

Factors Affecting Retweetability: An Event-Centric Analysis on Twitter

Research-in-Progress

Nargis Pervin

National University of Singapore
Singapore
nargisp@comp.nus.edu.sg

Hideaki Takeda

National Institute of Informatics
Tokyo, Japan
takeda@nii.ac.jp

Fujio Toriumi

The University of Tokyo
Tokyo, Japan
tori@sys.t.u-tokyo.ac.jp

Abstract

In Twitter information primarily propagates through retweet mechanism. While a massive amount of tweets gets generated everyday, only a handful of them get retweeted widely. In this study, we have investigated the impact of user-roles in retweet phenomena. We have introduced the concept of “Information Diffusion Impact” (IDI) and identified three important user roles, namely “information starter”, “amplifier”, and “transmitter”. Retweetability has been modeled using IDI impact for different user roles along with the content features like presence of hashtag, URL etc. Further, the effect of a major event on the factors affecting retweetability has been investigated. Our findings demonstrate that retweetability is significantly affected by amplifiers and information-starters and these effects change substantially due to event. We have also reexamined our model in another dataset of the Boston marathon bomb blast, 2013 and the outcome of this analysis is in good agreement with our findings from Japan earthquake dataset.

Keywords: Microblogs, Retweetability, Crisis

Introduction

Twitter has turned out to be the most popular microblogging service by far which can disseminate up-to-the-minute information rapidly. It endows users to share information in real time beyond geographic constraint and has gained uprising attention for political campaigning (Abel et al., 2011), news media, crime information (Chu et al., 2010), and disaster communication (Hughes and Palen, 2009; Mendoza et al., 2010). Research has been conducted in the line of diffusion of information on Twitter, particularly, in the context of adoption cascades (Gruhl et al., 2004) and trending topic detection (Pervin et al., 2013). However, little attention has been paid on how information diffuses and who participates in the diffusion process on Twitter, which demands further investigation. This is very significant, particularly in product advertising and campaigning through social media where brands or companies seek attention from large audiences very rapidly. This demands recognition of the potential and influential target audiences on the Twitter network, who in turn can promote the product by tweeting/retweeting product related information to his friends and followers. Therefore, it is very important to identify the communicators in the diffusion process and investigate their roles in diffusion mechanism.

The principal factor of information diffusion on Twitter, the so-called act of retweeting, allows users to broadcast someone else's tweet to their own set of friends and followers. In fact the users can use the official retweet button to share the content in one click. Though the practice of retweeting does not follow standard rules, the most common practice of giving attribution to the user is adding *RT @* before the Twitter handle of the user. The construction and analysis of retweet network is not a straightforward task. Due to the limited 140 characters in a tweet, users frequently tend to delete or modify the tweet content to meet the character limit and this adds complications in the construction and analysis of retweet network.

Recently, a surge of interest has been observed to unfold the factors impacting the retweetability. Boyd et al. (2010) have stated retweeting as a practice of participating in a conversation and studied the conventions and diverse reasons people retweet. On Twitter, information flows in a large network through the cascades of followers. To explode the social shares the tweet needs to reach out the correct users timely and should attract them by its content. Suh et al. (2010) have shown that inclusion of hashtags and URLs in the tweet content increases its share count. While content features are important, retweetability mainly depends on who is seeing your tweet and eventually participating in the process. For instance, in the time of campaigning for new product launch the companies try to reach out the journalists and the celebrities to acquire involvement of more audiences in it.

While unstandardized retweet practices not only make the construction of retweet network non-trivial, consideration of only the tweet content to build retweet network also adds bias in the analysis. In this paper we have focused to study the factors impacting retweetability considering both network variables and content variables of tweets. We present a systematic way to build the retweet network and then discuss the factors impacting retweetability. The users in the retweet chain have been classified into three distinct classes according to their roles in information propagation and finally incorporated all these factors into the model.

In addition, we have checked whether an external event can change the probability of retweets. Our dataset (2011 Great Eastern Japan earthquake Twitter data) revolves around a major event and hence, allows us specifically to address this research question. Next, we have used a Twitter dataset of the Boston marathon bomb blast (2013) to check whether the results obtained in both the events follow a similar pattern. For modeling the factors that affect retweetability, we have used regression technique. Furthermore, to check the effect of the event on retweetability, difference in difference estimator (DID) have been employed using three time windows centering the event in the dataset. The results obtained from both the datasets indicate that user's roles in information diffusion differ in the time of event as compared to the pre-event time window. Users with comparatively less number of followers, i.e., not so famous in Twitter, participate in the information diffusion process at the time of the event and play significant role in information diffusion in the time of crisis.

The contributions of the paper are as follows:

- i. We define and classify user-roles in information diffusion directly grounded on the impact that the users have on the network. The classified user roles determine how they participate in the diffusion process and proposed new metrics for user's impact score on network.

- ii. We analyze the retweet network along with the follower network to understand the factors impacting retweetability. Herein, we check whether the classified user roles have significant impact on retweetability along with other factors.
- iii. We investigate the effect of a major external event (e.g., earthquake) on these factors.

Literature Review

While a massive amount of information is available on Twitter, 40% of them are white noise (Chu et al., 2010) and according to current findings it proves to be even larger (DishDaily, 2014). In practice only a small percentage of tweets get retweeted. What are the reasons for a tweet to get retweeted? What kind of contents get shared by the people (Boyd et al., 2010)? Researchers have investigated that bad news travels faster on Twitter (Naveed et al., 2010). In an early work Kwak et al. (2010) have done a quantitative study of information diffusion on Twitter and investigated the relation between the authors' in-degree and their reachability in the network. They argue that users having less than 1000 followers tend to have on average same number of additional recipients of the tweet. With the increase in the number of followers the mean of number of additional recipients increase. This suggests the clear correlation of in-degree of tweet author and the number of users reached on the network.

Several researches have been carried out to investigate the retweetability (Suh et al., 2010; F. Samantha and M. Eni., 2013; Wu et al. 2011; Bastos et al., 2012). Suh et al. (2010) have examined a set of features that can predict the retweetability of a tweet. Applying Generalized Linear Model (GLM), they show that contextual features like hashtags, URLs or mentions affect the probability of a tweet getting retweeted. They also showed that if the original poster of a tweet has many followers and followees, the probability increases. Yang et al. (2010), attempt to predict the information propagation considering properties like historical mentions of users using survival analysis modeling.

While the inclusion of features like URL, hashtag, mentions or question marks in the tweet steers more attention, we claim that features like number of new people user makes aware of (not necessarily the number of followers), the position of the user in the retweet chain, and time of retweeting should also be considered. Particularly, the effect of a major external event like an earthquake on the retweetability has previously not been investigated. In this study, we have first classified the users based on their diffusion impact on the network and generated user score for each role. Next, we have examined the impact of these roles on retweetability of a tweet. Moreover, using two event-centric datasets, we investigate the effect of an event on various factors affecting retweetability.

Dataset Description

In this study, we have used two datasets from two separate emergency events - 2011 Japan earthquake and 2013 Boston Marathon bomb-blast. Both the datasets are described in turns.

2011 Great Eastern Japan Earthquake Dataset

Tweet Data: We used a Twitter dataset collected during the earthquake in 2011 described thoroughly in Toriumi et al. (2013). Data collection procedure has been discussed briefly here:

- First, a set of tweets has been collected from Twitter streaming API (Application Programming Interface) for tweets and the data was collected during the event.
- Next, for all these tweets the user details have been crawled using the same API along with the follower IDS.
- For all these users the tweets have been collected for 20 days.

The dataset covers a period of 20 days (from 5th March, 2011 to 24th March, 2011), and consists of 362,435,649 tweets posted by 2,711,473 users in Japan. This dataset is remarkable by its completeness: 80% to 90% of all published tweets of these users were present in this dataset. It should be noted that the dataset consists of tweets of Japanese Twitter users. Hence, a major proportion of tweets (98%) in the dataset are written in Japanese.

Follower Network Data: In Twitter, follower network depicts the social relationships among the users. Follower information has been collected by crawling Twitter API for the active users who have been mentioned more than 20 times in 20 days. Follower network dataset consists of 300,104 users and 73,446,260 relationships information.

2013 Boston Marathon Bomb-blast Dataset

Tweet Data: We have collected a month's Twitter data of Boston-marathon bomb-blast in 2013 in a similar fashion from 1st April, 2013 to 30th April, 2013. In this dataset there are 112,93,215 tweets posted by 30,000 users.

Follower Network Data: Investigating the tweet content we found the users who participated in tweeting and retweeting in the time of crisis. For 30,000 users follower information is collected using Twitter API. Follower network consists of 30,000 users with 73475897 relationships.

Solution Details

In this work, we have investigated the activity network and the static follower network of the Twitter users simultaneously. In Twitter, the retweet functionality allows users to share information with their friends and followers, generating a network of retweeters. Here, this retweet network is considered as activity network. In our case, to analyze the information diffusion for each tweet, we were interested in the retweet sequence of each tweet, which we name as retweet chain.

How to find Retweet Chain

In the most recent work (Tinati et al., 2013), the diffusion of information is directly extracted from the content of the tweets. For instance, if a tweet published by user u_i is composed of the following pattern: *RT @ u_o tweet*, then one considers that the information diffused directly from u_o to u_i . Retweet chains are identified by tweets containing several references, i.e., *RT @ u_1 RT @ u_o tweet*, or consecutive citations, such as u_1 posting the retweet: *RT @ u_o tweet* followed by u_2 posting *RT @ u_1 tweet*.

However, in practice, after one step of citation, there are two important biases:

- users tend to keep only the original author of the tweet, and not intermediate, in particular to meet the 140 character limit of Twitter. Even when using the official retweet function of Twitter, only the initial poster is kept. This will strongly increase the number of direct retweet and in turn the apparent role of the original poster in the diffusion of information.
- it has been shown in Toriumi et al., (2013) that users frequently retweet after seeing a tweet several times. As a result the user cited as source might not be fully representative of the information diffusion.

In this work, to characterize the diffusion of information, we will therefore adopt a combination of both the follower network and the retweeter information from tweets. A retweet chain is simply defined as the sequence of all tweets published containing the original content, ordered by their publication time. To consider the information flow, we combine this information with the fact that, each time a user publishes a tweet, all his followers can see the information. We can therefore know by whom the user might have been informed, independently of the user who appears as a source in the tweet itself. To build retweet chain of an original tweet, we searched in the dataset for their retweets and further follower information is used to mitigate the bias.

User Classification

We classify user roles in the light of information propagation through retweeting and introduced the concept of *Information Diffusion Impact (IDI)*, namely for a user u_i , the number of users he made aware of an information i . Therefore, making 10 people aware of one information and making one person aware of 10 different pieces of information result in the same *IDI* value. By analyzing the retweet chains the users are classified into three categories, “*information starter*”, “*amplifier*”, and “*transmitter*”.

- *Information-starters* are the users who are able to launch new information which will spread broadly in the network. They are the users whose information will reach many.
- *Amplifiers* are users who do not publish interesting content by themselves, but who have the potential to diffuse information published by others to many new people.
- *Transmitters* are users who act as bridges between several communities in the network. If an *information-starter* publishes an interesting tweet in a given community of the network, *amplifiers* will spread this tweet in the same community, but *transmitters* are necessary to reach other communities, which in turn will result in transmission of the information more broadly.

Information Starter: *Information starter* can be conceptualized as the one who creates the original information. Information starters are important as their information is retweeted by others and depending on the importance of the content it is diffused further in the network. We compute the information starter score as the total number of people made aware of the tweet by all the people who participated in the retweet process.

Amplifier: *Amplifiers* are considered as the individuals who share others' information and make many people aware of it. They are important as they are followed by many users and as a result, *amplifier* makes a large fraction of users aware of the information. To compute the power of *amplifier*, unlike *information starter*, here we calculate the direct impact of the user in the network.

We should note that this value is usually less than the number of followers of u , as some of his followers have already been made aware of the tweet. Therefore, users who appear early in the retweet chain, or who tend to inform users following few people will naturally have higher amplifier scores.

Transmitter: It is now accepted that the most social networks have a strong community structure (Girvan and Newman, 2010). The Twitter follower network is no exception, and its analysis reveals clearly defined modules. In this paper, we used the Fast OSLOM algorithm (Lancichinetti, 2011) to detect overlapping communities in our follower network. The algorithm found 8 communities in our follower network with an average size of 44668 nodes per community. By manual investigation, we found obvious meanings for some communities, such as a community of foreigners and a community of users related to nightlife (disc jockeys, hip-hop celebrities, etc.). After classifying the users into different roles we check their impacts on retweetability.

Table 1. Factors Affecting Retweetability

	Variable	Meaning
Dependent Variable	RetweetCount	Number of rewteets by a user per unit time
Network variable	NumFollowers	Number of followers of the user/retweeter
	AmplifierScore	How many new people he can make aware of
	InformationStarterScore	How many people made aware for the tweet he is the author
	TransmitterScore	Number of people made aware in different communities
	PeopleAware(t-1)	Number of people aware in previous time window in the retweet chain (at time t-1)
User specific variable	TweetCount	Total count of tweets by a user
	AveragePosition	Position of the user in the retweet chain, taking all tweets by a user we compute the average position of the user in the retweet chain to indicate early/late retweeters
Content variable	Hashtag	An indicator variable to specify whether a tweet has hashtag
	URL	An indicator variable to specify whether a tweet has URL

Control Variable	TimeOfDay	24 hours have been divided into 5 time-window, morning(7am-10am), noon (11am-3pm), afternoon (4pm-7pm), evening (8pm-11pm), night (12am -6am)
	DayOfWeek	Day of the week is coded as a dummy variable
	Age	Time since the tweet is composed

Data Preparation and Analysis

Data Preparation

Using Twitter dataset, we randomly selected 10,000 widely retweeted tweets. For all these 10,000 tweets, retweet chains have been formed, which are basically the chain of users in chronological order of their retweets of the original posts. Here, we want to add that by random selection of the tweets the network structure is not broken as we are considering retweet chain for each tweet separately. Our tweet dataset (5th March-24th March) have been divided into three time windows, pre-earthquake (5th March-10th March), during-earthquake (11th March-16th March), and post-earthquake (17th March-24th March). For each tweet, we first build the retweet network, i.e., we find all the users who have retweeted it along with the timestamp (the calendar date and time) of the retweet action.

Data Analysis

Using the follower network information the number of followers, and number of new people user make aware of in a retweet chain have been computed. The number of new people he makes aware of determines his own contribution in the retweet process. Notably, the number of followers of a user and the number of new people he makes aware of are not the same because there will be overlaps among the followers of the retweeters. Thus, the user's action of retweeting will contribute to the awareness of the tweet depending on the position of the user in the retweet chain.

Users who are participating in the retweet process also play very crucial role. The one who starts the tweet (or “*information starter*” as defined earlier) does not necessarily have many followers. But if the tweet gets noticed by a highly influential user it will be retweeted by many. Besides network structure and the users' participation in the retweet actions, tweet content also needs to be considered to understand retweetability. Usage of hashtag is very common and it allows the user to follow or search in Twitter the related information regarding the topic of the hashtag. Previous researchers (Suh et al., 2010, Yang et al. 2010) have shown that inclusion of URLs and hashtags increases the chance of retweetability. In our dataset among the retweeted tweets 26.5% of the tweets have URL and 10.3% of the tweets contain hashtags. We have revisited impact of URL or hashtag in retweetability in our model. To model the factors affecting the retweetability of the tweet, we have considered the variables described in Table 1. Regression technique has been used to model retweet frequency of a tweet and hence the dependent variable considered is computed as the number of retweets a tweet gets per minute. Retweet model is given below

$$\begin{aligned}
 RetweetCount_{it} &= \alpha + \beta_1 PeopleAware_{i,t-1} + \beta_2 NumFollowers_i(u) + \beta_3 AmplifierScore_i(u) \\
 &+ \beta_4 InformationStarterScore_i(u) + \beta_5 TransmitterScore_i(u) + \beta_6 AveragePosition_i(u) \\
 &+ \beta_7 isHashTag_i + \beta_8 isURL_i + \beta_9 isURL_i \times isHashTag_i + \beta_{10} Age_{it} + \beta_{11} TweetCount_t(u) \\
 &+ \beta_{12} DayOfWeek_i + \beta_{13} TimeOfDay_i + \varepsilon
 \end{aligned}$$

For the three distinct time-windows in Japan earthquake dataset, the model has been tested using panel data regression model and the results are shown in Table 2. Number of people the users make aware in the previous time-units (here previous minutes), i.e., $PeopleAware(t - 1)$ does not have any significant impact on retweet frequency in the pre-event time window. However, in during-event and post-event time windows this impact becomes positive. In Twitter, the same tweets get retweeted from several sources (followers), and people might retweet it after seeing it from more than one source. In normal situation (when there is no event), people may not retweet it immediately. However, in the time of emergency if more users see the tweet (more people are aware of the event) it increases the retweetability, whereas in normal situation this effect is much more complex. Interestingly, users with high number of followers

have a positive impact on the retweet frequency for both pre- and post-event window, but during the event the effect is opposite. This indicates that more users with low in-degree (number of followers) participated in the retweet process.

Also, the impact of *amplifier* score and *information-starter* score on retweet frequency is very high during the event. This suggests that during the event the highly retweeted tweets were retweeted mostly by the low in-degree users. However, by the definition of our *amplifier* and *information-starter* score, the users having low in-degree can have high *amplifier* scores or *information-starter* score if he makes a large number of audiences aware of the information. To put it in simpler words, in the time of crisis the users who usually creates and shares tweets are not so famous Twitter users. Surprisingly, *transmitter* score has negative impact on the retweet frequency. This result is counterintuitive as we hypothesized that if a tweet is transmitted to many communities the tweet will be retweeted more which needs further investigation. Like previous works (Suh et al., 2010; Yang et al. 2010), our model suggests that inclusion of hashtag and URL have significant positive impact on the retweet frequency in pre-event time window. However, the effect does not hold at the time of crisis. This might be due to the fact that the tweets get retweeted based on the actual content of the tweet rather than trending hashtags or URLs in it. For the feature *URL* in a tweet, the effect can be explained in a similar way. The effect persists even in the post-event time window. However, when we considered the effect of the interaction term *HashTag*URL*, the effect was positive in all the time windows, i.e., inclusion of both hashtag and URL in the tweet increases its probability of getting retweeted.

	Pre-event	During-event	Post-event
NumFollowers	.0002**	-.0002**	.000014
PeopleAware(t-1)	.00003	.00008***	.00013***
AmplifierScore	-1.10e-08	2.20e-06 ***	5.79e-07***
InformationStarterScore	-.0095***	.098***	.0254***
TransmitterScore	-0.127***	-0.036***	-0.040***
AveragePosition	-.698***	-4.11e-08***	-.614***
HashTag	1.646***	-2.55***	-.658***
URL	.849***	-1.708 ***	-2.053***
HashTag*URL	39.9 ***	2.646 ***	2.097***
* - $p < 0.10$, ** - $p < 0.05$, *** - $p < 0.01$			

The event-centric (here earthquake/bomb blast) nature of our datasets allows us to systematically partition the time range into three distinct time windows (pre-, during-, and post-earthquake) and enables to understand the changes in the effects of the factors inherently for these three time periods. To check the effect of the event we investigate whether there is a significant difference in the retweet frequency in the three time periods. We used difference in difference estimator to compare the effect of the factors in different time windows which has not been presented due to space-constraint. In the model we considered the pre-event time period as the base for comparison. Compared to pre-event time window, number of followers have negative impact on retweetability for both during and post-event time window. On the other hand, during and post-event time window the *amplifier score* and *information-starter score* have a higher positive impact on retweetability during the event. Similarly, inclusion of *hashtag* and *URL* have negative impact in during-event time window. All these aspects give us a signal that during the event the impacts of the factors affecting retweetability are very different in comparison with normal time. Interestingly, some of these effects have long term impact on retweetability in post-event time window like inclusion of *hashtag* and *URL*.

Furthermore, we have reexamined our model using a different Twitter dataset of the Boston marathon bomb blast, which happened on 15th April, 2013. Table 3 suggests that inclusion of *hashtag* and *URL* have significant positive impact on the retweet frequency in pre-event time window. However, this effect is not significant at the time of crisis. In the post-event time window the effect varies. To check the effect of the

event we investigate whether there is a significant difference in the retweet frequency in the three time periods (pre-, during-, and post-bomb blast). Impact of *amplifier* is positive and impact of *information-starter* is negative for all the three time periods. Another interesting finding is that follower count of a user at the time of crisis is not of much importance. Users with a comparatively low number of users tend to participate in the information diffusion significantly.

	Pre-event	During-event	Post-event
NumFollowers	6.61e-08**	-5.91e-08	5.91e-08***
PeopleAware(t-1)	.00003***	.0002***	.00002***
AmplifierScore	2.45e-7***	3.10e-06***	1.77e-07***
InformationStarterScore	-2.88e-08***	-2.13e-06***	-1.61e-08***
TransmitterScore	-0.127***	-0.036***	-0.040***
AveragePosition	-.698***	-.736***	-.136***
HashTag	0.01399***	-.609	-.02896***
URL	0.0605***	-.0578	.00917
HashTag*URL	-.01399***	7.0403***	.02556***
* - p < 0.10, ** - p < 0.05, *** - p < 0.01			

By a direct comparison of the outcomes from both the datasets, it is apparent that most of the variables have similar effects in retweetability in all the time windows for both the datasets except for the *Hashtag*URL* in pre-time window and for information starter in during- and post-time windows. This agreement is quite significant considering the vast disparity in the nature of the two events.

Conclusion and Future Work

Retweet is the core mechanism for information diffusion in Twitter. In this work we have studied the retweet phenomenon to understand the factors affecting retweetability. Earlier researchers (Suh et al., 2010; Yang et al. 2010) have shown that content factors like hashtags or URLs increase the likelihood for a tweet to get retweeted. However, our findings reveal that along with these content features, network features like how many people are made aware in the network (people aware) are very crucial. Users who are present in the beginning of the retweet chain (early retweeters) can make aware most of their followers for the first time and hence contribute largely in the diffusion process. Using datasets of two distinct events, the Great Eastern Japan earthquake and the Boston marathon bomb blast, we examine the effects of these factors in pre-, during-, and post-event time windows and the results obtained from both the datasets are in good agreement. While hashtag and URL have significant positive impacts in pre-event time-window, during the event the effects are opposite. However, the inclusion of hashtag and URL both in the tweet increases the probability of getting retweeted. These changes of effect of the factors in three time-periods demonstrate the influence of the event on retweetability and difference and difference estimator supports these findings. Further, the results show that during the event people do not necessarily retweet the users who have high indegree. In fact, during the event low-indegree users participate in information diffusion significantly as compared to users with large number of followers.

However, the work is still a research-in-progress and we would like to extend this the following ways. In our current study, we have not considered the network variables like eigenvalue-, betweenness-centrality of the users which might be interesting to investigate. We plan to extend our study by incorporating the centrality measures of the users on retweet network, which mostly evolves over time. Also, we have considered two datasets from emergency situations. Here, the evident question that arises is whether it is different from any trending topics, e.g., the death of a celebrity? We plan to collect Twitter data on such non-actionable events to understand whether the effects are different in such scenarios.

References

- A. L. Hughes and L. Palen. Twitter Adoption and Use in Mass Convergence and Emergency Events. In ISCRAM Conference, May 2009
- A. Lancichinetti, F. Radicchi, J. J. Ramasco, and S. Fortunato. Finding Statistically Significant Communities in Networks. *PLoS ONE*, 6(5), 2011.
- B. Suh, L. Hong, P. Pirolli, and E. H. Chi, “Want to be retweeted? large scale analytics on factors impacting retweet in twitter network,” in Social Computing (SocialCom), 2010 IEEE Second International Conference on. IEEE, 2010, pp. 177–184
- D. Boyd, S. Golder, and G. Lotan, “Tweet, tweet, retweet: Conversational aspects of retweeting on twitter,” in System Sciences (HICSS), 2010 43rd Hawaii International Conference on. IEEE, 2010, pp. 1–10
- D. Gruhl, R. Guha, D. Liben-Nowell, and A. Tomkins. Information diffusion through blogspace. In Proceedings of the 13th international conference on World Wide Web, WWW ’04, pages 491–501, New York, NY, USA, 2004. ACM
- F. Samantha and M. Eni. Visualizing Co-retweeting Behavior for Recommending Relevant Real-time Content, Proceedings of the 4th International Workshop on Modeling Social Media, 2013 Paris, France, ACM.
- F. Toriumi, T. Sakaki, K. Shinoda, K. Kazama, S. Kurihara, and I. Noda. Information sharing on twitter during the 2011 catastrophic earthquake. In Proceedings of the 22nd international conference on World Wide Web companion, WWW ’13 Companion, pages 1025–1028, Republic and Canton of Geneva, Switzerland, 2013. International World Wide Web Conferences Steering Committee
- F. Abel, Q. Gao, G.-J. Houben, and K. Tao. Analyzing user modeling on twitter for personalized news recommendations. In Proceedings of the 19th international conference on User modeling, adaption, and personalization, UMAP’11, pages 1–12, Berlin, Heidelberg, 2011. Springer-Verlag.
- GaggleAMP, “The 48 hours life of a tweet”, <https://gaggleamp.com/blog/the-48-hour-life-of-a-tweet/> Last seen on November 4, 2013
- H. Kwak, C. Lee, H. Park, and S. Moon, “What is twitter, a social network or a news media?” in Proceedings of the 19th international conference on World wide web. ACM, 2010, pp. 591–600
- J. Yang and S. Counts, “Predicting the speed, scale, and range of information diffusion in twitter.” *ICWSM*, vol. 10, pp. 355–358, 2010
- J. Leskovec, L. A. Adamic, and B. A. Huberman. The dynamics of viral marketing. *ACM Transactions on the Web (TWEB)*, 1(1):5, 2007
- J. Li and H. Rao, “Twitter as a rapid response news service: An exploration in the context of the 2008 china earthquake,” *The Electronic Journal of Information Systems in Developing Countries*, vol. 42, 2010
- L. Frederic, “The short lifespan of a tweet: retweets only happen in the first hour” http://readwrite.com/2010/09/29/the_short_lifespan_of_a_tweet_retweets_only_happen#awesm=~op3d4Mm7rMSKDB, 2010, last seen on November 4, 2013
- M. T., Bastos, R. Travizki, C. Puschmann. What Sticks With Whom? Twitter Follower-Followee Networks and News Classification. Sixth International AAAI Conference on Weblogs and Social Media. AAAI Press, 2012.
- M. Girvan and M. E. J. Newman. Community structure in social and biological networks. *Proceedings of the National Academy of Sciences*, 99(12):7821–7826, June 2002.
- M. Mendoza, B. Poblete, and C. Castillo, “Twitter under crisis: Can we trust what we rt?” in Proceedings of the first workshop on social media analytics. ACM, 2010, pp. 71–79
- N. Naveed, T. Gottron, J. Kunegis, A. C. Alhadi, Bad News Travels Fast: A Content-based Analysis of Interestingness on Twitter In Proc. Web Science Conf. 2011
- N. Pervin, F. Fang, A. Datta, K. Dutta, and D. E. VanderMeer. Fast, scalable, and context-sensitive detection of trending topics in microblog post streams. *ACM Trans. Management Inf. Syst.*, 3(4):19, 2013
- R. Tinati, L. Carr, W. Hall, and J. Bentwood. “Identifying communicator roles in twitter” Proceedings of the 21st international conference companion on World Wide Web, WWW ’12 Companion, pages 1161–1168, New York, NY, USA, 2012. ACM
- S. Wu, J. M. Hofman, W. A. Mason, & D. J. Watts, 2011. Who Says What to Whom on Twitter? New York: Yahoo! Research.

- T. Sakaki, M. Okazaki, and Y. Matsuo, "Earthquake shakes twitter users: real-time event detection by social sensors," in Proceedings of the 19th international conference on World Wide Web. ACM, 2010, pp. 851–860
- TheDishDaily, "Twitter is Just White Noise", <http://thedishdaily.com/2014/01/27/social-medias-white-noise/> published on January 1, 2014
- Z. Chu, S. Gianvecchio, and H. Wang, "Who is Tweeting on Twitter: Human, Bot, or Cyborg?," in Proceedings of the 26th Annual Computer Security Applications Conference, 2010, pp. 21-30.
- Z. Yang, J. Guo, K. Cai, J. Tang, J. Li, L. Zhang, and Z. Su, "Understanding retweeting behaviours in social networks," in Proceedings of the 19th ACM international conference on Information and knowledge management. ACM, 2010, pp. 1633–1636