

Final accepted manuscript (Feb 1, 2018):

Alperin, J.P., Gomez, C.J. & Haustein, S. (2018). Identifying diffusion patterns of research articles on Twitter: A case study of online engagement with open access articles. *Public Understanding of Science*. DOI: <https://doi.org/10.1177/0963662518761733>

Research Paper

**Corresponding Author:**

Juan Pablo Alperin, Harbour Centre Suite 3580, Simon Fraser University, 515 West Hastings Street, Vancouver, BC, V6B 5K3, Canada  
Email: [juan@alperin.ca](mailto:juan@alperin.ca)

# Identifying diffusion patterns of research articles on Twitter: A case study of online engagement with open access articles

## Juan Pablo Alperin

Juan Pablo Alperin is an Assistant Professor in the School of Publishing at Simon Fraser University, where he also directs the Scholarly Communications Lab. He uses a combination of computational techniques, innovative methods, and traditional qualitative methods to investigate how knowledge is produced, disseminated, and used. Read more at [scholcommmlab.ca](http://scholcommmlab.ca)

## Charles J. Gomez

Charles J. Gomez is an Assistant Professor of sociology at the City University of New York, Queens College. He studies the globalization of scientific knowledge by applying social network analysis and topic modeling to publication data. Read more at [charliegomez.com](http://charliegomez.com)

## Stefanie Haustein

Stefanie Haustein is an Assistant Professor in the School of Information Studies at the University of Ottawa. Her research focuses on scholarly communication, bibliometrics, altmetrics and open science and analyzes the role of social media in academia. Read more at [stefaniehaustein.com](http://stefaniehaustein.com)

## Abstract

The growing presence of research shared on social media, coupled with the increasing availability of freely available research, invites us to ask to whether scientific articles shared on platforms like Twitter diffuse beyond the academic community. We explore a new method for answering this question by identifying 11 articles from two open access biology journals that were shared on Twitter at least 50 times and analyzing the follower network of users who tweeted each article. We find that diffusion patterns of scientific articles can take very different forms, even when the number of times they are tweeted is similar. Our small case study suggests that most articles are shared within single connected communities with limited diffusion to the public. The proposed approach and indicators allows for automated can serve those interested in

the public understanding of science, science communication, or research evaluation to identify when research diffuses beyond insular communities.

## Keywords

Science communication, social media, altmetrics, public use of research, Twitter, diffusion, open access, social network analysis

## Introduction: Science, social media, and the public

Science has long been of public interest, as well as a site of intellectual and political contestation (Gauchat, 2012). However, it is only in recent years that social media—particularly Twitter, blogs and academic social networks such as ResearchGate—has become part of scientific communication, as researchers increasingly turn to online and social media platforms to diffuse and discuss scholarly work (Rowlands et al., 2011; van Noorden, 2014). Whereas specialized publication outlets (i.e., peer-reviewed scholarly journals) follow a unidirectional broadcast model targeted at the academic community, social media offers a new set of networked opportunities and new contexts for scientific communication from various types of people and institutions (Büchi, 2016; Nielsen, 2012; Veletsianos, 2016).

Since, by design, social media platforms facilitate conversations and create connections between users with a common interest, they also ought to increase the likelihood that scholarly work reaches new audiences, including the general public with an interest in research and scientific discoveries. In this way, social media platforms offer an online space where any given piece of research has an opportunity to reach beyond its intended audience (i.e., the scholarly community) by propagating through the social media network. Just as importantly, unlike peer-reviewed journals and conferences, in which only researchers are afforded the opportunity to communicate, social media provides opportunities to those outside the academic community to interact and engage with researchers in real-time. This provides a means for professional scientists and members of the public to directly interact with each other, and thus enables a two-way exchange between the scientific community and the public.

However, very little is known about how research literature diffuses on social media, what online communities it reaches, or whether these communities (e.g., academics and the public) interact with one another. In this paper, we seek to explore a novel approach to measuring and visualizing the diffusion patterns of primary research literature (i.e., peer-reviewed scientific articles) by studying the connections between individuals and organizations who have shared the same research article on Twitter. Specifically, we study the networks created by the follower and following relationships of Twitter accounts that shared at least one of eleven articles from the journals *BMC Biology* and *BMC Evolutionary Biology*. In studying these networks, we identify the existence of different network characteristics and diffusion patterns that we believe will serve to detect when research articles disseminate beyond the academic community or lead to exchanges with the general public.

Although limited in scope, our work offers initial insight into the role that primary research literature plays in the public's understanding and public engagement with science. To that end,

we offer a proof of concept as to how knowledge diffuses in the social media age, and how we might use the digital traces left on social media platforms to identify and assess the reach of research and scholarship among members of the general public. Our hope here is to start a wider dialogue amongst PUS researchers to engage with social media data and how research diffuses to lay audiences.

## Communities on Twitter: Reshares beyond researchers

### *How Research Moves between Twitter Audiences*

Scientists increasingly turn to social media outlets like Twitter to reach audiences beyond the confines of the ivory tower. Whereas other social media sites are designed for more interpersonal interactions (Facebook) or for sharing images (Instagram), Twitter is ideal because as a microblog, its designed around succinct prose, a natural extension for scientists. As such, it is also a natural extension of their current public teaching engagements with lay audiences (You, 2014). With a service like Twitter, scientists can readily reach audiences hundreds of times larger than a symposia or academic conference, allowing for their ideas and opinions to diffuse. Scientists across multiple fields and at different career stages are active on Twitter, with some possible overrepresentation of social and computer scientists and medical research more readily transcending the divide between academics and lay audiences (Ke et al., 2017).

Understanding this digital interface is important because little is known about *how* research moves from highly insular research communities of scientists to the general public. Indeed, diffusing scientific research beyond tight-knit academic fiefdoms is often difficult, in part due to how researchers and the public interface. Nevertheless, this interface is well documented in studies of scientific communication. Ludwick Fleck (1935) was one of the first to distinguish between academic communities—what he referred to as esoteric communities—and lay audiences that consume academic knowledge at a perfunctory level, or exoteric communities. Both sects engage in a dialogue which informs and delegates the dissemination (and eventual acceptance) of new knowledge. Over the past few years however, this interface has increasingly moved to social media. For instance, public audiences engage with scientific debates via their social media (Funk et al., 2017) and the Internet (Brossard, 2013; Brossard and Scheufele, 2013), in particular those with policy consequences (Myers et al., 2017; Vraga et al., 2015). While traditional media could serve as the nexus for lay audiences and scientists (Nisbet and Fahy, 2015), social media sidesteps the Fourth Estate in many ways. For instance, scientists can turn to Twitter to correct misinterpretations or distortion of scientific research by politicians, like in the case of climate change (Klemetti, 2017). More recent work on the relationship between scientists and Twitter has largely focused on linking usage with personal clout in the field (Eysenbach, 2011; Shuai et al., 2012; Thelwall et al., 2013; Weller and Puschmann, 2011). These efforts offer Twitter to supplement or complement traditional bibliometric measures, like citations. However, to date, we lack a proof of concept to even demonstrate this potential. Twitter and social network analysis could provide this opportunity.

### *Twitter, Social Network Analysis, and Research Diffusion*

Mapping and analyzing the structures of academic communities traditionally employs social networks analyses (SNA) of publication metadata. SNA is a multidisciplinary framework that

analyzes the structure of relations, called links, between entities in it called nodes, such as people (e.g., researchers), ideas (e.g., publications), and their online avatars (e.g., Twitter accounts) (Scott, 2017; Wasserman S. and Faust, 1994). For example, the relationships between researchers who coauthor on papers together can be transformed into a social network of researchers, connected by their joint-work on a published paper (i.e., coauthoring). The same logic can be extended to citations between journal articles, or collaborations on research grants (de Solla Price, 1965; Leydesdorff and Wagner, 2008; Moody, 2004; Wagner and Leydesdorff, 2005). With SNA, we can characterize both the larger geometric structure of the network with measures of how clustered (or clumped together) it is, and individual measures of how central certain people are relative to others in the wider structure. We can also measure how and why the network structure changes over time, as new ties are forged and dissolved. Networks of these data are often used as a proxy for the social interactions and relations between and amongst researchers within and across fields that define academic communities (*ibid*). Yet, little attention is paid to how research is broadcast, not only to disciplinary audiences (Fleck's esoteric circle), but to the general public at-large (exoteric circle).

Data from social media platforms like Twitter might show how ideas are diffused through their respective social networks that come to define their user-base (Goel et al., 2015). Whereas co-authorship and citation networks are provincially-oriented and bound within the esoteric sphere of an academic community, social media bridges the divide to reach lay audiences, as connections on social media require no *a priori* membership into the field aside from general interest. For instance, publishing an academic paper in a journal or conference is traditionally restricted to professional researchers, with limited opportunities for lay-people to engage in these conversations. Juxtaposed, a non-academic researcher can readily follow a university professor on Twitter due to interest in their research.

A social media platform like Twitter could serve as an interface between scientific and lay audiences. Twitter is an ideal venue for scientists to actively and widely engage with public audiences, as its users tend to be more educated than the overall population (Greenwood et al., 2016) and no one particular demographic group or geographic locale dominates its user-base (Mislove et al., 2011). It should be noted that this does imply that Twitter is representative of the general public, simply that its users come from a wide range of demographics.

As tweets and retweets of academic articles move through audiences that transcend the traditional venues for academic knowledge dissemination (e.g., conferences and journals that limit participation to specialists through gatekeeping practices), they facilitate increased awareness of research in the broader public sphere. The public not only has the opportunity to gain awareness that the research exists, but also access the knowledge contained therein, as an increasing number of full texts are freely available online through open access publishing models (Archambault et al., 2014; Piwowar et al., 2017).<sup>1</sup> Although Twitter is used by both members of the general public and in academia, the extent to which the microblogging platform does, in fact, bridge these two communities is little understood or explained.

### *Research Article Propagation Using Twitter: A Proof of Concept*

---

<sup>1</sup> As a point of reference, of 26% of online health seekers asked to pay for access, only 2% did so (Fox and Duggan, 2013).

Social media has created new opportunities and niches for non-academics to participate in scientific discourse or contribute as citizen scientists, and it has become an important tool for contemporary science communication (Bucchi, 2013; Nielsen, 2012; Veletsianos, 2016). To that end, we offer a proof of concept that gauges how various Twitter users consume and propagate scientific research that will further our understanding of its role in developing public awareness and understanding of science. As such, this study is a means to understand how science is circulated by various stakeholders and of how ideas diffuse through the relatively new frontier of social media.

Recent research has attempted to study the prevalence, volume, and meaning of sharing of research on various social media platforms (for a review of the literature see Sugimoto et al., 2017). Of the non-academic social media platforms, Twitter shows the highest activity related to scientific journal articles. On average, around 20% of recent publications are shared on Twitter, but large variations can be observed between publication years, disciplines, topics and journals as well as document types or national affiliation of authors (Alperin, 2015a; Costas et al., 2014, 2015; Hammarfelt, 2014; Haustein et al., 2015). Most work in this area (known as *alternative metrics*, or *altmetrics* for short), has concentrated on identifying the relationship between the number of times an article is shared on Twitter and the number of times an article is eventually cited, finding low but positive correlations between the two (Costas et al., 2014; Haustein et al., 2015). Some hope that altmetrics, and in particular research-related Twitter activity, has the potential to capture public interest and measure societal impact of research (Bornmann, 2014a, 2014b). However, studies found that the majority of scientific articles seem to be shared by academics instead of interested members of the public (Alperin, 2015b; Haustein, Tsou, et al., 2016; Tsou et al., 2015). It is therefore essential to look beyond counting the number of times articles have been shared and look towards both the meaning of the sharing (Haustein, Bowman, et al., 2016) and the demographics of the users behind the sharing.

This research analyzes these footprints for a small set of articles as a proof of concept that shows how SNA on social media data can be used to identify how scientific research moves through an online community and, in some cases, reaches the public. More specifically, based on a sample of open access articles in biology diffused on Twitter, this study shows how SNA can be used to answer the following research questions:

1. How do articles propagate through the follower network of Twitter users?
2. Are articles diffused within the same community or do they reach heterogeneous audiences?

## Data and methods

To analyze the diffusion and use of scientific articles, data about highly tweeted articles from the journals *BMC Evolutionary Biology* and *BMC Biology* were identified. The BMC journals were chosen because they are open access (i.e., freely available to the public) and cover topics that have the potential to be of interest to citizen scientists and to the public at large. Tweets “mentioning” BMC articles (i.e., contain a link to the article or have the article’s digital object identifier in the main text of the tweet) were obtained from Altmeteric LLC, a company tracking online activity around scholarly research outputs that has been continuously collecting tweets to scientific papers with DOIs through the Twitter firehose since 2011 and therefore

presents a valuable data source to analyze past Twitter activity referring to scientific journal articles.

We identified seven *BMC Biology* (Biol1-7) and four *BMC Evolutionary Biology* (Evol 1-4) articles published in 2014 which were tweeted or retweeted at least 50 times between the time of publication and June 30<sup>th</sup>, 2016 (Table 1A) (out of 294 articles from 2014 were tweeted at least once, from a total of 371 articles published across both journals that year). The threshold of 50 tweets was chosen to ensure a minimum number of users which allows for analyzing diffusion patterns and follower-following networks. All tweets mentioning these eleven papers were extracted from the Altmetric database, including the tweet ID, the user's Twitter handle, the date and time of the tweet, as well as the tweet content.<sup>2</sup>

The resulting dataset contained 1,590 tweets mentioning 11 articles. These tweets were sent by 1,287 unique Twitter users and contained 546 original tweets and 1,044 retweets (i.e., original tweets that users forwarded/retweeted). To analyze how these users were connected on Twitter, follower-following networks were constructed based on information obtained through the Twitter API. In order to collect follower-following information for a given Twitter user, the unique user identifier is needed, which is not included in the data obtained from Altmetric. Using the tweet IDs, these identifiers were retrieved were obtained by fetching the full tweet information from the Twitter REST API between November 23 and 29, 2016.<sup>3</sup> Using the *statuses/show* endpoint and the tweet ID from Altmetric, user information for 1,474 of the 1,590 tweets could be retrieved. The remaining 116 tweets were not available, either because the tweet or the user account had been deleted, or because the user account was private. To obtain the follower-following information for these users, the Twitter handle collected by Altmetric was used to query the user ID<sup>4</sup>. This approach yielded an additional 71 users, which, when combined with the users from the 1,474 tweets, resulted in 1,244 unique user IDs, corresponding to 97% of the users who tweeted the BMC articles in the sample.

These user IDs were required to reliably collect the followers and friends (i.e., following) of each of the 1,244 accounts that tweeted the 11 BMC articles. Collecting the list of users who each person follows, and the list of users who follows them (i.e., friends), it is possible to construct a network of connections between the users who shared articles. These lists of followers and friends were collected from the Twitter API (using the *followers/ids* and *friends/ids* endpoints) between November 23 and 29, 2016. These lists were then used to construct a directed follower/followee network for each of the eleven articles, where each node represents a single user who tweeted the article, and each arc represents a "following" relationship.<sup>5</sup> These networks therefore represent the established Twitter relationships that exist

---

<sup>2</sup> The choice of BMC journals and 50 tweet thresholds were arbitrarily chosen based on the logic explained above. The choices here were not intended to be representative of a wider universe of articles, but rather to demonstrate the kind of insight that the research method could yield.

<sup>3</sup> <https://dev.twitter.com/rest/public>

<sup>4</sup> Note that this is a less reliable way to collect the user ID than via the full tweet information, since it is possible, although unlikely, that the original user changed their screen name and that a new user has subsequently created an account with that name.

<sup>5</sup> In the majority of cases, the friends and follower information between two users was redundant, since they represent two sides of the same relationship, but both were used to fill in the data where a user's information could not be found.

between people who shared the same articles. The follower/friends network can provide a sense of whether users who share the same content are already aware of each other, or how many degrees of separation exist between them.

It should be pointed out that the follower/following relationships were collected as they appeared in November 2016, and not at the time of the original tweet (which, as described above, are no later than June 30<sup>th</sup>, 2016). It is expected, therefore, that since people follow and unfollow each other with some regularity (Myers and Leskovec, 2014), networks at the time of each tweet would have looked slightly different than the final resulting network, which we report on here.

We subsequently used the *python pandas* and *python igraph* modules<sup>6,7</sup> to calculate various network and diffusion statistics for each of the 11 networks. These statistics include the number of nodes, network density, network diameter, size of the largest weakly connected component, mean shortest path length, proportion of tweets versus retweets, time between the first and last tweet, and the tweet half-life.<sup>8</sup> These indicators help to characterize and compare diffusion patterns and network structures of the 11 articles. The description of the indicators, along with the ways in which each informs our understanding of diffusion and network structure, are explained in more detail in the appendix Table 2A.

To visualize the follower/followee networks, the Gephi software package was used by loading the list of users as nodes and the list of follower relationships as arcs. The Yifan Hu layout algorithm with an optimal distance between 1,000 and 1,300, relative strength of 0.2, initial step size of 20, step ratio of 0.95, quadtree max level of 10 and Theta of 1.2 was used. In the network graphs nodes are labeled with the screen name of the user and colored according to the order in which they tweeted from light to dark green (lighter nodes shared the article earlier than the darker ones). Node size corresponds to the number of followers who shared the same article, that is, the number of incoming arcs (in-degree) in the network graph. Three of the resulting networks can be seen in Figures 2-4, and all can be found in high resolution and with their corresponding Gephi files in the supplementary materials.

## Findings

The following results are based on the follower relationship between 1,240 unique Twitter users who shared seven BMC Biology and four BMC Evolutionary Biology articles published in 2014. Networks are analyzed separately and compared in terms of network structure and temporal diffusion patterns using the indicators described above.

### Resulting network structure

Analyzing the structure of the respective follower networks, we find that the great majority of users for each of the 11 articles were loosely connected with one another in the largest component. On average, 86.9% of all users tweeting an article were part of the largest (or giant) component (Figure 1), which means that majority of users sharing the same article were already

---

<sup>6</sup> <http://pandas.pydata.org/>

<sup>7</sup> <http://igraph.org/python/>

<sup>8</sup> The code used to calculate these statistics is available at [https://github.com/ScholCommLab/PUS\\_paper\\_2018](https://github.com/ScholCommLab/PUS_paper_2018)

connected to one another through common followers; for 5 of the 11 articles, the number of users connected to one another was over 90%. At the same time, a high percentage of users in the largest component indicates that only a small minority of users were isolated and had no connection to the core group of Twitter users. For 10 out of the 11 articles, less than one fifth of users tweeting about it were isolated without any or just a few connections to other users interested in the same paper. In fact, except for the occasional dyad (two connected users) or triad (three connected users), the majority of users outside the largest component were isolates. This means that within their personal Twitter follower network, they were the only one sharing the article. *Biol6* represents an exception to the strong concentration of users in the largest component, as it contains only 68.5% of users. The remaining Twitter users were, however, hardly connected, as 30% were isolates.

The extent to which users were following each other, varied between articles, with some networks having many users directly following each other, and others having only sparse ties between them. One way to observe this variance is by looking at the density of each network, where articles with lower density exhibit relatively sparse networks (i.e., those that tweeted the articles do not tend to all follow each other). Density in the 11 networks ranged from 0.01 to 0.14. The highest density was obtained by *Biol5*, an article that reports about the Earth Microbiome Project. The network graph shows that the core consists of well-connected users who follow each other.

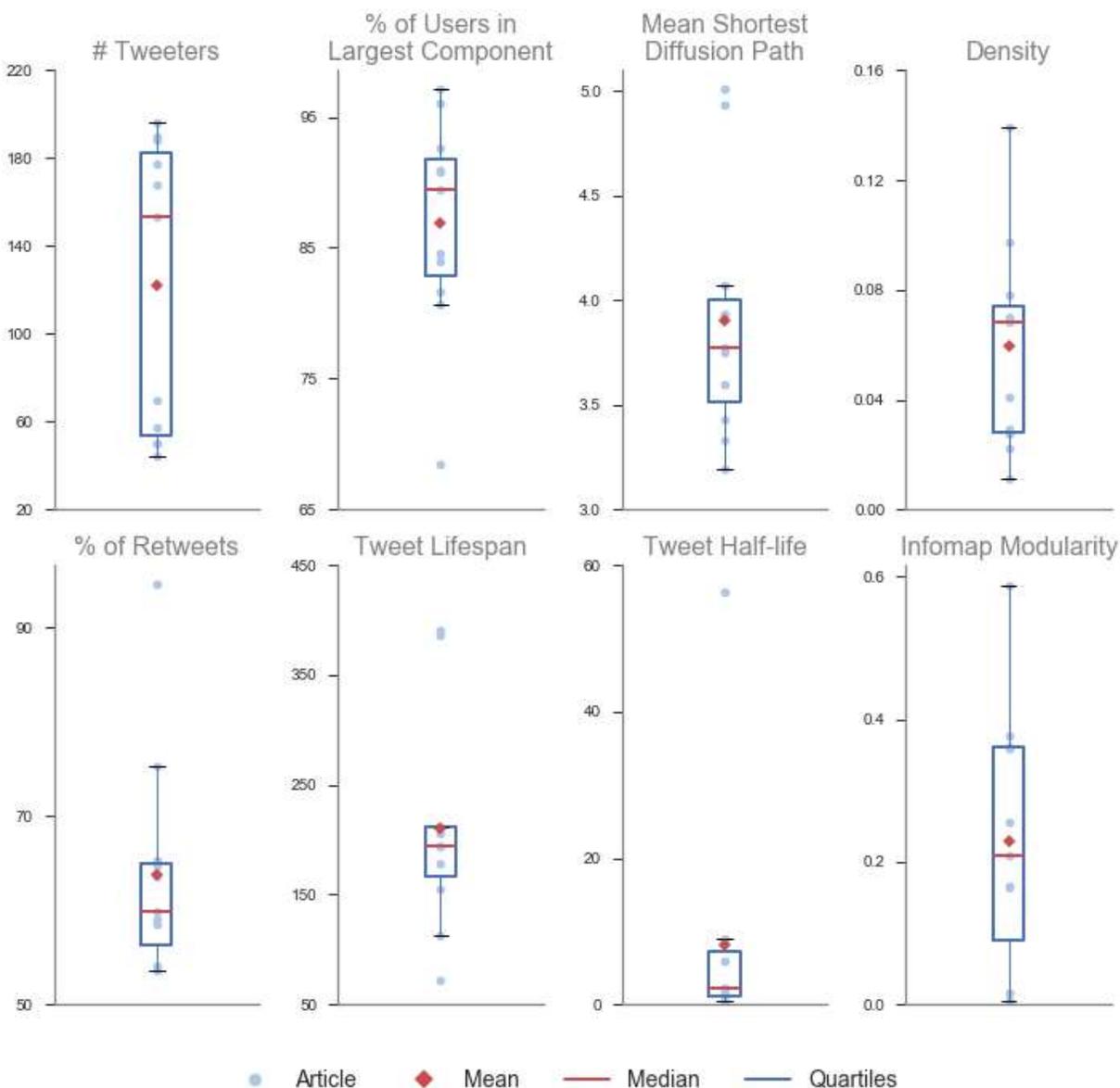


Figure 1. Distribution of network variables for the selected BMC articles

Another way to observe the extent of the ties between those that have shared an article is to use a community detection algorithm and then measure the modularity of the resulting clustering. This measure provides an indication of how well the network can be divided into distinct subgroups. The 11 networks in our sample show considerable heterogeneity in modularity scores, ranging from 0.00 (*Biol7*, very densely connected, Figure 2) and 0.59 (*Biol1*, sparsely connected, Figure 3). The wide range of modularity scores suggests different diffusion patterns of papers, with some being shared within a single tight-knit group of users, and others moving through various sub-communities. However, it appears that the modularity is not guided by network size. Although the 11 papers can be clearly divided into two groups in terms of Twitter activity (those tweeted by less than 80 users, and those tweeted by more than 140), there

are both high and low modularity networks in each group. Density and modularity provide an indication of the network structure between all those who saw and shared each paper, but they provide little insight into the diffusion pattern that lead to reaching these users.

To understand how the paper spread across the network within and between sub-communities it is necessary to look at the kind of sharing that took place (i.e., original tweet or retweet), and the timeline of when the article was tweeted.

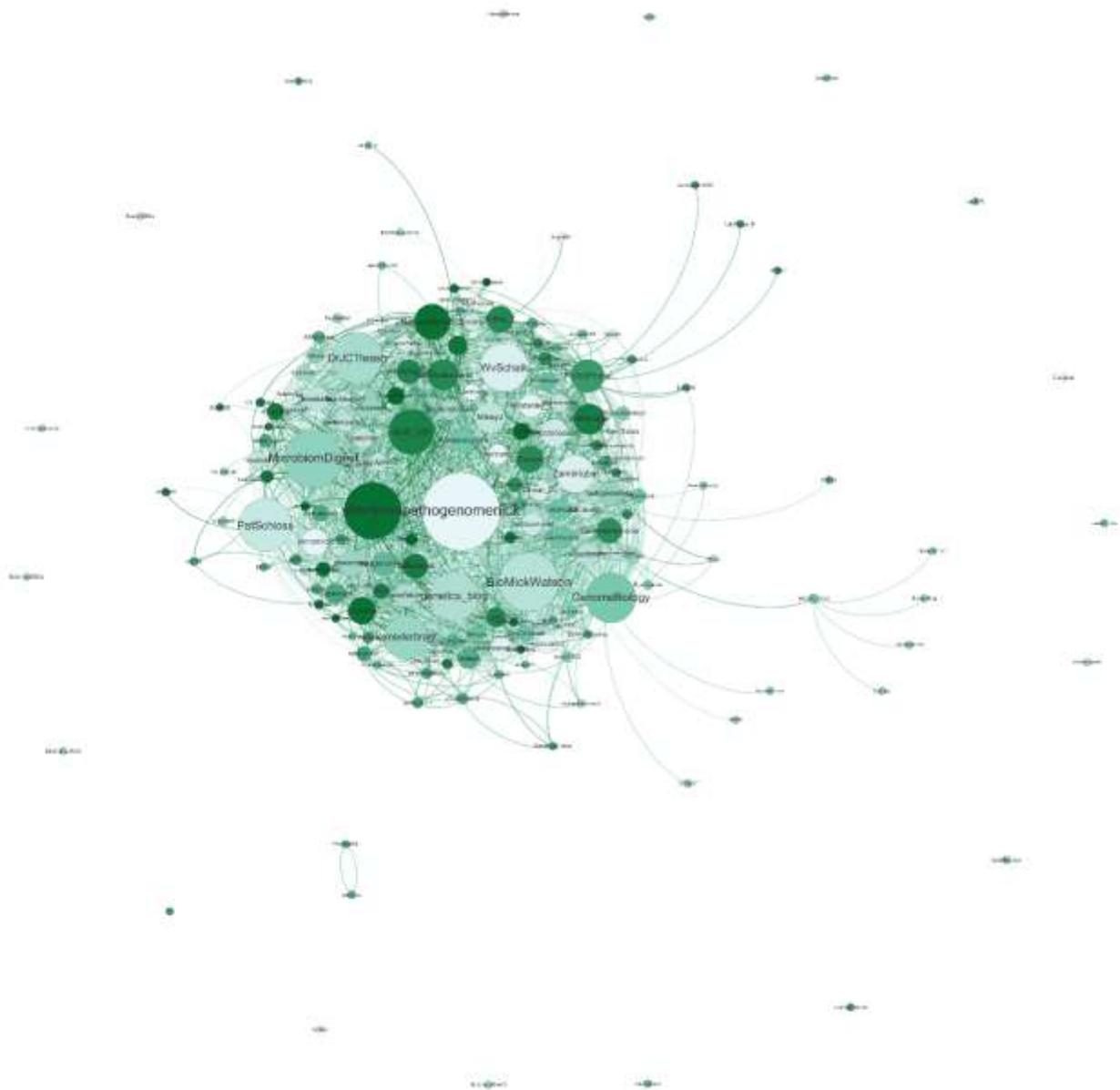


Figure 2. Follower network for Biol7 (example of a densely connected network)

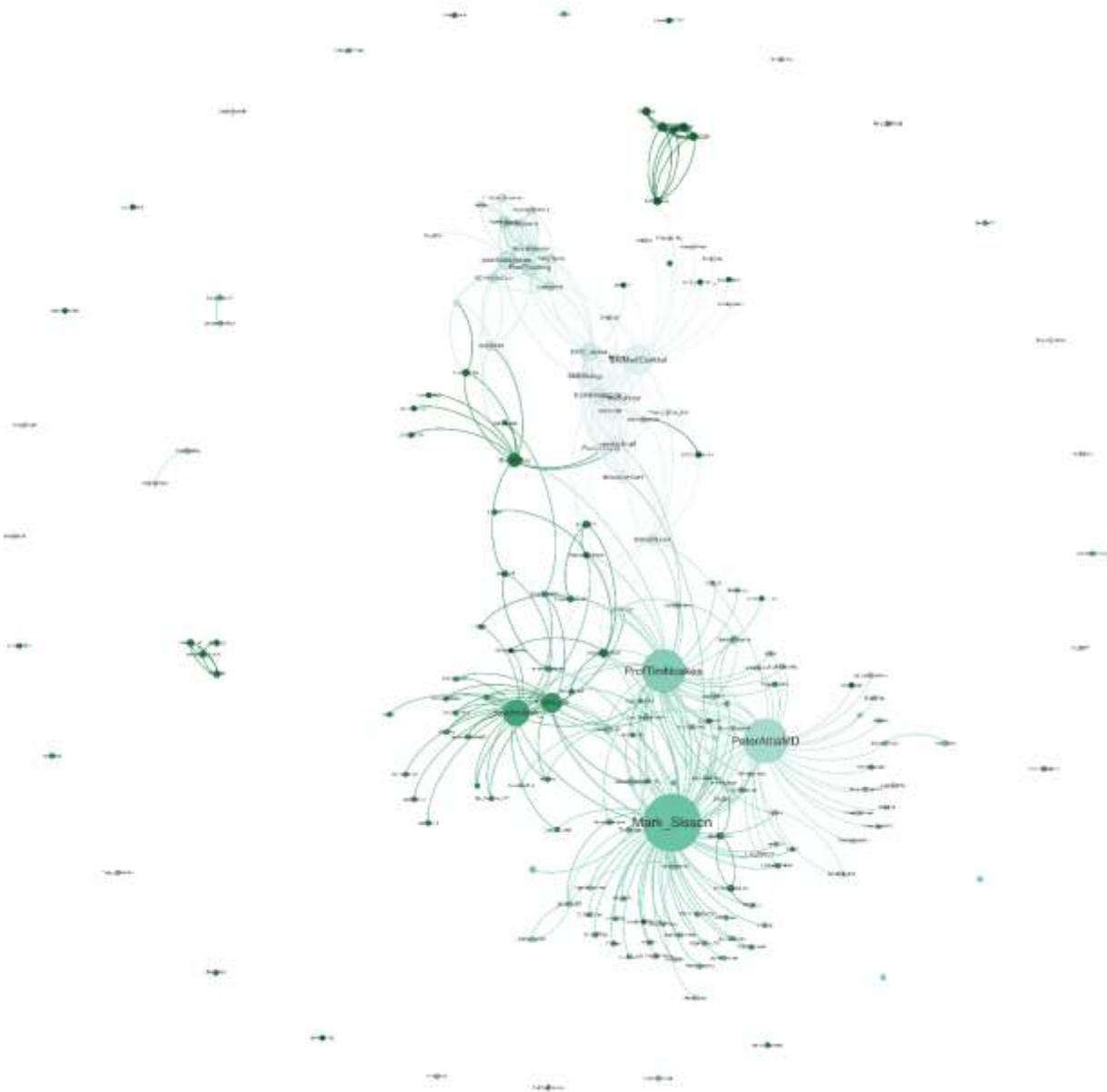


Figure 3. Follower network for Bio11 (example of a sparsely connected network)

## Temporal diffusion

To further investigate the temporal diffusion of the 11 articles on Twitter, we studied both the Twitter lifespan and the Twitter half-life (as defined in Table 2A). The time-sensitive nature of Twitter contributes to most adoptions of articles occurring during or the day after publication (Eysenbach, 2011; Shuai et al., 2012). Since Twitter involves a real-time updating of the latest content, it should be expected that users seldom dig-deep into other's archives to access

information. Temporal diffusion patterns might be affected by the “In case you missed it” function introduced in February 2016.<sup>9</sup>

Although most tweets in our dataset occur shortly after the initial appearance of an article, there is still heterogeneity in the diffusion patterns in social media of different published scientific articles. In the 11 papers studied, Twitter half-lives ranged from 0.4 to 56.3 days (Figure 1), meaning that it took between 10 hours and 8 weeks for half of all tweets mentioning an article to occur. However, 10 of the 11 papers had a half-life of less than 10 days (mean: 8.1, median: 2.2, standard deviation: 16.3). As shown in Table 3, *Bioll* is a clear outlier in this regard, with a half-life of 56.3.

Even though the majority of tweets occur within days of publication, the Twitter lifespan shows that some users still share them much later. Twitter lifespan varies between 72 and 391 days per paper, with a mean of 210 days and a standard deviation of 98 days. Eight of the papers have a lifespan within a standard deviation of the mean, with *Evol1* having a shorter lifespan (72 days) and *Bioll* (386 days) and *Biol2* (391 days) staying relevant for over a year.

One might expect that temporal patterns are reflected in the network structures. For example, high density and shorter paths between users enables faster diffusion or isolated users sharing the article later than the core of the network. We examine this relationship in two ways: by looking at the percentage of tweets that are retweets, and by looking at the mean shortest diffusion path. The percentage of users who shared the paper by retweeting someone else’s tweet, rather than initiating a new tweet of their own, helps us to explore whether the article propagated through Twitter using the retweet mechanism, or by users independently linking to the article. Retweets can only happen when a user sees an original tweet, which, by the nature of Twitter, generally only happens if the retweeting user is online at the same time or shortly after one of the users they follow tweets or retweets the original. The percentage of tweets that are retweets range from 54% (*Biol3*, *Biol5*, and *Evol1*) to 95% (*Evol4*), with a mean of 65% and a standard deviation of 12% (Table 3). One might expect, given the need to be online around the time of a previous tweet, that papers with higher proportions of retweets would have shorter half-lives, but there is no such relationship among the 11 papers studied. The percentage of retweets varies across the 11 papers, with *Bioll*, the paper with the highest half-life, exhibiting an average percentage of retweets (65%).

We find that the mean shortest diffusion path, which indicates the smallest distance a paper could have “traveled” through the follower network, ranges from 3.19 (*Biol5*) to 5.01 (*Evol4*), and among the sample of 11 papers, is strongly correlated with network density (Spearman  $p=0.75$ ,  $p\text{-value}<0.01$ ) and modularity of the largest component ( $p=0.54$ ,  $p\text{-value}<0.01$ ). This relationship between the length of the shortest diffusion paths and the network density and modularity of the largest component suggests a link between the timing, type of sharing, and the network structure. However, no relationship is evident when looking at the percentage of retweets or the tweet half-life, neither of which have statistically significant correlations among the eleven articles selected. Larger samples should be analyzed to be able to generalize these findings.

---

<sup>9</sup> Twitter’s explanation on how the Twitter timeline works can be found in their blog at the following URL: [https://blog.twitter.com/official/en\\_us/a/2016/never-miss-important-tweets-from-people-you-follow.html](https://blog.twitter.com/official/en_us/a/2016/never-miss-important-tweets-from-people-you-follow.html)

## Discussion

This study described how the measures proposed can be used to show that the diffusion of scientific articles on Twitter can take very different forms, even when the number of tweets and retweets is similar. Among the 11 articles studied, we saw large differences in the percentage of nodes in the largest component, network density, modularity scores, shortest diffusion paths, as well as tweet lifespans and half-lives, even among articles with similar number of tweets published in the same journal. For instance, the majority of networks had a half-life of less than 2.5 days, largest components with fairly low modularity, between 0.0 and 0.4 and mean shortest diffusion path lengths between 3 and 4; and. Such a network represents quick diffusion, primarily within a single, already connected community. The high percentage of nodes in the largest component, over 80% in all but one of the networks, coupled with the low modularity scores, suggest that many articles diffuse largely within a single community that is often tightly connected (i.e., with dense social ties between them). This is indicative of an already established community—a ready audience—for almost all of the papers.

Given the nature of these articles, it is reasonable to assume that the users who are already following each other on Twitter are part of an established community that has an interest in the topic of the paper and that this community is likely to consist mostly of academics. Although this remains to be verified by looking at the nature of the user accounts, it would suggest that many papers, even those with over 100 tweets, have not “broken out” of the scientific community and into the public but are shared by its traditional audience using a new medium. As a result, the findings here give us reason to be cautious of claims that social media metrics can be used as indicators of public and social impact of research (Bornmann, 2014b, 2015; Robinson-Garcia et al., 2017). Instead, these results suggest that metrics beyond counts are needed to identify the nature of the communities that a given piece of research has reached.

At the same time, a few articles (notably *Bioll*) differ by some key measures, and demonstrates how modularity, shortest diffusion paths, and Twitter half-lives may be useful to identify papers that have gone beyond a single community. This article, an interview with a researcher who studies the links between sugar and cancer, appears to have a community of early adopters (Figure 3) that includes the publisher’s Twitter account, indicating the community is composed of users who regularly read BMC journals. It also has a second community that discovers the article later (slightly darker) with several nutrition and fitness authors, such as @Mark\_Sisson, @ProfTimNoakes, @PeterAttiaMD and @\_sarahwilson\_, whose many followers share and retweet the paper, causing the high half-life. A few smaller communities discovered the article even later (darkest). The presence of these non-academics in the network suggests that this article had repercussions outside the academy in a way that is less evident in any of the other 10. Thus, high modularity (0.57), a long mean diffusion path (4.9) and a long half-life (56.3) may be good indicators of this reach beyond the academic community.

By looking at the network and diffusion patterns of these articles, we can determine that the number of tweets—the measure typically used when look at Twitter metrics for academic articles—is not able to reflect how the scientific article was shared. Network indicators are able to provide richer information about the visibility of scientific articles on Twitter. However, the results for the 11 articles reveal that none of the indicators alone are sufficient to understand the

diffusion patterns of each article. For example, *Evol4* also exhibits a relatively high modularity score (0.36); when visualized (Figure 4) it shows that the article was diffused among users in two connected communities, with the smaller one on the right tweeting earlier (light green) than the larger one on the left (darker green). Unlike *Bioll* (Figure 3), the half-life of *Evol4* was only 2.2 days, showing that the diffusion between one community and another can happen quickly. When inspecting the tweets and Twitter bios, it becomes apparent that both clusters seem to be made up of academic users but that geographical and language barriers separate the smaller cluster (English) from the larger one (Japanese).

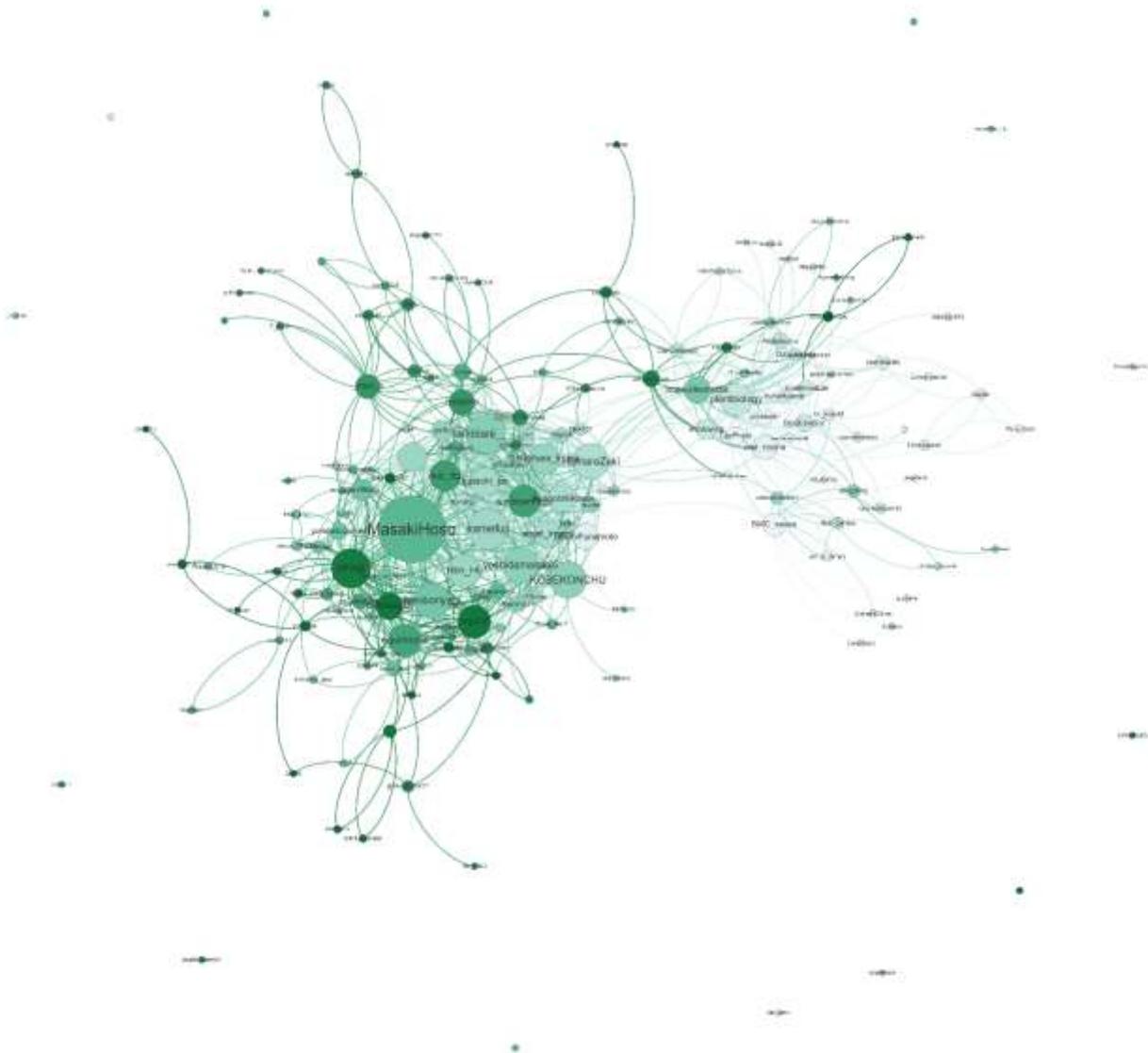


Figure 4. Follower network for *Evol4* (example of two connected community structure)

Most tweet activity occurs in the first two and a half days since a paper's initial appearance, showing that different diffusion patterns and eventual number of tweets can vary even within a short time frame. It seems, however, that a later tweet from a well-connected user can bring an article to life again, as is the case for *Bioll*. Such “sleeping beauties” may be more the exception than the rule for diffusion, but highlights how diffusion on Twitter is governed by the idiosyncratic sharing behaviours of users, as well as by time. It is unclear from this sample if such cases are always associated with longer diffusion paths and higher modularity.

Retweets represent a specific form of information diffusion and seem to play a significant role in sharing scientific papers on Twitter. More than half of all tweets for each of the papers were retweets, which seems to be in line with 54% for all Biology journal articles published in 2015 and indexed in the Web of Science (Haustein, 2018), but exceeds that observed for general Twitter users (boyd et al., 2010) and that of academics (Letierce et al., 2010; Priem and Costello, 2010) or science journalists on Twitter (Büchi, 2016). Since, when retweeting, Twitter users do not necessarily have to add anything new themselves to help propagate a new idea or article, the engagement with the content might be lower, with some users not having looked at or read the paper.

Overall, this study suggests that it is possible to detect different diffusion patterns, both in terms of the structure of the network reached and in terms of the time it takes to reach them. These differences are not visible by the number of tweets or distinct users, which are currently being used as altmetric indicators of Twitter visibility. For example, while *Biol2* and *Biol6* are similar in size (157 and 176 users respectively), modularity (0.25 and 0.37), mean shortest diffusion path length (both 3.9), half-lives (6.0 and 8.5), and percentage of retweets (59.0% and 53.6%), *Biol6* only has 68.5% of users in the largest component, as opposed to the more common 90.8% exhibited by *Biol2*. This indicates that *Biol6* reached a much higher proportion of isolated users that have no connection to any other users sharing the same paper, which suggests diffusion through a Twitter hashtags, through different media channels, or directly from the journal article.

## Conclusions

Social media is often heralded as an effective means to reach public audience with scientific knowledge that is normally confined to the esoteric musings of academic researchers. Indeed, the esoteric and exoteric circles that demarcate these scientific communities from the wider lay audience ought to be brought into dialogue because of social media platforms like Twitter. These platforms allow new and controversial ideas, concepts, and perspectives to be broadcasted to diverse audience instantaneously. As such, studying the diffusion patterns of published research ought to reveal how published papers traverse these audiences. However, our case study of several open access biology article re-shares on Twitter suggests the opposite. Despite varying patterns of diffusion over time, the reach of these papers rarely extended beyond the same users who form a well-connected community.

While the results of our study cannot be generalized beyond the small sample of biology articles studied, it is offered here as a proof of concept of the types of analysis that are possible across a much wider set of fields and platforms. Nevertheless, our finding has implications for those interested in science communication and public understanding of science, as well as research evaluation and management. For instance, we observe a high degree of heterogeneity in how articles are shared on social media, suggesting that better metrics for evaluating the circulation of research on social media are needed to capture the diversity of diffusion patterns. For instance, merely “counting” tweets is not sufficient for understanding how research diffuses on Twitter. Indeed, we observe that diffusion patterns often remain highly clustered, which would have been missed if re-tweets were simply counted. Our work emphasizes the importance of looking at the relationships and interactions between users who have engaged with research, as suggested by Robinson-Garcia et al. (forthcoming).

Our work also offers a conceptual foundation for an additional set of indicators that can be used to assess the visibility of research within and beyond the scientific community. Specifically, it may allow us to observe how and when research reaches the periphery of traditional scholarly communities and moves into the public sphere. It may also reveal when the direction is reversed: when citizen scientists bring debates and controversies to the footsteps of scholarly communities (see Newman, 2017). For instance, our work suggests that the majority of Twitter activity related to scientific publications stems from within the scholarly community, something that needs to be confirmed empirically by calculating the proposed indicators for a large and representative group of articles across fields. To that end, future work ought to scale the application of these proposed network analyses to examine the extent to which similar diffusion patterns are generalizable to other journals (including those which are not freely available to the public), scientific fields, and research topics. While this case study demonstrates the value of using social network analysis to detect different modes of propagation and instances where articles reach heterogeneous communities, a further cataloguing of how the corresponding network indicators behave across fields, journals, and topics will help both researchers and scientists better understand how—under what conditions—scientific work reaches different communities.

Future work ought to also focus on whom is playing a larger role in diffusing articles. Perhaps a high-profile Twitter account—irrespective if they are a scientist or not who is entrenched in these communities on Twitter—is more important to an article’s wide-spread diffusion, as was the case for *Bioll* in our case study (cf. Newman 2017 for the counterexample). To that end, we plan to triangulate findings on the network structure with characteristics about users and tweet content to further differentiate diffusion of scientific articles on Twitter.

Alternatively, it may be the article’s characteristics that hamper, catalyze, or otherwise affect its diffusion. For instance, applying techniques from computational linguistics to the text in articles may reveal what rhetorically or topically drives diffusion, and how it relates and shapes the network indicators of the resulting diffusion. For example, controversial subjects in a field (or to the wider societal zeitgeist) might catalyze diffusion between scientists and the public more readily. It may also be the case that articles that use certain forms of rhetoric that makes their arguments more accessible to lay audiences have a broader reach. These efforts also ought

to control for or link to trends in traditional media that may spur diffusion in social media (see O'Neill et al., 2015).

Social media like Twitter has substantial leverage as to how information is broadcast. As such, better understanding the patterns of diffusion has the potential to expand the dialogue between scientists and the public. Our work is a first step to develop richer and more robust measures of diffusion between scholars and public audiences on social media.

## References

- Alperin JP (2015a) Geographic variation in social media metrics: an analysis of Latin American journal articles. *Aslib Journal of Information Management*, Emerald Group Publishing Limited 67(3): 289–304. Available from: <http://www.emeraldinsight.com.ezproxy.lib.ucalgary.ca/doi/full/10.1108/AJIM-12-2014-0176> (accessed 26 May 2015).
- Alperin JP (2015b) Moving beyond counts: A method for surveying Twitter users. In: *2nd Altmetrics Conference, 2AM*, Amsterdam, Netherlands. Available from: [http://altmetrics.org/wp-content/uploads/2015/09/altmetrics15\\_paper\\_3.pdf](http://altmetrics.org/wp-content/uploads/2015/09/altmetrics15_paper_3.pdf) (accessed 12 August 2017).
- Archambault É, Amyot D, Deschamps P, et al. (2014) *Proportion of open access papers published in peer-reviewed journals at the European and world levels—1996–2013*. *Science-Metrix - European Commission*. Available from: [http://science-metrix.com/sites/default/files/science-metrix/publications/d\\_1.8\\_sm\\_ec\\_dg-rtd\\_proportion\\_oa\\_1996-2013\\_v11p.pdf](http://science-metrix.com/sites/default/files/science-metrix/publications/d_1.8_sm_ec_dg-rtd_proportion_oa_1996-2013_v11p.pdf) (accessed 15 January 2016).
- 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. Available from: <http://www.sciencedirect.com/science/article/pii/S1751157714000868> (accessed 3 October 2014).
- 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. Available from: <http://www.sciencedirect.com/science/article/pii/S1751157714000881> (accessed 20 October 2014).
- Bornmann L (2015) Usefulness of altmetrics for measuring the broader impact of research. Stefanie Haustein, Dr Cassidy R. Su D (ed.), *Aslib Journal of Information Management*, Emerald Group Publishing Limited 67(3): 305–319. Available from: <http://www.emeraldinsight.com.proxy.lib.sfu.ca/doi/full/10.1108/AJIM-09-2014-0115> (accessed 14 June 2015).
- boyd D, Golder S and Lotan G (2010) Tweet, Tweet, Retweet: Conversational Aspects of Retweeting on Twitter. In: *43rd Hawaii International Conference on System Sciences*, IEEE, pp. 1–10. Available from: <http://ieeexplore.ieee.org/document/5428313/> (accessed 12 August 2017).
- Brossard D (2013) New media landscapes and the science information consumer. *Proceedings of the National Academy of Sciences*, National Academy of Sciences 110(Supplement 3): 14096–14101. Available from: [http://www.pnas.org/content/110/Supplement\\_3/14096.short](http://www.pnas.org/content/110/Supplement_3/14096.short) (accessed 23 November 2017).
- Brossard D and Scheufele DA (2013) Social science. Science, new media, and the public. *Science (New York, N.Y.)*, American Association for the Advancement of Science 339(6115): 40–1. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/23288529> (accessed 23 November 2017).
- Bucchi M (2013) Style in science communication. *Public understanding of science* 22(8): 904–15. Available from: <http://journals.sagepub.com/doi/10.1177/0963662513498202> (accessed 12 August 2017).

- Büchi M (2016) Microblogging as an extension of science reporting. *Public Understanding of Science*: 96366251665779. Available from: <http://journals.sagepub.com/doi/10.1177/0963662516657794> (accessed 12 August 2017).
- Costas R, Zahedi Z and 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*, Digital Libraries, Leiden. Available from: <http://doi.wiley.com/10.1002/asi.23309> (accessed 22 January 2014).
- Costas R, Zahedi Z and Wouters P (2015) The thematic orientation of publications mentioned on social media. Stefanie Haustein, Dr Cassidy R. Su D (ed.), *Aslib Journal of Information Management*, Emerald Group Publishing Limited 67(3): 260–288. Available from: <http://www.emeraldinsight.com/doi/10.1108/AJIM-12-2014-0173> (accessed 12 August 2017).
- de Solla Price DJ (1965) *Little science, big science and beyond*. Columbia University Press, 1986.
- 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*, *Journal of Medical Internet Research* 13(4): e123. Available from: <http://www.jmir.org/2011/4/e123/> (accessed 27 February 2014).
- Fleck L (1935) *Genesis and development of a scientific fact*. University of Chicago Press, 1981.
- Fox S and Duggan M (2013) *Health Online 2013*. Pew Research Center Fact Sheet. Available from: <http://www.pewinternet.org/2013/01/15/health-online-2013/> (accessed 12 August 2017).
- Funk C, Gottfried J and Mitchell A (2017) Science News and Information Today. *Pew Research Center: Journalism and Media*. Available from: <http://www.journalism.org/2017/09/20/science-news-and-information-today/> (accessed 23 November 2017).
- Gauchat G (2012) Politicization of Science in the Public Sphere. *American Sociological Review*, SAGE Publications Sage CA: Los Angeles, CA 77(2): 167–187. Available from: <http://journals.sagepub.com/doi/10.1177/0003122412438225> (accessed 12 August 2017).
- Goel S, Anderson A, Hofman J, et al. (2015) The Structural Virality of Online Diffusion. *Management Science*, INFORMS: 150722112809007. Available from: <http://pubsonline.informs.org/doi/10.1287/mnsc.2015.2158> (accessed 21 June 2017).
- Greenwood S, Perrin A and Duggan M (2016) Social Media Update 2016. *Pew Research Center: Internet & Technology*. Available from: <http://www.pewinternet.org/2016/11/11/social-media-update-2016/> (accessed 28 November 2017).
- Hammarfelt B (2014) Using altmetrics for assessing research impact in the humanities. *Scientometrics* 101(2): 1419–1430. Available from: <http://link.springer.com/10.1007/s11192-014-1261-3> (accessed 30 May 2014).
- Haustein S (2018) Scholarly Twitter Metrics. In: Glaenzel W, Moed HF, Schmoch U, et al. (eds), *Handbook of Quantitative Science and Technology Research*, Springer.
- Haustein S, Costas R, Larivière V, et al. (2015) Characterizing Social Media Metrics of Scholarly Papers: The Effect of Document Properties and Collaboration Patterns. *PLOS ONE* 10(3): e0120495. Available from: <http://dx.plos.org/10.1371/journal.pone.0120495> (accessed 12 August 2017).
- Haustein S, Tsou A, Minik V, et al. (2016) Identifying Twitter user communities in the context of altmetrics. In: *3rd Altmetrics Conference 3:AM*, Bucharest, Romania.
- Haustein S, Bowman TD and Costas R (2016) Interpreting ‘altmetrics’: viewing acts on social media through the lens of citation and social theories. In: Sugimoto CR (ed.), *Theories of Informetrics and Scholarly Communication*, Berlin: de Gruyter Mouton, pp. 372–405. Available from: <https://www.degruyter.com/view/books/9783110308464/9783110308464-022/9783110308464-022.xml> (accessed 18 August 2015).
- Ke Q, Ahn Y-Y and Sugimoto CR (2017) A systematic identification and analysis of scientists on Twitter. Bornmann L (ed.), *PLOS ONE*, Public Library of Science 12(4): e0175368. Available from: <http://dx.plos.org/10.1371/journal.pone.0175368> (accessed 23 November 2017).
- Klemetti E (2017) Follow All of These Earth Scientists on Twitter Right Now. *Wired*. Available from: <https://www.wired.com/2017/01/follow-earth-scientists-twitter-right-now/> (accessed 23 November 2017).

- 2017).
- Letierce J, Passant A, Decker S, et al. (2010) Understanding how Twitter is used to spread scientific messages. In: *Proceedings of the WebSci10: Extending the Frontiers of Society On-Line*, Raleigh, USA. Available from: <http://journal.webscience.org/314/> (accessed 12 August 2017).
- Leydesdorff L and Wagner CS (2008) Is the United States losing ground in science? A global perspective on the world science system. *Scientometrics* 78(1): 23–36. Available from: <http://www.springerlink.com/index/10.1007/s11192-008-1830-4> (accessed 26 July 2012).
- Mislove A, Lehmann S, Ahn Y-Y, et al. (2011) Understanding the Demographics of Twitter Users. In: *Fifth International AAAI Conference on Weblogs and Social Media*, Barcelona, Spain. Available from: <https://www.aaai.org/ocs/index.php/ICWSM/ICWSM11/paper/viewFile/2816/3234> (accessed 12 August 2017).
- Moody J (2004) The Structure of a Social Science Collaboration Network: Disciplinary Cohesion from 1963 to 1999. *American Sociological Review* 69(2): 213–238. Available from: <http://journals.sagepub.com/doi/10.1177/000312240406900204> (accessed 12 August 2017).
- Myers SA and Leskovec J (2014) The bursty dynamics of the Twitter information network. In: *Proceedings of the 23rd international conference on World wide web - WWW '14*, New York, New York, USA: ACM Press, pp. 913–924. Available from: <http://dl.acm.org/citation.cfm?doid=2566486.2568043> (accessed 7 August 2017).
- Myers TA, Kotcher J, Stenhouse N, et al. (2017) Predictors of trust in the general science and climate science research of US federal agencies. *Public Understanding of Science*, SAGE PublicationsSage UK: London, England 26(7): 843–860. Available from: <http://journals.sagepub.com/doi/10.1177/0963662516636040> (accessed 27 November 2017).
- Newman TP (2017) Tracking the release of IPCC AR5 on Twitter: Users, comments, and sources following the release of the Working Group I Summary for Policymakers. *Public Understanding of Science*, SAGE PublicationsSage UK: London, England 26(7): 815–825. Available from: <http://journals.sagepub.com/doi/10.1177/0963662516628477> (accessed 26 January 2018).
- Nielsen MA (2012) *Reinventing discovery the new era of networked science*. Princeton University Press.
- Nisbet MC and Fahy D (2015) The Need for Knowledge-Based Journalism in Politicized Science Debates. Suhay E and Druckman JN (eds), *The ANNALS of the American Academy of Political and Social Science*, SAGE PublicationsSage CA: Los Angeles, CA 658(1): 223–234. Available from: <http://journals.sagepub.com/doi/10.1177/0002716214559887> (accessed 23 November 2017).
- O'Neill S, Williams HTP, Kurz T, et al. (2015) Dominant frames in legacy and social media coverage of the IPCC Fifth Assessment Report. *Nature Climate Change* 5. Available from: [https://pdfs.semanticscholar.org/732a/6a630b7e2c24c81b09bf29cc6283e6c63b02.pdf?\\_ga=2.161569598.305728434.1517007091-1944712879.1517007091](https://pdfs.semanticscholar.org/732a/6a630b7e2c24c81b09bf29cc6283e6c63b02.pdf?_ga=2.161569598.305728434.1517007091-1944712879.1517007091) (accessed 26 January 2018).
- Piwowar H, Priem J, Larivière V, et al. (2017) The State of OA: A large-scale analysis of the prevalence and impact of Open Access articles. *PeerJ Preprints*. Available from: <https://peerj.com/preprints/3119/> (accessed 12 August 2017).
- Priem J and Costello KL (2010) How and why scholars cite on Twitter. *Proceedings of the American Society for Information Science and Technology*, Wiley Subscription Services, Inc., A Wiley Company 47(1): 1–4. Available from: <http://doi.wiley.com/10.1002/meet.14504701201> (accessed 12 August 2017).
- Robinson-Garcia N, van Leeuwen TN and Rafols I (2017) Using Almetrics for Contextualised Mapping of Societal Impact: From Hits to Networks. *Science and Public Policy*. Available from: [https://papers.ssrn.com/sol3/papers.cfm?abstract\\_id=2932944](https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2932944) (accessed 12 August 2017).
- Rowlands I, Nicholas D, Russell B, et al. (2011) Social media use in the research workflow. *Learned Publishing* 24(3): 183–195. Available from: <http://doi.wiley.com/10.1087/20110306> (accessed 12 August 2017).
- Scott J (2017) *Social network analysis*. SAGE.
- Shuai X, Pepe A and Bollen J (2012) How the scientific community reacts to newly submitted preprints: article downloads, Twitter mentions, and citations. Ouzounis CA (ed.), *PloS one*, Public Library of

- Science 7(11): e47523. Available from: <http://dx.plos.org/10.1371/journal.pone.0047523> (accessed 21 January 2014).
- Sugimoto CR, Work S, Larivière V, et al. (2017) Scholarly use of social media and altmetrics: A review of the literature. *Journal of the Association for Information Science and Technology*.
- Thelwall M, Haustein S, Larivière V, et al. (2013) Do altmetrics work? Twitter and ten other social web services. *PloS one*, Public Library of Science 8(5): e64841. Available from: <http://dx.plos.org/10.1371/journal.pone.0064841> (accessed 21 January 2014).
- Tsou A, Bowman TD, Ghazinejad A, et al. (2015) Who tweets about science? In: *Proceedings of the 2015 International Society for Scientometrics and Informetrics*, pp. 95–100.
- van Noorden R (2014) Online collaboration: Scientists and the social network. *Nature* 512(7513): 126–129. Available from: <http://www.nature.com/doi/10.1038/512126a> (accessed 12 August 2017).
- Veletsianos G (2016) *Social media in academia : Networked scholars*. New York: Routledge, Taylor & Francis Group.
- Vraga EK, Thorson K, Kligler-Vilenchik N, et al. (2015) How individual sensitivities to disagreement shape youth political expression on Facebook. *Computers in Human Behavior*, Pergamon 45: 281–289. Available from: <https://www.sciencedirect.com/science/article/pii/S0747563214007419> (accessed 23 November 2017).
- Wagner CS and Leydesdorff L (2005) Network structure, self-organization, and the growth of international collaboration in science. *Research Policy* 34(10): 1608–1618. Available from: <http://linkinghub.elsevier.com/retrieve/pii/S0048733305001745> (accessed 26 July 2012).
- Wasserman S. and Faust K (1994) *Social Network Analysis: Methods and Applications*. Cambridge University Press. Available from: <http://www.cambridge.org/ca/academic/subjects/sociology/sociology-general-interest/social-network-analysis-methods-and-applications?format=PB&isbn=9780521387071#hBmOCTyW8UzTSvI2.97> (accessed 23 November 2017).
- Weller K and Puschmann C (2011) Twitter for Scientific Communication: How Can Citations/References be Identified and Measured. In: *Proceedings of the webSci11*. Available from: <http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.398.330&rep=rep1&type=pdf> (accessed 23 November 2017).
- You J (2014) The top 50 science stars of Twitter. *Science News*. Available from: <http://www.sciencemag.org/news/2014/09/top-50-science-stars-twitter> (accessed 23 November 2017).