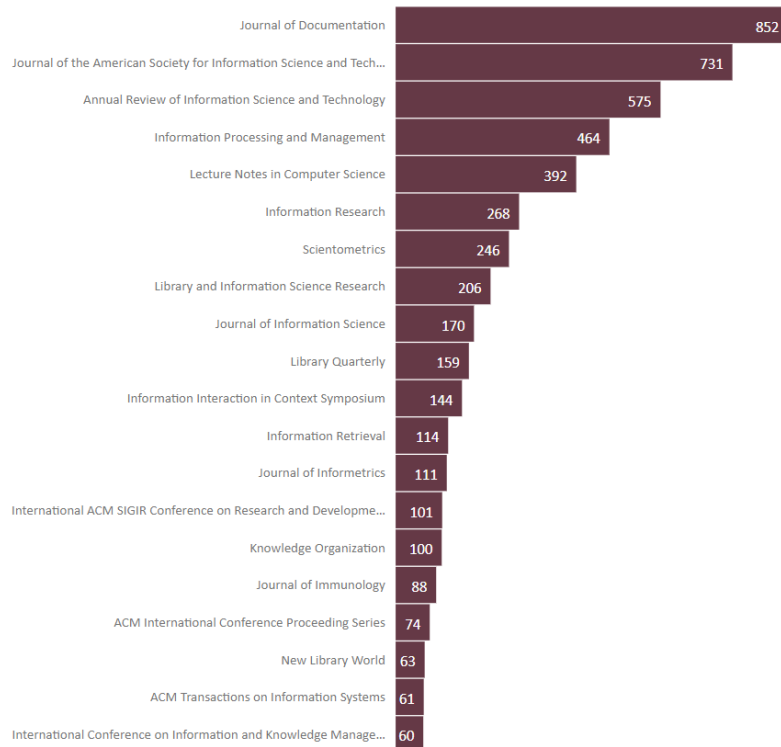
Mielenkiintoista on tarkastella julkaisukanavia sen perusteella kuinka paljon niissä julkaistut julkaisut ovat keränneet. Viitatuimpien 20 julkaisukanavan joukossa on vieläkin seitsemän konferenssijulkaisusarjaa, mutta viittausten perusteella järjestetty kaavio 11 näyttää, että järjestys julkaisumäärään verrattuna on erilainen. Viitatuin lehti on Journal of Documentation (852) kun taas viitatuin konferenssijulkaisusarja on Lecture Notes in Computer Science alle puolella viitatuimman viittauksista (392).

### Julkaisujen jakautuminen julkaisukanaviin yliopistoittain



Kaavio 10. Julkaisujen jakautuminen julkaisukanaville, 20 eniten julkaisuja kerännyttä.

### Viittausten jakautuminen julkaisukanaviin yliopistoittain



Kaavio 11. Viittausten jakautuminen julkaisukanaviin.

## 7 YHTEENVETO

Tutkielmani pääkysymys oli kuinka paljon kolme pohjoismaista yliopistoa ovat julkaisseet kirjasto- ja informaatiotutkimuksen julkaisuja vuosina 2005–2015 ja miten ne ovat kerryttäneet viittauksia. Lisäksi halusin vertailla miten Tampereen yliopiston julkaisut suhteutuvat Boråsin yliopiston ja Kööpenhaminan yliopiston julkaisuihin viittausmäärältään sekä miten viitatuimmat tutkijat ovat jakautuneet yliopistojen välille. Aineiston perusteella halusin myös tarkastella kahden viittaustietokannan eroja kattavuudessa sekä tarkastella julkaisujen jakautumista eri julkaisukanaviin.

Julkaisuaktiivisuus oli odotettavasti suurinta Kööpenhaminan yliopiston Royal School of Library and Information Sciencella sillä olihan se alustavien tutkimusteni perusteella selkeästi suurin yliopisto kirjasto- ja informaatiotutkimuksen alalla NORSLIS-verkoston yliopistoista. Julkaisuaktiivisuuden kehitys tarkastelemallani ajanjaksolla oli kuitenkin yllättävää, sillä Kööpenhaminan yliopiston julkaisuaktiivisuus oli lähes koko tarkasteltavan ajan laskussa. Tanskassa otettiin kansallinen suorituskykymittaristo käyttöön vuonna 2008, joka otti yhtenä osana huomioon myös bibliometriset indikaattorit (Ingwersen, Larsen, 2014). Suomen rahoitusmalli pohjaa samaan norjalaiseen rahoitusmalliin kuin Tanskan malli. Ingwersenin ja Larsenin (2014) mukaan suorituskykymittareiden tultua käyttöön tanskalaisten yliopistojen julkaisuaktiivisuus sekä niiden viittausmäärät ovat olleet jyrkässä nousussa 2008–2012 vuosien välillä. Tampereen yliopiston julkaisuaktiivisuus sen sijaan heilahteli hieman tarkasteltavan ajanjakson aikana, ja vaikka muutos oli pientä, niin kokonaisuudessaan se oli laskusuhteista. Boråsin yliopiston aineisto saattoi olla puuttellisesti kerätty tai Boråsin yliopistolla oli tapahtunut merkittäviä rakenteellisia muutoksia 2011. Yliopiston julkaisumäärä ensimmäiseltä kuudelta vuodelta 2005–2010 oli hyvin pieni, etenkin suhteessa siihen, miten se nousi vuonna 2011 edellisvuoden kahdeksasta julkaisusta 72 julkaisuun. Keräsin aineiston kansallisesta DIVA-tietokannasta. Boråsin yliopiston The Swedish School of library and Informationin sivuilla kerrotaan, että kaikkien laitoksen julkaisujen pitäisi löytyä DIVA-tietokannasta. Huima ero julkaisujen määrässä vuosina 2010 ja 2011 antaa kuitenkin sen vaikutelman, että osa julkaisuista puuttuisi aiemmilta vuosilta. Kokonaisaktiivisuus kuulostaa kyllä vertailukelpoiselta Tampereen yliopiston ja Kööpenhaminan yliopiston kanssa. Tästä johtuen tein tarkastelut julkaisuaktiivisuudessa ja viittausmäärissä pitkälti koko aineistolle, mutta myös

vuosien 2011–2015 julkaisuille erikseen. Halusin ottaa huomioon Boråsin yliopiston julkaisujen painottumisen vuodesta 2011 eteenpäin ja tarkastella ainoastaan näiden vuosien julkaisuja. Borås oli näiden vuosien osalta julkaisuaktiivisuudeltaan Tampereen yliopistoa aktiivisempi, mutta Tampereen yliopiston julkaisut kuitenkin keräsivät enemmän viittauksia. Viittauss määrä seuraa pitkälti julkaisumäärää, eli enemmän julkaissut yliopisto on kerännyt myös enemmän viittauksia. Yksittäisiä merkittävästi enemmän viittauksia keräyttäneitä julkaisuja oli aineiston joukossa, mutta siitä huolimatta kokonaisuutena yliopiston tasolla julkaisuaktiivisuus kertoo myös viittausten määrästä. Mielenkiintoista olisi ollut vertailla julkaisuaktiivisuutta yksittäisen tutkijan tasolla myös alkuperäisten julkaisuaineiston perusteella, mutta tämä ei ollut mahdollista aineistojen puutteellisten tietojen osalta. Kööpenhaminan yliopiston aineistosta puuttui julkaisun kirjoittaja 16 % julkaisuista. Viittaustietokannoista löydettyjen julkaisujen osalta hyvin moneen Kööpenhaminan ja Boråsin yliopistoilta kerättyihin julkaisutietoihin oli merkattu vain ensimmäinen kirjoittaja tai vain yliopiston omat kirjoittajat.

Tampereen yliopistoa verratessa kahteen muuhun yliopistoon oli odotettavaa, että enemmän julkaiseva yliopisto myös kerää enemmän viittauksia. Koko aineiston osalta tilanne oli tämä, mutta kun valittiin vain vuosien 2011–2015 julkaisut, Tampereen yliopiston keräämät viittaussmäärät ylittivät Boråsin yliopiston viittaussmäärän, vaikka Borås olikin julkaissut enemmän. Tampereen yliopistolla myös pienempi joukko kirjoittajia on kerännyt enemmän viittauksista kuin kahdessa muussa yliopistossa. Verrattaessa yksittäisiä tutkijoita viittaussmäärältään, olivat kolmen yliopiston kymmenen viitatuuta tutkijaa hyvin tasaisesti jakautuneet. Tampereen yliopiston Reijo Savolainen oli koko aineiston osalta viitatuin tutkija sekä ositetuissa että osittamattomissa viittaussmäärissä. Sen sijaan tarkasteltaessa vain vuosien 2011–2015 julkaisuja Kööpenhaminan yliopiston Birger Hjørland taas oli viitatuin tutkija. Tampereen yliopiston tutkijat olivat kuitenkin kaikissa vertailuissa hyvin edustettuina viitatuimpien kymmenen tutkijan joukossa.

Viittaussaineiston keräsin Scopuksesta ja Google Scholarista. Ero tietokantojen kattavuudessa oli häkellyttävä. Google Scholarista löytyi 86,2 % kaikista julkaisuista, kun vastaava luku Scopuksen tietokannassa oli 21,4 %. Kattavuus ei ole kuitenkaan kaikki kaikessa. Google Scholarin viittausten painoarvo ei välttämättä ole vastaava kuin Scopuksen aineiston, sillä Google Scholar kerää viittauksia myös muista kuin tieteellisistä julkaisuista. Sille riittää, että lähde on niin sanotusti hyvämaineinen Googlen standardien mu-



kaan. Google Scholarissa on myös paljon ongelmia viitetietojen tarkkuudessa. Dataa kerätessä sekä sen jälkikäsitellyssä joutuu käyttämään huomattava määrä tunteja datan läpikäymiseen. Julkaisujen nimissä, julkaisukanavissa, sekä kirjoittajien nimissä on hyvin paljon virheitä. Etenkin kirjoittajien ja julkaisun nimien virheellisyys tuottaa sen, että Google Scholarin kautta kerätyt viittaukset eivät välttämättä rekisteröidy oikealle kirjoittajalle, julkaisulle tai julkaisukanavalle. Julkaisuilla oli väärin merkittyjä duplikaatteja huomattava määrä. Tällaisia olivat muun muassa sellaiset julkaisut joiden nimissä ajatusviiva oli vaihtunut kaksoispisteeksi. Julkaisun avatessa huomaa, että kyseessä on sama julkaisu, mutta Google Scholar on rekisteröinyt nämä kahdeksi eri julkaisuksi. Kirjoittajan nimellä hakuja Google Scholariin tehdessä osa viittauksista saattaa myös jäädä piiloon, jos kirjoittajan nimi ei ole tunnistettavassa muodossa indeksoitunut. Tämän työn yhteydessä vastaan tuli esimerkiksi Tampereen yliopiston kirjoittajissa Jyri Saarikoski, jonka nimi oli muutaman julkaisun osalta lyhentynyt muotoon S Jyri. Muita mielenkiintoisia oli Ilkka Mäkisen sekä muutaman muun tutkijan etunimien muuttuminen muotoon T yliopisto Mäkinen tai T yliopisto Kettunen. Yleisesti ottaen myös ääkköset ja diakriittiset merkit ovat vaikeita Google Scholarille. Sama ongelma tosin on myös Scopuksen aineistossa, mutta Scopuksen sisällä julkaisu- ja viittaustietoja tarkasteltaessa ongelma on ratkaistu linkittämällä eri nimet toisiinsa. Vastaavia ongelmia oli muidenkin yliopistojen kirjoittajien nimissä, jossa etunimet vaihtuvat sukunimiksi tai ääkköset ja diakriittiset merkit puuttuvat.

Julkaisujen ja viittausten jakautuminen julkaisukanaviin toteutettiin käyttämällä Scopuksen aineistoa. Mielekästä olisi ollut tarkastella julkaisujen jakautumista julkaisukanaviin alkuperäisellä julkaisuaineistolla, mutta kuten julkaisujen kirjoittajienkin kohdalla oli, niin julkaisukanavien tiedot olivat puutteelliset Boråsin ja Kööpenhaminan yliopistojen julkaisuaineistoissa. Scopuksen aineistolla voitiin kuitenkin toteuttaa tämän työn osalta luotettavasti tarkastelu julkaisujen ja viittausten jakautumisesta julkaisukanaviin. Odotettavaa oli Bradfordinkin lain perusteella, että julkaisut jakautuvat muutamaa merkittäviin julkaisukanaviin ja suureen määrään sellaisia muutaman julkaisun kanavia.

Tässä työssä tarkastelin ainoastaan viittausmääriä yliopiston, julkaisukanavan sekä tutkijan tasolla. Jatkossa aihetta voisi laajentaa tarkastelemalla sitä, mistä maista ja julkaisukanavista viittaukset tulevat. Tarkasteluun voisi ottaa myös yhden yliopiston ja tarkastella pidemmällä aikavälillä, miten tutkimusaiheet ovat muuttuneet julkaisujen asiansanojen perusteella sekä kenen ja kanssa yksittäiset yliopiston tutkijat tekevät yhteistyötä. Tämän

työn perusteella Scopuksesta saatavat tiedot viittauksista, tutkijoista, affiliaatioista ja julkaisujen asiasanoista antaisivat mahdollisuuden erittäin mielenkiintoisille analyyseille. Abstraktien sanojen tai asiasanojen perusteella voisi tehdä tutkimusaiheiden kehittymisestä analyysejä tai tarkastella miten yliopiston tutkimusaiheiden profiili on muuttunut vuosien saatossa. Aiheita voisi toki tarkastella muutaman yksittäisen tutkijankin osalta pidemmällä aikavälillä. Verkostoanalyysillä voitaisiin tutkia sitä, miten yliopistot ja tutkijat tekevät yhteistyötä toistensa kanssa. Tarkastelemalla yhteistyössä tehtyjen julkaisujen viittauksia voitaisiin nähdä miten julkaisun tutkijoiden laaja maantieteellinen ulottuvuus vaikuttaa siihen mistä viittaukset julkaisulle lopulta kertyvät. Google Scholarilla on paikkansa bibliometrisessä analyysissä varmasti jatkossakin, mutta verrattuna maksullisiin tietokantoihin, sen tarkkuus viitetiedoissa on kyseenalainen ja viittausten painoarvo verrattuna maksullisiin kantoihin ei ole täysin samanlainen. Viittaustietokantojen vertailua on tehty paljon, etenkin Publish or Perish -ohjelman luonut Harzing on julkaissut paljon vertailuja eri tietokantojen välillä. Vuoden 2017 kesällä avattu uudelleen suunniteltu Microsoft Academic ja sen kattavuus on kuitenkin vielä toistaiseksi tutkimaton aihe kirjasto- ja informaatiotutkimuksen alalla.

# LÄHTEET

- Andrés, A. (2009). *Measuring academic research: how to undertake a bibliometric study*. Oxford: Chandos Publishing.
- Bibliometrinen tieteenalaluokitus 2016. Opetushallitus ja opetus- ja kulttuuriministeriö. Noudettu osoitteesta <https://vipunen.fi/fi-fi/layouts/download.aspx?SourceUrl=http://vipunen.fi/fi-fi/Raportit/Bibliometrinen%20tieteenalaluokitus%202016.pdf>
- The Clarivate Analytics Impact Factor. <https://clarivate.com/essays/impact-factor/>. Viitattu 21.10.2017
- Coile, R. C. (1975). Lotka and Information Science. *Journal of the American Society for Information Science (pre-1986)*; New York, 26(2), 133–134.
- De Bellis, N. (2009). *Bibliometrics and citation analysis: from the Science citation index to cybermetrics*. Lanham, Md: Scarecrow Press.
- Egghe, L. (2005). Power laws in the information production process. Bingley, U.K.: Emerald. Noudettu osoitteesta <http://www.emeraldinsight.com/1876-0562/05>
- Ehdotus ammattikorkeakoulujen rahoitusmalliksi 2017 alkaen, Opetus- ja kulttuuriministeriön työryhmämuistioita ja selvityksiä, 2015:18. Noudettu osoitteesta <http://julkaisut.valtioneuvosto.fi/bitstream/handle/10024/75159/tr18.pdf>
- Ehdotus yliopistojen rahoitusmalliksi 2017 alkaen, Opetus- ja kulttuuriministeriön työryhmämuistioita ja selvityksiä, 2015:19. Noudettu osoitteesta <https://julkaisut.valtioneuvosto.fi/bitstream/handle/10024/75157/tr19.pdf?sequence=1>
- Forsman, M. (2016). *Julkaisut ja tieteen mittaaminen: bibliometriikan käännekohtia*. Enoston.
- Garfield, E. (2006). Citation indexes for science. A new dimension in documentation through association of ideas. *International Journal of Epidemiology*, 35(5), 1123–1127. <https://doi.org/10.1093/ije/dyl189>
- Glänzel, W. (2003). *Bibliometrics as a research field. A course on theory and application of bibliometric indicators. Course Handouts*. Noudettu osoitteesta [http://www.cin.ufpe.br/~ajhol/futuro/references/01%23\\_Bibliometrics\\_Module\\_KUL\\_BIBLIOMETRICS%20AS%20A%20RESEARCH%20FIELD.pdf](http://www.cin.ufpe.br/~ajhol/futuro/references/01%23_Bibliometrics_Module_KUL_BIBLIOMETRICS%20AS%20A%20RESEARCH%20FIELD.pdf)
- Gmür, M. (2003). Co-citation analysis and the search for invisible colleges: A methodological evaluation. *Scientometrics*, 57(1), 27-57. doi:10.1023/A:1023619503005

- Goldfinch, S., & Yamamoto, K. (2012). *Prometheus assessed?: research measurement, peer review, and citation analysis*. Oxford: Chandos.
- Google Scholar About, Google, Noudettu osoitteesta: <https://scholar.google.com/intl/en/scholar/about.html>, Viitattu 23.11.2017.
- Harzing, A.-W., & Alakangas, S. (2017/1). Microsoft Academic: is the phoenix getting wings? *Scientometrics*, *110*(1), 371–383. <https://doi.org/10.1007/s11192-016-2185-x>
- Harzing, A.-W., & Alakangas, S. (2017/2). Microsoft Academic is one year old: the Phoenix is ready to leave the nest. *Scientometrics*, *112*(3), 1887–1894. <https://doi.org/10.1007/s11192-017-2454-3>
- Harzing, A., & van der Wal, R. (2008). Google Scholar as a new source for citation analysis. *Ethics in Science and Environmental Politics*, *8*, 61–73. <https://doi.org/10.3354/esep00076>
- Harzing, A.-W. (2010). *The Publish or Perish Book*. Noudettu osoitteesta <https://harzing.com/popbook/index.htm>, viitattu 26.11.2017
- Hirsch, J. E. (2005). An index to quantify an individual's scientific research output. *Proceedings of the National Academy of Sciences*, *102*(46), 16569–16572. <https://doi.org/10.1073/pnas.0507655102>
- Hirsch, J. E. (2010). An index to quantify an individual's scientific research output that takes into account the effect of multiple coauthorship. *Scientometrics*, *85*(3), 741–754. <https://doi.org/10.1007/s11192-010-0193-9>
- Hirsch, J. E., & Buela-Casal, G. (2014). The meaning of the h-index. *International Journal of Clinical and Health Psychology*, *14*(2), 161–164. [https://doi.org/10.1016/S1697-2600\(14\)70050-X](https://doi.org/10.1016/S1697-2600(14)70050-X)
- How are CiteScore metrics used? [https://service.elsevier.com/app/answers/detail/a\\_id/14880/kw/citescore/supporthub/scopus/](https://service.elsevier.com/app/answers/detail/a_id/14880/kw/citescore/supporthub/scopus/). Viitattu 21.10.2017.
- InCites Journal Citation Reports Help. <http://ipscience-help.thomsonreuters.com/helpios.uta.fi/incitesLiveJCR/overviewGroup/overviewJCR.html>. Viitattu 21.10.2017.
- Ingwersen, P., & Larsen, B. (2014). Influence of a performance indicator on Danish research production and citation impact 2000–12. *Scientometrics*, *101*(2), 1325–1344. <https://doi.org/10.1007/s11192-014-1291-x>
- Julkaisufoorumi, Luokitteluperusteet. (2017), <http://www.julkaisufoorumi.fi/fi/arvioinit/luokitteluperusteet>. Viitattu 21.10.2017

- Julkaisufoorumi, Tausta (2016), <http://www.julkaisufoorumi.fi/fi/julkaisufoorumi/tausta>. Viitattu 29.10.2017
- Kivistö, S., & Philström, S. (2015). Akateemisen elämänmuodon mahdollisuuden ehdoista. 3, 2015. Noudettu osoitteesta <http://journal.fi/tt/article/view/50769/15446>
- Kärki, R., & Kortelainen, T. (1996). *Johdatus bibliometriikkaan*. Tampere: Informaatiotutkimuksen yhdistys.
- Lariviere, V., Kiermer, V., MacCallum, C. J., McNutt, M., Patterson, M., Pulverer, B., ... Curry, S. (2016). A simple proposal for the publication of journal citation distributions. <https://doi.org/10.1101/062109>
- Lehvo, A., & Nuutinen, A. (2006). *Finnish science in international comparison*. Noudettu osoitteesta [http://www.aka.fi/globalassets/awanhat/documents/tiedostot/julkaisut/15\\_06-finnish-science-in-international-comparison.pdf](http://www.aka.fi/globalassets/awanhat/documents/tiedostot/julkaisut/15_06-finnish-science-in-international-comparison.pdf)
- Meho, L. I., & Yang, K. (2007). Impact of data sources on citation counts and rankings of LIS faculty: Web of science versus scopus and google scholar. *Journal of the American Society for Information Science and Technology*, 58(13), 2105–2125. <https://doi.org/10.1002/asi.20677>
- Merton, R. K. (1968). The Matthew Effect in Science. *Science*, 159(3810), 56–63. Noudettu osoitteesta <http://www.garfield.library.upenn.edu/merton/matthew1.pdf>
- Microsoft Academic (2017), Whats new in version 2.0, <https://academic.microsoft.com/#/faq>, Viitattu 26.11.2017
- Mongeon, P., & Paul-Hus, A. (2016). The journal coverage of Web of Science and Scopus: a comparative analysis. *Scientometrics*, 106(1), 213–228. <https://doi.org/10.1007/s11192-015-1765-5>
- Newman, M. (2005). Power laws, Pareto distributions and Zipf's law. *Contemporary Physics*, 46(5), 323–351. <https://doi.org/10.1080/00107510500052444>
- Opetus- ja kulttuuriministeriön julkaisu : 2016: 16 Lyytinen, A., Muhonen, R., Koskinen, O., & Eskola, O. (2016.). *Ulkomaiset tutkijat ja julkaisuaktiivisuus*. Noudettu osoitteesta <http://urn.fi/URN:ISBN:978-952-263-402-3>
- Pan, R. K., & Fortunato, S. (2014). Author Impact Factor: tracking the dynamics of individual scientific impact. *Scientific Reports*, 4. <https://doi.org/10.1038/srep04880>
- Poropudas, O., & Pölönen, J. (2015). Kohti tieteenalojen tasa-arvoa. *Tieteessä tapahtuu*, 2015(3). Noudettu osoitteesta <http://journal.fi/tt/article/view/50771/15448>
- Puuska, H.-M. (2014). Scholarly Publishing Patterns in Finland - A comparison of disciplinary groups. Noudettu osoitteesta <http://urn.fi/URN:ISBN:978-951-44-9480-2>

- Scopus. (2017). Content coverage guide. Noudettu osoitteesta [https://www.elsevier.com/\\_data/assets/pdf\\_file/0007/69451/0597-Scopus-Content-Coverage-Guide-US-LETTER-v4-HI-singles-no-ticks.pdf](https://www.elsevier.com/_data/assets/pdf_file/0007/69451/0597-Scopus-Content-Coverage-Guide-US-LETTER-v4-HI-singles-no-ticks.pdf). Viitattu 20.11.17
- Scopus (2018). Content policy and selection. Noudettu osoitteesta <https://www.elsevier.com/solutions/scopus/content/content-policy-and-selection>. Viitattu 2.4.18
- Sivertsen, G. (2014). Scholarly publication patterns in the social sciences and humanities and their coverage in Scopus and Web of Science. *Proceedings of the science and technology indicators conference 2014 Leiden*, 598–604. Noudettu osoitteesta: [https://s3.amazonaws.com/academia.edu.documents/42806714/Research\\_quality\\_characteristics\\_of\\_publ20160218-10100-1wsawux.pdf?AWSAccessKeyId=AKIAIWO-WYYGZ2Y53UL3A&Expires=1511707328&Signature=BdXwgD%2BA6do-Fik%2F4y5UBx%2BqNGmo%3D&response-content-disposition=inline%3B%20filename%3DResearch\\_quality\\_characteristics\\_of\\_publ.pdf#page=610](https://s3.amazonaws.com/academia.edu.documents/42806714/Research_quality_characteristics_of_publ20160218-10100-1wsawux.pdf?AWSAccessKeyId=AKIAIWO-WYYGZ2Y53UL3A&Expires=1511707328&Signature=BdXwgD%2BA6do-Fik%2F4y5UBx%2BqNGmo%3D&response-content-disposition=inline%3B%20filename%3DResearch_quality_characteristics_of_publ.pdf#page=610)
- Small, H. (1973). Cocitation in the scientific literature: A new measure of the relationship between two documents. *Journal of the American Society for Information Science and Technology*, 24(4). Noudettu osoitteesta <http://scimaps.org/exhibit/docs/Small1973cocit.pdf>
- Suomen Akatemia. Tieteen tila. <http://www.aka.fi/fi/tiedepoliittinen-toiminta/tieteen-tila/bibliometriset-analyysit/>. Viitattu 19.10.2017.
- Thelwall, M. (2008). Bibliometrics to webometrics. *Journal of Information Science*, 34(4), 605–621. <https://doi.org/10.1177/0165551507087238>
- van Eck, N. J., & Waltman, L. (2010). Software survey: VOSviewer, a computer program for bibliometric mapping. *Scientometrics*, 84(2), 523–538. <https://doi.org/10.1007/s11192-009-0146-3>
- van Leeuwen, T (2004) Descriptive versus Evaluative Bibliometrics. Teoksessa Handbook of Quantitative Science and Technology Research by Moed, Glänzel, Schmoch. Springer Netherlands. s. 373–389
- Voos, H. (1974). Brief Communications Lotka and Information Science. *Journal of the American Society for Information Science (pre-1986)*; *New York*, 25(4), 270–272.