Highly tweeted science articles – who tweets them? An analysis of Twitter user profile descriptions

Vainio, Julia [1]

Holmberg, Kim [1]

Abstract In this study we examined who tweeted academic articles that had at least one Finnish author or co-author affiliation and that had high altmetric counts on Twitter. In this investigation of national level altmetrics we chose the most tweeted scientific articles from four broad areas of science (Agricultural, Engineering and Technological Sciences; Medical and Health Sciences; Natural Sciences; Social Sciences and Humanities). By utilizing both quantitative and qualitative methods of analysis, we studied the data using research techniques such as keyword categorization, co-word analysis and content analysis of user profile descriptions. Our results show that contrary to a random sample of Twitter users, users who tweet academic articles describe themselves more factually and by emphasizing their occupational expertise rather than personal interests. The more field-specific the articles were, the more research-related descriptions dominated in Twitter profile descriptions. We also found that scientific articles were tweeted to promote ideological views especially in instances where the article represented a topic that divides general opinion.

Keywords Twitter ; Twitter profile ; Altmetrics ; Scholarly Communication

Vainio, Julia
e-mail: jukava@live.com

Holmberg, Kim
e-mail: kim.j.holmberg@utu.fi

[1] Research Unit for the Sociology of Education, University of Turku, Finland

Introduction

Social media platforms such as Twitter have become important means of communication and dissemination of information. Due to the easy access and low cost for users, they have become spaces that are used not only for pastime and to maintain personal connections but also for systematic purposes. Twitter has been used, for example, for political election campaigns (Vergeer, Hermans & Sams 2011), climate activism (Segerberg & Bennett 2011) and for scientific conferences (Desai, Shariff, Shariff, Kats, Fang, Christiano & Ferris 2012). Related to the latter, there is a growing incentive to study scholarly communications and dissemination of scientific articles online, in order to understand whether the attention certain research articles and other outputs receive online could tell something about the value or popularity of that research.

In the past, science-society communication strategies have been in most cases very top-down led. The public was seen as ignorant and needing to be educated by the professionals. (ICSU 2005, 20.) However, in recent years, this authoritative view has given way for a more flexible process of communications. Scientists are seen as having both the right and the responsibility to contribute to society and to the public. Methods for this include inter alia a two-way dialogue and an aim to incorporate researchers into public discussions of science. (Bird 2014; ICSU 2005.) This change in communications is also visible in requirements for funding in some cases. As Bird (2014) has summarized, an organization such as The National Science Foundation in the US has a criterion of “broader impacts” for the merit evaluation and funding of grant proposals whereas the National Institutes of Health in the US includes a criteria of “significance” in their evaluations of research proposals.

These changes in funding requirements signal a growing importance on considering the societal impact of science as a valid ground for research funding. This, in turn, has led to a growing interest and demand for altmetric analysis of scientific articles during the last decade. Altmetrics (short for alternative metrics) provide an alternative or a complement perspective to traditional measures on scholarly impact such as the number of citations peer-reviewed scientific articles attract (Brigham 2014; Holmberg 2014). As Holmberg (2014) notes, many of the traditional scholarly impact metrics were created for libraries for purposes of collection development and for researchers to track the dissemination and use of their work, not as methods for research assessment or to assess which researchers or research groups are the most deserving of competitive research funding. Altmetrics may be able to provide means of understanding the impact research has on the mainstream, as it focuses on measuring scientific impact based on online tools and activities by the public beyond academia (Brigham 2014; Priem, Groth, & Taraborelli 2012). Social media platforms such as Twitter provide with relative ease both a possibility for scholarly communication with the public and between researchers as well as the dissemination of research and specifically research articles to a wider audience.

It has been argued that quantifiable research such as correlation analyses do not actually reveal the meaning of altmetrics and that research should instead focus more on qualitative content analysis of social media. (Bornmann 2016.) This would mean a more descriptive approach on altmetrics where instead of focusing on the statistics of the online mentions of research articles we should focus on the quality and other characteristics of the mentions as well as on people who mention these research articles. As Bornmann (2016) suggests, this sort of research is vacant in altmetrics, save for a few studies (see eg. Thelwall, Tsou, Weingart, Holmberg, & Haustein 2013).

In this work we analyze how users who share or mention scientific research articles on Twitter describe themselves on their profiles. We have chosen a set of scientific articles from four broad areas of science that have been frequently shared and authored or co-authored by at least one researcher with a Finnish affiliation, thus providing also a view on the applicability of altmetrics at a national level. The user profiles of the people mentioning these articles were analyzed utilizing research techniques such as keyword categorization, co-word analysis, and content analysis of user profile descriptions. By using a combination of both quantitative and qualitative content analysis techniques we are able to triangulate the results and to form a contextualized view of the data and find common patterns between users. This will give us further information over the impact and distribution of scientific articles in general, as well as contribute to increased understanding of the methods and ways in which the participants in online scholarly communication can be examined.

Previous research

Twitter is more commonly used in broadcasting information rather than as a platform for more personal interaction among individuals (Neiger, Thackeray, Burton, Giraud-Carrier, & Fagen 2013). A typical tweet about a scientific article appears to be quite factual in its nature with little or no opinions expressed (Thelwall et al. 2013). One reason for this is the restrictive 140 character limit for each tweet (although Twitter appears to be increasing the limit). Even though general research on Twitter has been plentiful, the research focusing on understanding the types of Twitter users and their behavior has been scarcer. There have been, however, a few studies that have focused on what Twitter users expose on their profile (see eg. Mislove, Lehmann, Ahn, Onnela & Rosenquist, 2011; Semertzidis, Pitoura & Tsaparas 2013; Uddin, Imran & Sajjad 2014). These studies have mainly dealt with random samples of Twitter messages and Twitter users collected using the Twitter APIs (Semertzidis et al. 2013; Uddin et al. 2014). In some of the related work, machine learning technique was used in order to classify Twitter users into different classes (Uddin et al. 2014), while another study focused on the demographics of Twitter users discovering that in the US, Twitter users are predominantly male and that users are significantly over-represented in densely populated regions (Mislove et al. 2011).

Academic activity on Twitter has been researched through examining how scholars and academics tweet (see e.g. Priem & Costello 2010; Holmberg & Thelwall 2014, Hadgu & Jäschke 2014, Haustein, Bowman, Holmberg, Peters & Larivière 2014). In these studies the studied disciplines and researchers were usually selected before-hand. These studies have limitations, such as small and biased sample-sizes, where the more well-known or popular scientists may be overrepresented (Ke, Ahn & Sugimoto 2016). Holmberg and Thelwall (2014) studied researchers selected based on their productivity from ten different disciplines. They found variations in scholarly communications on Twitter between different disciplines; fields such as biochemistry, astrophysics and cheminformatics used Twitter for scholarly communications frequently, while sociologists, on the other hand, seemed to neglect it. In Haustein et al. (2014), the focus was on the behavior of 37 astrophysicists. Their findings suggested a low negative correlation between active tweeting and publication frequency, showing that, at least in the case of astrophysicists, the researchers who published more did not tweet frequently and vice versa. Previous studies have shown that even though scientists on Twitter tend to tweet mostly heterogeneously about a range of issues, they also tend to favor certain scholarly sites, such as generalist publications (i.e. *Nature* and *Science*), more than non-scholarly tweeters. It seems that even though the content that scientists and researchers tweet is highly heterogeneous, the tweeters of scholarly articles are more likely to be experts on their fields of study. (Ke et al., 2016.)

Scholars are not the only ones tweeting articles published in generalist publications. As one study found (Thelwall et al. 2013), nearly three quarters of the total amount of tweets mentioning articles from *Science* and almost 90% of the total amount of tweets mentioning articles from *Nature* contained the exact titles of the articles in question. This led researchers to suggest that *Science* and *Nature* organize or support tweeting of their articles despite the fact that these tweets originate from different people. Some of the tweets were also automatically tweeted by journal publishers, supporting the presumption that publishers follow a systematic tweeting strategy. Tweets concerning research articles are not only shared in communication about the actual research, but also for marketing (or even spam) purposes by the publishers and authors themselves (Nelhans & Gunnarsson Lorentzen 2016). This suggests that academic research might be shared on Twitter not just for the universal sake of disseminating research but also in order to profit from the research or from the attention it receives in some way. This notion indicates the need for further examination of Twitter users who tweet scientific articles.

Data and Methods

Data

Metadata of scientific publications by authors with Finnish affiliation were retrieved from the national VIRTA research publication database. Focusing on publications by authors with Finnish affiliation gives us the possibility to examine the applicability of altmetrics at a national level. The data were categorized according to OECD main categories of which we merged *Agricultural, Engineering and Technological Sciences* into one new category (AETS) and Social Sciences and Humanities into another new category (SSH). *Medical and Health Sciences* (MHS) and *Natural Sciences* (NS) were left intact. Using the DOIs of the articles the altmetric events to the publications were matched from data provided by Altmetric.com. From each area of science the articles were ranked based on how many times they had been tweeted and the five articles with the most Twitter mentions were chosen. To investigate who is tweeting about science the profile information of tweeters that had tweeted about any of the selected articles were downloaded and analyzed. There are considerable differences between the total numbers of contributing authors from one paper to another. As the prerequisites stated, for the paper to be involved in this study, there were to be at least one author with a Finnish affiliation, the paper was to be published between 2012 and 2014 and it had to qualify as being in the top 5 most tweeted article in its category at the time of data retrieval. These prerequisites lead to a situation where the internal variation between the articles in each category is quite considerable with one article having up to 2900 authors and another having just two authors.

Getting reliable demographic information on Twitter users is difficult due to the fact that users are not required to provide accurate information about themselves. However, we were able to deduct some information on the geographical distribution of the tweeters. The total amount of tweets about the top 5 articles from each category amounted up to 9493 tweets. Of these, 63.5 percent were by users who had a distinguished country code attached to their Twitter account. A majority, 26.0 percent, of these tweets originated from the US while the UK came as the second largest tweeting base with

21.2 percent of all tweets. Canada, Spain and Japan each comprised between 5.3 and 6.0 percent of the tweets. Besides these states, there were no other states clearly deviating from the rest. In total, we were able to detect 113 different nationalities from the data.

The distribution of altmetric events across research articles is highly skewed (see eg. Eysenbach 2011; Liu, Xu, Wu, Chen, & Guo 2013), with only a few articles receiving the majority of the attention, while the majority of articles receive only little or no attention at all. By focusing on the scientific articles that have received the most attention, we are also focusing on the cases that have had most impact, assuming that the level of attention is equivalent to the level of impact. As Thelwall et al. (2013) illustrated, altmetric count correlations with citations might only be useful when used to measure the altmetrics of an above the average article rather than by analyzing articles which have very little or any altmetric counts. The distortion between the popular few and the unnoticed many among research articles is not only an issue in altmetric counts. Studies show (see e.g. Larivière & Gingras 2009) how non-citation rates vary considerably among different fields from around 12 percent of medicine articles up to 82 percent of humanities articles not being cited. However, the accumulation of citation counts differs considerably from altmetric counts as the majority of altmetric counts usually occur immediately or in the becoming weeks or days after the article is published (Holmberg 2014).

It is notable that seven articles out of twenty in our data were either published in *Nature* or *Science*, or in a *Nature*-related journal, as it is shown in Table 1. This notion supports Ke et al.'s (2016) results in which scientists on Twitter seemed to favor certain scholarly sites, such as *Nature* or *Science*.

Table 1 *top five articles in each of the categories: Agricultural, Engineering and Technology (AETS), Medical and Health Sciences (MHS), Natural Sciences (NS), and Social Sciences and Humanities (SSH)*

Title	DOI	Year	Journal	Category	Twitter posts
Bodily maps of emotions	10.1073/pnas.1321664111	2014	PNAS	AETS	1218
GWAS of 126,559 Individuals Identifies Genetic Variants Associated with Educational Attainment	10.1126/science.1235488	2013	Science	AETS	420
Biological insights from 108 schizophrenia-associated genetic loci	10.1038/nature13595	2014	Nature	AETS	359
Synaptic, transcriptional and chromatin genes disrupted in autism	10.1038/nature13772	2014	Nature	AETS	331
Approaching a state shift in Earth's biosphere	10.1038/nature11018	2012	Nature	AETS	308
Nurse staffing and education and hospital mortality in nine European countries: a retrospective observational study	10.1016/S0140-6736(13)62631-8	2014	Lancet	MHS	2114
Duodenal Infusion of Donor Feces for Recurrent Clostridium difficile	10.1056/NEJMoa1205037	2013	New England Journal Of Medicine	MHS	1206
Arthroscopic Partial Meniscectomy versus Sham Surgery for a Degenerative Meniscal Tear	10.1056/NEJMoa1305189	2013	New England Journal Of Medicine	MHS	904
Cultural Bias in the AAP's 2012 Technical Report and Policy Statement on Male Circumcision	10.1542/peds.2012-2896	2013	Pediatrics	MHS	431

Cancer survival in Europe 1999-2007 by country and age - results of EUROCARE--5-a population-based study	10.1016/S1470-2045(13)70546-1	2014	Lancet Oncology	MHS	311
Observation of a new boson at a mass of 125 GeV with the CMS experiment at the LHC	10.1016/j.physletb.2012.08.021	2012	Physics Letters B	NS	431
Defining the role of common variation in the genomic and biological architecture of adult human height	10.1038/ng.3097	2014	Nature Genetics	NS	294
Continental-scale temperature variability during the past two millennia	10.1038/NGEO1797	2013	Nature Geoscience	NS	234
Ca-48+Bk-249 Fusion Reaction Leading to Element Z=117: Long-Lived alpha-Decaying (270)Db and Discovery of Lr-267	10.1103/PhysRevLett.112.172501	2014	Physical Review Letters	NS	198
Meta-analysis of 74,046 individuals identifies 11 new susceptibility loci for Alzheimer's disease	10.1038/ng.2802	2013	Nature Genetics	NS	179
Atheists Become Emotionally Aroused When Daring God to Do Terrible Things	10.1080/10508619.2013.771991	2014	International Journal For The Psychology Of Religion	SSH	180
Music reduces pain and increases functional mobility in fibromyalgia	10.3389/fpsyg.2014.00090	2014	Frontiers In Psychology	SSH	120
When does evidence-based policy turn into policy-based evidence? Configurations, contexts and mechanisms	10.1332/174426514X13990433991320	2014	Evidence & Policy	SSH	96
Happiness: Before and After the Kids	10.1007/s13524-014-0321-x	2014	Demography	SSH	82
The individualization of class: a case of working life coaching	10.1111/1467-954X.12209	2014	Sociological Review	SSH	76

Of Agricultural, Engineering and Technology Sciences category, the article *Bodily maps of emotions* attracted the most Twitter posts (1218 posts on Twitter, see Table 1). In the article researchers show how the most common emotions such as embarrassment or joy trigger strong bodily sensations and how these sensations are not felt culturally but biologically. Whereas *GWAS of 126,559 Individuals Identifies Genetic Variants Associated with Educational Attainment* (420 posts) was a GWAS (Genome-Wide Association Studies) study which paved way for research regarding the study of genetic variants that could be associated with educational attainment, *Biological insights from 108 schizophrenia-associated genetic loci* (359 posts) used GWAS to identify large numbers of risk loci in molecular genetic study of schizophrenia. Contributing to the genetic research on autism, *Synaptic, transcriptional and chromatin genes disrupted in autism* (331 posts), helped the science community to identify different relationships between the categories of genes associated with autism. The study that attracted the fifth most posts in the category, *Approaching a state shift in Earth's biosphere* (308 posts) places a bleak outlook on the global conditions in biosphere and biological resources due to the actions of humans.

Regarding Medical and Health Sciences' articles, *Nurse staffing and education and hospital mortality in nine European countries: a retrospective observational study* had the most Twitter activity (2114 posts). The study in question found a direct link between nursing cutbacks and higher patient death rates in hospitals. The second most tweeted article, *Duodenal Infusion of Donor Feces for Recurrent Clostridium difficile* (1206 posts), showed how the installation of healthy donor's stool into a patient has treated recurrent Clostridium difficile infection with high success rates. *Arthroscopic Partial*

Meniscectomy versus Sham Surgery for a Degenerative Meniscal Tear (904 posts) showed that patients with persistent knee pain without knee arthritis or a torn medial meniscus without recent trauma did not in fact benefit from a knee surgery. The fourth most shared article in this category with 431 posts was *Cultural Bias in the AAP's 2012 Technical Report and Policy Statement on Male Circumcision*, which was a response to a report by the American Academy of Pediatrics whose arguments over the perceived health benefits of newborn male circumcision were questionable and had little public health relevance. The fifth article, the EUROCARE-5 report found that even though the number of adults surviving for at least five years after a cancer diagnosis has risen across Europe, cancer survival rate in the region still has large disparities between different countries (*Cancer survival in Europe 1999–2007 by country and age: results of EUROCARE-5-a population-based study*, 311 posts).

Observation of a new boson at a mass of 125 GeV with the CMS experiment at the LHC (431 posts) from the Natural Sciences' category concentrated on the first observations of a new particle relating to the Standard Model Higgs boson. With 294 posts, *Defining the role of common variation in the genomic and biological architecture of adult human height* was a GWAS study providing a more comprehensive look at the biology of height and how common genetic variant affect height. *Continental-scale temperature variability during the past two millennia* (234 posts) showed how, in the past, climate change varied considerably between different regions in the world. *Ca-48+Bk-249 Fusion Reaction Leading to Element Z=117: Long-Lived alpha-Decaying (270)Db and Discovery of Lr-267* (198 posts) demonstrated for the first time the production of a superheavy element with atomic number Z=117 that marks an important step towards the observation of "island of stability" in nuclear physics. The fifth most mentioned article in this category was *Meta-analysis of 74,046 individuals identifies 11 new susceptibility loci for Alzheimer's disease* (179 posts), another GWAS study that identified new genetics risks for Alzheimer's disease.

The most Twitter activity (180 posts) in Social Sciences and Humanities category was around the article *Atheists Become Emotionally Aroused When Daring God to Do Terrible Things*, which found that just as self-identified believers, self-identified atheists began to sweat and feel discomfort when reading aloud sentences asking God to do terrible things. The article attracting the second most posts found that music could perhaps be used as a treatment for chronic pain in fibromyalgia which could thus reduce the risk of disability (*Music reduces pain and increases functional mobility in fibromyalgia*, 120 posts). *When does evidence-based policy turn into policy-based evidence? Configurations, contexts and mechanisms* (96 posts) studied how authority relations and cultural contexts influence on validating and understanding expertise and evidence. The fourth most tweeted article in the category was *Happiness: Before and After the Kids* with 82 posts. The study found that having children at a later age is associated with higher satisfaction levels and that having a third child decreases parent's happiness levels. *The individualization of class: a case of working life coaching* (76 posts) examined the process of transforming the relations between capital and labor as forms of collective entities towards being parts of personal characteristics or as differences between individuals.

While preparing the profile descriptions for analysis, duplicate instances were removed to avoid possible over-emphasis of some very active Twitter users. Common stop-words, special characters and non-English words were then removed from the profile descriptions. It should be noted that by disregarding the non-English tweets some information about the tweeters were lost, as for example the article *Atheists Become Emotionally Aroused When Daring God to Do Terrible Things* gained significant attention outside the Anglo-Saxon world.

Methods

A specific aspect of the English grammar needs to be taken into account before analyzing word frequencies or conducting co-word analysis. As Carstairs-Mccarthy (2002, 16–21) points out, there is a specific area of grammar in English language that is focused on the structure of words and relationships between these words. As a normal word would have a free root in which affixes might attach themselves to give the word a new meaning (like *–ance* in *performance* or *en-* in *enlarge*), compounds are complex words that can be put together from two (or more) free roots. Telling the difference between two free roots (or a phrase, such as *black board* or *white house*) and a compound word (*blackboard, (the) White House*) can be difficult. Even though it is not completely possible to tell these two apart, there are some characteristics that help with identification. First, there is usually a visible difference in vocalization corresponding to the difference in meaning. Second, compound words tend to have a meaning that is more idiosyncratic or hypersensitive than phrases. It is also noteworthy that compound words can occur as compound verbs, compound adjectives or compound nouns. (Carstairs-Mccarthy 2002, 59-63.) To sum up, compound words consist of individual words that, when combined, form a unit that has a different meaning than either of the individual words that constitute it. Therefore, when counting word frequencies we need to keep in mind the possibility of existing compound words and what these compound words could mean in terms of interpretation. In co-occurrence analysis the compound words should appear as having stronger connections between the words, as they are more frequently mentioned together.

With the possibility of compound words taken into account, we will focus both on the qualitative content analysis to reveal unique themes surrounding the researched phenomenon and on the statistical significances of the occurrence of particular texts or concepts. In this research both approaches –qualitative and quantitative– were utilized as we used four approaches to investigate who is tweeting science; word frequencies, coding words, co-word analysis, and coding of the profile descriptions.

In the first step of the analysis the frequencies of the words used in the profile descriptions were calculated in order to understand the overall topics mentioned or descriptions used in the profiles. In the second step we built upon Semertzidis et al.'s (2013) article that studied people’s descriptions of themselves on Twitter and we constructed eight general themes for the coding, as seen in Table 2. The most frequently used words in the profiles were coded according to this categorization. In the analysis, we approached the words in a very literate sense that did not take a possible inductively produced context into account. Each word was analyzed as an individual word despite the possibility of it being a part of a compound word. The possibility of compound words was taken into consideration at the stage of co-word analysis.

Table 2 *categories used for coding the 100 most frequently mentioned words in users' Twitter profile descriptions*

Category	Description	Examples of words
Research related words	words that either describe research fields or are closely connected with doing research	*education, research, university, science, audiology, methods*

Academic occupations and Education levels	words that are used to describe the academic occupation or education level of the user	*academic, fellow, professor, researcher, student*
Occupation (not including academic occupations and education levels)	words that are used to describe the occupation of the user	*doctor, consultant, author, assistant, work, blogger*
Interests/Preferences/ Hobbies	words that are used to describe interests and past-time activities of the user	*art, books, geek, music, beer*
Personal info	words that are used to describe personal information	*care, father, love, uk, family*
Social networks/Internet	words that are used to describe activities online and on social media platforms	*account, blog, follow, tweet, tweeting*
Descriptive	words that, as such, do not reveal anything about the user of the profile (Semertzidis et al., 2013)	*advocate, free, good, senior, cultural*
Miscellaneous	words that do no fall into any of the other categories	*brain, health, chair, mental, policy, medicine, just, development*

In the third step co-word analysis was used. Co-word analysis is a content analysis technique that takes into account both the frequencies with which words appear in a given text and the frequency with which the words appear together, i.e. how the words are connected to each other (Courtial 1994). Based on the frequency and strength of these connections the words can then be drawn and clustered in a network map, from which even large quantities of data can, with relative ease, be interpreted more qualitatively. In the resulting network maps the size of a node represents the number of times a word has appeared in the profile descriptions in relation to other words. The thickness of the edges or the links between the nodes is determined by the strength of the connection between the two words. The thicker the line, the more the words have appeared together in different profile descriptions. The position of a node in relation to all the other nodes is determined by both the number and the frequency with which it has been used together with other words; more connections to other words means thus a more central position on the map. By analyzing the appearance of co-occurrences we are also able to determine the possible compound words that are otherwise missed in one word frequency analysis. The co-word maps were drawn with Gephi and the layout of the maps were calculated with Force Atlas, a built-in algorithm of Gephi. A community detection algorithm was used to identify tightly connected clusters among the words. These were then color-coded in the maps. (Blondel, Guillaume, Lambiotte, & Lefebvre 2008; Lambiotte, Delvenne, & Barahona 2008.)

In the fourth step the profile descriptions were coded manually using a qualitative content analysis technique. This technique utilizes a systematic classification process of coding and identifying emerging themes or patterns from the data and comparing them together in order to find significance (Hsieh & Shannon, 2005; Zhang & Wildemuth 2009). Qualitative content analysis uses primarily an inductive approach to the research at hand, creating categories inductively from the data and adjusting the research methods as the investigation proceeds (Zhang & Wildemuth 2009). Starting with categories for academic positions and position related to health care professions, the codebook was built inductively during the coding.

Full list of categories and their descriptions can be seen in Table 3. Several of the coded Twitter profile descriptions were ambiguous and could have thus seen to represent several themes (Zhang & Wildemuth 2009) found in our codebook. However, as Lincoln and Guba (1985 in Zhang & Wildemuth 2009, 4) note, categories should be defined so that they are internally as homogeneous as possible and externally as heterogeneous as possible. This way in the coding process we stressed the beginning of the description with the presumption that what users describe themselves first with is the most relevant aspect for themselves.

Table 3 *content analysis categories, descriptions and examples*

Category	Description	Example of profile description in the category
1	Student	*"PhD-ing on women in Shakespeare. Writer. Feminist. Does other stuff. My own views, re/tweets not endorsements or agreements."*
2	Researcher	*"Social policy, public admin & SDoH researcher (Aust Nat Uni). I tweet about cross-sectoral relationships, governance, policy & politics. Also, I own a giant dog"*
3	Post-Doc/PHD	*"#???? #PEMFC, PhD in Engineering, Post Doctoral Researcher at CBNU."*
4	Professor	*"Visiting Prof of Empirical Social Research @ Uni Duisburg-Essen. Co-organizer of the Cologne R User Group"*
5	Health care professional	*"Psychiatrist and lecturer in social neuroscience interested in social cognition from an interactor's rather than from an observer's point of view."*
6	Position of expertise	*"RCN Head of Policy and International Affairs. All tweets in a personal capacity."*
7	Writer/Editor/Journalist	*"Chief Editor, @NaturePhysics"*
8	Other profession	*"RECRUITER - Host, Recruiting Animal Show -- SENSITIVE? DON'T FOLLOW ME -- Feel free to criticize me in public - http://PSYCHOLOGYOFJOBHUNTING.com"*
9	Company	*"Northwest Neuro Pro, LLC neurofeedback services for performance optimization. We use the latest high tech toys to help you train your brain.*
10	Entrepreneur	*"Me? INTJ, entrepreneur, digiphile, guerilla marketer, & programmer. Let's grab a beer and change the world!"*
11	Publisher	*"Academic journal dealing with Sociology in all forms. Publishing high quality and innovative articles for over 100 years."*
12	Publication (not peer-reviewed)	*"News, trends, & conversation about global health & development | We blog at Goats and Soda. Check us out: http://www.npr.org/blogs/goatsandsoda/"*
13	Non-profit organization/ Non-profit group	*"The Calgary Centre of the Royal Astronomical Society of Canada (RASC) is one of 28 such clubs for amateurs and professionals. Our meetings are free to everyone"*
14	Government organization/ Universities	*"The Institute for Science, Society and Policy (ISSP) carries out research, teaching and public outreach on issues of emerging science and technology."*
15	Other	*"Mummy of two, British living in USA"*
16	Opinion/ Propaganda	*"Circumcision Removes 20,000 Nerve Endings & Contributes To Erectile Dysfunction. Infant Circumcision & Resale Of Stolen Foreskin Is A Billion Dollar Industry."*
17	Librarian/ Other academic	*"single mom, emerg-tech librn, ehealth, informatics, ebhc, searchengines, web2.0, MODERATE, ?, quilts/yarn/origami, food, GF, ASD, iaido. SL: Perplexity Peccable"*

As Patton (2002) acknowledges, qualitative content analysis is always subjective. The nature of the research is fundamentally interpretive and researcher's personal and theoretical understanding will undoubtedly influence research.

Thus it is common practice to use two or more coders and measure their inter-coder reliability using for instance Cohen's Kappa in order to increase the overall reliability of the results.

Results

Word frequencies

There seems to be a considerable cohesion between the top-20 most frequent words between each category, as is seen on Table 4. Words such as *health*, *university* and *research* appear in all of the lists as most frequent words. Taking a closer look, the top-20 most frequent words include a number of academic titles and words related to research, such as *science*, *research*, *PhD*, *university*, and *professor*. Only the category of Social Sciences and Humanities makes an exception, as the top-20 most frequent words list has a lack of several of academic titles and words related to research that the other categories include. Examples include such as the before mentioned *PhD*, *professor* or a *student*, for that matter.

Table 4 *top-20 most frequent words in each category, F= frequency*

Rank	F	AETS	F	MHS	F	NS	F	SSH
1	148	science	471	health	146	science	87	hearing
2	106	research	212	views	60	university	44	health
3	88	health	209	care	56	research	37	research
4	85	phd	198	nurse	55	professor	32	aids
5	84	tweets	190	medical	51	climate	31	life
6	77	scientist	184	research	50	student	29	psychology
7	77	university	177	medicine	48	tweets	28	social
8	75	life	174	nursing	43	genetics	27	university
9	74	student	155	tweets	43	views	27	quality
10	66	social	152	university	40	physics	23	technology
11	62	interested	112	science	39	health	22	goal
12	57	professor	110	student	37	scientist	21	policy
13	56	writer	109	phd	35	genomics	20	restore
14	55	genetics	108	sports	32	interested	20	offering
15	54	human	103	professor	31	news	20	number
16	51	researcher	98	life	30	human	20	newest
17	50	lover	98	researcher	30	phd	19	public
18	50	director	93	healthcare	30	working	17	researcher
19	48	news	31	physician	30	opinions	17	sociology

20	48	love	30	clinical	29	researcher	16	student

Coding of the most frequently used words

Even though the frequency of the words varies from one category to another, there are only a few words in the top-20 lists in each category that express sentimentality or specifically reflect on user's hobbies or preferences. Rather it seems that a majority of the words are very factual and might express professional interest. Examples of these include words such as *health, climate, hearing* and *genomics*. In each of the categories, the single most frequently mentioned word has a considerably larger frequency rate than the rest of the words in that category, suggesting a power law like distribution of the word frequencies. In Semertzidis et al.'s research (2013), in contrast, the top-3 most frequent words found on Twitter user profiles were *love, life* and *music*, implying that a random set of Twitter user profile descriptions might contain more sentimentality and expressions of self.

In order to define common patterns from tweeters profile descriptions, we then looked at the 100 most frequently mentioned words in the profiles. When creating their profile's, users who tweeted the top 5 articles from each main category tended to mention academic occupations and research related details on their profiles considerably more often (AET 29.9 percent, MHS 18.6 percent, NS 34.2 percent and SSH 21.6 percent) than their personal information and interests (AET 10.6 percent, MHS 13.5 percent, NS 8.9 percent and SSH 8.4 percent) (Figure 1). This result is consistent with Semertzidis et al.'s (2013) findings in which users tend to talk more about their occupation in their profile descriptions rather than interests. Majority of the coded words belonged to the miscellaneous or general group as was the case in Semertzidis et al.'s (2013) coding of a random sample of Twitter profiles. Although, in contrast to the random sample in Semertzidis et al. (2013), where words categorized as general or miscellaneous provided for 81.5% of the coded words, in our coded data this varied between 45% and 32.7% (SSH and NS respectively), as shown in Figure 1. This implies that users tweeting scientific articles were more inclined to define themselves on Twitter more specifically, often through professional and interest related keywords.

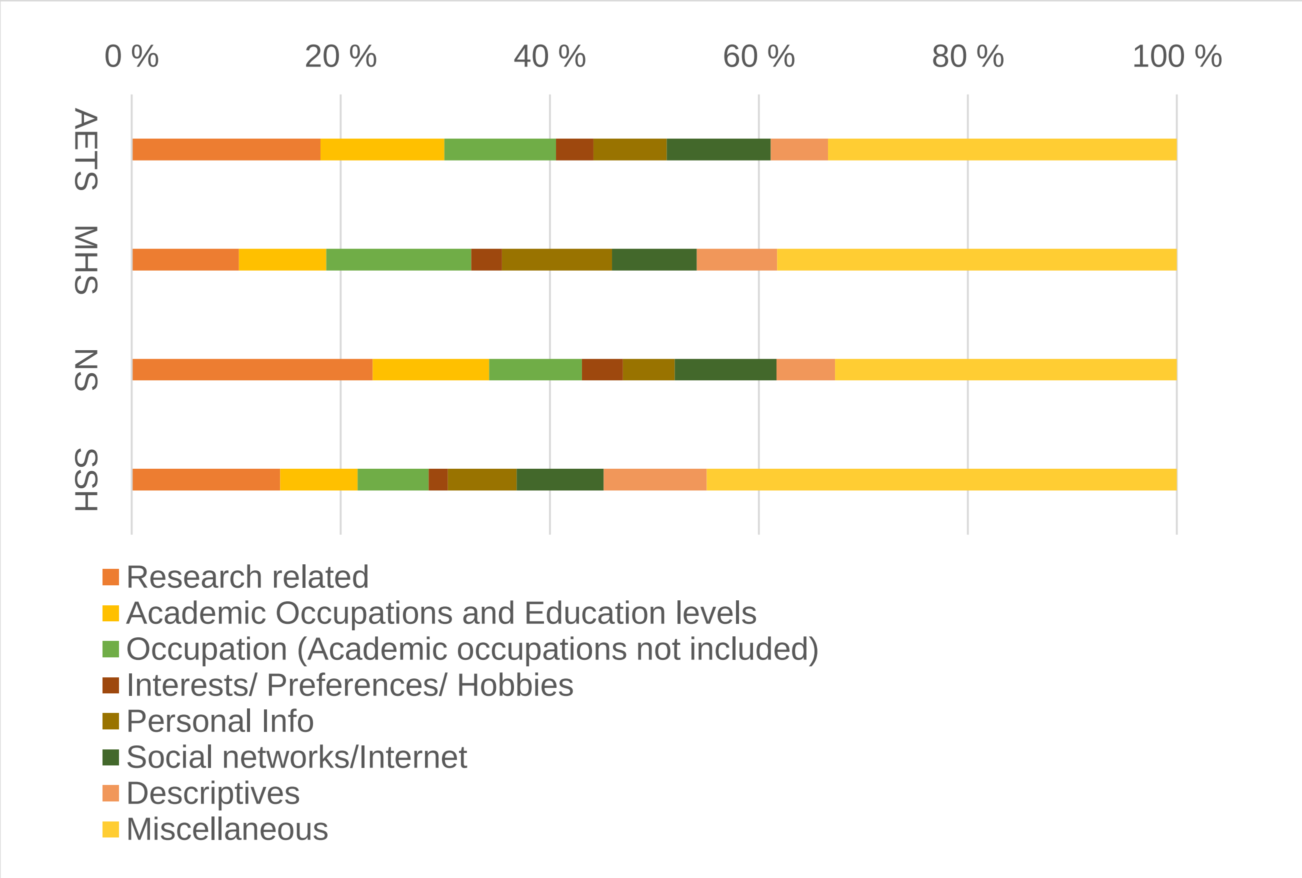

Fig. 1 *word categorization of most frequently used words*

There are some differences between the coded frequencies between categories. As shown in Figure 1, Natural Sciences' category had the most research related words and the least amount of words describing personal information. Medical and Health Sciences' category had the least research related words but most of both occupation and personal information related words. In Social Sciences and Humanities the amount of both miscellaneous and descriptive words was the highest and the relative amount of words describing interests, preferences or hobbies was more than twice smaller than in Medical and Health Sciences.

Co-word analysis

The co-word networks were filtered in order to focus on the most frequently co-mentioned words in the profile descriptions of the tweeters of the scientific articles in each of the four broad areas of science. The color coding in figures 2–5 illustrates the most tightly connected clusters, i.e. words that have been frequently mentioned together.

As discovered with the earlier methods used, the profile information of people tweeting articles in Agricultural, Engineering and Technology Sciences is dominated by words such as *science*, *research*, *health*, and *university*, whereas words describing personal information and interests are more outliers. The most frequently mentioned words do not seem to appear as often as compound words in profile descriptions, as they are rather used as descriptive terms that have their own root (Carstairs-Mccarthy 2002). In addition to the frequently mentioned words, we can see some clear clusters of frequently co-mentioned words. Some of the clusters are clearly connected to research and universities, like in the upper part of figure 2, while one of the clusters in figure 2 is connected to mental health, public health and health policy.

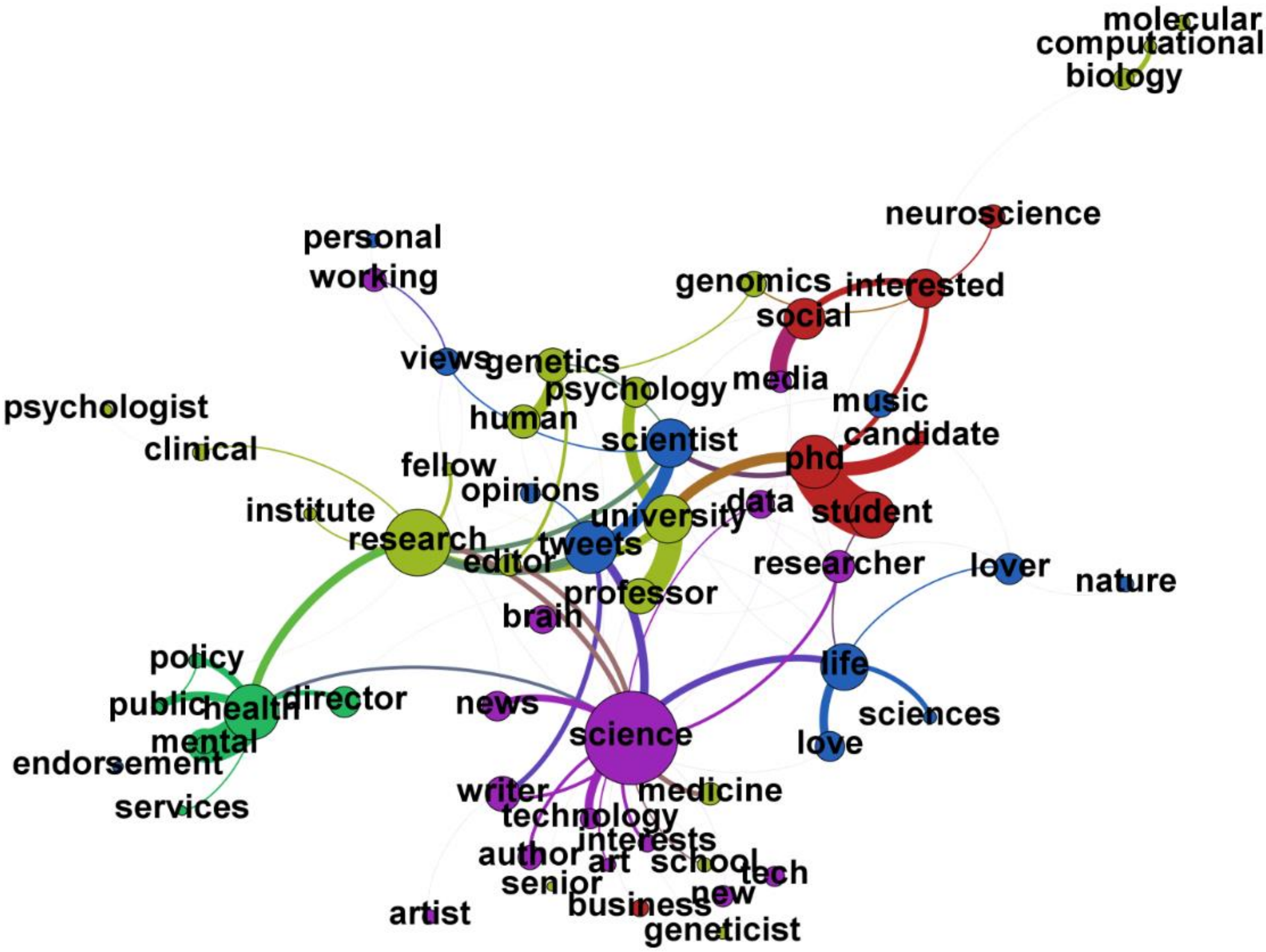

Fig. 2 *co-word map of the most frequently co-mentioned words in Twitter profiles of people tweeting papers from Agricultural, Engineering and Technology Sciences*

In Figure 3, *health* stands out as the largest and most connected node in the profile descriptions of people tweeting articles from Medical and Health Sciences. Compound word *public health* appeared in 76 descriptions, followed by other compound words such as *health care* (60 times), *health views* (59 times), *health research* (48 times) and *health policy* (47 times). Of all the words that had at least 20 co-occurrences in profile descriptions, 23 of these 42 combinations had the word *health* in them. This, in part, explains the significant size of the node *health* in Figure 3. It is noticeable though that all of these co-occurrences were not compound words, such as words *nurse* and *health* (mentioned 34 times together). While clusters connected to research and universities on one hand, and to health care professions on the other hand, were visible, the map is dominated by mentions of health and other words connected to the term.

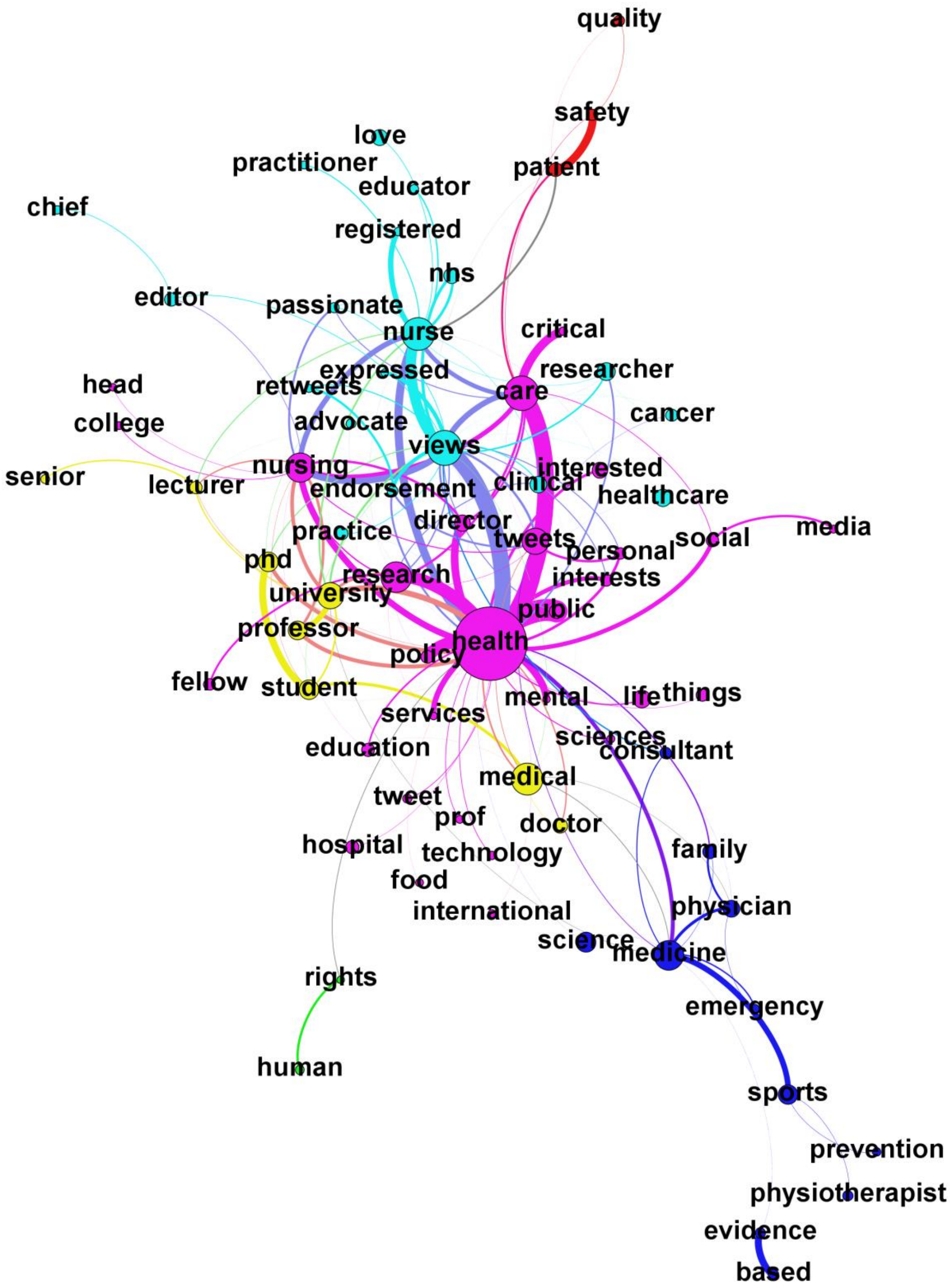

Fig. 3 *co-word map of the most frequently co-mentioned words in Twitter profiles of people tweeting papers from Medical and Health Sciences*

As with the co-word map for profile descriptions of tweeters in Agricultural Sciences, Engineering and Technology, so too is the map for Natural Sciences dominated by the word *science* and the words connected to it. Other visible clusters form around the compound word *climate change* (23 mentions together), *university professor* (19 mentions together), and around words connected to *genetics*.

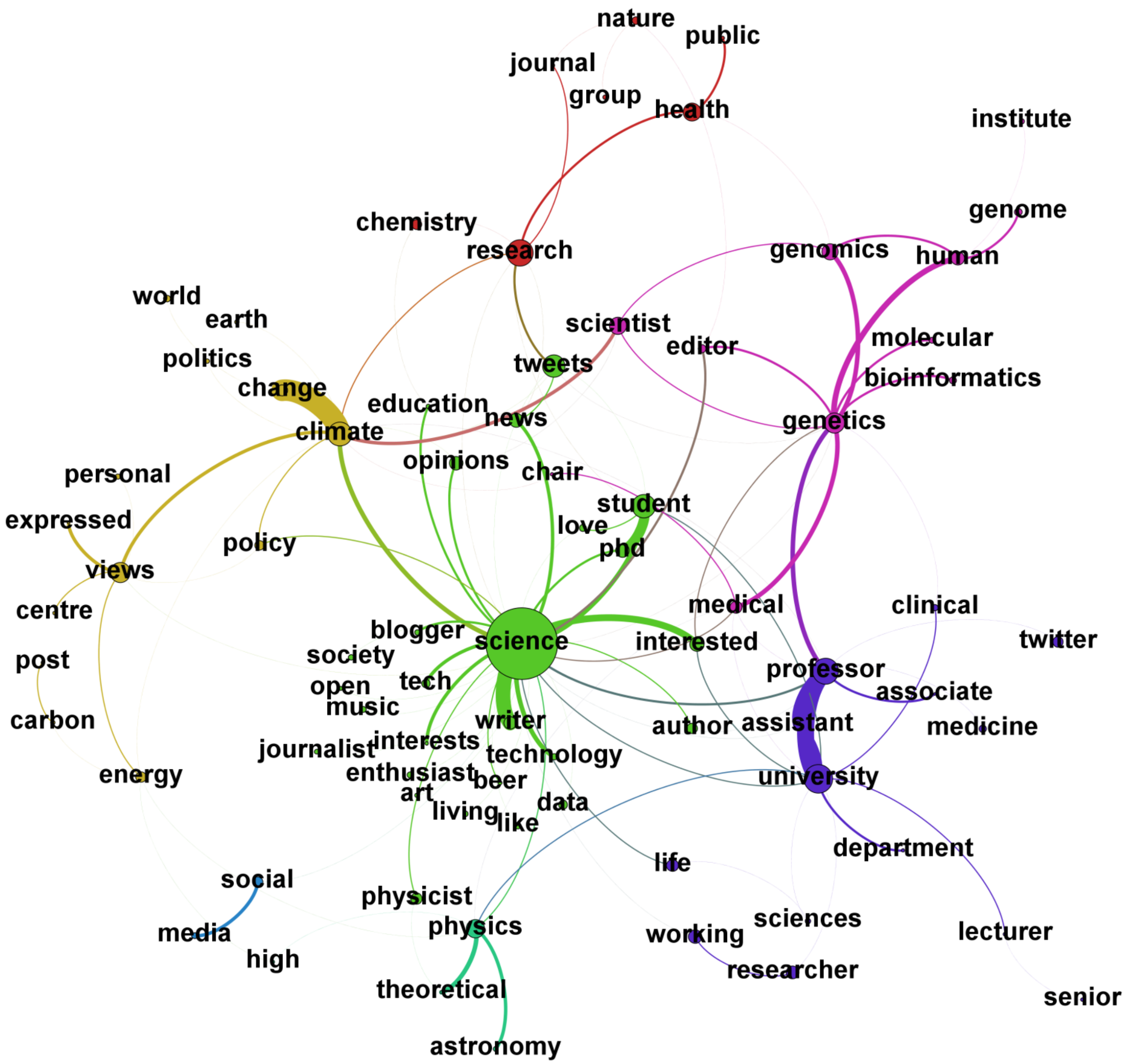

Fig. 4 *co-word map of the most frequently co-mentioned words in Twitter profiles of people tweeting papers from Natural Sciences*

As Figure 5 below shows, the co-word map of profile descriptions in Social Sciences and Humanities has two separate clusters, of which one is further divided into two clusters. The larger and more prominent of these is located on the top of the figure and it includes the largest node *hearing* as its center, which also further divides the cluster into two rather distinct clusters. While the lower cluster is related to health care and specifically to hearing, the upper part is related to companies providing hearing aids. The words *hearing* and *aids* were mentioned together most often with 51 mentions. Of the words in this upper cluster, words such as g*oal* and *hearing* as well as *quality hearing* were mentioned together 22 times. In the lower half of the graph we find words that are related to research about *health*, *public health*, *social health*, and *psychology*. Compound word *public health* appeared 21 times together.

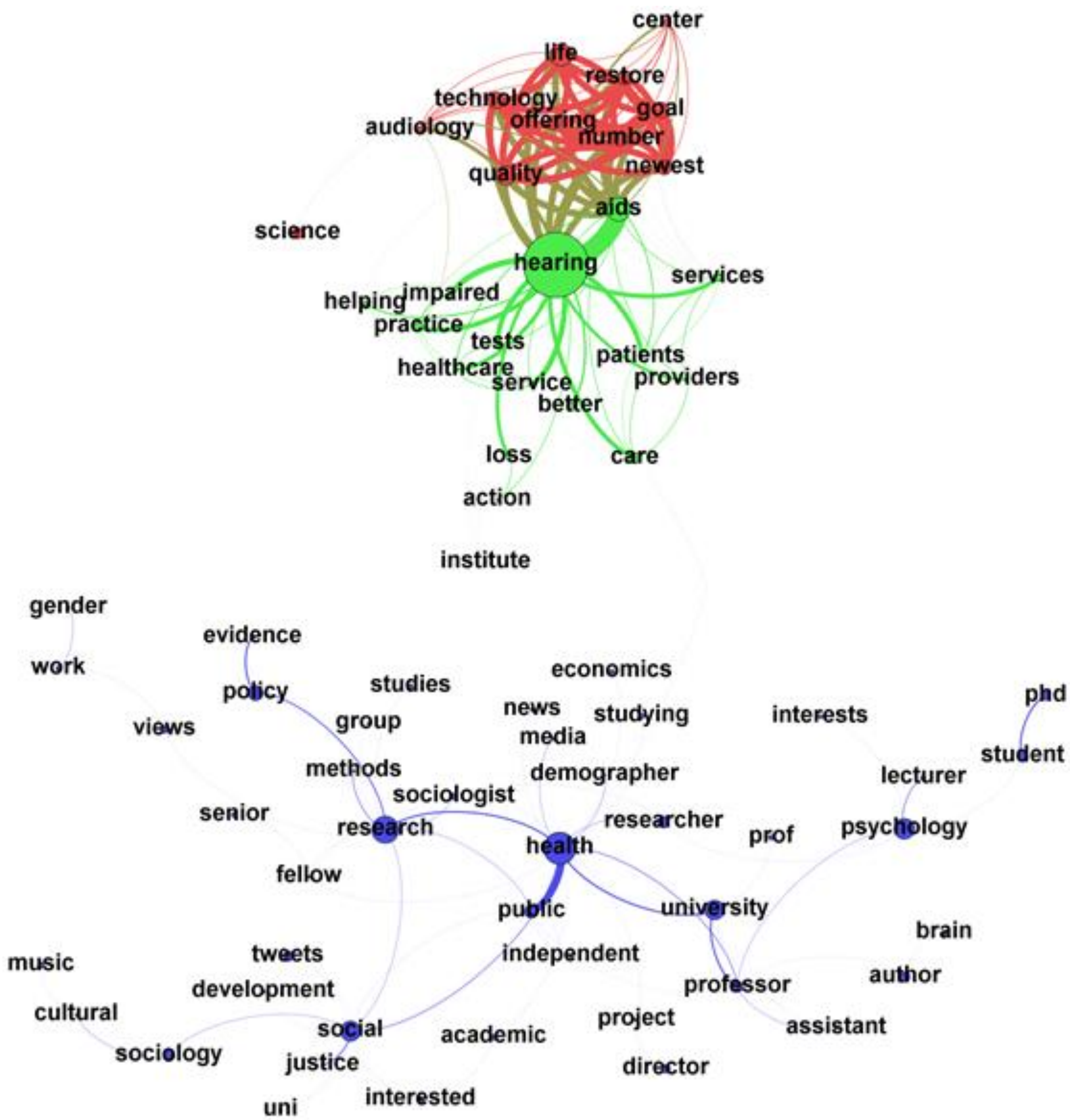

Fig. 5 *co-word map of the most frequently co-mentioned words in Twitter profiles of people tweeting papers from Social Sciences and Humanities*

Overall, the maps paint a picture of a very heterogeneous group of tweeters, where professional and personal matters are both mentioned in the profiles, suggesting at a very mixed use of Twitter for professional, in some cases academic, and personal purposes and communications.

Classifying tweeter profile

A random sample of 100 Twitter profile descriptions from each of the four categories were taken for manual coding. To confirm the representativeness of the random sample compared to the full sample we checked how the tweeters were distributed among the articles in both cases. In all the other categories the distributions differed +- 5 percent, except for Social Sciences and Humanities. In SSH, *Atheists Become Emotionally Aroused When Daring God to Do Terrible Things* had the highest Twitter posts count with over 32 percent (180 posts, see Table 1) of the posts, but in the random sample it had only 14 percent (14 tweets) of the tweets. In consequence, *Music reduces pain and increases functional mobility in fibromyalgia* had 22 percent (120 posts, see Table 1) of the posts but gained 30 percent (30 tweets) of the tweets in the random sample. The tweeters of these two articles are thus under- and overrepresented respectively in the random sample. The first author categorized all profile descriptions in the random sample and the second author categorized 50 random descriptions from each category. Cohen's Kappa was run to determine inter-coder reliability. There was substantial agreement between the results, K=0,769 ($p<0.05$). Figure 6 represents the distribution of profile descriptions.

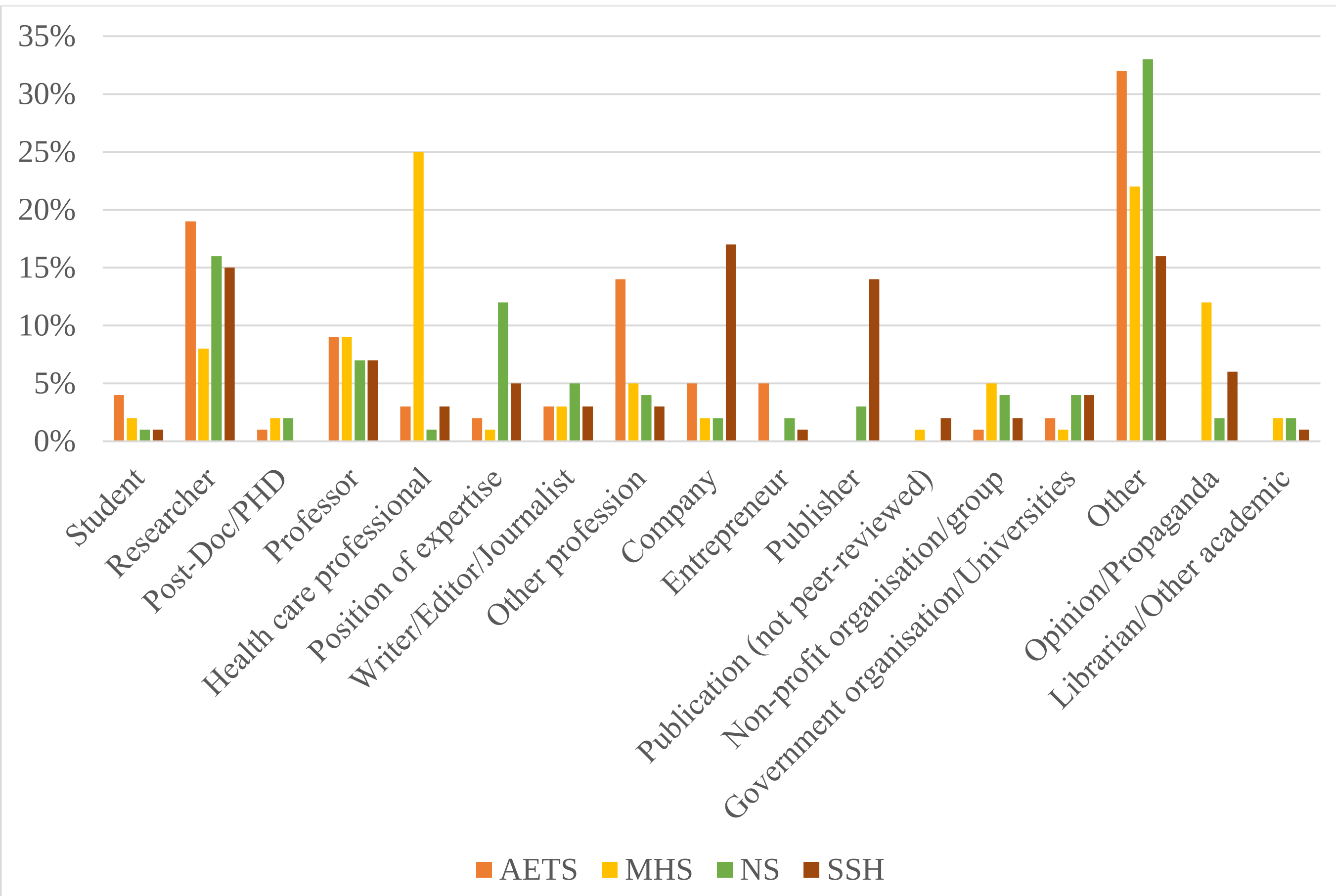

Fig. 6 *distribution of Twitter profile descriptions according to categorization*

In Agricultural, Engineering and Technology Sciences (32 percent of all analyzed descriptions) as well as in Natural Sciences (33 percent) most of the profiles were coded in "other" description, as seen in Figure 6. These included descriptions such as *"one of the better-known cyberandy"*, "*BEWARE: This twitter is about pop culture, Russian politics and sometimes genomics"* and *"a process, policy, and politics junkie".* Descriptions mentioning the user's position as a researcher were second most popular in both categories (AETS 19 percent and NS 16 percent). In Medical and Health Sciences, the most popular category was "health care professionals" that included descriptions such as *"Husband / Father / Doctor of PT / Proud UC & UD Graduate / Life Long Cincinnati Reds, Bengals & Bearcats Fan"* and *"Research Midwife, Care Maker, Parent Educator. Perineal trauma Lead, Perineal suturing instructor. Commenced an MSc in Education, Exciting times!"* The "other" category, including descriptions such as *"Thrown away the keys, and locked out of the comfort zone"* was the second most popular in MHS with 22 percent of the descriptions. Besides the high categorization rate of researchers, profile descriptions that implied the user to be a professor were also relatively high in each category (AETS 9 percent, MHS 9 percent, NS 7 percent and SSH 7 percent).

Social Sciences and Humanities were an exception as the distribution of profile descriptions was more wide-ranging. Most of the descriptions were categorized as being company profile descriptions (17 percent) such as *"The Trusted Name For Better Hearing!"* and *"Our practice is devoted to helping the hearing impaired"*. Descriptions such as *"Former Marine. Policy nerd. Runner. Detroit FBK. RTs/links imply interest, not necessarily endorsement"* that coded into "other" description category followed close behind (16 percent). Third most descriptions were categorized as researcher profiles (15 percent) with examples like *"Lecturer and Researcher in Psychology"* and *"sociologist at Ryerson University, feminist, Torontonian, one-time collector of yo-yos"* and fourth most for publishers (14 percent), like *"Comprehensive and critical assessment of the relationship between research evidence and the concerns of policy makers and practitioners, as well as researchers"*. In other categories the share of profile descriptions about publishers ranged from zero occurrences to 3 percent.

In Agricultural, Engineering and Technology Sciences third most descriptions (14 percent) fell under "other profession" categorization, such as *"Piano teacher for over 25 years. Qualified through the Associated Board Of The Royal Schools Of London"* or "*Designer, illustrator, maybe a photographer, possibly a Wookiee, definitely a fattie.. I mean foodie."* In Natural Sciences the third most descriptions were "position of expertise". This categorization entailed profile descriptions such as *"Physicist"* or *"Cowboy coder in the business of digital enlightenment, cloud computing, cryptocurrencies, transparency & the coming age of trust. F from ::1 @stackape @utterio."* Descriptions labeled as "opinion/propaganda" appeared as the third largest description group in Medical and Health Sciences (12 percent) with examples such as *"It's your family's penis and they can do what they want to it. circumcision #i2"* and *"Interested in environment, politics Frustrated over world injustice and double standard. Norwegian and English tweets. Free Chelsea Manning #SnowdenPeacePrize"* and the sixth largest group in Social Sciences and Humanities (6 percent) where examples included *"Pray Hard, Keep Watch, Stand Firm, Love God, Live Crucified, Serve Others, Share Christ, Shun sin, Be Holy, Walk Humbly, Be Ready, Finish Strong & Overcome!!!"* and *"Vote for what is good for USA and NOT for what is good for party. Free Film: http://www.stealingamericathemovie.org"*.

Discussion

This exploratory research analyzed people tweeting popular scientific articles that had at least one Finnish author affiliation. We set out to discover whether there were observable user groups who tweeted scientific articles. The results demonstrated that even though there are considerable similarities among the tweeters from one main category to the other, we were also able to define different tweeting profiles.

The most significant observation is that people who share scientific articles that are popular on Twitter tend to describe themselves through their occupation and expertise. This result was in contrast to Semertzidis et al.'s (2013) findings of a random sample of tweeters - that people describe themselves on profile descriptions through occupation and hobbies. Unlike in a random sample of a set of users and their profile description's (see Semertzidis et al. 2013), words describing personal information or interests did not top the list as the most frequently mentioned words.

The themes of the articles in each main category varied considerably, which was reflected in the profile descriptions as well: the more field-specific the articles were, the more research-related words were found in profile descriptions. The category of Natural Sciences category was an evident example of this. The top 20 most frequent words in the category consisted almost solely of words that were categorized as research-related or as academic occupations and education levels. On the other hand, the data from Medical and Health Sciences category show that words describing occupations appear more frequently in this category than in the rest. This finding is supported by the high frequency of analyzed profile descriptions that were classified as promoting *health care professional*ism. In our research the volume of descriptive and miscellaneous words was also considerably lower than in Semertzidis et al.'s (2013) findings. This suggests that users who share or discuss scientific articles tend to present themselves more precisely and with less filler words than Twitter users in general.

Another result was that scientific articles were used to promote ideological views. In news media, the success of news is often measured by the content of the piece. If the news "include conflict, novelty, geographic or cultural proximity to the audience, prominence of individuals, impact or personal relevance to the audience, and timeliness" it is considered to be valuable news (Badenschier & Holger 2012 in Dahlstrom 2014, 13615). In social media, these news values could also be applied to the dissemination of academic articles, such as in Medical and Health Science articles *Nurse staffing and education and hospital mortality in nine European countries: a retrospective observational study* proving how nursing cutbacks directly affect death rates in hospitals and *Cultural Bias in the AAP's 2012 Technical Report and Policy Statement on Male Circumcision* arguing that the pediatrics in the US were unable to objectively observe the benefits and risks in circumcision of young males. Both of these articles could be seen to meet several of the before-mentioned news values despite being science articles. This case can also be argued for Social Sciences and Humanities articles such as *Atheists Become Emotionally Aroused When Daring God to Do Terrible Things* and *Happiness: Before and After the Kids* that even have very self-explanatory headlines. A random sample of content analysis on profile descriptions found both Medical and Health Sciences and Social Sciences and Humanities to have an unusually high degree of Twitter accounts that have profile descriptions that are clearly written to promote ideological views, such as a ban to the circumcision of male infants or to promote a Christian lifestyle. This implies that Twitter users disseminate scientific research in order to promote their beliefs and world-views. By utilizing peer-reviewed papers in order to gain significance for the prominence of their message, Twitter users might be able to portray an ideological view as an evident factual knowledge. It is to be noted that while preparing the data for analysis, we had already eliminated all duplicate Twitter accounts. It is therefore quite possible that one or more individuals orchestrate propaganda through several different accounts that all tweet the same content. This, the spread of ideological viewpoints through the dissemination of research in social media, is an area of research that might benefit largely if observed by using altmetric methods.

Medical and Health Science articles gained a significant amount of Twitter attention compared to the other articles, as seen in Table 1. This might reflect a sign of personal relevance to the audience. Notable are also the significantly lower tweeting counts received by the top five Social Sciences and Humanities articles. In general, several of the top five articles were classified as genome-wide association studies (GWAS), or were performed using meta-analysis, where a study combines the results from several dozens or even hundreds of individual research papers. This sort of meta-analysis combines data from different independent studies in a new analysis and thus aims to strengthen the understanding of a particular topic. As there were several GWAS studies among the most tweeted articles in our data as well, it is presumable that the impact these GWAS' attract is purposefully shared online as well, where these studies have the ability to attract the attention of an even larger public.

Nelhans & Gunnarsson Lorentzen (2016) suggest that research articles are shared not only for communication purposes but also for marketing and even spam, usually by publishers. In our study, except for Social Sciences and Humanities, we found little evidence of this. The frequency of both private sector tweeters and observed academic publishers was notable only in Social Sciences and Humanities, where both profiles were clearly represented in profile description content analysis. In SSH, the marketing seemed to culminate on the article *Music reduces pain and increases functional mobility in fibromyalgia*, as most of the observed profile descriptions marked portraying themselves as *Company* were promoting hearing center practices and better hearing care. In table 4, for example, words such as *hearing* and *aids* were among the most frequent words in the SSH category even though there is no evident connection between the words and SSH category in itself. The most probable reason for these words to trend like they did is because one of the most tweeted articles (*Music reduces pain and increases functional mobility in fibromyalgia)* was used by Twitter profile accounts promoting hearing aids as an advertise promoting for their services. This division between commercial profiles and the rest was also evident from the data visualization in Figure 5.

The data used for this study was not a random sample but pre-determined based on the popularity of the research articles on Twitter that had at least one Finnish author affiliation. As the selected articles do not offer a statistically valid sample of data, these results are not applicable for further scientific use as such. The data used for this research was also cleared of all tweets that were not written in English. This might have had an impact on the analysis, since some of the articles had a considerable amount of tweets written in other languages besides English. In addition, the selected articles included single author articles and articles with hundreds of authors. Thus, for specific articles it might even be difficult to find a country that was not represented among the authors. How this might influence the results of an altmetric study at a national level is unclear, but it raises also the question of how articles with multiple authors should be treated in altmetrics. Should for instance fractional counting be applied as is sometimes the case in bibliometrics with citations? Further research is required on this topic.

However, the field of altmetrics is predominantly occupied by statistical and quantitative analysis. As the field grows and gains importance, the demand for qualitative, explanatory theories rises. Big data offers us an invaluable path towards developing new ways to measure scientific impact on different online services. However, crunching numbers only takes us up to a certain point. The results need to be not only explained but also interpreted. As Patterson and Monroe (1998, 320) put it, "*[w]hat can be explained through falsifiable hypotheses is necessarily limited. What can be explained also must be interpreted and understood.*"

Acknowledgments The authors would like to thank MSc. Jonne Lehtimäki for insightful comments on the article. This research was financed by The Finnish Ministry of Education and Culture's Open Science and Research Initiative 2014–2017 (funding number: OKM/33/524/2015).

References

Badenschier, F., Holger, W. (2012). Issue selection in science journalism: towards a special theory of news values for science news? *Sociology of the Sciences Yearbook*, *28* (February), 273–289. https://doi.org/10.1007/978-94-007-2085-5

Bird, S. J. (2014). Socially responsible science is more than "good science". *Journal of Microbiology & Biology Education*, *15*(2), 169–72. https://doi.org/10.1128/jmbe.v15i2.870

Blondel, V. D., Guillaume, J.-L., Lambiotte, R., & Lefebvre, E. (2008). Fast unfolding of communities in large networks. *Journal of Statistical Mechanics: Theory and Experiment*, *10008*(10), 6. https://doi.org/10.1088/1742-5468/2008/10/P10008

Bornmann, L. (2016). What do altmetrics counts mean? A plea for content analyses. *Journal of the Association for Information Science and Technology*. https://doi.org/10.1002/asi.23633

Brigham, T. J. (2014). An introduction to altmetrics. *Medical Reference Services Quarterly*, *33*(4), 438–47. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/25316077

Carstairs-Mccarthy, A. (2002). An Introduction to English Morphology: Words and Their Structure. *Linguistics*, 160.

Courtial, J. P. (1994). A coword analysis of scientometrics. *Scientometrics*, *31*(3), 251–260. https://doi.org/10.1007/BF02016875

Dahlstrom, M. F. (2014). Using narratives and storytelling to communicate science with nonexpert audiences. *Proceedings of the National Academy of Sciences of the United States of America*, *111*(Supplement_4), 13614–13620. https://doi.org/10.1073/pnas.1320645111

Desai, T., Shariff, A., Shariff, A., Kats, M., Fang, X., Christiano, C., & Ferris, M. (2012). Tweeting the meeting: An in-depth analysis of Twitter activity at kidney week 2011. *PLoS ONE*, *7*(7).

https://doi.org/10.1371/journal.pone.0040253

Eysenbach, G. (2011). Can tweets predict citations? Metrics of social impact based on Twitter and correlation with traditional metrics of scientific impact. *Journal of Medical Internet Research*, *13*(4). https://doi.org/10.2196/jmir.2041

Hadgu, A. T., & Jäschke, R. (2014). Identifying and analyzing researchers on twitter. In *CEUR Workshop Proceedings* (Vol. 1226, pp. 164–165). https://doi.org/10.1145/2615569.2615676

Haustein, S., Bowman, T. D., Holmberg, K., Peters & Larivière, V. (2014). Astrophysicists on Twitter: An in-depth analysis of tweeting and scientific publication behavior. *Aslib Journal of Information Management*, *66*(3), 279–296. https://doi.org/10.1108/AJIM-09-2013-0081

Holmberg, K. (2014). The meaning of altmetrics. *IATUL Annual Conference Proceedings*, (35), 1–11. Retrieved from http://search.ebscohost.com/login.aspx?direct=true&db=llf&AN=97787073&site=ehost-live

Holmberg, K., & Thelwall, M. (2014). Disciplinary differences in Twitter scholarly communication. https://doi.org/10.1007/s11192-014-1229-3

Hsieh, H.-F., & Shannon, S. E. (2005). Three Approaches to Qualitative Content Analysis. *Qualitative Health Research*, *15*(9), 1277–1288. https://doi.org/10.1177/1049732305276687

ICSU. (2005). *Science and Society : Rights and Responsibilities*. *Science*.

Ke, Q., Ahn, Y.-Y., & Sugimoto, C. R. (2016). A Systematic Identification and Analysis of Scientists on Twitter. Digital Libraries; Physics and Society. Retrieved from http://arxiv.org/abs/1608.06229

Lambiotte, R., Delvenne, J.-C., & Barahona, M. (2008). Laplacian Dynamics and Multiscale Modular Structure in Networks. *arXiv Preprint arXiv:0812.1770*, 1–29. https://doi.org/10.1109/TNSE.2015.2391998

Larivière, V., & Gingras, Y. (2009). The Decline in the Concentration of Citations , 1900 – 2007. *Journal of the American Society for Information Science and Technology*, *60*(4), 858–862. https://doi.org/10.1002/asi

Lincoln, Y. S., & Guba, E. G. (1985). *Naturalistic Inquiry*. *Naturalistic Inquiry*. https://doi.org/10.1177/1473325006070288

Liu, C. L., Xu, Y. Q., Wu, H., Chen, S. S., & Guo, J. J. (2013). Correlation and interaction visualization of altmetric indicators extracted from scholarly social network activities: dimensions and structure. *Journal of Medical Internet Research*, *15*(11). https://doi.org/10.2196/jmir.2707

Mislove, A., Lehmann, S., Ahn, Y., Onnela, J., & Rosenquist, J. N. (2011). Understanding the Demographics of Twitter Users. *Artificial Intelligence*, 554–557. Retrieved from http://www.aaai.org/ocs/index.php/ICWSM/ICWSM11/paper/viewFile/2816/3234

Neiger, B. L., Thackeray, R., Burton, S. H., Giraud-Carrier, C. G., & Fagen, M. C. (2013). Evaluating Social Media's

Capacity to Develop Engaged Audiences in Health Promotion Settings: Use of Twitter Metrics as a Case Study. *Health Promotion Practice*, *14*(2), 157–162. https://doi.org/10.1177/1524839912469378

Nelhans, G., & Gunnarsson Lorentzen, D. (2016). Twitter conversation patterns related to research papers. *Information Research*, *21*(2). Retrieved from http://www.informationr.net/ir/21-2/SM2.html

Patton, M. Q. (2002). *Qualitative research and evaluation methods*. *Qualitative Inquiry* (Vol. 3rd). https://doi.org/10.2307/330063

Priem, J., & Costello, K. L. (2010). How and why scholars cite on Twitter. In *Proceedings of the ASIST Annual Meeting* (Vol. 47). https://doi.org/10.1002/meet.14504701201

Priem, J., Groth, P., & Taraborelli, D. (2012). The Altmetrics Collection. *PLoS ONE*, *7*(11). https://doi.org/10.1371/journal.pone.0048753

Segerberg, A., & Bennett, W. L. (2011). Social Media and the Organization of Collective Action: Using Twitter to Explore the Ecologies of Two Climate Change Protests. *The Communication Review*, *14*(3), 197–215. https://doi.org/10.1080/10714421.2011.597250

Semertzidis, K., Pitoura, E., & Tsaparas, P. (2013). How people describe themselves on Twitter. *Proceedings of the ACM SIGMOD Workshop on Databases and Social Networks - DBSocial '13*, 25–30. https://doi.org/10.1145/2484702.2484708

Thelwall, M., Tsou, A., Weingart, S., Holmberg, K., & Haustein, S. (2013). Tweeting links to academic articles. *Cybermetrics*, *17*(1), 1–8.

Uddin, M. M., Imran, M., & Sajjad, H. (2014). Understanding Types of Users on Twitter. *arXiv Preprint*, 6. Retrieved from http://arxiv.org/abs/1406.1335

Vergeer, M., Hermans, L., & Sams, S. (2011). Is the voter only a tweet away? Micro blogging during the 2009 European Parliament election campaign in the Netherlands. *First Monday*, *16*(8). https://doi.org/10.5210/fm.v16i8.3540

Zhang, Y., & Wildemuth, B. M. (2009). Qualitative Analysis of Content. *Applications of Social Research Methods to Questions in Information and Library Science*, 421. https://doi.org/10.1002/hbm.20661