

RESEARCH ARTICLE

# Mention effect in information diffusion on a micro-blogging network

Peng Bao<sup>1\*</sup>, Hua-Wei Shen<sup>2</sup>, Junming Huang<sup>3,4</sup>, Haiqiang Chen<sup>5</sup>

**1** School of Software Engineering, Beijing Jiaotong University, Beijing, China, **2** CAS Key Laboratory of Network Data Science and Technology, Institute of Computing Technology, Chinese Academy of Sciences, Beijing, China, **3** CompleX Lab, Web Sciences Center and Big Data Research Center, University of Electronic Science and Technology of China, Chengdu, China, **4** Center for Complex Network Research, Northeastern University, Boston, MA, United States of America, **5** China Information Technology Security Evaluation Center, Beijing, China

\* [baopeng@bjtu.edu.cn](mailto:baopeng@bjtu.edu.cn)



**OPEN ACCESS**

**Citation:** Bao P, Shen H-W, Huang J, Chen H (2018) Mention effect in information diffusion on a micro-blogging network. PLoS ONE 13(3): e0194192. <https://doi.org/10.1371/journal.pone.0194192>

**Editor:** Kazutoshi Sasahara, Nagoya University, JAPAN

**Received:** January 24, 2017

**Accepted:** February 14, 2018

**Published:** March 20, 2018

**Copyright:** © 2018 Bao et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

**Data Availability Statement:** Data are available from WISE 2012 Challenge at <http://www.wise2012.cs.ucy.ac.cy/challenge.html>. Others are able to access the data in the same manner as the authors and the authors did not have any special access privileges.

**Funding:** This work was funded by the National Natural Science Foundation of China under grant number 61702031, 61472400 and 61673150, the Beijing Excellent Talents Supporting Program under grant number 2017000020124G054, and the Fundamental Research Funds for the Central

## Abstract

Micro-blogging systems have become one of the most important ways for information sharing. Network structure and users' interactions such as forwarding behaviors have aroused considerable research attention, while *mention*, as a key feature in micro-blogging platforms which can improve the visibility of a message and direct it to a particular user beyond the underlying social structure, is seldom studied in previous works. In this paper, we empirically study the mention effect in information diffusion, using the dataset from a population-scale social media website. We find that users with high number of followers would receive much more mentions than others. We further investigate the effect of mention in information diffusion by examining the response probability with respect to the number of mentions in a message and observe a saturation at around 5 mentions. Furthermore, we find that the response probability is the highest when a reciprocal followship exists between users, and one is more likely to receive a target user's response if they have similar social status. To illustrate these findings, we propose the response prediction task and formulate it as a binary classification problem. Extensive evaluation demonstrates the effectiveness of discovered factors. Our results have consequences for the understanding of human dynamics on the social network, and potential implications for viral marketing and public opinion monitoring.

## Introduction

Nowadays, the subject of information diffusion is central to an information era from knowledge database to online media [1–4]. Information diffusion is a fundamental process in social network, capturing behaviors that cascade from node to node like an epidemic or chain reaction. Recent studies devote to investigating the diffusion process of different type of information, such as text [5], image [6], video [7], etc. On the micro-blogging platforms such as Twitter and Sina Weibo, users can post any topic of messages no longer than 140 characters and follow any other users to receive their messages. Moreover, with various sharing features

Universities under grant number 2015RC031. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

**Competing interests:** The authors have declared that no competing interests exist.

on the platform, every user owns the power to effectively spread information beyond the underlying followship structure [8, 9]. In recent years, researchers have paid great effort on user interest modeling [10–12], influential users identification [13–15] and recommendations [16–19]. Understanding the mechanisms of information diffusion is especially critical in a wide range of areas, such as human dynamics [20–22], popularity prediction [5, 23–26], and viral marketing [27, 28].

The characteristics of network structure and user relationship of micro-blogs have aroused considerable research interests in the past few years [29, 30]. Recently, researchers have paid extensive attention on characterizing information cascades [31, 32], discovering structural and temporal patterns [33–35], and further predicting individual behaviors and popularity dynamics [36–39]. However, previous works mainly focused on the re-tweeting or forwarding behaviors, assuming that information diffusion relies on the underlying followship network among users. One can be exposed to a message only when he/she has already followed the publisher or spreader of the message. Therefore, the scale of the diffusion would be limited due to the visibility restriction [40, 41].

However, as a key feature in micro-blogging platforms, *mention* can improve the visibility of a message and direct it to a particular user beyond the underlying social structure. A user uses the “@username” to mention another user in the body of a message, so that the mentioned user will see the message in his/her personal mention tab. One’s followers would also easily miss a message in time if there’s no notification [13]. Therefore, with the proper usage of mention, an ordinary user has the potential to break through the visibility barrier and spread his/her messages broadly. In recent years, the essential question of whom-to-mention in a message has been studied extensively. Most previous works formulate the problem as a ranking based recommendation task [42–46], while some researchers take it as a link prediction problem [47] or an unbalance assignment problem [48]. Besides, different aspects of factors have been investigated, such as content [44, 46, 49], social influence [46, 50], spatiotemporal information [44, 48, 51], and the interests of users [50, 52]. However, the underlying microscopic factors governing the effectiveness of mention still need to be explored. Therefore, it is still an open problem and of great interest to present an in-depth study of mention effect in information diffusion on social networks.

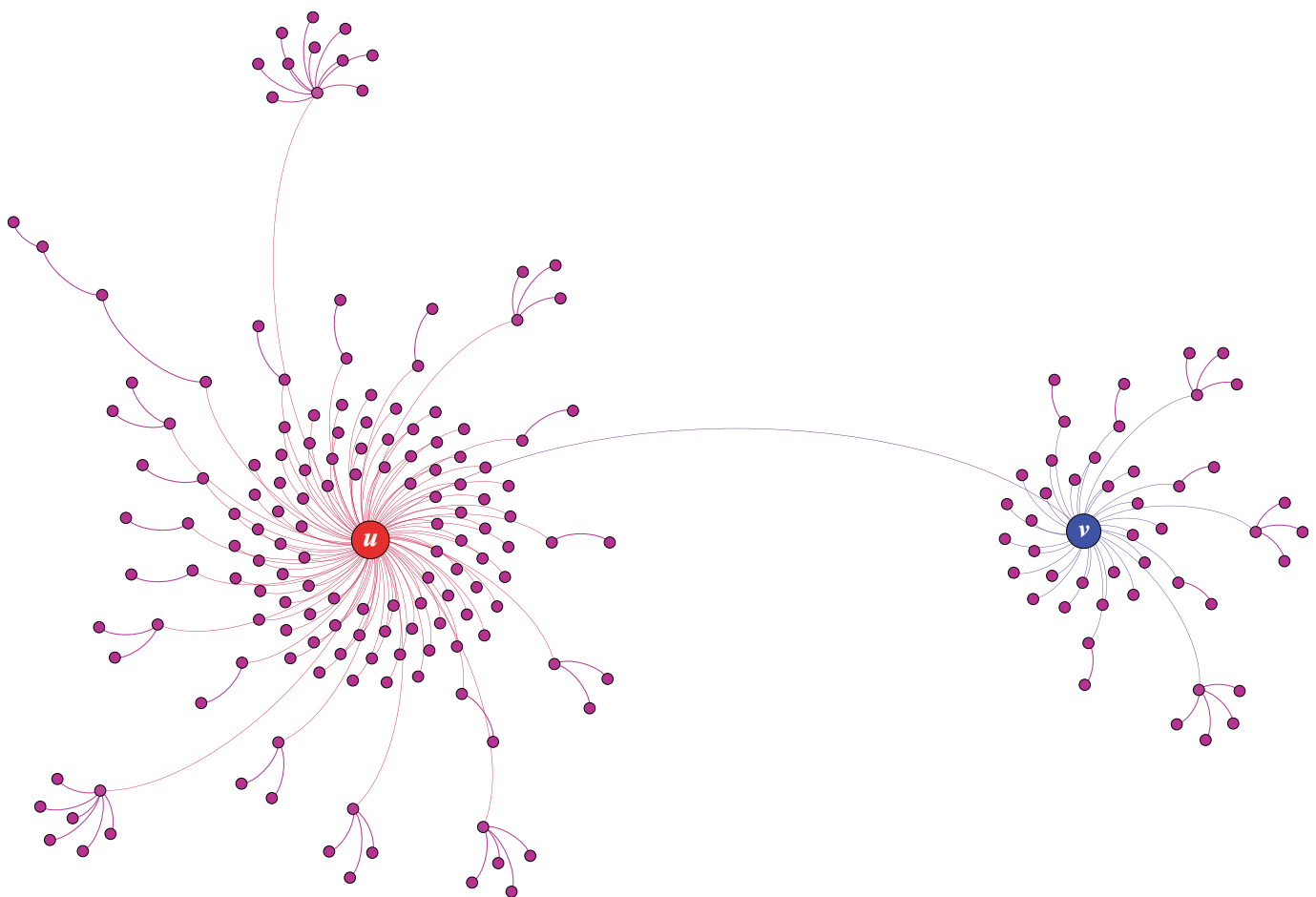
In this paper, in order to investigate the mention effect in information diffusion on micro-blogging networks, a comprehensive empirical study is conducted on the most popular micro-blogging website in China, namely Sina Weibo. Note that here the unit of information refers to a message on the micro-blogging network, and we use the forwarding behaviors among users as the proxy for information diffusion, which is also widely adopted in previous studies [23, 53]. We start with the statistical characteristics of mention in information diffusion represented as diffusion network. We find that users with high number of followers would receive much more mentions than those small-degree users, and meanwhile it brings the problem of mention overload. We further investigate the effect of mention in information diffusion by examining the response probability from the perspective of mention count and network structure respectively. We observe a saturation at around 5 mentions in a message which means that with each additional mention, a message is more and more likely to receive response from the mentioned users, up to a point. When a message contains more than 5 mentioned users, the response probability increase marginally. Furthermore, we examine the response probability with respect to the network structure between users. We find that the response probability is the highest when a reciprocal followship exists between users. In addition, one is more likely to receive a target user’s response if they have similar social status. To illustrate these findings, we propose the response prediction task and formulate it as a binary classification problem. Extensive evaluation demonstrates the effectiveness of discovered factors.

## Results

### Diffusion network

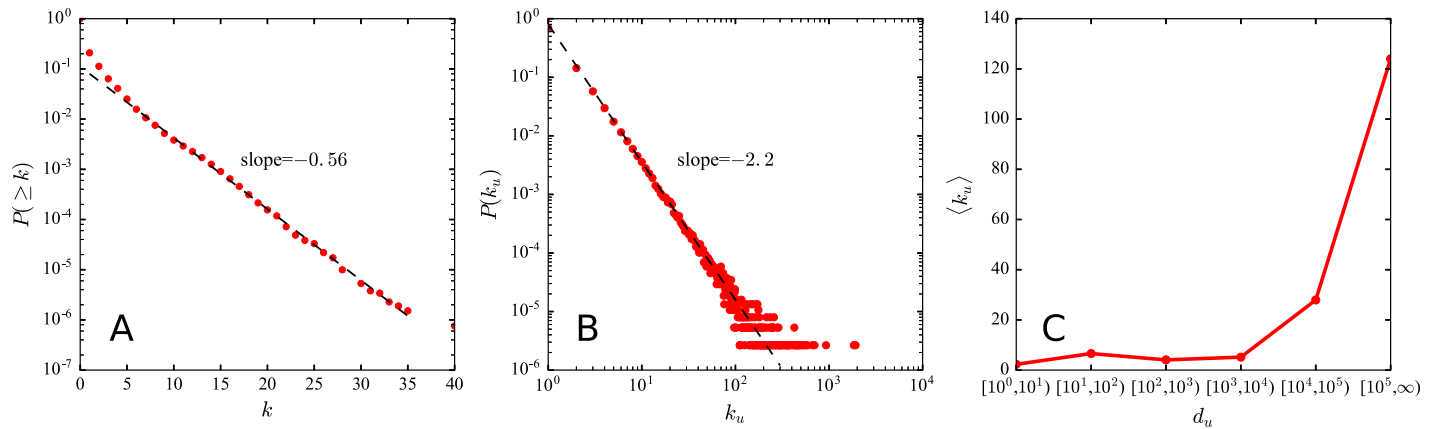
To begin our analysis, the cascade of message is represented as a diffusion network which characterizes the relationship among users who involve in the diffusion process. In this paper, we construct the diffusion network as a directed network where each node represents an involved user in the diffusion process and each link denotes an observed forwarding behavior between two users, as done in [53]. In addition, we call the node which corresponds to the source user of message the root node of diffusion network. Fig 1 gives an example of diffusion network of a message containing mentions, which is derived from a real cascade. The root node  $u$  initiated the cascade of the message with a mentioned user which is marked by the blue node  $v$ . Soon after being mentioned, node  $v$  forwarded the message and further triggered a new spread of the message. In this paper, we define the be-mentioned user's forwarding behavior towards the message as *response*.

We first investigate the statistics of the number of '@' in a message. As depicted in Fig 2A, more than 20% of messages are posted with at least one mentioned user. Due to the strict length restriction of a message, only a small number of users can be mentioned in a message, which follows an exponential distribution with an exponent 0.56. We further examine the



**Fig 1. Diffusion network of a message containing mentions.** This example is derived from a real cascade. The node  $u$  initiated the cascade of the message and mentioned the node  $v$ . We observed that the be-mentioned node  $v$  forwarded the message and further triggered a new spread of the message.

<https://doi.org/10.1371/journal.pone.0194192.g001>



**Fig 2. Statistical characteristics of mention.** (A) Cumulative distribution  $P(\geq k)$  where  $k$  denotes the number of '@' in a message. The cumulative distribution of  $k$  is exponential with an exponent 0.56. We can also observe that more than 20% of messages are posted with at least one mentioned user. (B) Distribution  $P(k_u)$  where  $k_u$  denotes the number of '@' user  $u$  received. It indicates a power law interdependence with an exponent 2.2. (C) Average number of received '@'  $\langle k_u \rangle$  versus the number of followers  $d_u$  for each user  $u$ . We classify users into six categories according to  $d_u$ . We find that users with large number of followers receive much more '@' than small-degree users.

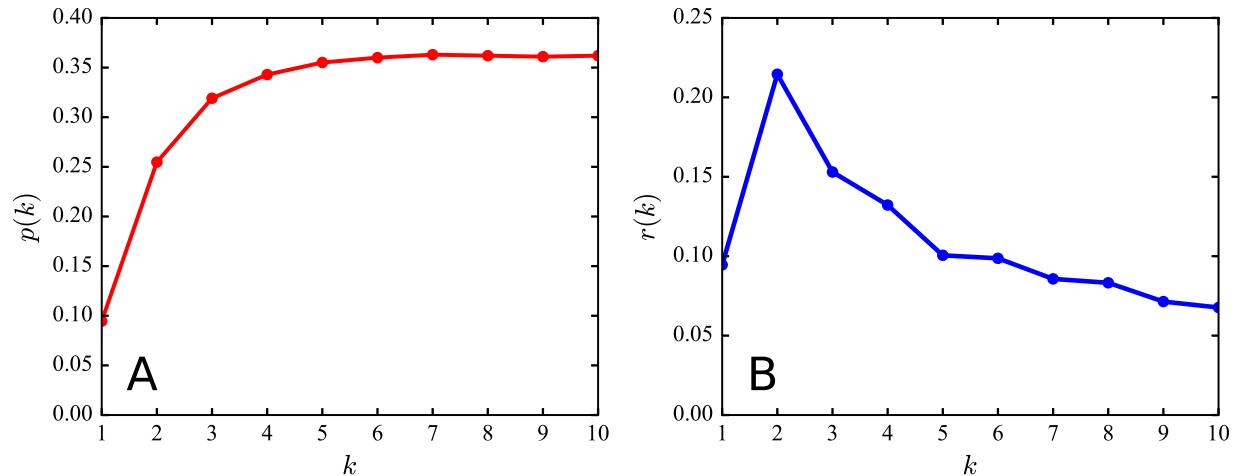
<https://doi.org/10.1371/journal.pone.0194192.g002>

distribution of the number of '@' a user would receive. From Fig 2B, we can observe a power law distribution with an exponent 2.2, indicating that mention is allocated in a rather asymmetric way, with a majority of users getting a few mentions, whereas a few receive a disproportionate number of mentions. Users with high number of followers on the explicit followship network are usually called “opinion leaders” [54] and are indispensable to the popularity of a message. Hence they would receive much more mentions than those small-degree users, as shown in Fig 2C. However, a user being mentioned too many times will suffer from the severe mention overload problems. Tons of mention notifications will interrupt user’s daily use of micro-blogs and decrease user’s interest in forwarding.

### Effect of mention count

Given the statistical characteristics of mention in information diffusion, here we ask: How about the effect of mention? Are there any factors that would affect mention effect? Can we predict the response of those be-mentioned users? To address these questions, we first quantify the effect of mention in terms of **response probability**. Specifically, for a given message with  $k$  mentioned users, we define response probability  $p(k)$  as the probability that at least one of the  $k$  mentioned users will forward the message. Here we just consider the forwarding behavior as a sign of direct response. We denote  $M(k)$  the number of messages with  $k$  be-mentioned user, and  $R(k)$  the number of messages that receive response by at least one of the  $k$  mentioned users. We then conclude that  $p(k) = \frac{R(k)}{M(k)}$  is the corresponding response probability.

One would expect that a message is more likely to receive response if it contains more mentions. On the other hand, one would also think that there is a saturation point. With the above definition, we empirically study the response probability  $p(k)$  using all messages forwarded by more than 10 users. Taking the activity pattern of users on the platform into consideration, we thus only consider messages with the post time between 10am and 10pm per day, as done in [53]. Fig 3A shows  $p(k)$  with respect to the number of mentions  $k$  in a message. We observe a saturation at around 5 mentions. This means that with each additional mention, a message is more and more likely to receive response from the mentioned users, up to a point. When a message contains more than 5 mentioned users, the response probability increase marginally.



**Fig 3. Effect of mention count.** (A) Response probability  $p(k)$  versus mention count  $k$  in a message. We find that response probability increases with more and more mentions in a message, up to a saturation point around  $k = 5$ . (B) Response ratio  $r(k)$  versus mention count  $k$  in a message. We can observe a peak in response ratio at 2 mentions in a message and then a slow drop.

<https://doi.org/10.1371/journal.pone.0194192.g003>

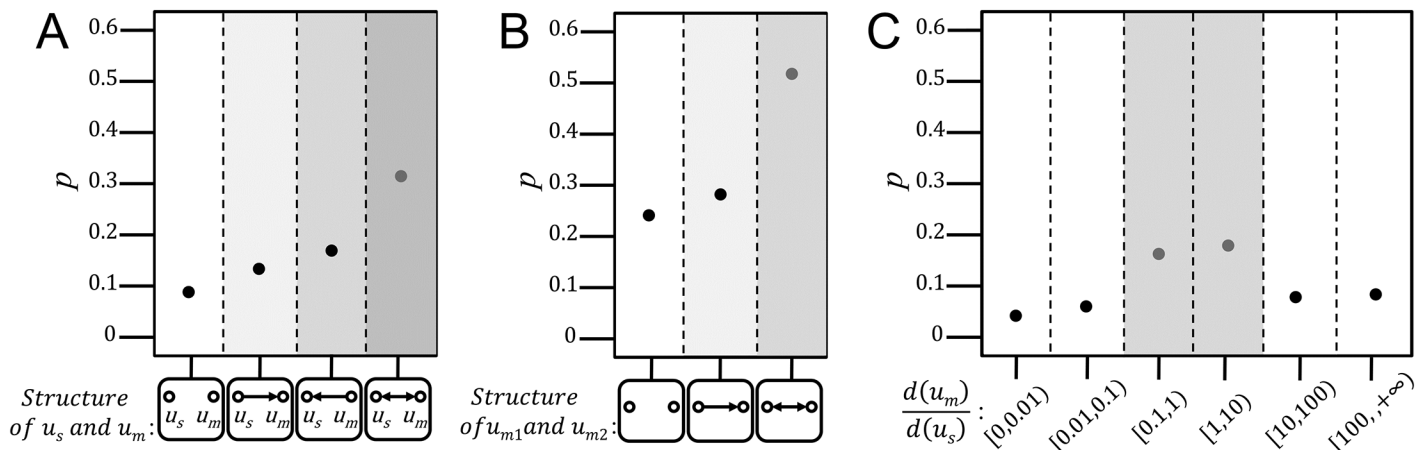
Furthermore, we examine the percentage of response users among all be-mentioned users in a message. For a message with  $k$  mentions, and  $n_r$  is the number of be-mentioned users who give response to the message. We define the *response ratio* of the message as  $n_r/k$ . We use  $r(k)$  to represent the average response ratio over all messages with  $k$  mentions. As shown in Fig 3B, we can observe a peak in response ratio at 2 mentions in a message and then a slow drop. This implies that if a message contains more than two mentions, the be-mentioned users are less likely to give response to it, possibly because a message mentioning a lot of users is likely to be supposed as a spam, which will decrease others' interest in forwarding it. We further investigate the response ratio with respect to different kinds of be-mentioned users, which reveals that users with large number of followers are less likely to respond to a mention probably due to the overload of mentions (S1 Fig).

### Effect of network structure

When going beyond the mention effect on the mention count in a message, we continue to wrestle with response probability with respect to the network structure between users.

We start with the topological structure between source user  $u_s$  and be-mentioned user  $u_m$ . According to their followship on the social network, we have four types of structures: (a) no followship between  $u_s$  and  $u_m$ , (b)  $u_s$  follows  $u_m$ , (c)  $u_m$  follows  $u_s$ , and (d) reciprocal followship between  $u_s$  and  $u_m$ . Note that followship offers a proxy for tie strength while the reciprocal followship represents a strong tie of friendship between users [55]. Fig 4A demonstrates the response probability with respect to the four types of structures. We can observe that the response probability of the reciprocal followship is significantly higher than the others, demonstrating that the stronger tie strength between two users, the more likely for one user to receive response via mentioning the other in a message.

Moreover, we investigate the interdependence between the response probability and the network structure among be-mentioned users. Among all the cases where a message contains multiple mentions, 2-mention case is the most frequent one and the study of response probability for 2-mention case can be easily extended to other cases of multiple mentions. Therefore, in this paper, we only focus on the 2-mention case. For convenience, we denote the two



**Fig 4. Effect on network structure.** (A) Response probability  $p$  versus structure of source user  $u_s$  and be-mentioned user  $u_m$ . (B) Response probability  $p$  versus structure of two be-mentioned users  $u_{m1}$  and  $u_{m2}$ . From (a)(b), we can both observe that the response probability is the highest when a reciprocal following exists between users. (C) Response probability  $p$  versus degree ratio  $\frac{d(u_m)}{d(u_s)}$ , where  $d(u_m)$  represents the number of followers of be-mentioned user  $u_m$  and  $d(u_s)$  is the number of followers of source user  $u_s$ . We find that the response probability of the degree ratio between [0.1, 10] is higher than the others.

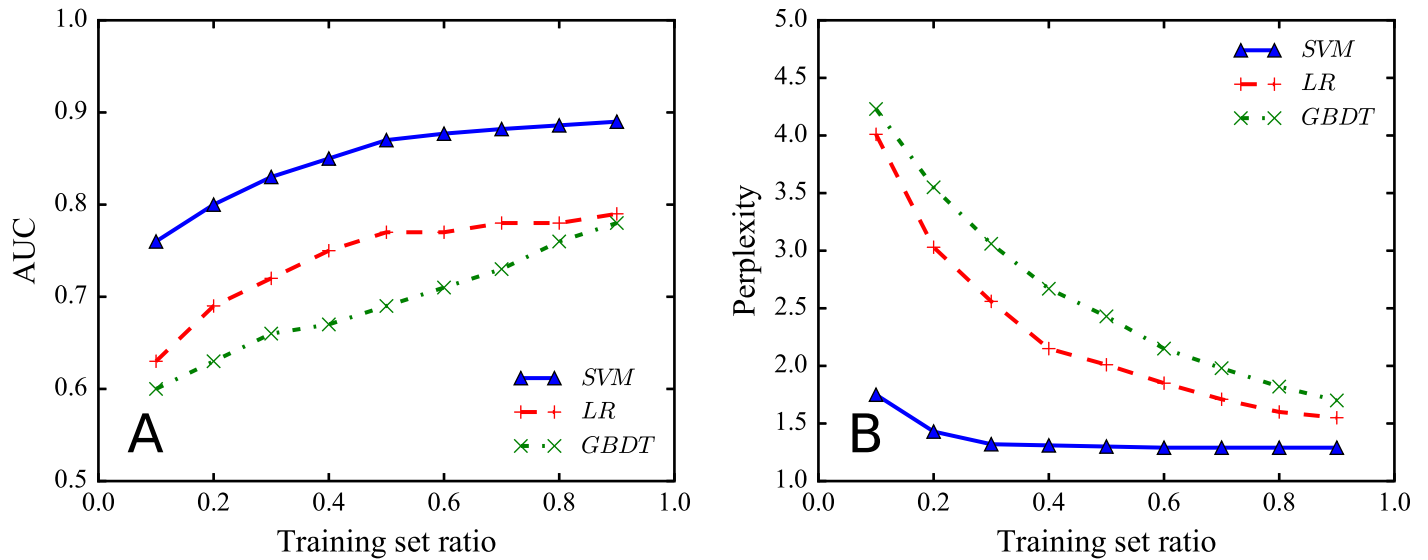
<https://doi.org/10.1371/journal.pone.0194192.g004>

be-mentioned users as  $u_{m1}$  and  $u_{m2}$ . Then, according to the following between  $u_{m1}$  and  $u_{m2}$ , we have three types of structures: (a) no following between  $u_{m1}$  and  $u_{m2}$ , (b)  $u_{m1}$  follows  $u_{m2}$  or  $u_{m2}$  follows  $u_{m1}$ , and (c) reciprocal following between  $u_{m1}$  and  $u_{m2}$ . Intuitively, the more number of followships among users, the more overlapped their interest will be, and therefore the more probable they will interact with each other. From Fig 4B, we find that the response probability of the reciprocal following structure is the highest. This finding implies that mentioning connected users in a message would probably motivate them to participate in the discussion and therefore respond to it.

Finally, we also examine whether the status difference between two users will affect the mention effect. Here we adopt the number of followers as a measure of a user’s status. We define the *degree ratio* as  $\frac{d(u_m)}{d(u_s)}$ , where  $d(u_m)$  represents the number of followers of the be-mentioned user  $u_m$  and  $d(u_s)$  is the number of followers of the source user  $u_s$ . As shown in Fig 4C, the response probability of the degree ratio between [0.1, 10] which could be viewed as similar status between  $u_m$  and  $u_s$ , is higher than the others. This means that one is more likely to receive a target user’s response if they have similar social status. One possible explanation for these findings is that people are living in status groups and they are only supposed to engage with people of like status [56].

### Response prediction

To illustrate the empirical findings, we turn to the question: Can we predict the response of those be-mentioned users? We call this problem as “response prediction” (RP). Formally, given a message  $d$ , the source user  $u_s$  and a mention candidate  $u_m$  pair, we try to predict whether  $u_m$  will give a response to the message. This prediction task can be formulated as a binary classification problem. Firstly, based on the observed interdependence between the response probability and the network structure, we extract two types of factors which would affect the prediction performance: (a) Structure factors, including whether or not  $u_s$  follows  $u_m$  and whether or not  $u_m$  follows  $u_s$ ; (b) Influence factors, including the logarithmic of the number of followers of  $u_s$  and  $u_m$  respectively, the average number of forwardings for each message from  $u_s$  and  $u_m$  respectively. In addition, according to our previous studies [53], we also



**Fig 5. Prediction performance.** (A) AUC of the three algorithms. AUC measures the area under the ROC curves. (B) Perplexity of the three algorithms when predicting response behaviors, against the training set ratio.

<https://doi.org/10.1371/journal.pone.0194192.g005>

consider: (c) Content factors, including whether or not the message contains an embedded URL, whether or not the message is annotated with certain events. Then, we employ three widely used machine learning models for classification task: Support Vector Machine with an RBF kernel (*SVM-RBF*), Linear Regression (*LR*), and Gradient Boosted Decision Trees (*GBDT*).

To evaluate the prediction performance, we adopt two widely used metrics for classification task: *AUC* and *perplexity*. A higher *AUC* and a lower *perplexity* indicate better prediction performance. See Section [Materials and methods](#) for details. [Fig 5A](#) reports *AUC*. We find that the *SVM-RBF* classifier obtains the best performance, raising *AUC* to nearly 90%. We then report perplexity on the testing set with respect to the training set ratio. As shown in [Fig 5B](#), the *SVM-RBF* classifier also achieves the lowest perplexity among all tested classifiers. These results indicate that because of the complexity of the factors that would affect individual’s response behavior, it is more suitable to capture them using a nonlinear feature space mapping. Therefore, we select the *SVM-RBF* classifier to further evaluate the importance of the proposed factors.

Furthermore, in order to analyze how each factor contributes to the prediction, we design a contrast experiment by eliminating one factor at a time and observe how the prediction performance changes. Here we use the *SVM-RBF* classifier for training and testing. We take 50% of all the samples as the training set and the rest 50% as the testing set. As shown in [Table 1](#), we find that when we leave out the content factors (*No\_Content*), the *AUC* suffers from a 4.6% decline and the perplexity suffers from a 4.2% increase. This finding shows us that the message content can affect a be-mentioned user’s response behavior, but the effect is very limited. One possible explanation is that because the length of each message in micro-blogging network is restricted to no larger than 140 characters, it is still a challenge to reveal the semantics from sparse and noise short texts [57]. In comparison, when we take out the influence factors (*No\_Influence*) from our model, the *AUC* decreases 9.1% and the perplexity increases 8.5%. This result indicates that the interpersonal influence plays a more important role than message content in information diffusion, which is consistent with empirical findings in previous works [13, 14, 19]. More importantly, when we eliminate the structure factors (*No\_Structure*),

**Table 1. Comparison on how different factors affect the performance.**

	ALL	No_Content	No_Influence	No_Structure
AUC	0.87	0.83	0.79	0.61
Perplexity	1.30	1.35	1.41	1.64

ALL: all factors are combined. No\_Content: content factors are eliminated. No\_Influence: influence factors are eliminated. No\_Structure: structure factors are eliminated. We can observe that although message content and user influence help to improve the response prediction result, structure-dependent factors play a much more significant role in the prediction.

<https://doi.org/10.1371/journal.pone.0194192.t001>

we observe a 29.9% decrease of the AUC and a 26.2% increase of the perplexity. This result shows that although message content and social influence help to improve the response prediction result, the structure factors play a much more significant role in the prediction. This result is consistent with our empirical findings about the mention effect of network structure in previous section. The network structure among users, such as structural diversity or structural hole, plays a key role in predicting the individual's behavior that underlies the social contagion processes [34, 58].

## Discussion

In this paper, the key feature of mention in micro-blogging platform has been investigated comprehensively. We conduct our study on a population-scale dataset from the most popular Chinese micro-blogging network, namely Sina Weibo. We study the statistical characteristics of mention in information diffusion represented as diffusion network. In fact, a significant proportion of these cascades contains mentions in content. We find that users with large number of followers on social network would receive much more mentions than those small-degree users, and meanwhile it brings the problem of mention overload. It will not only interrupt user's daily use of micro-blogs, but also result in frustration and decrease user's interest in forwarding. These findings provide us insight and guidance in proposing a new recommendation scheme to maximize the spread of influence.

To further investigate the effect of mention in information diffusion, we examine the response probability from the perspective of mention count and network structure respectively. We observe a saturation at around 5 mentions in a message which means that with each additional mention, a message is more and more likely to receive response from the mentioned users, up to a point. Then we study the response ratio among all mentions in a message and observe a peak at 2 mentions. Beyond the mention effectiveness on the mention count in a message, we further examine the response probability with respect to the network structure between users. We find that the response probability is the highest when a reciprocal followship exists between users. Furthermore, one is more likely to receive a target user's response if they have similar social status. From the perspective of machine learning, the discovered correspondence provides predictive factors to estimate response probability. To illustrate these findings, we propose the response prediction task and formulate it as a binary classification problem. By adopting features including message content, user influence and topological structure between users, a machine learned prediction function is trained. Extensive evaluation demonstrates the effectiveness of discovered factors.

To understand the variation of response probability for different messages, we also classify messages into different categories according to the content and compare the response probability curves of each category. Due to strict length restriction of a message, many different viewpoints and additional context can be expressed through embedded URL sharing and annotated event keywords, which represent important features of the content of messages [59,



60]. As done in our previous studies [53], we classify messages according to their content, i.e., whether containing embedded URLs or event keywords. The comparison of response probability curves indicate that be-mentioned users are not prone to respond messages containing embedded URLs or events (S2 Fig). This could be explained from the perspective of psychology that people in China are sometimes conservative and protective on online social network while facing social events.

As future work, we will devote to deep investigation on the interdependence between the popularity of a message and the structural characteristics of multiple mentioned users. We will further study whether there are some kinds of significant patterns existing in the information diffusion process. Moreover, it is of great interest to model the individual behaviors from the micro-perspective and uncover the information cascading process with behavioral dynamics.

## Materials and methods

### Data

The data used in this paper are collected from Sina Weibo, which is the most popular micro-blogging platform in China. It includes basic information about messages (time, user ID, message ID etc.), mentions (user IDs appearing in messages), forwarding paths, and whether containing embedded URLs or event keywords. In addition, it also contains a snapshot of the following network of users (based on user IDs). This data is also used in our previous studies [53]. It is now available from the WISE 2012 Challenge (<http://www.wise2012.cs.ucy.ac.cy/challenge.html>). The results we present are produced using messages that were originally posted between July 1, 2011 and July 31, 2011. We cleaned the data by removing inactive users and unpopular messages. We also removed spam users who abnormally forward a single message for hundreds of times. To alleviate the effect from activity pattern of users, we only consider the messages posted between 10am and 10pm per day, which is the active period in Sina Weibo system. In total, there are 2.6 million messages. And for each message, its forwarding information between July 1, 2011 and August 31, 2011 is also collected. Detailed statistics of the dataset is reported in Table 2.

### Comparison algorithms and evaluation metrics

We denote with a tuple  $(u_s, u_m, d)$  the sample that a user  $u_s$  (called the source user) mentions another user  $u_m$  in a message  $d$ . Each time  $u_m$  sees the message that he has not forwarded before, we say  $\delta_{s,m,d} = 1$  if  $u_m$  forwarded  $d$ , forming a positive example indicating  $u_s$  successfully activates  $u_m$  to give a response to  $d$ ; otherwise  $\delta_{s,m,d} = 0$  for a negative example if  $u_m$  neglects  $d$ .

To compare the performance of response prediction, three mainstream classification algorithms are implemented to estimate and predict response probabilities on all samples, including Support Vector Machine with an RBF kernel, Linear Regression, and Gradient Boosted

Table 2. Data statistics.

Statistics	Value
Num of users	43,378,576
Num of followships	198,347,101
Num of cascades	2,636,198
Num of cascades with '@'	393,772
Num of positive (response) examples	119,845
Num of negative (no response) examples	273,925

<https://doi.org/10.1371/journal.pone.0194192.t002>

Decision Trees. Some other widely used models are not compared because those models require exogenous such as message content or user profiles that are absent in this scenario.

In this paper, we use *AUC* and *perplexity* as evaluation metrics. *AUC* measures the area under the Receiver Operating Characteristic curve, which represents the probability that a model correctly distinguishes a randomly selected positive sample from a randomly selected negative sample. The perplexity measures how the testing samples surprise a trained model, as shown in Eq (1). A higher *AUC* and a lower perplexity indicate better prediction performance. The definition of perplexity is as follows:

$$perplexity = e^{-\frac{\sum_{(s,m,d) \in D_{test}} \delta_{s,m,d} \ln \hat{P}(\delta_{s,m,d}=1) + (1-\delta_{s,m,d}) \ln (1-\hat{P}(\delta_{s,m,d}=1))}{|D_{test}|}}, \quad (1)$$

where  $D_{test}$  represents the testing set, and  $\hat{P}(\delta_{s,m,d} = 1)$  is the estimated response probability.

## Supporting information

**S1 Fig. Response ratio  $r(k)$  versus the average degree of be-mentioned users  $\langle d(u_m) \rangle$ .** We classify all messages into two categories:  $\langle d(u_m) \rangle \in [1, 100)$ , and  $\langle d(u_m) \rangle \in [10000, +\infty)$ . We observe a peak in response ratio at 2 mentions and then a slow drop in both categories. Moreover, we find that the response ratio of  $\langle d(u_m) \rangle \in [1, 100)$  is higher than that of  $\langle d(u_m) \rangle \in [10000, +\infty)$ .

(EPS)

**S2 Fig. The variation of response probability for different kinds of messages.** (A) response probability  $p(k)$  versus mention count  $k$  for messages with and without embedded Events. (B) response probability  $p(k)$  versus mention count  $k$  for messages with and without embedded URLs.

(EPS)

## Author Contributions

**Conceptualization:** Peng Bao, Hua-Wei Shen, Junming Huang, Haiqiang Chen.

**Data curation:** Peng Bao.

**Formal analysis:** Peng Bao.

**Funding acquisition:** Peng Bao.

**Investigation:** Peng Bao, Hua-Wei Shen.

**Methodology:** Peng Bao, Hua-Wei Shen, Junming Huang, Haiqiang Chen.

**Project administration:** Peng Bao.

**Resources:** Peng Bao.

**Writing – original draft:** Peng Bao.

**Writing – review & editing:** Peng Bao, Hua-Wei Shen, Junming Huang, Haiqiang Chen.

## References

1. Salganik M, Dodds P, Watts D. Experimental study of inequality and unpredictability in an artificial cultural market. *Science* 311: 854–856 (2006). <https://doi.org/10.1126/science.1121066> PMID: 16469928
2. Watts D, Dodds PS. Influentials, networks, and public opinion formation. *J. Consum. Res.* 34: 441–458 (2007). <https://doi.org/10.1086/518527>

3. Lazer D, Pentland A, Adamic L, Aral S, Barabási AL, Brewer D, et al. Computation social science. *Science* 323: 721–723 (2009).
4. Muchnik L, Aral S, Taylor SJ. Social influence bias: a randomized experiment. *Science* 341: 647–651 (2013). <https://doi.org/10.1126/science.1240466> PMID: 23929980
5. Szabo G, Huberman BA. Predicting the popularity of online content. *Commun. ACM* 53: 80–88 (2010).
6. Khosla A, Sarma AD, Hamid R. What makes an image popular? *Proc. WWW '14*: 867–876 (2014).
7. Pinto H, Almeida JM, Goncalves MA. Using early view patterns to predict the popularity of YouTube videos. *Proc. WSDM '13*: 365–374 (2013).
8. Pastor-Satorras R, Vespignani A. Epidemic spreading in scale-free networks. *Phys. Rev. Lett.* 86: 3200–3203 (2001). <https://doi.org/10.1103/PhysRevLett.86.3200> PMID: 11290142
9. Romero DM, Meeder B, Kleinberg J. Differences in the mechanics of information diffusion across topics: idioms, political hashtags, and complex contagion on twitter. *Proc. WWW '11*: 695–704 (2011).
10. Hong L, Davison B. Empirical study of topic modeling in twitter. *Proc. SIGKDD '10 on SMA*: 80–88 (2010).
11. Michelson M, Macskassy S. Discovering users' topics of interest on twitter: a first look. *Proc. CIKM '10*: 73–80 (2010).
12. Wu S, Hofman J, Mason W, Watts D. Who says what to whom on twitter. *Proc. WWW '11*: 705–714 (2011).
13. Ye S, Wu S. Measuring message propagation and social influence on twitter. *Social Informatics*: 216–231 (2010). [https://doi.org/10.1007/978-3-642-16567-2\\_16](https://doi.org/10.1007/978-3-642-16567-2_16)
14. Bakshy E, Hofman J, Mason W, Watts D. Everyone's an influencer: quantifying influence on twitter. *Proc. WSDM '11*: 65–74 (2011).
15. Cha M, Haddadi H, Benevenuto F, Gummadi K. Measuring user influence in twitter: The million follower fallacy. *Proc. ICWSM '11*: 10–17 (2011).
16. Pazzani M, Billsus D. Learning and revising user profiles: The identification of interesting web sites. *Machine learning* 27(3):313–331 (1997). <https://doi.org/10.1023/A:1007369909943>
17. Guy I, Zwerdling N, Carmel D, Ronen I, Uziel E, Yogev S, et al. Personalized recommendation of social software items based on social relations. *Proc. RecSys '09*: 53–60 (2009).
18. Xu B, Bu J, Chen C, Cai D. An exploration of improving collaborative recommender systems via user-item subgroups. *Proc. WWW '12*: 21–30 (2012).
19. Huang J, Cheng XQ, Shen HW, Zhou T, Jin X. Exploring social influence via posterior effect of word-of-mouth recommendations. *Proc. WSDM '12*: 573–582 (2012).
20. Barabási AL. The origin of bursts and heavy tails in human dynamics. *Nature* 435: 207–211 (2005). <https://doi.org/10.1038/nature03459> PMID: 15889093
21. Crane R, Sornette D. Robust dynamic classes revealed by measuring the response function of a social system. *Proc. Natl. Acad. Sci.* 105(41): 15649–15653 (2008). <https://doi.org/10.1073/pnas.0803685105> PMID: 18824681
22. Gleeson JP, Cellai D, Onnela J, Porter MA, Reed-Tsochas F. A simple generative model of collective online behavior. *Proc. Natl. Acad. Sci.* 111(29): 10411–10415 (2014). <https://doi.org/10.1073/pnas.1313895111> PMID: 25002470
23. Bao P, Shen HW, Huang J, Cheng XQ. Popularity prediction in microblogging network: a case study on sina weibo. *Proc. WWW '13*: 177–178 (2013).
24. Cheng J, Adamic L, Dow A, Kleinberg J, Leskovec J. Can cascades be predicted? *Proc. WWW '14*: 925–936 (2014).
25. Bao P. Modeling and predicting popularity dynamics via an influence-based self-excited Hawkes process. *Proc. CIKM '16*: 1897–1900 (2016).
26. Bao P, Zhang X. Uncovering and predicting the dynamic process of collective attention with survival theory. *Scientific Reports* 7: 2621 (2017). <https://doi.org/10.1038/s41598-017-02826-6> PMID: 28572618
27. Kempe D, Kleinberg J, Tardos E. Maximizing the spread of influence through a social network. *Proc. SIGKDD '03*: 137–146 (2003).
28. Leskovec J, Adamic L, Huberman BA. The dynamics of viral marketing. *ACM Trans. Web* 1:5 (2007).
29. Onnela JP, Saramäki J, Hyvönen J, Szabó G, Lazer D, Kaski K, et al. Structure and tie strengths in mobile communication networks. *Proc. Natl. Acad. Sci.* 104(18): 7332–7336 (2007). <https://doi.org/10.1073/pnas.0610245104> PMID: 17456605
30. Barabási AL. The network takeover. *Nat. Phys.* 8: 14–16 (2012). <https://doi.org/10.1038/nphys2188>

31. Kwak H, Lee C, Park H, Moon S. What is twitter, a social network or a news media? Proc. WWW '10: 591–600 (2010).
32. Gomez-Rodriguez M, Leskovec J, Schölkopf B. Modeling information propagation with survival theory. Proc. ICML '13: 666–674 (2013).
33. Lü L, Chen DB, Zhou T. The small world yields the most effective information spreading. New J. Phys. 13(12): 123005 (2011). <https://doi.org/10.1088/1367-2630/13/12/123005>
34. Ugander J, Backstrom L, Marlow C, Kleinberg J. Structural diversity in social contagion. Proc. Natl. Acad. Sci. 109: 5962–5966 (2012). <https://doi.org/10.1073/pnas.1116502109> PMID: 22474360
35. Delvenne J, Lambiotte R, Rocha LEC. Diffusion on networked systems is a question of time or structure. Nat. Commun. 6: 7366 (2015). <https://doi.org/10.1038/ncomms8366> PMID: 26054307
36. Song C, Qu Z, Blumm N, Barabási AL. Limits of predictability in human mobility. Science 327: 1018–1021 (2010). <https://doi.org/10.1126/science.1177170> PMID: 20167789
37. Wang C, Huberman BA. How random are online social interaction? Scientific Reports 2: 633 (2012). <https://doi.org/10.1038/srep00633> PMID: 22953054
38. Shen HW, Wang D, Song C, Barabási AL. Modeling and predicting popularity dynamics via reinforced poisson processes. Proc. AAAI '14: 291–297 (2014).
39. Zhao Q, Erdogdu MA, He HY, Rajaraman A, Leskovec J. SEISMIC: a self-exciting point process model for predicting tweet popularity. Proc. SIGKDD '15: 1513–1522 (2015).
40. Hopcroft J, Lou T, Tang J. Who will follow you back? reciprocal relationship prediction. Proc. CIKM '11: 1137–1146 (2011).
41. Ratkiewicz J, Fortunato S, Flammini A, Menczer F, Vespignani A. Characterizing and modeling the dynamics of online popularity. Phys. Rev. Lett. 105: 15870 (2010). <https://doi.org/10.1103/PhysRevLett.105.15870>
42. Yang J, Counts S. Predicting the speed, scale, and range of information diffusion in Twitter. Proc. ICWSM '10: 355–358 (2010).
43. Wang B, Wang C, Bu J, Chen C, Zhang W, Cai D. Whom to mention: expand the diffusion of tweets by recommendation on micro-blogging systems. Proc. WWW '13: 1331–1340 (2013).
44. Tang L, Ni Z, Xiong H, Zhu H. Locating targets through mention in Twitter. World Wide Web 18(4): 1019–1049 (2015). <https://doi.org/10.1007/s11280-014-0299-8>
45. Zhou G, Yu L, Zhang CX, Liu C, Zhang ZK, Zhang J. A novel approach for generating personalized mention list on micro-blogging system. Proc. ICDMW '15: 1368–1374 (2015).
46. Li Q, Song D, Liao L, Liu L. Personalized mention probabilistic ranking—recommendation on mention behavior of heterogeneous social network. Proc. WAIM '15: 41–52 (2015).
47. Jiang B, Sha Y, Wang L. Predicting user mention behavior in social networks. Proc. NLPCC '15: 146–158 (2015).
48. Ding Z, Zou X, Li Y, He S, Cheng J, Qiao F, *et al.* Mentioning the optimal users in the appropriate time on Twitter. Proc. APWeb '16: 464–468 (2016).
49. Gong Y, Zhang Q, Sun X, Huang X. Who will you “@”? Proc. CIKM '15: 533–542 (2015).
50. Pramanik S, Wang Q, Danisch M, Bandi S, Kumar A, Guillaume J, *et al.* On the role of mentions on tweet virality. Proc. DSAA '16: 204–213 (2016).
51. Li Y, Ding Z, Zhang X, Liu B, Zhang W. Confirmatory analysis on influencing factors when mention users in Twitter. Proc. APWeb '16: 112–121 (2016).
52. Huang H, Zhang Q, Huang X. Mention recommendation for Twitter with end-to-end memory network. Proc. IJCAI '17: 1872–1878 (2017).
53. Bao P, Shen HW, Chen W, Cheng XQ. Cumulative effect in information diffusion: empirical study on a microblogging network. PLoS ONE 8(10): e76027 (2013). <https://doi.org/10.1371/journal.pone.0076027> PMID: 24098422
54. Katz E. The two-step flow of communication: an up-to-date report on a hypothesis. Public Opin. Quart. 21: 61–78 (1957). <https://doi.org/10.1086/266687>
55. Petersen AM. Quantifying the impact of weak, strong, and super ties in scientific careers. Proc. Natl. Acad. Sci. 112(34): 4671–4680 (2015). <https://doi.org/10.1073/pnas.1501444112>
56. Weber M. Weber's rationalism and modern society. Waters, T. & Waters, D.: Macmillan (2015).
57. Yan X, Guo J, Lan Y, Cheng XQ. A bitern topic model for short texts. Proc. WWW '13: 1445–1456 (2013).
58. Burt RS. Structural holes: the social structure of competition. Harvard University Press (1992).

59. Cao C, Caverlee J, Lee K, Ge H, Chung J. Organic or organized? exploring URL sharing behavior. *Proc. CIKM '15*: 513–522 (2015).
60. Cheng T, Wicks T. Event detection using Twitter: a spatio-temporal approach. *PLoS ONE* 9(6): e97807 (2014). <https://doi.org/10.1371/journal.pone.0097807> PMID: 24893168