Highly tweeted articles with Finnish affiliation: Who are sharing them and what types of articles are shared the most?

Master’s thesis in the International Master’s Programme in Information and Knowledge Management

Supervisors: Kim Holmberg and Gunilla Widén

Faculty of Social Sciences, Business and Economics

Åbo 2019
Abstract
Title: Highly tweeted articles with Finnish affiliation: Who are sharing them and what types of articles are shared the most?

Supervisor: Kim Holmberg and Gunilla Widén

At present, growing numbers of researchers are interacting with different social media tools such as Twitter, Facebook, and Mendeley for their professional communication. Nowadays, social media platforms such as Twitter have become an essential part of scholarly communication. Scholars are using Twitter not only as a means of communication but also as a platform for disseminating research information. Alternative metrics (known as Altmetrics) refers to online metrics that may measure the societal impact of research. Bibliometrics and peer review are considered as a standard method for measuring the scientific impact whereas altmetrics may be used to measure other forms of impact. By using social media data, altmetrics measures the impact or visibility of research on social media. In this study, the author investigates who shared scientific articles with at least one Finnish author affiliation and that had high altmetrics counts on Twitter. The types of articles and the profile description of the tweeters are also examined in this study. Moreover, the citation number of these shared academic articles is studied. The author of this study also investigated whether the highly tweeted articles also are highly cited. For the data analysis, mixed research methods, combining quantitative research and qualitative study, are used; quantitative analysis based on Co-Word analysis, Spearman Correlation analysis and content analysis of the user profile description are preformed. The results of the study show the occupational expertise, gender of the users who are sharing academic articles on Twitter, and what types of articles with Finnish affiliation are shared on Twitter. From this study we can see that scientific articles ,with at least one Finnish author, are often shared by men and organizations, and the individuals are often working in academia. Moreover, this study found out that highly tweeted articles are not always highly cited scientific articles.
Table of Contents

Abstract ........................................................................................................................................... 1
List of Figures .................................................................................................................................. 4
List of Tables ................................................................................................................................... 5
List of Abbreviation ....................................................................................................................... 6
1. Introduction .................................................................................................................................. 7
  1.1 Motivation of the study ............................................................................................................ 7
  1.2 General Use of Twitter .......................................................................................................... 7
  1.3 The overall aim of the thesis and research questions ............................................................ 8
  1.4 Limitation of the study .......................................................................................................... 9
  1.5 Structure of the thesis .......................................................................................................... 9
2.0 Literature Review ..................................................................................................................... 11
  2.1 Scholarly communication and the development of scholarly communication ................. 11
    2.1.1 Formal and Informal scholarly communication ........................................................... 12
  2.2 Scholarly communication process and challenges ............................................................... 13
  2.3 Metrics research ................................................................................................................... 14
    2.3.1 Informetrics .................................................................................................................... 14
    2.3.2 Bibliometrics ................................................................................................................ 15
    2.3.3 Scientometrics .............................................................................................................. 16
    2.3.4 Webometrics ............................................................................................................... 17
  2.4 Citation analysis .................................................................................................................... 18
3. Social Media, scholarly communication in Twitter and Altmetrics ......................................... 20
  3.1 Social Media ........................................................................................................................ 20
    3.1.1 Analyzing Social Media ............................................................................................... 21
  3.2 Previous research on Twitter ................................................................................................ 22
  3.2 Altmetrics ............................................................................................................................. 24
Author: Anik Dutta

3.2.1 Potential and limitations of Altmetrics ......................................................24

4.0 Data and Methodology .................................................................................28
  4.1 Research Design .........................................................................................28
  4.2. Research Methodology .............................................................................28
    4.2.1. Quantitative and qualitative research method ........................................29
  4.3 Data ...........................................................................................................30
  4.4 Methods of the study .................................................................................32

5.0 Results .........................................................................................................36
  5.1 Analysis and Results ..................................................................................36
  Q.1. Who is tweeting about research publications? .........................................36
  Q.2. What kind / types of articles professionals are shared most on Twitter? ....38
  Q.3. Do highly tweeted articles also receive citations? ....................................40

6.0 Discussion .....................................................................................................45
  6.1 Discussion of the research question ..............................................................45
    Q.1. Who is tweeting about research publications? ........................................45
    Q.2. What kind / types of articles professionals are shared most on Twitter? ....46
    Q.3. Do highly tweeted articles also receive citations? ....................................47

7.0 Future work .................................................................................................49

8.0 Conclusions .................................................................................................50

9.0 References ...................................................................................................51

10. Appendix ......................................................................................................67
    Appendix 1: Overall summary of the quantitative research: .......................67
List of Figures

<table>
<thead>
<tr>
<th>Figure</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Figure 1</td>
<td>The cycle of scholarly communication.</td>
<td>14</td>
</tr>
<tr>
<td>Figure 2</td>
<td>The relationships between different metrics-based research area</td>
<td>18</td>
</tr>
<tr>
<td>Figure 3</td>
<td>Data collection process</td>
<td>32</td>
</tr>
<tr>
<td>Figure 3</td>
<td>Co-word map of most frequent mentioned word in Twitter profile of people who is twitting about scientific article</td>
<td>43</td>
</tr>
</tbody>
</table>
Author: Anik Dutta

List of Tables

Table 1: Codebook of Gender 33

Table 2: Codebook of occupation 34

Table 3: Gender of the Tweet authors 36

Table 4: Occupation of the Tweets authors 37

Table 5: open or close access article 38

Table 6: Percentage of open or close access article and types of article 39

Table 7: Who is sharing what types of articles 40

Table 8: Number of citations of the tweeted article 41

Table 9: Correlation between number of time articles with at least one Finnish author mentioned on a tweets and number of times articles was cited. 42
Author: Anik Dutta

List of Abbreviation

WOS- Web of Science

API- Application programing interface

ABE-Agriculture, Biological science and Engineering

MHS- Medical and Health Science

SSH-Social Sciences and Humanities

NS- Natural Science
1. Introduction

This chapter introduces the motivation for the research of the thesis, as well as the research scope and research questions. It also explains the limitations of the study and the overall structure of the thesis.

1.1 Motivation of the study

Now a day’s social media has changed the way we interact and share information daily. Social media has not only changed our everyday information behaviour but has also influenced work practices such as in the context of research. At present, many researchers use social media to discover new research ideas, to disseminate research information or to share the research with others. One of the social media platforms that a researcher can use for scholarly communication or identify the social impact of the study is Twitter. According to Holmberg and Thelwall, (2014, p.2), "Twitter is a real-time microblog network where users can publish their opinions, ideas, stories and news in messages that are up to 140 characters long". Many researchers use this platform to disseminate research information to their followers. It means social media has impacted on the pattern of scholarly communication. To understand the online scholarly impact, new online metrics are developed, which are commonly known as Altmetrics (Haustein, 2016). Altmetrics are used as an indicator of impact and scientific performance. There are various notions and definition of scholarly communication are available, but there are not enough definitions of Altmetrics are available. Since there was a lack of definition of Altmetrics the author was driven to dig deep into the reason of this lacking and find a way to contribute in the field of Altmetrics. Nevertheless, few studies are concentrating on finding the reasons for using Twitter as a platform for sharing research publications. Therefore, this topic is chosen to fill in this gap.

1.2 General Use of Twitter

Twitter is a popular short messaging service which is accessible through a Web page, desktop or mobile software (Dann, 2010). Twitter is a real-time microblog network where users can publish their opinions, ideas, stories, and researchers can use it in scholarly communication (Holmberg and Thelwall, 2014). Twitter is an efficient and rapid tool for sharing emergency information; for example, sharing information about protest and movement or sharing valuable news has gained popularity among its users (Ash, 2011; Badenschier and Holger 2012). The best practice of Twitter has been studied for instance in the field of health community (Berger, 2009), public libraries (Cahill, 2009; Cuddy, 2009), different political campaigns (Cetina,
2009; Henneburg, et al., 2009), business and commercial for example: organizations nowadays using Twitter for advertisement (Power and Forte, 2008), journalism/reporting (Ettama, 2009), social campaign (Galer–Unti, 2009), and live coverage of different events (Gay, et al., 2009). Moreover, this technology has also been used by the government (Macintosh, 2008) to follow different activities. Furthermore, Twitter is known as one of the most frequently used social media tools that mention academic documents (Costas et al., 2015; Haustein et al., 2014; 2014a). Twitter has three unique features that support communication; the first feature is retweets or forwarded tweets; a second feature is the use of @username which means a person is mentioned or tagged in a tweet, and the last feature is hashtag which is a short keyword, prefixed with a symbol # and acts as a way of coordinating a broader discussion between large groups of users (Bruns and Burgess, 2011; Dann, 2010; Holmberg and Thelwall, 2014). Structurally, retweeting is like email forwarding where users post messages initially posted by others, and it brings new people into a thread, inviting them to involve without directly addressing them (Boyd et al., 2010). Earlier research (Boyd et al., 2010) found that people retweet for spreading information among new viewers or followers. "@" is the symbol which is used to mention someone else within a tweet followed by the intended username (Sriram et al., 2010). As Holmberg and Thelwell (2014) said, it is used to send a message to another Twitter user. By using the # character in front of a word in a tweet, users can gather a group of tweets about the same topic. Moreover, hashtags are regularly used at scientific conferences as a way to collect all the tweets about the session because users can set up real-time monitoring of hashtags through Twitter (Holmberg and Thelwall, 2014). This means that Twitter has the potential for promoting different research ideas and it is essential to study to what extent researchers use this channel for scholarly communication.

1.3. The overall aim of the thesis and research questions
In the last few years, the use of social media is increasing in various spheres of society. At present scientist are using social media for combined authoring, scheduling different meetings, disseminating research findings and discovering new information and research ideas (Rowland, 2002). Nowadays many scholars are using Twitter for scholarly communication.

This research aims to contribute to the field of Altmetrics research. The primary target of this study is to identify who is sharing scientific articles on Twitter authored by researchers with Finnish affiliation. Secondly, the analysis of the tweets of various researchers helps to figure out which fields are more prominent when it comes to sharing research articles. For instance, according to the blog of altmetric.com (2016), in 2011 the most popularly published articles
on Twitter belong to the field of medical science. Thirdly, the study also identifies the professional background of the publisher of tweets such as students, researchers, and professors. Fourthly, the results of the research elaborate on the social impact of sharing articles with Finnish affiliation on Twitter. Twitter is a platform where many individuals share articles with different authors. To concise the focus the author decided to research articles with at least one author with Finnish affiliation to study Altmetrics applicability at a national level (Vainio and Holmberg, 2017). Lastly, this study will help to find which types of articles that are used for future research.

In other words, the overall aim of the research is to answer the following question,
1. Who is tweeting about research publications?
2. What kinds of articles professionals are shared most on Twitter?
3. Do highly tweeted articles also receive more citations?

1.4 limitation of the study

The primary emphasis of this study is to find out who is the professional who is sharing articles on Twitter. New possibilities emerged when the author started to dig deeper into the topic. For instance, is there a possibility to study why different professionals are sharing scientific articles on Twitter? What is the reason for sharing scientific articles on Twitter? To find this answer the author of this study tried to contact Twitter profile owners via Twitter. To investigate the purposes of sharing scientific articles on Twitter, interviewing the tweet owners would have been a logical method to use. However, due to anonymous use of Twitter and because of time restriction and lack of resources, this was not feasible to do in this research. Because of the limitations, the author could not answer questions like what is the reason behind sharing mainly medical and health-related articles on Twitter, why are different professionals sharing articles on Twitter? As data gathered from VIRTA and Altmetrics.com, the author focused on studying the professionals who are sharing articles on Twitter.

1.5. Structure of the thesis

The previous section of this chapter focused on the limitation of this research. This section focus on elaborates the overall structure of this research. Chapter 2 starts with a discussion of scholarly communication, explaining different types of scholarly communication which leads up to the description of the scholarly communication process. Altmetrics is a new concept of scholarly communication. To see the larger picture of how Altmetrics has developed, theories
of all the metrics developed before Altmetrics are discussed in the later part of this chapter. Chapter 3 explores the context of social media in scholarly communication. Various aspects of social media analysis and earlier research on Twitter are presented in this chapter. The latter part of this chapter includes theories of Altmetrics, followed by the potential and the challenges with Altmetrics. Chapter 4 is all about the methodology of the thesis. Moreover, this chapter outlines how data is collected for this study. All the methods which are used in this thesis are described in this chapter. Chapter 5 This chapter includes the data analysis of the research question and the results of those analyses. Chapter 6 Includes a discussion of the results. This chapter includes a detailed overview of the findings. Chapter 7 Presents a summary of the work and suggestions for further research. Chapter 8 Includes a conclusion followed by the reference to the work and appendix.
2.0 Literature Review

Scholarly communication is a process by which different scholars are communicating with each other as they create new knowledge and publish that knowledge for future research to build upon. As technology is developing, new forms of scholarly communication are emerging. For instance, nowadays Twitter is a popular social media platform, and many researchers are now using Twitter for different purposes (e.g., Rowlands et al., 2011). It is essential to know how the process of scholarly communication works to study how Twitter might be used as a channel for scholarly communication in our changing information landscape. Altmetrics offers a new way to map scholarly communication, and it aims to measure the impact of science on social media. To understand the development of Altmetrics, it is necessary to know previous metrics-based research on scholarly communication. This chapter discusses scholarly communication and its processes, along with the definition of different metrics on scholarly communication.

2.1 Scholarly communication and the development of scholarly communication

Merton (1973) writes that the “institutional goal of science is the extension of the certified knowledge”. The contribution of the certified knowledge of a researcher is recognized through the formal scientific publication of a researcher. A research paper establishes the position of a researcher in the field of science, and through the publication, a researcher can claim the intellectual ownership of his or her idea (Holmberg, 2015, p. 11). The contribution of certified knowledge or common knowledge can be both formal and informal. The term “scholarly communication” defines the procedure of sharing the research work and publishing of research work and outcomes in formal scientific publications (Borgman and Furner, 2002). Scholarly communication refers to the method by which scholars communicate with each other as they create new knowledge (Thorin, 2006). Moreover, Case (2002), defines scholarly communication as a process by which scholars and scientists conduct the research work and make the results of their work known for other scholars and scientists. Borgman (2000) states that scholarly communication is the study of scholars of the various fields such as physical, biological, social, and behavioral sciences, humanities and technology to use and disseminate the information of the research through formal and informal channels. Holmberg (2015) says that both formal and informal contribution to the common knowledge is an integral part of scholarly communication. Scholarly communication is considered as an essential part of scientific research (Borgman and Furner, 2002; Barjak, 2006). Moreover, it is a valuable
framework when studying the impact of new participatory context and its influence on the research work (Hurd, 2000). Through scholarly communication, the research work becomes available to a broader academic community and beyond (Halliday, 2001). Scholarly communication is an essential system to create scholarly writing and to expand broad access to research publication for future research. Furthermore, it encourages researchers to publish and disseminate the research work in two ways:

1) To advance intellectual progress in their subject
2) To establish the publishing right of the author (Swan and Brown, 2004).

2.1.1 Formal and Informal scholarly communication

Traditionally, scholarly communication has been divided into two types; formal scholarly communication and informal scholarly communication (Meadows, 1998). Formal communication is impersonal in nature, and it takes place in scientific journals, citations, books, and to some extent at conferences (Barjak, 2006). Formal scholarly communication uses public and permanent vehicles such as books, and it leads to the assimilation of research results (Lancaster and Smith, 1978; Mukherjee, 2008; Zhang, 1998). The formal scientific publication is accepted as certified scientific knowledge through peer review (Holmberg, 2015, p. 11). Peer review is applied in many scholarly activities, but the primary function is in the publication of the formal journal (Rownald, 2002). The scientific peer review has been defined as the assessment of research findings for competence, implication, and originality by qualified experts or researchers (Brown, 2006). The formal scientific publication is crucial in every discipline and the publication made available for other researchers or scholars to use in the research work as a reference (Holmberg, 2015, p. 11). Scientific journals are the primary focus area of formal scholarly communication. Moreover, conference papers and monographs are also considered as essential parts of formal scholarly communication (Garvey et al., 1972; Zhang, 1998).

Informal scholarly communication refers to the communication that takes place at any time in any format such as exchanging letters, reports and texts (Aufi and Fulton, 2015). Barjak (2006) states that informal scholarly communication represents the prominent feature of scientific knowledge production rather than a mere feature of the dissemination of results. Traditionally, informal communication is known as personal and social in nature, but with the advent of the internet, it has involved for instance email, multimedia, social networking, and video conferencing (Barjak, 2006; Holmberg, 2015, p. 15). The scientific publication of a researcher
signifies the formal contribution of the certified knowledge and the conversation at a conference or sharing knowledge on Twitter, or other social networking sites are considered as an informal contribution to the common knowledge (Holmberg, 2015, p. 11). Gresham (1994) states the efficiency of the research can be influenced and facilitated by the informal interactions and communication among scholars and researchers of different disciplines.

2.2. Scholarly communication process and challenges

The process of scholarly communication starts with a research idea which is usually acquired from reading the work of other researchers, and that is undoubtedly grounded on the work of others (Holmberg, 2015, p. 11; Thorin, 2006, p.1). The steps of the cycle are followed by the research work and writing the article which builds upon the research work, and ends with a formal scientific publication (Holmberg, 2015, p. 11; Thorin, 2006, p.1). Michelson and Rothenberg (1992) state that, scholarly communication is a five-step process- 1) identify the sources, (2) communicate with the colleagues, (3) analyse the data, (4) dissemination of research findings, and (5) documentation and instruction for future scholars. Overall different scholars refer to the similar scholarly communication process.

During 18th and 19th centuries scientific journals were considered as replication of the fragmented knowledge (Day, 1999). On that time scientific journals were assumed to be an additional function of registering ownership over scientific innovation (Ramalho et al., 2005). Publication was assumed as prime indicator of professional qualification researchers and organizations (Schauder, 1994). However, scholarly communication started to face new challenges with the invention of new technology. Because of the invention of electronic journal, scholarly communication has been experiencing new challenges such as; slow broadcasting of research and development outcomes among peers, undesirable delays in publication, and excessive price hikes in journal subscription rates and so on (Harnad, 1998, 1999, 2000; liedes, 1997; Sompel and Lagozeet, 2000; Tenopir and King, 2000). Between 1985 to 2005 journal expenditure of American research increased by 273% (Venkadesan, 2009). Venkadesan (2009) also mentioned that earlier the manuscript of the articles was owned by the university and could be accessed by the university staff whenever required. However, in this digital era, many articles are uploaded online and they can only be accessed by paying a fee. Therefore, only the institutions, like universities, research institutes etc, that can afford to pay the subscription fees can access the articles. On the other hand, informal scholarly communication helps distribute articles among scholars in a diverse manner. For example, if
someone tweets any article on Twitter, it gives the opportunity for others to access and read the article.

![The cycle of scholarly communication](image)

Figure 1: The cycle of scholarly communication. Source: (Holmberg, 2015, p. 11).

2.3 Metrics research

The rise of social media and its uptake by researchers has prompted the development of Altmetrics. These new metrics can, in principle, be utilized in an evaluative role, to give an early indication of later impact of the publications (Sud and Thelwall, 2014). Altmetrics is a new way to map scholarly communication. In the following earlier metrics based research is discussed and the development of altmetrics is presented.

2.3.1 Informetrics

Informetrics is known as one of the core research areas within the field of library and information science. It is the umbrella term that covers the measurement of the information phenomena and some other areas of research that are concerned with evaluating the
Tague-Sutcliffe (1992, p.1) defined informetrics as “the study of quantitative aspects of information in any form, not just records or bibliographies, and in any social group, not just scientists”. Ingwersen and Christensen (1997, p.13) present the following definition:

“The term informetrics designates a recent extension of the traditional bibliometric analyses also to cover non-scholarly communities in which information is produced, communicated, and used.”

Informetric research is not only focused on specific types of information such as scientific communication or online information, rather than it studies all kinds of information and focuses on the discovery of mathematical models to explain properties of information (Holmberg, 2015, p. 16; Martin-Martin et al., 2016).

2.3.2 Bibliometrics

Bibliometrics is recognized as an essential field of information science because it represents a unique set of techniques for the monitoring and analysis of information resources (Patra et al., 2006). Bibliometrics is known as one of the oldest metrics-based research areas in the field of the library and information science. It is the statistical analysis of books, articles or other publications (Holmberg, 2015, p. 16).

The term "Bibliometrics" is a combination of two words: "biblio" and "metrics". The term "biblio" is derived from the combination of a Greek and Latin word "biblion", and it means book. On the other hand, the word “metrics”, is retrieved either from the Latin or Greek word "metricus" or "metrikos", respectively each means the measurement (Sengupta,1992). Pritchard (1969, p.349) defined bibliometrics as “the application of mathematics and statistical methods to books and other media of communication”. Schrader (1981, p.151) one of the teachers of bibliometrics defines bibliometrics as "the scientific study of recorded discourse". Hawkins (1977), describes bibliometrics, as the implementation of the quantitative analysis in the bibliographical references of the body of literature. According to Broadus (1987), bibliometrics is the quantitative study of physically published units such as books, journals or of bibliographic units or the substitutes of others. The most cited definition of bibliometrics comes from Tague-Sutcliffe (1992, p.1), who defined it as “the study of the
quantitative features of the production, dissemination, and use of the recorded information”. White and McCain (1989, p.119) defined bibliometrics as “the quantitative study of literature as they reflected in bibliographies”.

Moreover, Fairthorne (2005) suggested, bibliometrics is the quantitative study of the properties of recorded discourse and behaviour appertaining to it. The bibliometric research focuses on the scientific structures, and the study aims to develop mathematical models, measurement, and indicators for the process (Holmberg, 2015, p. 16). Moreover, the aim of the bibliometric study is "to shed light on the processes of written communication and of the nature and course of development of a discipline, by means of counting and analyzing the various facets of written communication” (Pritchard, 1969, p. 348).

2.3.3 Scientometrics

The concept of scientometrics overlaps considerably with bibliometrics. For instance, bibliometrics studies focus on the scientific structures and scholarly activities whereas scientometrics research concentrates on the actual publication (Holmberg, 2015, p. 16). The purpose of bibliometrics is to “improve scientific documentation, Information and communication activities by quantitative analysis of library collections and Services” (Osareh, 1996, p.150). Whereas the purpose of scientometrics is to help in “understanding the mechanism of scientific study as a social activity, a quantitative analysis of generation, and utilization of scientific Information” (Osareh, 1996, p.150). Bookstein (1997) defined scientometrics as "the science of measuring science". Nalimov and Mulchenko (1969) state that those quantitative methods deal with the analysis of science observed as an Information process referred to as scientometrics. Moreover, scientometrics research studies on the quantitative studies of science and technology (Van Ran, 1997). Scientometrics study is a part of the sociology of science and has an application to science policymaking (Jacobs, 2010). Tague-Sutcliffè (1992, p. 1) described scientometrics as “the study of quantitative aspects of science as a discipline or economic activity”. Vinkler (2010) defined, scientometrics as a scientific field which is dealing with the quantitative aspects of individuals or group of people, matters and phenomena in science, and their relationships. Scientometrics research includes the quantitative study of scholarly communication and other scientific studies such as publishing practices and research evaluation using citation analysis (Holmberg, 2015, p. 16).
2.3.4. Webometrics

Webometrics has several similarities with informetrics, and scientometric studies and with the application of the common bibliometric methods (Bjorneborn and Ingwersen, 2001). The term webometrics was first defined by Almind and Ingwersen (1997). Webometrics is known as the quantitative study of Web in the field of library and information science (Holmberg, 2015, p. 37, Thelwall et al., 2005). Bjorneborn and Ingwersen (2004, p.1217), defined webometrics as “the study of the quantitative aspects of the construction and use of information resources, structures and technologies on the Web drawing on bibliometric and informetric approaches”.

In other words, webometrics is the study of the content, structures, and use of the web using primarily quantitative bibliometric methods (Holmberg, 2015, p. 37). Webometrics has several visible contributions, but one of the most significant outputs of the webometric method is the ranking of world universities based on university websites and online impact (Aguillo et al., 2006). Webometrics research includes a broad range of studies such as studies of visualization’s (Ortega et al. 2007; Ortega and Aguillo 2009) longitudinal quantitative analysis of Web pages (Koehler 2002), citation analysis of web (Kousha and Thelwall 2006; Vaughan and Shaw, 2005), business or organization Web sites (Vaughan 2005; Vaughan et al. 2009), digital library studies (Zuccala et al. 2007), social network site investigations or studies (Ackland 2009; Thelwall 2008) and theoretical contributions (Bar-Ilan 2001; Björneborn 2006; Björneborn and Ingwersen 2004; Thelwall 2008). Moreover, webometrics studies include link analysis, web citation analysis, evaluation of the search engine and descriptive studies of the web (Thelwall, 2008).

It all started with the bibliometric method and due to the technical development along with the emergence of widened fields of scientific research and been named differently to reflect their focus and speciality (Holmberg, 2015, p. 17). Cybernetics is another method in the string of specialized research areas that has roots in bibliometric research, and it concentrates on the quantitative analysis of the website components and concepts, measure the growth and analyze the information on the internet (Holmberg, 2015; Sen, 2004).
2.4. Citation analysis

A scientific paper does not stand alone and it is supported by citation and reference of prior publications (Ziman, 1968; Smith, 1981). Moed (2015), states that bibliometric indicators are commonly accepted for the assessment of research at large scale and these indicators are most appropriate and relevant when aggregated and used for the analysis of large volumes of papers. Holmberg (2015) states that the simplest bibliometric measure is the number of publications a researcher or a group of researchers has produced. But, the number of publications can only tell about the productivity, while citations can say something about the scientific impact of the papers (Holmberg, 2015, p. 21).

Citations are a crucial part of scholarly communication. Citations are considered as links between current research and earlier research. They show the use of the previous study in new research, and it's argued that citation indicates the value of the cited study (Holmberg and Thelwall, 2014). Sandison (1989) defined citations as not only bibliographic data but also as representing the relation between author writing the current research and author writing previous research. Shaw (1979) defined citations as follows; citation creates a relationship among writers and measure of the extent to which writers connect indirectly through the literature.

“Bibliometric indicators are based on citation and widely accepted for scholarly work evaluation in decision making about governmental funding for universities, promotions of
researchers and assessing research proposal” (Holmberg, 2015, p. 21). The general assumption is that most cited work is recognized as more valuable work. Based on this assumption, citations have been used as an indicator to measure and assess different features of scholarly work and research work (e.g., Moed et al., 1985, Moed et al., 1995; Cole, 2000; Borgman, 2000). Moreover, citations are part of the academic reward system (Merton, 1973), which means that highly cited authors of different research are recognized as significant contributors to science.

“Counting citations is at the core of scientometric methods” (Holmberg and Thelwall, 2014, p.1027). It is used to determine the influence of scholarly work and to map collaboration networks between researchers (Moed et al., 1995; Cole, 2000; Borgman, 2000). However, citations can be studied for various reasons (Borgman and Furner, 2002); they can be used to investigate the impact of scholarly communication, to map the development of science, to analyze the geographical and organizational distribution of science (Holmberg, 2015, p. 22).

As there are disciplinaries differences both publishing and citation traditions vary among disciplines, new approaches are desired to measure the visibility and impact of studies. In this situation, social media may produce new ways to investigate scientific output (Priem and Hemminger, 2010). One of the new social networking services that scholars can use in scholarly communication and that has some potential to generate new ways to measure research impact is Twitter (Holmberg and Thelwall, 2014).
3. Social Media, scholarly communication in Twitter and Altmetrics

The arrival of social media has changed many aspects of scholarly communication. Now researchers can present their ideas in blogs, discuss and share their research with others on different social media platforms such as Facebook, Twitter, and Mendeley. Research has been moved out from the closed scientific ecosystem to the open web (Holmberg, 2015, p.1). A new form of scholarly communication needs new methods of measurement of impact. This chapter presents the emergence of social media, scholarly communication in Twitter and the development of altmetrics.

3.1 Social Media

In the early days of the web most of the web users were consumers because creating a website requires at least some basic knowledge about HTML, but nowadays technological development makes it easier to also be a producer in addition to consumer of contents on the web in Wikipedia, blogs and social networking sites, for instance, Facebook, Twitter, Instagram. The development of the technologies leads to a new version of a network which is known as web 2.0 (O’Reilly, 2005). On the web, 2.0, people are not only consumers but also producers (Holmberg, 2015, p. 45). The change in the role of web users helps to create new types of websites. The purpose of these websites is “to let people interact with each other and create content in collaboration with each other” (Holmberg, 2015, p.45). These sites are known as social networking sites because these websites enable people to interact with each other and maintain their social network online. There are many definitions of social networking sites, but most of them acknowledge that social networking sites enables its users to share or create content and helps them to participate in social networking (e.g., Boyd and Ellison, 2010; Thelwall, 2009). Some of the popular social networking sites are Facebook, Twitter, Wikipedia, YouTube and LinkedIn. Social networking sites are the most popular sites on the web. We receive news and entertainment as a recommendation of content through different social networking sites. Social networking sites give us a platform to connect with people. Today people participate in online conversations and create content on the web. Social networking sites depend on user-produced content to even exist. Twitter would not exist without our tweets; Facebook would not survive if anyone doesn’t share any content and connect with their friends on it and Instagram would not survive if nobody shared any pictures with hashtags. In other word, people are creating a social web and helping to the expansion of web content by generating a massive amount of substances (Holmberg, 2015, p. 47). Social
media has exploded as a type of online discourse where people creates different content, share it, bookmark it and network at a remarkable rate (Asur and Huberman, 2010).

The web in general and different social media, has become a rich source of data. Researchers and scholars of different disciplines can analyze these data and generate new knowledge about the social media user’s behaviour. Ward and Barker (2013) state that social media generates enormous amounts of data. Recently, the term big data has become a term ubiquitous. It is a term which not only refers to the immense amount of data that can be processed and shaped into patterns and trends but also occasionally to the technologies and practices used to analyze the data (Boyd and Crawford 2012, p. 63).

The content and the users connect in social media. An integral part of social media is to create content. This content includes blogs, tweets, status updates, likes, messages, emojis, tags, data from location-based social networks, wikis, videos, photos, tags, bookmarks etc. Baribier and Liu, (2011) said that the social media dataset is (1) large, (2) noisy, and (3) dynamic in nature. The datasets collected from different social media sites can be massive. For instance, consider that over 500 million tweets are sent every day on Twitter, or 1.4 billion users are sharing and creating content on Facebook. Because of an active global use of social media, “social media datasets can be noisy and contain spam and other false positives to the intended data collection” (Holmberg, 2015, p. 49). Social media datasets are dynamic which means that they are changing all the time; new content is created while old content is shared, modified and even deleted. This raises new challenges for researchers and new tools and new methods to analyze social media data are needed.

3.1.1. Analyzing Social Media

There are various methods for analyzing social media, but among the most frequently used methods are for instance 1. text analysis or text mining, 2. social network analysis, 3. trend analysis (Stieglitz et al., 2014).

The aim of text analysis is to generate mining options and attitudes from online texts either through sorting of texts or automated analysis of texts. Text analysis includes investigation about online communication and public opinions about policies, trending political topics and so on. In this research, a co-word analysis is conducted based on the profile description of the Twitter authors.
In social network analysis, the purpose of the researchers is to investigate either a specific node’s position in the network or analyze the overall structure of the network (Holmberg, 2015, p. 50). Using social network analysis researchers can find influential people or organizations based on the number of connections they have to other people in the network and analysis of the structure of the networks can help to reveal some information about the communication patterns or the patterns of the information flow in the network (Holmberg, 2015, p. 50). Different researchers have used social network analysis for different purposes such as: for market communication (Jansen et al., 2009; Hennig-Thurau et al., 2010), climate change communication (Pearce et al., 2014; Kirilenko and Stepchenkova, 2014), political conversations (Golbeck et al., 2010; Kim and Park, 2011; Stieglitz and Dang-Xuan, 2013), dissemination of health information (Brownstein et al., 2009; Scanfeld et al., 2010), and news (Larmena and Ghosh, 2010; Bruns and Burgess, 2011). In this research social network analysis used to find the influential persons or organizations who shared most of the research articles written by at least one Finnish researcher on Twitter.

Trend analysis is the most complex of the three types of analysis. The development of computing power and statistics have made it possible to detect trends in real time and able to predict the emerging trend of social media based on the social media data. By analyzing trends in social media data researchers have for instance predicted influenza epidemic (Signorini et al., 2011) and Box office revenues (Asur and Huberman, 2010).

3.2 Previous research on Twitter

Microblogging has emerged as a new form of communication in which users can, for instance, express their opinions and share information in short posts (Java et al., 2007). Microblogging also a useful platform for various academic purposes and contexts (Herwig et al., 2009; Mollett et al., 2011). Twitter is a popular microblogging service (http://www.Twitter.com/), it was launched in 2006 and users were able to write messages or tweets that were 140 characters or less (Priem et al., 2010). Twitter is more commonly used in disseminating information rather than as a platform for more personal connection among individuals (Neiger et al., 2013). At present scholars are increasingly using Twitter as a communication platform (Priem et al., 2010). A lot of research has focused on Twitter, but the research is concentrating on understanding the categories of Twitter users and their behaviour (Vainio and Holmberg, 2017). However, some studies have focused on Twitter user’s profile descriptions (see, e.g., Mislove et al., 2011; Semertzidis et al., 2013; Uddin et al., 2014). These studies are primarily
based on random samples of Twitter messages and Twitter users, collected using Twitter APIs (Semertzidis et al., 2013; Uddin et al., 2014). In some of the related work, machine learning methods have been used to categorize Twitter users into different classes (Uddin et al., 2014).

Moreover, there are no comprehensive studies to show how, why and in what ways scholars are using Twitter (Mahrt et al., 2014). However, some researcher has found that scholars are actively using Twitter (see, e.g., Bader, Fritz, and Gloning, 2012; Gerber, 2012; Harley et al., 2010). Moreover, Gu and Widén-Wulff (2011) found that 14% of the respondent scholars at a Finnish university are using “micro-blogs” frequently. Different researchers (See, e.g., Priem and Costello 2010; Holmberg and Thelwall 2014, Hadgu and Ja¨schke 2014, Haustein et al., 2014), have been researching academic activities on Twitter through the examination of how scholars are using Twitter and scholarly tweets. These studies have some limitations such as small sample size, and some have biased samples by age as well as academic disciplines.

In Holmberg and Thelwall’s (2014) study, the scholars were chosen from ten different disciplines. In that study, they found dissimilarities in scholarly communications on Twitter between different subjects; scholars from biochemistry, astrophysics and cheminformatics used Twitter for scholarly communications more frequently, whereas sociologists, were not actively using Twitter for that purpose.

Some of the studies show which types of scholarly sites are more popular than others. For instance, Ke et al.’s (2016) results show that scientist favour’s certain disciplines such as natural science or health science on Twitter. Vainio and Holmberg (2017) study investigated different people’s tweeting of scientific articles that had at least one Finnish author. In their research, they studied the occupation of the user groups who tweeted scientific articles, and their findings suggested that Medical and Health Science 13.5% attract significant numbers of attention compared the articles from Social science articles.

Gerber (2012) found Twitter to be among the most well-known Web 2.0 applications among scholars. In one study (Thelwall et al., 2013), approximately three-quarters of the total number of tweets stating articles from Science and nearly 90% of the total amount of tweets mentioning articles from natural science. Academic research is not only shared on for disseminating research information, but academic research can be shared for marketing purposes by the publishers and authors themselves (Nelhans and Lorentzen, 2016). This demonstrates that there is a need for further investigation of Twitter users who tweet scientific articles and what types of scientific articles they are sharing on Twitter.
The arrival of social media has changed many aspects of scholarly communication. The idea behind Altmetrics is to find out the trace of scholarly communication and to track the attention research has received in social media and discover the impact of the research (Holmberg, 2015, p. 1).

Any research evaluation aims to assess the value or quality of the research in comparison to other research. As quality is subjective and difficult to measure; citations are used as a proxy to evaluate the quality of research. An essential part of scholarly communication is citations, and it is considered as a significant component of research evaluation, with the assumption that highly cited work is more valuable than the less cited research work (e.g., Moed et al., 1985, Moed et al., 1995; Borgman, 2000). Citations are used to indicate the scientific impact, but recently funders and other stakeholders are demanding not only scientific impact but also societal implications; Altmetrics is the new method to evaluate the data and indicate the effect that research has made on various audience.

The term “Altmetrics” comes from “alternative metrics” (Holmberg, 2015, p. 4). Altmetrics is understood as online metrics that measure the impact or visibility of science and act as an alternative to traditional citation (Haustein, 2016). Priem (2014, p. 266) defined Altmetrics as the “study and use of scholarly impact measures based on the activity in online tools and environment”. Moed (2015, p.1) defines Altmetrics as “‘traces of the computerization of the research process’”. Adie and Roe (2013, p.12) wrote that “Altmetrics offers an alternative to the current practice of relying only on citation counts and journal impact factors for the quantitative analysis of impact by introducing new complementary methods and sources of data”. This emphasizes that Altmetrics is an alternative view on research indicators and a new approach to research evaluation.

3.2.1 Potential and limitations of Altmetrics

As discussed, earlier Altmetrics are known as online indicators of research impact. These indicators can be a tweet containing a URL of a scientific article or Facebook post containing a URL of a journal or an article which is bookmarked on Mendeley or other research product that have mentioned in any social media sites.
Potential of Altmetrics

At present, it is not clear how to examine the impact of research on other areas of society than science. While bibliometrics and peer review have been considered as standard methods for measuring the impact of scientific research, on the other hand, alternative metrics can be viewed as an exciting option for measuring the societal impact of research (Bornmann, 2014). Last few years researchers from different disciplines have published articles and editorials presenting on Altmetrics (e.g., Galligan and Dyas-Correia, 2013; Osterieder, 2013; Brigham, 2014; Crotty, 2014; Galloway et al., 2013). A standard feature of most of these works is that the researchers see a lot of potential and possibilities with Altmetrics. Some of the researchers express their concerns and suggest cautious uptake while others predict the end of citation-based research evaluation.

Citation-based research evaluation and the journal impact factor have been criticized for several things such as the time it takes for citations to accumulate and the disciplinary differences in publishing and citing traditions, which places some disciplines in a disadvantaged situation. Because of this, subjects such as humanities and social sciences have been underrepresented in citation databases and left out from citation-based assessment (Holmberg, 2015, p. 65).

Now, in social media, scholars regardless of discipline can share information about their research for instance in blogs and Twitter and they can do so long before their study has been formally published in a traditional scientific journal. One of the benefits of altmetrics is that it can measure the impact of research product on social media. Another benefit of Altmetrics is that it uses openly available data. Because Altmetrics are using publicly available data it gives a level of transparency which is not possible when using citation-based indicators.

Overall altmetrics have the potential to:

**Measure broader impact.** One of the benefits of Altmetrics is that it helps to capture the impact and visibility of a much wider range of research product. Altmetrics have the potential to measure broader impact of research (Priem et al., 2012). Altmetrics has the potential to provide transparent description of interest, usage, influence of scholarly products and it can provide diverse form of impact analysis than the traditional metrics like bibliometrics (Fausto et al., 2012; Taylor, 2013; Waltman and Costas, 2014). According to Fenner (2013), altmetrics measure the impact of reasearch in more practical fields, and papers better than with citations.
Assessment of different disciplines. Altmetrics can be beneficial for the assessment of disciplines such as humanities and social science (Hammarfelt, 2014; Haustein et al., 2015). It can help to understand how researchers of different disciplines are using social media in general (Rowlands et al., 2011) and Twitter in particular.

Speed. One of the main drawbacks of citation analysis is that it can provide the measurement of the reliability and impact of a journal several years after publication (Wang, 2013). Whereas altmetrics take only few days or weeks to measure the impact of a paper or other product after it has appeared (Haustein et al., 2014; Mohammadi and Thelwall, 2014). At present, several Web 2.0 tools offer real-time access to organized altmetric data through application programming interfaces (APIs) (Priem and Hemminger, 2010) and by using these tools the impact of the scientific paper can be tracked at any time after publication.

Wider adoption of open science. Altmetrics are close to open science movement and it favours open science publications (Adie, 2014). In this scenario researchers who publish online paper and institutions with open access repositories have more advantages and altmetrics may favour them. Ovadia (2013) said that, if altmetrics become more accepted than faculty with more openly accessible work might find themselves in an advantageous position than those faculty whose work behind a subscription-based paywall. Altmetrics may provide a massive incentive for broader and rapid adoption of open science movement.

With these potentials, Altmetrics may be able to deliver much more nuanced and diverse view of research than citation alone can provide. Citation analysis can tell only about the scientific impact and attention whereas Altmetrics have the potential to scholarly communication and research impact at an unprecedented speed and level of detail (Holmberg, 2015, p. 72).

limitations of Altmetrics
Besides having great potential Altmetrics have some disadvantages as well. For example, lack of understanding of what Altmetrics means and what this new indicator can tell about the research impact. One of the serious critiques of Altmetrics is mainly related to the data and lack of understanding about what the data represents. According to Priem (2014), lack of definition, ease of gaming and possible biases are the significant three limitations of Altmetrics.

Some of the main critiques of Altmetrics:
Manipulation. One of the disadvantages of altmetrics is ease of manipulation. It is easier to manipulate altmetrics than bibliometrics (Rousseau and Y Ye, 2013; Thelwall et al., 2013) and
there are many opportunities to manipulate altmetrics. Since “social websites do not have quality control and do not have a formal procedure to link users to offline identities, it would be easy to systematically generate high altmetric scores for any given researcher or set of articles” (Thelwall, 2014, p. 4). For instance, Twitter mentions can be automatically generated through fake accounts (Darling et al., 2013; Liu and Adie, 2013; Galligan and Dyas-Correia, 2013).

**Biases.** The measurement of impact is related to a specific sample of people who have mentioned a paper on social networking sites (Bornmann, 2014). This sample can be biased; for instance, sample has a biased towards younger people or people who have professional interests (Priem, 2014; Neylon et al., 2014). There are no accurate sample description or user statistics for each social networking sites is available. The lack of a standard for data and method means the biasness of the sample cannot be quantified.

**Multiple version.** Publications often exist in numerous forms for instance preprints on arXiv.org and post-prints from a publisher (Bornmann, 2014; Holmberg, p.72, 2015). Therefore, using altmetrics to measure their impact would result in ambiguity and redundancy (Liu and Adie, 2013).

**Lack of standard for data and method.** Altmetrics does not have the standard rules for data collecting, aggregating and presenting the data. As there are no standards, a provider of altmetrics data can receive data from different sources. For instance, one service provider collects Facebook “likes” whereas other service providers can collect only the mentions of research products on Facebook status. While one service can count page view and other services can check unique visitor on the web page. It means differences have been found in altmetrics from different service providers (Zahedi, Fenner, and Costas, 2014) and it creates confusion among those interested to use altmetrics in their work.

**Lack of understanding.** Lack of proper definition is another limitation of altmetrics. A theoretical problem arises from a lack of understanding about what is the meaning of altmetrics and what it does (Holmberg, p.73, 2015). There is uncertainty about what altmetrics measure (Bornmann, 2014).
4.0 Data and Methodology

This chapter describes the data collection process and different methods used in this research. Throughout this chapter, mixed research methods will be introduced, along with quantitative and qualitative research method. Moreover, this chapter includes an explanation of the fundamental motivation for choosing this approach for this research.

4.1 Research Design

A research design is one of the essential parts of the research. Kothari (2004, p.31), said that “A research design is the arrangement of conditions for collaborating and analyzing of data in a manner that aims to combine relevance to the study purpose with economy in procedure”. In fact, “the research design is known as the conceptual design within which the research is conducted; it prepares the blueprint for the collection, measurement and analysis of data” (Kothari, 2004, p.31). Research design enables the smooth sailing of various research operations such as designing the research questions, data collection process, research methodology and data analysis etc. which helps to make the research as efficient as possible.

4.2. Research Methodology

"Research methodology is known as a way to scientifically solve the research problem” (Kothari, 2004, p. 8). To conduct a legitimate study and ensure reliable findings from the research, it is essential to select an appropriate method for the data analysis, which meets the requirements set by the core research questions and the purpose of the study.

There are two conventional methods for collecting data and analysis are regularly used by the researchers, quantitative and qualitative method. Along with quantitative and qualitative research, mixed research methods are becoming increasingly popular, and recognized as a third significant research approach (Johnson et al., 2007). As Creswell (2007, p.2), defined “mixed methods research as a research methodology where the researcher collects, analyzes, and mixes (integrates or connects) both quantitative and qualitative data in a single study or a multiphase program of inquiry”. Mixed method research is a type of research method in which the researcher blends or combines quantitative and qualitative study techniques, methods, approaches, concepts or notions into a single study or a set of related studies (Johnson et al., 2007). This research aims to identify professionals who are sharing research articles on Twitter and to meet that aim mixed methods approach is suitable for this study.
4.2.1. Quantitative and qualitative research method

Quantitative research is based on the analysis of numerical data whereas qualitative analysis is usually based on the analysis of the non-numerical data. Most of the analysis of this study is based on quantitative data, and quantitative research method is used to answer the following research questions:

Q.1. Who is tweeting about research publications? Q.2. What kinds/types of articles shared most on Twitter? Q.3. Do highly tweeted articles also receive citations?

To better understand the quantitative research methodology, a detailed overview of the method is given below:

“Empiricism” (Leach, 1990) and “Positivism” (Duffy, 1985) are the terms that can describe the quantitative research methods. Quantitative research methods are primarily methods that allow researchers or scholars to explore and investigate data and findings numerically or quantified statically. This research method is an “objective, formal, systematic process in which numerical data are used to quantify or measure phenomena or findings” (Carr, 1994). In the words of Aliaga and Gunderson (2002), quantitative research is “explaining phenomena by collecting numerical data that are analyzed using mathematically based techniques”. Quantitative research methods are frequently used to accumulate numerical data which can be transformed into statistical information which is applied to validate or invalidate a hypothesis or theory (Gerrish and Lacey, 2010).

According to Muijs (2010), a quantitative research method is applied when a researcher attempts to answer research questions which start with “how many” and that generate numerical answers. The outcome attained from quantitative research can be used for making a prediction, predicting trends and describe the affiliation between variables (Gerrish and Lacey, 2010). Quantitative analysis is frequently implemented to study numerical changes.

The primary step of the quantitative approach is to decide about the selection of a case study or samples (Neuman, 2013). Next step of the process is data collection. A quantitative researcher must be very cautious about recording the data and verifying the information and transferring the data into a computer readable format (Neuman, 2013). “Quantitative data helps to establish the correlation between variables and outcomes; quantitative data allow others to authenticate original findings by independently duplicating the analysis” (Choy, 2014, Dudwick, et al., 2006: 3). After that, the research ends with a computer-generated output which
The qualitative data analysis method is one of the most crucial parts of the research process. Qualitative research is categorized by its objectives, and qualitative research generates words rather than numbers as for data analysis (McCusker and Gunaydin, 2015). Qualitative methodology refers to the wider sense of research that produces descriptive data (Taylor et al., p.7, 2015). Qualitative research produces descriptive types of data and to produce such kinds of data researcher needs to listen and closely observe the source of data (Taylor and Bogdan, 1998). Choy (2014), said that in the qualitative analysis a wide range of techniques or methods is used for data collection and analysis. For instance: Content analysis is one of the most common analysis that is used to interpret qualitative data (Zhang and Wilderrmuth, 2016). Content analysis requires an inductive approach where the raw data of the research are compressed into different categories (Zhang and Wilderrmuth, 2016), and it is applied for the semi-structured interviews.

Some parts of this study use qualitative analysis. Co-word analysis is used was used to answer this research question Who is tweeting about research publications? Co-word map of Twitter profile of people who are tweeting about scientific articles is used to answer this question. Co-word analysis is a quantitative method but the interpretation of the results of Co-word analysis required qualitative approach.

4.3 Data

In this study quantitative data was derived from the VIRTA database and Altmetrics.com. After that the data was analyzed by the means of basic statistical methods on SPSS and Excel.

**VIRTA database**

Since 2011 Ministry of Education and Culture in Finland collects publication information from Finnish educational and research institution (Nikkanen, 2019). VIRTA publication information system is an innovative data warehouse solution to assimilate institutional data at the national level. It is a data hub which complies bibliographic information of all scientific articles from Finnish research institutions (Nikkanen, 2019). VIRTA provides up to date metadata to higher educational institutes and other research organizations. Moreover, the annual data is publicly available in Vipunen (statistical information) and Juuli (bibliographic data) services. (Nikkanen, 2019). In this study, data is collected from Juuli bibliographic data service. It contains information on publications by the Finnish research institution. Finnish National
Everyday thousands of messages about scholarly content are published online. “Altmetric tracks a range of sources to gather and store this activity, helping researcher to monitor and report on the attention surrounding the work one care about” (Altmetric, 2019). The company provides reliable and transparent data to publishers, research institutions, researchers and funders (Altmetric, 2019). Altmetric.com have a functioning data sharing policy and multiple research projects have used the data that altmetric.com have provided for instance (Vainio and Holmberg, 2017; Robinson et al., 2014; Zahedi et al., 2014).

The aim of focusing the data collection to authors with a Finnish affiliation give us the opportunity to investigate the applicability of altmetrics at a national level. After collecting the data, the articles were categorised into Web of Science (WOS)main categories and then merged into Agriculture, Biological science and Engineering (ABE), Medical and Health Science (MHS), Social Sciences and Humanities (SHS), and Natural Science (NS). After that, the next step was to use the DOI of the publications to match the publications with the data provided by Altmetric.com. The next stage of the process was to select 400 random tweets about publications of at least one author with Finnish affiliation from the data for data analysis. Data of more than 50000 publications was stored in the dataset and from that 400 publications randomly chosen by using a formula in Excel. The process of data collection and sample selection was thus as depicted in figure 3.
4.4 Methods of the study

In this research both quantitative and qualitative methods were used to investigate who is tweeting about science. To answer the research questions approaches such as coding gender of the tweet authors, coding the profile descriptions and co-word analysis was used.

In the first step of the statistical analysis, the codebook of the gender of the tweets was created to answer who is tweeting about research publications? Qualitative content analysis was used to categorize the gender of the tweet authors. Hsieh and Shannon (2005, p.1278) defined qualitative content analysis as "a study technique for the subjective interpretation of the content of text data through the systematic classification process of coding and identifying themes or patterns”.

The profiles of the Tweet authors were coded based on the following four categories. All the categories can be seen in Table 1. The author manually checked the profile name, profile
picture and profile description. The male or female was identified based on the profile descriptions, profile picture and names of the Tweet author. By checking Tweet authors profile description, profile picture and name author of this research found some Tweets were posted on behalf of on behalf of an organization, company or institution. To categorize these, the author decided to introduce the category “Not Applicable”. In addition to that, the author found some profile which was not used and some of the profile could not be distinguished as an individual or organization. Therefore, the author had to come up with another category called “Unknown”.

<table>
<thead>
<tr>
<th>Categories</th>
<th>Gender</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Male</td>
<td>Tweet author is male</td>
</tr>
<tr>
<td>2</td>
<td>Female</td>
<td>Tweet author is Female</td>
</tr>
<tr>
<td>3</td>
<td>Unknown</td>
<td>Gender of the Tweet is not identified.</td>
</tr>
<tr>
<td>4</td>
<td>Not applicable</td>
<td>Tweet author is not one single person, they are Tweeting on behalf of an organization, company or institution.</td>
</tr>
</tbody>
</table>

Table 1: Codebook of gender

The next step of the process is to find out professions of the tweet authors. In this study the profile description of the tweet authors manually checked and coded by using qualitative content analysis technique. After analyzing the profile information, the occupation or background of the Twitter author is identified. Full list of categories and occupation list can be seen in Table 2.
Table 2: Codebook of occupation.

The third step of the study is using co-word analysis. In this study Co-word analysis is used to analyze the qualitative data. Co-word analysis is a content analysis technique which “uses forms of co-occurrence of sets of objects (i.e., words or noun phrases) in a corpus of texts to recognize the relationship between ideas within the subject areas presented in texts” (He, 1999). Callon et al., (1991) said that it is a content analysis which uses both the frequency of items and their connections or existing relationships between them. The relationships are extracted from the co-occurrences of pairs of words and the frequencies of the co-occurrences.
are used to determine the strength of these connections. The connections of the items or words are then gathered and used to draw a network map for more qualitative interpretation.

Three steps are involved in Co-word analysis (Callon et al., 1991; He, 1999):

1. Extracting key words from selected materials or database
2. Creating the co-occurrence matrix and drawing the network maps
3. Interpretation of the network maps and the data they represent.

The first step of the process is most crucial for the outcome of the analysis. It is crucial because extracting and selecting of data can influence the data. When indexing is done researcher needs to complete a level of generalization of the indexing words. The next step of the process is to create the co-word matrix and converting the raw data into co-relation matrix (Holmberg, et, al., 2009). The creation of the matrix and the visualization of the data into maps are done by the computer software’s such as Gephi, BibExcel (Persson, 2006). The last step of the process is to analyze and interpret the network maps. In this study all these three steps are followed and presented in the result chapter.
5.0 Results

This chapter includes the analysis of the data based on the research question and results as well as provided the clear overview of the research analysis.

5.1 Analysis and Results

Q.1. Who is tweeting about research publications?

To examine who is tweeting about scientific articles, the profile information of the tweet authors was manually checked. To answer our research question better, the data was categorized according to gender and profession of the tweet authors.

Gender of the authors of the tweeters

After the analysis of the tweeter profile, we found out who is tweeting about articles with Finnish affiliation. The details result of the analysis can be seen in Table 3.

<table>
<thead>
<tr>
<th>Categories</th>
<th>Gender</th>
<th>N</th>
<th>Percent</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Male</td>
<td>161</td>
<td>40.3%</td>
</tr>
<tr>
<td>2</td>
<td>Female</td>
<td>58</td>
<td>14.5%</td>
</tr>
<tr>
<td>3</td>
<td>Unknown</td>
<td>48</td>
<td>12.0%</td>
</tr>
<tr>
<td>4</td>
<td>Not applicable</td>
<td>133</td>
<td>33.3%</td>
</tr>
<tr>
<td></td>
<td>Total</td>
<td>400</td>
<td>100.0%</td>
</tr>
</tbody>
</table>

Table 3: Gender of the tweets Authors

From the analysis, we can see that 161, or 40.3% percent of the tweeters in the random sample are male and 14.5% are female. 133 or 33.3 % of the authors of the tweets were either institution or an organization or there are more than two persons who are responsible for that
Author: Anik Dutta

account. Moreover, 12% of the sample the gender or the type of account could not be determined.

**Occupation of the authors of the tweets**

In this study the author manually checked the profile information of the profiles of the tweeters. After analyzing the profile information, the occupation or background of the tweeter was identified. It was challenging to extract profile information of some users due to deactivation of some accounts and lack of information in the profile description.

The result of the analysis of the tweeters occupation was as follows:

<table>
<thead>
<tr>
<th>Categories</th>
<th>Occupation</th>
<th>N</th>
<th>Percent</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Professors (Associate, Assistant etc.)</td>
<td>39</td>
<td>9.75</td>
</tr>
<tr>
<td>2</td>
<td>University research/teaching staff</td>
<td>7</td>
<td>1.75</td>
</tr>
<tr>
<td>3</td>
<td>PhD student/candidate</td>
<td>2</td>
<td>0.5</td>
</tr>
<tr>
<td>4</td>
<td>Blogger</td>
<td>52</td>
<td>13</td>
</tr>
<tr>
<td>5</td>
<td>CEO/Entrepreneur</td>
<td>3</td>
<td>0.75</td>
</tr>
<tr>
<td>6</td>
<td>Researchers</td>
<td>39</td>
<td>9.75</td>
</tr>
<tr>
<td>7</td>
<td>Health care</td>
<td>53</td>
<td>13.25</td>
</tr>
<tr>
<td>8</td>
<td>Journal Editor</td>
<td>29</td>
<td>7.25</td>
</tr>
<tr>
<td>9</td>
<td>Columnist/Author</td>
<td>17</td>
<td>4.25</td>
</tr>
<tr>
<td>10</td>
<td>Organization</td>
<td>76</td>
<td>19</td>
</tr>
<tr>
<td>11</td>
<td>Student/Medical student</td>
<td>5</td>
<td>1.25</td>
</tr>
<tr>
<td>12</td>
<td>Unknown</td>
<td>56</td>
<td>14</td>
</tr>
<tr>
<td>13</td>
<td>Other</td>
<td>22</td>
<td>5.5</td>
</tr>
<tr>
<td>Total</td>
<td></td>
<td>400</td>
<td>100.0</td>
</tr>
</tbody>
</table>

Table 4: Occupation of the tweet authors

From the above table we can see that around 12% of the tweeters were working as Professors (Associate, Assistant etc.), or University/teaching staff or PhD student/candidate and 22% of the tweeters were working in academia. 76 out of 400 (19%) of the tweet accounts belonged to different organization such as: Journal publishing companies, Educational institutes and
Research Institutes. 13.3 % of the tweeter profile are from health care professionals and 14% of the total sample are unknown.

Q.2. What kind/types of articles professionals are shared most on Twitter?

In the next stage of this research, the different types of articles shared on Twitter by different professionals were investigated.

Every tweet author in the sample shared different kinds of articles on their Twitter profile. Some of the articles are shared by multiple authors. This study focused on how many times articles were shared on Twitter rather than how many unique articles were shared on Twitter.

The articles can be either open access or close access. Open access means that one can read an article anytime without paying for access. For reading these types of the journal, the person does not need to subscribe to any publication sites. On the other hand, closed access articles are articles which cannot be found online for free. One needs to pay for access these types of articles. Educational or research institute usually buy the subscription of the publication sites.

In the random of 400 tweets most articles are openly accessible (table 5). A total of 370 tweets of the random sample mentioned that were open access while 30 mentioned about subscription-based articles.

<table>
<thead>
<tr>
<th>Categories</th>
<th>N</th>
<th>Percent</th>
</tr>
</thead>
<tbody>
<tr>
<td>Open access</td>
<td>370</td>
<td>92.5</td>
</tr>
<tr>
<td>Close access</td>
<td>30</td>
<td>7.5</td>
</tr>
<tr>
<td>Total</td>
<td>400</td>
<td>100</td>
</tr>
</tbody>
</table>

Table 5: open or close access articles.

Next it was investigated which research area subjects gets more attention on Twitter. The articles were categorized in four categories Agriculture and biological science and engineering (ABE), Medical and health science (MHS), Natural science (NS) and Social sciences and Humanities (SSH). A comparison between open or closed access articles by area of science is given below:
Types of Articles

<table>
<thead>
<tr>
<th>Categories</th>
<th>ABE (n)</th>
<th>MHS (n)</th>
<th>SSH (n)</th>
<th>NH (n)</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Open access</td>
<td>15.25%</td>
<td>50.25%</td>
<td>4.25%</td>
<td>22.75%</td>
<td>92.5%</td>
</tr>
<tr>
<td></td>
<td>(61)</td>
<td>(201)</td>
<td>(17)</td>
<td>(91)</td>
<td>(370)</td>
</tr>
<tr>
<td>Closed access</td>
<td>0.75%</td>
<td>5.75%</td>
<td>0.25%</td>
<td>0.75%</td>
<td>7.5%</td>
</tr>
<tr>
<td></td>
<td>(3)</td>
<td>(23)</td>
<td>(1)</td>
<td>(3)</td>
<td>(30)</td>
</tr>
<tr>
<td>Total</td>
<td>16%</td>
<td>56%</td>
<td>4.5%</td>
<td>23.5%</td>
<td>100%</td>
</tr>
<tr>
<td></td>
<td>(64)</td>
<td>(224)</td>
<td>(18)</td>
<td>(94)</td>
<td>(400)</td>
</tr>
</tbody>
</table>

Table 6: Percentage of open or close access article and types of article

From the above table we can see that 16% of articles are ABE, MHS articles shared 224 times or 56%, NS articles shared 94 times (23.5%) and SSH articles shard 18 times (4.5%). Vainio and Holmberg (2017) 7 out of 20 articles studied were from natural science articles. Ke et al.’s (2016) results showed that scientist favor’s natural science articles were more frequently shared. In this research it was found that MHS or NS articles received more attention than SSH articles.

The aim of this study is better understanding what types of articles with Finnish affiliation are shared in Twitter and by whom these articles are shared. This is presented in table 7 below:
Based on the results health professionals frequently shared Medical and health-related articles. It is also noted that different organization’s share mostly Medical and health-related articles in their tweets. Moreover, it was discovered that SSH articles get less attention among professionals. We can see that SSH articles were shared only 18 times and academic professionals shared most of the SSH articles 10 times.

**Q3. Do highly tweeted articles also receive citations?**

Citation analysis is frequently used as a tool for impact evaluation (Garfield, 1972). The more the articles are cited the more impact they have. The number of citations to the analyzed articles were counted and analyzed. Number of citations of the tweeted articles are shown in table 8 below:

<table>
<thead>
<tr>
<th>Occupation</th>
<th>ABE</th>
<th>MHS</th>
<th>NS</th>
<th>SSH</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Academia (Professors university researcher, PhD, Medical student, Researcher)</td>
<td>20</td>
<td>31</td>
<td>31</td>
<td>10</td>
<td>92</td>
</tr>
<tr>
<td>Blogger</td>
<td>8</td>
<td>33</td>
<td>10</td>
<td>1</td>
<td>52</td>
</tr>
<tr>
<td>CEO/Entrepreneur</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>3</td>
</tr>
<tr>
<td>Health care</td>
<td>2</td>
<td>45</td>
<td>6</td>
<td>0</td>
<td>53</td>
</tr>
<tr>
<td>Journal Editor</td>
<td>5</td>
<td>18</td>
<td>3</td>
<td>3</td>
<td>29</td>
</tr>
<tr>
<td>Columnist/Author</td>
<td>3</td>
<td>8</td>
<td>6</td>
<td>0</td>
<td>17</td>
</tr>
<tr>
<td>Organization</td>
<td>13</td>
<td>49</td>
<td>13</td>
<td>1</td>
<td>76</td>
</tr>
<tr>
<td>Unknown</td>
<td>7</td>
<td>26</td>
<td>21</td>
<td>2</td>
<td>56</td>
</tr>
<tr>
<td>Other</td>
<td>5</td>
<td>13</td>
<td>3</td>
<td>1</td>
<td>22</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td><strong>64</strong></td>
<td><strong>224</strong></td>
<td><strong>94</strong></td>
<td><strong>18</strong></td>
<td><strong>400</strong></td>
</tr>
</tbody>
</table>

Table 7: Who is sharing what types of articles
Table 8: Number of citations of the tweeted articles.

In table 8, result shows that around 82.5% of the articles were cited around 0-20 times whereas only one (0.3%) of the articles were cited around 1001-2000 times.

In this study spearman correlation was conducted to find out if highly tweeted articles are also highly cited or not. Because of the skewed nature of the data the spearman rank correlation was suitable for this analysis. The spearman rank correlation is a nonparametric technique which is used to determine correlation between to independent variable (Gauthier, 2002).
Table 9: Correlation between number of time articles with at least one Finnish author mentioned on a Tweets and number of times articles was cited.

In Spearman, analysis correlation is significant at the 0.05 level. In this study, the number of times articles were tweeted and cited articles were negatively correlated. The correlation value between the number of times scientific articles were tweeted and cited is -.124. The correlation value of this analysis is not statistically significant. It means highly tweeted articles were not highly cited.

The third step of the analysis is co-word analysis. The co-word network map was analyzed with BibExcel and visualized using Gephi. The map is focused on the most frequently mentioned word (at least five times or more) in the profile description of the tweeters of the scientific articles. The thickness of the edges or links between the words reflect how often the words were mentioned together. In figure 4 only those connections or links between the nodes that appeared twice or more together are visible. The layout algorithm used to position the words in the network was Force Atlas, which is a spring-based algorithm treating the edges as if they were springs between the nodes, pulling the nodes closer and trying to find balance in the network system.
Figure 4: Co-word map of the most frequently mentioned words in Twitter profiles of people who were tweeting about scientific articles

In figure 4, the size of the nodes represents how often the words were mentioned together. In the graph words like health, research, Twitter, professor and science stand out as the most frequently used word in the profile description of the authors of tweeting articles. As discovered from the earlier analysis of occupation of tweet users, in table 3 we can see that medical and health science articles are mostly shared in Twitter. What is clear from the map is
Author: Anik Dutta

that it supports other results; it is clear how many of the words are connected to health and medicine. Apart from the most frequently used words, we can see some clear cluster of most co-mentioned words in the upper part of the network is a cluster relating to professors, associates and assistants. While some of the clusters are related to health. In Figure 4 the words management, neuroscience describing personal information and interests are more outliers.
6.0 Discussion

In this chapter, the findings of the study are discussed and the research questions are answered.

6.1 Discussion of the research question

Q.1. Who is tweeting about research publications?

This research analyzed people who are tweeting about a scientific publication that had at least one Finnish author affiliation. The study set out to discover the professional background of the tweet user who shared articles on Twitter. The study of Semertzidis et al., (2013), focused on the profile description, occupation and hobbies of a random sample of tweeters. Semertzidis et al., (2013), found that Twitter bios describes some interest in instance “music”, “fan”, “food” of the user. Moreover, Semertzidis et al., (2013), found that several users stated the occupation, for example, “marketing”, “author”, “artist”, “designer on their profile description. In this study, the tweeters professional background and gender profile were analyzed. This study will help researchers understand which professionals are sharing scientific articles on Twitter.

It was observed that 40.3% of tweeters were male, while 14.5% are female. However, 33.3% of the tweets were posted by organizations and gender was not assigned to them. The data was gathered from VIRTA database and Altmetrics.com, so it was not feasible to dig deep into this matter due to data limitations. This study gives us the gender ratio from the data but not the reasons for it. In the future study researchers can look into the possible reasons for this.

At present people use social media to share information, create content and help to participate in social networking (Boyd and Ellison, 2010; Thelwall, 2009). Nowadays many scholars are using Twitter for communication or disseminating purpose (Priem et al., 2010). In this study, it was discovered that 23% of the sample are working in academia and using Twitter to share a scientific article. Moreover, 13.25% of the sample of this study are medical health professionals. Furthermore, a co-word map of the most frequently mentioned words in Twitter profiles provided evidence for the former statement. Words like health, research, Twitter, professor and science stood out as largest nodes in the maps and thus being the most frequent word in the profile descriptions of the tweeters. This study showed us that organizations (Medical Institute, universities, Journal publishing company etc.) also shared scientific research articles on Twitter. 33.3% shared scientific articles on Twitter are shared by organizations.
From this study we can see that scientific articles, with at least one Finnish author, are frequently shared by men and organizations, and the individuals are often working in academia. Thus, tweets may not reflect the societal impact of research rather new forms of scholarly communication.

Q2. What kind/types of articles professionals are shared most on Twitter?

Another result of the study is to find out whether open science articles are shared more than articles behind paywalls on Twitter. The results showed us that 92.5% of the articles shared on Twitter are openly accessible and 7.5% articles shared on Twitter are behind paywall. It is beneficial for the followers to find an open-access article if they do not have the access rights of subscription-based articles.

This study also investigated which kinds of articles such as agricultural, biological science and engineering articles, medical and health science articles, natural science articles or social and humanities articles are being shared on Twitter. Around 56% of the articles which are shared on Twitter are from Medical and health science. Out of this 56%, 50.25% of the articles are openly accessible.

Vainio and Holmberg (2017, p.348) state that “the distribution of the altmetric event across different research articles is highly skewed”. It means a few types of articles receive greater attention while many articles do not receive at all attention. In this study, we can see that after medical and health science articles, Natural science articles (94 times) get more attention on Twitter whereas social science and Humanities articles get less attention on Twitter. According to the study of Vainio and Holmberg (2017), 7 out of 20 articles belonged to natural sciences. In addition to that, Ke (2016) reflected that scholars in Twitter preferred natural science or health articles and journals. According to our findings, social science articles were tweeted the least among all other professionals 18times (Table 7).

From this study, we can see that open access articles and MHS articles were frequently shared on Twitter. It was discovered from this study that some of the tweet authors mention scientific articles about diabetics, breast cancer, cancer and obesity. One possibility of sharing such MHS articles on Twitter is to create self-awareness among followers. Interest on MHS articles by different professionals can be another reason. We know that around 14.50% (table 4) of the
sample are medical health or medical student by profession and thus have an interest in medical and health science articles.

Q3. Do highly tweeted articles also receive citations?

Ziman (1968) and Smith (1981), state that a scientific paper does not stand alone; it needs citations. The citation is an essential part of scholarly communication. Citation analysis is known as a tool for journal evaluation. Highly cited articles tend to have a higher impact, and the author of that journal tends to be highly recognized.

This study tells us about the number of times the tweeted articles were cited. Around 82.5% of the articles shared on Twitter have been cited for 0 to 20 times (Table: 8). Some of the articles are more popular than others and they were mentioned for 201 to 500 times (Table:8). From this study, we can see that one article has a very high impact compared to other articles and this article was cited for 1001 to 2000 times.

This study calculates the correlation between highly tweeted articles and the number of times the articles were cited. The correlation among variables is significant at the 0.05 level (Table: 9). The correlation value of this study is -.124 which is not highly significant. This value refers to the fact that highly tweeted articles were not necessarily highly cited. From table 10 we can see that “Nurse staffing and education and hospital mortality in nine European countries: a retrospective observational study” is one of the highly tweeted articles of this study. This MHS article was tweeted 14times by different tweet authors and this article was cited only one time. “Duodenal Infusion of Donor Feces for Recurrent Clostridium difficile” is another MHS article which was tweeted ten times and cited one times. It means these highly tweeted articles were not solely shared for citation purpose.

Scientific articles could be used to promote ideological views. In news media, the success of the news is measured by the content of the story. Social media can be used for spreading valuable news (Ettama, 2009). An earlier study of Vainio and Holmberg (2017) discovered that in social media valuable news could be applied to disseminate academic article. The article “Nurse staffing and education and hospital mortality in nine European countries: a retrospective observational study” investigated how nursing cutbacks directly death rate in different hospitals and culture. Despite being a scientific article, this article can meet several
of the before mentioned news values (Badenschier and Holger 2012). This result is in line with the earlier research of Vainio and Holmberg (2017). Other possible reasons for highly tweeting articles is to raise social and self-awareness autism, mental health and animal health etc. among followers.

This study investigated professionals who shared scientific articles with at least one Finnish affiliation on Twitter. It was also investigated what types of articles are shared most on Twitter. This study also discovered which professional are sharing what types of articles. From this study we can see that scientific articles with at least one Finnish author are often shared by men and organizations and the individuals are often working in academia. Moreover, this study discovered that the highly tweeted scientific articles were not highly cited articles. However, this study has some limitations that needs to be acknowledged. This study is based on quantitative data. Therefore, it was not possible to dig deep to solve some questions. For example:

1. It was not possible to answer why people shared articles with at least one Finnish author on Twitter?
2. Why scientific articles were more often shared by men?
3. Why medical and health science articles were shared more on Twitter?
4. It was not possible to answer why social science and humanities articles were shared less on Twitter?

The arrival of social media has changed many aspects of scholarly communication. Now scholars can present their thoughts in blogs, discuss and share their research with others on different social media platform for instance, Facebook and Twitter. Different professionals are now using informal scholarly communication, for instance, social media for disseminating research. From this study, we can see that different professionals such as academic professionals, medical health care professionals, organizations and bloggers are using Twitter for sharing scientific articles with at least one Finnish author affiliation. Moreover, this study discovered that highly tweeted articles are not highly cited articles. This refers to the fact that professionals are using informal scholarly communication medium such as Twitter to create awareness and to share ideas and knowledge.

Author: Anik Dutta
7.0 Future work

The results represented in this thesis have raised several interesting questions that require further investigation. One of the interesting topics would be to find out the reasons behind sharing scientific articles in Twitter. Another interesting question could be to investigate what are the feedback of the followers about that post? Why scientific articles were more often shared by men? Another interesting question could be why articles of certain disciplines share more on Twitter?

Another theoretical point is from an Information and Knowledge Management perspective, that is, when a professional’s share scientific articles on Twitter should we call it knowledge sharing, or it should call information sharing. In one point of view, we can say that when someone is sharing an article it will be named as knowledge sharing because someone from the author’s follower can read that article and gain knowledge. On the other hand, someone calls it as knowledge is created when someone receives an article on Twitter reads it and understand it. Information is shared but knowledge is created when information received and understood. Somebody can believe that only information can be shared knowledge can not be shared. As because it is not clear which is more correct there may be an opportunity to work further on it.

The findings of this thesis provide insights into articles with authors with Finnish affiliations and that one shared on Twitter. Future work could investigate this in other countries too, to see the in these findings are similar or not.
8.0 Conclusions

Overall, this study helps us to understand what kind of articles are shared in Twitter and who shares them. This study found out that men and organizations and the individuals are often working in academia share more scientific articles with at least one Finnish author on Twitter. It was also discovered that medical and health related journals were shared more often on Twitter. From this we can see that highly tweeted articles were not highly cited articles. Furthermore, it would also be efficient for the publishers to understand what kind of professionals usually share articles on Twitter. However, our research does not stop here. The author believes that future study is needed to answer questions such as why different professionals shared articles on Twitter? What is feedback of the followers of the tweet author?
9.0 References


Author: Anik Dutta


Author: Anik Dutta


Author: Anik Dutta


Author: Anik Dutta


Author: Anik Dutta


Author: Anik Dutta


Author: Anik Dutta


alysis_of_Mendeley_User_Categories


Author: Anik Dutta


Author: Anik Dutta


http://www.dlib.org/dlib/february00/vandesompel-oai/02vandesompeloai.html


Author: Anik Dutta


Author: Anik Dutta


Author: Anik Dutta


### Appendix 1: Overall summary of the quantitative research:

<table>
<thead>
<tr>
<th>Gender * Type of Articles * Occupation</th>
<th>Types of Articles</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Agriculture</td>
<td>Biological</td>
</tr>
<tr>
<td></td>
<td>Sc.</td>
<td>Engineering</td>
</tr>
<tr>
<td>Professors Associate, Assistant etc.</td>
<td>Gender</td>
<td>Male</td>
</tr>
<tr>
<td></td>
<td>% within Gender</td>
<td>20,6%</td>
</tr>
<tr>
<td></td>
<td>Female</td>
<td>Count</td>
</tr>
<tr>
<td></td>
<td>% within Gender</td>
<td>40,0%</td>
</tr>
<tr>
<td></td>
<td>Total</td>
<td>Count</td>
</tr>
<tr>
<td></td>
<td>% within Gender</td>
<td>23,1%</td>
</tr>
<tr>
<td>University research teaching staff</td>
<td>Gender</td>
<td>Male</td>
</tr>
<tr>
<td></td>
<td>% within Gender</td>
<td>60,0%</td>
</tr>
<tr>
<td></td>
<td>Female</td>
<td>Count</td>
</tr>
<tr>
<td></td>
<td>% within Gender</td>
<td>50,0%</td>
</tr>
<tr>
<td></td>
<td>Total</td>
<td>Count</td>
</tr>
<tr>
<td></td>
<td>% within Gender</td>
<td>57,1%</td>
</tr>
<tr>
<td>PhD student/candidate</td>
<td>Gender</td>
<td>Male</td>
</tr>
<tr>
<td></td>
<td>% within Gender</td>
<td>50,0%</td>
</tr>
<tr>
<td></td>
<td>Total</td>
<td>Count</td>
</tr>
<tr>
<td></td>
<td>% within Gender</td>
<td>50,0%</td>
</tr>
<tr>
<td>Bloggers</td>
<td>Gender</td>
<td>Male</td>
</tr>
<tr>
<td></td>
<td>% within Gender</td>
<td>11,1%</td>
</tr>
<tr>
<td></td>
<td>Female</td>
<td>Count</td>
</tr>
<tr>
<td></td>
<td>% within Gender</td>
<td>0,0%</td>
</tr>
<tr>
<td></td>
<td>Unknown</td>
<td>Count</td>
</tr>
<tr>
<td></td>
<td>% within Gender</td>
<td>37,5%</td>
</tr>
<tr>
<td></td>
<td>Not applicable</td>
<td>Count</td>
</tr>
<tr>
<td></td>
<td>% within Gender</td>
<td>12,5%</td>
</tr>
<tr>
<td></td>
<td>Total</td>
<td>Count</td>
</tr>
<tr>
<td>Author: Anik Dutta</td>
<td></td>
<td></td>
</tr>
<tr>
<td>-------------------</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Gender</th>
<th>Count</th>
<th>% within Gender</th>
<th>Male</th>
<th>Female</th>
<th>Not applicable</th>
<th>Total</th>
<th>% within Gender</th>
</tr>
</thead>
<tbody>
<tr>
<td>CEO Entrepreneur</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Male</td>
<td>1</td>
<td>15,4%</td>
<td>1</td>
<td>33,3%</td>
<td></td>
<td>3</td>
<td>33,3%</td>
</tr>
<tr>
<td>% within Gender</td>
<td></td>
<td></td>
<td>1</td>
<td>33,3%</td>
<td></td>
<td>3</td>
<td>33,3%</td>
</tr>
<tr>
<td>Total</td>
<td>1</td>
<td>15,4%</td>
<td>1</td>
<td>33,3%</td>
<td></td>
<td>3</td>
<td>33,3%</td>
</tr>
<tr>
<td>Female</td>
<td>1</td>
<td>25,0%</td>
<td>2</td>
<td>50,0%</td>
<td></td>
<td>4</td>
<td>50,0%</td>
</tr>
<tr>
<td>% within Gender</td>
<td></td>
<td></td>
<td>1</td>
<td>50,0%</td>
<td></td>
<td>2</td>
<td>50,0%</td>
</tr>
<tr>
<td>Total</td>
<td>2</td>
<td>25,0%</td>
<td>1</td>
<td>50,0%</td>
<td></td>
<td>3</td>
<td>50,0%</td>
</tr>
<tr>
<td>Researcher</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Male</td>
<td>6</td>
<td>28,6%</td>
<td>5</td>
<td>23,8%</td>
<td></td>
<td>11</td>
<td>26,9%</td>
</tr>
<tr>
<td>% within Gender</td>
<td></td>
<td></td>
<td>1</td>
<td>38,1%</td>
<td></td>
<td>2</td>
<td>34,6%</td>
</tr>
<tr>
<td>Female</td>
<td>1</td>
<td>25,0%</td>
<td>2</td>
<td>50,0%</td>
<td></td>
<td>4</td>
<td>50,0%</td>
</tr>
<tr>
<td>% within Gender</td>
<td></td>
<td></td>
<td>1</td>
<td>25,0%</td>
<td></td>
<td>2</td>
<td>25,0%</td>
</tr>
<tr>
<td>Not applicable</td>
<td>0</td>
<td>0,0%</td>
<td>0</td>
<td>0,0%</td>
<td></td>
<td>0</td>
<td>0,0%</td>
</tr>
<tr>
<td>Total</td>
<td>7</td>
<td>28,6%</td>
<td>8</td>
<td>50,0%</td>
<td></td>
<td>15</td>
<td>30,8%</td>
</tr>
<tr>
<td>Health care</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Male</td>
<td>2</td>
<td>5,4%</td>
<td>31</td>
<td>83,8%</td>
<td></td>
<td>33</td>
<td>84,9%</td>
</tr>
<tr>
<td>% within Gender</td>
<td></td>
<td></td>
<td>4</td>
<td>10,8%</td>
<td></td>
<td>8</td>
<td>12,5%</td>
</tr>
<tr>
<td>Female</td>
<td>0</td>
<td>0,0%</td>
<td>14</td>
<td>87,5%</td>
<td></td>
<td>16</td>
<td>11,3%</td>
</tr>
<tr>
<td>% within Gender</td>
<td></td>
<td></td>
<td>2</td>
<td>12,5%</td>
<td></td>
<td>1</td>
<td>10,0%</td>
</tr>
<tr>
<td>Total</td>
<td>2</td>
<td>5,4%</td>
<td>45</td>
<td>87,5%</td>
<td></td>
<td>47</td>
<td>96,2%</td>
</tr>
<tr>
<td>Journal Editor</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Male</td>
<td>0</td>
<td>0,0%</td>
<td>2</td>
<td>66,7%</td>
<td></td>
<td>2</td>
<td>66,7%</td>
</tr>
<tr>
<td>% within Gender</td>
<td></td>
<td></td>
<td>1</td>
<td>33,3%</td>
<td></td>
<td>1</td>
<td>33,3%</td>
</tr>
<tr>
<td>Female</td>
<td>1</td>
<td>16,7%</td>
<td>5</td>
<td>83,3%</td>
<td></td>
<td>6</td>
<td>83,3%</td>
</tr>
<tr>
<td>% within Gender</td>
<td></td>
<td></td>
<td>0</td>
<td>0,0%</td>
<td></td>
<td>0</td>
<td>0,0%</td>
</tr>
<tr>
<td>Not applicable</td>
<td>4</td>
<td>16,7%</td>
<td>11</td>
<td>83,3%</td>
<td></td>
<td>15</td>
<td>83,3%</td>
</tr>
<tr>
<td>Total</td>
<td>5</td>
<td>16,7%</td>
<td>18</td>
<td>55,0%</td>
<td></td>
<td>23</td>
<td>55,0%</td>
</tr>
<tr>
<td>Columnist/author</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Male</td>
<td>3</td>
<td>23,1%</td>
<td>6</td>
<td>46,2%</td>
<td></td>
<td>9</td>
<td>46,2%</td>
</tr>
<tr>
<td>% within Gender</td>
<td></td>
<td></td>
<td>4</td>
<td>30,8%</td>
<td></td>
<td>4</td>
<td>30,8%</td>
</tr>
</tbody>
</table>

| Total            | 5     | 16,7%           | 18   | 55,0%  |               | 23    | 55,0%           |

68
<table>
<thead>
<tr>
<th>Role</th>
<th>Gender</th>
<th>Count</th>
<th>Female</th>
<th>Male</th>
<th>Total</th>
<th>% within Gender</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Female %</td>
</tr>
<tr>
<td>Author</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Gender</td>
</tr>
<tr>
<td>Organization</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Gender</td>
</tr>
<tr>
<td>Scientists</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Gender</td>
</tr>
<tr>
<td>Student/Medical</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Gender</td>
</tr>
<tr>
<td>Photographer</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Gender</td>
</tr>
<tr>
<td>Unknown</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Gender</td>
</tr>
<tr>
<td>Not applicable</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Gender</td>
</tr>
<tr>
<td>Gender</td>
<td>Male</td>
<td>Count</td>
<td>7</td>
<td>26</td>
<td>21</td>
<td>2</td>
</tr>
<tr>
<td>--------</td>
<td>--------</td>
<td>-------</td>
<td>----</td>
<td>----</td>
<td>----</td>
<td>----</td>
</tr>
<tr>
<td></td>
<td>% within Gender</td>
<td>12,5%</td>
<td>46,4%</td>
<td>37,5%</td>
<td>3,6%</td>
<td>100,0%</td>
</tr>
<tr>
<td>Female</td>
<td>Count</td>
<td>3</td>
<td>6</td>
<td>2</td>
<td>1</td>
<td>12</td>
</tr>
<tr>
<td></td>
<td>% within Gender</td>
<td>25,0%</td>
<td>50,0%</td>
<td>16,7%</td>
<td>8,3%</td>
<td>100,0%</td>
</tr>
<tr>
<td>Female</td>
<td>Count</td>
<td>1</td>
<td>4</td>
<td>1</td>
<td>0</td>
<td>6</td>
</tr>
<tr>
<td></td>
<td>% within Gender</td>
<td>16,7%</td>
<td>66,7%</td>
<td>16,7%</td>
<td>0,0%</td>
<td>100,0%</td>
</tr>
<tr>
<td>Unknown</td>
<td>Count</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>% within Gender</td>
<td>100,0%</td>
<td>0,0%</td>
<td>0,0%</td>
<td>0,0%</td>
<td>100,0%</td>
</tr>
<tr>
<td>Total</td>
<td>Count</td>
<td>5</td>
<td>10</td>
<td>3</td>
<td>1</td>
<td>19</td>
</tr>
<tr>
<td></td>
<td>% within Gender</td>
<td>26,3%</td>
<td>52,6%</td>
<td>15,8%</td>
<td>5,3%</td>
<td>100,0%</td>
</tr>
<tr>
<td>Total</td>
<td>Male</td>
<td>Count</td>
<td>28</td>
<td>84</td>
<td>39</td>
<td>10</td>
</tr>
<tr>
<td></td>
<td>% within Gender</td>
<td>17,4%</td>
<td>52,2%</td>
<td>24,2%</td>
<td>6,2%</td>
<td>100,0%</td>
</tr>
<tr>
<td>Female</td>
<td>Count</td>
<td>7</td>
<td>36</td>
<td>14</td>
<td>1</td>
<td>58</td>
</tr>
<tr>
<td></td>
<td>% within Gender</td>
<td>12,1%</td>
<td>62,1%</td>
<td>24,1%</td>
<td>1,7%</td>
<td>100,0%</td>
</tr>
<tr>
<td>Unknown</td>
<td>Count</td>
<td>7</td>
<td>21</td>
<td>19</td>
<td>1</td>
<td>48</td>
</tr>
<tr>
<td></td>
<td>% within Gender</td>
<td>14,6%</td>
<td>43,8%</td>
<td>39,6%</td>
<td>2,1%</td>
<td>100,0%</td>
</tr>
<tr>
<td>Not applicable</td>
<td>Count</td>
<td>22</td>
<td>83</td>
<td>22</td>
<td>6</td>
<td>133</td>
</tr>
<tr>
<td></td>
<td>% within Gender</td>
<td>16,5%</td>
<td>62,4%</td>
<td>16,5%</td>
<td>4,5%</td>
<td>100,0%</td>
</tr>
<tr>
<td>Total</td>
<td>Count</td>
<td>64</td>
<td>224</td>
<td>94</td>
<td>18</td>
<td>400</td>
</tr>
<tr>
<td></td>
<td>% within Gender</td>
<td>16,0%</td>
<td>56,0%</td>
<td>23,5%</td>
<td>4,5%</td>
<td>100,0%</td>
</tr>
</tbody>
</table>