How do scientific papers with different levels of journals spread online? Exploring the temporal dynamics in the diffusion processes

Renmeng Cao\textsuperscript{a}, Xiaoke Xu\textsuperscript{c}, Yunxue Cui\textsuperscript{a}, Zhizhao Fang\textsuperscript{b}, Xianwen Wang\textsuperscript{a}

\textsuperscript{a} WISE Lab, Faculty of Humanities and Social Sciences, Dalian University of Technology, Dalian 116085, China
\textsuperscript{b} Centre for Science and Technology Studies (CWTS), Leiden University, Leiden, The Netherlands
\textsuperscript{c} College of Information and Communication Engineering, Dalian Minzu University, Dalian 116600, China

Abstract

Social media has become an important channel for publicizing academic research, which provides an opportunity for each scientific paper to become a hit. Employing a dataset of about 10 million tweets of 584,264 scientific papers from 2012 to 2018, this study investigates the differential diffusion of elite and non-elite journal papers (divided by Average journal impact factor percentile). We find that non-elite journal papers are diffused deeper and farther than elite journal papers, showing a diffusion trend with multiple rounds, sparse, short-duration and small-scale bursts. In contrast, the bursts of elite journals are characterized by a small number of persistent, dense and large-scale bursts. We also discover that elite journal papers are more inclined to broadcast diffusion while non-elite journal papers prefer viral diffusion. Elite journal papers are generally disseminated to many loosely connected communities, while non-elite journal papers are diffused to several densely connected communities.

1. Introduction

Social media has become a dispensable part of the scholarly communication system in a fundamental way (Sugimoto et al., 2017). An increasing number of journal publishers and authors are using social media to promote their research outputs, which not only boosts the social and scientific impact of scientific publications but also enhances the interactions between the scientific community and the public (Bik & Goldstein, 2013; Darling et al., 2013; Zheng et al., 2019). According to a survey by Hitlin & Olmstead (2018), millions of users read science-related information on Facebook feeds or elsewhere on social media, which makes it a reality for scientific papers to obtain large-scale attention. Just like breaking news, scientific papers can also cumulate much attention in a very short time, which in turn creates bursty diffusion with profound effect (Cao et al., 2021; Cui et al., 2019; Zakhlebin & Horvát, 2020). As scientific research increasingly shape public discourse and impact the decision-making of both individuals

\textsuperscript{1} Corresponding author: Xianwen Wang. xianwenwang@dlut.edu.cn
and policymakers, there is a growing interest in studying how scientific findings are diffused online (Yin et al., 2021; Zakhlebin & Horvát, 2020).

1.1 Temporal dynamics of online information

Information diffusion on social media exhibits many complex characteristics. Burst is a salient temporal dynamic in the process of information diffusion. When information receives a lot of attention, the information appears as a burst and shows a significant rising and falling trend on the diffusion curve (Crane & Sornette, 2008; Matsubara et al., 2012; Yang & Leskovec, 2011). Studying the bursty nature of online information not only enhances the comprehension of the information diffusion process but also helps to disclose the underlying mechanism by which online information goes viral or spreads widely (Cao et al., 2021; Cheng et al., 2016; Leskovec et al., 2009).

To explore the burst mechanism behind online information diffusion, many researchers have measured the bursty characteristics of various types of online content such as videos, news, and phrases. For example, Broxton et al. (2013) observed that highly viewed videos on YouTube rose to, and fell from, their peak popularity more quickly than less viewed videos. Wu et al. (2011) found that the peak decay rate of news information containing more words related to negative emotion, actions and more complicated cognitive processes was more quickly than those containing more words related to positive emotion, leisure, and lifestyle. By tracking 1.6 million mainstream media sites and blogs over three months, Leskovec et al. (2009) reported a common lag of 2.5 hours between the peaks of attention to a phrase in news media and blogs. Besides, many scholars have distinguished between the burst patterns of online information. For example, Yang and Leskovec (2011) developed a clustering algorithm to explore how the popularity of online content grows and fades over time and clustered their temporal patterns into six different burst patterns. Crane and Sornette (2008) analyzed the time series of daily views for videos on YouTube and divided their burst patterns into three classes. Matsubara et al. (2012) argued that there was only one type of burst pattern, and accordingly proposed a model that can fit all patterns found in real data simply by changing the parameter values.

In addition to the characterization of the burst patterns, unraveling the origin of bursts of online content has also attracted much attention in recent years. By analyzing the online activity of millions of users on Twitter, De Domenico & Altmann (2020) argued that the heterogeneity of social interactions and the preferential attention towards influential users were the two important mechanisms driving the bursts in collective attention. Cheng et al. (2016) found that the cascades of image memes on Facebook exhibit multiple peaks of popularity over long time scales, and bursts took place inside a very well connected region, followed by later bursts in different parts of the network. Further study showed that a small fraction of users who share content more than once played an important role in catalyzing multiple bursts of information cascades (Almanza et al., 2021). Zhan et al. (2020) reported that cross-community diffusion is the main mechanism driving online information experiencing multiple bursts of popularity.

To sum up, while considerable research has investigated the bursty nature of
different types of online information, there is less research exploring the bursty nature of online diffusion of scientific publications. In other words, it remains unknown how scientific papers burst (go viral or succeed) online and whether scientific papers share some commonalities with other types of online information in the way of bursting?

1.2 Online diffusion of scientific papers

Exploring online diffusion patterns of scientific papers has become a question of great interest to many researchers in recent years. Many studies have investigated the structural properties of diffusion networks of scientific papers. For example, Didegah and Thelwall (2018) constructed co-tweeted networks of research articles, finding that the full-text accessibility of an article may affect the likelihood of it being tweeted. Based on the technique of social network analyses (SNA), Alperin et al. (2019) analyzed the following network of users who tweeted research articles, finding that most papers were shared within single-connected communities with limited diffusion to the public. Imran M. et al. (2018) perform a community detection for Twitter mention networks of scholarly articles from different disciplines, finding that the interconnectivity degree of communities in a network varies in disciplines. Based on the study of Imran M. et al. Said et al. (2019) discovered that large communities are dominated by organizational accounts associated with journals, while small communities are dominated by experts in the field.

In addition to the diffusion network structure, some studies focused on the bursty nature of online diffusion of scientific papers. For instance, Cui et al. (2019) found that burstiness is a typical feature of online diffusion of highly retweeted articles, and the diffusion pattern of these articles was more inclined to broadcast diffusion, rather than viral diffusion. Cao et al. (2021) further explored the cause of multiple bursts of highly retweet articles and found that overlapping users could drive scholarly articles to experience multiple rounds of bursts. Zakhlebin and Horvát (2020) studied the bursty diffusion of scientific articles across online platforms, finding bursts within each platform routinely occur more than once, and recurring bursts become smaller and less frequent.

Some studies investigated the communication effect (i.e., altmetrics) of different levels of papers on online platforms (Haustein et al., 2016; Waltman & Costas, 2014). For example, Costas et al. (2015) found that those papers published in high-impact journals such as Nature or Science obtained substantial attention on Mendeley. Two prior works showed that papers published in the most prestigious scientific journals such as Lancet or New England Journal of Medicine were generally highly tweeted on Twitter (Haustein et al., 2014). Moreover, the cumulative advantage of high-impact journals is also shown in other online platforms such as F1000 and blogs (Shema et al., 2015; Waltman & Costas, 2014).

Overall, the mentioned studies provide a comprehensive view to understand the online diffusion of scientific papers. Nevertheless, less effort has been made to disclose the diffusion mechanism of scientific papers, especially from the perspective of the journal levels. More specifically, how papers published in different levels of journals spread online and how people interact with them is still unclear.
1.3 Research questions

By constructing diffusion (retweet) networks of scientific papers, this study aims to compare the online diffusion of papers published in different levels of journals. Specifically, we seek to address the following questions:

RQ1. What communication effect do papers published in different levels of journals exhibit?

RQ2. What kind of bursty trends do they exhibit?

RQ3. What are the differences in the diffusion mechanisms behind them?

Answers to these questions not only further enhance the comprehension of the diffusion process of scientific papers, but also help to improve the diffusion of scientific knowledge.

2 Data and methods

2.1 Data description and preprocessing

Altmetric.com is one of the main providers of altmetric data. It tracks mentions to scholarly outputs from a selection of online sources such as social media platforms, blogs, news media, and online reference managers. Twitter is the predominant source of social media data in Altmetric.com (Robinson-Garcia et al., 2014). Altmetric.com tracks original tweets and retweets that contain direct links to scholarly outputs in real-time.

This study uses a Twitter mention data of scientific papers provided by Altmetric.com, which records the retweeting activity of scientific papers between 2012 and 2018. This dataset involves 300 journals (not including review journals), 584,264 research papers, 4,136,308 original tweets, 5,873,464 retweets and 4,625,396 users. For each paper, altmetric id, title, publication date, DOI and journal name are collected. For each original tweet, original author id, tweet id, posting time are included. For each retweet, tweeter id, retweeter id, tweet id, retweeted id, reposting time are collected.

2.2 Defining elite and non-elite journal papers

Considering that the papers in our dataset are from different subject areas, the metric of the journal impact factors (IF) is not suitable to rank the journals. Alternatively, we use the Average Journal Impact Factor (JIF) Percentile metric\(^3\), which is computed as follows:

\[
\text{Average JIF Percentile} = \frac{\text{JIF Percentile}_1 + \cdots + \text{JIF Percentile}_N}{N},
\]

\(^3\)https://help.altmetric.com/support/solutions/articles/6000235926-twitter (Accessed October 12, 2021).

\(^4\)http://help.incites.clarivate.com/incitesLiveJCR/glossaryAZgroup/g4/9995-TRS.html (Accessed October 12, 2021).
where $JIF \text{ percentile}_N$ denotes the JIF Percentile of a subject area a journal belongs to, and $N$ denotes the number of subject areas a journal belongs to. Compared with impact factor, Average JIF Percentile improves the relative value of impact factor, has a smaller coefficient of variation and the data distribution is closer to a normal distribution, which is good for horizontal comparison among journals (Yu & Yu, 2016). Given the fluctuations of Average JIF Percentile at different years for the same journal, this study uses the mean of Average JIF Percentile between 2012 and 2018 ("P7" hereafter) as a proxy for the ranking of journals. Figure 1 shows the distribution of P7 values among the 300 journals. From the figure, it can be seen that these scientific journals come from different impact levels. Moreover, the distribution is very uneven, ranging from 11.08 to 99.72. The median and mean of the distribution are 76.87 and 80.02, respectively.

To compare the online diffusion of papers with different levels of journals, this paper pays particular attention to highly retweeted papers published between 2012 and 2018 on Twitter. First, we give a uniform standard of highly retweeted papers from different journal levels, which are defined as the papers that were retweeted larger than 100 times. Next, we rank the scientific journals of these highly retweeted papers by P7 values in descending order. “Elite” journals are defined as the top 10% journals, and “non-elite” journals are defined as the journals with P7 values lower than the median (80.23). Then, 30 “elite” journals and 150 “non-elite” journals are obtained. The P7 values of “elite” journals range from 98.74 to 99.71, and that of “non-elite” journals are distributed between 11.80 and 80.09. The median of the former is 98.74, while that of the latter is 65.68. Finally, using the 180 journals to match scientific papers, we obtain 4,126 elite journal papers and 1,158 non-elite journal papers. To construct a control group for the study, an equal number of elite and non-elite journal papers (1,158) are chosen.
2.3 Constructing the diffusion network

Retweeting is the key mechanism in the information diffusion in Twitter, providing valuable traces for the study of information diffusion in Twitter (Kupavskii et al., 2012; Suh et al., 2010). When user A retweets an original tweet posted by user B, we argue that the information spread by user B is received/adopted by user A. To characterize the diffusion paths of each paper, we construct a directed diffusion network $G_R = (U_R, E_R)$ based on the retweeting relationships amongst Twitter users, where each user $u \in U_R$ is represented as a node and each directed edge $e(u_s, u_t) \in E_R$ from $u_s$ to $u_t$ is established when $u_t$ retweets the original tweet posted by $u_s$. Figure 2(a) shows the diffusion network of a scientific paper example (Altmetric ID: 3775741). The black nodes are the initiators of paper diffusion, which are the users who posted original tweets that link to the paper. The gray nodes are the users who retweeted original tweets (i.e., retweeters). The network consists of multiple subgraphs of different sizes. However, these subgraphs only describe the diffusion path of original tweets that link to the paper, while the diffusion paths from the paper to the initiators are missing. Hence, we add a virtual node (i.e., a paper node) and new edges into $G_R$ for each network, forming a complete diffusion network $G'_R$. The newly added node and edges not only make the diffusion network complete but also helps to characterize the importance of papers in diffusion processes (Lü et al., 2011). Figure 2(b) displays the complete diffusion network of the scientific paper example (Altmetric ID: 3775741). The red node is the paper node and the blue edges between the paper node and initiators are the newly added edges, which are used to represent the diffusion paths from the paper to initiators. The number of newly added edges is equal to that of initiators.

Figure 2. The topological structure of the diffusion network of a paper example (Altmetric ID: 3775741): (a) The network without diffusion paths from the paper to initiators; (b) The network
To compute the number of temporal motifs (detailed descriptions of temporal motifs and temporal network will be given in Section 2.4), we construct a temporal network $G_T = (U_T, E_T, W_{\Delta T})$ for each paper, where each user $u \in U_T$ is represented as a node and each directed edge $e(u_s, u_t) \in E_T(u_s, u_t) \in E_T$ from $u_s$ to $u_t$ is established when $u_t$ retweets the original tweet posted by $u_s$. The weight of each edge $w_{\Delta t} \in W_{\Delta T}$ is represented as the response time between the retweeting timestamp $r_t$ and posting timestamp $p_t$ (i.e., $w_{\Delta t} = r_t - p_t$). Likewise, we add a paper node and new edges between the paper node and initiators for each paper, forming a complete temporal network $G'_T$. The weights of these newly added edges represent the response time from the initiators discovering scientific papers to posting original tweets that contain links to the papers. This paper set the weights of these edges to zero.

2.4 Measuring the diffusion network

(1) Diffusion metrics

We measure the diffusion patterns of elite and non-elite journal papers from the dimensions of Scale, Breadth and Average depth. To exemplify the definition of these three dimensions, Figure 3 shows a sketch of a diffusion network in a tree layout. The pink node represents the root node (i.e., a scientific paper). The blue nodes are the communicators of the paper on Twitter, including both initiators and retweeters. These nodes are classified into different levels based on their shortest distance from the root node. Therefore, for a specific paper:

- **Scale** is defined as the total number of communicators of the paper;
- **Breadth** is defined as the maximum number of communicators amongst levels;
- **Average depth** is defined as the average shortest distance between the root node and other nodes.

Take the diffusion network in Figure 3 as an example, the scale, breadth and average depth of the network are 13, 5, and 2.08, respectively.

![Figure 3. A schematic representation of a diffusion network in a tree layout.](image-url)
(2) Temporal motif

Temporal networks are commonly used to represent systems where connections between elements are active only for restricted periods of time, such as networks of telecommunication, neural signal processing and information diffusion (Holme & Saramäki, 2012; Karsai et al., 2014). Temporal motifs, small subgraphs in temporal networks, consist of several nodes with patterns of interconnection, and links between nodes only occur at specific periods of time (Kovanen et al., 2013; Paranjape et al., 2017). The count of temporal motifs is crucial to understanding the structure, evolution and function of the complex systems modeled by the graph (Kovanen et al., 2013; Milo et al., 2002; Zhao et al., 2010).

Broadcast and viral diffusion are two main modes of information diffusion (Goel et al., 2015). Figure 3 shows a schematic depiction of the broadcast and viral diffusion. Broadcasting describes a large number of individuals receiving information from a single user, that is, a “one to many” process. In this process, the source nodes could be a social media user with a large number of followers, such as news outlets or celebrities (Wang et al., 2019). Virality describes that a piece of information reaches many individuals through multiple generations of peer-to-peer dissemination, that is, a “one-to-few” process. The source nodes in this process may be those users with a small number of followers (Wang et al., 2019).

To investigate diffusion patterns of elite and no-elite journal papers, we build a temporal network by the data of retweeting relationship and select four types of temporal motifs to characterize broadcast and viral effects. As shown in Figure 4, from Motif I to Motif IV, the average depth $\langle D \rangle$ of these motifs gradually increase, which symbolizes the transition from broadcast and viral diffusion. All users acquire information from a root user in Motif I. In Motif II, most users acquire information from a root user, few users receive information from an intermediate user. Most users receive information from an intermediate user in Motif III. Information in Motif IV is diffused by the peer-to-peer approach.

![Figure 4](image.png)

Figure 4. Four-node temporal motifs that are used to characterize the broadcast and viral mechanism

(3) Peak shape

To measure the bursty dynamics of elite and non-elite journal papers, we design several metrics based on the shape of the peaks. The definition of these metrics are as follows:

- **Peak number**: the number of peaks in the diffusion curve of a scientific paper.
- **Peak width**: the time interval between the end time and the start time of a peak.
• **Peak audience**: the number of users who participated in the spread of scientific papers between the start time and end time of a peak.

• **Peak interval**: the time interval between two adjacent peak time points.

(4) **Modularity**

*Modularity* is a measure of the structure of networks that evaluates the quality of network division (Newman, 2006; Newman & Girvan, 2004). It is defined as the number of edges falling within groups minus the expected number in an equivalent network with edges placed at random. The modularity is computed as follows:

\[
Q = \sum_i (e_{ii} - \alpha_i^2)
\]

where \(e_{ii}\) represents the fraction of edges in the network that connect nodes in the same community (i.e., intra-community links), and \(\alpha_i\) represents the fraction of edges that are connected to vertices in community \(i\). The range of modularity ranges from -0.5 to 1. Modularity is positive if the number of edges within a community exceeds the expected fraction in a random network. Networks with high modularity have dense connections between the nodes within modules (also called groups, clusters or communities) but sparse connections between nodes in different modules.

3 **Results**

This section consists of three parts: the first part presents the communication effect of elite and non-elite journal papers. The second part shows the bursty trends of the two categories of papers. The last one focuses on the differences in mechanisms by which elite and non-elite journal papers go viral.

3.1 **Communication effect**

To make a comparison of the diffusion characteristics of elite and non-elite journal papers, we use the complementary cumulative distribution function (CCDF) to display the distribution of variables, which describes the probability that a variable value is larger than \(x\). Figure 5a plots the CCDFs of diffusion scale for elite and non-elite journal papers. Of elite journal papers, there are 96.7% of papers less than 1000 in scale and 3.3% larger than 1000 in scale. Of non-elite journal papers, the proportion of papers with a scale smaller than 1000 is 96.2% and that of the papers whose diffusion scale exceeds 1000 is 3.8%. The median diffusion scale for elite journal papers is 186 and that of non-elite journal papers is 157. In terms of breadth, the proportion of elite and non-elite journal papers with a breadth less than 1000 is 96.7% and 97.6%, respectively (Figure 5b). The median breadth of elite journal papers is 158 and that of non-elite journal papers is 137. In the terms of average depth, the proportion of elite and non-elite journal papers with an average depth over 1.9 is 72.1% and 47.1%, respectively (Figure 5c). Figure 5d and Figure 5e show the mean average depth of elite and non-
elite journal papers at every scale and depth, respectively. The two figures display almost the same trends: the values of mean average depth of non-elite journal papers are mainly distributed between 0.4 and 0.6, while that of elite journal papers range from 0.1 to 0.3.

In summary, the gap of diffusion scale and breadth for elite and non-elite journal papers is narrow, while that of average depth for the two kinds of papers is significant. In other words, reaching the same level of scale and breadth with elite journal papers need to cross a longer network distance.

Figure 5. Complementary cumulative distribution functions (CCDFs) of scale, breadth and average depth of diffusion network for elite and non-elite journal papers. (a) Scale. (b) Breadth. (c) Average depth. (d) The mean average depth of elite and non-elite journal papers at every scale. (e) The mean average depth of elite and non-elite journal papers at every breadth.

3.2 Bursty trends

In order to compare the bursty trends of elite and non-elite journal papers, we characterize the bursty trends of the two categories of papers by the peaks in the diffusion curves and conduct a peak detection by the find_peaks function of python package Signal processing. This package not only finds the location of a peak but also return the start and end time of a peak. All parameters are set to default values except for setting the parameters of prominence to 10. We resample the retweet time series of scientific papers in one-day intervals such that each peak can be relatively well detected.

---

4 https://docs.scipy.org/doc/scipy/reference/signal.html
Figure 7a shows the distribution of the number of peaks for elite and non-elite journal papers. From the figure, we can see that nearly 41% of non-elite journal papers have a peak number larger than 3 while the proportion of elite journal papers with the number of peaks greater than 3 is about 33%, suggesting that the spread of non-elite journal papers show more peaks in their diffusion curves. In Figure 7b, about 45% of peaks of elite journal papers last longer than 6 days, whereas the peaks of non-elite journal papers with a duration time longer than 6 days account for 30%, which means that the peaks of elite journal papers last longer than non-elite journal papers. In Figure 7c, the proportion of elite journal papers with an audience size larger than 90 account for 31% while that of non-elite journal papers with an audience size greater than 14%. The median of the former is 52 and that of the latter is 35, which demonstrates that the peak height of elite journal papers is much higher than that of non-elite journal papers. In the aspect of peak interval, it takes an average of 30 days for elite journal papers to obtain a renewed burst of retweets while the average time for non-elite journal papers to experience a renewed burst of retweets is 83 days (Figure 7d). That is to say, non-elite journal papers experience a longer waiting time to obtain a new round of bursts of retweets. Overall, the diffusion trends of elite journal papers show a pattern of a small number of persistent, dense and large-scale bursts, while that of non-elite journal papers is characterized by multiple short-lived, sparse and small-scale bursts.

![Figure 7. CCDFs of the burstiness of elite and non-elite journal papers: (a) The number of peaks in the diffusion curve; (b) The width of peaks; (c) The audience size of peaks; (d) The interval between two adjacent peaks](image)
3.3 Diffusion mechanisms

The results in the above two subsections indicate that there are clear differences in diffusion network structure and bursty trends between elite and non-elite journal papers. In the following subsections, we try to disclose the underlying mechanism behind the diffusion of elite and non-elite papers and the differences in diffusion mechanisms.

(1) Broadcast and viral diffusion

Broadcasting and virality are two main diffusion modes of information diffusion (Goel et al., 2015; Rogers, 2010). This study depicts the two diffusion modes of elite and non-elite journal papers by temporal motifs. Figure 6 shows the growth of the number of four types of temporal motifs we choose. The x-axis represents the time windows, $T$, with different time scales, for example, $T$ equal to $x$ means the size of the time window ranges from 0 to $x$. The y-axis represents the number of temporal motifs within corresponding time windows. When the window of observation time is 10 seconds, the count of Motif I is larger than 0, while that of the other three types of motifs is equal to 0 (Figure 6a-d). As the time scales increase to $10^3$ seconds, the counts of Motif I, II, III show a significant increasing trend, whereas no growth is visible in Motif IV. When the time scale is $10^5$, the count of four types of motifs experiences a significant surge, which demonstrates that the two classes of papers are diffused via both pure broadcast and viral mechanisms, as well as essentially all conceivable combinations of the two (Figure 6e-h).

In terms of Motif I and II, the motif counts of elite journal papers are generally larger than those of non-elite journal papers. In terms of Motif III and IV, the motif counts of non-elite papers are larger than those of elite journal papers, which shows that elite journal papers are more inclined to broadcast diffusion, while non-elite journals are more inclined to viral diffusion. We also find that when the time window is smaller than $10^4$, the count of Motif I and II of elite journal papers is larger than that of non-elite journal papers (Figure 6a and Figure 6b). The count of Motif III and IV of non-elite journal papers is greater than that of elite journal papers when the window of observation time is larger than $10^4$ (Figure 6c and Figure 6d). This means that elite journal papers are more inclined to broadcast diffusion at smaller time scales, and non-elite journal papers prefer viral diffusion at larger time scales.
Figure 6. The growth of the number of four types of temporal motifs at different time scales. \( T \) is the time window. (a)–(d) Cumulative growth of Motif I, II, III and IV counts. (d)–(e) Net growth of Motif I, II, III and IV counts

(2) **Cross-community patterns**

Previous studies show that community structure plays an important role in information diffusion (Galstyan & Cohen, 2007; Gleeson, 2008; Grabowicz et al., 2012). For example, Weng et al. (2013) found that the more communities a meme permeated, the more viral it is. To compare the cross-community patterns of elite and non-elite journals, we perform community detection on diffusion networks of elite and non-elite journal papers by Lovin algorithms (Blondel et al., 2008) and use the metrics of *modularity* to measure the quality of the community division (Newman, 2006; Newman & Girvan, 2004). A network with high modularity has dense connections between the nodes within communities but sparse connections between nodes in different communities. Figure 8 shows the patterns of cross-community diffusion of elite and non-elite journal papers. Figure 8a compares the number of communities of elite and non-elite journal papers at each burst. From the figure, we can see that elite journal papers are diffused to many communities at the first bursts, but only to a few communities in subsequent bursts. Non-elite journal papers are disseminated to a few communities in each burst, but they experience more rounds of bursts than the elite. Their final cumulative number of communities exceeds that of elite journal papers in the subsequent bursts. Figure 8b shows the community size of elite and non-elite journal papers at each burst. Elite and non-elite journal papers are adopted by a larger number of people at the first bursts, while they receive a little attention in the subsequent bursts. The final cumulative community size of non-elite journal papers exceeds that of elite journal papers as the number of bursts increases. We further compare the modularity and the inter-community links of elite and non-elite journal papers. As shown in Figure 8c, a greater
fraction of non-elite journal papers has modularity distributing between 0 and 0.6, while a greater fraction of elite journal papers has modularity ranging from 0.6 to 1.0. That is to say, elite journal papers show a clearer community structure than the non-elite. Figure 8d shows the CCDFs of the number of inter-community links for elite and non-elite journal papers. About 31% of elite journal papers have inter-community links larger than 10, while nearly 46% of non-elite journal papers do.

To sum up, the number and size of community for elite and non-elite journal papers decays as the round of bursts increases, while the final cumulative community number and size of non-elite journal papers are larger than that of the elite. Elite journal papers show a clearer community structure than the non-elite. The audience of elite journal papers comes from many communities with sparse links in between, while that of non-elite journal papers comes from several communities with dense links in between.

Figure 8 Patterns of cross-community of elite and non-elite journal papers. (a) The number of communities at the kth burst. (b) The community size at the kth burst. (c) The distribution of community modularity. (d) CCDFs of the number of inter-community links.
4 Conclusion and discussion

This study compares the online diffusion of scientific papers published in different levels of journals. The results show that in order to obtain large-scale attention, non-elite journals paper need to pay more effort than elite journal papers: they need to spend longer network distance and spreading time to cumulate audience, exhibiting a diffusion trend with multiple rounds, sparse, short-duration and small-scale bursts. Besides, elite journal papers tend to broadcast diffusion, while non-elite journal papers are inclined to viral diffusion. In other words, the diffusion of elite journal papers may be dominated by some users with a large number of followers. With their enormous social influence and penetration, they could attract a large audience from different backgrounds easily and quickly. By contrast, the diffusion of non-elite journal papers may originate from some grassroots users with a small number of followers. Only by constantly being accepted and disseminated by their friends could they reach a massive audience.

Our findings have two-sided implications. On the one hand, online diffusion of scientific papers shares some commonalities with that of other online content such as news, videos and picture (Goel et al., 2015, 2012): elite and non-elite journal papers are diffused via a mixture of broadcast and viral mechanism. On the other hand, our experimental results show that the cumulative advantage of high-impact journals can be extended from offline to online channels, making a great difference in the dissemination of scientific papers. Specifically, elite journal papers show a clear cumulative advantage in terms of peak, density, and recurrence of bursts, as well as communities links, and seem to reach their target audience without much effort. In contrast, it is not easy for non-elite journal papers to succeed on social media because they have to experience more processes and rely on stronger community relationships to further spread during the diffusion cycle.

Despite exerting a persistent influence in the social media environment, the cumulative advantage brought by journal impact seems to have been weakened to a certain extent. For example, a recent study showed that there is very little overlap between very highly cited papers and those that received the highest altmetric scores (Banshal et al., 2018). We assume that this phenomenon could be attributed to two aspects. On the one hand, papers are available from many digital platforms in the era of Web 2.0, no longer physically tied to journals. Hence, papers can be read and cited based on their own merits, independently of the journal’s physical availability, reputation, or impact factor (Lozano et al., 2012). On the other hand, on many social media platforms like Twitter and Facebook, the willingness of a user to spread (tweet or retweet) a paper strongly depends on his/her online social network and the overlap between the content of the paper and personal interest rather than scientific merits (Haustein et al., 2014; Lagnier et al., 2013; Zhang et al., 2015). Therefore, we encourage scholars to positively share their research outputs online, even if those works are not published in those prestigious journals.

Some limitations should be clarified in this study. First, we do not consider the differences in the distribution of subject areas. For example, do papers from the subject
areas of biomedicine are easier to attract large-scale online attention? Besides, the differences between key opinion leaders who drive the diffusion of scientific papers are unclear. Second, although we disclosed the differences of online diffusion of elite and non-elite journal papers based on our current dataset (180 journals), employing a larger-scale dataset may make our results more robust. Finally, the metric we used for journal ranking considers self-citations, and the value of the metric depends on the size of subject areas, which may not truly reflect the journal impact to some extent. Future studies will further perfect our work and try to conduct a systematical analysis of factors driving scientific papers to go viral online, including the content of scientific papers, temporal trends, network structure and other aspects. Besides, we plan to use the current research paradigm to compare the differences of other types of information (e.g. true and false news) in temporal dynamics (Juul & Ugander, 2021; Vosoughi et al., 2018).

Acknowledgments

We gratefully acknowledge the Grant from the National Natural Science Foundation of China (71673038, 71974029), thank Jinlian Zhou for discussing the paper and Altmetric.com for providing the altmetric data of scientific publications.

References

Almanza, M., Lattanzi, S., Panconesi, A., & Re, G. (2021). Twin Peaks, a Model for Recurring Cascades. Proceedings of the Web Conference 2021, 681–692. https://doi.org/10.1145/3442381.3449807

Alperin, J. P., Gomez, C. J., & Haustein, S. (2019). Identifying diffusion patterns of research articles on Twitter: A case study of online engagement with open access articles. Public Understanding of Science, 28(1), 2–18. https://doi.org/10.1177/0963662518761733

Banshal, S. K., Basu, A., Singh, V. K., & Muhuri, P. K. (2018). Scientific vs. Public Attention: A Comparison of Top Cited Papers in WoS and Top Papers by Altmetric Score. In M. Erdt, A. Sesagiri Raamkumar, E. Rasmussen, & Y.-L. Theng (Eds.), Altmetrics for Research Outputs Measurement and Scholarly Information Management (pp. 81–95).
Bik, H. M., & Goldstein, M. C. (2013). An Introduction to Social Media for Scientists. *PLOS Biology, 11*(4), e1001535. https://doi.org/10.1371/journal.pbio.1001535

Blondel, V. D., Guillaume, J.-L., Lambiotte, R., & Lefebvre, E. (2008). Fast unfolding of communities in large networks. *Journal of Statistical Mechanics: Theory and Experiment, 2008*(10), P10008. https://doi.org/2008101003130400

Broxton, T., Interian, Y., Vaver, J., & Wattenhofer, M. (2013). Catching a viral video. *Journal of Intelligent Information Systems, 40*(2), 241–259. https://doi.org/10.1007/s10844-011-0191-2

Cao, R., Wang, X., Xu, X., & Zhou, J. (2021). Multiple bursts of highly retweeted articles on social media. *Scientometrics*. https://doi.org/10.1007/s11192-021-03970-7

Cao, R.-M., Liu, X. F., & Xu, X.-K. (2021). Why cannot long-term cascade be predicted? Exploring temporal dynamics in information diffusion processes. *Royal Society Open Science, 8*(9), 202245. https://doi.org/10.1098/rsos.202245

Cheng, J., Adamic, L. A., Kleinberg, J. M., & Leskovec, J. (2016). Do cascades recur? *Proceedings of the 25th International Conference on World Wide Web*, 671–681. https://doi.org/10.1145/2872427.2882993

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. https://doi.org/10.1108/AJIM-12-2014-0173

Crane, R., & Sornette, D. (2008). Robust dynamic classes revealed by measuring the response
function of a social system. *Proceedings of the National Academy of Sciences*, **105**(41), 15649–15653. https://doi.org/10.1073/pnas.0803685105

Cui, Y., Xu, X., Cao, R., Fang, Z., Zhou, J., & Wang, X. (2019). Burst diffusion of highly retweeted scholarly articles in Social Media. In G. Catalano, C. Daraio, M. Gregori, H. F. Moed, & G. Ruocco (Eds.), *17th International Conference on Scientometrics & Informetrics (issi2019)*, Vol I (pp. 1166–1177). Int Soc Scientometrics & Informetrics-Issi. https://www.webofscience.com/wos/alldb/full-record/WOS:000508217900115

Darling, E. S., Shiffman, D., Côté, I. M., & Drew, J. A. (2013). The role of twitter in the life cycle of a scientific publication. *ArXiv:1305.0435 [Physics, q-Bio].* http://arxiv.org/abs/1305.0435

De Domenico, M., & Altmann, E. G. (2020). Unraveling the Origin of Social Bursts in Collective Attention. *Scientific Reports*, **10**(1), 4629. https://doi.org/10.1038/s41598-020-61523-z

Didegah, F., & Thelwall, M. (2018). Co-saved, co-tweeted, and co-cited networks. *Journal of the Association for Information Science and Technology*, **69**(8), 959–973. https://doi.org/10.1002/asi.24028

Galstyan, A., & Cohen, P. (2007). Cascading dynamics in modular networks. *Physical Review E*, **75**(3), 036109. https://doi.org/10.1103/PhysRevE.75.036109

Gleeson, J. P. (2008). Cascades on correlated and modular random networks. *Physical Review E*, **77**(4), 046117. https://doi.org/10.1103/PhysRevE.77.046117

Goel, S., Anderson, A., Hofman, J., & Watts, D. J. (2015). The Structural Virality of Online Diffusion. *Management Science*, **62**(1), 180–196.
Goel, S., Watts, D. J., & Goldstein, D. G. (2012). The structure of online diffusion networks. *Proceedings of the 13th ACM Conference on Electronic Commerce*, 623–638. https://doi.org/10.1145/2229012.2229058

Grabowicz, P. A., Ramasco, J. J., Moro, E., Pujol, J. M., & Eguiluz, V. M. (2012). Social Features of Online Networks: The Strength of Intermediary Ties in Online Social Media. *PLOS ONE, 7*(1), e29358. https://doi.org/10.1371/journal.pone.0029358

Haustein, S., Bowman, T. D., & Costas, R. (2016). Interpreting ‘Altmetrics’: Viewing Acts on Social Media through the Lens of Citation and Social Theories. In *Interpreting ‘Altmetrics’: Viewing Acts on Social Media through the Lens of Citation and Social Theories* (pp. 372–406). De Gruyter Saur. https://doi.org/10.1515/9783110308464-022

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. https://doi.org/10.1002/asi.23101

Hitlin, P., & Olmstead, K. (2018, March 21). The Science People See on Social Media. *Pew Research Center Science & Society.* https://www.pewresearch.org/science/2018/03/21/the-science-people-see-on-social-media/

Holme, P., & Saramäki, J. (2012). Temporal networks. *Physics Reports, 519*(3), 97–125. https://doi.org/10.1016/j.physrep.2012.03.001
Imran M., Akhtar A., Said A., Saﬁder I., Hassan S.U., & Aljohani N.R. (2018). Exploiting Social Networks of Twitter in Altmetrics Big Data. *STI 2018 Conference Proceedings*, 1339–1344. http://hdl.handle.net/1887/65219

Juul, J. L., & Ugander, J. (2021). Comparing information diffusion mechanisms by matching on cascade size. *Proceedings of the National Academy of Sciences, 118*(46). https://doi.org/10.1073/pnas.2100786118

Karsai, M., Perra, N., & Vespignani, A. (2014). Time varying networks and the weakness of strong ties. *Scientific Reports, 4*(1), 4001. https://doi.org/10.1038/srep04001

Kovanen, L., Kaski, K., Kertész, J., & Saramäki, J. (2013). Temporal motifs reveal homophily, gender-speciﬁc patterns, and group talk in call sequences. *Proceedings of the National Academy of Sciences, 110*(45), 18070–18075. https://doi.org/10.1073/pnas.1307941110

Kupavskii, A., Ostroumova, L., Umnov, A., Usachev, S., Serdyukov, P., Gusev, G., & Kustarev, A. (2012). Prediction of retweet cascade size over time. *Proceedings of the 21st ACM International Conference on Information and Knowledge Management*, 2335–2338. https://doi.org/10.1145/2396761.2398634

Lagnier, C., Denoyer, L., Gaussier, E., & Gallinari, P. (2013). Predicting Information Diffusion in Social Networks Using Content and User’s Proﬁles. In P. Serdyukov, P. Braslavski, S. O. Kuznetsov, J. Kamps, S. Rüger, E. Agichtein, I. Segalovich, & E. Yilmaz (Eds.), *Advances in Information Retrieval* (pp. 74–85). Springer. https://doi.org/10.1007/978-3-642-36973-5_7

Leskovec, J., Backstrom, L., & Kleinberg, J. (2009). Meme-tracking and the dynamics of the
news cycle. *Proceedings of the 15th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 497–506. https://doi.org/10.1145/1557019.1557077

Lozano, G. A., Larivière, V., & Gingras, Y. (2012). The weakening relationship between the impact factor and papers’ citations in the digital age. *Journal of the American Society for Information Science and Technology*, 63(11), 2140–2145. https://doi.org/10.1002/asi.22731

Lü, L., Zhang, Y.-C., Yeung, C. H., & Zhou, T. (2011). Leaders in Social Networks, the Delicious Case. *PLOS ONE*, 6(6), e21202. https://doi.org/10.1371/journal.pone.0021202

Matsubara, Y., Sakurai, Y., Prakash, B. A., Li, L., & Faloutsos, C. (2012). Rise and fall patterns of information diffusion: Model and implications. *Proceedings of the 18th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 6–14. https://doi.org/10.1145/2339530.2339537

Milo, R., Shen-Orr, S., Itzkovitz, S., Kashtan, N., Chklovskii, D., & Alon, U. (2002). Network Motifs: Simple Building Blocks of Complex Networks. *Science*, 298(5594), 824–827. https://doi.org/10.1126/science.298.5594.824

Newman, M. E. J. (2006). Modularity and community structure in networks. *Proceedings of the National Academy of Sciences of the United States of America*, 103(23), 8577–8582. https://doi.org/10.1073/pnas.0601602103

Newman, M. E. J., & Girvan, M. (2004). Finding and evaluating community structure in networks. *Physical Review E*, 69(2), 026113. https://doi.org/10.1103/physreve.69.026113
Paranjape, A., Benson, A. R., & Leskovec, J. (2017). Motifs in Temporal Networks. *Proceedings of the Tenth ACM International Conference on Web Search and Data Mining*, 601–610. https://doi.org/10.1145/3018661.3018731

Robinson-García, N., Torres-Salinas, D., Zahedi, Z., & Costas, R. (2014). New data, new possibilities: Exploring the insides of Altmetric.com. https://doi.org/10.3145/epi.2014.jul.03

Rogers, E. M. (2010). *Diffusion of Innovations, 4th Edition*. Simon and Schuster.

Said, A., Bowman, T. D., Abbasi, R. A., Aljohani, N. R., Hassan, S.-U., & Nawaz, R. (2019). Mining network-level properties of Twitter altmetrics data. *Scientometrics, 120*(1), 217–235. https://doi.org/10.1007/s11192-019-03112-0

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

Sugimoto, C. R., Work, S., Larivière, V., & Haustein, S. (2017). Scholarly use of social media and altmetrics: A review of the literature. *Journal of the Association for Information Science and Technology, 68*(9), 2037–2062. https://doi.org/10.1002/asi.23833

Suh, B., Hong, L., Pirolli, P., & Chi, E. H. (2010). Want to be Retweeted? Large Scale Analytics on Factors Impacting Retweet in Twitter Network. *2010 IEEE Second International Conference on Social Computing*, 177–184. https://doi.org/10.1109/SocialCom.2010.33

*The Science People See on Social Media | Pew Research Center*. (n.d.). Retrieved September 11, 2021, from https://www.pewresearch.org/science/2018/03/21/the-science-
people-see-on-social-media/

Vosoughi, S., Roy, D., & Aral, S. (2018). The spread of true and false news online. *Science*, *359*(6380), 1146–1151. https://doi.org/10.1126/science.aap9559

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. https://doi.org/10.1002/asi.23040

Wang, X., Chen, L., Shi, J., & Peng, T.-Q. (2019). What makes cancer information viral on social media? *Computers in Human Behavior*, *93*, 149–156. https://doi.org/10.1016/j.chb.2018.12.024

Weng, L., Menczer, F., & Ahn, Y.-Y. (2013). Virality Prediction and Community Structure in Social Networks. *Scientific Reports*, *3*(1), 2522. https://doi.org/10.1038/srep02522

Wu, S., Tan, C., Kleinberg, J., & Macy, M. (2011). Does Bad News Go Away Faster? *Proceedings of the International AAAI Conference on Web and Social Media*, *5*(1), 646–649.

Yang, J., & Leskovec, J. (2011). Patterns of temporal variation in online media. *Proceedings of the Fourth ACM International Conference on Web Search and Data Mining*, 177–186. https://doi.org/10.1145/1935826.1935863

Yin, Y., Gao, J., Jones, B. F., & Wang, D. (2021). Coevolution of policy and science during the pandemic. *Science*, *371*(6525), 128–130. https://doi.org/10.1126/science.abe3084

Yu, L., & Yu, H. (2016). Does the average JIF percentile make a difference? *Scientometrics*, *109*(3), 1979–1987. https://doi.org/10.1007/s11192-016-2156-2

Zakhlebin, I., & Horvát, E.-Á. (2020). Diffusion of Scientific Articles across Online Platforms.
Zhan, C., Wu, F., Huang, Z., Jiang, W., & Zhang, Q. (2020). Analysis of collective action propagation with multiple recurrences. *Neural Computing and Applications, 32*(17), 13491–13504. https://doi.org/10.1007/s00521-020-04756-3

Zhang, J., Tang, J., Li, J., Liu, Y., & Xing, C. (2015). Who Influenced You? Predicting Retweet via Social Influence Locality. *ACM Transactions on Knowledge Discovery from Data, 9*(3), 25:1–25:26. https://doi.org/10.1145/2700398

Zhao, Q., Tian, Y., He, Q., Oliver, N., Jin, R., & Lee, W.-C. (2010). *Communication motifs: A tool to characterize social communications* (p. 1648). https://doi.org/10.1145/1871437.1871694

Zheng, H., Aung, H. H., Erdt, M., Peng, T.-Q., Sesagiri Raamkumar, A., & Theng, Y.-L. (2019). Social media presence of scholarly journals. *Journal of the Association for Information Science and Technology, 70*(3), 256–270. https://doi.org/10.1002/asi.24124