Social Media in Scholarly Communication

A Review of the Literature and Empirical Analysis of Twitter Use by SSHRC Doctoral Award Recipients

Study commissioned by:

Social Sciences and Humanities Research Council

Conducted by:

Canada Research Chair on the Transformations of Scholarly Communication

University of Montreal

Authors:

Samantha Work
Stefanie Haustein
Timothy D. Bowman
Vincent Larivière

Executive Summary

This report has been commissioned by the Social Sciences and Humanities Research Council (SSHRC) to analyze the role that social media currently plays in scholarly communication as well as to what extent metrics derived from social media activity related to scholarly content can be applied in an evaluation context.

Scholarly communication has become more diverse and open with research being discussed, shared and evaluated online. Social media tools are increasingly being used in the research and scholarly communication context, as scholars connect on Facebook, LinkedIn and Twitter or specialized platforms such as ResearchGate, Academia.edu or Mendeley. Research is discussed on blogs or Twitter, while datasets, software code and presentations are shared on Dryad, Github, FigShare and similar websites for reproducibility and reuse. Literature is managed, annotated and shared with online tools such as Mendeley and Zotero, and peer review is starting to be more open and transparent. The changing landscape of scholarly communication has also brought about new possibilities regarding its evaluation. So-called altmetrics are based on scholarly social media activity and have been introduced to reflect scholarly output and impact beyond considering only peer-reviewed journal articles and citations within them to measure scientific success. This includes the measurement of more diverse types of scholarly work and various forms of impact including that on society.

This report provides an overview of how various social media tools are used in the research context based on 1) an extensive review of the current literature as well as 2) an empirical analysis of the use of Twitter by the 2010 cohort of SSHRC Doctoral Award recipients was analyzed in depth. Twitter has been chosen as one of the most promising tools regarding interaction with the general public and scholarly communication beyond the scientific community. The report focuses on the opportunities and challenges of social media and derived metrics and attempts to provide SSHRC with information to develop guidelines regarding the use of social media by funded researchers as well support the informed used of social media metrics.

Opportunities and Challenges of Social Media Use

Academics face different opportunities and challenges when using social media in a research context. In general, the use by and presence of researchers on social media is increasing. Differences can be identified between different platforms and types of tools as well as between scientific disciplines. The following key findings were identified based on the literature review:

- Social media platforms are designed to facilitate sharing and engaging with content and other users. They can help researchers to establish new connections, disseminate and discover as well as discuss their research. They may enable research activities to take place more quickly and efficiently and facilitate international and interdisciplinary collaboration. This includes crowdsourcing efforts for scientific discoveries, and soliciting feedback from experts in a network before submission to a journal or giving a conference presentation.
- Sharing research outputs on social media could increase the visibility of the work and attract publicity, not just among scientists but the general public as well. Using the functionality of different platforms properly could amplify its reach by increasing discoverability, such as utilizing hashtags on Twitter or writing a blog post to explain complex research in layman's terms.

- The professional use of social media by researchers is increasing, which suggests that scholarly communication is becoming more heterogeneous, but also more visible to those outside the scientific community.
- There is little evidence, however, that the research process itself has been fundamentally altered by social media and even though new communication genres and tools develop, the peer-reviewed journal article remains the central and most important form of scholarly communication.
- Due to missing incentives and recognition by funding agencies, many researchers lack motivation to use social media and consider it a waste of time. Only a few funding agencies have started to acknowledge new types of output. Social media requires time commitment and the large amount of different platforms available require researchers to prioritize and select the most suitable ones.
- The ease and speed of communication on social media can also lead to damaged reputations quickly. Serious repercussions resulting from the improper use of social media may make researchers hesitant to adopt it, particularly in the absence of usage guidelines from universities and funders. To minimize these occurrences, funders can provide best practice guidelines for specific platforms.
- Usage among platforms is extremely varied. While specialized social networks like Academia.edu and ResearchGate cater to an academic audience, they are used mostly passively; Facebook has a large uptake but usage is mostly private; Twitter has the potential to reach a broad audience but suffers from low uptake among researchers; Mendeley behaviour correlates with traditional citation measures but its networking functionality is more or less ignored. Certain demographics are underrepresented on social media, including by country, age and research discipline, and thus user bias transfers to discussions and collaborations on the platforms.
- As scholarly communication is becoming more diverse and open, the evaluation system needs to adapt. By including new forms of research output and impact, metrics based on social media activity have the potential to complement evaluations based on the number of peer-reviewed journal articles and citations. However, low uptake and the underrepresentation of certain demographics of social media platforms reflects in the metrics derived from them and the ability of currently captured social media metrics to reflect scientific or societal impact has yet to be validated.

SSHRC Doctoral Award Recipients on Twitter

The analysis of the Twitter use of the 2010 cohort of SSHRC Doctoral Award recipients yielded the following key findings:

- A Twitter profile could be identified for slightly less than one-third of the SSHRC Doctoral Award recipients, representing a comparably high uptake compared to other studies of researchers' use of Twitter.
- The majority of award recipients used the account description to identify as an academic, mentioning their academic disciplines, titles or universities. Fewer than 10% of the Twitter bios were strictly non-academic. This implies that SSHRC Doctoral Award recipients considered using Twitter, at least to some extent, professionally.
- Active Twitter use as reflected in the number of tweets, followees and followers was skewed with a few very active users. Although some differences between fields could be observed regarding tweeting activity, none were significant.

- The majority of tweets contained user mentions, indicating direct interactions between users and one-third were retweets or contained hashtags, respectively. About half of the award recipients' tweets contained URLs linking to external sources. News websites were the most popular sources, even when discussing academic topics, which suggests that their topics are indistinguishable from current events, or that users either targeted a more general audience instead of the specialized academic community. Scholarly publications were the second most popular link category among academic tweets with links.
- The majority of the tweets sent by SSHRC Doctoral Award recipients from History, Modern Languages and Literature, Education and Political Science did not refer to academic topics, while 21% were related to their discipline, 5% to academic life in general and 4% to the topic of their PhD thesis.

Content

List of Figures

List of Tables

1 Introduction

With the advent of the Internet and particularly the rise of the social web, scientific communication is becoming more diverse, progressively transparent and open. Although the scientific journal remains the central communication channel of the scholarly community 350 years after its creation, research results are increasingly diffused and discussed outside scientific journals. Social media have found their way into academia with an increasing number of scientists using platforms such as ResearchGate, LinkedIn, and Facebook (Van Noorden, 2014). Research results and ideas are published in blog posts and discussed on Twitter; datasets, software code, and presentations are shared across online on platforms such as Dryad, Github, FigShare, and SlideShare; literature is stored, managed, and annotated with social bookmarking and reference managers such as Mendeley, Zotero, and CiteULike; and peer review is beginning to be performed openly online by anonymous as well as non-anonymous referees. With the increasing use of social media tools comes the pressure for scholars to manage their online personas and social capital on various social media platforms. Scholars and others social groups (such as athletes, politicians, and musicians, to name just a few) have faced scrutiny and punishment for controversial communications they have made in online and social media contexts, as their (often personal) contributions were framed in such a way as to have an effect on their professional lives. With regards to scholars specifically, there have been those that were disciplined, placed on leave, or had their job offer rescinded (Berrett, 2010; Herman, 2014; Ingeno, 2013; Jaschik, 2014; Rothschild & Unglesbee, 2013) for messages posted within the context of social media.

The diversification and increasing openness of the scientific communication system also entails the potential to change the way science is being evaluated. So-called altmetrics—short for alternative (to citation) metrics—were developed to provide new filters and counterbalance the oversimplification of research evaluation based on bibliometric indicators. Instead of relying only on peer-reviewed journal articles and citations within them, the metric-based evaluation system should adjust and become more diverse to include all types of research products (Piwowar, 2013) and various types of impact on the scientific community, as well as on society at large (Priem & Hemminger, 2010; Priem, Taraborelli, Groth, & Neylon, 2010). Even though the term was coined and popularized in 2010 (Priem, 2010), the idea behind the concept of altmetrics, i.e. that of capturing traces of scholars and their documents on the web to measure more diverse impact of science, had been discussed years before (Almind & Ingwersen, 1997; Cronin, 2005; Cronin, Snyder, Rosenbaum, Martinson, & Callahan, 1998; Thelwall, Vaughan, & Björneborn, 2005). The term 'altmetrics' has been the subject of many debates and criticized (Rousseau & Ye, 2013), yet there is still no accepted consensus regarding the proper terminology for the set of metrics that measure scholarly activities on social media. Altmetrics are usually based on activity on social media platforms, which relates to scholars or scholarly content. Typical examples of altmetrics include tweets, mentions in blog posts, readership counts on Mendeley, posts, likes and shares on social networks such as Facebook and Google Plus, and recommendations and ratings on F1000. However, altmetrics also comprise mentions in mainstream media or policy documents, as well as usage metrics such as full text views and downloads, although these have been available long before the concept of altmetrics was introduced. The common denominator of these heterogeneous metrics is that they exclude, and are opposed to, 'traditional' bibliometric indicators. It has been suggested that it might be more useful to consider all of these metrics—citations, downloads, various altmetrics—together as different tools of the scholarly metrics toolbox (Haustein, Sugimoto, & Larivière, 2015). As this report focuses on the subset of altmetrics, which are based on activities on social media, the preferred

term in this report is 'social media metrics' as introduced by Haustein, Larivière, Thelwall, Amyot, and Peters (2014) because it emphasizes the origin of the data instead of intent or meaning.

Motivation and Structure of Study

This report was commissioned by the Social Sciences and Humanities Research Council (SSHRC) in order to provide an overview of the use of social media by researchers, as well as an in-depth analysis of the current Twitter use by the recipients of SSHRC Doctoral Award recipients for the competition year 2010. Twitter is one of the most promising tools regarding outreach to a wider audience, as it reaches beyond the closed scientific community.

More specifically, the study explores the opportunities and challenges of social media metrics in research evaluation and supports SSHRC in identifying guidelines regarding the use of social media by funded researchers. The report contains a comprehensive review of the literature on social media in scholarly communication across all domains (Section 2), and an empirical section (Section 3) that analyzes the uptake and specific use of Twitter by the 2010-2011 cohort of SSHRC Doctoral Award recipients, which is supported by a comparison with and discussion based on results found in the reviewed studies. The report provides sufficient background to answer the following questions raised by SSHRC, based on the literature review:

- RQ.1 Opportunities and Challenges of Social Media: How can academics use social media in a research context? What are the opportunities and challenges?
- RQ.2 Role of Funding Agencies: How can a funding agency, such as SSHRC, recognize and encourage the proper use of social media to communicate research findings?
- RQ.3 Potential of Social Media Metrics: What metrics based on social media activity related to scholarly documents or researchers do currently exist, what is known about their meaning and how can they be captured?

The analysis of the Twitter use of the 2010-2011 SSHRC Doctoral Award recipients helps to answer the following questions:

- RQ.4 Twitter Uptake: how far do SSHRC Doctoral Award recipients already use Twitter?
- RQ.5 Twitter Use: What is the tweeting behaviour of SSHRC Doctoral Award recipients? How much do they use it in an academic context?
- RQ.6 Field Differences: How does Twitter use of SSHRC Doctoral Award recipients differ between fields?

Detailed methods regarding literature review (Section 2) and Twitter data analysis (Section 3) are provided in the Appendix.

2 Literature Review

The literature review (see Appendix A.1 for a description of methods) is based on 139 documents relevant to social media and scholarly communication and related metrics. The key findings from these documents are organized into three sections: Section 2.1 focuses on how academics use social media in a research context, Section 2.2 discusses how much academic content and research findings are disseminated on social media, and Section 2.3 examines how different social media metrics compare with each other and traditional citation measurements. The literature review concludes with addressing the main limitations of social media metrics.

2.1 How Academics Use Social Media for Research

The extent to which social media has transformed the academic landscape is continuously being assessed as norms and perceptions change within science. In 2010, a survey of over 1,000 UK academics revealed that 76% believed it was likely that new media formats and types of online publication would become more important in the next five years (R. N. Procter et al., 2010). Published in the same year, in-depth interviews examining the perceptions of publishing on web 2.0 platforms revealed that traditional publishing still carried the most importance in determining academic advancement, with no evidence that younger academics had taken advantage of new platforms for dissemination (Harley, Acord, Earl-Novell, Lawrence, & King, 2010). Interviewees in Acord and Harley (2012) conceded that while collaboration had become technically easier through the adoption of web 2.0 tools, certain fields had actually become less collaborative and more protective of ideas, with traditional channels of dissemination still considered essential for the recognition of research contributions.

National differences can also be observed. For instance, Nicholas et al. (2014) found that attitudes towards social media differed between researchers from the UK and the US. Researchers in the UK had little social media involvement and did not trust it as a credible academic source, although younger researchers were more likely to consider it useful, whereas US researchers had more experience with social media and mentioned its usefulness as a way to communicate, publicize their work, and obtain ideas. Although new communication genres and tools develop, the peer-reviewed journal article remains the central and most important mean of scientific communication. Furthermore, while social media may enable research activities to take place more quickly and efficiently, there is little evidence that the research process itself has been fundamentally altered by social media adoption. Less than half of UK doctoral students surveyed in a large study agreed that social media made new kinds of research possible (British Library et al, 2012).

Still, there are indications that the way in which social media is viewed by the academe is changing. The norms of different scientific cultures may dictate how academics within each discipline use social media, but these are being challenged by the increased adoption and ubiquity of social media. Recently criticism of a much publicized genomics paper was posted to Twitter, which resulted in calls from other researchers for a response from the authors, and in some cases, a retraction. The lead author on the paper chastised the use of Twitter to discuss the concerns (Woolston, 2015). In the context of research evaluation, Piwowar (2013) points to the US National Science Foundation's decision to recognize scholars' research "products" rather than just their publications as indicative of the increasing acceptance of researchers' online and social media output. A similar decision was also reached by the UK Medical Research Council (Viney, 2013). Among researchers, however, acceptance of new types of outputs and impacts in an evaluative context has been indecisive: slightly under half of the respondents

in Procter et al. (2010) expected reader ratings and citations to be considered in tandem with peer review in the next five years.

The following provides an overview of the various social media platforms academics tend to use for research-related purposes, of how they use them, as well as of the motivations behind their usage. Studies investigating the general uptake of various social media platforms among researchers are described first, followed by analyses focussing on specific platforms. Types of specific platforms described include social bookmarking and reference managers, social networking sites, blogging, microblogging, open peer review and commenting, and Wikipedia and wikis. Note that results from specific platforms included in the general studies section are not repeated in these subsections.

General Social Media Studies

Social media is used by academics to disseminate their own research, discover others' work, and connect with other scholars in the interest of collaboration and idea generation. Faulkes (2014) likened the feedback and peer-review facilitated by social media to that at a research conference, a sentiment echoed by other academics ((British Library et al, 2012). Behaviour can be specific to the social media platform itself, including how academics choose to present themselves—either adopting a professional, personal, or mixed identity—as some social media platforms are public facing, while other are intended for specifically academic audiences. Factors such as age, gender, academic status, field, and culture have been put forth to explain the variance in the attitudes towards social media and how it is used. There have been several large-scale studies with regards to the usage of social media for research purposes. A 2011 study found that 63% of researchers used collaborative authoring tools, 27% social networking tools (almost half being Facebook), 15% blogs, and 9% microblogs (almost all being Twitter) (Rowlands, Nicholas, Russell, Canty, & Watkinson, 2011). A similar level of Twitter use were found in Procter et al (2010), whose survey also revealed that while 73% of UK academics used Google Scholar, 69% Wikipedia, 29% YouTube, and 24% Facebook, the majority never actually contributed original content to blogs, articles and wikis, nor posted slides, text, or videos publicly. This spectator-like behaviour was confirmed by a 2013 survey at UK research universities, which found that about 25% of respondents did not use social media for work, and more than 75% did not create original content on the social media platforms (Tenopir, Volentine, & King, 2013).

Along these lines, a study of the presence of highly cited European researchers on social networking sites (SNS) and reference managers found 49% of social sciences researchers, 43% in engineering, 40% in health sciences, 30% in life sciences, and 28% in physical sciences had a presence on one or more of the examined platforms, with LinkedIn being the most popular, ranging from 18% to 29% of researchers on it depending on the discipline (Mas-Bleda, Thelwall, Kousha, & Aguillo, 2014). Procter et al. (2010) found that Computer Science and Mathematics researchers were disproportionately represented as frequent users of social media tools for professional reasons. The degree of social media adoption was positively associated with males, older respondents, more senior staff, and involvement in collaborative research. However, discipline was not significantly related to the creation of social media content in Tenopir, Volentine and King (2013). Haustein, Peters, Bar-Ilan, et al. (2014) analyzed the social media presence of 57 presenters at a bibliometric conference and found that LinkedIn was the most popular platform (70%), followed by ResearchGate (58%), Academia.edu (30%), ORCID (25%), Google Scholar Citation profiles (23%) and Twitter (16%). In a follow-up study including a

survey among conference participants (Haustein, Peters, Bar-Ilan, et al., 2014), 73% used Facebook (mostly privately), 68% used LinkedIn and 44% Twitter and the social media presence of the 57 presenters had increased particularly for Google Scholar Citation profiles, Twitter and LinkedIn. About half of survey respondents used reference management software, with Mendeley being the most commonly used social bookmarking service: 26% used Mendeley, while 77% had heard about it. CiteULike had 73% awareness but only 13% usage. For this small group of bibliometrics conference attendants, gender and age did not appear to have an impact on Mendeley usage. Google Scholar was most used for publications listings and citation checking. Moreover, 86% of the conference participants indicated that altmetrics (including downloads) could be useful in the evaluation of authors and article. Bowman (2015b) asked US professors to indicate which social media platforms they used and 1,639 responded, with results showing that Facebook (70%), LinkedIn (58%), and Google+ (50%) were by far the most reported general social media platforms used by the scholars. With regards to academic-related social platforms, respondents reported a low use of Mendeley (7%), while a higher proportion reported using ResearchGate (26%) and Academia.edu (22%).

In a broad and recent survey conducted by *Nature*, 3,500 researchers indicated that having a social media profile was most useful for publicity and the dissemination of their work: 41% said it was quite useful for raising the profile of their work, 39% found it quite useful for sharing papers and datasets, and 31% found it quite useful for attracting collaborators (Van Noorden, 2014). Having a social media presence was considered less useful for attracting funding, however: 34% said it was not at all useful and 26% said their social media presence was not very useful for attracting future employers (though the other responses were quite close: 22% quite useful, 21% not at all useful, and 25% didn't know).

In general, the use by and presence of researchers on social media is increasing. Differences can be identified between different platforms and types of tools as well as between scientific disciplines. In the following, the findings regarding social media behaviour of researchers are further examined by platform.

Social Bookmarking and Reference Managers

Reference managers and bookmarking tools allow users to store and save bibliographic information and organize and annotate full texts of relevant publications for future use. Social reference managers such as Mendeley, CiteULike, BibSonomy, or Zotero have a built-in networking component to share and comment on references with others and connect with users with similar interests. However, there is yet little evidence to support the use of these social functions by academics in this way. In a large study of UK doctoral students, 66% of those in Arts and Humanities and Social Sciences used social media to manage references and research notes at least once a month (British Library et al, 2012). However, 85% of the Arts, Humanities and Social Sciences students did not use social bookmarking sites in the course of their research, and only 4% who did found them useful. A majority (87%) of 864 Mendeley users surveyed reported that they used the platform as a reference manager, mostly to include the publications they saved as citations in their own research, while only 15% indicated they used it as a social networking site (Mohammadi, Thelwall, & Kousha, In-press). Similarly, a smaller study found that most Zotero users did not make connections on the site (Jordan, 2014). There were significant differences in Mohammadi et al (In-press) between occupation and how Mendeley was used—academic staff used it more to publicize their own publications and master's students used it more to search for publications. Haustein and Larivière (2014) found that PhD students were the largest reader group on Mendeley for a sample of over a million papers in health fields, and that PhD students' reader counts were the most correlated with

citation rates. For most fields, PhD students were also the largest reader group, followed by postgraduate students and postdoctoral students (Mohammadi, Thelwall, Haustein, & Larivière, 2015). Thelwall and Maflahi (2014) found that Mendeley users were more likely to read papers written by authors from their own countries.

Social Networking Sites

Social networking sites enables users to interact with others, generate and access content, and share information. Platforms include those aimed at and used by the general public (e.g., Facebook and Google+) as well as those targeted at academics (e.g., Academia.edu and ResearchGate). Nández and Borrego (2013) found that the primary motivations behind 293 Academia.edu users were to contact other researchers (67%), disseminate research (61%), and follow the research of others (59%). This is consistent with the results in Jordan (2014) where most respondents indicated that an Academia.edu profile functioned like an online business card, and in a *Nature* survey where users of ResearchGate and Academia.edu simply maintained their profiles for contact information purposes (Van Noorden, 2014). Many of these profiles, however, contain little information, even in cases of researchers with many followers and profile views (Thelwall & Kousha, 2014a). Although ResearchGate was more popular than Academia.edu, Van Noorden (2014) found that Academia.edu was more prominent among researchers from the Social Sciences, Arts and Humanities than those from Science and Engineering. There was no significant correlation between the number of Scopus publications and documents researchers had on Academia.edu, which may indicate that active authors do not update their Academia.edu profiles with the full coverage of their papers (Thelwall & Kousha, 2014a). Despite an increase in ResearchGate uptake (i.e., 13% increase in users and 7% in uploaded publications from August to October 2013), there was no indication that the use of ResearchGate benefits researchers (Thelwall & Kousha, 2014b). Only a small portion, 27%, of UK doctoral students had used academic social networking sites in the past year with Arts, Humanities and Social Sciences students at exhibiting the highest use at 34% (British Library et al, 2012). By contrast, 43% had used general audience social networking sites such as Facebook for research, including 47% of Arts, Humanities and Social Sciences students, the highest usage among fields.

LinkedIn, a social networking site with a more general audience of professionals, exhibited similar usage patterns as the academic social networking platforms. Sixty-eight percent of researchers maintained a profile in case someone wanted to contact them about their research, which represented the most frequently given professional use of LinkedIn (the same percentage as those who used ResearchGate in the same way), while 30% had never posted work-related content (Van Noorden, 2014). Similarly, 68% of presenters at the 2010 STI Conference used LinkedIn (Haustein, Peters, Bar-Ilan, et al., 2014), although other field-specific studies found smaller uptake percentages, including 46% of urologists (Loeb et al., 2014), and 24% of geographers surveyed (Wilson & Starkweather, 2014). For Facebook use, 56% of surveyed researchers with Facebook accounts did not use Facebook professionally. For those who did use Facebook professionally, the most frequently reported use was posting work-related content (Van Noorden, 2014).

While social media can be used to collect data from subjects, there are issues concerning the ongoing consent of data collection and what limitations should be placed on it. Côté (2013) examined the ethical issues raised by political scientists' Facebook interactions with their subjects when conducting fieldwork, particularly given the risks posed to maintaining their anonymity and the dangers they may face in places where freedom of speech is

not guaranteed. Côté posited that ethics boards had not yet devised sufficient guidelines to address problems that might arise when using social media in research.

Blogging

Academics may write blog posts as a commentary or critique to published research, or use them to disseminate findings to the public (Bonetta, 2007; Puschmann, 2014). Mewburn and Thomson (2013) found that 41% of the 100 academic blogs they examined contained "academic culture critique" and 40% contained "research dissemination" content. In another survey, 21% of 215 academic health policy researchers had blogged about their research in the past year (Grande et al., 2014). However, among UK doctoral students, while 24% had passively used blogs in their research in the past year, only 16% had actively contributed to them (British Library et al, 2012). This varied across disciplines with Arts, Humanities and Social Sciences students having the highest active level of use (16%) and Engineering and Computer Science students having the highest passive level of use (30%). In a study regarding the authorship practices of 126 academic blogs, most of the authors were male, 32% had a PhD, 27% were graduate students, over half were affiliated with an academic institution, and 84% authored their blogs using their real names (Shema, Bar-Ilan, & Thelwall, 2012). As academic blogs such as those indexed by ResearchBlogging.org might be increasingly rewarded by university administrators and research councils, it makes sense that authors would choose to use their names and indicate their institutional ties.

Blogging has provided benefits for the public by overcoming the problems with science journalism and the politicization of science in the public sphere, and for scientists by offering a platform to network and communicate with other scientists (Wilkins, 2008). Mahrt and Puschmann (2014) found that most researchers who blogged did so because they enjoyed writing (79%) and wanted to "present my discipline to the public" (78%). Only 7% said that they did it to advance their academic careers, and only 2% said that blog readers sometimes left comments to indicate the desire to collaborate academically, while 36% said it rarely happened and 62% said never.

Microblogging

Microblogging platforms limit publications by character length, resulting in very short texts. These short texts or posts are diffused to users that follow the posting user, but are typically available to all users of the platform through search functionality. While the boundaries between professional and personal identity can blur on many social media platforms, this may be particularly pronounced on general audience microblogging platforms such as Twitter, due to their publicly facing nature and the ease with which one can publish quickly. Among social media platforms, microblogging platforms such as Twitter exhibit lower uptake among researchers. While 85% of surveyed researchers were aware of Twitter, only 14% were regular visitors (Van Noorden, 2014). Similar findings were obtained in a Finnish study, in which two-thirds of respondents indicated they knew about microblogging platforms but only 23% used them; the lowest knowing-to-using ratio for all examined platforms (Gu & Widén-Wulff, 2011). The same was discovered for faculty members of German universities: Pscheida, Albrecht, Herbst, Minet, and Köhler (2013) reported that 97% of respondents knew Twitter but as few as 15% used it and only 10% did so professionally. In a recent study of 1,910 US professors surveyed, Bowman (2015a, 2015b) found that 32% (613) reported having a profile on Twitter. The way in which Twitter was used varied, as respectively 29% of the 553 professors completing the survey used it either strictly personally or professionally and 42% used it for both personal and professional communications.

Some academic users choose to incorporate their professional identity into their Twitter biography. Bowman (2015b) found that the majority of tweets were personal (78%) as opposed to professional (19%) for a sample of 445 professors' accounts, and that 69% of professional tweets contained links compared with 15% of personal tweets. Higher instances of work-related use were found in Van Noorden (2014) where only 21% of researchers did not use Twitter professionally[1]. Researchers mostly used it by following research issue discussions (49%), tweeting work-related content (47%), discovering others in their field (42%), and finding paper recommendations (40%). Interestingly, the majority of these activities are passive rather than active content creation. Similarly, just over half of respondents in a study of Romanian researchers indicated they used microblogging platforms to search for information (Grosseck & Holotescu, 2011). According to a 2010 survey of UK doctoral students, both passive and active microblogging use for research was low, with 7% having used it passively in the past year and 6% actively (British Library et al, 2012). The rate of usage was higher for Arts, Humanities and Social Sciences students, with 8% passive use and 7% active use. However, the 2011 survey results from the same longitudinal study found that 11% of students posted on a microblog at least once a month, and that 18% followed someone else's microblog at least once a month.

Identifying professionally may have networking benefits, but also negative consequences. Those who were the most influential users in the network had at least 50 connections with other emergency physicians (Lulic & Kovic, 2013). In a study of 250 self-identified physicians, most of whom went by their full names, Chretien, Azar, and Kind (2011) categorized 3% of their tweets as unprofessional due to their use of profanity, sexually explicit or discriminatory content, or having the potential to violate patient privacy. Some professional organizations have adopted guidelines for social media best practices to try and avoid these types of violations from occurring (e.g., the American Urological Association (Loeb et al., 2014)).

Open Peer Review and Commenting

Peer review is a central part of the scholarly communication system as it functions as a quality control and gatekeeping mechanism. It has traditionally been closed but in the context of open science is becoming more transparent and moving online. Ford (2013) identified eight of peer review: signed, disclosed, editor-mediated, transparent, crowd-sourced, pre-publication, synchronous, and post-publication review. Recently, a few journals have experimented with open peer review and readers' comments. An early trial includes *BMJ*, where reviewers' identities are revealed to authors beginning with a study in 1999. In a randomized controlled trial the review process was analyzed regarding quality, length, time to review as well as editorial decisions. No significant difference was found regarding review quality, decision, or time to completion, but referees were more likely to decline to review (van Rooyen, Godlee, Evans, Black, & Smith, 1999), which suggests that a totally transparent form of evaluation might not be acceptable to all. *BMJ* nonetheless started to reveal the identity of reviewers soon after the study (Smith, 1999). New journals such as *PeerJ*, *eLife* and *F1000Research* have since followed.

[1] The two percentages might not comparable, however, as the unit of analysis of Bowman (2015a, 2015b) is the tweet, while the unit analyzed by Van Noorden (2014) is the researcher.

Along these lines, the post-publication peer review platform F1000Prime names experts to identify and summarize important publications to help academics identify worthwhile literature. With open, non-anonymous peer review and recommendations, the review report develops into a new, previously invisible, form of scholarly output that researchers can be credited for.

However, journals with open review still represent the exception rather than the rule. The uptake of open and post-publication peer review has not yet been systematically analyzed. Besides the platform F1000Prime, which is based on post-publication peer review by selected experts, there are also some journals and publication platforms that incorporate or allow for open peer review (for instance, *F1000Research* and *PeerJ*). A proposed supplement to peer-review is the commenting system, implemented on many blogging and journal platforms—such as that of PLOS—allowing readers to comment on published documents. However, the commenting system has not been so successful, leading some journals to remove the commenting function, either because of a lack of comments or because of their non-scientific nature. For example, *Nature* held a four-month trial phase opening articles up to online commenting, in which authors were asked if they would allow their papers to be open to technical commenting online, in addition to the regular peer review process (Anonymous, 2006). Only 5% of authors (71) of papers during this time allowed their papers to be open for comment, and 38 papers received at least one comment, for a total of 92 technical comments. In a follow-up survey with 27 of the authors, 11 stated that they preferred the open peer review system and 10 of the authors whose work received comments found them at least 'somewhat' useful. Interestingly, the commenting trial did result in larger traffic to the *Nature* website, although this did not convert into comments, and the option to open papers up for comment was subsequently abandoned.

Wikipedia and Wikis

Interviews with 51 members of the American Society for Information Science and Technology revealed that wikis, including Wikipedia, were the most commonly used social media tool for research (42% of interviewees) (Gruzd, Staves, & Wilk, 2012). In another small survey of 71 attendees of the 17th International Conference on Science and Technology Indicators, 34% of respondents indicated that Wikipedia citations of articles would be a potential alternative metric for evaluating authors or papers, and 27% said that the same about a Wikipedia article about themselves. The respondents in both of these studies may be more predisposed to positive attitudes towards social media use, given their fields of research (Haustein, Peters, Bar-Ilan, et al., 2014). In a large survey of UK PhD students, the levels of wiki usage for research and their perceived value differed among fields, with those in Arts and Humanities, Social Sciences and Medicine, Dentistry and Health using them less and finding them more limited in value than other fields (British Library et al, 2012). Only 25% of PhD students in Arts, Humanities and Social Sciences used a wiki to collaborate online while 44% of computer scientists and engineers had; students in the biological and biomedical sciences used it the least. As for its use as an source of scholarly information, a study of 59 journal articles in the field of Medical Informatics found that 22% cited articles in wikis and that 21% cited articles in Wikipedia (Rousidis, Garoufallou, & Balatsoukas, 2013). However, preserving the material as originally cited may be difficult on wiki sites where continual crowd-sourced edits to articles are a fundamental aspect of the medium.

2.2 How research findings are disseminated on social media

The dissemination of scholarly publications on social media occurs through user behaviours such as linking to an article, saving and bookmarking it, tweeting about it, and commenting on or rating it. This section focuses on the diffusion of research findings on social media, which may be performed by researchers and academic institutions, funding agencies and policy makers, lecturers and students as well as practitioners or by the general public. In fact, one of the central goals of the altmetrics movement is to measure the use of research by the general public and, it is assumed, a certain form of societal impact. The demand of accounting for the impact of research on society was expressed through the decision of the British Higher Education Funding Council's (HEFCE) to consider "all kinds of social, economic and cultural benefits and impacts beyond academia" (Higher Education Funding Council for England, 2011, p. 4). Time has been shown to affect the uptake of research on social media, and impacts each platform differently. Therefore, attention should be paid to the publication dates of both the research publications being analyzed and the studies themselves. The findings of general studies on the topic are presented first, followed by platform-specific studies following the structure of Section 2.1.

General Large-Scale Studies

In a study of 19,772 publications indexed in the Web of Science (WoS) and published between 2005 and 2011, Zahedi, Costas, and Wouters (2014) found that Mendeley provided the highest coverage, with 63% of the publications saved at least once in the reference manager. On other platforms, documents were only sparsely represented: over 98% of the publications had no presence on Wikipedia, Twitter, or Delicious. Using another dataset of 500,229 WoS-indexed publications from 2011, Costas, Zahedi, and Wouters (2014) found that 15% had at least one social media mention, with 13% being mentioned on Twitter, 2.5% on Facebook, and 1.9% on blogs. For 1.3 million WoS papers published in 2012, coverage had increased to 21.5% for Twitter and 4.7% for Facebook, while blog coverage remained stable at 1.9% (Haustein, Costas, & Larivière, 2015).. Twitter, Facebook and blog coverage differed between disciplines. Disciplines that have stronger ties to society, such as SSH, or deal with specific concerns of people's everyday lives, such as health or environmental issues, have a higher probability of appearing on social media platforms than those from more technical disciplines. A potential explanation for this might be that that the general public is more able to understand and relate to "softer" science topics than complex technical issues from Physics, as the latter generally makes use of a very formal language (Costas, Zahedi, & Wouters, 2015; Haustein, Costas, et al., 2015)(Haustein, Costas, & Larivière, 2015). Costas, Zahedi, and Wouters (2015) came to similar conclusions; they showed that papers in General Medicine, Psychology, and Social Sciences—fields that are considered to have greater social impact—were much more visible on Twitter than papers in other fields suggesting that tweets could, to a certain extent, reflect impact on the general public. Priem, Piwowar, and Hemminger (2012) analyzed 21,096 research articles published from 2003 to 2010 in *PLOS ONE* and found that 80% were bookmarked in Mendeley and 31% on CiteULike. Delicious bookmarks, tweets, Facebook shares and PLOS comments were found for 10-12% of the articles, 7.5% were mentioned in blogs or were rated in F1000, and 5% were cited in Wikipedia, or liked or commented on Facebook. In a small study of 310 journal articles from the Swedish publications index, 62% were mentioned on Google Scholar, 61% of the articles were on Mendeley, 21% on Twitter, 5% on CiteULike, 3% on Facebook, and 2% on blogs (Hammarfelt, 2014). Much lower Mendeley save rates were obtained for a newer sample of 516,150 publications indexed from 2011 to 2013 in WoS (Robinson-García, Torres-Salinas, Zahedi, & Costas, 2014). Only

12% were saved in Mendeley, while 16% were mentioned on Twitter. These lower rates are due to Altmetric.com only capturing Mendeley reader counts for documents with at least one other altmetric event and thus underestimate the actual Mendeley coverage. This also applies to Knight (2014), where 13% of 6,981 articles on solid organ transplantation were saved on Mendeley according to Atlmetric.com. A total of 19% of the articles on solid organ transplantation had at least one social media count, and a similar breakdown among platforms was found. Twitter had the most mentions (18% of articles were tweeted about at least once, and 4% more than once), 2% were mentioned on Facebook, 1% on CiteULike, 1% were on F1000, and less than 1% on blogs. This highlights an important difference amongst methods of data collection, an issue which is discussed further in Section 2.4.

Social Bookmarking and Reference Managers

Among the social reference managers and bookmarking sites, Mendeley appears to provide the most comprehensive coverage of scientific publications. Of 219,326 articles published in 2008 in Clinical Medicine, Engineering and Technology, Social Science, Physics, Chemistry and indexed in WoS, 46% had at least one reader on Mendeley (Mohammadi et al., 2015). For 27,558 Social Sciences and 1,914 Humanities articles published in 2008, Mohammadi and Thelwall (2014) found that 44% of Social Sciences articles and 13% of Humanities articles had been saved on Mendeley at least once. Psychology and Linguistics were the best-covered specialties in each discipline (54% and 34% respectively), and Education and Literature had the lowest coverage (39% and 4% respectively). Haustein, Larivière, et al. (2014) also found Psychology to be the best represented field on Mendeley, with 81% of papers published from 2010-2012 having a reader in the social reference manager, and Humanities the lowest with 40.7% coverage. It is important to note that authors make use of different discipline and field classification systems and that platforms such as Mendeley often have their own, which makes true cross-wise comparison difficult and thus should be interpreted with caution. For instance, where Psychology is a sub-field of the discipline Social Sciences in Mohammadi and Thelwall (2014), Psychology and Social Sciences are their own separate disciplines in Haustein, Larivière, et al. (2014). On the whole, these results highlight important field differences in the coverage of scientific publications on Mendeley, with Psychology obtaining higher scores, and fields within the Humanities obtaining lower scores.

Higher coverage has been found in more specialized samples. Bar-Ilan et al. (2012) examined 1,136 publications authored by the presenters at the 2010 STI Conference and found 82% of publications were saved at least once in Mendeley (mean of 9.5 users per publication), and that only 28% were bookmarked in CiteULike (mean of 2.4 users per publication). Mendeley favoured more recent articles, with 88% of documents published since 2000 being saved in the reference manager, compared to 44% of those published before 1990. Torres-Salinas, Cabezas-Clavijo, and Jiménez-Contreras (2013) found similar patterns of coverage when they compared Mendeley, CiteULike, and Twitter counts for the 10 most cited communication articles in WoS from 2010-2012. Mendeley had the highest coverage, with 57-62% of the articles saved on the platform compared to 23-30% for CiteULike (Twitter counts for both cited and uncited articles were low). Articles published in the high-ranking journals *Nature* and *Science* since 2007 were comprehensively covered in Mendeley (94% and 93% respectively), and to a lesser extent by CiteULike (62% and 59%)(Li, Thelwall, & Giustini, 2012). Mendeley also contained the largest number of publications by authors who also wrote blogs on Scienceblogs.com and Scienceblogs.de (53% for .com 42% for .de authors) when compared with BibSonomy and CiteULike (Weller & Peters, 2012). It is interesting to note that while Mendeley coverage of 1.4 million journal papers indexed in WoS and PubMed published from 2010-2012 was 66%, the coverage slightly decreased with each publication year, from 70% in

2010 to 57% in 2012 (Haustein, Larivière, Thelwall, et al., 2014) suggesting that social reference managers behave in a manner similar to citations, for which one has to wait a few years to obtain a signal. However, the lag on Mendeley seems shorter than for citations (Thelwall & Sud, 2015).

In an earlier study, which investigated the presence of physics papers indexed in WoS and published between 2004 and 2008 on the social bookmarking sites BibSonomy, Connotea and CiteULike, Haustein and Siebenlist (2011) found that CiteULike had the most complete coverage for journals in the field (78% of bookmarks and 84% of users). The frequency of unique tags assigned by bookmarkers appeared to follow a power-law distribution, with a few tags occurring many times and many tags only being used once. Good, Tennis, and Wilkinson (2009) analyzed the tagging for 19,118 PubMed articles and found CiteULike had 45,525 'tagging events' and Connotea only 28,236. Most of the PubMed publications appearing on CiteULike and Connotea were only tagged by a few people, and there were low levels of tag usage agreement among users.

Social Networking Sites

Social networking sites for academics—such as Academia.edu or ResearchGate—encourage researchers to upload their publications to their profiles. Academics in philosophy, for instance, were found to have a mean of 1.05 papers listed on their Academia.edu profiles, while students had a mean of 0.23 (Thelwall & Kousha, 2014a). Faculty members had their uploaded publications viewed more (median 9.12, mean 384.39) than users who were students (median 0, mean 103.78). Another study found only a small number of Academia.edu users amongst 1,515 highly cited European scientists, half of whom had uploaded at least one publication to their profile (Mas-Bleda et al., 2014). Of those users, approximately half had at least 250 document views.

Thelwall and Kousha (Thelwall & Kousha, 2014b) looked at the publications uploaded on ResearchGate on an institution and country level and compared them with publications in WoS. The lowest ratio of ResearchGate publications to WoS publications was 0.01 and the highest was 1.9, with a mean of 0.9. Looking on a per-country basis, the United States had the largest ratio with over 14 ResearchGate documents per WoS document, an indication that many American academics upload their papers there, followed by Japan, Sweden, and then Canada, while China had the fewest. The authors did find an increase of 7% in the total number of publications on ResearchGate in just a two-month period, and thus the platform may become a more comprehensive depository of research in the future, especially for medical research. On the whole, these results suggest that the usage of such platforms varies significantly by country.

Blogging

Although it is one of the oldest social media forums, there is some evidence that blogging rates may be in decline, at least for academics. Fausto et al. (2012) found that the number of posts on ResearchBlogging.org were highest in 2010 compared to 2011 and 2009, and similarly Shema, Bar-Ilan and Thelwall (2015) found the number of posts decreasing from 2010 to 2011 and further from 2011 to 2012.

Fausto et al. (2012) investigated the relationship between posts from blogs indexed in ResearchBlogging.org and the scientific journal articles they link to for 19,000 blog posts published between 2007-2011. The posts linked to 26,154 articles in 3,350 journals, with the journals *Science*, *Nature*, *PNAS,* and *PLOS ONE* being cited more than 1,000 times. Biology was the most common discipline for blog posts (36%) and Health Sciences the second

(15%). Open access journals were disproportionately represented, with 11.7% of the articles cited in blogs coming from them despite representing only 7.2% of the total journals covered, a statistically significant difference. Shema and Bar-Ilan (2014) looked at over 10,000 blog posts also on ResearchBlogging.org from 2009 and 2010 and confirmed bloggers' preference for the biological and medical sciences, with 67% of referenced articles in 2009 and 74% in 2010 coming from journals in the two subject areas. Highly cited journals also had high levels of blog coverage—in 2009, *Psychological Science* had 21% of its articles reviewed in a blog, *Science* had 18%, *Nature* 14% and *PLOS Biology* 13%. In 2010, the *New England Journal of Medicine* (*NEJM*) had 31% of articles reviewed in a blog, *Psychological Science* had 25%, *Nature* had 23% and *Science* had 20%. Similar results were found in Shema, Bar-Ilan and Thelwall (2015) where the journals cited in blog posts on health were mostly highly cited multidisciplinary or general and internal medicine journals, and in Groth and Gurney (2010) where 71% of the publications referenced in 295 chemistry blog posts came from the top 20 high-impact chemistry journals. Costas et al. (2014, 2015) showed that papers published in multidisciplinary high impact journals such as *Science*, *Nature* and *PNAS* were overrepresented among papers mentioned in blogs.

The preference for multidisciplinary journals appears to be shared by blogs and mainstream news (Costas et al., 2015). Shema and Bar-Ilan (2014) found that 81% of *NEJM*'s articles which were blogged about in 2009 and 53% in 2010 were also covered by mainstream media outlets. Although the news articles were typically published before the blog posts, this indicates similar interests among both journalists and science bloggers. Content-wise, Shema, Bar-Ilan and Thelwall (2015) found that health blog posts citing research were often explanatory in nature, with 44% "discussing factors which influence a health condition/life style", 28% "discussing social phenomena" and 23% "discussing a public health issue". Thirty-percent of the blog posts were critiques of the research itself, including concerns about methodology and conclusions.

Microblogging

The coverage of academic publications on Twitter varies among disciplines. Examining 1.4 million journal papers published from 2010-2012 indexed in both PubMed and WoS, Haustein, Peters, Sugimoto, Thelwall, and Larivière (2014) and Haustein, Larivière, Thelwall, Amyot, and Peters (2014) found that 9% had been tweeted about, and professional fields (e.g. law, business, education) had the highest article coverage on Twitter with 17% of papers receiving at least one tweet, while physics had the lowest with 1.8%. For those papers which were on Twitter, 63% were mentioned only once. The authors found that for this set of health-related papers, Twitter coverage increased each year from 2010 to 2012. This could be due to the relative newness of Twitter as a platform and should be a potential consideration for data collection and comparison purposes. The tendency of Twitter to favour newer publications was confirmed in Eysenbach (2011), who found that 44% of tweets linking to academic articles published in the Journal of Medical Internet Research occurred on the day the article was published, 16% the next day, and 6% the day after that.

Haustein, Bowman, Macaluso, Sugimoto, and Larivière (2014) investigated the difference in Twitter activity between articles published in journals and their corresponding e-print on arXiv. Putting a preprint on arXiv before the article is officially published in an academic journal is common in Physics, Mathematics, and Computer Science. Of the original arXiv documents, 39% of 84,374 documents submitted to arXiv in 2012 in the fields of physics, mathematics, computer science, statistics, quantitative biology, and quantitative finance received at least one tweet. The percentage of tweeted papers varied among subject fields, from 26% (Statistics) to 50% (Quantitative Finance). The authors discovered that the high Twitter coverage in some fields is caused by

automated Twitter accounts, which automatically publish the arXiv RSS feed on Twitter. These arXiv bots are closer examined in Haustein, Bowman, et al. (2015), identifying 43 platform feeds that automatically distributed submissions to an arXiv section or subsection. The authors warn against potential gaming and its influence on the evaluation of science.

Looking at the way academics use Twitter to cite external sources, an analysis of the URLs contained in 2,322 tweets sent by 28 academics found 6% of the tweets contained links to peer-reviewed scholarly articles, 52% of which were direct links and 48% indirect links (i.e. second order events, such as links to a news item or post about the articles, rather than directly to the articles themselves) (Priem & Costello, 2010). The motivation for tweeting indirect links was identified as ease of use and a way to circumvent paywalls. Accordingly, over half of the direct links were to open access publications, while a quarter of indirect links were to open access articles, a difference found to be statistically significant. Holmberg & Thelwall (2014) similarly found that scholarly tweets linked less to actual articles and more to blogs or items which then linked to the articles. Thelwall, Tsou, Weingart, Holmberg, and Haustein (2013) analyzed 270 tweets linking to academic articles in digital libraries and large multidisciplinary journals and discovered that many of the tweets referencing the journal *Science* were modified retweets of those coming from the journal's Twitter account, and that tweets referencing *Nature* were largely in the format provided by the share button found on each article on the journal's website.

Open Peer Review and Commenting

Adie (2009) analyzed *PLOS ONE* comments and compared them to those on BioMedCentral. While 39% of *PLOS ONE* articles were found to have comments, that percentage dropped to 18% if limited to only comments from readers or authors (excluding *PLOS ONE* editors), which means that the vast majority of papers do not receive comments. Of the 2,389 comments, authors contributed 21% and editorial staff contributed 48%. Seventeen percent of the author/user comments are discussions about the results and interpretations, 11% requests for clarification, 13% criticisms (flaws or errors).

F1000 had 132,662 recommended publications in January 2013 (Waltman & Costas, 2014), and by January 2014 it had increased to 149,227 publications (Bornmann, 2014c). Waltman and Costas (2014) found that the majority of included publications had only one recommendation (81%), and less than 3% had more than three recommendations. Fifty-nine percent of recommended papers had a score of "Good", 35% "Very Good", and 7% "Exceptional". Similar distributions of scores were found by Bornmann (2014c) for papers published in PLOS with F1000 recommendations. Only 3% of papers published by PLOS had F1000 recommendations.

Wikipedia and Wikis

Evans and Krauthammer (2011) found 0.5% of all Wikipedia entries cited a PubMed Central article, for a total of 161,155 PubMed articles cited on Wikipedia, representing approximately 0.1% of all PubMed articles. Just over half of the Wikipedia-cited articles were published between 2001 and 2010, with open access journals accounting for 2.8% of the articles. One-fifth of the articles were referenced by more than one entry on Wikipedia. Higher levels of coverage was found for 110,129 articles published in the open access PLOS journals, with 4% of PLOS articles found on Wikipedia, and 47% of the references in non-English Wikipedia articles (Lin & Fenner, 2014).

2.3 How social media metrics compare with traditional metrics

The activity on social media platforms associated with scholarly content is used as the basis for social media metrics. This activity is seen as a series of indicators that are used to measure the use and visibility of this content on social media. This includes, for instance, the number of tweets linking to a paper on Twitter, the number of blog posts or Wikipedia entries citing an article, the saving of a publication by Mendeley users, or liking a Facebook post regarding research. As described above, this set of metrics is in constant flux adapting to changes caused by uptake and use by scholars, changes in social media platform interfaces, as well as technical possibilities (particularly the availability of APIs). Social media metrics are based on the measurement of online social media activity related to scholars or scholarly content and, as such, can be considered as a subset of webometrics and scientometrics.

A variety of tools have been developed to measure social media impact, either by creating their own impact scores or aggregating other measures. Some of the most commonly referred to tools include Altmetric.com, Plum Analytics, Impact Story (formerly Total-Impact), and PLOS Article-Level Metrics. For additional overviews of the various measurement tools along with their benefits and weaknesses, see, for example, Bornmann (2014a), Brigham (2014), Das and Mishra (2014), Kwok (2012), Neylon and Wu (2009), Galligan and Dyas-Correia (2013), Galloway, Pease and Rauh (2013), Priem and Hemminger (Priem & Hemminger, 2010), and Sweet (Sweet, 2014).

In order to better understand social media metrics, many researchers compared them with traditional bibliometric impact measures; such as the number of citations a paper receives. This is often done by determining statistical association and the ability for one set of metrics to predict the other. Due to the varied nature of the scope of each study, methodology involved, and the relative newness of social media platforms, whose adoption rates vary between years, it is difficult to draw definitive conclusions regarding the social media metrics' relationship with traditional bibliometric measures. Often similar types of metrics correlate better with each other than with citations. For instance, factors which affect how traditional citations are calculated, such as normalization, may not pose the same problem for social media counts. Bornmann (2014b) concluded that normalization could take place on the level of topic rather than journal subject, as certain topics are of more general public interest (e.g. evolution), but field normalizations for ResearchGate metrics were deemed essential by Thelwall and Kousha (2014b).

After providing a brief overview of data providers and aggregators collecting various altmetrics, this section summarizes the findings of studies that focus on the analysis and comparison of social media metrics. We focus on general studies first which include different social media metrics, followed by platform-specific studies under their own subsections. These include: social bookmarking and reference managers, social networking sites, blogging, microblogging, open peer review and commenting, and Wikipedia and wikis. Note that results from specific platforms included in the general studies section are not repeated in these subsections.

General Social Media Studies

Most studies that compare citations with different social media metrics find that correlations are positive but very low, with the majority of papers not being mentioned on various social media platforms. The social reference manager Mendeley is an exception with moderate to strong positive correlations with citations.

In a recent comprehensive analysis examining 1.3 million WoS-indexed papers published in 2012 and social media counts from Altmetric.com, Haustein, Costas, and Larivière (2015) examined the main patterns of five social media metrics as a function of document characteristics (i.e., discipline, document type, title length, number of pages and references) and collaborative practices. Findings showed that both citations and social media metrics increase with the extent of collaboration and the length of the references list. On the other hand, while editorials and news items are seldom cited, these types of documents are the most popular on Twitter. Similarly, while longer papers typically attracted more citations, an opposite trend was seen on social media platforms. While only one-fifth of 2012 papers had received tweets, two-thirds had already been cited by the end of 2013, which contradicts the belief that social media metrics could overcome citation delay. On the whole, these findings suggest that factors driving social media events and citations are different (Haustein, Costas, et al., 2015).

Focusing on the subject areas of half a million papers published in 2011, (Costas et al., 2015) visualized a map of science with topics that were popular on Twitter, Mendeley, Facebook, blogs, and mainstream news. Surprisingly, in most disciplines, readership counts exceeded citation rates; this was especially true for the Social Sciences. They concluded that in the Social Sciences (much more than in the Humanities and Natural Sciences), where the use of citations is more problematic, readership counts could be used as an alternative to citations as a marker of scientific impact.

Thelwall, Haustein, Larivière, and Sugimoto (2013) compared citations for over 300,000 articles indexed in PubMed with their social media counts compiled by Altmetric.com. Positive significant (but small) associations were found between citations and Nature Research Highlights, mainstream media mentions, Facebook posts, and blogs, while Twitter had a negative significant correlation, leading the authors to conclude that the discrepancy between the quick uptake of social media and the delay before citations start accruing demonstrates the biases of using correlation coefficients.

In a similar large-scale study covering 500,229 documents from 2011 indexed in WoS, the traditional bibliometric average journal citation score (JCS) was better at identifying the top 1% most cited publications than social media counts, but the social media metrics outperformed the JCS at the 5% recall mark (Costas et al., 2014). However, over half of the most-cited publications had no associated social media metrics scores. A small study of 10 drug development research articles also showed that the Journal Impact Factors were not related to social media metrics (Altmetric.com scores, Twitter activity, Mendeley readers) (Huang, Davies, Joseph, & Wadyka, 2014). Among social media counts, citations correlated moderately with reference manager saves for 21,096 research articles published from 2003 to 2010 in PLOS. Interestingly, the citation counts from 2011 correlated better than 2010 citation counts with all social media metrics, possibly indicating an increase in the uptake of social media for the purposes of disseminating research (Priem et al., 2012).

Social Bookmarking and Reference Managers

In general, Mendeley fares better than most other social media in correlating with citations, usually with moderate strength. Contrary to other social media metrics like tweets and blog posts, the average number of Mendeley readers often exceeds the number of citations for recent papers. Among papers recommended on F1000, CiteULike and Mendeley user counts, Mendeley had the highest correlation with citations and journal impact factors (JIF) for 1,397 highly rated articles published in the field of genomics and genetics (Li & Thelwall,

2012). In a comparison of Mendeley and Twitter counts for 1.4 million journal papers published from 2010 to 2012 indexed in PubMed and WoS, the correlation between documents with at least one Mendeley reader and their citations was moderate, while the correlation between documents with at least one tweet and their citations was much lower, and the correlation between the number of Mendeley readers and the number tweets for a document was also low (Haustein, Larivière, Thelwall, et al., 2014). For the Social Sciences and Humanities, Mohammadi and Thelwall (2014) found that correlation between Mendeley readership and citations was moderate. However, when including articles not found in Mendeley, the correlation dropped as many frequently cited articles were not saved to Mendeley. The median number of citations for Social Science papers was 2 for articles not in Mendeley and 5 for those which were; for Humanities it was 0 for those not in Mendeley and 1 for those which were.

In a small case study, Bar-Ilan et al. (2012) found a similar moderate correlation between Mendeley and citations, with a lower correlation between CiteULike and citations for the work of 57 researchers who presented at the STI 2010. Mendeley had the highest correlation with citations and BibSonomy the lowest, for 678 publications by researchers who also blogged (Weller & Peters, 2012), and for all papers published in Nature and Science in 2007, moderate correlations were found between Mendeley saves and citations, higher than the correlations found for CiteULike, although CiteULike and Mendeley counts had the highest correlation (Li et al., 2012). CiteULike bookmarks were also found to have the largest correlation with Mendeley saves in Li and Thelwall (2012). Haustein and Larivière's (2014) study of 1.2 million documents in PubMed and WoS, Clinical Medicine, Health, and Psychology also found a moderate correlation between the number of Mendeley readers and citation rate, with three out of the four disciplines showing correlations above the average for the entire sample. Sud and Thelwall (2014) looked at 9,608 Biochemistry articles published in 2011 and found citations and Mendeley reader count exhibited moderate correlation, higher than those for the authorship factors also examined in the study. They found that the number of countries involved in a publication does not matter, but the specific countries involved do (i.e. some countries have higher impact for readership and citations than others). Collaboration with researchers from the USA may increase the impact of publications from international authors. Thelwall and Maflahi (2014) reported that international collaboration may have a positive effect on Mendeley readership, by drawing a larger audience. Similarly, moderate correlation was found between reader counts and citations, and readers who were PhD students and post-doctorate students had the largest effect size of all reader types on citations (Mohammadi et al., 2015).

However, comparing Mendeley with the traditional measures article downloads and citations for two information systems journals, Schlögl, Gorraiz, Gumpenberger, Jack, and Kraker (2014) found that Mendeley readership and citations exhibited the lowest correlation for articles published from 2002 to 2011. The relative newness of Mendeley (introduced in 2009) was put forth as a potential explanatory factor for the low correlation. Additionally, the majority of Mendeley readers for the two journals were students, with a possible explanation that younger readers preferred more recent articles. Similar findings are presented in the authors' earlier publication (Schlögl, Gorraiz, Gumpenberger, Jack, & Kraker, 2013). Similarly, Bar-Ilan (2013) analyzed 100 astrophysicists' publications and found only a weak correlation between citations in Scopus and Mendeley readership.

The research on the other social bookmarking sites is less clear on the nature of their relationship with citations. Haustein and Siebenlist (2011) compared citations with bookmark metrics from BibSonomy, Connotea and CiteULike. On the level of journals, they showed that Journal Impact Factors had a moderate correlation with the

usage ratio of journals (i.e., number of bookmarked articles divided by total articles published), but no correlation with the number of users on the bookmarking sites. There was also no correlation between the number of tags used by bookmarking users on the same platforms and the number of citations received by 724 documents in physics journals in Haustein and Peters (2012). Finally, in a study of 33,128 articles from PLOS that did not include Mendeley, bookmarking in CiteULike was the most highly correlated of the social media measures, although it exhibited only a low correlation with PDF downloads and citation counts (Liu, Xu, Wu, Chen, & Guo, 2013).

Social Networking Sites

A sample of 250 faculty members on Academia.edu found no statistically significant correlation between the users' popularity on the site, as measured by the number of times their profile was viewed, and the number of times their uploaded documents were viewed, and the number of citations they received in Scopus or their H-index (Thelwall & Kousha, 2014a). For 74 higher education institutions on ResearchGate, Thelwall and Kousha (2014b) attempted to determine if ResearchGate metrics correlated with rankings of academic institutions. The ResearchGate Score (based on users' reputations and interactions on the platform), the number of document downloads and profile views had low correlations with the institution ratings lists. Correlations decreased when these numbers were normalized by the number of ResearchGate users per institution, and the authors concluded that field and size normalizations are essential for social media metrics. However, they found that the lists of rankings had the strongest correlations with one another.

Blogging

Shema and Bar-Ilan (2014) examined the relationship between the articles referenced in blog posts and their citations, using 10,130 blog posts from 2009 to 2010 indexed on ResearchBlogging.org. Comparing the median citation rate of journal articles mentioned in a blog to that of those not blogged about, citation impact was higher for the blogged articles for most of the journals; however, these differences were only statistically significant for seven out of 12 journals. Similarly, Fausto and colleagues (2012) investigated the relationship between 19,000 posts from blogs indexed in ResearchBlogging.org and the 26,154 articles to which they linked. The relationship between the citation-based journal metrics for the blogged article was compared with blog metrics such as the number of times a blog post was viewed or the article was cited in a blog). Correlation was higher between the different citation measures than it was between citation and the blog metrics (Fausto et al., 2012). A high Impact Factor did not translate into being more frequently cited by blogs. However, overlap between the articles mentioned in blog posts and those found in high-impact journals has been found in several studies (Costas et al., 2014, 2015; Groth & Gurney, 2010; Shema & Bar-Ilan, 2014; Shema et al., 2015). Blogs also showed higher precision than Twitter in identifying highly-cited publications in Costas et al. (2014, 2015).

Allen, Stanton, Di Pietro, and Moseley (2013) attempted to measure the impact that blogging about journal articles had on the articles' web page views and PDF downloads. No significant relationship was found between the social media interactions (people who had liked, shared, or otherwise engaged with the blog promotion on social media) and the increase in web page views or PDF downloads. However, this study was based on as few as 16 documents. There was also no correlation between citations received a year later and social media counts. As with other studies, the only significant association found was between the web page views of the blog post and PDF downloads of the article.

Microblogging

Early findings showed a significant association between highly tweeted articles and highly cited articles, with highly tweeted articles identifying 75% of highly cited articles and non-highly tweeted articles referring to 93% of non-highly cited articles (Eysenbach, 2011). However, these findings were based on as few as 55 articles published in the Journal of Medical Internet Research, which itself has a strong Twitter presence. Subsequent studies captured more modest associations between tweets and citations. Shuai, Pepe, and Bollen (2012) analyzed a sample of 1,710 articles submitted to arXiv and linked to on Twitter. Twitter mentions were better correlated with citations (moderate correlation) than arXiv downloads. Excluding the two most tweeted articles lowered the correlation, however. Haustein, Peters, Sugimoto, Thelwall, and Larivière (2014) investigated the presence of biomedical papers on Twitter and analyzed the Twitter citation rate (i.e., mean number of tweets per tweeted article) and the coverage rate for journals on Twitter (i.e., percentage of articles tweeted) along with traditional journal metrics, including impact factors, immediacy indexes, Eigenfactor and article influence scores. Small positive significant correlations were found for all of them, and correlation was slightly higher with Twitter coverage than Twitter citation rate. Correlations between the number of tweets and citations on the article level were generally low but positive for a majority of specialties within biomedicine; however, only 26 of the 61 correlations at the specialty level were significant. The most frequently tweeted documents had humorous or unusual content or wider health implications, rather reflecting the general public audience of Twitter more than their academic impact.

Eysenbach (2011) also developed several metrics specifically for determining the impact of Twitter citations, including Twimpact Factor (total number of tweets accumulated in the 7 days following publication), Tweeted Half-Life (day when half of all tweets received by an article have occurred), and Twindex (rank percentile of an article in comparison with other articles ranked by twimpact factor on a scale of 0 to 100). However, these measures have not yet caught on to an extent which would make them useful and further research confirming their efficacy is needed.

Open Peer Review and Commenting

In an editorial in *Nature Neuroscience,* the quantitative analysis of F1000 reviews revealed that 11 journals were the source of two-thirds of the recommended neuroscience papers (Anonymous, 2005). The F1000 ratings and Journal Impact Factor for these journals were highly correlated, indicating that this new alternative to peer review yields similar results to the old system. Results from Waltman and Costas (2014) point to a different conclusion, however. The authors compared F1000 ratings with citations and journal citation scores (JCS) for 38,369 articles in WoS published between 2006 and 2009. All documents (1.7 million) published during the same time period were used for comparison. The majority of included publications had only one recommendation (81%), and less than 3% had more than three recommendations. The more recommendations a publication had on F1000, the more citations, although increasing numbers of recommendations had diminishing returns on citations for those with four or more recommendations. The highest recommendation score a paper received had low correlation with JCS, and its correlation with citations was even lower. This was the same for the number of recommendations. JCS was a better predictor of publications with the most citations than F1000 recommendations. Again, a lack of correlation with citations is not necessarily problematic; the authors point out that this may be evidence that F1000 recommendations have a different type of impact.

Examining whether F1000's impact correlated with other social media platforms, Li and Thelwall (2012) compared F1000 article factors, Mendeley and CiteULike usage, Journal Impact Factors, and citations for 1,397 highly rated articles published in 2008 in the field of Genomics and Genetics. Citation counts had a low correlation with F1000 ratings and the number of evaluators. Mohammadi and Thelwall (2013) also found a low correlation between F1000 article ratings and the number of citations received for a random sample of 344 articles from 2007 and 533 articles from 2008 rated in F1000 medicine. Differences among F1000 labels were not found to be significant with citations or F1000 ratings, except for papers labelled "Changes to clinical practice", which showed a significant difference for F1000 scores but not for citation rates, and "New Findings", which showed a significant difference in the median number of citations those papers received. Bornmann (2014c) also investigated the relationship between 1,082 papers' labels in F1000 and their shares on Facebook, Figshare, Mendeley, and Twitter, and found a few predictive associations. The tag "Good for teaching" was found to be statistically significant in the Twitter and Facebook models, meaning that the predicted number of counts was higher for papers with that label on these social media platforms. The "Technical Advance" label was significant in the Mendeley model, meaning papers with that label should receive a higher number of reader counts in Mendeley, and "New Finding" was significant in the Facebook model.

Bornmann and Leydesdorff (2013) examined 125 papers published in 2008 in cell biology and immunology with at least one German author. They compared F1000 ratings with bibliometric measures. Percentile ranks for citations in subject area and F1000 score had the highest correlation, and a medium effect size correlation was also found for the number of citations and the subject area normalized citation rate, although the results are limited by the small and not random sample of this study.

Wikipedia and Wikis

Citations in Wikipedia articles are compared with citations in traditional publishing platforms, such as journals, as a way of assessing the usefulness of Wikipedia as an information source and investigating whether its citation patterns are similar to academic publishing. Studies have investigated the extent to which academic content is cited in Wikipedia articles long before the altmetrics movement. Nielsen (2007) analyzed 30,368 journal citations in Wikipedia against academic citations and concluded that such citations in Wikipedia will continue to grow, making it more valuable as background reading. The total number of each journal's Wikipedia citations was highly correlated with the total citations for each journal, but there was less correlation with a journal's Impact Factor. High-impact journals, such as *Nature*, *Science,* and *NEJM,* were slightly over-cited in Wikipedia. The fact that reference managers like Zotero incorporate a Wikipedia citation function highlights its increasing legitimacy (F. A. Nielsen, 2007). Along these lines, Evans and Krauthammer (2011) analyzed 4,905 PubMed Central journal articles referenced in Wikipedia entries and F1000 and found that citation counts and F1000 scores were higher for the Wikipedia-cited articles than for a control sample of articles.

2.4 Limitations of Social Media Metrics

Social media metrics are often said to have the potential to make scholarly evaluation fairer and more inclusive, so that various types of research products and broader types of impact are considered, including impact on society. However, such claims have yet to be backed with empirical evidence, and many studies have actually highlighted the limitations of such metrics.

First, the behaviour of users, as well as the representativeness of such users among the entire scientific community, is not well understood yet. While traditional bibliometrics, which social media metrics are supposed to complement, are based on publishing and citing, which are the fundamental elements of scholarly communication, social media metrics are derived from events on new and continuously changing platforms whose use and user communities are not entirely understood. Moreover, metrics can only capture use which leaves traces, excluding "invisible" activities. As described in Section 2.2, a considerable amount of social media use is passive. While social rules and norms exist within the scientific community of how, when, and what to cite, these norms are currently lacking with regard to social media as the ecosystem is constantly changing (Haustein, Bowman, & Costas, In-press). Making things even more complex is the fact that a certain proportion of social media activity might not actually be performed by humans; Haustein et al. (2015) found that 9% of tweets related to arXiv articles in 2012 came from automated accounts ("bots").

Second, social media metrics are based on different platforms with different levels of user engagement but often they are considered as one type of impact. Taraborelli (2008) wrote that, while social media metrics would not replace traditional ones such as peer review, they had certain advantages including being better at signifying intended usage behaviour for publications. On the other hand, some aggregators categorize social media events based on what they *hope* to measure rather than what is *actually* measured. For example, ImpactStory categorizes HTML views as views by the public, while PDF downloads are considered as being made by scholars (Haustein, Sugimoto, et al., 2015). Similarly, the platform considers tweets as coming from the general public, although—except in the cases of humorous tweets—the majority of tweets associated with scientific papers are likely to be made by researchers (Tsou, Bowman, Ghazinejad, & Sugimoto, 2015). Moreover, it is difficult to determine audiences on Twitter, given the sparse information available via account descriptions For instance, Desai, Patwardhan, and Coore (2014) found that 58% of the 132 Twitter users in their sample could not be categorized, given the insufficient and ambiguous information found in their Twitter profiles. Along these lines, the imprecision of the concepts measured by social media metrics has led to various issues of data interpretation. Are tweets indicators of scientific impact—as some authors have found them to be correlated with citations—or are they indicators of social impact—as funding agencies would like demonstrate. Dinsmore, Allen, and Dolby (2014) posited that much additional work is needed on the subject to fully understand and create meaningful, consistent metric scores to aid funders and determine the full impact of scientific research.

Third, important issues of data collection have been raised in the literature. The majority of currently collected altmetrics rely on the presence of certain identifiers, above all the DOI. Although the uptake of this unique identifier is increasing, the percentage of DOIs is particularly low among journal articles from the Social Sciences and Humanities (Haustein, Costas, et al., 2015). The focus on DOIs also excludes many types of documents from the outset. Although the scholarly journal article remains the most important type of formal scholarly publication, it seems to be less discussed on social media platforms such as Twitter, where only a small fraction of tweets by academics mention scholarly papers (e.g., Holmberg and Thelwall (2014) and Priem and Costello (2010); see also Section 3.3.3 for the tweet topic analysis of SSHRC Doctoral Award recipients). Moreover, different methods regardomg how social media metrics are collected by aggregators can result in different social media counts—as shown by Torres-Salinas, Cabezas-Clavijo and Jiménez-Contreras (2013) for counts obtained by Altmetric.com and ImpactStory as well as Zahedi, Fenner and Costas (2014) for a comparison of Altmetric.com and PLOS metrics. For example, as described in Section 2.2, Mendeley coverage is underestimated if collected through Altmetric.com (Knight, 2014; Robinson-García et al., 2014). As a consequence, data collection methods may affect outcomes if certain social media counts are more difficult to obtain, or if they are underestimated using certain

methods (Chamberlain, 2013; Zahedi, Bowman, & Haustein, 2014). In Sud and Thelwall (2013), the authors discussed methodological concerns when conducting research on the correlation between social media metrics and citations, and provide a good reference base for methodological techniques used in the field. Along these lines, Wouters and Costas (2012) identified the lack of ease of downloading and managing data among most tools as a concern. However, as the usage of these platforms increase and reaches a critical mass, it is quite likely that ways to obtain the data would improve as well.

3 SSHRC Doctoral Award Recipients on Twitter

This section aims to better understand the Twitter use of SSHRC Doctoral Award recipients from the competition year 2010, who were identified using the SSHRC Awards Search Engine available online[2]. A total of 509 doctoral students from 25 main fields in the Social Science and Humanities received the SSHRC Doctoral Award in the fiscal year 2010-2011, 393 of which were enrolled at 34 Canadian universities, while 116 were not affiliated with a university in Canada. For approximately one third (31.4%, Table 1) of the SSHRC Doctoral Award recipients, a Twitter account could be identified based on the available information (see Appendix A.2.1) for detailed methods to identify Twitter accounts). Based on the Twitter account information, the Twitter uptake and self-presentation of these 160 award recipients were analyzed (Table 1).

The study assessed how many SSHRC Doctoral Award recipients were present on the microblogging platform, how many people they interacted with, and whether they identified themselves as academics on Twitter. Twitter uptake and self-presentation were analyzed in Section 3.1, which was based on 85,185 of 175,085 tweets accessible and collected via the Twitter API. In addition, the tweeting behaviour and Twitter use of the recipients were analyzed in Section 3.2. Section 3.3 focuses on the four fields with the highest number of SSHRC Doctoral Award recipients: History, Modern Languages and Literature, Education and Political Science. Based on tweets published during the last 12 months (before data collection in February 2015), the tweet content for award recipients from these four fields was examined in detail by analyzing hashtags (Section 3.3.1) and URLs (Section 3.3.2) as well as through manual coding (Section 3.3.3) of a representative sample of 2,079 original tweets (see Appendix A.2.6 for a detailed description of the coding scheme). Results of the analysis are described and interpreted, including findings from previous studies as discussed in the literature review (Section 2). Results were tested for statistical significance by field, the results of which can be found in Appendix A.2.3.

3.1 Twitter Uptake and Self-Presentation

As shown in Table 1, 160 (31.4%) of the 509 SSHRC Doctoral Award recipients funded in 2010 and 2011 were identified to have a Twitter account. Overall, award recipients from the Humanities (34.4%) were slightly more present on Twitter than those from the Social Science fields (29.7%). However, differences in terms of Twitter uptake were much larger among fields within domains. Note that fields highlighted in grey in Table 1 are based on fields with less than 10 applicants and should thus be interpreted with care. Focusing on those fields with at least 10 successful applicants, only 1 (9.1%) out of 11 PhD students in Law had a Twitter account, while 53.3% of the 15 Communications and Media Studies applicants were on Twitter. History (21 applicants with Twitter account), Modern Languages and Literature (21), Education (15), and Political Science (14) were the fields with the largest absolute number of SSHRC-funded PhD students on Twitter.

Comparing the overall result of 31% of doctoral students on Twitter with findings of previous studies, the Twitter uptake of the 2010 SSHRC Doctoral Award recipients is comparably high, although similar to the percentage that

[2] http://www.outil.ost.uqam.ca/CRSH/RechProj.aspx

(Bowman, 2015b) found for US professors. Although results of previous studies varied by field, the proportion of academics using Twitter was generally around 10% (e.g. see Section 2.1 for a summary). The high percentage of SSHRC Doctoral Award recipients might be caused by several factors such as disciplinary and age differences, as well as the fact that Twitter uptake is generally increasing. Moreover, not all academics with a Twitter profile use the microblogging platform in a professional context (Pscheida et al., 2013; Tenopir et al., 2013).

Table 1. Number of SSHRC Doctoral Award recipients on Twitter and self-presentation based on Twitter account description per domain and field.

Domain and main field	Number of recipients			accounts w/ description								
	n	w/ Twitter account		n	%	self presentation based on account description						
		n	%			academic	non-academic	both	topic of PhD	discipline	academic title	university
all fields	*509*	*160*	*31.4%*	*114*	*71.3%*	*46.5%*	*8.8%*	*44.7%*	*31.6%*	*73.7%*	*64.0%*	*49.1%*
Humanities	186	64	34.4%	48	75.0%	45.8%	6.3%	47.9%	25.0%	85.4%	62.5%	45.8%
Archaeology	10	2	20.0%	1	50.0%	0.0%	0.0%	100.0%	0.0%	100.0%	0.0%	0.0%
Classics, Classical & Dead Languages	7	1	14.3%	0	0.0%	*n/a*	*n/a*	*n/a*	*n/a*	*n/a*	*n/a*	*n/a*
Fine Arts	31	7	22.6%	6	85.7%	66.7%	0.0%	33.3%	16.7%	100.0%	50.0%	50.0%
Folklore	2	2	100.0%	1	50.0%	100.0%	0.0%	0.0%	100.0%	100.0%	100.0%	100.0%
History	47	21	44.7%	18	85.7%	50.0%	5.6%	44.4%	27.8%	83.3%	66.7%	50.0%
Law	11	1	9.1%	1	100.0%	100.0%	0.0%	0.0%	100.0%	100.0%	100.0%	0.0%
Mediaeval Studies	3	2	66.7%	2	100.0%	0.0%	0.0%	100.0%	0.0%	100.0%	100.0%	50.0%
Modern Languages and Literature	56	21	37.5%	14	66.7%	35.7%	7.1%	57.1%	21.4%	85.7%	71.4%	42.9%
Philosophy	19	7	36.8%	5	71.4%	40.0%	20.0%	40.0%	20.0%	60.0%	20.0%	40.0%
Social Science	323	96	29.7%	66	68.8%	47.0%	10.6%	42.4%	36.4%	65.2%	65.2%	51.5%
Anthropology	26	7	26.9%	6	85.7%	33.3%	16.7%	50.0%	0.0%	66.7%	66.7%	50.0%
Communications and Media Studies	15	8	53.3%	7	87.5%	57.1%	0.0%	42.9%	42.9%	71.4%	100.0%	85.7%
Criminology	7	1	14.3%	0	0.0%	*n/a*	*n/a*	*n/a*	*n/a*	*n/a*	*n/a*	*n/a*
Demography	1	0	0.0%	*n/a*	*n/a*	*n/a*	*n/a*	*n/a*	*n/a*	*n/a*	*n/a*	*n/a*
Economics	15	5	33.3%	4	80.0%	25.0%	25.0%	50.0%	0.0%	75.0%	50.0%	50.0%
Education	51	15	29.4%	10	66.7%	60.0%	20.0%	20.0%	50.0%	30.0%	30.0%	20.0%
Geography	17	6	35.3%	3	50.0%	66.7%	33.3%	0.0%	33.3%	33.3%	33.3%	33.3%
Interdisciplinary Studies	2	0	0.0%	*n/a*	*n/a*	*n/a*	*n/a*	*n/a*	*n/a*	*n/a*	*n/a*	*n/a*
Linguistics	17	6	35.3%	3	50.0%	33.3%	0.0%	66.7%	33.3%	100.0%	33.3%	66.7%
Management, Business, Administrative Studies	25	7	28.0%	4	57.1%	75.0%	25.0%	0.0%	25.0%	75.0%	75.0%	75.0%
Political Science	43	14	32.6%	11	78.6%	36.4%	0.0%	63.6%	45.5%	100.0%	90.9%	63.6%
Psychology	39	9	23.1%	7	77.8%	57.1%	14.3%	28.6%	42.9%	42.9%	57.1%	28.6%
Religious Studies	12	2	16.7%	1	50.0%	0.0%	0.0%	100.0%	0.0%	0.0%	0.0%	0.0%
Social Work	8	1	12.5%	1	100.0%	100.0%	0.0%	0.0%	0.0%	100.0%	100.0%	100.0%
Sociology	36	11	30.6%	6	54.5%	33.3%	0.0%	66.7%	50.0%	66.7%	66.7%	50.0%
Urban and Regional Studies, Environmental Studies	9	4	44.4%	3	75.0%	33.3%	0.0%	66.7%	66.7%	66.7%	100.0%	66.7%

In order to investigate the extent to which Twitter plays a role in the professional academic life of the SSHRC Doctoral Award recipients, Twitter account descriptions were analyzed. These descriptions can be written in free text up to 160 characters and provides Twitter users with an opportunity to present information about themselves. The descriptions help other users to identify interesting and relevant accounts to follow or correspond with. A study of 632 emergency physicians on Twitter found that the text could lead to social networking benefits on Twitter, as users who included work-related information had more followers, including other emergency physicians (Lulic & Kovic, 2013). Bowman (2015b) found that among US professors with Twitter accounts, 87% added their professional title and place of work to their profiles.

Of the 160 identified Twitter accounts of SSHRC Doctoral Award recipients, 114 (71.3%) provided a description on their Twitter profiles. Descriptions were coded into three categories: academic, non-academic, or both. These texts were further analyzed regarding mentions of the topic of the PhD, the discipline, an academic title, or a university affiliation. On average, 46.5% profiles included strictly academic self-presentations, 8.8% were strictly non-academic, and 44.7% included both academic and non-academic content. Thus, 91.2% of all SSHRC Doctoral

Award recipients provided an account description, in which they identified themselves in some way as academic. Including those accounts without an account description, this amounts to two-thirds of all award recipients with a Twitter account. The majority of award recipients on Twitter also mentioned their discipline (73.7%) and academic title (64.0%) in their account description, while nearly half (49.1%) mentioned an affiliation with a university and less than one third (31.6%) specified their PhD topic. It should be noted that the account information was retrieved in February 2015, four years after the funding period, such that some of the applicants had finished their PhD and were now working as assistant professors, lecturers, or in other professions. In these cases, it is less likely that they would mention the topic of their PhD thesis. The presence of academic disciplines, titles and universities, as well as the fact that more than 90% of all Twitter account descriptions included some academic content, suggests that the majority of 2010 SSHRC Doctoral Award recipients felt it appropriate to present themselves on Twitter in an academic context. One scenario that might explain this is the understanding that in most instances Twitter profiles are searchable and available to vast invisible audiences, which can have an influence the way young scholars are perceived when applying for jobs, attempting to collaborate with other scholars, and applying for funding. Thus, these Twitter users are presenting themselves in a professional way by associating themselves with their universities, their academic accomplishments, and their area of study.

Comparing the two domains (Humanities and Social Science), award recipients from the Social Science provided slightly less Twitter account descriptions (68.8% vs. 75.0%) and were more likely to present themselves as strictly non-academic (10.6% vs. 6.3%). They were also less likely to include their discipline in the self-description (65.2% vs. 85.4%), but more likely to mention the topic of their PhD (36.4% vs. 25.0%) and a university (51.5% vs. 45.8%) than award recipients from the Humanities. On the level of main fields, the interpretation focuses on the comparison on History, Modern Languages and Literature, Education and Political Science, as other fields had too few users to interpret results. With only two-thirds of SSHRC Doctoral Award recipients including information in the Twitter bio section, award recipients in Modern Languages and Literature as well as Education were less likely to provide a self-description. The percentage of accounts including a description was above average in History (85.7%) and Political Science (78.6%). Historians were also more likely to include their discipline (83.3% vs. 73.7%). The majority of award recipients in Education presented themselves strictly as academics (60% vs. 46.5% for all fields), but were also more likely to include strictly non-academic descriptions (20.0% vs. 8.8%). They were also much less likely to mention Education as their discipline (30.0% vs. 73.7%). This suggests that these users with a PhD in Education were less likely to discuss research and work-related topics on Twitter. Based on self-presentation in the Twitter bio, they seemed less likely to mix professional and private issues: 20% (vs. 8.8% overall) included strictly non-academic descriptions, while 60.0% (vs. 46.5%) presented themselves as strictly academic. Modern Languages and Literature as well as Political Science PhD students were less likely to describe themselves in a strictly academic fashion but more likely to include both academic and non-academic information in their Twitter profiles. In fact, award recipients in Political Science always included some kind of academic information and never appeared in a strictly non-academic way, which suggests that Twitter is particularly used for academic discussions by analyzed Political Science award recipients. This is supported by the finding that all of the eleven political scientists providing an account description mentioned their discipline and 90.9% their academic title. Percentages of mentioning the university (63.6%) and PhD topic (45.5%) were also above average in Political Science. However, with the exception of including the discipline, differences between fields regarding self-presentation based on the Twitter account description were not statistically significant according to the Chi-square test (see Appendix A.2.3 for detailed results).

3.2 Twitter Use

As the sheer presence of a Twitter account does not necessarily reflect actual usage of a platform, Twitter use of the SSHRC Doctoral Award recipients was analyzed based on the number of tweets, number of followers and following as well the extent to which certain affordances (i.e., retweets, hashtags, user mentions and URLs) were used on Twitter. In order to compare average tweeting activity, the mean number of tweets was computed taking into account the date a user opened the Twitter account (see Appendix A.2 for a detailed description of the methods). Thus tweeting activity could be compared independently of when a user joined Twitter.

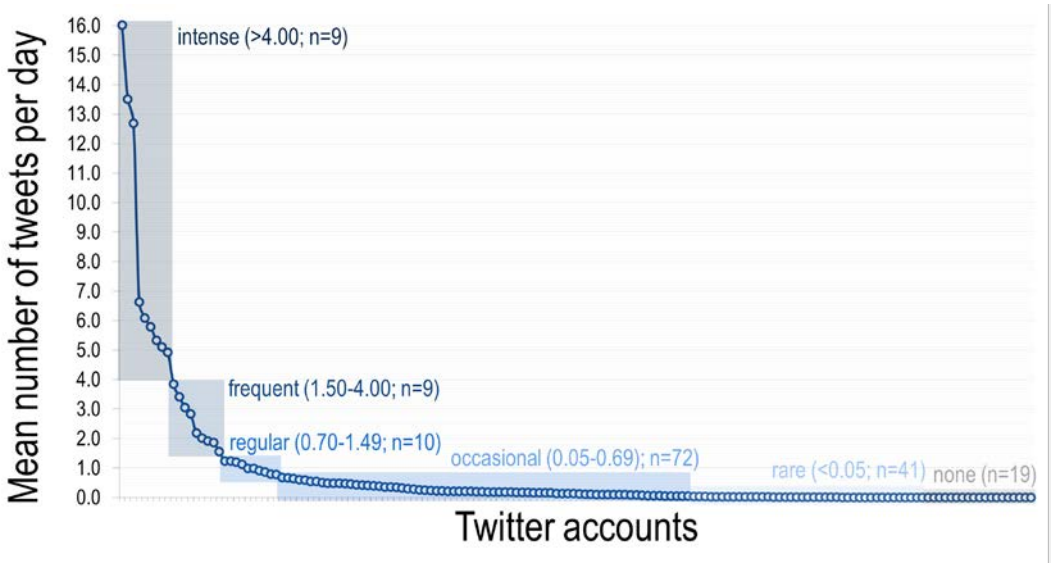


Figure 1. Average tweeting activity per user.

As shown in Figure 1, tweeting activity varied among the 160 accounts and ranged from 0 to an average of 16 tweets per day. As to be expected, tweeting activity is distributed according to a Pareto distribution, where a few users tweet the most—17.5% tweet at least regularly—and the majority of users tweet only occasionally (45%), rarely (25.6%), or not at all (11.9%). Similar distributions of tweets per day per user were found in Haustein, Bowman, Holmberg, Peters, and Larivière (2014) for a group of astrophysicists on Twitter. Among US professors, those from the social sciences averaged a higher number of mean tweets per day than the scholars from the natural sciences (1.40 compared to 0.61 respectively), although the difference was not statistically significant (Bowman, 2015b). Users were grouped into six groups according to mean tweeting activity: intense (more than 4 tweets per day), frequent (1.50-4.00), regular (0.70-1.49), occasional (0.05-0.69), rare (less than 0.05 tweets per day) and none (no tweets). More intense and frequent tweeters were found in the Humanities, while the Social Science had less frequent tweeters. These differences were, however, not significantly different according to the Mann-Whitney U test (see Appendix A.2.3). Users who tweeted on average more than 4 times per day (intense use) did their PhDs in History (n=2), Political Science (n=2), and Modern Languages and Literature (n=1), while users from Education and Sociology had no user tweeting more than occasionally (0.69 tweets per day). Users from Modern Languages and Literature represented the highest number of Twitter accounts (n=4) who had not tweeted at all. There is a statistically significant difference between Education and Political Science based on the absolute number of tweets sent during the last 12 months before data collection with award recipients from Political Science tweeting significantly more than those from Education (see Appendix A.2.3).

The number of followers indicates how many Twitter users show an interest in the tweeting activity of the SSHRC Doctoral Award recipients as they follow their tweets on Twitter. In the study by Mou (2014) Twitter profiles of professors with many followers were rated more highly for professionalism, likability, and credibility compared to those with fewer. As shown in Figure 2, the number of followers varied among the 160 accounts and ranged from 0 to over 25,000. Just as with tweeting activity, the amount of followers follows a Pareto distribution, where a few users have the highest number followers and the majority of users have less than 100 followers. Overall Twitter users from Humanities had more followers than those in the Social Sciences. Only users from History (n=1) and Political Science (n=1) had accounts with over 1,500 followers. Among the five fields with more than 10 users, Sociology (n=2) users represented the highest number of accounts without followers.

Comparing the two domains, users from Humanities were shown to tweet more on average per day (1.12) and to follow others (271) more than those in the Social Sciences, whereas users from Social Sciences had a much larger mean number of followers (441). Focusing on the four main fields with the highest number of SSHRC Doctoral Award recipients, users from History had the highest average mean tweets per day (1.30) while Sociology users had the lowest (0.13). Political Science users had the highest mean number of followers (434) and mean number of those they were following (414).

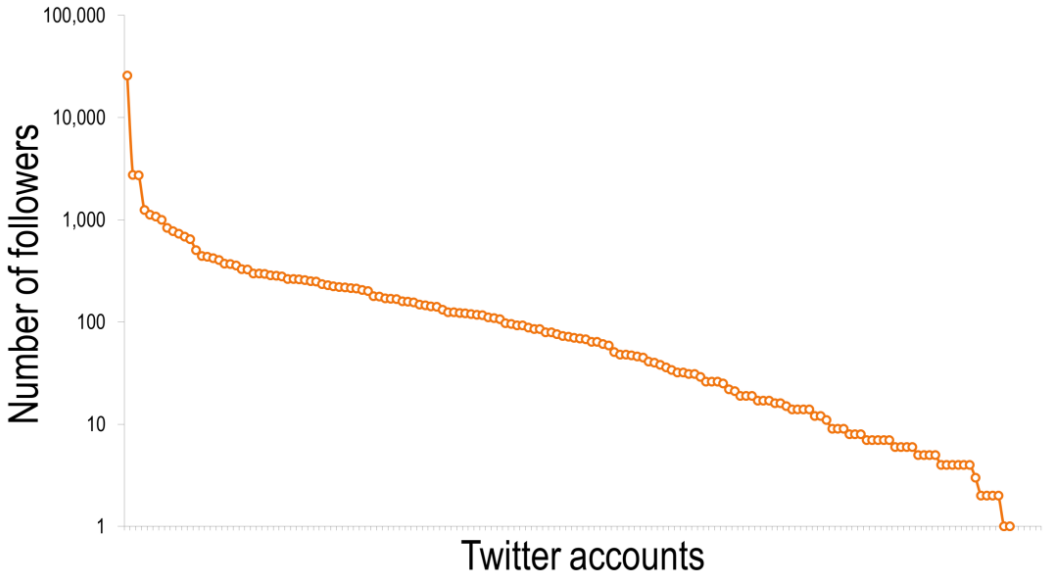


Figure 2. Number of followers per Twitter account (logarithmic scale).

Table 2 summarizes the average Twitter activity per Twitter user per domain and main field. The analysis of the tweeting activity was restricted to the last 12 months to improve comparability. The statistics focus on the use of Twitter affordances (Bowman, 2015b) such sent and received retweets, as well as the percentage of tweets containing hashtags, user mentions and URLs. On average, SSHRC Doctoral Award recipients had 340 followers and were following 244 other Twitter users. However, the mean ratio of followers and following was 0.86, which indicates that for most of the 160 accounts the number of accounts they followed exceeded the number of users following them. On average, users sent, on average, slightly more than one tweet per day during the last year and 9.7% were retweeted by others.

As described above, results for fields with a small number of accounts should be interpreted with care. Because of this limitation, the description focuses on fields with at least 10 accounts, namely History, Modern Languages

and Literature, Education, Political Science and Sociology. Overall users from Humanities (1.12 to 1.47 mean number of tweets per day per main field) and Social Sciences (0.58 to 0.85) both showed an increase in tweeting activity over the last 12 months compared to the entire period analyzed (Figure 1). Users from Political Science (1.10 to 2.34) had the largest increase in average mean tweets per day per user, while users from History (1.30 to 1.09) showed a decrease in activity. Comparing the total number of tweets between the four largest main fields, the difference in the median number of tweets sent during the 12-month period was significant between Education (Mdn = 37.0) and Political Science (Mdn = 255.0), but not between any other combination of the four fields. For detailed results of the statistical tests refer to Appendix A.2.3.

Table 2. Average Twitter account activity and affordance use during the last 12 months per user per domain and field.

Domain and main field	Recipients with Twitter account	Mean number of followers	Mean number of following	Mean ratio followers/ following	Tweets (last 12 months)						
					sum	mean tweeting activity	mean % RTs	mean % tweets retweeted	mean % tweets w/ hashtag	mean % tweets w/ user mentions	mean % tweets w/ URLs
all fields	*160*	*340*	*244*	*0.86*	*43,176*	*1.09*	*36.7%*	*9.7%*	*32.2%*	*72.4%*	*47.4%*
Humanities	64	187	271	0.63	21,329	1.47	36.6%	9.8%	34.2%	71.8%	42.9%
Archaeology	2	11	42	0.40	79	0.22	6.3%	0.0%	57.0%	15.2%	10.1%
Classics, Classical & Dead Languages	1	21	89	0.24	34	0.09	55.9%	2.9%	44.1%	91.2%	41.2%
Fine Arts	7	224	364	1.20	1,086	0.75	33.1%	10.9%	51.5%	68.0%	40.8%
Folklore	2	424	168	1.83	2,864	7.85	46.4%	12.3%	49.4%	85.7%	43.5%
History	21	247	307	0.57	7,064	1.09	36.6%	11.5%	28.2%	75.6%	46.2%
Law	1	3	8	0.38	0	0.00	*n/a*	*n/a*	*n/a*	*n/a*	*n/a*
Mediaeval Studies	2	291	467	0.63	1,523	2.09	21.4%	17.7%	19.5%	68.7%	40.2%
Modern Languages and Literature	21	122	222	0.49	5,844	1.68	38.1%	6.9%	35.4%	69.9%	40.0%
Philosophy	7	170	322	0.51	2,835	1.94	43.7%	10.6%	35.2%	71.7%	49.7%
Social Science	96	441	226	1.01	21,847	0.85	36.8%	9.6%	30.9%	72.9%	50.2%
Anthropology	7	295	250	1.53	2,350	1.07	38.5%	14.2%	32.5%	74.9%	66.0%
Communications and Media Studies	8	163	127	1.47	1,047	0.36	25.7%	10.9%	22.2%	54.9%	57.0%
Criminology	1	4	22	0.18	3	0.03	33.3%	33.3%	66.7%	66.7%	66.7%
Economics	5	39	94	0.36	83	0.08	27.9%	6.4%	19.6%	89.9%	18.1%
Education	15	75	141	0.90	794	0.19	36.3%	10.3%	30.2%	78.9%	49.8%
Geography	6	173	263	0.56	2,061	1.49	51.4%	9.0%	51.7%	78.1%	40.1%
Linguistics	6	120	156	0.83	1,594	0.88	16.9%	5.1%	15.5%	61.9%	25.2%
Management, Business, Administrative Studies	7	3,915	406	4.25	3,498	1.60	33.1%	8.1%	30.9%	59.5%	82.0%
Political Science	14	434	413	0.80	8,556	2.34	39.6%	13.4%	29.4%	80.0%	41.1%
Psychology	9	143	325	0.53	761	0.52	30.3%	5.6%	42.8%	79.0%	54.6%
Religious Studies	2	51	178	0.28	4	0.01	50.0%	0.0%	100.0%	50.0%	25.0%
Social Work	1	14	69	0.20	25	0.07	72.0%	8.0%	52.0%	76.0%	88.0%
Sociology	11	51	100	0.41	350	0.14	55.0%	5.3%	25.1%	76.2%	45.4%
Urban and Regional Studies, Environmental Studies	4	126	178	0.62	721	0.66	37.3%	8.5%	23.7%	82.2%	51.7%

As shown in Table 2, 72.4% of the tweets sent by SSHRC Doctoral Award recipients during the last year before data collection contained user mentions, which indicates that they use Twitter to connect with and interact with other users. This percentage is much higher than what was found for a small group of astrophysicists on Twitter, where about half of the users' tweets contained other Twitter handles (Haustein, Bowman, Holmberg, et al., 2014). In fact, affordance use by the SSHRC award recipients exceeded that of the astrophysicists in all categories. The percentage of retweets sent by SSHRC Doctoral Award recipients (36.7%) was also much higher than that by astrophysicists (13.1%), while the difference between hashtags (32.2% vs. 22.8%) and URLs (47.4% vs. 36.7%) was smaller. The share of tweets with URLs was comparable to that found by Priem and Costello (2010) for 26 academics.

Comparing affordance use among award recipients, those from the Humanities had a slightly higher mean average of tweets with hashtags (34.2%) than Social Sciences (30.9%), whereas doctoral students in Social Sciences had a higher average use of URLs (50.2%) than those from the Humanities (50.2%). Sociologists had a

higher percent average of retweets (55%) than the other domains. However, the differences in the percentages of tweets containing hashtags, mentions, RTs, and URLs among the fields of History, Modern Languages and Literature, Education and Political Science were not found to be statistically significant (see Appendix A.2.3 for methods and details). In fact, there were no statistical differences found for affordance use (i.e., hashtag, mention, RT, and URL usage) between the two domains or four fields for the analyzed 12-month period.

Other than what was found for the group of astrophysicists, where the percentage of RTs and user mentions increased with tweeting activity and rare tweeters hardly used hashtags (Haustein, Bowman, Holmberg, et al., 2014), no association between number of tweets and affordance use was found for the SSHRC Doctoral Award recipients (see scatterplots in Figure 5 in Appendix A.2.3).

3.3 Tweet Content

The tweet content analysis focuses on the four fields with the highest number of Twitter users identified, namely History, Modern Languages and Literature, Education and Political Science. The analysis was restricted to tweets published during the last 12 months before data collection in February 2015 to focus on the most recent discussions and ensure some stability of contents. Content was assessed analyzing the most popular hashtags (Section 3.3.1) and URLs (Section 3.3.2) tweeted by the SSHRC Doctoral Award recipients as well as by coding a representative sample of tweets to analyze the tweeted topics (Section 3.3.3). The analysis was aimed at identifying to what extend the award recipients used Twitter in a scholarly context, for example to discuss topics relevant to their field of study. The content analysis was carried out separately for History, Modern Languages and Literature, Education and Political Science to distinguish particular differences between the four fields.

3.3.1 Hashtag Analysis

Hashtags are one of the most popular affordance on Twitter. These keywords indicated by a # sign often represent particular topics and allow users to identify related tweets through the Twitter search functionality. In an academic context, hashtags are particularly used at conferences where conference-specific hashtags help attendees and others to follow discussions during the conference in real-time. Thus conference hashtags can function as a retrieval tool, as well as for networking (Weller & Puschmann, 2011). The analysis of most frequently used hashtags can thus reveal certain topics discussed by the SSHRC Doctoral Award recipients. A social network graph (Figure 3) can visualize the connections between award recipients through commonly used hashtags and thus reveal whether they interact by discussing common topics on Twitter.

Of the 22,258 tweets sent by the SSHRC Doctoral Award recipients from February 2014 to February 2015, 24.5% contained at least one hashtag and 0.3 per tweet (Table 3). Note that these statistics are not compiled per person but as in Section 3.2 and thus differ slightly from results reported in Table 2. These tweets contained 2,779 different hashtags, which occurred 2.8 times on average. Not all of the award recipients from History, Modern Languages and Literature, Education, and Political Science made use of hashtags (88.7%). Although the percentage of tweets with hashtags was highest (33.2%), Education was the main field with the lowest percentage of Twitter users making use of hashtags (75.0%).

Table 3. Hashtag use of users in four selected fields during the last 12 months.

	tweets	tweets with hashtags		mean hashtags per tweet	unique hashtags	mean tweets per hashtag	unique users using hashtags	
		n	%				n	%
all fields	22,258	5,444	24.5%	0.3	2,779	2.8	47	88.7%
Humanities	12,908	3,288	25.5%	0.4	1,856	2.5	29	93.5%
History	7,064	1,843	26.1%	0.4	924	2.8	17	94.4%
Modern Languages and Literature	5,844	1,445	24.7%	0.4	1,045	2.1	12	92.3%
Social Science	9,350	2,156	23.1%	0.3	1,118	2.7	18	81.8%
Education	794	264	33.2%	0.5	150	2.6	9	75.0%
Political Science	8,556	1,892	22.1%	0.3	990	2.7	9	90.0%

The 7,064 tweets published during the 12-months period by the 17 award recipients in History (4 of the 21 users did not tweet in the last 12 months) contained 924 unique hashtags, which were mentioned 2,551 times. On average, the 26.1% of the tweets contained hashtags. The connections between users (yellow squares) and hashtags (blue circles) are visualized in the 2-mode network graph in Figure 3A (see Appendix A.2.5 for detailed methods). Frequency of hashtags is indicated by the size of nodes. The majority of the most frequently used hashtags referred to places. *Toronto* and *Canada* indicate the involvement in local and Canadian conversations of the SSHRC Doctoral Award recipients. The hashtags *Israel*, *Palestine* and *Gaza*, *BDS* (Boycott, Divestment and Sanctions) as well as *Ferguson* reflect discussions about current political events and conflicts. *AAA2014* is the official hashtag of the 2014 Annual Meeting of the American Anthropological Association and although mentioned 43 times, it was used by only 1 of the 17 History students, who likely tweeted content from the conference. The same applies to *vegan*. In fact, as can be seen in the lower right of Figure 3A, both hashtags were tweeted by the same user. This user was also the most active in terms of hashtags use, mentioning 265 hashtags 1,048 times, which represents 28.7% of unique hashtags and 41.1% of the overall hashtag used by the 17 award recipients working on a PhD in History. The hashtag connecting most of the 17 users is *twitterstorians* (in the center of the network in Figure 3A), which is a term particularly used to connect historians[3] on Twitter. During the time of analysis it was used 47 times by 11 different users. The hashtags *history*, *cdnhistory,* and *phdchat* also suggest that the SSHRC Doctoral Award recipients from History use Twitter to discuss and connect with other Historians and PhD students. *Cdnpoli* (Canadian politics), *MMIW* (missing and murdered indigenous women) and *onpoli* (Ontario politics) identified current political discussions in Canada. *BellLetsTalk* referred to a charitable program initiated by Bell that is dedicated to mental health.[4]

The 21 Modern Languages and Literature PhD students with Twitter accounts sent a total of 5,844 tweets during the 12-month period before data collection in February 2015. Tweets with hashtags accounted for 24.7% of the total number of tweets during this period. Twelve users were responsible for sending the tweets, which contained 1,045 unique hashtags. The hashtags tweeted by the largest number of users largely referenced news and current

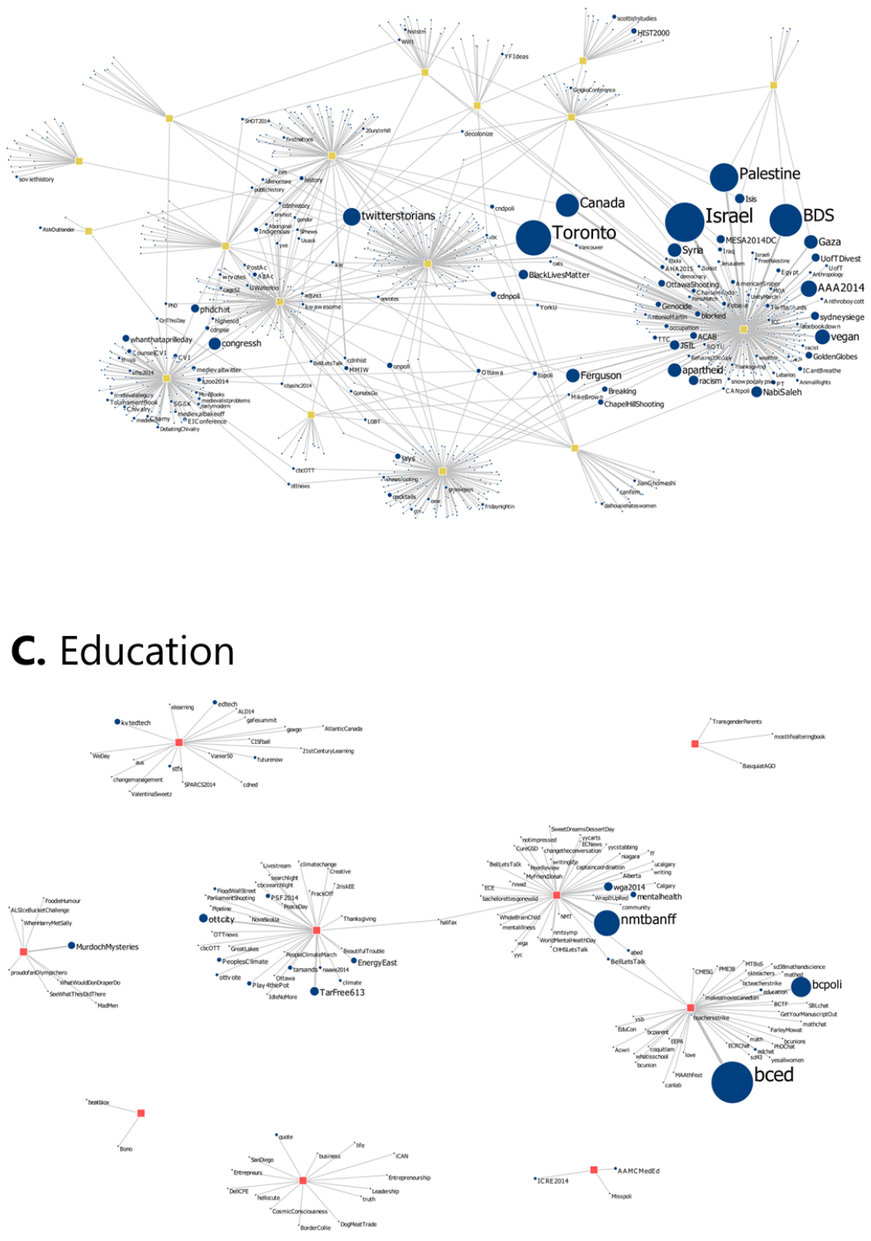
[3] https://thevieweast.wordpress.com/2010/09/07/history-on-twitter-happy-anniversary-twitterstorians/

[4] http://letstalk.bell.ca/en/our-initiatives/

events, often with a political or activist angle. The top three hashtags included *Ferguson* (referencing civil unrest in Ferguson), *cdnpoli* (Canadian politics), and *cdnpse* (Canadian post-secondary education) (Figure 3B). While the hashtags with the most users had some overlap with the most frequently occurring hashtags (*Ferguson* was ninth with 24 tweets, *cdnpse* was third with 32 tweets), the majority of the most frequently tweeted hashtags were only sent by one user each indicating that the SSHRC Doctoral Award recipients in Modern Languages and Literature did not focus on the same topics. As shown in Figure 3B, the most frequently used hashtags were mainly references to either current events or conferences. Conferences included *VSAWC2014* (Victorian Studies Association of Western Canada 2014 conference), *MLA15* (2015 Modern Language Association Annual Convention), *dhsi2014* (Digital Humanities Summer Institute 2014), *UOconference* (Upgrading Online Annual Conference), and *acifa2014* (Alberta Colleges and Institutes Faculties Association 2014 conference). Only the MLA15 conference tag was tweeted by more than two users, however. The use of conference hashtags in particular, and hashtags in general, could help SSHRC Doctoral Award recipients to increase audiences and their collaboration networks through Twitter. For instance, Sopran, Rey, Butler, and Shneiderman (2012) showed that the vast majority of the Twitter users tweeting about the Theorizing the Web 2012 conference were not actually attending the conference in person, which highlights the potential of conference hashtags to involve larger audiences. However, the Twitter-related networking benefits were statistically associated with being present at the conference, with speakers and in-person attendees gaining significantly more followers than remote participants.

A total of 12 Education PhD students sent 794 tweets during the 12-month analysis period. Of these, 33.2% contained hashtags, which is above the Social Sciences average of 23.1% (Table 3). Nine users sent tweets with hashtags for a total of 264 tweets and 150 unique hashtags. The most frequently occurring were *bced* (BC education), *nmtbanff* (Neurosequential Model Symposium), *bcpoli* (BC politics), *ottcity* (city of Ottawa), and *TarFree613* (expansion of oil pipeline into Ontario). There was no overlap between the most frequently occurring hashtags and those sent by the most users, however. The top 10 most frequently tweeted hashtags were each only sent by one user, and only three hashtags were tweeted by more than one unique user (and none more than two unique users). These included *BellLetsTalk* (a mental health awareness initiative from the company Bell), *ABed* (Alberta education), and *halifax*. This discordance among the number of occurrences of hashtags and the number of users can be seen in the distinct separation of user hashtag clusters in Figure 3C. Only one user is connected to two others through one and two common hashtags, respectively. The Twitter users of award recipients in Education did thus not connect through commonly used hashtags.

Political Science award recipients sent the most tweets during the 12-month period of all the analyzed disciplines and at 2.34 tweets per day showed the highest average tweeting activity (Table 2). An average of 29% of tweets included hashtags, which indicated an interest in current politics. Over half of the most frequently occurring hashtags were related to political events and news, and the top four referenced Canadian politics at the federal, provincial and municipal levels: *cdnpoli*, *topoli*, *onpoli and ableg* stand for Canadian, Toronto and Ontario politics and Alberta legislature. There was decent overlap between the most frequently tweeted hashtags and those with the most users, with half of the top ten hashtags also having the most users and the most frequently occurring hashtag *cdnpoli* was the hashtag used by the most users. This central linking node is visually reflected in the network graph Figure 3D. Other political news and events related hashtags were also tweeted by multiple users, including the top four (*cdnpoli*, *onpoli*, *TOpoli*, *Ferguson*). Comparing the frequency distributions of hashtags of the four main fields, the Political Science network was also most concentrated with the top four hashtags used by more than half of the users and accounting for 30% of all hashtag occurrences.



History	most occurring				most users				
# hashtag	occurrence n=2,551		user n=17		# hashtag	user n=17		occurrence n=2,551	
	n	%	n	%		n	%	n	%
1 Israel	108	4%	2	12%	1 twitterstorians	11	65%	47	2%
2 toronto	95	4%	4	24%	2 history	6	35%	7	0%
3 BDS	87	3%	2	12%	3 cdnpoli	5	29%	14	1%
4 Palestine	78	3%	3	18%	MMIW	5	29%	10	0%
5 Canada	64	3%	2	12%	5 toronto	4	24%	95	4%
6 twitterstorians	47	2%	11	65%	phdchat	4	24%	21	1%
7 AAA2014	43	2%	1	6%	onpoli	4	24%	11	0%
8 vegan	40	2%	1	6%	Ottawa	4	24%	7	0%
9 Ferguson	37	2%	3	18%	cdnhistory	4	24%	6	0%
Gaza	37	2%	2	12%	BellLetsTalk	4	24%	4	0%

Modern Languages	most occurring				most users				
# hashtag	occurrence n=2,144		user n=12		# hashtag	user n=12		occurrence n=2,144	
	n	%	n	%		n	%	n	%
1 VSAWC2014	92	4%	1	8%	1 Ferguson	5	42%	24	1%
2 charte	46	2%	1	8%	2 cdnpoli	4	33%	11	1%
3 cdnpse	32	2%	3	25%	3 cdnpse	3	25%	32	2%
4 MLA15	30	1%	3	25%	MLA15	3	25%	30	1%
Haiti	30	1%	1	8%	YesAllWomen	3	25%	22	1%
6 dhsi2014	29	1%	1	8%	writing	3	25%	19	1%
7 withaPhD	27	1%	1	8%	phdchat	3	25%	11	1%
8 UOconference	25	1%	1	8%	blacklivesmatter	3	25%	7	0%
9 Ferguson	24	1%	5	42%	9 *52 hashtags mentioned by 2 users each*				
10 acifa2014	23	1%	1	8%					
QC2014	23	1%	1	8%					

Education	most occurring				most users				
# hashtag	occurrence n=396		user n=9		# hashtag	user n=9		occurrence n=396	
	n	%	n	%		n	%	n	%
1 bced	58	15%	1	11%	1 BellLetsTalk	2	22%	4	1%
2 nmtbanff	35	10%	1	11%	2 ABed	2	22%	3	1%
3 bcpoli	27	9%	1	11%	3 halifax	2	22%	2	1%
4 ottcity	12	4%	1	11%	4 *147 hashtags mentioned by 1 user each*				
TarFree613	12	5%	1	11%					
6 wga2014	11	4%	1	11%					
7 MurdochMysteries	10	4%	1	11%					
8 mentalhealth	8	3%	1	11%					
EnergyEast	8	4%	1	11%					
10 kvtedtech	7	3%	1	11%					

Political Science	most occurring				most users				
# hashtag	occurrence n=2,649		user n=9		# hashtag	user n=9		occurrence n=396	
	n	%	n	%		n	%	n	%
1 cdnpoli	329	12%	7	78%	1 cdnpoli	7	78%	329	12%
2 TOpoli	209	9%	5	56%	2 onpoli	6	67%	59	3%
3 ableg	136	6%	2	22%	3 TOpoli	5	56%	209	9%
4 onpoli	59	3%	6	67%	Ferguson	5	56%	32	2%
5 bioethics	56	3%	1	11%	5 yyc	4	44%	42	2%
6 IVF	52	3%	1	11%	toronto	4	44%	24	1%
7 yyc	42	2%	4	44%	SCC	4	44%	24	1%
8 thucydidesthursday	34	2%	1	11%	8 cbc	3	33%	11	1%
9 Ferguson	32	2%	5	56%	indyref	3	33%	11	1%
10 yql	28	1%	1	11%	phdchat	3	33%	10	1%
					law	3	33%	10	1%
					cdnpse	3	33%	7	0%
					Canada	3	33%	7	0%
					queensu	3	33%	5	0%
					WorldPride	3	33%	4	0%
					yeg	3	33%	4	0%
					canlab	3	33%	4	0%
					s	3	33%	3	0%
					science	3	33%	3	0%
					CPC	3	33%	3	0%
					inequality	3	33%	3	0%
					FairElectionsAct	3	33%	3	0%

Figure 3. 2-mode networks of hashtags (blue circles) and users (squares) in History (A), Modern Languages and Literature (B), Education (C) and Political Science (D) during the last 12 months.

3.3.2 Link Analysis

Twitter users often include links in their tweets as it allows them to distribute more information than the 140 character limit of a tweet affords. The link analysis is carried out at the domain level of URLs that appeared in tweets of the four main fields History, Modern Languages and Literature, Education and Political Science, since as few as 17 URLs were shared by more than one user. The use of links was more common than the use of the Twitter specific hashtags, as 38.0% of the 22,258 tweets contained a URL (Table 4). Note that these statistics are compiled as averages per field and not per person as in Section 3.2 and thus differ slightly from results reported in Table 2. This is comparable to the percentages reported for 26 academics (Priem & Costello, 2010). In Education and History links were used more than in Political Science and Modern Languages. The SSHRC Doctoral Award recipients from the four main fields linked to 2,529 unique domains and each domain was referred to 3.4 times on average.

Table 4. Use of links by users in four selected fields during the last 12 months.

	tweets	tweets with URLs		mean URLs per tweet	unique domains	mean tweets per domain	unique users using domain	
		n	%				n	%
all fields	22,258	8,449	38.0%	0.4	2,529	3.4	49	92.5%
Humanities	12,908	5,023	38.9%	0.4	1,595	3.2	28	90.3%
History	7,064	3,018	42.7%	0.4	916	3.3	17	94.4%
Modern Languages and Literature	5,844	2,005	34.3%	0.3	849	2.4	11	84.6%
Social Science	9,350	3,426	36.6%	0.4	1,216	2.9	21	95.5%
Education	794	379	47.7%	0.5	192	2.0	12	100.0%
Political Science	8,556	3,047	35.6%	0.4	1,068	3.0	9	90.0%

Focusing on the field of History, the 7,064 tweets published by the SSHRC Doctoral Award recipients referred to 916 unique domains (Table 4). The most popular websites distributed by History award recipients were news and social media sites such as *theguardian.com*, *cbc.ca*, *haaretz.com*, *972mag.com*, *thestar.com*, *theglobeandmail.com*, *nytimes.com*, *bbc.co.uk* as well as *youtube.com* and *facebook.com* (Table 5). This suggests that the most popular URLs shared on Twitter by the 17 historians are general news and not specific sites referring to historical research. This corresponds to findings by Holmberg and Thelwall (2014), who identified less than 3.5% of sampled tweets by tenure-track researchers as part of scholarly communication. Among the SSHRC Doctoral Award recipients from History, 31 URLs linked to a '.edu' website, most of which refer to university websites or *academia.edu* linking to profiles or publications on the academic social networking site, while 30 times websites of Canadian universities were tweeted. Wiley (4 mentions), Taylor and Francis (3), UBC Press (2) and Penguin Random House (2) could be identified as academic publishers, to which the historian linked more than once. 26 domains referred to 'history' in the domain name, the most popular of which is *activehistory.ca*, which was mentioned by 3 users in 27 tweets. Compared to the strong presence of news websites, academic and historical content was, however, hardly linked to by the recipients. Looking at the way academics cite external sources on Twitter, Priem and Costello (2010) found that 6% of the tweets by 26 university faculty, postdocs, or doctoral students contained links to peer-reviewed scholarly articles, 52% of which were direct links and 48% indirectly linked to a site that referenced the article. Holmberg and Thelwall (2014) similarly found that scholarly tweets linked less to actual articles and more to blogs or items which then linked to the articles.

Table 5. Most popular domains of URLs used in History.

most occurring						most users				
#	domain	occurrence *n=3,045*		user *n=17*		#	domain	user *n=17*		occurrence *n=3,045*	
		n	%	n	%			n	%	n	%
1	theguardian.com	194	6%	9	53%	1	youtube.com	11	65%	105	3%
2	uapp.ly	141	5%	1	6%		theglobeandmail.com	11	65%	46	2%
3	youtube.com	105	4%	11	65%	3	cbc.ca	10	59%	82	3%
4	facebook.com	93	3%	9	53%	4	theguardian.com	9	53%	194	7%
5	cbc.ca	82	3%	10	59%		facebook.com	9	53%	93	4%
6	haaretz.com	80	3%	3	18%	6	nytimes.com	8	47%	27	1%
7	972mag.com	78	3%	1	6%		bbc.co.uk	8	47%	11	0%
8	thestar.com	60	2%	7	41%	8	thestar.com	7	41%	60	2%
	electronicintifada.net	51	2%	1	6%	9	slate.com	6	35%	24	1%
10	theglobeandmail.com	46	2%	11	65%		ow.ly	6	35%	6	0%

Table 6. Most popular domains of URLs used in Modern Languages and Literature.

most occurring						most users				
#	domain	occurrence *n=2,035*		user *n=11*		#	domain	user *n=11*		occurrence *n=2,035*	
		n	%	n	%			n	%	n	%
1	youtube.com	121	6%	9	82%	1	youtube.com	9	82%	121	6%
2	cbc.ca	48	2%	7	64%	2	cbc.ca	7	64%	48	3%
3	trib.al	38	1%	6	55%	3	trib.al	6	55%	38	2%
4	facebook.com	33	1%	6	55%		facebook.com	6	55%	33	2%
	theguardian.com	33	1%	6	55%		theguardian.com	6	55%	33	2%
6	huffingtonpost.com	26	1%	6	55%		huffingtonpost.com	6	55%	26	1%
7	chronicle.com	25	1%	3	27%	7	theglobeandmail.com	5	45%	22	1%
	lapresse.ca	25	1%	1	9%		twitter.com	5	45%	19	1%
9	untappd.com	24	1%	1	9%		washingtonpost.com	5	45%	10	1%
10	theglobeandmail.com	22	1%	5	45%	10	slate.com	4	36%	14	1%
							buzzfeed.com	4	36%	12	1%
							npr.org	4	36%	12	1%
							thestar.com	4	36%	11	1%
							openculture.com	4	36%	8	0%
							academia.edu	4	36%	6	0%
							chroniclevitae.com	4	36%	6	0%
							bbc.co.uk	4	36%	5	0%

The percentage of tweets containing links for award recipients from Modern Languages and Literature was slightly higher than those with hashtags, at 34.3%. This represents 2,005 tweets sent by 11 users which linked to 849 unique domains. As shown in Table 6, the top six most frequently linked domains were the same as the top six domains with the most users linking to them. Half of these were news websites (*cbc.ca*, *guardian.com*, *huffingtonpost.com*), two were social media sites (*youtube.com*, *facebook.com*) and one was a link shortening site (*trib.al*). The remaining top domains, both in terms of frequency of tweets and users, were also dominated by news sites. Thus the award recipients' linking behaviour seemed to mirror the current events focus of their hashtag usage. As with Historians, Twitter users in Modern Languages and Literature do not appear to be referencing academic sources in their links, although two of the domains with the highest number of Twitter users were academic-specific social media sites (*academia.edu*, *chronicleviate.com*) suggesting that some users do reference their professional identity on Twitter.

Education students had the highest share of link usage in their tweets, with almost half of the 794 tweets containing URLs. These were tweeted by 12 users and contained 192 domains. Similarly to other fields, social media and news sites made up the majority of the top domains by number of users (Table 7). Two link shortening sites were also included (*trib.al* and *ow.ly*). The domains that had overlap between the highest number of users and highest number of mentions included social media sites (*facebook.com*, *youtube.com*) and news sites (*cbc.ca*, *blogs.kqed.org*, *huffingtonpost.com*). The two most frequently tweeted domains were professional services websites, tweeted by only one user.

Table 7. Most popular domains of URLs used in Education.

most occurring						most users				
#	domain	occurrence *n=387*		user *n=12*		#	domain	user *n=12*		occurrence *n=387*	
		n	%	n	%			n	%	n	%
1	fivestepstohappy.com	41	11%	1	8%	1	facebook.com	4	33%	9	2%
2	howdogshelpkids.com	20	5%	1	8%		cbc.ca	4	33%	8	2%
3	youtube.com	12	3%	3	25%	3	youtube.com	3	25%	12	3%
4	facebook.com	9	3%	4	33%		theglobeandmail.com	3	25%	5	1%
	thetyee.ca	9	3%	1	8%	5	blogs.kqed.org	2	17%	7	2%
6	cbc.ca	8	2%	4	33%		huffingtonpost.com	2	17%	7	2%
	theatlantic.com	8	2%	1	8%		nytimes.com	2	17%	6	2%
8	blogs.kqed.org	7	2%	2	17%		trib.al	2	17%	5	2%
	huffingtonpost.com	7	2%	2	17%		chronicle.com	2	17%	4	1%
	slate.com	7	2%	1	8%		music.cbc.ca	2	17%	3	1%
							ow.ly	2	17%	3	1%
							news.nationalpost.com	2	17%	2	1%
							vimeo.com	2	17%	2	1%

News websites also dominate the top URLs sent by SSHRC Doctoral Award recipients from Political Science, both by frequency and number of users tweeting them. As shown in Table 8, seven of the top ten most frequently tweeted were news site domains reflecting local, national and international news (*theglobeandmail.com*, *thestar.com*, *nytimes.com*, *cbc.ca*, *theguardian.com*, etc.). The highest number of users for any URL was The Toronto Star (*thestar.com*) with nine users, occurring 80 times.

The link analysis reveals that the most frequently shared sources are not academic, but represent general news websites. However, it cannot be directly inferred from the domains whether or not the content of the shared news sites was irrelevant to the award recipients' field of study. Academically relevant topics could have been discussed based on their coverage in mainstream media. However, one can conclude from the link analysis that SSHRC Doctoral Award recipients hardly share academic resources and original research content such as academic articles, datasets or university websites.

Table 8. Most popular domains of URLs used in Political Science.

most occurring						most users				
#	domain	occurrence *n=3,142*		user *n=9*		#	domain	user *n=9*		occurrence *n=3,142*	
		n	%	n	%			n	%	n	%
1	theglobeandmail.com	141	4%	7	78%	1	thestar.com	9	100%	80	3%
2	thestar.com	80	3%	9	75%		nytimes.com	8	89%	78	3%
3	nytimes.com	78	3%	8	67%		washingtonpost.com	8	89%	52	2%
4	cbc.ca	74	3%	5	42%	4	theglobeandmail.com	7	78%	141	5%
5	youtube.com	66	2%	7	58%		youtube.com	7	78%	66	2%
6	practicalparticularism.tu	65	2%	1	8%	6	theguardian.com	6	67%	53	2%
7	theguardian.com	53	2%	6	50%		ottawacitizen.com	6	67%	24	1%
8	washingtonpost.com	52	2%	8	67%		facebook.com	6	67%	11	0%
	twitter.com	52	2%	5	42%	9	cbc.ca	5	56%	74	3%
10	theatlantic.com	39	2%	3	25%		twitter.com	5	56%	52	2%
							trib.al	5	56%	35	1%
							macleans.ca	5	56%	24	1%
							vox.com	5	56%	22	1%
							globalnews.ca	5	56%	11	0%
							telegraph.co.uk	5	56%	11	0%

3.3.3 Topic Analysis

In order to analyze the topics discussed by the SSHRC Doctoral Award recipients on Twitter beyond hashtags and URLs in more detail, a sample of tweets was coded intellectually (see Appendix A.2.6 for a detailed description of the sampling, coding scheme and inter-coder reliability). Table 9 provides an overview of the stratified random sample of 2,079 out of 15,302 original tweets, for which the content was coded into categories, which helped to identify tweets related to an award recipient's thesis, discipline or academic life in general. Based on the coding of the random sample of tweets sent by SSHRC Doctoral Award recipients from the main fields of History, Modern Languages and Literature, Education and Political Science, the majority (69.7%) of tweets were not

related to academic topics. One-quarter of tweets could be identified as referring to an award recipient's thesis, discipline or academic life in general. Comparing the two domains, award recipients from the Social Sciences tweeted more about academic topics (51.4% unrelated) than those from the Humanities (81.0%). On the level of main fields, the percentage of unrelated tweets was highest in Modern Languages and Literature and lowest in Political Science. Among tweets that contained academic topics, the majority referred to the discipline of the tweeter (20.7%), while less than 5% respectively referred to their thesis or academic life in general. It should be noted, that inter-coder reliability was 'almost perfect' according to Cohen's Kappa for the categories *discipline*, *academic life*, *unrelated,* and *unknown,* but only 'fair' for *thesis* (see Table 14 in Appendix A.2.6), which reflects the difficulty to determine whether tweets were related to the topic of the PhD based only on the tweet and the title of the PhD project funded by SSHRC. Humanities tweets were less often relevant to thesis or discipline, while the opposite was true for tweets from Social Science. Results for tweets from award recipients in Education were particularly striking as one third of tweets referred to the topic of their theses and 40.7% were relevant to their discipline. On the other hand, tweets from Modern Languages and Literature hardly referred to the award recipients' thesis topics. These differences in tweet content by the sender's field were found to be statistically significant. Although the majority of tweets were unrelated, comparing these results to findings of other studies, SSHRC Doctoral Award recipients seem to use Twitter more to discuss scholarly topics. For example, Holmberg and Thelwall (2014) reported that only 3.5% tweets of sampled researchers were scholarly, while Haustein, Bowman, Holmberg, Peters, and Larivière (2014) found a low overlap between the terms used by a group of astrophysicists in tweets with the abstract terms of their publications.

Table 9. Content analysis for a random sample of tweets.

Domain and main field	Number of tweets coded	Tweet relates to				
		thesis	discipline	academic life	unrelated	unknown
all fields	*2,079*	*4.4%*	*20.7%*	*4.9%*	*69.7%*	*4.6%*
Humanities	1,289	2.8%	10.5%	5.2%	81.0%	3.3%
History	753	3.9%	9.8%	6.0%	79.5%	4.6%
Modern Languages and Literature	536	1.3%	11.4%	4.1%	83.0%	1.3%
Social Science	790	7.1%	37.5%	4.4%	51.4%	6.7%
Education	59	33.9%	40.7%	0.0%	54.2%	5.1%
Political Science	731	4.9%	37.2%	4.8%	51.2%	6.8%

Among those tweets that contained a URL and were related to either the SSHRC Doctoral Award recipients' thesis, discipline or academic life in general, the type of link contents was further classified and categorized as blogs, general or academic event websites, news, publications, social network or other websites that were considered as general or academic (Table 10, Appendix A.2.6 for detailed coding methods). This analysis aimed at identifying the type of academic content that SSHRC Doctoral Award recipients distributed on Twitter. As shown in Table 10, one third of URLs in academic tweets linked to news websites, which reflects the results based on the analysis of most frequent domains (Section 3.3.2). These results emphasize that, although the websites are not primary academic sources, award recipients discuss academic topics covered in mainstream media and news. Among award recipients obtaining their PhD in History, news websites were less popular (21.6%).

Table 10. Classification of types of links for tweets related to thesis, discipline or academic life.

Domain and main field	Number of tweets coded	Tweet links to:							
		blog	event website		news	publication	social network	other website	
			general	academic				general	academic
all fields	*290*	*14.1%*	*2.1%*	*3.4%*	*32.1%*	*17.2%*	*14.8%*	*7.6%*	*8.6%*
Humanities	108	20.4%	0.9%	6.5%	26.9%	7.4%	14.8%	7.4%	15.7%
History	74	18.9%	1.4%	8.1%	21.6%	9.5%	13.5%	8.1%	18.9%
Modern Languages and Literature	34	23.5%	0.0%	2.9%	38.2%	2.9%	17.6%	5.9%	8.8%
Social Science	182	10.4%	2.7%	1.6%	35.2%	23.1%	14.8%	7.7%	4.4%
Education	24	20.8%	8.3%	8.3%	33.3%	8.3%	0.0%	4.2%	16.7%
Political Science	158	8.9%	1.9%	0.6%	35.4%	25.3%	17.1%	8.2%	2.5%

Overall, publications were the second most popular link category (17.2%). Among academic tweets with URLs, publications appeared more often in tweets by social scientists (23.1%) than humanists (7.4%) and were particularly often referenced by SSHRC Doctoral Award recipients from Political Science (25.3%), while they hardly played a role in tweets in Modern Languages and Literature (2.9%). It should be noted that the 50 of 290 academic tweets with links referring to publications represent 2.4% percent of the total 2,079 sampled tweets. This is in line with 3% direct links[5] to peer-reviewed resources reported by Priem and Costello (2010). This low percentage of tweets to publications is particularly interesting from the social media metrics perspective, where tweets or other social media events referring to scholarly articles are the focus of studies and tools such as Altmetric.com (see Sections 2.2 and 2.3 for an overview of the literature). This suggests that current altmetric indicators reflect only a small minority of academic discussions on Twitter.

Award recipients often linked to content from social network sites (14.8%) and blogs (14.1%), but there were differences between main fields. Education award recipients did not link to social network sites, and blogs were less popular among political scientists. Academic websites such as those of universities and academic institutions and websites of academic events including conferences and workshops were particularly popular among SSHRC Doctoral Award recipients from History (18.9%, 8.1%) and Education (16.7%, 8.3%), while they played a much smaller role in Modern Languages and Literature (8.8%, 2.9%) and Political Science (2.5%, 0.6%). The differences in link source between fields were found to be statistically significant with a moderate effect size, indicating that the Twitter user's field had some association with the type of links in academic tweets they were more likely to send.

It can be concluded that SSHRC Doctoral Award recipients use Twitter to discuss academic topics to a certain extent, although the majority of tweets are unrelated to their professional lives. The percentage of academic tweets was, however, higher than those reported by previous studies. Among tweets related to academic life in

[5] Priem and Costello (2010) identified almost the same amount of indirect, that is second-order links to peer-reviewed resources. These included, for example, links to blog posts that would link to peer-reviewed publications. They reported that a total of 6% of sampled tweets referred to peer-reviewed resources. The coding of the links in tweets by SSHRC Doctoral Award recipients was limited to first-order events.

general, the field of study or the thesis of the award recipient, news sites were the most popular resource that they linked to, followed by publications and blog posts. The popularity of news sites suggests that SSHRC Doctoral Award recipients targeted a more general audience when discussing academic topics on Twitter instead of the specialized academic community. Overall, academic publications were linked to in 2.4% of sampled tweets.

4 Conclusions

This section provides answers to the questions presented in the introduction of the report, based on two sources of evidence: a literature review on the use of social media by academics (Section 2) and an empirical analysis of the tweeting behaviour of SSHRC Doctoral Award recipients for the competition year 2010 (Section 3). A special emphasis is put on the opportunities and challenges of social media in the context of scholarly communication to support SSHRC in developing future guidelines for funded researchers and inform on the role of social media in academia.

Each of the six main research questions raised by SSHRC is answered by summarizing and interpreting information provided throughout the report. As shown in the literature review (Section 2), the uptake and use of particular social media platforms as well as the availability of derived metrics vary heavily among disciplines, which often prevent the generalization of findings. When available, the conclusions will therefore emphasize and focus on results related to Social Sciences and Humanities.

RQ.1 Opportunities and Challenges of Social Media

How can academics use social media in a research context? What are the opportunities and what are the challenges?

Four primary research-related activities conducted via social media were identified according to the literature review: making connections within and outside of the scholarly community as well as disseminating and discovering as well as discussing research using social media. These activities can be carried out using specific social media tools but often platforms combine more than one type of use. Goodier and Czerniewicz (2015) recommended researchers to take control of their online presence and focus on certain platforms, from which to expand strategically. The behaviour and use is often specific to the platform, including how academics choose to present themselves—either adopting a professional, personal, or mixed identity—as some social media platforms are public facing, while others are intended for specifically academic audiences.

The specific opportunities and challenges for researchers to connect with others, disseminate and discover as well as discuss research on social media are described in the following.

Connection

Collaboration is fundamental to scientific progress. Both international collaboration and interdisciplinarity have been found to increase scientific impact (Larivière, Haustein, & Börner, 2015). Especially for early-career researchers, it is crucial to connect with colleagues. A study of UK doctoral students found that most were optimistic that social media could aid them in making connections. More than two-thirds of survey participants obtaining their PhDs in Arts, Humanities and Social Sciences agreed that social media allowed for new collaborations and to learn from others (British Library et al, 2012).

Opportunities. Social media platforms are by design intended to facilitate conversations and interactions. They can thus help to forge new connections and strengthen existing ones within the research community and have

the potential to increase national, international and interdisciplinary collaboration. One-third of researchers surveyed by *Nature* found social media useful for attracting collaborators (Van Noorden, 2014). Before trying to connect to other researchers or an interested audience outside the scholarly community, Goodier and Czerniewicz (2015) recommend that researchers set up Google Alerts and ImpactStory and Altmetric.com profiles to track their current personal online presence as well as the visibility of their work. As there are many different platforms with similar functions, researchers need to prioritize and select the most suitable ones to connect with particular audiences. Users need to acknowledge that social media involves time commitment and it is thus recommended to use a few platforms properly rather than multiple platforms poorly (Goodier & Czerniewicz, 2015).

In order to connect with potential collaborators, researchers need to determine on which platforms their peers are active. Among specialized social networking sites, LinkedIn, ResearchGate and Academia.edu were identified as popular platforms for researchers from the Social Sciences, Arts and Humanities. However, most users seemed to use these specialized networking sites only if contacted, which suggests that researchers need to actively approach users to be noticed (Van Noorden, 2014). Facebook—used by 1.5 billion people representing half of the world's online population (Hope, 2015)—was used much more and more actively but rather in a private context and less in an academic one. It might thus make it a valuable platform to connect with colleagues to strengthen existing ties rather than to establish new connections. Arts, Humanities and Social Sciences exhibited the highest usage of social networking sites for research purposes among fields for UK doctoral students. One-third used academic social networking sites, while almost half used general social networking sites like Facebook for research (British Library et al, 2012), which might indicate a shift towards more professional usage of Facebook among younger researchers. Although social bookmarking platforms such as Mendeley and Zotero have social networking functions, they seem to be used less for making new connections (Mohammadi et al., In-press).

Twitter can be identified as particularly valuable for connecting and engaging with conference participants (Weller & Puschmann, 2011) by using a designated conference hashtag. Hashtags can also be used to find people tweeting about certain topics. As the majority of Twitter users and tweets are public and the tool is open for anyone to use, it represents the most promising social media channel to engage directly with an interested public beyond the closed scientific community. Using the Twitter account description to identify as a researcher and specify particular research interests helps to be found by other users and identifying professionally may have networking benefits (Lulic & Kovic, 2013). Many academics reveal their full names and identify professionally in their Twitter account descriptions (Bowman, 2015b; Chretien et al., 2011), and so do the majority of SSHRC Doctoral Award recipients (Section 3.1).

Social media platforms can also be used to directly involve the general public in research. On the one hand, they can be used to create samples and find interview partners for social science studies, for example via Facebook

(Côté, 2013), while citizen science projects like galaxy zoo[6] or eBird[7] are based on the participation of amateurs in data collection and classification.

Challenges. The main challenge in using social media to connect to other users is that usage is not universal and uptake is low among researchers and many similar platforms exist. Despite a high level of awareness, Twitter uptake is particularly low with approximately 10% of researchers using it professionally (Gu & Widén-Wulff, 2011; Pscheida et al., 2013; Rowlands et al., 2011; Van Noorden, 2014). Among UK doctoral students from the Social Sciences and Humanities, professional Twitter use was higher than in other fields, but passive use prevailed (British Library et al, 2012). Similarly, predominant passive use was identified for Academia.edu, ResearchGate and LinkedIn, where researchers logged in only when contacted directly (Van Noorden, 2014). This passive behaviour might make it more difficult to establish new connections.

Users are unevenly distributed in terms of discipline, age, academic status and geographical location. If researchers focus on connecting with colleagues via social media, they need to be aware that they are excluding those who are not present on these platforms. This might apply particularly to senior researchers and certain countries such as China and Iran, where Twitter and other platforms are blocked. Many researchers do not trust social media, and are skeptical of its added value given the time required to learn and utilize new platforms. Attitudes towards social media differ between country, age groups and academic rank or experience (Gruzd et al., 2012; David Nicholas et al., 2014).

Barriers to social media entry include the lack of incentives to use it, pushback from peers and negative consequences if used improperly. The absence of administrative and faculty support for the usage of social media means that usage is not recognized or rewarded and thus, absent personal volition, the adoption of social media may be seen as a time-wasting endeavour (Harley et al., 2010). However, some funding agencies have started to recognize alternative forms of scholarly output and impact (Higher Education Funding Council for England, 2011; Piwowar, 2013; Viney, 2013), which might motivate researchers to become more active on social media in the future. Some scientific cultures are more permissive than others regarding the use of social media. For example, Acord and Harley (2012) showed that researchers in neuroeconomics were sharing less research online and becoming more secretive about their work, as they feared that their ideas might be stolen. While some fields have clear cultural norms regarding accepted communication channels, others are more ambiguous, which can create obstacles for those who wish to adopt social media but experience a lack of buy-in from their peers (Acord & Harley, 2012; Cheverie, Boettcher, & Buschman, 2009).

Particularly in a professional context, there is the risk that social media might be improperly used as it can be very easy to broadcast something quickly and without giving adequate consideration to how it may be perceived (Chretien et al., 2011). Reputations can be easily damaged, and the repercussions can be far-reaching, including being fired (Berrett, 2010; Herman, 2014; Ingeno, 2013; Jaschik, 2014; Rothschild & Unglesbee, 2013). Therefore,

[6] http://www.galaxyzoo.org/

[7] http://ebird.org/content/ebird/

reputation management remains a critical issue for ensuring proper social media use. The need to avoid saying things which could be controversial or affect hiring committees, particularly given the absence of specific university policies to govern its use, is another reason why academics may be hesitant to use social media professionally (Grande et al., 2014).

Dissemination and Discovery

The dissemination and discovery of research on social media largely overlap, as they represent the same processes, but from the perspectives of the sending and receiving users.

Opportunities. Social media platforms are built to make it as easy as possible to engage in, produce and share contents with others. Goodier and Czerniewicz (2015) recommend that researchers share their publications, datasets and presentation slides on online repositories such as the Social Science Research Network, arXiv, SlideShare, Figshare and other institutional repositories and platforms relevant to their fields and highlight the importance of descriptive metadata including tags. Using site-specific affordances, such as hashtags on Twitter, to increase the discoverability of various forms of research outputs.

The advantages of using social media platforms to disseminate research findings include increased transparency and openness, possible reuse and the fact that they can reach interested audiences directly and without delay. Nielsen (2012) even suggested that social media could amplify collective intelligence speeding up research and discovery in previously unknown dimensions through hyperspecialization and having experts focus on micro tasks (e.g., Polymath projects[8]). More than half of UK doctoral students felt that social media could enable one to more quickly discover and filter quality research by crowdsourcing in their network (British Library et al, 2012). Researchers surveyed by *Nature* confirmed the usefulness of social media to share papers and datasets and to raise the profile of their work (Van Noorden, 2014). Among UK doctoral students, 41% of those in Arts, Humanities and Social Sciences shared references with other researchers via social media at least once a month (British Library et al, 2012). As a large number of recent journal articles are shared on social reference managers such as Mendeley, these platforms are useful for discovering relevant literature with the number of readers as a potential filtering tool, particularly considering that most Mendeley users saved publications to cite later (Mohammadi et al., In-press).

Journal articles from the Social Sciences in particular have the potential to attract a general audience on Twitter. They are disseminated at a greater rate than those from more technical disciplines, as they are more relevant to the general public and more easily understandable and relatable than complex technical issues (Costas et al., 2015; Haustein, Costas, et al., 2015). By using social media which is also used by the general public, researchers can communicate their findings directly with an interested audience beyond the scientific community. Twitter is the most promising tools regarding outreach to the society at large, while blogs provide room for more in-depth discussions, allowing researchers to get complex topics across to interested laymen (Bonetta, 2007; Puschmann,

[8] http://polymathprojects.org/

2014). Presenting their discipline to the public was among the most frequent motivations of bloggers surveyed by Mahrt and Puschmann (2014). Research blogging is particularly popular among male doctoral and graduate students in health-related fields (Shema et al., 2012). One-quarter of Arts, Humanities and Social Sciences doctoral students from the UK had passively used blogs, while 16% had actively contributed to them, which was the highest active use of blogs among disciplines (British Library et al, 2012). The usefulness of social media for publicity was confirmed by researchers surveyed by *Nature* (Van Noorden, 2014). Goodier and Czerniewicz (2015) particularly recommend blogging and reference managers to share research. However, the number of passive users mostly surpasses active contributors for various social media platforms, suggesting that engagement with research on social media more often takes the form of invisible, unmeasurable actions.

Even though the scholarly journal remains the most important communication channel, publications on social media such as blogs posts and shared datasets are gaining importance and are starting to be recognized by important funders such as the US National Science Foundation (Piwowar, 2013).

Challenges. The full potential of social media for scholarly communication has not been reached as researchers currently use the new channels mostly to disseminate traditional content such as peer-reviewed journal articles. This may be due to a lack of incentives and motivations, because, with a few exceptions, researchers are not credited for new types of research outputs. If the dissemination of research findings via social media is desired by research managers and funding bodies, rewarding it could increase the uptake among researchers.

As found by previous studies (e.g., Priem and Costello, (2010)) and corroborated by the analysis of SSHRC Doctoral Award recipients (Section 3), researchers hardly disseminate scholarly contents on Twitter and most use of social media is passive rather than active (British Library et al, 2012). Even though Social Science research is more frequently shared on social media in comparison to other disciplines, the great majority of recent papers never get mentioned on Twitter, Facebook or in blog posts (Costas et al., 2015; Haustein, Costas, et al., 2015). While new communication genres and tools develop, some having been presented as better alternatives to aspects of traditional publishing such as the peer review process, there is no research that conclusively quantifies the advantages of social media over traditional publishing. Moreover, researchers still consider social media less useful to attract funding or future employers (Van Noorden, 2014). The peer-reviewed journal article remains the central, most valued and most important channel of scientific communication and social media plays only a complementary role (Grande et al., 2014; David Nicholas, Watkinson, Rowlands, & Jubb, 2011). Interviews with academics regarding non-traditional publishing including open access and social media revealed that traditional publishing still carried the most importance in determining academic advancement, with no evidence that younger academics have been able to take advantage of new platforms. Indeed, pre-tenure academics faced the largest pressure to publish in high-impact journals and limit alternative forms of dissemination such as blogs (Harley et al., 2010). Concerns about using social media professionally were raised; researchers felt the need to avoid controversial topics, particularly given the absence of specific university policies to govern its use (Grande et al., 2014).

Within the mass of content produced on social media platforms, it might be challenging to gain users' attention with scholarly topics or find relevant contents. Curious and humorous studies and papers with funny titles or entertaining stories have been shown to receive a lot of attention on Twitter (Haustein, Peters, Sugimoto, et al., 2014; Neylon, 2014). Correlation studies reflect that papers which are popular on Twitter, the channel that is assumed to have great potential to reflect societal impact, do not reflect high scientific impact as measured by

citations (Costas et al., 2014; Haustein, Costas, et al., 2015). In searching for relevant studies and research data on social media, users need to bear in mind that anyone can contribute without any gatekeeping or quality control.

One of the most important limitations of using social media for scholarly communication is the plurality of tools as well as the underrepresentation of certain disciplines, geographical and cultural regions and age groups. Using these platforms to disseminate content thus entails excluding certain audiences, such as researchers in China and Iran, where Twitter is blocked. From the point-of-view of discovery these biases imply that research conducted by underrepresented user groups is less likely to be found on social media.

Discussion

Discussing research on social media implies that relevant content is diffused and discovered and that users have connected with others to discuss it. The opportunities and challenges described above are thus amplified when it comes to actively engaging in research-related discussions on social media. Additional opportunities and challenges are discussed below.

Opportunities. Social networks like Facebook and ResearchGate facilitate discussions by enabling users to ask questions and crowdsource answers by obtaining opinions from their network. Twitter allows users engage in dialogues by using the Twitter handle but the 140-character restriction of tweets hinders complex arguments. Blogs include the possibility to comment on single posts, representing a popular platform to discuss research topics. They provide the space for in-depths discussions as well as for commenting or critiquing published studies. Three-quarters of UK doctoral students from the Arts, Humanities and Social Sciences agreed that social media enabled them to communicate and discuss ideas with peers in their fields, while over half considered them as a way to improve their work by presenting ideas and writing informally first before submitting to a journal or presenting at a conference. Around one-quarter of UK doctoral students from the Arts and Humanities and Social Sciences reported that they commented on other people's research-related content on social media at least once a month and used a wiki to collaborate online, while 23% maintained their own blog and 29% contributed to someone else's (British Library et al, 2012). This demonstrates the potential of using social media to discuss scholarly work and emphasizes its benefits, particularly among early-career researchers. Social media platforms are considered as venues similar to academic conferences but allow ideas and research to be discussed regardless of geographic location.

Challenges. The general low uptake and even lower active engagement of scholars on social media represent the greatest challenges of discussing scholarly contents online. If there are not enough participants on particular platforms, comments may never be seen. Approaching researchers on social media to discuss their work also does not guarantee that they will reply using the same platform (Woolston, 2015). Discussions might cease or not develop at all if users fear negative consequences when they engage in controversial topics on social media (Grande et al., 2014). Along these lines, "trolling" is a social-media specific problem that could also provide a challenge in the context of scholarly communication.

RQ.2 Role of Funding Agencies

How can a funding agency, such as SSHRC, recognize and encourage the proper use of social media to communicate research findings?

There are several possibilities for a funding agency such as SSHRC to encourage the use of social media among researchers to establish and strengthen connections within and outside the scholarly community and to use these tools to disseminate and discuss research outputs. If a funding agency wishes to increase the importance of social media in scholarly communication and, in turn, support open science, it should recognize new forms of communication and actively encourage use of social media by supporting beneficial use and mitigating negative effects. Rather than being prescriptive, the answers to this research question are intended to inform SSHRC's next steps in developing a more formalized policy for the use of social media. More research, which analyzes the social media use of SSHRC-funded researchers beyond Twitter and the 2010 cohort of SSHRC Doctoral Award recipients, are needed.

Recognizing social media as new forms of communication

The current perception of researchers appears to be that funding organizations are, at best, ambivalent towards social media outputs, and, at worst, discount their value entirely. Researchers did not consider their social media profiles to be useful for attracting funding (30% not very useful, 34% not at all useful) (Van Noorden, 2014). Instead social media profiles were considered most valuable for raising the profile of their work in the research community. There is insufficient motivation presently for researchers to invest time into utilizing these platforms professionally (Dave Nicholas et al., 2015). Crediting researchers for their activity on social media may increase the motivation to use them. Therefore funding organizations need to communicate their desire to recognize and encourage the use of social media to researchers.

A first step for SSHRC and other research funding organizations to support open science and strengthen the role of social media in the scientific community would be to acknowledge new types of scholarly communication such as blog posts, review reports, openly shared datasets and software codes as valuable outputs and credit researchers for these contributions. The US National Science Foundation has taken a step into this direction and decided to recognize researchers' "products", rather than just publications (Piwowar, 2013). A similar decision was also reached by the UK Medical Research Council (Viney, 2013), while the UK Higher Education Funding Council has decided to consider "all kinds of social, economic and cultural benefits and impacts beyond academia" (Higher Education Funding Council for England, 2011, p. 4). The Wellcome Trust explored the opportunities of social media metrics for funders and suggested that for younger academics who have had less time to publish and accumulate traditional citations, the new metrics may provide an alternative way of assessing their impact (Dinsmore et al., 2014). Of course, acknowledging new forms of outputs and impacts does not replace traditional forms of scholarly communication. The peer-reviewed journal article, monographs and edited books remain central means of scholarly communication. Other forms of scholarly work could nevertheless be acknowledged and considered complementary to traditional forms in an evaluation context.

The second step is to identify the motivations for recognizing the use of social media in communicating research findings. Is communicating SSHRC-funded research to the public the primary goal? Is it disseminating SSHRC-funded research to other scientists in the hopes of elevating the profile of the work? Is it to strengthen and

increase SSHRC-funded researchers' networks in order to facilitate collaboration? To this end, the opportunities and challenges of social media as laid out in Research Question 1 of the conclusion should help guide SSHRC in defining its priorities. Once this is more tangibly defined, various different social media metrics can be examined and assessed.

Encouraging the use of social media

If SSHRC decides to recognize social media outputs, it should also encourage and support researchers in using it. The following provides an overview of the actions funding organizations might consider taking in order to foster an increase in proper social media use.

Create guidelines. Researchers have discussed their reluctance to use social media professionally for fear of negative consequences given the absence of institutional social media policies (Grande et al., 2014). In order to avoid improper use of social media and its repercussions, some professional organizations have put in place guidelines for social media best practices (Loeb et al., 2014). Another oft-repeated reason as to why academics do not utilize social media was the time required to learn how to use them and then do it (Dave Nicholas et al., 2015). To this end, funding organizations can reduce this barrier to use by creating guidelines with recommendations on which platforms to target and best practices for their usage.

Different platforms reach different audiences. Public platforms like Twitter, Facebook, blogs and Wikipedia can be used to share SSHRC-funded research with broad audiences. Blog posts for example, may be written by academics to comment or critique published research, or use them to discuss findings with the public (Bonetta, 2007; Mewburn & Thomson, 2013; Puschmann, 2014; Shema et al., 2015). Academically focused platforms such as ResearchGate and Academia.edu can act as a business card, and researchers could be encouraged to simply establish a presence on these platforms for contact purposes and listing research. Researchers have shown preferences for different type of information seeking on different platforms. For instance, when researchers used social media to discover funding opportunities, they were more likely to use Twitter than any other platform to do so (Van Noorden, 2014). As SSHRC already has an established profile on Twitter, it has an advantage in connecting with researchers and promoting the proper use of the platform and should consider focusing its efforts there. Based on the findings of the SSHRC Doctoral Award recipients on Twitter, uptake and use was not dependent upon discipline, and so it does not seem necessary that Twitter usage guidelines be discipline-specific but rather could apply to all recipients of SSHRC Doctoral Award. Identifying which platforms are beneficial to researchers for specific activities could help increase their adoption.

Diffuse information through social media. Many academic institutions already have their own social media accounts to disseminate research as a part of their communications strategy. SSHRC now requires publications resulting from its funded grants to be made openly available online as part of the *Tri-Agency Open Access Policy on Publications.* Prioritizing open access is part of a greater trend among funding organizations globally (Goodier & Czerniewicz, 2015) and using an organization's social media profile to broadcast publicly available research could be used to reach a greater audience. Diffusing relevant content on specific platforms might increase the motivation for funded as well as unfunded researchers to participate. The adoption of social media by funding organizations in this manner could be seen as a validation of their legitimacy by researchers, thus encouraging uptake.

Help researchers to connect. Skewed distributions of followers on Twitter and the number of tweets shows that the majority of SSHRC Doctoral Award recipients appear only in the periphery of the Twitter network. SSHRC could help them to increase their visibility and create connections on Twitter by introducing and promoting specific hashtags for funded researcher of specific research fields. The English (@SSHRC_CRSH) and French (@CRSH_SSHRC) SSHRC Twitter accounts already maintain "Universities and Colleges" lists with 521 and 129 Twitter accounts, respectively. A list of SSHRC Doctoral Award recipients or other groups of funded researchers might increase their visibility and help them to connect to each other. Other institutions, such as the London School of Economics and Political Science have created similar lists[9] by crowdsourcing, i.e. asking academics to submit which other academics they follow. A similar strategy could be used for other platforms that SSHRC identifies as relevant, for example specific groups on Facebook or Mendeley.

Educate researchers about social media metrics. Certain methods of linking to research on social media are more useful for ensuring that the events are captured by altmetric aggregators. Researchers should be aware that, if diffusing or discussing scholarly documents on Twitter, Facebook or in blogs for instance, linking to the publisher website and/or using the DOI is preferable. The importance of accurate metadata for online publications should also be emphasized. Academics on Twitter will link to secondary sources rather than publications directly if they are not open access (Priem & Costello, 2010). Since open access has been made a priority by SSHRC, it could encourage funded researchers to link to the publications themselves by making them aware of the benefits. It is currently too early to adopt social media metrics as a way to assess the impact of research and there is much research that still needs to be conducted before they could help research funders in identifying and quantifying the impact of research outputs (Dinsmore et al., 2014; Haustein, Costas, et al., 2015). However, if certain social media measures should be validated and applied in an evaluation context, the use of platforms from which they are derived will increase.

RQ.3 Potential of Social Media Metrics

What metrics based on social media activity related to scholarly documents or researchers currently exist, what do we know about their meaning and how can they be captured?

Even though traditional peer-reviewed publications remain crucial to the diffusion of knowledge, the scholarly communication system is becoming more diverse, open and transparent. The evaluation system needs to adapt to these changes. Altmetrics have been developed to overcome oversimplified evaluation methods that reduce scientific productivity and success to the number of journal articles and citations within them. They include various types of research products and broader types of the impact of research, potentially including impact on society. These metrics thus have the potential to make scholarly evaluation more complete by complementing

[9] http://blogs.lse.ac.uk/impactofsocialsciences/2011/09/02/academic-tweeters-your-suggestions-in-full/

traditional bibliometric indicators and including all types of research products (Piwowar, 2013). Typical examples of social media events which could be potentially used for metrics include tweets, mentions in blog posts, readership counts on Mendeley, posts, likes and shares on Facebook and recommendations and ratings on F1000. However, besides adding more numbers to the evaluation, it is not clear whether currently captured social media activity around scholarly content reflects scientific or societal impact; empirical studies have highlighted the biases and limitations of metrics derived from social media activity. Dinsmore, Allen, and Dolby (2014) concede that much more research is needed before social media data can be useful for funders. Initially, social media metrics had been discussed as better and faster filters than citations and, in turn, were considered as predictors of scientific impact. In the absence of a definitive correlation between the two, social media metrics are touted as complements to traditional metrics rather than replacements, while most researchers demand more qualitative studies to investigate the underlying processes to validate the usefulness of these metrics. Moreover, the lack of a conceptual framework around them makes it difficult to define precisely what these indicators are metrics of (Haustein et al., In-press)

One of the main limitations which applies to almost all social media metrics is the low uptake of social media by researchers along with the biases towards certain demographics such as age and academic status, scientific field and country. Metrics derived from certain social media platforms inherit the biases of the user community and exclude non-users. While citations are a fundamental element of scholarly communication, social media platforms are just entering the scientific community and it is unclear whether scientific norms are in effect (Haustein et al., In-press). Moreover, the landscape of social media tools is changing fast; functions that affect user behaviour change over time and there is no guarantee that current platforms will be around in the future.

Furthermore, metrics only capture social media use which leaves traces, excluding "invisible" activities, which, given the amount of passive users, might prevail on social media. Invisible use might include reading a document without saving, linking, sharing, or commenting on it, but also applies to technical limitations in capturing acts. The majority of altmetrics aggregators rely on certain identifiers such as DOIs to retrieve mentions of documents, completely ignoring documents without these identifiers as well as mentions that do not include them. This is particularly problematic in the Social Sciences and Humanities, where one-third of journal papers in the Web of Science do not have a DOI (Haustein, Costas, et al., 2015); even if documents have identifiers that are used for retrieving social media events, users do not necessarily include them (Zahedi, Bowman, et al., 2014).

The heterogeneity of social media metrics represents opportunities and challenges at the same time. For example, events currently used for social media metrics range from 140-character tweets to entire blog posts, from a like on Facebook to an expert report on F1000 and from a save on Mendeley to a Wikipedia citation. All of these platforms differ regarding level of engagement as well as active user groups, and even within platforms these can differ. For example, a tweet mentioning a journal article can be sent by its author for self-promotion, forwarded by a researcher, discussed by an interested layman or be recommended by a funding agency. Currently all these tweets are counted equally. Another limitation of social media metrics is that not all events reflect usage by humans but might be generated automatically by bots (Haustein, Bowman, et al., 2015). Although traditional bibliometric indicators are not free from manipulation, gaming on social media is easier than in scholarly publishing, where gatekeeping and quality control mechanisms are in place.

One major concern regarding currently captured social media metrics is that they are strongly tied to the availability of APIs and often the ease of data collection seems to prevail over significance. At the same time,

different retrieval strategies may lead to different outcomes, creating issues of data quality and consistency among aggregators. While it makes sense to focus on journal articles as they represent the most important form of traditional scholarly interaction, they only represent a minor fraction of scholars' interactions on Twitter (e.g., Holmberg and Thelwall (2014), Priem and Costello (2010); see also Table 10). Second-order events, such as tweets to news articles or blog posts mentioning a researcher's work, might be better suited to reflect broader interest.

Keeping this long list of limitations in mind, some metrics seem to have potential to capture certain types of use of research that goes beyond the community of citing authors. New forms of output are emerging which should be recognized as complementary to the traditional academic publication. These include blog posts, datasets and open peer review reports. The most promising indicator to capture broad academic use of research are user counts form social bookmarking and reference managers. Among reference managers that currently provide access to their usage data, Mendeley is by far the one with the most activity. Correlation studies showed that Mendeley reader counts closely aligns with article downloads and citations (Schlögl et al., 2014). The fact that recent documents accumulate a larger number of reader counts, which often exceed citations and are generated by a largely academic user population makes Mendeley an attractive source to measure academic use of documents. Since students represent a large share of Mendeley users, reader counts incorporate use that is potentially broader or has another focus than citations in peer-reviewed journal articles. It should be noted that the Mendeley coverage was particularly low for Humanities documents (Mohammadi & Thelwall, 2014), reflecting a potential field bias. Scholars might use other tools, not use reference managers at all, or documents in certain domains such as Humanities may take longer to acquire readers on Mendeley. Zotero might be an additional data source in the future[10].

General audience platforms seem appealing as a potential measure of societal impact, but due to their heterogeneous use and user groups, it is not yet possible to determine the various types of uses of scholarly content. It would be logical to assume that tools with general audiences, such as Twitter, Facebook and Wikipedia, could be used to ascertain the societal impact of research, the specifics of which types of actions should be measured and what they mean for "impact". For instance, while Twitter may be a valid communication strategy for communicating SSHRC-funded research to a broad audience, and statistics are relatively easy to capture for the platform, there is no agreed upon way of interpreting the impact reflected by these interactions. The high popularity of curious and humorous topics on Twitter, for example, questions the validity of the number of tweets to indicate scientific or societal impact of scholarly documents. This is in addition to the limitations inherent to each platform, as discussed throughout this report.

[10] https://www.zotero.org/blog/funding-for-altmetrics-research-and-expanded-api/

RQ.4 Twitter Uptake

To what extent do SSHRC Doctoral Award recipients already use Twitter?

With slightly less than one-third having a Twitter account, Twitter uptake among the 2010 SSHRC Doctoral Award recipients was higher than what was reported by most other studies. While Bowman (2015b) found almost the same percentage of Twitter accounts among 1,910 US professors, the share of award recipients with a Twitter account was still comparatively high, as most studies found that around 10% to 15% of researchers used Twitter (Grande et al., 2014; R. N. Procter et al., 2010; Pscheida et al., 2013; Rowlands et al., 2011; Tenopir et al., 2013; Van Noorden, 2014).

This high uptake might be caused by several factors, such as disciplinary and age differences, as well as the fact that Twitter use is generally increasing. The sample of Arts, Humanities and Social Sciences doctoral students from the UK surveyed for the Researchers of Tomorrow report (Carpenter, Wetheridge, Tanner, & Smith, 2012) represented the most comparable reference; 15% reported to use Twitter in the research context (British Library et al, 2012). Comparable to findings by other studies analyzing academics on Twitter (Bowman, 2015b; Chretien et al., 2011), most SSHRC Doctoral Award recipients used their full names and many included their professional identities in their account descriptions, mentioning academic disciplines, titles and universities. This means that other Twitter users could find and identify them as researchers even if they do not tweet actively. Being identifiable, their activity on Twitter could potentially influence the way they are perceived when applying for jobs, attempting to collaborate with other scholars, and applying for funding. Less than one-third of account descriptions specified the PhD topic, possibly because many had finished their PhD four years after the end of the funding period and were now working as assistant professors, lecturers, or in other professions. Fewer than 10% of Twitter account descriptions were strictly non-academic. The fact that two-thirds of all award recipients on Twitter used the account description to present themselves as academics, suggests that they consider it an academic social media tool.

RQ.5 Twitter Use

What is the tweeting behaviour of SSHRC Doctoral Award recipients? How much do they use it in an academic context?

As described previously, a considerable amount of Twitter users use the platform passively rather than actively. This was also reflected in the tweeting activity of SSHRC Doctoral Award recipients, where more than 80% of users tweeted only occasionally or less, while a small number of users contributed the majority of tweets. On average, 340 Twitter users were following the Twitter activity of award recipients, while they were following 244 accounts, even though the mean follower-following ratio per recipient was 0.86. This indicates that the distributions of followers and followees were also extremely skewed. In fact, these Pareto distributions are typical not only for social media activity but also in traditional scholarly communication, for example the number of papers per author or citations per paper. Because of these distributions, the majority of SSHRC Doctoral Award recipients appear only on the periphery of the Twitter network.

SSHRC Doctoral Award recipients tweeted on average slightly more than once per day during the 12-month period analyzed, and 10% of their tweets were retweeted by other Twitter users. Their use of Twitter affordances (user mentions, ULRs, retweets and hashtags) exceeded that of researchers found in other studies (Haustein, Bowman, Holmberg, et al., 2014), and there was no association between tweeting activity and the use of retweets, user mentions and hashtags. This suggests that the user of these Twitter specific affordances among SSHRC Doctoral Award recipients is not influenced by how often one tweets but rather may be due to personal preferences regarding whether or not to retweet, contact other users through mentions or to use hashtags. This is corroborated by the finding that personal initiative was the most influential factor in a researcher's social media use in Nicholas et al (2011). Each of the award recipients included user mentions in 72% of their tweets on average. This particularly high share of tweets including other user names indicates that SSHRC Doctoral Award recipients used Twitter not only as an information dissemination tool but mostly for direct communication with other users.

Almost half of the users' tweets contained a URL to link to external information, allowing them to overcome the limited space of 140 characters. Analyzing links tweeted by doctoral students from the four main fields of History, Modern Languages and Literature, Education and Political Science, few URLs were shared by more than one award recipient. The domain analysis revealed that certain websites that were referred to frequently by a number of users, however. These included mostly general news websites as well as Youtube and Facebook. SSHRC Doctoral Award recipients hardly linked to academic resources and original research content such as academic articles. However, the content analysis of a sample of tweets identified one-quarter as relevant to an award recipient's thesis, discipline or academic life in general. Of these academic tweets containing a URL, news websites were the most popular category of links, which suggests that, although they are not academic sources as such, mainstream news media were used by SSHRC Doctoral Award recipients to support and engage in academic discussions on Twitter. By linking to news articles rather than scholarly publications, SSHRC Doctoral Award recipients might have targeted a more general audience instead of the specialized academic community when discussing academic topics on Twitter.

Publications were the second most popular link category among the tweets referring to an award recipient's thesis, discipline or academic life, followed by links to social networking sites and blogs. Scholarly publications represented 17% of academic tweets with links and 2% of the total 2,079 sampled tweets, which is in line with the 3% direct links to peer-reviewed resources reported by Priem and Costello (2010) as well as findings by Holmberg and Thelwall (2014). This suggests that the Twitter metrics currently collected by data aggregators such as Altmetric.com and ImpactStory, which focus on counting the number of tweets to scientific publications, report on only a small share of academic discussions on Twitter, as well as of general discussions academics have on the platform.

SSHRC Doctoral Award recipients used hashtags in one-third of their tweets. Hashtags often represent particular topics and allow users to identify relevant tweets through the Twitter search function. In an academic context, hashtags are frequently used at conferences for networking (Weller & Puschmann, 2011). However this particular hashtag use occurred only rarely among SSHRC Doctoral Award recipients from the four main analyzed fields. Award recipients could increase their visibility among participants of relevant conferences by using designated conference hashtags to take part in discussions, regardless of whether they attend in person or not. Hashtags also allowed SSHRC Doctoral Award recipients to connect with one another on Twitter, as demonstrated in the network graphs. However, opposed to the shared websites, which were similar between the four analyzed main

fields, hashtags were more diverse. Historians, for example, connected through the *twitterstorians* hashtag, which was created specifically for this purpose.

RQ.6 Field Differences

How does Twitter use of SSHRC Doctoral Award recipients differ between fields?

Previous studies had highlighted differences in Twitter uptake and use among disciplines, which need to be accounted for, particularly in an evaluation context. Special emphasis was thus placed on comparing tweeting behaviour of SSHRC Doctoral Award recipients between the two domains Humanities and Social Sciences as well as main fields in order to identify and account for discipline-specific differences. The four main fields of History, Modern Languages and Literature, Education and Political Science were also examined in further detail as these fields had comparable numbers of Twitter users and the other fields were too sparsely populated for comparison. Most of the differences analyzed were not found to be statistically significant, indicating that the field of study does not determine Twitter uptake and tweeting activity for this particular population of SSHRC Doctoral Award recipients. Subsequently, there is no evidence from this study that SSHRC needs to develop discipline-specific Twitter policies for its Doctoral Award recipients. However, it needs to be emphasized that the study is based on the Twitter use of the 2010 cohort of SSHRC Doctoral Award recipients only and might not reflect the use of more recent cohorts or other researchers funded by SSHRC.

This section summarizes the findings where differences were statistically significant. There was a statistically significant association between domain and including discipline in Twitter account descriptions, with SSHRC Doctoral Award recipients from the Humanities being more likely to mention their discipline than those from the Social Sciences. Significant differences were also found between the four main fields, with award recipients from Education mentioning their discipline less in their Twitter profile than those from the other fields.

During the 12-month period analyzed, SSHRC Doctoral Award recipients from Political Science were found to tweet significantly more than those from Education as based on the absolute number of tweets. The median number of tweets per user was 37 in Education and 255 in Political Science. This needs to be taken into account if initiating communication on Twitter and in particularly in the context of evaluation. If Political Scientists tweet more, their research topics are also more likely to be discussed on Twitter, while the opposite is the case for award recipients from Education.

Analyzing the tweet content of 2,079 sampled tweets from SSHRC Doctoral Award recipients from the four main fields, statistically significant differences were found between fields with respect to tweeting about academic topics (including their discipline, PhD thesis or academic life in general). Award recipients from Education were more likely to tweet about their PhD thesis. Award recipients from History tweeted the most about academic life in general, although this only constituted 6% of the field's sampled tweets. Education and Political Science tweeted more about their fields of study than History and Modern Languages and Literature. Those from Modern Languages and Literature had the lowest percentage of tweets related to discipline, thesis or academic life. Interestingly, the tweets from award recipients in Political Science had the highest rate of ambiguity; that is, coders were unable to determine if the tweet was related to discipline, thesis or academic life for 7% of sampled

tweets. Given that there may be a large overlap between current events and this field specifically, this result is perhaps not surprising.

Focusing on the sampled tweets containing links and whose content was related to academic life, discipline or thesis, a statistically significant association was found between the field of SSHRC Doctoral Award recipients and the type of linked source, with a moderate effect size. This suggests that the award recipients' field had some association with the type of links in academic tweets they were more likely to send. For example, scholarly publications appeared more often in tweets by Social Scientists (23%) than Humanists (7%). Comparing the four fields, they were particularly often referenced by award recipients from Political Science (25% of sampled tweets), while they hardly appeared in tweets in Modern Languages and Literature (3%). Education award recipients did not link to social networking sites at all, and blogs and non-mainstream news sources were the least popular among Political Scientists. Academic websites, such as those of universities and academic institutions and academic events including conferences and workshops, were particularly popular among SSHRC Doctoral Award recipients from History (19%, 8%) and Education (17%, 8%), while they played a much smaller role in Modern Languages and Literature (9%, 3%) and Political Science (3%, 1%). Mainstream news media was by far the most preferred link type for all fields, ranging from 22% in History to 38% in Modern Languages and Literature.

Bibliography

Acord, S. K., & Harley, D. (2012). Credit, time, and personality: The human challenges to sharing scholarly work using Web 2.0. *New Media & Society*, *15*(3), 379–397. http://doi.org/10.1177/1461444812465140

Adie, E. (2009, February 11). Commenting on scientific articles (PLoS edition) [Nature.com Blog]. Retrieved from http://blogs.nature.com/nascent/2009/02/commenting_on_scientific_artic.html

Allen, H. G., Stanton, T. R., Di Pietro, F., & Moseley, G. L. (2013). Social media release increases dissemination of original articles in the clinical pain sciences. *PloS One*, *8*(7), e68914–e68914. http://doi.org/10.1371/journal.pone.0068914

Almind, T. C., & Ingwersen, P. (1997). Informetric analyses on the world wide web: methodological approaches to "webometrics." *Journal of Documentation*, *53*(4), 404–426. http://doi.org/10.1108/EUM0000000007205

Anonymous. (2005). Revolutionizing peer review? *Nature Neuroscience*, *8*(4), 397. http://doi.org/doi:10.1038/nn0405-397

Anonymous. (2006). Overview: Nature's trial of open peer review. *Nature*. http://doi.org/10.1038/nature05535

Bar-Ilan, J. (2013). Astrophysics publications on arXiv, Scopus and Mendeley: a case study. *Scientometrics*, *100*(1), 217–225. http://doi.org/10.1007/s11192-013-1215-1

Bar-Ilan, J., Haustein, S., Peters, I., Priem, J., Shema, H., & Terliesner, J. (2012). Beyond citations: Scholars ' visibility on the social Web. In *Proceedings of the 17th International Conference on Science and Technology Indicators Montreal Canada 58 Sept 2012* (Vol. 52900, pp. 98–109). Montreal. Retrieved from http://arxiv.org/abs/1205.5611

Berrett, D. (2010, February 26). ESU professor suspended for comments made on Facebook page. *Pocono Record*. East Stroudsburg. Retrieved from http://www.poconorecord.com/apps/pbcs.dll/article?AID=/20100226/NEWS/2260344

Bonetta, L. (2007). Scientists Enter the Blogosphere. *Cell*, *129*(3), 443–445. http://doi.org/10.1016/j.cell.2007.04.032

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

Bornmann, L. (2014b). Validity of altmetrics data for measuring societal impact: A study using data from Altmetric and F1000Prime. *Journal of Informetrics*, *8*(4), 935–950. http://doi.org/10.1016/j.joi.2014.09.007

Bornmann, L. (2014c). Which kind of papers has higher or lower altmetric counts ? A study using article-level metrics from PLOS and F1000Prime. *arXiv Print*. Retrieved from http://arxiv.org/abs/1409.2863

Bowman, T. D. (2015a). Differences in personal and professional tweets of scholars. *Aslib Journal of Information Management*, *67*(3), 356–371. http://doi.org/10.1108/AJIM-12-2014-0180

Bowman, T. D. (2015b, July). *Investigating the use of affordances and framing techniques by scholars to manage personal and professional impressions on Twitter* (Dissertation). Indiana University, Bloomington, IN, USA. Retrieved from http://www.tdbowman.com/pdf/2015_07_TDBowman_Dissertation.pdf

Brigham, T. J. (2014). An introduction to altmetrics. *Medical Reference Services Quarterly*, *33*(4), 438–47. http://doi.org/10.1080/02763869.2014.957093

British Library et al. (2012). *Researchers of Tomorrow, 2009-2011* (No. UKDA study number: 7029). Colchester, Essex: UK Data Archive. Retrieved from http://dx.doi.org/10.5255/UKDA-SN-7029-1

Carpenter, J., Wetheridge, L., Tanner, S., & Smith, N. (2012). Researchers of Tomorrow. Retrieved from http://www.jisc.ac.uk/publications/reports/2012/researchers-of-tomorrow

Chamberlain, S. (2013). Consuming Article-Level Metrics: Observations and Lessons. *Information Standards Quarterly*, *25*(2), 4. http://doi.org/10.3789/isqv25no2.2013.02

Cheverie, J. F., Boettcher, J., & Buschman, J. (2009). Digital Scholarship in the University Tenure and Promotion Process: A Report on the Sixth Scholarly Communication Symposium at Georgetown University Library. *Journal of Scholarly Publishing*, *40*(3), 219–230. http://doi.org/10.1353/scp.0.0044

Chretien, K., Azar, J., & Kind, T. (2011). Physicians on Twitter. *JAMA*, *305*(6), 566–568. http://doi.org/10.1001/jama.2011.68

Costas, R., Zahedi, Z., & Wouters, P. (2014). Do "altmetrics" correlate with citations? Extensive comparison of altmetric indicators with citations from a multidisciplinary perspective. *Journal of the Association for Information Science and Technology*, n/a–n/a. http://doi.org/10.1002/asi.23309

Costas, R., Zahedi, Z., & Wouters, P. (2015). The thematic orientation of publications mentioned on social media: Large-scale disciplinary comparison of social media metrics with citations. *Aslib Journal of Information Management*, *67*(3), 260–288. http://doi.org/10.1108/AJIM-12-2014-0173

Côté, I. (2013). Fieldwork in the Era of Social Media: Opportunities and Challenges. *PS: Political Science & Politics*, *46*(03), 615–619. http://doi.org/10.1017/S1049096513000565

Cronin, B. (2005). *The hand of science: academic writing and its rewards*. Lanham, Md: Scarecrow Press.

Cronin, B., Snyder, H. W., Rosenbaum, H., Martinson, A., & Callahan, E. (1998). Invoked on the web. *Journal of the American Society for Information Science*, *49*(14), 1319–1328. http://doi.org/10.1002/(SICI)1097-4571(1998)49:14<1319::AID-ASI9>3.0.CO;2-W

Das, A. K., & Mishra, S. (2014). Genesis of altmetrics or article-level metrics for measuring efficacy of scholarly communications: Current perspectives. *Journal of Scientometric Research*, *3*(2). http://doi.org/10.2139/ssrn.2499467

Desai, T., Patwardhan, M., & Coore, H. (2014). Factors that contribute to social media influence within an Internal Medicine Twitter learning community. *F1000Research*, *3*, 120–120. http://doi.org/10.12688/f1000research.4283.1

Dinsmore, A., Allen, L., & Dolby, K. (2014). Alternative Perspectives on Impact: The Potential of ALMs and Altmetrics to Inform Funders about Research Impact. *PLoS Biology*, *12*(11), e1002003.

Evans, P., & Krauthammer, M. (2011). Exploring the Use of Social Media to Measure Journal Article Impact. In *AMIA Annual Symposium Proceedings 2011* (pp. 374–381). American Medical Informatics Association.

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), e123–e123. http://doi.org/10.2196/jmir.2012

Faulkes, Z. (2014). The vacuum shouts back: postpublication peer review on social media. *Neuron*, *82*(2), 258–60. http://doi.org/10.1016/j.neuron.2014.03.032

Fausto, S., Machado, F. A., Bento, L. F. J., Iamarino, A., Nahas, T. R., & Munger, D. S. (2012). Research blogging: indexing and registering the change in science 2.0. *PloS One*, *7*(12), e50109–e50109. http://doi.org/10.1371/journal.pone.0050109

Ford, E. (2013). Defining and characterizing open peer review: A review of the literature. *Journal of Scholarly Publishing*, *44*(4), 311–326. http://doi.org/10.3138/jsp.44-4-001

Galligan, F., & Dyas-Correia, S. (2013). Altmetrics: Rethinking the Way We Measure. *Serials Review*, *39*(1), 56–61. http://doi.org/10.1080/00987913.2013.10765486

Galloway, L. M., Pease, J. L., & Rauh, A. E. (2013). Introduction to Altmetrics for Science, Technology, Engineering, and Mathematics (STEM) Librarians. *Science & Technology Libraries*, *32*(4), 335–345. http://doi.org/10.1080/0194262X.2013.829762

Goodier, S., & Czerniewicz, L. (2015). *Academics' online presence: a four-step guide to taking control of your visibility*. University of Cape Town. Retrieved from http://hdl.handle.net/11427/2652

Grande, D., Gollust, S. E., Pany, M., Seymour, J., Goss, A., Kilaru, A., & Meisel, Z. (2014). Translating Research For Health Policy: Researchers' Perceptions And Use Of Social Media. *Health Affairs*, *33*(7), 1278–1285. http://doi.org/10.1377/hlthaff.2014.0300

Grosseck, G., & Holotescu, C. (2011). ACADEMIC RESEARCH IN 140 CHARACTERS OR LESS. In *Conference proceedings of "eLearning and Software for Education" (eLSE)* (Vol. 2, pp. 84–94). Bucharest, RO: Universitatea Nationala de Aparare Carol I.

Groth, P., & Gurney, T. (2010). Studying Scientific Discourse on the Web Using Bibliometrics : A Chemistry Blogging Case Study. In *Proceedings of the WebSci10: Extending the Frontiers of Society On-Line*. Raleigh, North Carolina.

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

Gu, F., & Widén-Wulff, G. (2011). Scholarly communication and possible changes in the context of social media. *The Electronic Library*, *29*(6), 762–776. http://doi.org/10.1108/02640471111187999

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

Harley, D., Acord, S. K., Earl-Novell, S., Lawrence, S., & King, C. J. (2010). *Assessing the future landscape of scholarly communication: an exploration of faculty values and needs in seven disciplines.* Berkeley: The Center for Studies in Higher Education, Univ Of California Press. Retrieved from http://escholarship.org/uc/item/15x7385g

Haustein, S., Bowman, T. D., & Costas, R. (In-press). Interpreting "altmetrics": Viewing acts on social media through the lens of citation and social theories. In C. R. Sugimoto (Ed.), *Theories of Informetrics and Scholarly Communication*. Retrieved from http://arxiv.org/abs/1502.05701

Haustein, S., Bowman, T. D., Holmberg, K., Peters, I., & 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. http://doi.org/10.1108/AJIM-09-2013-0081

Haustein, S., Bowman, T. D., Holmberg, K., Tsou, A., Sugimoto, C. R., & Larivière, V. (2015). Tweets as impact indicators: Examining the implications of automated "bot" accounts on Twitter. *Journal of the Association for Information Science and Technology*. http://doi.org/10.1002/asi.23456

Haustein, S., Bowman, T. D., Macaluso, B., Sugimoto, C. R., & Larivière, V. (2014). Measuring Twitter activity of arXiv e-prints and published papers. In *Altmetrics14*. http://doi.org/10.6084/m9.figshare.1041514%20

Haustein, S., Costas, R., & Larivière, V. (2015). Characterizing social media metrics of scholarly papers: The effect of document properties and collaboration patterns. *PLoS ONE*, *10*(3), e0120495. http://doi.org/10.1371/journal.pone.0120495

Haustein, S., & Larivière, V. (2014). Mendeley as the source of global readership by students and postdocs? Evaluating Article Usage by Academic Status. In *IATUL Conference, Espoo, Finland, June 2-5 2014.* Retrieved from http://docs.lib.purdue.edu/cgi/viewcontent.cgi?article=2033&context=iatul

Haustein, S., Larivière, V., Thelwall, M., Amyot, D., & Peters, I. (2014). Tweets vs. Mendeley readers: How do these two social media metrics differ? *It - Information Technology*, *56*(5), 207–215. http://doi.org/10.1515/itit-2014-1048

Haustein, S., & Peters, I. (2012). Using social bookmarks and tags as alternative indicators of journal content description. *First Monday*, *17*(11), 1–28. http://doi.org/10.5210/fm.v17i11.4110

Haustein, S., Peters, I., Bar-Ilan, J., Priem, J., Shema, H., & Terliesner, J. (2014). Coverage and adoption of altmetrics sources in the bibliometric community. *Scientometrics*, *101*(2), 1145–1163. http://doi.org/10.1007/s11192-013-1221-3

Haustein, S., Peters, I., Sugimoto, C. R., Thelwall, M., & Larivière, V. (2014). Tweeting biomedicine: an analysis of tweets and citations in the biomedical literature. *Journal of the Association for Information Science and Technology*, *65*(4), 656–669. http://doi.org/10.1002/asi.23101

Haustein, S., & Siebenlist, T. (2011). Applying social bookmarking data to evaluate journal usage. *Journal of Informetrics*, *5*, 446–457. http://doi.org/10.1016/j.joi.2011.04.002

Haustein, S., Sugimoto, C., & Larivière, V. (2015). Guest editorial: social media in scholarly communication. *Aslib Journal of Information Management*, *67*(3). http://doi.org/10.1108/AJIM-03-2015-0047

Herman, B. (2014, September 5). Steven Salaita Twitter Scandal: University Offers Settlement, But Free Speech Questions Linger. *International Business Times*. Retrieved from http://www.ibtimes.com/steven-salaita-twitter-scandal-university-offers-settlement-free-speech-questions-linger-1678854

Higher Education Funding Council for England. (2011). *Decisions on assessing research impact. Research Excellent Framework (REF) 2014*. Retrieved from http://www.ref.ac.uk/media/ref/content/pub/decisionsonassessingresearchimpact/01_11.pdf

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

Hope, K. (2015, July 29). Facebook now used by half of world's online users. *BBC*. Retrieved from http://www.bbc.com/news/business-33712729

Huang, T., Davies, J., Joseph, S., & Wadyka, M. (2014). Measuring publication impact using article-level metrics (altmetrics). *Current Medical Research and Opinion*, *30*(S1), S9–S23. http://doi.org/10.1185/03007995.2014.896156

Ingeno, L. (2013). *Outrage over professor's Twitter post on obese students*.

Jaschik, S. (2014). Out of a job. *Inside Higher Ed*. Retrieved from https://www.insidehighered.com/news/2014/08/06/u-illinois-apparently-revokes-job-offer-controversial-scholar

Jordan, K. (2014). Academics and their online networks: Exploring the role of academic social networking sites. *First Monday*, *19*(3). http://doi.org/http://dx.doi.org/10.5210/fm.v19i11.4937

Knight, S. R. (2014). Social media and online attention as an early measure of the impact of research in solid organ transplantation. *Transplantation*, *98*(5), 490–6. http://doi.org/10.1097/TP.0000000000000307

Kwok, R. (2012). Altmetrics make their mark. *Nature*, *500*, 491–493. http://doi.org/10.1038/nj7463-491a

Larivière, V., Haustein, S., & Börner, K. (2015). Long-Distance Interdisciplinarity Leads to Higher Scientific Impact. *PLoS ONE*, *10*(3), e0122565. http://doi.org/10.1371/journal.pone.0122565

Lin, J., & Fenner, M. (2014). An analysis of Wikipedia references across PLOS publications. Retrieved from http://files.figshare.com/1546358/Altmetrics14_PLOS_v2.pdf

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), e259–e259. http://doi.org/10.2196/jmir.2707

Li, X., & Thelwall, M. (2012). F1000, Mendeley and traditional bibliometric indicators. In *Proceedings of the 17th International Conference on Science and Technology Indicators* (Vol. 2, pp. 451–551). Montreal: Citeseer. Retrieved from http://2012.sticonference.org/Proceedings/vol2/Li_F1000_541.pdf

Li, X., Thelwall, M., & Giustini, D. (2012). Validating online reference managers for scholarly impact measurement. *Scientometrics*, *91*(2), 461–471. http://doi.org/10.1007/s11192-011-0580-x

Loeb, S., Bayne, C. E., Frey, C., Davies, B. J., Averch, T. D., Woo, H. H., … Eggener, S. E. (2014). Use of social media in urology: data from the American Urological Association (AUA). *BJU International*, *113*(6), 993–998. http://doi.org/10.1111/bju.12586

Lulic, I., & Kovic, I. (2013). Analysis of emergency physicians' Twitter accounts. *Emergency Medicine Journal : EMJ*, *30*(5), 371–6. http://doi.org/10.1136/emermed-2012-201132

Mahrt, M., & Puschmann, C. (2014). Science blogging: an exploratory study of motives, styles, and audience reactions. *Journal of Science Communication*, *13*(3), A05.

Mas-Bleda, A., Thelwall, M., Kousha, K., & Aguillo, I. F. (2014). Do highly cited researchers successfully use the social web? *Scientometrics*, *101*(1), 337–356. http://doi.org/10.1007/s11192-014-1345-0

Mewburn, I., & Thomson, P. (2013). Why do academics blog? An analysis of audiences, purposes and challenges. *Studies in Higher Education*, *38*(8), 1105–1119. http://doi.org/10.1080/03075079.2013.835624

Mohammadi, E., & Thelwall, M. (2013). Assessing non-standard article impact using F1000 labels. *Scientometrics*, *97*(2), 383–395. http://doi.org/10.1007/s11192-013-0993-9

Mohammadi, E., & Thelwall, M. (2014). Mendeley readership altmetrics for the social sciences and humanities: Research evaluation and knowledge flows. *Journal of the Association for Information Science and Technology*, *65*(8), 1627–1638. http://doi.org/10.1002/asi.23071

Mohammadi, E., Thelwall, M., Haustein, S., & Larivière, V. (2015). Who reads research articles? An altmetrics analysis of Mendeley user categories. *Journal of the Association for Information Science and Technology*. http://doi.org/10.1002/asi.23286

Mohammadi, E., Thelwall, M., & Kousha, K. (In-press). Can Mendeley Bookmarks Reflect Readership? A Survey of User Motivations. *Journal of the Association for Information Science and Technology*, *In-press*. http://doi.org/10.1002/asi.23477

Mou, Y. (2014). Presenting professorship on social media: from content and strategy to evaluation. *Chinese Journal of Communication*, *7*(4), 389–408. http://doi.org/10.1080/17544750.2014.938669

Nández, G., & Borrego, Á. (2013). Use of social networks for academic purposes: a case study. *The Electronic Library*, *31*(6), 781–791. http://doi.org/10.1108/EL-03-2012-0031

Neylon, C. (2014, October 3). Altmetrics: What are they good for? Retrieved from http://blogs.plos.org/opens/2014/10/03/altmetrics-what-are-they-good-for/

Neylon, C., & Wu, S. (2009). Article-Level Metrics and the Evolution of Scientific Impact. *PLoS Biology*, *7*(11), e1000242. http://doi.org/10.1371/journal.pbio.1000242

Nicholas, D., Herman, E., Jamali, H. R., Osimo, D., Pujol, L., & Porcu, F. (2015). Analysis of Emerging Reputation and Funding Mechanisms in the Context of Open Science 2.0. Retrieved from http://www.ciber-research.eu/download/20150521-Reputation_Mechanisms-Final_report-JRC94952.pdf

Nicholas, D., Watkinson, A., Rowlands, I., & Jubb, M. (2011). Social Media, Academic Research and the Role of University Libraries. *The Journal of Academic Librarianship*, *37*(5), 373–375. http://doi.org/10.1016/j.acalib.2011.06.023

Nicholas, D., Watkinson, A., Volentine, R., Allard, S., Levine, K., Tenopir, C., & Herman, E. (2014). Trust and Authority in Scholarly Communications in the Light of the Digital Transition: setting the scene for a major study. *Learned Publishing*, *27*(2), 121–134. http://doi.org/10.1087/20140206

Nielsen, F. A. (2007). Scientific citations in Wikipedia. *First Monday*, *12*(8). http://doi.org/10.5210/fm.v12i8.1997

Nielsen, M. (2012). *Reinventing discovery: the new era of networked science*. Princeton; NJ: Princeton Univ. Press.

Piwowar, H. (2013). Value all research products. *Nature*, *493*, 159. http://doi.org/10.1038/493159a

Priem, J. (2010, September 28). I like the term #articlelevelmetrics, but it fails to imply *diversity* of measures. Lately, I'm liking #altmetrics. Retrieved from https://twitter.com/jasonpriem/status/25844968813

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

Priem, J., & Hemminger, B. M. (2010). Scientometrics 2.0: Toward new metrics of scholarly impact on the social Web. *First Monday*, *15*(7). Retrieved from http://pear.accc.uic.edu/ojs/index.php/fm/rt/printerFriendly/2874/2570

Priem, J., Piwowar, H. A., & Hemminger, B. M. (2012). Altmetrics in the wild: Using social media to explore scholarly impact. *arXiv Print*, 1–17.

Priem, J., Taraborelli, D., Groth, P., & Neylon, C. (2010, October 26). Altmetrics: A manifesto [Alternative metrics tool]. Retrieved from http://altmetrics.org/manifesto/

Procter, R. N., Williams, R., Stewart, J., Poschen, M., Snee, H., Voss, A., & Asgari-Targhi, M. (2010). *If you build it, will they come? How researchers perceive and use Web 2.0*. London, UK: Research Network Information. Retrieved from http://wrap.warwick.ac.uk/56246

Procter, R., Williams, R., Stewart, J., Poschen, M., Snee, H., Voss, A., & Asgari-Targhi, M. (2010). Adoption and use of Web 2.0 in scholarly communications. *Philosophical Transactions. Series A, Mathematical, Physical, and Engineering Sciences*, *368*(1926), 4039–56. http://doi.org/10.1098/rsta.2010.0155

Pscheida, D., Albrecht, S., Herbst, S., Minet, C., & Köhler, T. (2013). *Nutzung von Social Media und onlinebasierten Anwendungen in der Wissenschaft*. ZBW–Deutsche Zentralbibliothek für Wirtschaftswissenschaften–Leibniz-Informationszentrum Wirtschaft. Retrieved from http://www.qucosa.de/fileadmin/data/qucosa/documents/13296/Science20_Datenreport_2013_PDF_A.pdf

Puschmann, C. (2014). Opening Science. *Opening Science*, 89–106. http://doi.org/10.1007/978-3-319-00026-8

Robinson-García, N., Torres-Salinas, D., Zahedi, Z., & Costas, R. (2014). New data, new possibilities: Exploring the insides of Altmetric.com. *El Profesional de La Información*, *23*(4), 359–366. http://doi.org/10.3145/epi.2014.jul.03

Rothschild, S., & Unglesbee, B. (2013). *Kansas University professor receiving death threats over NRA tweet*. Retrieved from http://www.shawneedispatch.com/news/2013/sep/24/kansas-university-professor-receiving-death-threat/

Rousidis, D., Garoufallou, E., & Balatsoukas, P. (2013). Metadata Requirements for Repositories in Health Informatics Research : Evidence from the Analysis of Social Media Citations. In *Metadata and Semantics Research* (Vol. 390, pp. 246–257). Springer International Publishing. Retrieved from http://link.springer.com/chapter/10.1007/978-3-319-03437-9_25

Rousseau, R., & Ye, F. Y. (2013). A multi-metric approach for research evaluation. *Chinese Science Bulletin*, *58*(26), 3288–3290. http://doi.org/10.1007/s11434-013-5939-3

Rowlands, I., Nicholas, D., Russell, B., Canty, N., & Watkinson, A. (2011). Social media use in the research workflow. *Learned Publishing*, *24*(3), 183–195. http://doi.org/10.1087/20110306

Schlögl, C., Gorraiz, J., Gumpenberger, C., Jack, K., & Kraker, P. (2013). Download vs . citation vs . readership data : the case of an information systems journal. *Proceedings of the 14th International Society of Scientometrics and Informetrics Conference*, 626–634.

Schlögl, C., Gorraiz, J., Gumpenberger, C., Jack, K., & Kraker, P. (2014). Comparison of downloads, citations and readership data for two information systems journals. *Scientometrics*, *101*(2), 1113–1128. http://doi.org/10.1007/s11192-014-1365-9

Shema, H., & Bar-Ilan, J. (2014). Do Blog Citations Correlate With a Higher Number of Future Citations ? Research Blogs as a Potential Source for Alternative Metrics. *Journal of the Association for Information Science and Technology*, *65*(5), 1018–1027. http://doi.org/10.1002/asi.23037

Shema, H., Bar-Ilan, J., & Thelwall, M. (2012). Research Blogs and the Discussion of Scholarly Information. *PLoS ONE*, *7*(5), e35869–e35869. http://doi.org/10.1371/journal.pone.0035869

Shema, H., Bar-Ilan, J., & Thelwall, M. (2015). How is research blogged? A content analysis approach: How is Research Blogged? A Content Analysis Approach. *Journal of the Association for Information Science and Technology*, *66*(6), 1136–1149. http://doi.org/10.1002/asi.23239

Shuai, X., Pepe, A., & Bollen, J. (2012). How the scientific community reacts to newly submitted preprints: article downloads, Twitter mentions, and citations. *PloS One*, *7*(11), e47523–e47523. http://doi.org/10.1371/journal.pone.0047523

Smith, R. (1999). Opening up BMJ peer review. *BMJ*, *318*(7175), 4–5. http://doi.org/10.1136/bmj.318.7175.4

Sopan, A., Rey, P. J., Butler, B., & Shneiderman, B. (2012). Monitoring Academic Conferences: Real-Time Visualization and Retrospective Analysis of Backchannel Conversations. *2012 International Conference on Social Informatics*, (SocialInformatics), 62–69. http://doi.org/10.1109/SocialInformatics.2012.20

Sud, P., & Thelwall, M. (2013). Evaluating altmetrics. *Scientometrics*, *98*(2), 1131–1143. http://doi.org/10.1007/s11192-013-1117-2

Sud, P., & Thelwall, M. (2014). Not all International Collaboration is Beneficial : The Mendeley Readership and Citation Impact of Biochemical Research team size. *Journal of the Association for Information Science and Technology*, *In press*. Retrieved from http://www.scit.wlv.ac.uk/~cm1993/papers/InternationalCollaborationBioChemistryPreprint.pdf

Sweet, D. J. (2014). Keeping Score. *Cell Stem Cell*, *14*(6), 691–692. http://doi.org/10.1016/j.stem.2014.05.015

Taraborelli, D. (2008). Soft peer review: Social software and distributed scientific evaluation. *Proceedings of the 2008 International Conference on the Design of Cooperative Systems*, 99–110.

Tenopir, C., Volentine, R., & King, D. W. (2013). Social media and scholarly reading. *Online Information Review*, *37*(2), 193–216. http://doi.org/10.1108/OIR-04-2012-0062

Thelwall, M., Haustein, S., Larivière, V., & Sugimoto, C. R. (2013). Do altmetrics work? Twitter and ten other social web services. *PloS One*, *8*(5), e64841–e64841. http://doi.org/10.1371/journal.pone.0064841

Thelwall, M., & Kousha, K. (2014a). Academia.edu: Social network or Academic Network? *Journal of the Association for Information Science and Technology*, *65*(4), 721–731. http://doi.org/10.1002/asi.23038

Thelwall, M., & Kousha, K. (2014b). ResearchGate: Disseminating, communicating, and measuring Scholarship? *Journal of the Association for Information Science and Technology*, *66*(5), 876–889. http://doi.org/10.1002/asi.23236

Thelwall, M., & Maflahi, N. (2014). Are scholarly articles disproportionately read in their own country? An analysis of mendeley readers. *Journal of the Association for Information Science and Technology*. http://doi.org/10.1002/asi.23252

Thelwall, M., & Sud, P. (2015). Mendeley readership counts: An investigation of temporal and disciplinary differences. *Journal of the Association for Information Science and Technology*, n/a–n/a. http://doi.org/10.1002/asi.23559

Thelwall, M., Tsou, A., Weingart, S., Holmberg, K., & Haustein, S. (2013). Tweeting Links to Academic Articles. *Cybermetrics: International Journal of Scientometrics, Informetrics and Bibliometrics*, *17*(1), 1–8.

Thelwall, M., Vaughan, L., & Björneborn, L. (2005). Webometrics. *Annual Review of Information Science and Technology*, *39*(1), 81–135. http://doi.org/10.1002/aris.1440390110

Torres-Salinas, D., Cabezas-Clavijo, Á., & Jiménez-Contreras, E. (2013). Altmetrics: New Indicators for Scientific Communication in Web 2.0. *Comunicar*, *21*(41), 53–60. http://doi.org/10.3916/C41-2013-05

Tsou, A., Bowman, T. D., Ghazinejad, A., & Sugimoto, C. R. (2015). Who tweets about science? In *Proceedings of the 2015 International Society for Scientometrics and Informetrics*. Istanbul, Turkey.

Van Noorden, R. (2014). Online collaboration: Scientists and the social network. *Nature*, *512*(7513), 126–129. http://doi.org/10.1038/512126a

van Rooyen, S., Godlee, F., Evans, S., Black, N., & Smith, R. (1999). Effect of open peer review on quality of reviews and on reviewers' recommendations: a randomised trial. *BMJ*, *318*(7175), 23–27. http://doi.org/10.1136/bmj.318.7175.23

Viney, I. (2013). Altmetrics: Research Council Responds. *Nature*, *494*(176). http://doi.org/10.1038/494176c

Waltman, L., & Costas, R. (2014). F1000 Recommendations as a Potential New Data Source for Research Evaluation: A Comparison With Citations. *Journal of the Association for Information Science and Technology*, *65*(3), 433–445. http://doi.org/10.1002/asi.23040

Weller, K., & Peters, I. (2012). Citations in Web 2 . 0. In *Science and the Internet* (pp. 209–222). düsseldorf university press.

Weller, K., & Puschmann, C. (2011). Twitter for Scientific Communication : How Can Citations / References be Identified and Measured ? In *Proceedings of the ACM WebSci'11*. Koblenz, Germany.

Wilkins, J. (2008). The roles, reasons and restrictions of science blogs. *Trends in Ecology & Evolution*, *23*(8), 411–413. http://doi.org/10.1016/j.tree.2008.05.004

Wilson, M. W., & Starkweather, S. (2014). Web Presence of Academic Geographers: A Generational Divide? *The Professional Geographer*, *66*(1), 73–81. http://doi.org/10.1080/00330124.2013.765290

Woolston, C. (2015). Potential flaws in genomics paper scrutinized on Twitter. *Nature*, (521), 397. http://doi.org/10.1038/521397f

Wouters, P., & Costas, R. (2012). Users, Narcissism and Control — Tracking the Impact of Scholarly Publications in the 21 st Century. In *Proceedings of 17th International Conference on Science and Technology Indicators* (Vol. 2, pp. 847–857). Retrieved from http://2012.sticonference.org/Proceedings/vol2/Wouters_Users_847.pdf

Zahedi, Z., Bowman, T. D., & Haustein, S. (2014). Exploring data quality and retrieval strategies for Mendeley reader counts. Presented at the SIG/MET Workshop, ASIS&T 2014 Annual Meeting, Seattle. Retrieved from http://www.asis.org/SIG/SIGMET/data/uploads/sigmet2014/zahedi.pdf

Zahedi, Z., Costas, R., & Wouters, P. (2014). How well developed are altmetrics? cross-disciplinary analysis of the presence of "alternative metrics" in scientific publications. *Scientometrics*, *101*(2), 1491–1513. http://doi.org/10.1007/s11192-014-1264-0

Zahedi, Z., Fenner, M., & Costas, R. (2014). How consistent are altmetrics providers? Study of 1000 PLOS ONE publications using the PLOS ALM, Mendeley and Altmetric. com APIs. In *altmetrics 14. Workshop at the Web Science Conference, Bloomington, USA*. Retrieved from http://files.figshare.com/1945874/How_consistent_are_altmetrics_providers__5_.pdf

A Appendix

The Appendix provides detailed descriptions of the methods used for the literature review (A.1) and the analysis of Twitter use of the 509 SSHRC 2010 Doctoral Award recipients (A.2).

A.1 Methods: Literature Review

The literature review was initially conducted in November 2014 and included relevant publications found in the Web of Science, Google Scholar and arXiv, as well as relevant references cited within them.

The initial search for peer-reviewed papers was done in the Web of Science (WoS) by using the following search phrases under the topic field:

- TS="social media" metric* (130 results)
- TS=academic NEAR "social media" (127 results)
- TS="social media" AND citation* (41 results)
- TS=bibliometric NEAR "social media" (3 results)
- TS=((scholarly commun*) OR (scientific commun*)) AND social media (724 broad results, first 500 checked).

A Google Scholar search was then conducted using the phrase “scholarly research” + "social media" and the results parsed until page 20 of 51,000 pages. A brief search was also carried out in arXiv using the term “altmetric*”. These searches were supplemented by reviewing the bibliographies of papers already in the literature review, and adding relevant works. Each document in the results was assessed to determine its relevancy regarding the topics. Additional relevant documents were identified from the reference lists of these documents. A second update to the literature review was conducted after the first draft submission, and central documents which had been published after November 2014 were added. In total, over 200 publications were reviewed for relevancy, 139 of which were ultimately included in this report.

A.2 Methods: Twitter Data

This section describes the methods involved in identifying Twitter accounts for the 509 SSHRC 2010 Doctoral Award recipients (A.2.1), data retrieval via Twitter API including data transformation and definition of specific indicators (A.2.2), coding Twitter account descriptions (A.2.4) as well as tweet content (A.2.6). Statistical testing is described in A.2.3, while the social network analysis used to visualize the 2-mode networks of users and hashtags is explained in A.2.5.

A.2.1 Identifying Twitter accounts

Each listed grant recipient was initially searched for on Twitter using their first and last name encased in quotes. Quotations often ruled out different spellings of names, but can also limited the results in the instance of nicknames, shortened names, or the use of a middle initial. The approximately top 40 results were scanned for information to identify the person behind the account in the Twitter account description, including location, field

of study, and or academic status. If the list of results was short (approximately 10) then the results were clicked on to look for related clues among the tweets.

If the list of search results was too long or information provided on Twitter was not sufficient enough to determine if an account belonged to the particular grant recipient, a Google search was performed. Finding professional profile pages for the recipients helped to correctly identify the associated Twitter account. The profile pages were located by searching for the recipient's name and field of study. In most cases the recipients had graduated and were affiliated with another institution, either as a professor, researcher or postdoctoral student. A few of the recipients had Twitter accounts listed on their academic profile pages or on a personal website, which were found through a general Google search. Many of the recipients had an Academia.edu account on which their Twitter account was listed. From these profile pages and website, it was apparent that the recipient went by a nickname, or their middle name. If their preferred name was different than the name originally searched for on Twitter, a second search on Twitter was performed.

If Twitter accounts were not listed on personal or profile pages, in some cases Google searches helped to identify the particular Twitter accounts. Most often these types of searches were helpful for recipients with common names. Google searches included the name, the term 'Twitter', and various combinations of terms associated with their field of study, location, university name, or 'phd'.

If Twitter accounts did not have the name of the recipient associated with them, other pieces of information were used. This included comparing Twitter profile pictures of particular accounts with those from profile pages or personal website and scanning tweets, followers and followees for topics relevant to the recipient's field of study. The account was assigned to the recipient if either the profile picture or at least two pieces of information—such as a university department related to the recipients fields of study and location—matched the information found.

If Twitter searches, profile page searches, and more general Google searches did not lead to any results or the available information was not sufficient to determine whether the grant recipient had a Twitter account, the search was terminated. In this case the recipient was recorded.

A.2.2 Determining Twitter activity

The Twitter profile information and a sample of tweets from each account were collected on February 20, 2015. A PHP program was written employing the Twitter API to query and retrieve the tweets from each of the accounts and save the data as JSON files. The Twitter API is robust and allows a programmer to establish a secure connection with the Twitter application using oAuth standards in order to request specific data. For this phase of the data collection, up to 3,200 tweets were retrieved from each of the accounts. If an account had created private tweets or private messages, the Twitter API did not allow the retrieval of these tweets. In addition, the Twitter API places a limit on the tweets retrieved such that the most recent 3,200 tweets will be collected for any account where the number of tweets exceed 3,200.

The Twitter API method 'GET statuses/user_timeline' was used to retrieve the tweet content. The data returned for each tweet contained a large amount of information about both the tweet and the account holder. Each retrieved JSON record contained information that the account holder had added to his or her Twitter profile (e.g.,

link to photo, name, location, website, language, etc.) at the time of retrieval and information about the tweets themselves (e.g., if a tweet was retweeted, how many mentions, hashtags, URLs it contained, etc.).

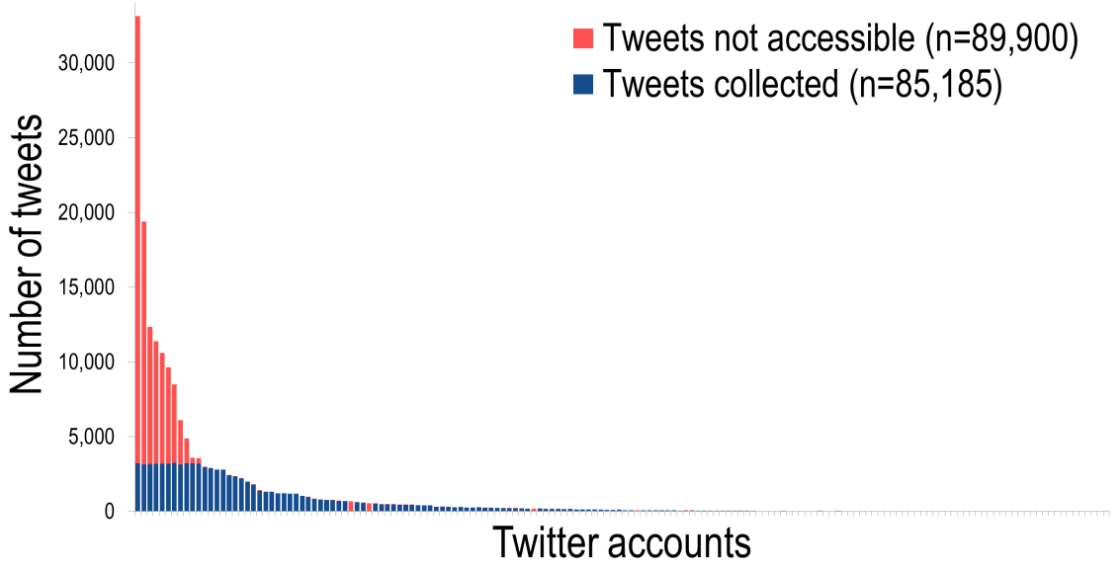


Figure 4. Number of tweets collected and not accessible via API.

The JSON files were then parsed using a second PHP script and the data was added to multiple MySQL database tables following a relational database model. The database schema included a separate table for the tweets, user profile information, hashtags (#), user_mentions (@), urls (long and shortened), media (photo or video files), and symbols (financial symbols).

Mean tweets per day

In order to compare tweeting activity, the mean number of tweets per day was computed based on the number of total tweets as provided in the account information divided by the number of days since the creation of the Twitter account. As shown in Figure 1, the distribution was a long tail distribution and thus there were many accounts with a lower number of tweets per day than the mean of 0.79 (median=0.12; mode=0). Because of this difference, the Twitter users were divided into six groups ranging from intense use to no use.

Retweets

Retweets, in general, are identified in Twitter by a specific convention that entails the use of "RT @username:" at the beginning of the tweet. To accurately reflect the number of retweets, all tweets beginning with the string "RT" were flagged as retweets in the database. The percentage of retweets sent per individual indicates the share of tweets that were not original but messages forwarded from other users. Retweets were excluded from coding tweet content in Section 3.3.3 to restrict the coding to original messages by the SSHRC-funded PhD students. An example of a retweet:

- "RT *@anonymized*: Media Advisory: Minister to Announce Funding for Renowned Ecological Site http://t.co/07Z2qOLJfA"

Received retweet counts

The Twitter API contains an attribute labeled 'retweet_count', which is the number of times a tweet has been retweeted, that is forwarded by other users. However, if the tweet itself is a retweet rather than an original tweet, then the count reflects the total number of times the original tweet (not made by the account holder) has been retweeted. If a user's tweet is not a retweet as defined above, the 'retweet_count' as provided in the Twitter JSON file indicates the number of times the user's original tweet (as made by the account holder) has been retweeted, that is forwarded by other users. This number indicates the popularity of the account holder's tweet on Twitter.

Hashtags

The Twitter API contains an attribute labeled 'entities', which contains the various built-in affordances available in a tweet (e.g., user mentions, hashtags, URLs, media, and symbols). One of these affordances is a list of hashtags (i.e., the pound sign followed by a keyword, e.g., #someTag) included in a tweet; this set of hashtags was stored in a hashtag table in the MySQL database and related to each tweet for each account holder. Based on this data, the number of hashtag occurrences was compiled for each account and each field. Because of the variety of languages and the case of the hashtags used (uppercase, lowercase, or mixed case), a PHP script using a regular expression was used to first remove any diacritic characters (e.g., ü or é) from the hashtags and second to make all hashtags lower case for comparison and counts. Note that Twitter itself does not distinguish between these various forms of hashtags, so that searching for "haiti" returns all results including "Haïti".

An example of a tweet containing hashtags:

- "Harper says new anti-terror laws are in the works http://t.co/A5nzr2GZxw #Canada #fascism"

Links

The URLs included in tweets are one of the Twitter API 'entities' provided as a list of URLs (http://www.domain.com/subdomain). This set of URLs was stored in a URL table in the MySQL database and related to each tweet for each account holder. Because Twitter is limited by 140 characters, users often either shorten their URLs using services (e.g., http://tinyurl.com/ or https://goo.gl/) or rely on Twitter's own built-in URL shortener (http://t.co). Therefore the URLs had to be expanded to obtain the destination domain. To accomplish this, a PHP script was written that first expanded the URL and then followed the link to the destination. Results from this process were stored in the URL database table and were subsequently used to determine the number of occurrences of a specific domain. In certain instances there existed multiple subdomains for a prominent website (e.g., mobile.nytimes.com, opinionator.blogs.nytimes.com, myaccount.nytimes.com, well.blogs.nytimes.com, parenting.blogs.nytimes.com, elections.nytimes.com, sinosphere.blogs.nytimes.com are all a part of the 'nytimes.com' domain), making it sometimes difficult to distinguish between parent domains and subdomains.

An example of a tweet containing URLs:

- "Écrivaine et professeure montréalaise d'origine haïtienne, Stéphane Martelly remporte Prix Michel Tournier jeunesse. http://t.co/L4Ky8vFBq2"

User mentions

User mentions (@username) included in tweets is another affordance listed in the Twitter JSON file retrieved via API. The set of user mentions was stored in a table in the MySQL database and related to each tweet for each account holder. In the report, user mentions are anonymized (displayed as *@anonymized*) except when they refer to public figures or organizations.

An example of a tweet containing a user mention of a public organization:

- "MLA report calls for Ph.D. program reform, including cutting time to degree @insidehighered http://t.co/FeKXCzoIo1"

An example of a tweet containing an anonymized user mention:

- *@anonymized* Thanks for joining in! :D

A.2.3 Testing for Statistical Significance Among Differences in Twitter Use

Tests for statistical significance in the differences between how domains and fields use Twitter were conducted.

Self-presentation based on Twitter account description

The association between the domains and fields of the SSHRC Doctoral Award recipients and the inclusion of information including PhD Topic, Discipline, Academic Title, and University in their Twitter account description were tested for statistical significance. The confidence level for significance was set at 95%, meaning that rejection of the null hypothesis occurred when $p < 0.05$.

Chi-square tests for association were conducted between the domain of the award recipients and inclusion of PhD topic, discipline, academic title and university in their Twitter account description, respectively. All expected cell frequencies were greater than five. There was a statistically significant association between domain and including discipline in Twitter bios ($\chi 2(1) = 5.462$, $p = .019$). An effect size of $\varphi = -.229$ was calculated. There was no statistically significant association between discipline and including the PhD topic ($\chi 2(1) = 2.214$, $p = .137$), academic title ($\chi 2(1) = 0.471$, $p = .492$) or university ($\chi 2(1) = 0.784$, $p = .376$).

Fisher's exact test for association was conducted between field of the award recipients and inclusion of PhD topic, discipline, academic title and university in their Twitter account description, respectively. Not all expected cell frequencies were greater than five. There was a statistically significant association between field and including the discipline in the Twitter account description ($p = .045$), but no statistically significant association between field and including PhD ($p = .548$), academic title ($p = .101$) or university ($p = .778$) according to the two-tailed Fisher's exact test.

Distribution of tweets sent over 12 months

A Kruskall-Wallis H test was run to determine if there were differences in the number of tweets sent over the past 12-month period between the fields History, Modern Languages and Literature, Education and Political Science. The Kruskall-Wallis H Test is a non-parametric test and was chosen to due to the non-normal distribution of the data and presence of outliers. The distributions of tweets were similar for all fields, as assessed by visual inspection of a boxplot. The confidence level for significance was set at the 95%, meaning that rejection of the null hypothesis (H_0: there is no statistically significant difference among the median number of tweets sent by field) occurred when $p < 0.05$. Median number of tweets was found to be statistically significantly different between groups, $\chi^2(3) = 8.589$, $p = .035$. Pairwise comparisons were performed using Dunn's procedure with a Bonferroni correction for multiple comparisons. Adjusted p-values are presented. This post hoc analysis revealed a statistically significant difference in the median number of tweets sent during the 12-month period between education (Mdn = 37.0) and political science (Mdn = 255.0) ($p = .028$), but not between any other field combination.

Affordance use

The differences in the use of retweets, hashtags, links (URLs), and user mentions among the different domains and fields were tested for statistical significance. Assumptions of normality and homogeneity of variances were first checked to determine the type of test required. The confidence level for significance was set at the 95%, meaning that rejection of the null hypothesis (H_0: there is no statistically significant difference among affordance usage) occurred when $p < 0.05$. For the two domains Humanities and Social Sciences as well as the fields History, Modern Languages and Literature, Education and Political Science the percentage of tweets containing the various affordances was analyzed for the 12-month period.

For the percentage of tweets containing one or more hashtags, award recipients from the Social Sciences did not display a normal distribution as assessed by Shapiro-Wilk's test ($p < .05$); however, inspection of a normal Q-Q plot indicated normality. There was homogeneity of variances, as assessed by Levene's test for equality of variances ($p = .794$). Therefore, an independent-samples t-test was run to determine if there were differences in percentages of tweets containing hashtags between Humanities and Social Sciences award recipients. Humanities Twitter users had a slightly higher percentage (M = 0.34, SD = 0.21) than Social Sciences users (M = 0.31, SD = 0.03), but this difference was not found to be statistically significant (M = 0.03, 95% CI [-0.05, 0.12], $t(113) = 0.797$, $p = .427$).

The percentage of tweets with hashtags per field was normally distributed, as assessed by Shapiro-Wilk's test ($p > .05$). Data is presented as mean ± standard deviation. The percentage of tweets with hashtags ranged from History (n = 18, .27 ± 0.18) to Modern Languages and Literature (n = 13, .35 ± .27), Education (n = 12, .30 ± .27) and Political Science (n = 10, .29 ± .22). There was homogeneity of variances, as assessed by Levene's test for equality of variances ($p = .375$). There were no statistically significant differences in percentage of tweets with hashtags between the different fields ($F(3, 49) = .264$, $p = .851$).

Regarding the percentage of tweets mentioning another Twitter, SSHRC Doctoral Award recipients from the Social Sciences did not display a normal distribution as assessed by Shapiro-Wilk's test ($p < .05$) which inspection of a normal Q-Q plot confirmed. Due to this violation of normality, a Mann-Whitney U test was run rather than

an independent-samples t-test to determine if there were differences in the percentage of tweets with mentions between Humanities and Social Sciences students. The distributions of tweets with mentions for both disciplines were visually assessed as being similar. Median percentage of tweets with a user mention was not statistically significantly different between the Humanities (mdn = 0.71) and Social Sciences (mdn = 0.76, U = 1719.00, z = 0.904, p = 0.366). The effect size was very small (r = 0.084).

For the analysis of tweets containing user mentions per field there were no outliers in the data, as assessed by visual inspection of a boxplot. The percentage of tweets with mentions was normally distributed, as assessed by Shapiro-Wilk's test (p > .05). The percentage of tweets with mentions ranged from History (n = 18, .76 ± 0.18), to Modern Languages and Literature (n = 13, .70 ± .22), Education (n = 12, .79 ± .17) and Political Science (n = 10, .80 ± .15). There was homogeneity of variances, as assessed by Levene's test for equality of variances (p = .477). There were no statistically significant differences in percentage of tweets with mentions between the different fields (F(3, 49) = .731, p = .539).

Analyzing the occurrence of retweets (RTs) per domain, neither Social Sciences nor Humanities displayed normal distribution for the percentage of RTs as assessed by Shapiro-Wilk's test (p < .05), which inspection of normal Q-Q plot confirmed. Therefore, a Mann-Whitney U test was run to determine if there were differences in percentage of tweets that were RTs between Humanities and Social Sciences award recipients. The distributions of retweets for both disciplines were visually assessed as being similar. Median percentage of retweets were not statistically significantly different between the Humanities (mdn = 0.085) and Social Sciences (mdn = 0.077, U = 1474.00, z = -5.08, p = 0.611).

Comparing fields, the percentage of RTs per user could not be determined as normally distributed, as assessed by Shapiro-Wilk's test (p < .05). A Kruskal-Wallis H test was run to determine if there were differences in the percentages of tweets with RTs in the fields of History, Modern Languages and Literature, Education and Political Science. Distributions of percentage of RTs were not similar for all fields, as assessed by visual inspection of a boxplot, and therefore the results could only be extended to differences in mean rank, rather than median. The mean ranks for percentage of tweets ranged from History (29.50) to Modern Languages and Literature (20.92), Education (25.83), and Political Science (31.80). The difference of mean rank for RTs was not statistically significant ($\chi 2(3) = 3.551$, p = .314).

Analyzing the use of links between domains, the Social Sciences did not display normal distribution as assessed by Shapiro-Wilk's test (p < .05); however inspection of normal Q-Q plot indicated normality. There was homogeneity of variances, as assessed by Levene's test for equality of variances (p = .083). An independent-samples t-test was run to determine if there were differences in the percentages of tweets containing URLs between Humanities and Social Sciences. Award recipients from the Humanities had a slightly lower percentage (M = 0.43, SD = 0.24) than those from the Social Sciences (M = 0.50, SD = 0.30), but this difference was not found to be statistically significant (M = -0.07, 95% CI [-0.18, 0.03], t(113) = -1.370, p = .173). The percentage of tweets with URLs was normally distributed among fields, as assessed by Shapiro-Wilk's test (p > .05). The percentage of tweets with URLs ranged from History (n = 18, .46 ± .27) to Modern Languages and Literature (n = 13, .40 ± .26), Education (n = 12, .50 ± .31) and Political Science (n = 10, .41 ± .24). There was homogeneity of variances, as assessed by Levene's test for equality of variances (p = .748). There were no statistically significant differences in percentage of tweets with URLs between the different fields (F(3, 49) = .336, p = .800).

Since neither domain nor field was found to be associated with affordance use, we hypothesized that tweet frequency may have an association with affordance use. A scatterplot was created, with percentage of tweets containing each affordance over the past 12 months on the y-axis and the number of tweets sent during the same time period on the x-axis. The scatterplots showed no discernible relation between the amount of Twitter activity as reflected in the number of total tweets and affordance use for each of the examined affordances, and thus no further statistical testing was conducted. The scatterplots are shown in Figure 5 below

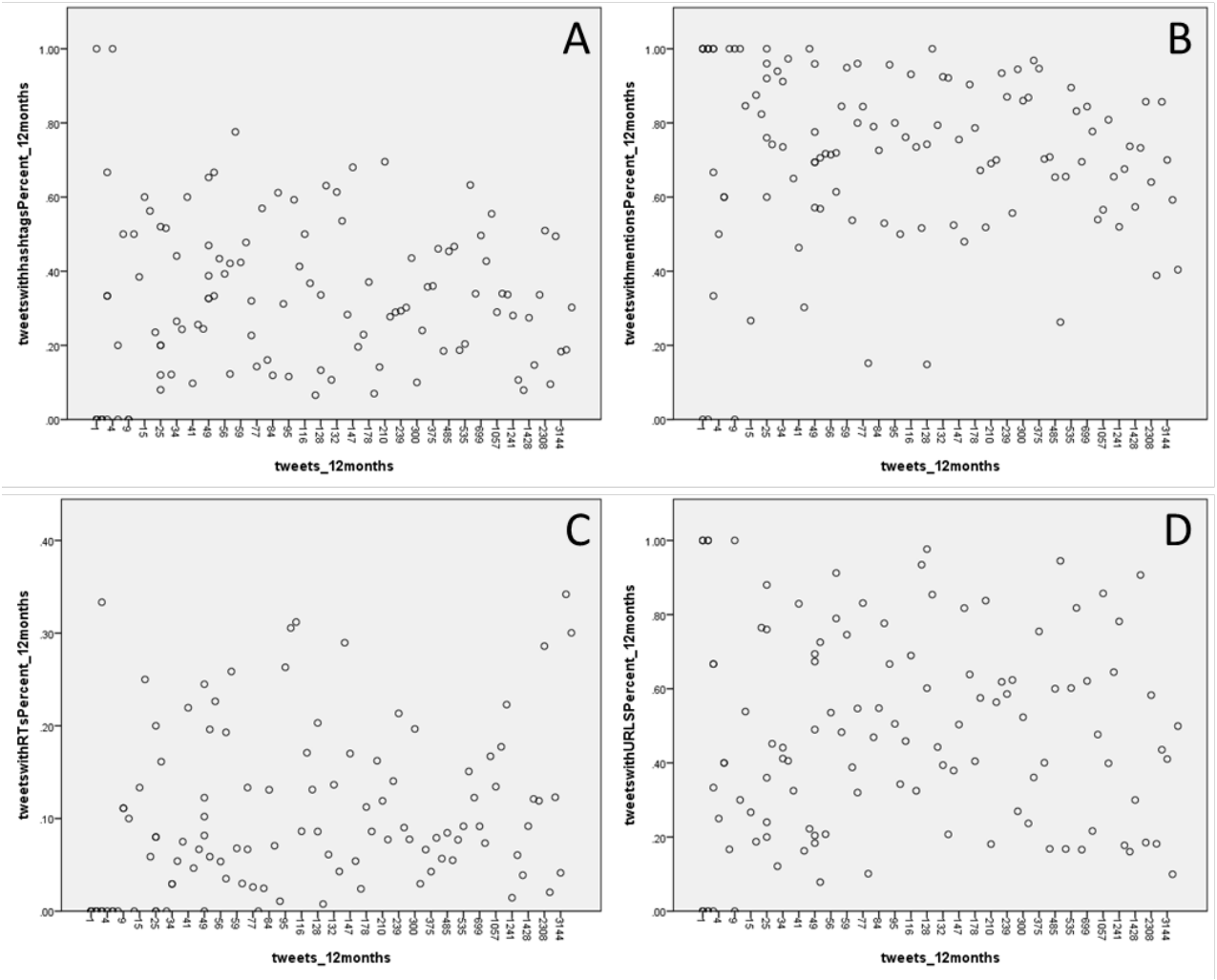


Figure 5. Scatterplots of the number of tweets during the last 12 months and the percentage of tweets with hashtags (A), user mentions (B), RTs (C) and links (D).

Coded Tweet Content

A random stratified sample of 2,079 tweets were coded based on whether their content was related to academics, their field of research, or their thesis topic (see Section A.2.6 for details). Tweets containing links and related to academics, field of research or thesis were further examined and their links classified based on the type of source they linked to. Differences among the type of tweet content as well as the type of source tweets linked to were tested for statistical significance by the Twitter user's field. The confidence level for significance was set at 95%, meaning that rejection of the null hypothesis (H_0: there is no statistically significant difference among affordance usage) occurred when $p < 0.05$.

Concerning whether a tweet's content was related to academic topics (including their discipline and thesis or academic life in general) or the sender's main field of study, Fisher's exact test for association was conducted between the field of the sender (Education, History, Modern Languages and Literature and Political Science) and the coded tweet content (related to academic life, related to field, not related to either, or unknown). Not all expected cell frequencies were greater than five. There was a statistically significant association between field and tweet content, $p < .000$, two-tailed Fisher's exact test. The effect size was small, Cramer's $V = 0.204$.

Regarding whether tweet content was related to the sender's PhD thesis (yes, no, or unknown), Fisher's exact test for association was conducted between field (Education, History, Modern Languages and Literature and Political Science). Not all expected cell frequencies were greater than five. There was a statistically significant association between field and whether content was thesis related, $p < .000$, two-tailed Fisher's exact test. The effect size was small, Cramer's $V = 0.196$.

Focusing specifically on tweets with links and whose content was related to academic life, field, or thesis, Fisher's exact test for association was conducted between the sender's domain (Humanities, Social Sciences) and the type of source linked to by the URL contained in the tweet (blogs and news items that are not published by mainstream news media; general events, meetings, workshops; scholarly conferences, e.g. conference websites, programs, event pages; mainstream news media including local, national and international newspapers, TV, radio, etc.; presentation slides, e.g. SlideShare; scholarly publications, e.g., scholarly journal articles, conference papers, books; social media, e.g., Facebook; general website not included in above categories; academic website). Not all expected cell frequencies were greater than five. There was a statistically significant association between domain and type of link source, $p < .000$, two-tailed Fisher's exact test. The effect size was moderate, Cramer's $V = 0.330$. Fisher's exact test for association was also conducted between the sender's field (Education, History, Modern Languages and Literature and Political Science) and the type of source linked to by the URL contained in the tweet. Not all expected cell frequencies were greater than five. There was a statistically significant association between field and type of link source, $p < .000$, two-tailed Fisher's exact test. The effect size was moderate, Cramer's $V = 0.268$.

A.2.4 Coding Twitter Account Descriptions

If available, Twitter account descriptions were coded according to five categories as shown in the coding scheme in Table 11. Based on the text provided in the Twitter bio compared to the information provided about the funded PhD student by SSHRC (main field, PhD thesis title, university), the self-presentation was categorized as strictly academic, strictly non-academic or including both, academic and non-academic content (Table 11A).

Examples of strictly academic self-presentations:

- "I am a PhD student in Political Science at Carleton University, currently working on my thesis on the politics of the future of sustainable energy."
- "Assistant professor of African American and African Studies at UC Davis"

Examples of both academic and non-academic self-presentations:

- "Teaching, researching (math ed), herding cats"

- "Mayor of Witless Bay. Memorial U Lecturer. Action Canada Fellow (2012-13). Ph.D. Candidate (Anthropology). New dad! Views expressed here are my own."

Examples of strictly non-academic self-presentations:

- "I'm just a guy on a bike. Aren't we all? Well, except for people like me who don't ride bikes."
- "I'm here to inspire you to Thrive, Shine, & Flourish! I love making music. Latest project: #TarFree613 music video! http://t.co/VszYKloKTR"

Table 11. Coding scheme for Twitter account descriptions.

A. account description is academic, non-academic or both	
acad	*self-presentation is strictly academic or professional without any non-academic or private content*
non-acad	*self-presentation is strictly non-academic, non-professional or private without any academic content*
both	*self-presentation contains both academic and non-academic contents*
B. account description contains topic of PhD of user	
yes	*description refers to PhD topic based on thesis title*
no	*description does not refer to PhD topic based on thesis title*
C. account description contains scientific discipline of user	
yes	*description refers to scientific discipline based on the main field*
no	*description does not refer to scientific discipline based on the main field*
D. account description contains academic title of user	
yes	*description includes academic title or status*
no	*description does not include academic title or status*
E. account description contains university of user	
yes	*description includes university*
no	*description does not include university*

In addition, descriptions were assessed as to whether they referred specifically to the PhD topic (as based on the thesis title), the scientific discipline (as based on the SSHRC main fields), contained an academic title (such as PhD, Postdoctoral Fellow, Professor, Lecturer etc.) or a university.

Example of an account description referring to the PhD topic:

- Political Science, thesis title: *Technologies of control: politics of economic and technological forecasting for sustainable energy*: "I am a PhD student in Political Science at Carleton University, currently working on my thesis on the politics of the future of sustainable energy."

Example of an account description referring to the discipline:

- Anthropology: "Future Research Partner and Social Anthropologist @Treefrog inc., Social Media Strategist, Phd Candiate YorkU Anthropology"

Example of an account description including an academic title:

- "Assistant Professor in the Department of Sociology at Trent University. Researching the socio-spatial dimensions of health and homelessness. All tweets my own."

Example of an account description including a university:

- "SSHRC postdoc researcher at UofGuelph, interested in self-injury, eHealth, social media, embodiment, narraive and research ethics"

A.2.5 Social Network Analysis of Users and Hashtags

To demonstrate the use and reuse of hashtags in the four fields analyzed in detail, History, Modern Languages and Literature, Education and Political Science, 2-mode networks of users and hashtags were created. A 2-mode network including connections between users and hashtags was generated based on the relational data described in Section A.2.2 separately for each of the four fields History, Modern Languages and Literature, Education and Political Science. Data was restricted to tweets published by PhD students from these fields within the 12 months before data collection (20 February 2014 to 19 February 2015) to ensure a certain stability of topics and hashtags. For each of the four fields networks were visualized in a 2-mode networks graph using UCInet and Netdraw. The spring embedding algorithm was used, which positions more central nodes—e.g., hashtags used by a large number of users and users mentioning a large number of hashtags used by other users—in the center and more specialized—hashtags with one user and users with few hashtags—on the periphery of the network graph, trying to find a global optimum for the network layout, Nodes are visualized as squares representing users and circles representing hashtags, where the size of circles indicates the number of times a certain hashtag was mentioned. The width of edges (i.e., connections between users and hashtags) are sized according to the number of times a certain user has mentioned the hashtag during the 12 months under analysis.

A.2.6 Coding Tweet Content

The sample for coding tweet content from SSHRC Doctoral Award recipients in History, Modern Languages and Literature, Education and Political Science was based on all original tweets published within the last 12 months (20 February 2014 to 19 February 2015) to allow for a stable and thus comparable timeframe. Retweets were excluded as the focus was on determining original content by the award recipients. Based on the 22,258 original tweets, a random stratified sample with a 95% confidence level and 2% confidence interval was determined (Table 12). This amounted to an overall sample of 2,079 tweets, selecting 14% of tweets from each of the four fields, i.e. 753 tweets from History, 536 from Modern Languages and Literature, 59 from Education and 731 from Political Science.

The sample was coded by two coders according to the coding scheme presented in Table 13. The scheme focused on determining whether the 2010 SSHRC Doctoral Award recipients used Twitter to discuss topics related to their thesis, field or general academic topics. Based on the information about applicants provided by SSHRC (i.e., SSHRC Main Field and title of the PhD thesis), coders tried to assess if a tweet was related to the research field (field); if it was unrelated to the field, they judged if the tweet discussed general academic topics (acad). Unrelated (unrel) tweets did not refer to the field or academic life in general. Topics of unknown tweets (unk) could not be clearly identified based on the tweet text or the link, if available.

Table 12. Stratified random sampling of original tweets in four selected fields.

Domain and main field	Applicants w/ Twitter account	Collected tweets (last 12 months)			
		all tweets	original tweets (w/out retweets)		
		n	n	stratified random sample confidence level=95% confidence interval=2%	
				n	%
all fields	*71*	*22,258*	*15,302*	*2,079*	*14%*
Humanities	42	12,908	9,485	1,289	14%
History	21	7,064	5,544	753	14%
Modern Languages and Literature	21	5,844	3,941	536	14%
Social Science	29	9,350	5,817	790	14%
Education	15	794	433	59	14%
Political Science	14	8,556	5,384	731	14%

Examples of tweets related to discipline:

- History: "Umm, the minister of justice in 1896 was A.R. Dickey. I wish I could think of a justice/pitching pun here."
- History: "Good morning, #medievaltwitter! I'm looking for mdv images of English royal justice, for teaching. Any suggestions? I have some FR ones."
- History: "Amazing Project! The importance of history. Quest for the sunken slave ship which claimed 664 lives | via @Telegraph http://t.co/gS9ymxhgVV"
- History: "I just wanted to read the article about George Etienne Cartier and the Senate, but that's behind a paywall, and that's not going to happen."
- Modern Languages and Literature: "if hedley existed when I started studying Shakespeare they would have inspired me to study Shakespeare b/c they write like Shakespeare"
- Modern Languages and Literature: "Effectivement un chef d'oeuvre, ce premier roman de #Frankétienne. "Quand Francketienne a Crevé" http://t.co/yBPfcQNVil #Haiti"
- Modern Languages and Literature: "You, Beloved, who are all / the gardens I have ever gazed at, / longing. - Rilke [You who never arrived]"
- Modern Languages and Literature: "Looking for a review I once read that complains about the fad of weakling heroes in the 1860s and cannot find it. Does it ring any bells?"
- Education: "Sometimes teaching is unteaching http://t.co/Cjc6iziDIa"
- Education: "Tech tools for problem based learning - am curious if these translate to younger grades http://t.co/HrdQ3kgIYs"
- Political Science: "ICYMI: What explains the patterns underlying the #TOpoli mayoral voting results map? #cdnpoli http://t.co/4INGz7hwGb http://t.co/Zy8PVToXm9"
- Political Science: "New issue of Telos on Secularism: http://t.co/IFbHG3dFVv"
- Political Science: "PC performance in Calgary-Elbow, relative to the rest of Alberta, 1971-2012. #ableg http://t.co/SceEJIXWLO"
- Political Science: "*@anonymized* I was wondering this yesterday - what's the best model to explain Hamas' behavior during the current crisis?"

Example of tweets related to academic life in general:

- “What a relief to be tidying up an article instead of a thesis for once. It's so...short!”
- “Good tips for surviving your PhD "Studying for a humanities PhD can make you feel cut off from humanity" http://t.co/8C1olgLqI2”
- “*@anonymized* This is how I choose to deal with 5 days of grading.”

Table 13. Coding scheme for tweet content.

A. tweet relates to:	
field	*related to field: topic can be generally associated with the person's field of study according to SSHRC Main Field*
acad	*related to academic life: topic does not concern the person's field of study but academic life in general, for example scholarly publishing, academic career, life as a researcher etc.*
unrel	*unrelated: does neither relate to the person's field of study nor to academic life in general*
unk	*unknown: topic cannot be clearly identified*
B. tweet relates to thesis:	
yes	*topic clearly relates to the PhD thesis as identified by thesis title*
no	*topic can be largely associated with the person's research field but does not relate to the PhD thesis as identified by thesis title*
unk	*unknown, topic cannot be clearly identified with PhD thesis*
C. type of source (if tweets is related to field, academic life or thesis and contains link):	
blog	*blogs and news items that are not published by mainstream news media*
event-gen	*general events, meetings, workshops*
event-acad	*scholarly conferences, e.g. conference websites, programs, event pages*
news	*mainstream news media (local, national and international newspapers, TV, radio)*
pres	*presentation slides, e.g. SlideShare*
pub	*scholarly publications, e.g., scholarly journal articles, conference papers, books*
social	*social media, e.g., Facebook, Google+*
site-gen	*general website (not included in above categories)*
site-acad	*academic website*

Examples of unrelated tweets:

- “Protesters Block Georgetown Streets over #Ferguson Decision | NBC4 Washington http://t.co/piX92zizFZ via @nbcwashington”
- “I'm caught up on Serial and it deserves the hype - it is great and very interesting. I have some thoughts though (of course there was a but)”
- “@anonymized and just in case you don't check facebook very often, have a fine birthday today!”

Examples of tweets with unknown topics:

- “@anonymized @anonymized Well, 25% is pretty good for the early 1990s!”
- “@anonymized @anonymized #mostlifealteringbook”
- “@anonymized אז אם כך, יש דברים וגב. ידוע לך אם הוא הגיע בטעות (חשה ונספח) או בכוונה? יש לבכל מאמצים לשחרר אותו? יש לך שם?”

Independently of the four categories mentioned above, coders tried to ascertain whether the tweet specifically referred to the topic of the sender’s PhD thesis. As only the title of the PhD thesis was available to access

relevancy, it was often difficult to determine, which is reflected in lower inter-coder reliability (fair agreement according to Cohen's Kappa, Table 14).

Examples for tweets related to thesis:

- History, *The rhetoric's of patronage: identity, job-seeking and the Canadian state, 1861-1896*: "I just wanted to read the article about George Etienne Cartier and the Senate, but that's behind a paywall, and that's not going to happen."
- History, *Giving and taking advice: mirrors for princes in medieval England and France*: "It's called "Counselling Charles VI of France: Christine de Pizan, Honorat Bovet, Philippe de Mézières, & Pierre Salmon." #CounselCVI"
- Modern Languages and Literature, *Victorian writing and the rhetoric of somatic corruption*: "Just learned of 2 excellent Victorian Studies websites by @JMU students: http://t.co/kZCH1lPqdV and http://t.co/yF9AzuvTI4"
- Modern Languages and Literature, *States of mourning: vacancies of law in Shakespeare's tragedies*: "@anonymized "if you deny me, fie upon your law" love that (if law doesn't satisfy MY will-to-power...) one of my fav lines"
- Education, *Authoring math: a study of middle school collectives as mathematics learners*: "Students use math, engineering for cardboard building projects http://t.co/RvgBwl4rBl"
- Education, *Children reading with a therapy dog and an adult mentor in a grade 2 classroom*: "Are you in the media and would like to cover a story about 125 children adopting dogs? Contact me today! http://t.co/2ZuM3O74pB"
- Political Science, *Controlling conception: the discursive construction of Canada's assisted human reproduction act*: "Quebec in vitro fertilization program to be scaled back - Montreal - CBC News http://t.co/yDZ756cSiW #IVF @anonymized @anonymized"
- Political Science, Institutional change in an age of judicial empowerment: reforms to the judicial appointment systems of Canada, Israel and Australia: "Intl. Commission of Jurists says fed. government "should review the law and practice for the appointment of judges" http://t.co/UfZK2QyVuu"

Type of source referred to the website linked to by a tweet and was only coded if the tweet contained a URL and referred to either the user's thesis, their field or academic life in general.

After the two coders had established and agreed upon the coding methods by discussing results for 100 of the 2,079 random tweets, another set of 100 tweets were coded independently by both coders to test inter-coder reliability. Cohen's Kappa was calculated based on the 100 independently coded tweets. As shown in Table 14 'almost perfect agreement' according to Cohen's Kappa was reached for A (tweet relates to: field, acad, unrel, unk) and C (type of source: blog, event-gen, event-acad, news, pres, pub, social, site-gen, site-acad), as coders agreed for 93 out of 100 tweets (A) and 9 out of 10 sources. Category B "relates to thesis" led to 15 disagreements out of 100 tweets, resulting in only fair agreement according to Cohen's Kappa. This reflects the difficulty of determining whether tweets were related to the topic of the PhD based on the tweet and the title of the PhD thesis only. The interpretation of category B should thus be conducted with care.

For the 200 of the 2,079 tweets that were coded by both coders, disagreements were discussed after computing inter-coder reliability to determine one result per tweet and category. If the two coders could not agree on one result was involved in the discussions and made the final decision.

Table 14. Inter-coder reliability for coding tweet content.

A. tweet relates to:

Cohen's Kappa: 0.84 almost perfect agreement
Percentage agreement: 93%

		coder 2				
	observed	field	acad	unrel	unk	
coder 1	field	16	0	1	1	18
	acad	0	2	0	1	3
	unrel	1	0	71	2	74
	unk	0	0	1	4	5
		17	2	73	8	100

Cohen's Kappa:	
< 0	poor agreement
0.01 – 0.20	slight agreement
0.21 – 0.40	fair agreement
0.41 – 0.60	moderate agreement
0.61 – 0.80	substantial agreement
0.81 – 1.00	almost perfect agreement

B. tweet relates to thesis:

Cohen's Kappa: 0.38 fair agreement
Percentage agreement: 85%

		coder 2			
	observed	yes	no	unk	
coder 1	yes	1	2	0	3
	no	1	80	8	89
	unk	2	2	4	8
		4	84	12	100

C. type of source (if tweets is related to field, academic life or thesis and contains link):

Cohen's Kappa: 0.86 almost perfect agreement
Percentage agreement: 90%

		coder 2								
	observed	blog	event-gen	event-acad	news	pub	social	site-gen	site-acad	
coder 1	blog	2	0	0	0	0	0	0	0	2
	event-gen	0	0	1	0	0	0	0	0	1
	event-acad	0	0	0	0	0	0	0	0	0
	news	0	0	0	5	0	0	0	0	5
	pub	0	0	0	0	1	0	0	0	1
	social	0	0	0	0	0	0	0	0	0
	site-gen	0	0	0	0	0	0	1	0	1
	site-acad	0	0	0	0	0	0	0	0	0
		2	0	1	5	1	0	1	0	10