INFLUENCE OF TITLE CHARACTERISTICS IN SCIENTIFIC LITERATURE ON TWEETING BEHAVIOUR

Lieu Phuong Phuong

Master thesis in the Information and Knowledge Management

Master's Programme

Supervisors: Kim Holmberg and Gunilla Widén

Faculty of Social Sciences, Business and Economics

Åbo Akademi University, 2017

Abstract

This study investigates the impact or influence of research products in the field of arts and humanities by Finnish authors on the public by investigating microblogging site Twitter. The impact is examined by analysing similarities between scientific articles' titles and tweets' content, as well as an exploration deeper into societal impact of tweeting behaviour in the online community. To establish a background for the study, different kinds of metrics from traditional ones such as usage statistics, peer review, and citations to altmetrics for research evaluation are reviewed before overviewing different aspects of science tweeting and features of academic titles such as length of title, type of title, and amusement of title. This study mainly uses exploratory methods and analyses Twitter data shared by a research project at Research Unit of Sociology and Education (RUSE), University of Turku. After filtering and analysing data, the results showed a weak connection between the number of retweets and the level of lexical difficulty of scientific titles in context of public attention. In addition, more than half of the number of tweets are summarization of the titles or contain personal comments instead of only quoting the titles. Moreover, this study revealed that Finnish family matters is the most popular topic in the interdisciplinary field of arts and humanities. In spite of limitations in methodology, this study concluded with discussions that explained implications, similarities as well as differences with previous studies, to present a new potential evidence for approaching advanced research of scholarly communication.

Acknowledgement

First, I would like to express a deep gratitude to two professors as well as my supervisors Kim Holmberg and Gunilla Widén who consulted me enthusiastically and encouraged me to finalize this study.

Kim oriented me dedicatedly right from the first moment since I tried to select my research topic. He provided me dataset to conduct research, assisted me to determine research direction, and gave comments and suggestions during the writing process. There is always a bunch of work on his desk, but the way he enjoys life and simultaneously keep being professional in working inspired me very much.

Gunilla was always there to listen to my thesis presentations. She guided me with all matters of a thesis work, even tiny details such as format, timeline, references. She reminded me in advance of the deadlines of final draft's language checking as well as graduation dates to secure my study plan. In addition, she also has to check up and give comments to other ten fellow students, but she always makes sure the quality of comments as well as opens new possibilities whenever students get stuck.

To my family, especially mom and dad, I want to send sincere thanks and appreciation for their love. Although they live apart from me for a long time, they always wanted me to feel like home in Finland and I also believe that they always stay beside me wherever I am. More importantly, they understood how hard I decided to leave my family to come to Finland where I feel totally free, that not many Asian parents could understand.

Lastly, I would like to thank my boyfriend, also my best friend and my nanny. Whenever I needed, he was always there to cheer me up, to cook for me, to carry me on his back going around. I could not imagine one day without him, I would feel empty.

Thanks for all encouragement and love!

Table of Contents

List of figures and tables

Chapter 1. Introduction

The motivation of this research initially comes from author's personal interest in Arts research and curiosity about the value of academic publications in Arts and Humanities to society. The author later had a chance to receive data from a project called "Measuring the societal impact of open science", by collaboration between researchers of Research Unit for the Sociology of Education, University of Turku and Information Studies, University of Oulu. The project aims to map the online attention of Finnish research from wider audience and develop an open system to collect and detect mentions of research output in Finland from various open sources.

Research assessment has become increasingly essential partly due to the growth in the number of researchers and research products, as well a need by the funders to evaluate research more carefully and strictly because of limited funds. Assessments based on peer review or citations are two methods that have been used widely for evaluating scientific impacts of research, even though both approaches have disadvantages. Peer review is still irreplaceable method in order to decide whether a manuscript is good enough to be published in scientific journals. Citations are used not only for intellectual protection (Moed, De Bruin, & Van Leeuwen, 1995; Borgman, 2000; Cole, 2000; Holmberg & Thelwall, 2014) but also as a part of scientific reward system (Merton, 1968). It is assumed that the more cites a research receives, the higher quality and more valuable it is. However, traditional metrics such as citations and peer review focus on only scientific impacts of research, whereas funders are increasingly curious about societal impact and how academic research proves valuable to society (Bornmann, 2014).

Open science has brought scholarly community to open web, where readers have more chance to read scientific research, share articles and discuss their own opinions. Many open access websites were born to allow researchers to publish their manuscripts without a strict peer review process, as well as give a chance for wider audience accessing these publications for free (e.g., Dynich and Wang, 2016). Open science is not only free access to scientific manuscript, but also about discussing and sharing of scientific ideas in online platforms and social media. Because the amount of scientific research that is available online has been increasing and there are many tracks audience left on online platform and social media from

discussion, altmetrics was born as potential approach to reveal hidden impact of research reflect the online mentions of research products. Whereas citation and peer review has been used to measure scientific impact of research, altmetrics has been investigated to trace these scholarly activities and conversations on social media, in order to understand what societal impact scientific research has. Although altmetrics is still in its infancy and not widely accepted, altmetrics has shown its potential to trace online attention toward research products, and has been studied to investigate more its uses as metrics for research evaluation outside of academia (Holmberg, 2015).

Among social media sites for research discussion, Twitter is still the most popular platform for distributing scientific products (Thelwall, Haustein, et al., 2013). While Mendeley is used mainly by researcher, Twitter has been reaching audiences widely. Twitter is also the second largest source after Mendeley in terms of providing altmetrics data (Haustein et al., 2014). However, these traces of scientific publications on Twitter are still being investigated, especially in art and humanities field. According to Thelwall and Delgado (2015), due to the transparent impacts in art and humanity field, contextual data should be used as impact evidence instead of evaluating research impact by any kinds of metrics. The data in this research is online mentions about Finnish research in Art and Humanities discipline from Twitter.

With an interest in exploring how people on Twitter +react to scientific research in Art and Humanities, the author started to analyse the given data and recognized the content of tweets toward scientific articles are mostly about title of articles or short summary of the articles, shortness perhaps due to the limit of 140 characters for each tweet. This raise a question about if the title is too difficult to understand, will it affect the number of retweets? In other words, the more difficult a title is, is there less attraction to wider audience? As such, this research aims to explore two broad research questions:

(1) Is there any connection between the difficulty of the scientific articles' titles and number of tweets?

(2) How do audiences on Twitter react to scientific publication in art and humanities field of Finnish research?

To answer the first question, the titles of publications or summarization of the articles' content were two typical types of retweets due to the limit of text in each tweet. VocabProfile tool was used to analyse the difficulty of each research title and to examine the correlation between the difficulty and number of tweets, with separation based on the previous coding scheme.

The second question attempts to explore the tweeting behaviour of readers of art and humanity research. The codebook was built to classify the tweets in order to clarify how users tweet and how they make conversation on Twitter.

To sum up the structure of the study, the theoretical part includes introduction and theoretical background, the empirical part comprises methodology, results and discussions before conclusion. The theoretical sections will discuss the relevant earlier research about scholarly communication, open science, metrics for research evaluation, potential of using altmetrics in evaluating art and humanities research, how science tweet and the importance of title in academic research. The empirical research starts by the explanation of methodology including overview of data and the used methods to conduct the study. Next, the findings of the study are presented based on analysing the difficulty as well as length of title and its correlation with the number of tweets. In the discussion chapter, all results will be discussed and reflected into the theoretical framework, literature review, answer the main research questions of this thesis. In addition, the contribution of the thesis as well as some limitations and future research needed are also discussed before finalizing the study.

Chapter 2. Theoretical framework

In this chapter, literature in a large scope is reviewed based on previous relevant research. The first three parts are overview of Finnish research, definition and activities within scholarly communication, and open science movement. The fourth section explains four metrics that are used to evaluate research impact and in which, introducing concept of altmetrics. Next, the potential of using altmetrics in evaluating art and humanities research is introduced. The following section discusses how scholars use Twitter and review the correlation between the presence of research on Twitter and citation counts in various discipline. Finally, four aspects of title which are considered as influencing factors to citation counts of research are given, such as length of titles, types of titles, specific markers as colons, and the level of amusement of titles.

2.1 Scholarly communication

Scholarly communication is the system of creating, evaluating, disseminating and preserving the value of research and other scholarly writings for future use (Principles and strategies for the reform of scholarly communication 1, 1996). This includes both formal communication, such as scientific publications in any formats that researchers present their ownership of intellectual ideas and contributions; and informal communication through conversations during conferences, or communicating via Twitter and other social media channels (Holmberg, 2015). In general, scholarly communication is a process starting with a new research idea or idea that builds on earlier research. After going through academic works and writing, these manuscripts will be assessed by peer review and published if they are considered to be good enough to make a contribution to science. These publications continue then to inspire other researchers and become a source of literature for future research. As a part of this formal scholarly communication, citations are used as an indication of use of earlier work and as contributing to the academic reward system. It is assumed that if a research receives many citations, it reflects higher value and contributes more to science. For this reason, its author could be consider as productive researcher who creates bigger impact than those who with less cited papers.

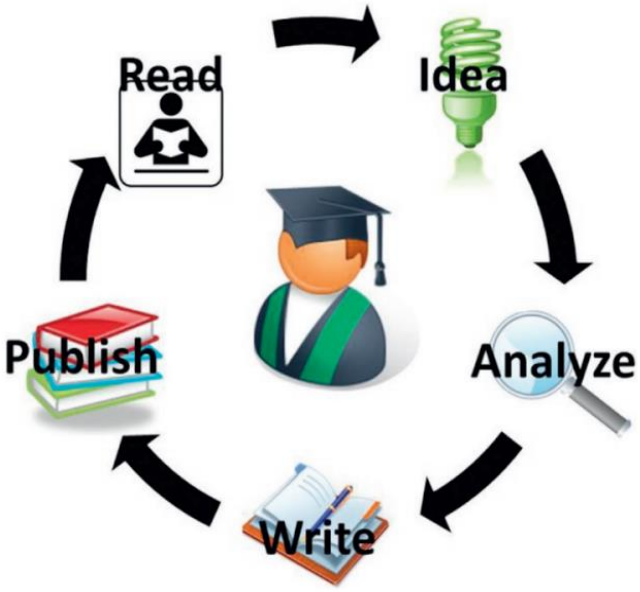


Figure 1. The cycle of scholarly communication (Holmberg, 2015)

For the development of the Internet and electronic format for research products, the dissemination, availability and accessibility of research outputs have been affected significantly (Walsh et al., 2000). Scholarly communication has been moving to the social media platform due to desire of knowledge sharing and public interest. Researchers therefore also engage in using available tools and services from social media in order to productively support the presence of their research online (Thorin, 2006; Widén- Wulff and Gu, 2011).

2.2 Open science

The advent of web 2.0 and social networking sites have changed the way of publishing, disseminating and sharing knowledge within scholarly community and to people outside academia. Along with the transfer from printed academic journals to electronic publishing, a demand from society to access scientific knowledge has increased. The concept of 'open science' was born to encourage scientists and researchers to share scientific articles to spread knowledge to wider audiences. Open science is defined as "the idea that scientific knowledge of all kinds should be openly shared as early as is practical in the discovery process" (Nielsen, 2011). There are still some concerns about what information should be shared for public uses, whether the public could misunderstand science data or science might be exploited for bad things (Galligan and Dyas-Correia, 2013; Ferrara et al., 2014; Holmberg, 2015). Despite of

these concerns, new organizations were born to adopt open science concept, for example, Public Library of Science (PLoS) with a collection of open access journals and scientific literature; F1000 Research with open publishing and open peer review for life-sciences (Wikipedia, 2016). Regarding Finnish research, as stated in reports of Ministry of Education and Culture, Finland (2014), Finland believes that open science and research under legal and ethical operating environments could increase the competitiveness and innovation, bring surprising discoveries and creative insights, and widen the impact of research.

Along with open science, social networking tools are used for exchanging information, building new relations and interacting with other people sharing the same interests, and also for various purposes by people outside academia (Tiryakioglu & Erzurum, 2010; Chen & Bryer, 2012; Gruzd, Staves, & Wilk, 2012; Al-Aufi & Fulton, 2014). Researchers use wikis and online documents to keep up-to-date and search for information, instant messages to interact with others, and other platforms for knowledge sharing (Gu & Widén-Wulff, 2011; Holmberg, 2015). Twitter has been used frequently by scholars for discussion, promoting individual works or finding collaboration (Letierce, Passant, Breslin, &Decker, 2010). Many conferences have been held with attendees discussing, commenting and sharing interesting links on Twitter (Ebner & Reinhardt, 2009; Letierce, Passant, Decker, & Breslin, 2010; Ross, Terras, Warwick, & Welsh, 2011; Weller, Dröge, & Puschmann, 2011; Weller & Puschmann, 2011; Homberg, 2015). Gu and Widen-Wulff (2011) found that scholars in Finland are familiar with using social network tools in everyday life. Although not all researchers actively use these tools for scholarly activities, most of them agree that the use of social media among academics is increasing. The advantages of social media sites have diversified the way of practicing information, interactive communication and expanding international and local collaboration with other researchers. Scholarly communication has shifted to the online world, and more people follow researcher's activities to discover any impact that has appeared on these online platforms between researchers or researchers and audiences.

2.3 Metrics for research evaluation

The movement of open science and the shift of scholarly communication online raise a potential to track and measure online outputs and impacts by identifying and tracking the interactions surrounding scientific outputs. Priem, Taraborelli, Groth and Neylon (2010) stated four methods for evaluating research impact, from traditional to emerging way.

Figure 2. Four ways to measure impact of scholarly products (J. Priem et al. 2010)

2.3.1 Usage statistics

Assessing research impact based on usage statistics looks at how many times the publication has been viewed or downloaded. The biggest advantage of this measure is that it can immediately show the impact of research as soon as it has been published online (Bollen, Van de Sompel, Hagberg, & Chute, 2009)

However, there is no way to know whether a reader actually read an article after downloading it or not (Neylon & Wu, 2009). Moreover, Haustein &Siebenlist (2011) pointed out that many publishers do not share their usage data, except some cases for specific collection of articles. Earlier research has, however, found a correlation between download counts and later citations (Brody, Harnad, & Carr, 2006; Watson, 2009; Bollen et al., 2009). This is explained that when an article has been downloaded, there is more chance to be read and leading to possibility to be cited later.

2.3.2 Peer review

Peer review is the process of evaluating manuscripts by researcher's peers, who usually are experts in the same area of research. After author writes article and submits to journal, editor sends copy to a peer (expert in the field) for reviewing. Before paper is accepted and published, it has been check carefully for accuracy and assess the validity of the research methodology and procedures (Elsevier, 2016). Despite criticisms, peer review is still the only widely accepted method for research validation across different disciplines. It has been a formal part of scientific communication since the first scientific journals appeared more than

300 years ago. There are three main types of peer review, namely single blind review, double blind review, and open review. Single blind review is the most common type so far, with the name of reviewers hidden from the author; whereas both reviewers and authors are anonymous to each other in double blind review. Open review is, as the name suggests, completely open. Some referees and authors believe it could prevent malevolent comments, stop plagiarism, and encourage honest review (Elsevier, 2016). Another opinion is that it would be less likely to be honest. For example, junior researchers could fear to give open and honest review to famous scientist's works in order to protect their own career and funding (Mulligan, 2005). In general, single-blind review has still been the most popular method since 1750s (Spier, 2002).

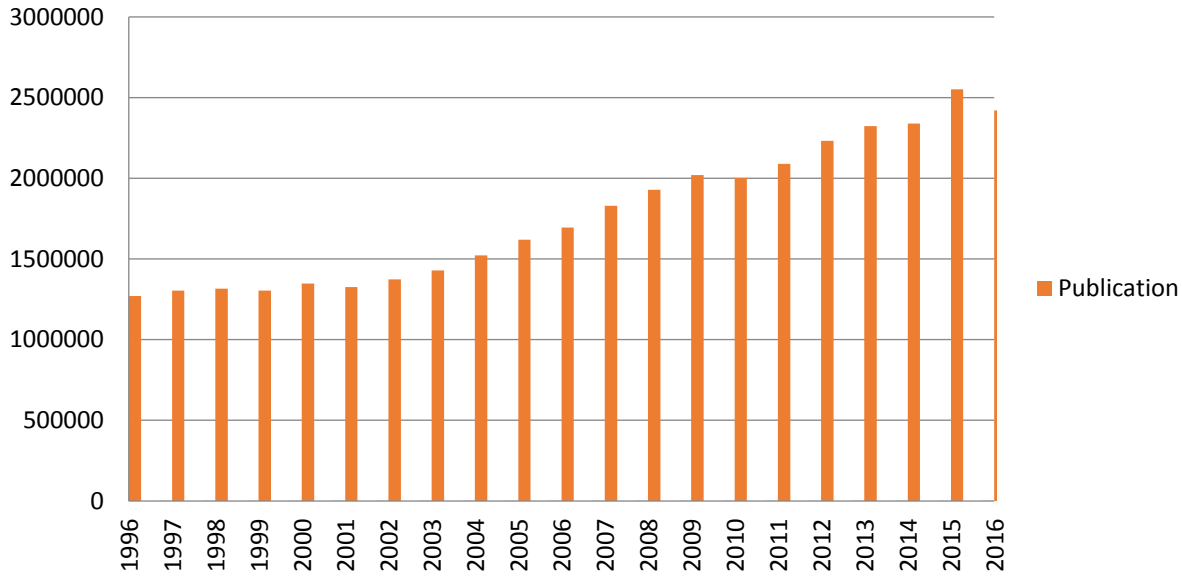


Figure 3. The increase of publication as archives by Web of Science between 1996 and 2016 (data retrieved from web of Science, February 26, 2017)

With a double increase in number of researchers and scientific publication during the past ten years, using peer review alone to evaluate this long queue of manuscripts may be impossible. Due to time-consuming and costly process, hence it cannot be used at a large scale (Holmberg, 2015). However, peer review still remains as a reliable method for vetting of knowledge claims (Lee et al., 2012). As such, a combination of using peer review and other bibliometrics indicators is necessary to assess impact of research in terms of both quality and quantity, especially at large scale.

2.3.3 Citation impact

Bibliometric is a statistical measure used to analyse the quantity and performance of publications. In 1969, Pritchard introduced the concept of bibliometric as "the application of mathematics and statistical methods to books and other media of communication" (p.2). Since then bibliometric has developed into a set of methods to assess scientific productivity and impact (Holmberg, 2015; Yu, Wu, Alhalabi, Kao & Wu, 2016).

In the early 1960s, Eugene Garfield from ISI (Institute for Scientific Information) came up with the idea of journal impact factor, by counting how many times articles in specific journal have been cited on average in the previous two years (Garfield, 2006). Since then, bibliometric indicators, mainly based on citation, have become the most popular tool for research evaluation. However, citation-counting metrics are criticized as they ignore differences among research fields and cannot alone measure the quality of the publications (Seglen, 1997). In 2005, Hirsch addressed the problem of bibliometric indicators that used only the total number of papers or only total number of citations to evaluate the research. Both of these statistics should be use together. Hirsch therefore introduced h-index for evaluating scientific performance of researchers by trying to take both quality and quantity into account. If a researcher has h number of publications receiving at least h citations each, then h-index of this researcher is h. For example, if researcher A has published 7 publications that all have more than 7 citations, then his h-index is 7. A high h-index would mean that a researcher has published many highly-cited papers. H index has been used, studied and criticized widely among academics, and it has become an inspiration for developing other citation-based indicators (Burrell, 2007; Abramo, D'Angelo, & Viel, 2013; Ferrara & Romero, 2013; Schreiber, 2013). Despite many arguments, citations have been considered as a tool to measure the scientific impact and even for evaluating the quality of research.

To respect and protect the ownership of original ideas in scholarly communication as well as to avoid plagiarism, researchers have to reference and give credit to prior work by citing articles they have used to build their own work upon. Apart from protecting intellectual ownership, citations are used to capture the accumulated knowledge, collaborations and flow of ideas in science (Moed, De Bruin, & Van Leeuwen, 1995; Borgman, 2000; Cole, 2000; Holmberg & Thelwall, 2014). However, scholarly communication slowly turns an idea into

a complete research product (Holmberg, 2015), especially if it requires collaboration by many researchers from different institutions worldwide. After successfully publishing, it may take even longer time for people to recognize the value of the research and cite. Citation measure for this research then also need some more time to accumulate enough citations for analysing. As such, it would take at least few years after a research published in order to get reliable citation evaluation. Such a slow process makes research ideas become outdated at the moment that value of research can be evaluated, as well as only showing value of research in the pass instead of current research.

Apart from protecting intellectual ownership, citations are part of scientific reward system (Merton, 1968), with highly cited research being considered as more valuable and its author as having more influence than those who do not receive as many citations. This strongly applies in choosing candidates for academic positions. Regardless of teaching quality, a teacher with greater publications is often evaluated higher than those who have better teaching quality but poor research outputs (Holmberg, 2015).

2.3.4 Altmetrics

Since the 1990s, as competition for research funding has increased, funders are not only interested in scientific impact of research but also of the wider societal impact, which varies from social, cultural, environmental, and to economic returns (Bornmann &Marx, 2013). For instance, Finland has recently focused on the importance of wider impact of scientific research in society, with specific on four research fields varied from (1) ecology, evolutionary biology and ecophysiology; (2) history; (3) material science and technology to (4) medical engineering and health technology. The impact analysis will attempt to capture the research impact beyond academia by using surveys and group interviews to collect information of influences, interactions and broader impacts of research (State of scientific research - Academy of Finland, 2016). Societal impact evaluation, however, is still in its infancy. While scientific impact can be measure by peer review and bibliometric methods and are widely accepted and used, there are no widely recognized established approaches for assessing societal impact of research yet. Bornmann & Marx (2013) stated that societal impact occurs when the content of a scientific article is address outside of academia and it could be measured by tracking attention that scholarly articles receive online. Altmetrics, short for

alternative metrics, is an interesting option to assess the societal impact of research. Altmetrics is defined by Shema, Bar-Ilan & Thelwall (2014, p.1019) as a "web-based metrics for the impact of scholarly material, with an emphasis on social media outlets as sources of data". In other words, altmetrics count all online attention related to the articles from views, downloads, saves, tweets, likes, bookmarks, comments to blog posts and online discussions. Priem & Hemminger (2010) summarized that there are seven major platforms in social media to track the impact of research, namely bookmarking sites, reference managers, recommendation services, comments on articles, microblogs, Wikipedia, and blogs (Bornmann, 2014). Whereas citations or journal impact factor take months or even years to accumulate to see the value or impact of the research, altmetrics occur fast and allow researchers to see, almost in real time, how an article or other research outputs are used. One more potential advantage of altmetrics is that it captures data from a variety of sources, not only in traditional academic publishing setting. In a way, this reflects 'real' connection to wider audiences beyond scholarly communication.

Earlier studies have found that there is an association between higher altmetrics counts and number of citations (Shema, Bar-Ilan & Thelwall, 2014), and between altmetrics counts and article views (Wang, Liu, Fang & Mao, 2014). Article views include abstract views, HTML views and downloads, in which pdf download counts has the most significant correlation with citations (Wang et al., 2014). However, these studies of correlations between citation counts and altmetrics might not be useful as the high association between them addresses the similarity of their functions for evaluating impact of research, which is not the goal of altmetrics. Altmetrics was born to investigate new kinds of impacts on wider audiences. In other words, the less altmetrics indicators correlate with citations counts or other bibliometric indicators, the more opportunities to discover uses of altmetrics there are beyond scholarly impact (Holmberg, 2015).

There are some applications that offer altmetrics views. Altmetric.com (*http://www.altmetric.com*) is a service that captures tweets, blog posts, news and other contents on different social media platforms and in the web in general that mention research outputs. Plum Analytic (plumanalytics.com) focus more on delivering traces of attention to institutions and research managers. Another service is ImpactStory (*www.impactstory.org*), which tracks the volume of discussions (good or bad) around research (Bornmann & Marx,

2013). There is a difference in the level of impact or level of engagement that different social media activities can reflect. Holmberg (2015) discussed the association of the level of engagement with each online activity and he suggested that the number of tweets, retweets, likes and shares are likely to have less impact than the number of downloads or bookmarks, while a blog post or news story which might generate more conversations could be considered as stronger indicator.

Apart from its potential, there are also many critics against altmetrics. Many researchers are sceptical and afraid of changing from traditional metrics such as citations to altmetrics, as they are strongly tied to academic reward system (Weller, M., 2011). Some have argued that altmetrics could easily be manipulated by software robots (Galligan & Dyas-Correia, 2013; Haustein, Bowman, Holmberg, Peters, & Larivière, 2014). Even though the programs for detecting those robots are also becoming more advanced and more able to decrease the threat of manipulation (Ferrara, Varol, Davis, Menczer, & Flammini, 2014) as well as there was lately a two-phase initiative from NISO Altmetrics Initiative to explore, identify, and standardize new potential metrics (Verma, 2016), altmetrics is still lacking of standards compared to citation metrics or peer review. No one can guarantee that all the social bots will be detected. The situation could be even worse if altmetrics would be standardized and considered as an evaluating factor for research funding, since people will probably start manipulating them (Holmberg, 2015). The second problem with altmetrics is that there is no firm understanding about what altmetrics mean and how the data collected should be interpreted. Unlike citations, altmetrics does not have an established score or a standard that apply for all altmetrics providers (e.g. altmetrics.com, plumanalytics.com, impactstory.org). The other problem is that the data sources lack in stability. If you for instance measure the number of tweets on Twitter or blogposts, they could change or disappear by time and would be impossible to retrieve; leading to fluctuating altmetrics score (Haustein & Siebenlist, 2011). It is undeniable that there are many defunct social media sites that are no longer used widely, such as Google Wave, Orkut, Hyves, Six Degrees and sites getting more niche and targeted such as MySpace and FourSquare. So it is difficult to determine which platforms would be stable and reliable over time. It is even possible that platforms such as Facebook, Twitter or Mendeley could be shut down some day and all data on them might disappear (Holmberg, 2015).

In general, altmetrics raises both interest and critics. It was born along with the movement of open science when general public has more chance to access scientific research for free, and with development of social media sites which allow open- discussion on scientific publication worldwide. It has been found that in some cases online visibility of research leads to more citation counts, proving that altmetrics also reflects similar aspects towards research products that citations do (Holmberg, 2015). On the other hand, beyond scholarly impact, altmetrics could reflect other aspects of societal impact such as public interest or attention towards research. In future, although current social media platforms are no longer existed or replaced by new ones, altmetrics (alternative metrics) is still able to develop as long as with the advancement of open science movement.

Altmetrics is still in its infancy and has no established standards that are widely accepted as other bibliometric such as citations do. It has diverse source of data, worldwide users and tremendous amount of content generating every single second toward research products. The mentions could be self-promoting or automated bots trying to manipulate altmetrics score. For such reasons, before we fully understand its potential, altmetrics should play a role as a complement to traditional metrics in observing public attention towards research product, rather than trying to replace them.

2.4 Potential of using altmetrics in evaluating art and humanities research

The definitions of arts and humanities are fuzzy. This heterogeneous domain generally covers religion, philosophy, arts, music, literature, linguistics, archaeology and history (Hellqvist, 2009).

There are many problems in evaluating impact of arts and humanities research with established bibliometrics methods such as citations. The main reason is because for social science as well as arts and humanities, there is citation data only for Psychological and Cognitive Science (Thelwall & Delgado, 2015). As summarised by Ardanuy (2013) from studies that have used citation analysis in humanities from 1951 to 2010, Web of science, Scopus and even Google Scholar are not good at covering social sciences or humanities. Differences in publishing behaviour mean that citations are not widely used in humanities. In addition, research publications in arts and humanities are mostly non-English and non-journal. As stated by Research Excellence framework (REF) 2014 in Main Panel D and sub-panels 27-36 artists produce a wide variety of 34 different types of outputs in the field of arts

and humanities, namely "advisory reports and evaluations, books (authored and edited), chapters in books, journal articles, published conference papers, electronic resources and publications, exhibition catalogues, translations and scholarly editions, compositions and musical scores, creative writing (libretti, film scripts, radio plays, novels, short stories, stage plays), databases, grammars, patents, digital and broadcast media, performances, films, video and media presentations, installations, designs and exhibitions, software design and development, working papers". These may generate meaning and knowledge, encourage practical applications, and contribute on happiness and democracy (Small, 2013), which influence the society at a subtle level which cannot be evaluated by traditional measures using citations or peer review.

The impact from arts and humanities research is not transparent or as visible as other types of research, such as medical research or technological innovation. As stated by Thelwall and Delgado (2015), we should not measure the social impact of arts and humanities research with traditional metrics but use contextualised data instead. In order to understand the wider impact of research, data itself should be used as an "indicator". For example, 'publication and sales figures, external funding, evidence of use of educational materials, tourism data, and business growth figures, such as income, or employment, critiques or citations from users, public engagement data (including numbers and descriptions), policy engagements, independent testimony and formal evaluations' (Thelwall & Delgado, 2015). Therefore, altmetrics alone is not reliable alone, but it could provide data for research evaluation in arts and humanities.

Altmetrics could help to capture a wider societal influence of research beyond citations that capture scholarly impacts. However, it is undeniable that the problems of using altmetrics in arts and humanity also relate to issues already addressed in the bibliometrics approaches. For instance, the diverse publication channels in arts and humanities, the reliance on print instead of e-publication as well as the low coverage of non-English research, are all challenges for altmetrics too. However, compared to citation-based metrics, altmetrics are still more diverse and able to capture real time activities related to research. As stated by Thelwall and Delgado (2015), altmetrics data could be used to understand how arts and humanities research reach the public, not for counting numbers of tweets or Mendeley readers, but to carefully look at

how scholarly publications are tweeted and blogged, analysing content and motivation for tweeting and blogging.

2.5 Science tweeting

Not only wider audience mention scientific research in social media for various reasons, but researchers within scholarly communication also promote their products online and seek for collaboration by using social media. It has been found that there is a relationship between social media visibility of scientific research and citations, revealing its potential as a rapid data source for analysing research impact (Eysenbach, 2011, Shuai, Pepe, and Bollen, 2012, Bar-Ilan et al., 2012, Thelwall, Haustein, et al., 2013). It could be assumed that if an article is mentioned frequently on Twitter or get a high number of retweets, it is more valuable and create more societal impact compared with publications with less social media visibility. Twitter is used by scholarly community for following discussion and posting work-relate content (Holmberg, 2015), and by general public for updating daily life events and activities, plus searching and sharing information (Java et al., 2009, Zhao, Rosson, and Beth, 2009). Twitter allows users to choose who they want to follow and what kind of tweets they want to feed, for that reason, network on Twitter tend to be clustered around with those who share the same interests. Even though not everyone is online and uses Twitter regularly, population on Twitter keeps growing rapidly with an ease of access to data. For these reasons, Twitter has gained a lot of attention as a source of data from researchers in many disciplines and lately become the most popular social media sites (Holmberge, 2015). Twitter may become a useful metric to measure attentiveness of public to a specific scientific publication. In other words, via Twitter we could know which topics are popular and attract attention from the public, even they did not receive many citations. It may support funding organization, journal editors, and research institution in making funding decision which topics resonate and trendy enough with the public, particularly Twitter population (Eysenbach, 2011).

2.5.1 Science tweeting- motivation and behaviour

Microblogging allows people to post concise public messages through social media platforms to potentially large audiences (Mollett, Moran & Dunleavy, 2011). Twitter is one of the most popular microblogging sites that has been used widely not only for sharing information but also in promoting activist groups and online campaigns (Jansen, Zhang, Sobel, & Chowdury,

2009). It has been proven as an efficient tool to spread online campaigns during political elections (Vergeer, Hermans, & Sams, 2011) or climate activism (Bennett & Segerberg 2011). There are only a handful of studies about how Twitter is used among scholars (Priem & Costello, 2010; Weller, Dröge, & Puschmann, 2011), however, most of them found that Twitter is not considered as a popular tool for scholarly dissemination. Although Twitter data for altmetric events was more extensive than data from other social media platforms (Thelwall, Haustein, Larivière, & Sugimoto, 2013), this bigger usage is less likely to contribute to a scholar's reputation (Cruz and Jamias, 2013). As studied by Kwak, Lee, Park, and Moon (2010), if a tweet gets retweeted, it could reach to more than 1000 users on average, which is a potential approach for wide information dissemination. There is, however, evidence that despite a large population of tweeters, the connections between individuals are not as strong as on other social media platforms, such as Facebook (Huberman, Romero, & Wu, 2008). Twitter is more information-based, for instance, during academic conferences, it is frequently used to share information about the event, comment on and share individual talks (Weller, Dröge, & Puschmann, 2011).

Less than 10% of researchers use microblogging (Rowlands, Nicholas, Russell, Canty, &Watkinson, 2011), while only 2.5% of scientists are active on Twitter (Priem, Costello, & Dzuba, 2011). When scholars tweet, almost 50% of the tweets are related to scholarly communication (Chretien, Azar, &Kind, 2011; Holmberg & Thelwall, 2014). There are three kinds of special characteristics or affordances on Twitter; hashtags, directed @messages and retweets (Boy, Golder & Lotan, 2010). Hashtags combine with a key word to indicate the topics and label the tweet. @messages are used to message directly to specific people (e.g.@amandapalmer) or reference other users (e.g. "I saw@oprah's show today"). Unlike @message and hashtag, retweeting is much more inconsistent. Ideally, a retweet should copy the original tweet, precede it with RT and mention the author with @. However, people usually change the content when retweeting and therefore lead to difficulty to track the spread of retweets. Retweets are contributed as "internal citations", while "external citations" appear when the users link to another website or outside information (Weller & Peters, 2013). Tweets are more likely to be retweeted if they contain links; and it has been showed that almost a third of academic tweets contain URLs (Peters, Beutelspacher, Maghferat, & Terliesner, 2012), even though it could vary greatly among disciplines (Holmberg &

Thelwall, 2014). These links generally lead to blogs or mass media channel, such as news or video sharing platforms, and to open publication website (Weller, Dröge, & Puschmann, 2011; Weller & Puschmann, 2011; Peters et al., 2012; Holmberg & Thelwall, 2014).

2.5.2 Tweeting and citations

Haustein and her co-authors (2013) conducted a large-scale study and found that "less than 10% of more than 1.4 million articles in WoS and PubMed are tweeted" (p.14). When articles were tweeted, each will be tweeted two and half times on average. Correlation between Twitter coverage and traditional citation rates were positive. Correlation between Twitter citations and formal citations for all publications in 2011 was low but positive. In other words, these two indicators appear to be somewhat related, but probably measure different kinds of impacts. As such, authors conclude that altmetrics should not be considered as alternative to traditional citation-based indicators, but as complementary. The results from Haustein et al. (2013) also showed that these correlations are lower than the relation between other metrics (readership and mentorship) and citations, as well as lower than the correlation between tweets and other scholarly metrics, such as Google scholar citations and downloads. They also concluded that highly tweeted articles were either related to health topics or contained humorous or curious contents, which means that tweets do not necessarily reflect intellectual impact.

In 2013, Thelwall and his co-authors conducted a content analysis study of how 270 tweets are linked to academic articles in four journals (PLOS ONE, PNAS; Science, and Nature). The authors found that tweets mainly contain article titles (42%) or a brief summary (41%). The summaries seem like more understandable for wider audience to read. The majority of tweets do not link to the name of authors, but some do and some are self-citations. Most of the tweets are neutral and have no comments along with them, except some tweets expressing excitement. For these reasons, while tweet counts might reflect popularity of scientific articles, the content of the tweets are unlikely to reveal insights about the reactions of readers, except in some special cases. However, the results were concluded from a small sample of tweets, and only from articles from four journals. There is still lack of studies about why scholarly articles are mentioned on Twitter and why some articles are tweeted more than others, particularly in arts and humanities.

Regarding humanities discipline, Holmberg and Thelwall (2014) found that scholars in digital humanities use Twitter for conversations with wider audiences. It has been stated previously that humanities journals are often used in conversational and negotiating manners (Hammarfelt, 2014; according to Hellqvist, 2009). In addition, tweets about science are mainly either title or its summary.

2.6 The power of title

Title is an important part of a scientific article that informs readers about the content of the paper and encourages them to read it (Subotic and Mukherjee, 2013). It also helps to attract public attention and affect research articles' impact (Wang and Bai, 2007). In the computer age, for the success of searching an article among thousands of others, it is necessary to include some key words relating to the research into the title (Hartley, 2008). According to Ball (2009, p.668), "a title should motivate the reader to read an article, give the readers a summary of the contents, give overview of the topics and finding discussed, and introduce the way in which the reported items are looked at". However, it is not easy to fulfil all of those functions as most titles are attractive but not informative, or vice versa, informative but not attractive (Kane, 1988).

More than 70% of citers re-cite the articles from lists of references just based on titles and without reading them or just skimming through them (Simkin and Roychowdhury, 2002). Previous research focused mostly on the correlation between citation counts and title characteristics, such as the length, title types, specific markers as colon, or extreme attention grabbing such as humour (Subotic & Mukherjee, 2013; Didegah, 2014). However, these prior studies are debatable, disciplines studied are limited and it is not reasonable to assume that simply changing the title will attract more citations.

2.6.1 The length of titles

The title length has been studied to discover its correlation with citation rates; however, the findings are highly contradictory and differ between disciplines and fields. Letchford, Moat and Preis (2015) investigated 20,000 most highly cited papers from 2007 to 2013 and found that shorter titles received more citations. However, this correlation depends on how famous the journals are, as papers published in several journals attracted more citations than papers published in others. They explained that shorter titles may be easier to understand and then

attract more readers, as well as high-impact journals might have a restriction on the length of their articles' title. Sharing the same results, Subotic and Mukherjee (2013) stated that short titles at around 10 ± 4 words received above median number of citations. Authors concluded that shorter titles, at least in psychology discipline, are easier to digest and therefore, attract more citations. Didegah and Thelwall (2013) also concluded that title length associated with decreased citations in Biology and Biochemistry and Social Sciences, whereas no significant association was found in Chemistry and weak positive correlation was found in Psychological articles. Generally, titles should be informative enough and reflect article's content. Title length is not an important factor to affect citations.

In contrast, in another study on 9,031 articles from 22 scientific journals in medical field, Habibzadeh and Yadollahie (2010) concluded that longer titles seem to be associated with higher citation rates. They explained that "longer titles are mainly those which include the study methodology and/or results in more detail and thus, attract more attention and citations" (p.169). They also suggested conducting more content analysis of the titles to understand what makes them longer; result, information about methods, expression of emotion, or other types of writing styles. Yitzhaki (1994, 2002) concluded that longer titles are cited more because they are normally bigger projects, more results have been found from that, more keywords included and therefore, more findable and visible in databases and thus they enable wider readership.

2.6.2 Types of titles

The number of substantive words in article titles in humanities research is lower than that found in journals in sciences and social sciences between 1940 and 1960; even the titles have become more and more informative in humanities (Yitzhaki, 1997). Yitzhaki (1997) supposed that a title is considered informative if it can help readers get the idea of what the article is about. This informative concept has been analysed based on the number of substantive words, which do not include stop words such as articles, prepositions, conjunctions, pronouns and auxiliary verbs. Yitzhaki also found in 1994 and 2002 a positive correlation between the title length, number of authors and word counts of articles. Another study in biomedical articles found that the articles with method-describing titles were cited less than result-describing titles (Paiva C.E, Paiva B.S.R, and Lima, 2012).

Jamali and Nikzad (2011) investigated more than 2,000 articles in six PLoS journals and analysed them based on the types of title, namely declarative, descriptive and question-title. Declarative titles summarize the finding of articles, whereas descriptive titles only describe the topic without adding any conclusions, and question titles indicate the subject of the paper in a form of questions. They found that even the number of articles with question mark in titles has increased four times within last 40 years in physics, medicine and life sciences, however, they are downloaded more but cited less in comparison with articles with descriptive or declarative titles. Subotic and Mukherjee (2013) supports this by the discovering that descriptive titles were the most common way to name psychology articles.

2.6.3 Specific markers as colons or question marks

According to Ramana, Jain, and Howlett (2013), all style of titling depends on individual's writing style. They added that, some authors try to explain titles in simple way, while others prefer complex ones; some writers use many commas, while others use both commas and colons; sometimes, authors also use brackets for additional explanations, or even subtitle, some emphasize their title with a question mark or exclamation mark to stimulate readers. Therefore, different individual writing styles affect their titling behaviour, and this may also affect the way authors want to deliver messages to audiences. Another studies mentioned that there was an increase in title length and using colon from 1981 to 2001 (Lewison and Hartley, 2005). Hartley (2007) found that there was a greater use of colons in arts than in science. He concluded that even though students and academics often prefer putting colons into titles, however, using colons in the titles of scientific articles, particularly in psychology discipline, had no effect on citation rates. However, in medical and life science discipline, Jamali and Nikazad (2011) found that titles with colon were longer and received fewer downloads and citations. The disciplines also vary in different ways of titling, and this finding of Jamali and Nikazad in 2011 implied that the long titles with colon were not attractive to people in citing or in usage statistics.

Ball (2009) concluded from his study of nearly 20 million scientific articles that titles with question marks increased significantly from 50% to 200% from 1966 to 2005. From his conclusion, the use of question mark in the title is considered as a marketing trick, which might appeal more to the readers and attract them to read. In overall, even there has been a

growing use of special markers such as colon in scientific research's titles, it has no or little influence on how the research be cited later.

2.6.4 Amusement or humour of titles

Humor has been considered as attention grabber, which is applied widely in marketing and even now in scientific titles to attract readers. Previous study found that humorous illustration in the text book for undergraduate students was associated positively with the level of enjoyment, but negatively associated with the credibility of the author and the persuasiveness of the text (Bryant et al., 1981). This finding suggested that humorous titles may harm the credibility and make readers treat its content less seriously. Sagi and Yechiam (2008) stated that the purpose of using humour in in scientific titles is to stimulate the audiences, but the title also has to make sense in a point of view to convince the audiences. On the other hand, they reviewed earlier research and indicated that the use of humour in scientific titles sometimes have the "side-effects" that reduce the tendency to read through the article or underestimate the quality of the article. In addition, they found from articles in psychology that number of citations is weakly associated with pleasantness of title which is "giving a sense of happy satisfaction or enjoyment; friendly and likeable" (p.681); and articles titled amusingly which cause readers to laugh or smile may receive only a few citations.

However, even though pleasantness seems to connect with amusement, only amusement feature was negatively associated with articles citation. They explained that articles with humorous title include less professional key words and thus do not provide enough information for viewers. Another reason is that as Bryant and his colleague found, humorous content could destroy the credibility of the paper as readers consider it as a signal of low quality. On the other hand, Subotic and Mukherjee (2013) showed that the level of amusement did not influence the level of citations, but that the articles with amusing titles were downloaded slightly more often. The authors conclude that humorous titles do not necessarily contribute to the quality of articles, but simply communicate better and they attract more attention and spark curiosity.

Chapter 3. Methodology

The aim of this study is to investigate the connection between research titles and tweets, that lead to a reflection of tweeting behaviour of public towards scientific products. This chapter describes data and methods used to achieve the research aim. Data section tells how the author obtained the raw data, then process of filtering data as well as choosing sample selection for the study. Methods section explains how the two core methods of title analysis and tweet analysis work and apply in this study. While title analysis analyses how difficult the titles are, tweet analysis codifies different kinds of the whole tweets based on their content. Due to only textual data used in this study, the main characteristics of methods are lexical exploratory as well as textual content analysis, plus correlation function to examine a connection between titles and tweets. As such, supportive tools were used in those methods such as VocabProfile program and Microsoft Excel.

3.1 Data

The data was provided by Research Unit for the Sociology of Education (RUSE), University of Turku, Finland, and it was extracted from their on-going research project "Measuring the societal impact of open science". This project investigates societal impact between open science, specifically Finnish research products and public by examining altmetrics data. In addition, the research project classified their data in different sections including arts and humanities which is relevant to the research field of this study in context of social networking analysis. The raw data comprises of mentions of research articles written by authors of which at least one had a Finnish affiliation on three social media platforms; Twitter, blogs, news; however, this study only use Twitter data for research. The research thus focuses on the attention Arts and Humanities research in Finland has received online.

The Twitter data includes 1023 Finnish scientific articles written in English in field of arts and humanities. According to the data, 934 of total 1023 articles were tweeted 4243 times (including retweets) on Twitter. In addition, the data contains records of articles with a DOI (Digital Object Identifier), altmetric attention score (calculated by Altmetric.com), author, title, journal, publisher, Twitter links, post's summary, publish time and other citations data.

Due to the large Twitter data, 200 articles were selected randomly that were mentioned on Twitter 989 times, making the sample for this study. Of the randomly selected 989 tweets, most were in English (887 tweets), the rest is various languages shown in Table 1.

Table 1. A summary of language of tweets

Language of tweets	Total (989)
English	887
Japanese	28
Finnish	18
Dutch	10
French	10
German	10
Spanish	9
Swedish	6
Norwegian	5
Chinese	2
Korean	2
Italian	1
Turkish	1

Additionally, Table 2 shows an overview of status of tweets' links. In general, most links still exist (around 89%), some links were no longer existing (around 9%) at the moment the author checked them. The rest was not found for different reasons stated on Twitter site such as "This account's Tweets are protected", or "Account suspended". Fortunately, content of all 989 tweets could still be used due to the availability of all posts' content in the raw data.

Table 2. Status of tweets' links

	Number of tweets	Percentage
Link is no longer existed or could not be found	92	9,3%

Active link	881	89,08%
Private account	12	1,21%
Account suspended	4	0,4%
Total	989	100%

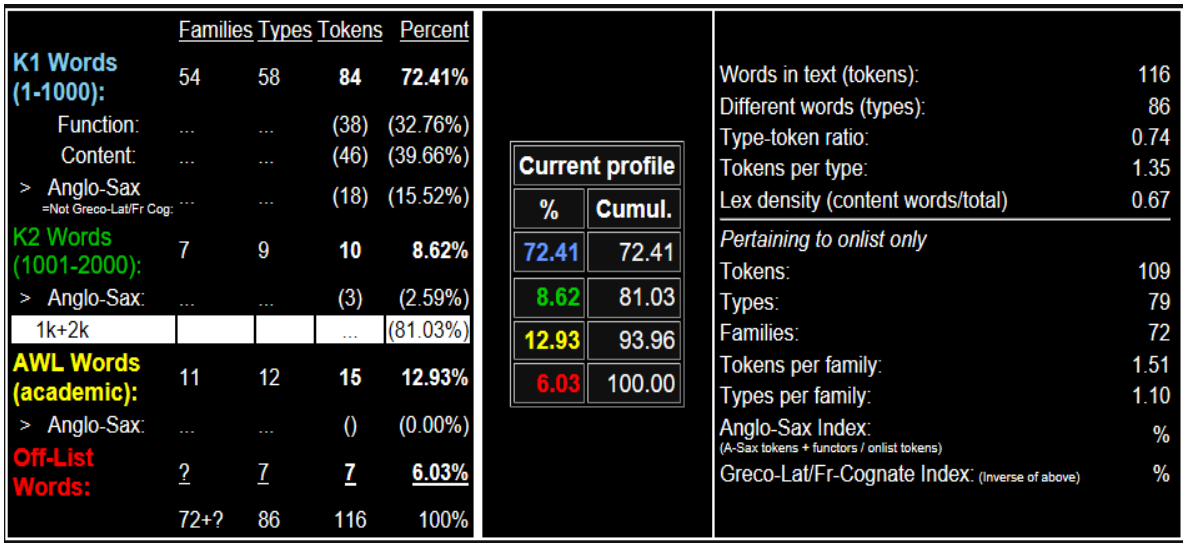
3.2 Methods

This section explains methods used in this study. The selection of methods based on the availability in analysing texts in titles and classifying types of tweets to explore findings that possibly answer research questions. It begins with title analysis with VocabProfile which was reviewed on its own website as a vast range of resources for analysing vocabulary and grammar, especially in English texts. This free computer program performs lexical text analysis, which will tell how many words the title contains from the following four categories: (1) percentage of academic words, (2) percentage of most frequent 1000 words, (3) percentage of most frequent 2000 words, and (4) off-list words (words that do not appear on the other lists). Next, in the tweet analysis or tweet content analysis, tweets are classified into three classes manually with support of VocabProfile: class 0 for tweets citing exactly the article title or that contain some additional information about authors or publishers, class 1 is similar with 0 but attaching hashtag or @ and personal comments, and class 2 for tweets that are totally different from article titles. The purpose of running these two methods is to explore similarities between article titles and tweets.

3.2.1 Title analysis

One of the core methods in this study is to analyse the level of difficulty of the chosen articles' titles. VocabProfile (VP) (http://www.lextutor.ca/vp/eng/) by Tom Cobb performs lexical text analysis and breaking texts down by word frequencies. This counts total number of words, number of different words, type-token ratio which means the higher number of this, the fewer times a word is re-used (e.g. type-token ratio of 0.5 means the analysed text occurred twice; similarly, the ratio of 0.25 means the word occurrences are 4). Type-token ratio is useful to determine common words, word families, common topics as well as hot topics among different analysed titles.

In addition, VP looks at four categories: (1) the percentage of academic words (AWL, in total of 550 most frequent words in academic texts in English across various topics), (2) the percentage of words in the most frequent 1,000 words in English and (3) the second most frequent 1,000 words in English (i.e. 1001 to 2000 words), and (4) off-list words which mean the rest words. Those categories are always calculated in percentage and these are the main criteria for the title analysis in this study. The calculation for all of the categories is executed, and the results are shown and compared at the same time. However, larger texts may affect the accuracy of the results, while short texts are recommended to test with this feature. Figure 4 shows typical results after analysing texts with VP.



	Families	Types	Tokens	Percent
K1 Words (1-1000):	54	58	84	72.41%
Function:	...	...	(38)	(32.76%)
Content:	...	...	(46)	(39.66%)
> Anglo-Sax =Not Greco-Lat/Fr Cog:	...	...	(18)	(15.52%)
K2 Words (1001-2000):	7	9	10	8.62%
> Anglo-Sax:	...	...	(3)	(2.59%)
1k+2k			...	(81.03%)
AWL Words (academic):	11	12	15	12.93%
> Anglo-Sax:	...	...	()	(0.00%)
Off-List Words:	?	7	7	6.03%
	72+?	86	116	100%

Current profile	
%	Cumul.
72.41	72.41
8.62	81.03
12.93	93.96
6.03	100.00

Words in text (tokens):	116
Different words (types):	86
Type-token ratio:	0.74
Tokens per type:	1.35
Lex density (content words/total)	0.67
Pertaining to onlist only	
Tokens:	109
Types:	79
Families:	72
Tokens per family:	1.51
Types per family:	1.10
Anglo-Sax Index: (A-Sax tokens + functors / onlist tokens)	%
Greco-Lat/Fr-Cognate Index: (Inverse of above)	%

Figure 4. An example of vocabulary profile output for texts by VocabProfile

Because VP does not allow input of several texts at once, each title was input separately and then separate results were produced as above for further analysis.

3.2.2 Tweet analysis

The main idea is to find correlation between number of tweets and the title difficulty. However, tweets were retweeted differently, such as repeating the exact title, or commenting own opinion, or using hashtag and direct message along with the title. For this reasons, this method is built to classify the contents of tweets, then comparing associations with title analysis according to the first method above. After analysing titles, Twitter data was categorised into three classes: 0 represents tweets as exactly the title, possibly along with

information about authors or publishers, 1 represents tweets as the title with hashtag or mentions of @usernames or personal comments that means sharing something to somebody, and category 2 represents tweets as completely different from the title (Table 3). The three classes are exclusive; for instance, if a tweet is codified in class 0, it can't appear in class 1 or 2. The coding process was done manually by the author, and to increase the reliability of the coding, VocabProfile was used to compare familiarity as common words between tweets and titles.

Table 3. A description of coding scheme

0	**1**	**2**
Title	Title + emoticon, opinion, comment (in English or other languages)	No title but summary of title and content
Part of title (because tweet length is limited within 140 characters)	Title +Hashtag or @username	No title but comment
Title + info as name of author, publisher, year of publish		No title, only the link

Each result of categories in title analysis method will be compared with number of tweets in each tweet class. For example, an article has 10 tweets including two tweets as code 0, three tweets as code 1 and five tweets as code 3, each number of tweets in every code will be compared with each categories of title analysis. The result is calculated by Spearman rank correlation function on Microsoft Excel that output a decimal number between -1 and 1. For example, comparing number of tweets in class 1 with percentage of academic words in title (AWL) gives a result of 0,75 that means fairly high positive correlation between them, or a correlation result of 0,16 for number of tweets in class 1 and percentage of off-list words means very low due to the outcome close to 0.

Chapter 4. Results

The chapter presents all outcomes of title analysis, tweet analysis, and the correlation results of those analyses together. First, title analysis shows the average results of lexical categories for each title and top 10 articles' titles which have the largest number of tweets. Next, tweet analysis displays results of the coding and then the correlation results between tweet and title is presented.

4.1 Title analysis

As described in the methods section, the title analysis outputs findings in a summary of titles' word counts in total of 200 titles which are tweeted 989 times, percentage of K1 (most 1000 frequent words), percentage of K2 (most 2000 frequent words), percentage of academic words, percentage of off-list words. Table 4 shows results of average title analysis including average number of tweets. The length of titles was well-distributed statistically; the longest title has 23 words versus the 2 words as the shortest one, and the average word counts of title is 12,03 which is very close to the average of 12,5. On the other hand, the numbers of tweets were not distributed well. The largest number of tweets for one article is 82 and the minimum number is 1, while average number of tweets is only 4,95. This shows that the distribution of tweets among articles is highly skewed, with few articles receiving a lot of tweets, and a lot of articles receiving only few tweets.

Table 4. A summary of title analysis

Categories	Average	Max	Min
Length of titles	12,03	23 words	2 words
Number of tweets	4,95	82 tweets	1 tweet
Percentage of K1	54,87%	90%	0%
Percentage of K2	8,18%	66,67%	0%
Percentage of academic words	14,08%	54,55%	0%
Percentage of off-list	22,87%	100%	0%

Among 200 titles, up to 64% contained most 2000 common words in average (54,87 of K1 + 8,18% of K2), followed by off-list at 22,87%, and academic words made more than 14%.

It is not surprising that K1 makes the highest percentage, while K2 positions at the lowest because the natural language processing in VocabProfile includes stop words in calculating most 1000 frequent words, while K2 has filtered those words and thus have less occurrences of typical words. Stop words refer to the most common short words in a language such as *the, is, that, this, and, which, on*, etc. in English. For instance, a title "this is the best one and the last one" would make 100% in K1, because all the words are among the 1000 most frequently used words in the English language. However, it is surprising that average percentage of academic words is not high (14.08%), even though those titles are considered as heading of scientific articles.

Table 5. Top 10 articles and journals according to top number of tweets

Journal	Title	Number of tweets	Title length	K1 (%)	K2 (%)	Academic words (%)	Off-list (%)
Demography	Happiness: Before and After the Kids	82	6	83,33	0	0	16,67
Autism: The International Journal of Research & Practice	Investigating the cross-cultural validity of DSM-5 autism spectrum disorder: Evidence from Finnish and UK samples	61	15	35,29	5,88	23,53	35,3
Archives of Sexual Behavior	Is a Woman's Preference for Chest Hair in Men Influenced by Parasite Threat?	41	13	61,54	30,77	0	7,69
Journal of Child Psychology & Psychiatry	Predictors of developmental dyslexia in European orthographies with varying complexity	34	10	30	0	20	50
Archives of Sexual Behavior	Finnish Women and Men Who Self-Report No Sexual Attraction in the Past 12 Months: Prevalence, Relationship Status, and Sexual Behavior History	30	20	57,14	14,29	14,29	14,28
Information Systems Research	The Effect of Customers' Social Media Participation on Customer Visit Frequency and Profitability: An Empirical Investigation	26	16	50	18,75	25	6,25
Frontiers in Psychology	Interplay between singing and cortical processing of music: A longitudinal study in children with cochlear implants	22	16	62,5	0	6,25	31,25
Journal of Learning Disabilities	Basic Auditory Processing Deficits in Dyslexia Systematic Review of the Behavioral and Event-Related Potential/ Field Evidence	22	16	52,94	11,76	17,65	17,65
Animal Behaviour	No direct relationship between human female orgasm rate and number of offspring	21	12	75	8,33	0	16,67

Additionally, Table 5 displays top 10 articles which were tweeted the most. The shortest title length (6 words) in this list were tweeted 82 times (including retweets), and this title also has the highest percentage in K1 among top 10, whereas the article with highest rate of academic words (25%) in its title has smaller number of tweets (26 tweets). Top 10 titles are mostly about social life and especially family matters among men, women, and their kids in Finnish

society. This pointed out that people are more interested in and tend to share more research about Finnish family matters as a sub-topic in the field of arts and humanities.

4.2 Tweet and title

Tweet analysis refers to classifying and codifying the types of the 989 tweets and then examining connections with results of title analysis. Basically, tweets are categorised into three classes: 0 for tweets as exact titles, probably including name of authors and publishers, 1 for tweets which contain titles and other sharing symbols such as hashtag or @, and title summary, and 2 for tweets which are totally different with the titles. The codifying stage was processed manually by the author.

Table 6. A summary of code

Code	Number of tweets	Number of tweets in percent
0	193	19,51%
1	222	22,45%
2	574	58,04%
Total	989	100%

Table 7 presents results of the coding, showing that category 2 dominates at more than 50% in overall, followed by categories 1 and 0 with 22,45% and 19,51% respectively. This showed that public is more likely to express their personal comments or introducing scientific articles along with linking rather than only mentioning the titles for mere sharing. Therefore, the results indicate a high level of engagement by people expressed towards scientific products in the field of arts and humanities on Twitter.

Moreover, the correlation between title and tweet was examined based on findings of title analysis and tweet analysis. Specifically, each code is tested with each category including length of title, K1 words, K2 words, academic words, and off-list words in the title analysis according to number of tweets for each title to output correlation results. For example, a title is tweeted 15 times which are categorised in number of 5 as code 0, 6 as code 1, and 4 as

code 2, then those numbers will be tested with results of that title's categories separately to produce the final correlation results.

Table 7. The example of data for correlation test

Citation:title	Number of tweets	Length of title	Percentage of K1	Percentage of K2	Percentage of Academic words	Percentage of off-list words	Code
The Effect of Custome	21	16	50	18,75	25	6,25	2
The Effect of Custome	2	16	50	18,75	25	6,25	1
The Effect of Custome	3	16	50	18,75	25	6,25	0
Education of research	1	10	30	20	30	20	2
Education of research	1	10	30	20	30	20	0
Undergraduate medica	9	13	53,85	23,08	7,69	15,38	2
Undergraduate medica	1	13	53,85	23,08	7,69	15,38	1
Undergraduate medica	2	13	53,85	23,08	7,69	15,38	0

Table 8 shows the data after analysing title by using VocabProfile and classifying and coding each tweet. The data will be sorted by code and number of tweets in order to run separate correlation tests.

Table 8. Correlation between title and tweet according to number of tweets

	Code 0	**Code 1**	**Code 2**	**Total number of tweets**
Number of tweets to length of title	0,20	0,09	0,14	0,13
Number of tweets to K1	0,02	-0,04	0,05	-0,03
Number of tweets to K2	0,00	0,08	-0,06	0,03
Number of tweets to academic words	0,01	-0,05	0,08	0,02
Number of tweets to off-list words	-0,05	0,01	-0,04	-0,04

Table 9 shows the correlation results between coding of the tweets and title's lexical analysis. In overall, the correlation results are not really meaningful that most of those are close to 0 in both positive and negative cases. The strongest positive correlation is between code 0 and

length of title (0,20), while code 2 and number of tweets to K1 made the most negative correlation (-0,06). Moreover, there are still slightly differences between codes in different categories. Length of title is correlated the most with code 0 at 0,20 as the largest correlation number in this table, followed by code 2 and code 1 as 0,14 and 0,09 respectively in the same context. This can be explained by how each code was codified. While code 0 means tweets as exact title, or part of title, or title with information of authors and publishers, code 1 and 2 are less relevant with title according to mutual words between title and tweets. Besides, code 1 is connected the most with K2 at 0,08, whereas K2 correlated with code 2 negatively at -0,06 and correlated neutrally with code 0 in the same category. This probably indicates that comments or opinions about the title in tweets as code 1 contain more texts in a range of 2000 common words than the other two codes. In overall, when combining all the three codes, most of correlation results were very close to neutral, only the correlation with length of title was different that received higher score positively.

Chapter 5. Discussions

In the previous chapter, this study presented results of title analysis and the correlations between title characteristics and number of tweet, according to type of tweets. This chapter provides discussions, limitations, suggestions for future research, and more importantly demonstrates concrete answers to the research questions based on the results and reflects them with previous studies in theoretical background. Beginning with reviewing correlation between the contents of tweets and title characteristics in scientific articles from the outcomes, that revealed a limited association between these two. The second research question is discussed based on a view of tweeting behaviour of public in earlier research and the findings. People tend to express personal comments or discussions along with linking the articles in tweets, that indicate a high level of engagement of tweeting behaviour generally. Finally, the contributions of this study to scholarly communication is stated, the limitations of methodology and suggestions for future research are discussed based on these mistakes as well as on applications to other fields of research.

5.1 Correlation between tweeting and titling

Previous studies (Haustein et al. 2013; Thelwall et al. 2013) concluded positive results of correlation between tweeting and citations. This study aimed to investigate another aspect of correlation between the features of title and number of tweet. The features of title comprise factors such as length of titles and standardized categories of Vocabprofile as 2000 common words, academic words, and off-list words. As reported in Table 9, all correlation results were below average (< |0,5|) that indicated weak correlations between these features of scientific articles' titles and number of tweet. This answered the first research question of how much connection between the difficulty of the scientific articles' titles in field of arts and humanities and number of tweet.

However, there were slightly differences among codes of tweet and the length of titles. Tweets as well as retweets that repeat titles received positive correlation results towards length of titles; for instance, code 0 and length of titles were given correlation number at 0,20 as the highest one. This means that the longer titles tend to be retweeted more if the tweets quoted the titles. It can be explained that long titles show clearly content of the articles as well as specific what the articles aimed to demonstrate, therefore, readers were attracted more

and they simply retweeted those titles. This matches with a study of Habizadeh and Yadollahie (2010) that the longer titles seem to be more associated with higher citation rates. Meanwhile long titles were less retweeted if the tweets only gave comments briefly about the titles or summarized the titles. By contrast, in a research of Letchford, Moat and Preis (2015), they concluded that within 20,000 most highly cited papers from 2007 to 2013, the shorter titles received more citations. However, the correlation also depends on the reputation of the journals as well as authors and field of research. Additionally, they insisted that it makes sense if a short title is easier to understand and to attract readers that encourage them to cite or to tweet.

Ball (2009, p.668) gave a definition for a good way of titling: "a title should motivate the reader to read an article, give the readers a summary of the contents, give overview of the topics and finding discussed, and introduce the way in which the reported items are looked at". The concept described plus point in writing a title, but it is not always easy to follow the rules in scientific title, especially in abstractive field such as arts and humanities. This indicated the less occurrences of academic words in titles, the more chances of retweets for those titles. For the tweets, which summarize titles or give comment by using academic language, this rule is even clearer due to more extreme in negative correlation (code 1 vs academic words), or close to neutral (code 0 vs academic words), while code 2 correlated more positively with academic words because it does not contain academic titles in the tweets. It could be explained that although authors of the tweets have rewritten the titles or summarized the content of research in their own language, when people click on the link and read the paper by themselves, they still ignore to retweet if the title is written in academic tone or hard to understand.

Along with open science movement, scientific research may be able to reach more and more audiences from any educational levels and backgrounds. Plus, the language on social media sites is not the same language using in academic context. As such, it is reasonable that social media users may hesitate to share a scientific article if they do not even understand what the title talks about. Academic words in titles are the drawbacks to reach audiences widely, and this affected the number of tweets or retweets. Although the data set of this research is small and the negative correlation between the number of tweets and the academic tone of titles is

not highly significant, researchers should think about who audiences are and how to write a paper which is informative but readily comprehensible to the average people.

5.2 Tweeting behaviour

This study revealed that a high level of engagement of public towards scientific articles on Twitter, that is somehow new and reverse with earlier studies. For instances, in 2013, Thelwall and his co-authors conducted a content analysis study of how 270 tweets are linked to academic articles in four journals (PLOS ONE, PNAS; Science, and Nature). The authors found that tweets mainly contain article titles (42%) or a brief summary (41%). The summaries seem like more understandable for wider audience to read. The majority of tweets do not link to the name of authors, but some do and some are self-citations. In addition, they also found that most of the tweets are neutral and no attached comments in the tweets, except some tweets in purpose of making joke about science or excitement. In humanities discipline, especially digital humanities, Holmberg and Thelwall (2014) found that many scholarly products were tweeted on Twitter to communicate with audiences. In this study, among 989 tweets collected in this study, more than half belonged to code 2 which includes tweets or retweets as summarized titles or personal comments about titles and different with the original based on common words. This showed that the public is more likely to express their personal comments or introducing scientific articles along with linking rather than only mentioning the titles for mere sharing. Therefore, the results indicate a high level of engagement by people expressed towards scientific products in the field of arts and humanities on Twitter. This could be an answer for the second research question which concerns the tweeting behaviour of social information users in arts and humanities research. As Thelwall and Delgado (2015) said: "we should not measure the social impact of arts and humanities research with traditional metrics but use contextualised data instead. In order to understand the wider impact of research, data itself should be used as an indicator". In addition, they added that altmetrics data could be used to understand how arts and humanities research reach the public, not for counting numbers of tweets or Mendeley readers, but to carefully look at how scholarly publications are tweeted and blogged, analysing content and motivation for tweeting and blogging. Although the average number of tweets was fairly little in overall of 200 articles (average of 4,95 between 82 tweets as maximum and 1 tweets as minimum per articles), the distribution of tweets among all articles varied. Only a few of

articles attract attention that received highly number of tweets, while many articles received only a few tweets. The study found out that those top attractive articles were mainly about Finnish family matters as a sub-topic in the field of arts and humanities. This indicated that people were interested in this topic and tend to share more research about Finnish family troubles, whereas other sub-topics were less attentive or neglected.

5.3 Contributions, limitations and suggestions for future research

The study has contributions to scholarly communication, specifically in the field of arts and humanities. The findings revealed that public is more likely to engaged in tweeting scientific titles that is different with earlier research. Thus, the largest part in number of tweets in this research tells more about the article such as personal comments, content's summary than only quoting the titles or brief summary of title as founded in previous studies. In addition, people seem to be interested in Finnish family matter mostly that make high level of tweets on the articles about this topic. Open scholarly discussion is communicated mostly by text, especially on social networking platform; therefore, the methodology in this study could be a good example for textual analysis in research about scholarly communication.

However, the samples selection from raw data was only a small set of research in art and humanity, that was limited to 200 random articles presented on Twitter. Therefore, the results are not necessarily a representative of the entire Finnish research in arts and humanities on social networking. Next, the method of title analysis was executed by the only software that based on their words' database, so the outcomes might not be standardized entirely. Another limitation is that Web VocabProfile does not allow users to input several texts at the same time and keep track of which texts are contributing to which parts of the profile. The author had to extract the result manually for each title. It was time-consuming and could lead to error when trying to extend the scale, for example, from 200 to 2000 articles. Moreover, the tool seemed to ignore the markers, colons or other kinds of mark in title that may affect the meaning of title's implication, for example, amusement or humour of titles. In addition, classification of the coding scheme in tweet analysis was sometimes subjective and the bias in analyzing tweet contents is inevitable, even with the assistance from VocabProfile. As such, those limitations may have affected the study's findings.

Future research might apply the methods to either larger scales or from other altmetrics providers. An important contribution of this study is that it revealed how and in what specific sub-topics in Finnish arts and humanities people are mostly interested in. In addition, the methodology could be applied into other fields of research or other topics that are more common than arts and humanities, so a bigger contribution would be possibly accomplished. In general, trending on social networking changes fast and altmetrics is still new that need more research to create a standard tool to evaluate scientific products on this communication platform widely.

Chapter 6. Conclusion

Due to the growing phenomenon of modern communication on social media platforms nowadays, there has been popularly research about this to investigate influences between academia and public audiences. Specifically, this study focused on Twitter platform and provided a new look of tweeting behaviour on scientific articles in field of arts and humanities and whether those article' titles are connected to the tweets' content.

It was argued that positive results of correlation between tweeting and citations in previous studies. However, this study extended to correlation between tweeting and title characteristics, that was evidenced as weak connection generally. Nevertheless, tweeting behaviour of information users in this research was revealed. More than half in total of the tweets tended to explain the article's titles instead of quoting the titles. People were more likely to explain the mentioned articles in their own words, but if the language use in those tweets sounds less academic, the chances of retweets were increased. It is understandable that to reach audience widely, language use matters. In addition, not all aspects in the field of arts and humanities were noticed, the outcomes revealed that Finnish family matters were the most attractive sub-topics. Furthermore, this study also agreed with previous studies that Twitter is different with other social media platforms in context of social interaction behaviour. Number of followers does not guarantee a person to receive a large number of retweets, while celebrities or popular people on Facebook usually receive much attention in their every post.

Despite some limitations in methodology, this study takes advantage of a unique circumstance in tweeting behaviour context as well as revealing what kinds of sub-topics in field of arts and humanities people were attracted, and a discovery of correlation between features of scientific titles and tweeting behaviours. This may contribute to sharpen similar research in this complex field in the future.

Bibliography

Abramo, G., D'Angelo, C.A. and Viel, F. (2013) 'Assessing the accuracy of the h- and g-indexes for measuring researchers' productivity', Journal of the American Society for Information Science and Technology, 64(6), pp. 1224–1234. doi: 10.1002/asi.22828.

Al-Aufi, A.S. and Fulton, C. (2014) 'Use of social networking tools for informal scholarly communication in humanities and social sciences disciplines ☆', Procedia - Social and Behavioral Sciences, 147, pp. 436–445. doi: 10.1016/j.sbspro.2014.07.135.

Ardanuy, J. (2013) 'Sixty years of citation analysis studies in the humanities (1951-2010)', Journal of the American Society for Information Science and Technology, 64(8), pp. 1751–1755. doi: 10.1002/asi.22835.

Ball, R. (2009) 'Scholarly communication in transition: The use of question marks in the titles of scientific articles in medicine, life sciences and physics 1966–2005', Scientometrics, 79(3), pp. 667–679. doi: 10.1007/s11192-007-1984-5.

Bar-Ilan, J., Haustein, S., Peters, I., Priem, J., Shema, H. and Terliesner, J. (2012) Title: Beyond citations: Scholars' visibility on the social web. Available at: https://arxiv.org/abs/1205.5611 (Accessed: 26 February 2017).

Bennett, W.L. and Segerberg, A. (2011) 'DIGITAL MEDIA AND THE PERSONALIZATION OF COLLECTIVE ACTION', Information, Communication & Society, 14(6), pp. 770–799. doi: 10.1080/1369118x.2011.579141.

Bollen, J., Van de Sompel, H., Hagberg, A. and Chute, R. (2009) 'A principal component analysis of 39 scientific impact measures', PLoS ONE, 4(6), p. e6022. doi: 10.1371/journal.pone.0006022.

Bornmann, L. (2014) 'Do altmetrics point to the broader impact of research? An overview of benefits and disadvantages of altmetrics', Journal of Informetrics, 8(4), pp. 895–903. doi: 10.1016/j.joi.2014.09.005.

Bornmann, L. and Marx, W. (2013) 'How should the societal impact of research be generated and measured? A proposal for a simple and practicable approach to allow interdisciplinary comparisons', Scientometrics, 98(1), pp. 211–219. doi: 10.1007/s11192-013-1020-x.

Boyd, D., Golder, S. and Lotan, G. (2010) 'Tweet, Tweet, Retweet: Conversational aspects of Retweeting on Twitter', 2010 43rd Hawaii International Conference on System Sciences, . doi: 10.1109/hicss.2010.412.

Brody, T., Harnad, S. and Carr, L. (2006) 'Earlier web usage statistics as predictors of later citation impact', Journal of the American Society for Information Science and Technology, 57(8), pp. 1060–1072. doi: 10.1002/asi.20373.

Bryant, J., Alan, D.B., Silberberg, R. and Elliott, S.M. (1981) 'Effects of humorous illustrations in college textbooks', Human Communication Research, 8(1), pp. 43–57. doi: 10.1111/j.1468-2958.1981.tb00655.x.

Burrell, Q.L. (2007) 'Hirsch index or Hirsch rate? Some thoughts arising from Liang's data', Scientometrics, 73(1), pp. 19–28. doi: 10.1007/s11192-006-1774-5.

Burrell, Q.L. (2013) 'Formulae for the h-index: A lack of robustness in Lotkaian informetrics?', Journal of the American Society for Information Science and Technology, 64(7), pp. 1504–1514. doi: 10.1002/asi.22845.

Cardona-Grau, D., Sorokin, I., Leinwand, G. and Welliver, C. (2016) 'Introducing the Twitter impact factor: An objective measure of Urology's academic impact on Twitter', European Urology Focus, 2(4), pp. 412–417. doi: 10.1016/j.euf.2016.03.006.

Chen, B. and Bryer, T. (2012) 'Investigating instructional strategies for using social media in formal and informal learning', The International Review of Research in Open and Distributed Learning, 13(1), pp. 104–87.

Chmiel, A., Sobkowicz, P., Sienkiewicz, J., Paltoglou, G., Buckley, K., Thelwall, M. and Hołyst, J.A. (2011) 'Negative emotions boost user activity at BBC forum', Physica A: Statistical Mechanics and its Applications, 390(16), pp. 2936–2944. doi: 10.1016/j.physa.2011.03.040.

Chretien, K., Azar, J. and Kind, T. (2011) 'Physicians on Twitter', JAMA., 305(6), pp. 566–8.

Cruz, F. and Jamias, S.B. (2013) 'Scientists' use of social media: The case of researchers at the university of the Philippines Los Baños', First Monday, 18(4). doi: 10.5210/fm.v18i4.4296.

Didegah, F. (2014) FACTORS ASSOCIATING WITH THE FUTURE CITATION IMPACT OF PUBLISHED ARTICLES: A STATISTICAL MODELLING APPROACH. Available at: http://wlv.openrepository.com/wlv/bitstream/2436/322738/1/Didegah_PhDthesis.pdf (Accessed: 24 January 2017).

Didegah, F. and Thelwall, M. (2013) 'Which factors help authors produce the highest impact research? Collaboration, journal and document properties', Journal of Informetrics, 7(4), pp. 861–873. doi: 10.1016/j.joi.2013.08.006.

Dynich, A. and Wang, Y. (2016) 'On the e-commercialization of science: A step toward the future?', Electronic Commerce Research and Applications, 20, pp. 183–188. doi: 10.1016/j.elerap.2016.11.001.

Ebner, M. and Reinhardt, W. (2009) Social networking in scientific conferences -- Twitter as tool for strengthen a scientific community. Available at: https://www.researchgate.net/publication/216015988_Social_networking_in_scientific_con

ferences_--_Twitter_as_tool_for_strengthen_a_scientific_community (Accessed: 29 July 2016).

Elsevier (2016) What is peer review? Available at: https://www.elsevier.com/reviewers/what-is-peer-review (Accessed: 24 December 2016).

Eysenbach, G. (2011) 'Can Tweets predict citations? Metrics of social impact based on Twitter and correlation with traditional metrics of scientific impact', Journal of Medical Internet Research, 13(4), p. 123. doi: 10.2196/jmir.2012.

Ferrara, E. and Romero, A.E. (2013) 'Scientific impact evaluation and the effect of self-citations: Mitigating the bias by discounting the h-index', Journal of the American Society for Information Science and Technology, 64(11), pp. 2332–2339. doi: 10.1002/asi.22976.

Ferrara, E., Varol, O., Davis, C., Menczer, F. and Flammini, A. (2014) Title: The rise of social Bots. Available at: http://arxiv.org/abs/1407.5225 (Accessed: 19 September 2016).

Galligan, F. and Dyas-Correia, S. (2013) 'Altmetrics: Rethinking the way we measure', Serials Review, .

Garfield, E. (2006) 'The history and meaning of the journal impact factor', JAMA, 295(1), p. 90. doi: 10.1001/jama.295.1.90.

Gruzd, A., Staves, K. and Wilk, A. (2012) 'Connected scholars: Examining the role of social media in research practices of faculty using the UTAUT model', Computers in Human Behavior, 28(6), pp. 2340–2350. doi: 10.1016/j.chb.2012.07.004.

Gu, F. and Widén- WulffGunilla (2011) 'Scholarly communication and possible changes in the context of social media', The Electronic Library, 29(6), pp. 762–776. doi: 10.1108/02640471111187999.

Habibzadeh, F. and Yadollahie, M. (2010) 'Are shorter article titles more attractive for citations? Cross-sectional study of 22 scientific journals', 51(2).

Hammarfelt, B. (2014) 'Using altmetrics for assessing research impact in the humanities', Scientometrics, 101(2), pp. 1419–1430. doi: 10.1007/s11192-014-1261-3.

Hartley, J. (2007) 'Planning that title: Practices and preferences for titles with colons in academic articles', Library & Information Science Research, 29(4), pp. 553–568. doi: 10.1016/j.lisr.2007.05.002.

Hartley, J. (2008) 'Academic writing and publishing', British Journal of Educational Technology, 40(4), pp. 775–776. doi: 10.1111/j.1467-8535.2009.00994_7.x.

Haustein, S., Bowman, T.D., Holmberg, K., Peters, I. and Larivière, V. (2014) 'Astrophysicists on Twitter', Aslib Journal of Information Management, 66(3), pp. 279–296. doi: 10.1108/ajim-09-2013-0081.

Haustein, S., Bowman, T.D., Holmberg, K., Tsou, A., Sugimoto, C.R. and Larivière, V. (2015) 'Tweets as impact indicators: Examining the implications of automated "bot" accounts on Twitter', Journal of the Association for Information Science and Technology, 67(1), pp. 232–238. doi: 10.1002/asi.23456.

Haustein, S., Peters, I., Sugimoto, C.R., Thelwall, M. and Larivière, V. (2013) 'Tweeting biomedicine: An analysis of tweets and citations in the biomedical literature', Journal of the Association for Information Science and Technology, 65(4), pp. 656–669. doi: 10.1002/asi.23101.

Haustein, S. and Siebenlist, T. (2011) 'Applying social bookmarking data to evaluate journal usage', Journal of Informetrics, . doi: 10.1016/j.joi.2011.04.002.

Hellqvist, B. (2009) 'Referencing in the humanities and its implications for citation analysis', Journal of the American Society for Information Science and Technology, , p. n/a–n/a. doi: 10.1002/asi.21256.

Holmberg, K. and Hellsten, I. (2016) 'Twitter campaigns around the fifth IPCC report', SAGE Open, 14(4), p. 2158244016659117. doi: 10.1177/2158244016659117.

Holmberg, K.J. (2015) Altmetrics for information professionals: Past, present and future. United Kingdom: Chandos Publishing (Oxford).

Holmberg, K. and Thelwall, M. (2014) 'Disciplinary differences in Twitter scholarly communication', Scientometrics, 101(2), pp. 1027–1042. doi: 10.1007/s11192-014-1229-3.

Huberman, B., Romero, D.M. and Wu, F. (2008) 'Social networks that matter: Twitter under the microscope', First Monday, 14(1). doi: 10.5210/fm.v14i1.2317.

Jamali, H.R. and Nikzad, M. (2011) 'Article title type and its relation with the number of downloads and citations', Scientometrics, 88(2), pp. 653–661. doi: 10.1007/s11192-011-0412-z.

Jansen, B.J., Zhang, M., Sobel, K. and Chowdury, A. (2009) 'Twitter power: Tweets as electronic word of mouth', Journal of the American Society for Information Science and Technology, 60(11), pp. 2169–2188. doi: 10.1002/asi.21149.

Java, A., Song, X., Finin, T. and Tseng, B. (2009) 'Why we Twitter: An analysis of a Microblogging community', in Advances in Web Mining and Web Usage Analysis. Springer Science + Business Media, pp. 118–138.

Kane, T.S. (1988) The new Oxford guide to writing. 16th edn. New York: Oxford University Press.

Kwak, H., Lee, C., Park, H. and Moon, S. (2010) 'What is Twitter, a social network or a news media?', Proceedings of the 19th international conference on World wide web - WWW '10, . doi: 10.1145/1772690.1772751.

Lahuerta-Otero, E. and Cordero-Gutiérrez, R. (2016) 'Looking for the perfect tweet. The use of data mining techniques to find influencers on twitter', Computers in Human Behavior, 64, pp. 575–583. doi: 10.1016/j.chb.2016.07.035.

Lee, C.J., Sugimoto, C.R., Zhang, G. and Cronin, B. (2012) 'Bias in peer review', Journal of the American Society for Information Science and Technology, 64(1), pp. 2–17. doi: 10.1002/asi.22784.

Letchford, A., Moat, H.S. and Preis, T. (2015) 'The advantage of short paper titles', Research, 2(8), p. 150266. doi: 10.1098/rsos.150266.

Letierce, J., Passant, A., Breslin, J. and Decker, S. (2011) Understanding how Twitter is used to spread scientific messages. Available at: http://journal.webscience.org/314/ (Accessed: 29 July 2016).

Letierce, J., Passant, A., G. Breslin, J. and Decker, S. (2010) Using Twitter uring an Academic Conference: The #iswc2009 Use-Case. Available at: http://www.aaai.org/ocs/index.php/ICWSM/ICWSM10/paper/view/1523/1877 (Accessed: 27 July 2016).

Lewison, G. and Hartley, J. (2005) 'What's in a title? Numbers of words and the presence of colons', Scientometrics, 63(2), pp. 341–356. doi: 10.1007/s11192-005-0216-0.

Merton, R.K. (1968) 'The Matthew effect in science: The reward and communication systems of science are considered', Science, 159(3810), pp. 56–63. doi: 10.1126/science.159.3810.56.

Mitrović, M., Paltoglou, G. and Tadić, B. (2010) 'Networks and emotion-driven user communities at popular blogs', The European Physical Journal B, 77(4), pp. 597–609. doi: 10.1140/epjb/e2010-00279-x.

Moed, H.F., De Bruin, R.E. and Van Leeuwen, T.N. (1995) 'New bibliometric tools for the assessment of national research performance: Database description, overview of indicators and first applications', Scientometrics, 33(3), pp. 381–422. doi: 10.1007/bf02017338.

Mollett, A., Moran, D. and Dunleavy, P. (2011) Using Twitter in university research, teaching and impact activities A guide for academics and researchers. Available at: http://blogs.lse.ac.uk/impactofsocialsciences/files/2011/11/Published-Twitter_Guide_Sept_2011.pdf (Accessed: 6 October 2016).

Mulligan, A. (2005) 'Is peer review in crisis?', Oral Oncology, 41(2), pp. 135–141. doi: 10.1016/j.oraloncology.2004.11.001.

Naveed, N., Gottron, T., Kunegis, J., Alhadi and Che, A. (2011) 'Bad news travel fast', , p. 8. doi: 10.1145/2527031.2527052.

Neylon, C. and Wu, S. (2009) 'Article-level metrics and the evolution of scientific impact', PLoS Biology, 7(11), p. e1000242. doi: 10.1371/journal.pbio.1000242.

Nielsen, M. (2011) Definitions of open science? Available at: https://lists.okfn.org/pipermail/open-science/2011-July/000907.html (Accessed: 13 August 2016).

Open science and research leads to surprising discoveries and creative insights (2014) Available at: http://openscience.fi/documents/14273/0/Open+Science+and+Research+Roadmap+2014-2017/e8eb7704-8ea7-48bb-92e6-c6c954d4a2f2 (Accessed: 6 July 2016).

Paiva, C.E., Paiva, B.S.R. and Lima, J.P. da S.N. (2012) 'Articles with short titles describing the results are cited more often', 67(5).

Peters, I., Beutelspacher, L., Maghferat, P. and Terliesner, J. (2012) 'Scientific bloggers under the altmetric microscope', Proceedings of the American Society for Information Science and Technology, 49(1), pp. 1–4. doi: 10.1002/meet.14504901305.

Priem, J., Costello, K. and Dzuba, T. (2011) 'First-year graduate students just wasting time? Prevalence and use of Twitter among scholars', Metrics 2011 Symposium on Informetric and Scientometric Research, New Orleans, LA, USA. Retrieved from http://jasonpriem. org/self-archived/5uni-poster. png, .

Priem, J. and Costello, K.L. (2010) 'How and why scholars cite on Twitter', Proceedings of the American Society for Information Science and Technology, 47(1), pp. 1–4. doi: 10.1002/meet.14504701201.

Priem, J. and Hemminger, B.H. (2010) 'Scientometrics 2.0: New metrics of scholarly impact on the social web', First Monday, 15(7). doi: 10.5210/fm.v15i7.2874.

Priem, J., Taraborelli, D., Groth, P. and Neylon, C. (2010) Altmetrics: A manifesto. Available at: http://altmetrics.org/manifesto/ (Accessed: 22 September 2016).

Principles and strategies for the reform of scholarly communication 1 (1996) Available at: http://www.ala.org/acrl/publications/whitepapers/principlesstrategies (Accessed: 7 July 2016).

Pritchard, A. (1969) Statistical Bibliography or Bibliometrics? Available at: https://www.researchgate.net/publication/236031787_Statistical_Bibliography_or_Bibliometrics (Accessed: 25 January 2017).

Ramanna, S., Jain, L.C. and Howlett, R.J. (2013) Emerging paradigms in machine learning. Available at: https://books.google.fi/books?id=i6FschwRXdgC&pg=PA454&lpg=PA454&dq=Specific+markers+as+colons+or+question+marks&source=bl&ots=lkrv-yS1yf&sig=qpj_5o0u2UhP3n9aZoZT2HVQAYY&hl=en&sa=X&ved=0ahUKEwjN8fWZo6zSAhWMEywKHTjdDK0Q6AEIOjAF#v=onepage&q=Specific%20markers%20as%20colons%20or%20question%20marks&f=false (Accessed: 25 February 2017).

Ross, C., Terras, M., Warwick, C. and Welsh, A. (2011) 'Enabled backchannel: Conference Twitter use by digital humanists', Journal of Documentation, 67(2), pp. 214–237. doi: 10.1108/00220411111109449.

Rowlands, I., Nicholas, D., Russell, B., Canty, N. and Watkinson, A. (2011) Social media use in the research workflow. Available at: http://discovery.ucl.ac.uk/1332803/ (Accessed: 6 October 2016).

Sagi, I. and Yechiam, E. (2008) 'Amusing titles in scientific journals and article citation', Journal of Information Science, 34(5), pp. 680–687. doi: 10.1177/0165551507086261.

Schreiber, M. (2013) 'Do we need the g-index?', Journal of the American Society for Information Science and Technology, 64(11), pp. 2396–2399. doi: 10.1002/asi.22933.

Seglen, P.O. (1997) 'Why the impact factor of journals should not be used for evaluating research', BMJ, 314(7079), pp. 497–497. doi: 10.1136/bmj.314.7079.497.

Shema, H., Bar-Ilan, J. and Thelwall, M. (2014) 'Do blog citations correlate with a higher number of future citations? Research blogs as a potential source for alternative metrics', Journal of the Association for Information Science and Technology, 65(5), pp. 1018–1027. doi: 10.1002/asi.23037.

Shuai, X., Pepe, A. and Bollen, J. (2012) 'Title: How the scientific community reacts to newly submitted Preprints: Article Downloads, Twitter mentions, and citations', . doi: 10.1371/journal.pone.0047523.

Simkin, M.V. and Roychowdhury, V.P. (2002) Title: Read before you cite! Available at: https://arxiv.org/abs/cond-mat/0212043 (Accessed: 8 October 2016).

Small, H.H. (2013) The value of the humanities. Oxford, United Kingdom: Oxford University Press.

Spier, R. (2002) 'The history of the peer-review process', 20(8), pp. 357–358. doi: 10.1016/S0167-7799(02)01985-6.

State of scientific research - Academy of Finland (2014) Available at: http://www.aka.fi/en/research-and-science-policy/state-of-scientific-research/ (Accessed: 6 July 2016).

State of scientific research - Academy of Finland (2016) Available at: http://www.aka.fi/globalassets/30tiedepoliittinen-toiminta/tieteentila/tt2016_kysely_vaikuttavuudesta_en.pdf (Accessed: 24 January 2017).

Subotic, S. and Mukherjee, B. (2013) 'Short and amusing: The relationship between title characteristics, downloads, and citations in psychology articles', Journal of Information Science, , p. 165551513511393. doi: 10.1177/0165551513511393.

Thelwall, M., Buckley, K. and Paltoglou, G. (2011) 'Sentiment strength detection for the social web', Journal of the American Society for Information Science and Technology, 63(1), pp. 163–173. doi: 10.1002/asi.21662.

Thelwall, M. and Delgado, M.M. (2015) 'Arts and humanities research evaluation: No metrics please, just data', Journal of Documentation, 71(4), pp. 817–833. doi: 10.1108/jd-02-2015-0028.

Thelwall, M., Haustein, S., Larivière, V. and Sugimoto, C.R. (2013) 'Do Altmetrics work? Twitter and Ten other social web services', PLoS ONE, 8(5), p. e64841. doi: 10.1371/journal.pone.0064841.

Thelwall, M., Tsou, A., Weingart, S., Holmberg, K. and Haustein, S. (2013) Tweeting links to academic articles. Available at: http://www.scit.wlv.ac.uk/~cm1993/papers/TweetingLinksAcademicArticles.pdf (Accessed: 27 September 2016).

Thorin, S. (2006). Global Changes in Scholarly Communication.

Tiryakioglu, F. and Erzurum, F. (2011) 'Use of social networks as an education tool', Contemporary Educational Technology, 2(2), pp. 135–150.

Vergeer, M., Hermans, L. and Sams, S. (2011) 'Online social networks and micro-blogging in political campaigning: The exploration of a new campaign tool and a new campaign style', Party Politics, , p. 1354068811407580. doi: 10.1177/1354068811407580.

Verma, H. (2016) Outputs of the NISO alternative assessment metrics project A recommended practice of the national information standards organization. Available at: http://www.niso.org/apps/group_public/download.php/17091/NISO%20RP-25-2016%20Outputs%20of%20the%20NISO%20Alternative%20Assessment%20Project.pdf (Accessed: 26 December 2016).

Walsh, J., Kucker, S., Maloney, N. and Gabbay, S. (2000). Connecting minds: Computer-mediated communication and scientific work. [Accessed 18 Mar. 2017]

Wang, X., Liu, C., Fang, Z. and Mao, W. (2014) From attention to citation, what and how does Altmetrics work? Available at: https://www.researchgate.net/publication/265687564_From_Attention_to_Citation_What_and_How_Does_Altmetrics_Work (Accessed: 4 September 2016).

Wang, Y. and Bai, Y. (2007) 'A corpus-based syntactic study of medical research article titles', System, 35(3), pp. 388–399. doi: 10.1016/j.system.2007.01.005.

Watson, A.B. (2009) 'Comparing citations and downloads for individual articles at the journal of vision', Journal of Vision, 9(4), pp. i–i. doi: 10.1167/9.4.i.

Weller, K., Dröge, E. and Puschmann, C. (2011) Citation analysis in Twitter: Approaches for defining and measuring information flows within Tweets during scientific conferences. Available at: http://ceur-ws.org/Vol-718/paper_04.pdf (Accessed: 29 July 2016).

Weller, K. and Peters, I. (2013) Citations in web 2.0. Available at: http://nfgwin.uni-duesseldorf.de/sites/default/files/Weller.pdf (Accessed: 6 October 2016).

Weller, K. and Puschmann, C. (2011) Twitter for scientific communication: How can citations/references be identified and measured? Available at: http://journal.webscience.org/500/1/153_paper.pdf (Accessed: 29 July 2016).

Weller, M. (2011) The digital scholar: How technology is transforming scholarly practice. Available at: http://oro.open.ac.uk/29664/ (Accessed: 3 March 2017).

Wikipedia (2016) 'Open science', in Wikipedia. Available at: https://en.wikipedia.org/wiki/Open_science (Accessed: 13 August 2016).

Yitzhaki, M. (1994) 'Relation of title length of journal articles to number of authors', Scientometrics, 30(1), pp. 321–332. doi: 10.1007/bf02017231.

Yitzhaki, M. (1997) 'Variation in informativity of titles of research papers in selected humanities journals: A comparative study', Scientometrics, 38(2), pp. 219–229. doi: 10.1007/bf02457410.

Yitzhaki, M. (2002) Scientometrics, 54(3), pp. 435–447. doi: 10.1023/a:1016038617639.

Yu, M.-C., Wu, Y.-C.J., Alhalabi, W., Kao, H.-Y. and Wu, W.-H. (2016) 'ResearchGate: An effective altmetric indicator for active researchers?', Computers in Human Behavior, 55, pp. 1001–1006. doi: 10.1016/j.chb.2015.11.007.

Zhao, D., Rosson and Beth, M. (2009) 'How and why people Twitter', , pp. 243–252. doi: 10.1145/1531674.1531710.