

# Retweet Influence on User Popularity Over Time: An Empirical Study

Yecely Aridaí Díaz-Beristain<sup>(✉)</sup>, Guillermo-de-Jesús Hoyos-Rivera,  
and Nicandro Cruz-Ramírez

Centro de Investigación en Inteligencia Artificial, Universidad Veracruzana,  
Sebastián Camacho no 5, 91000 Xalapa, Veracruz, Mexico  
yecely.diaz@gmail.com, {ghoyos,ncruz}@uv.mx

**Abstract.** Web-based Social Networks (W-bSN) have experienced a significant raise in terms of users, as well as the number of relationships among them. One crucial factor for this is the level of influence that a given user can have on other users, and how relationships emerge and disappear among users given the interest generated in a certain community by the posted commentaries. Twitter is the clearest case of W-bSN in which the relevance of the commentaries posted influences the way users create new relationships. In this paper, we analyze the cross influence among users, based on their area of interest, and the messages they post, and how relevant are these messages in the creation of new relationships.

**Keywords:** Social Network · Twitter · Data visualization influence

## 1 Introduction

Web-based Social Networks (W-bSN's) have enormously evolved since their dawn in the late 90's. Many of them had a relative success, such as Microsoft©'s MySpace® or Hi5®. They established the basis for developing a new generation of W-bSN's. Recently, there is a growing interest in two of the most used: Twitter® and Facebook®. Each of them has a considerable amount of users worldwide, and their activity represents many terabytes of information exchanged every day.

Twitter is considered to be a special case of W-bSN, called “micro-blogging”. It implements an asymmetrical model, since there is no need of an authorization in order, for a given user, to create a new relationship (*follow*) with another user. A user willing to receive updates of the activities of another user, e.g. the tweets he posts, only needs to *follow* him, with no action required by this latter. Users' main activity in Twitter is posting *tweets*, short messages of at most 140 characters long, expressing thoughts, ideas, feelings or opinions. Once a user tweets a message, all the following users are notified of this. Another action that can be performed in Twitter is the, so-called, *retweet* (RT), which corresponds to the action of spreading a message previously twitted or RT by another user. It is important to note that the retweeting users do not necessarily have to be following a given user in order for this to be able to retweet him.

Consequently, this model produces some tweets which, given their relevance, are widely spread, becoming eventually viral and, using the Twitter terminology, become

a *Trending Topic* (TT). Thus, a TT is a topic a lot of people is interested about and, in consequence, decides to spread it. Given these characteristics, Twitter has emerged as a suitable platform through which people can try to become popular, i.e., have more followers everyday. If this is the utterly intention, a given user will try to tweet on interesting or controversial subjects, so that people can get interested in them, and eventually can become his follower. In addition, there are passive users who only wish to keep informed about tweets of the people they follow, and do not have an active presence in Twitter. In this order of ideas, it becomes a subject of study to define the factors that can, eventually, allow a user to get a follower of someone else. A user with an important number of followers can be considered as an influential one, since each idea he expresses will automatically reach a big number of users of the platform.

In this paper, we present results of our work, which consisted in analyzing, the factors that may lead that a user gets more followers, and in this way, eventually become influential.

This paper is organized as follows. In Sect. 2 we describe the related work and what other authors have proposed. In Sect. 3 we present the basis for our experimentation. In Sect. 4 the elements necessary to initiate our experimentation. In Sect. 5 the results of our experiment. In Sect. 6 we draw some conclusions and propose some directions for future research. Finally, we present the references.

## 2 Related Work

Since its creation in 2006, Twitter has gained notoriety and popularity. Currently, it has about 313 million active users every month. It has become a near real-time information spreading way widely used worldwide, and its relevance has enormously growth. As explained before, some users tend to have more followers due to the fact that other users get interested in their opinions. These users are considered as influential ones. The relevant element to analyze here is knowing which factors can be considered as important so as to confer a user more followers, and how it influences the number of followers of other users. There are several works that study the influential or “valuable” users [1, 2], the impact of tweeting and retweeting [3, 4], viral marketing [5], etc., which are used to understand the spread of information and the level of the user influence.

Cha et al. [6] present an empirical study of the patterns of influence on users considered as popular. They study three main factors: in-degree (the numbers of users following the user under analysis), number of RTs, and number of mentions. They propose the existence of influential users, i.e., those who can make their tweets to be widely retweeted, as well as to receive a big amount of mentions. From their analysis they concluded that such sort of users tend to publish tweets on controversial subjects. Also, their study affirms that users who limit their tweets to a single topic show a greater increase in the level of influence.

Romero et al. [7] realize the intuitive idea of some users are harder to influence because they are not interested in creating or sharing information, argue that the most of Twitter users are passive. According with the author’s passivity is a barrier to propagation, while some users retweet a lot, others do not do it very often. They propose an

algorithm similar to Hyperlink-Induced Topic Search (HITS) and PageRank to measure the influence considering not only the number of followers, but also the RTs and mentions. They found that influential users are highly active, and as a consequence defined a new influence measure based on user activity.

Retweeting, as stated in [8], has a preponderant importance since the fact of executing such action indicates not only interest in a given tweet, but also the level of confidence deposited in the original publisher, as well as the agreement with the content. This is an important conclusion that helps us supporting our work later.

Another case, presented in [9], affirms that the propagation of the tweets tends to happen by users that have shown to be influential in the past, and who also have an important number of followers. They propose a formula allowing to measure the influence of users taking in consideration the number of RTs and number of mentions they have.

Other researchers have focused in proposing methods to measure the influence of users in subjects with similar topics. Anagnostopoulos et al. [10] define the level of influence as the fact that one individual can induce another individual to act in a similar way. Such type of users are called to be “active”. They present a probabilistic model that evaluates when a user becomes active in a period of time, and they assume that their friends (this is how the author refers to the followers) increase their own probability of becoming active too. They concluded that people are influencing each other every discrete time and estimated the maximum likelihood.

In [11] Crandall et al. study the influence of users based on the “homophily”. This term refers to the level of similarity of people that interact with each other. They divide it in social influence and selection. The first is when people pick up behaviors related with people they interact with, and the selection is when they seek out for similar users to interact with. They quantify the similarity of users over time consider the topic of interest of each user. The authors proposed a model of user behavior where individual users can interact with others and then select the users with a higher number of activities and interactions referred as to influential users.

The work of Weng et al. [12] consists in identifying influential users on Twitter. The strategy is similar to the PageRank algorithm. Their algorithm considers topics extracted from tweets. One of the main contributions is that they compute each user’s topic distribution based on their tweets using LDA, showing that topics of connected users are significantly correlated.

Compared to the previously presented works, ours can be considered as an experimental framework allowing us to analyze the real impact, using real data, that tweets, RTs, and mentions, may have in the level of popularity of Twitter users. We assume as our basic hypothesis that RTs and mentions made by influential users have an effect on the number of followers of a given user.

### 3 Basic Experiment Framework

The influence is the ability that an individual has to modify the perception or beliefs of other people. The reputation of a user has a direct effect on the perception and opinions

of other individuals, and can be effectively used to obtain advantages. The opinion expressed by an influential user can produce, as an effect, that other users change their mind about what they previously thought. For a long time, there have been studies in fields such as sociology, politics, and marketing about the influence that experts, or recognized people, may have, and in consequence, to understand why certain trends appear. For example a campaign turns to be more effective if a message related to it becomes viral. The theory of the traditional communication [13, 14] affirms that a minority of people belonging to a group, denominated distinguished influential, become natural leaders with the capacity of persuading others.

In W-bSN, particularly Twitter, we can also find outstanding users in certain fields and with a number of followers who maintain a more or less permanent interest in their opinions. These users become participatory entities, mentioning a user they are interested on, and sharing those ideas or opinions with their respective followers.

Our main interest consists in analyzing the influence patterns among Twitter users, and how the users considered as experts in a given field can promote the growth of the number of followers of other users, positioning this last through the use of RTs.

Taking into account the proposal of [15], we assumed 18 thematic categories of main subjects of interest (art and design, books, business, charity & deals, fashion, food and drinks, health, holidays & dates, humor, music, politics, religion, science, sports, technology, TV & movies, other news, other). Then we defined six linguistics values based on the number of followers of a given user, as shown in Table 1.

**Table 1.** Linguistic values and their range

Linguistics values	Range of followers	Figure
Unknown	0 to 1,000	Circle
Ordinary	1,001 to 10,000	Square
Outstanding 1	10,001 to 100,000	Triangle
Outstanding 2	100,001 to 1,000,000	Diamond
Outstanding 3	1,000,001 to 10,000,000	Pentagon
Famous	10,000,001 and more	Heptagon

As we can see on Table 1, the “Unknown” linguistic value represents users who either, have just created their accounts, or users with very little activity, generating null attraction for other users. The “Ordinary” users are those who start gaining some popularity and, in consequence, start having new followers who are interested in their timeline and RT him. For the “Outstanding” users, we have defined three different levels based on their number of followers, representing active user accounts that realize diverse posts during the day, so that still more users decide to follow them. According to our results we could infer that these kinds of users have opinions that are respected, so that their influence can be known as important. Finally, for “Famous” users we consider those users who have a huge number of followers. Some users of this type are: @katyperry (92,044,564), @justinbieber (86,775,392), and @BarackObama (76,878,181), to name some.

We found no information about the number of users belonging to each linguistic value. So, we obtained a sample of a million random users to figure out the proportions of the accounts: 98% correspond to the “Unknown” linguistic value, 1.53% to “Ordinary”, “Outstanding 1” has 0.35%, “Outstanding 2” the 0.072%, the “Outstanding 3” has 0.040% and just the 0.008% correspond to “Famous”.

## 4 Experiment Design

In this section, we present the design of the empirical study to analyze the importance of patterns through Twitter users and how they can endorse the increase the number of followers some other users have, by means of the use of RTs.

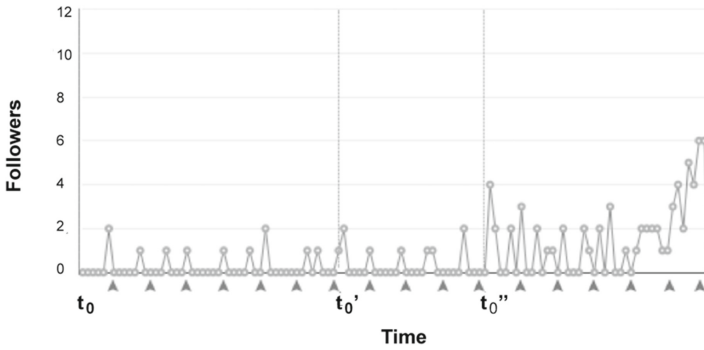
As the first part, we picked a user to be the subject of our analysis, which we will call in the following to be the *Root* user. Using the Twitter API we extracted all the information about the activity of *Root*, that is, tweets, RTs, mentions and new followers. At the beginning of the observation, *Root* had a total of 3,253 followers and his tweets were classified mainly as belonging to in the Technology category. Since the creation of his Twitter account, the growth of followers had a relative stable behavior, getting at most two or three new followers per week, also associated to the Technology category.

Then his behavior was modified through the diversification of his main publishing interests, as well as the use of additional resources to plain text, such as embedded images/videos, or URLs. The new categories in which this user newly participated included Sports and Music. We must explicitly mention that, even if this user used to publish mainly about Technology, tweets posted in other areas contained no relevant information, so other users generally ignored them. In this context, tweets having these other categories as the main subject, were analyzed by the algorithm of [15]. Once we were sure that they corresponded to the desired category, we included a commonly used hashtag (HT) in the tweet. This was made with the intention of making his tweets visible in currently existing conversation threads.

## 5 Experiment Results

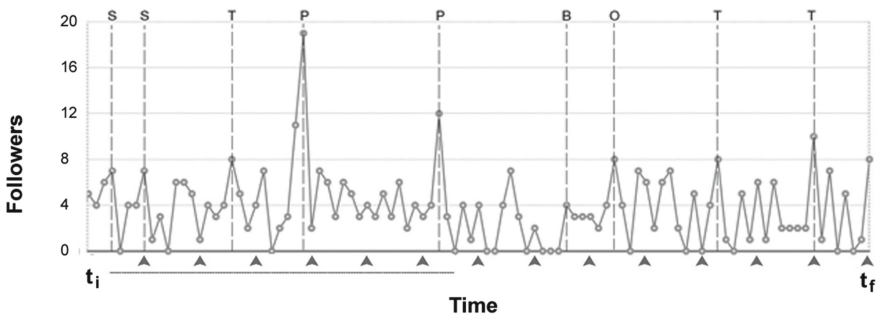
In our experiment we firstly illustrate in Fig. 1 the fact that, there was a regular behavior of *Root* during the seven previous weeks ( $t_0$ ) until the first phase of the experiment ( $t_0'$ ) corresponding to four weeks. Then, we increasing the amount of tweets ( $t_0''$ ) and diversified the subjects (Sports, Politics, Businesses and Others categories). As it can be clearly observed, the amount of new followers per unit of time (week) grows at the right side of the figure.

In Fig. 2, we present a zoom corresponding to the period between  $t_0'$  and  $t_0''$ , which corresponds to the period of fourteen weeks that lasted our experiment. As it can be seen, the number of new followers increased when *Root* published a tweet in the alternative categories. In this figure  $P$  correspond to Politics,  $T$  to Technology,  $B$  for business and  $O$  for others. Vertical lines headed by the letter of the chosen category represents each time *Root* posted a new tweet on a specific category, seeing then the amount of new followers associated to that tweet. It can be observed that tweets of the Politics



**Fig. 1.** Regular ( $t_0$ ), slight ( $t_0'$ ) and strong behavior ( $t_0''$ )

category obtained a substantial increase in the number of followers, because users with “Outstanding 1” linguistic value retweeted the original tweet, and their followers considered it as relevant. The horizontal line at the bottom of Fig. 2 corresponds to part illustrated in Fig. 3, where we can observe the correlation that exists between the RT action and the follow action.

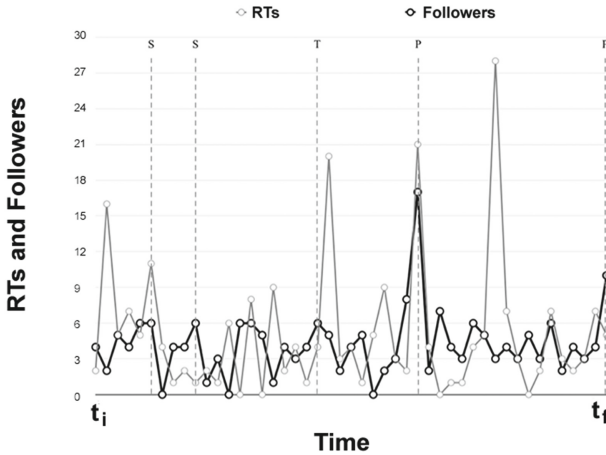


**Fig. 2.** The beginning ( $t_i$ ) and the conclusion ( $t_f$ ) of the accomplished experiment

In this context, we can see that each time *Root* tweets were retweeted (in gray), especially by users with great quantities of followers, his number of new followers increased.

As a collateral result of our experiment, it is important to note that not every user having retweeted a given *Root* tweet, decided to follow him. These users were reached through a RT by an intermediary user they follow, and then they considered the original opinion relevant enough to spread it, but in a first time they did not consider the *Root* user as interesting enough a to start a follow relation on him. Of all the users who did a RT only 37.5% decided to start following. It is important to mention that some users began following *Root* and at the end the 3.2% decided to retire the relation to *Root* user.

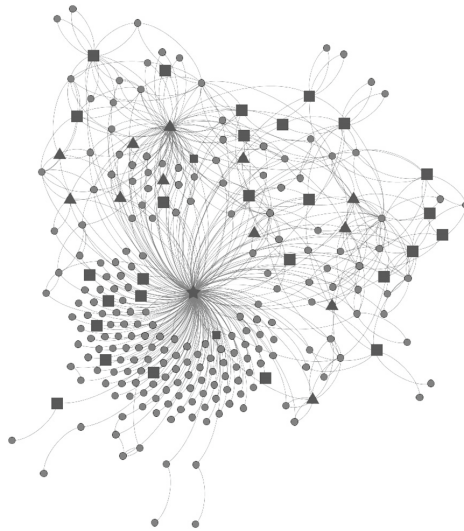
In Fig. 4 we present the users that started following *Root* after a RT of a *Root*'s tweet. By using different symbols, we represent the linguistic values of the other types of users. The star *Root*. Here it can be observed that, per example, the “Outstanding 1” (triangle)



**Fig. 3.** Correlation that exists between RTs and follows

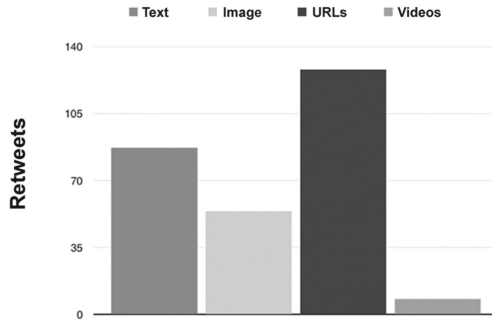
user produced more new followers for *Root* than those produced by an “Unknown” (circle) user. During the experiment corresponding to the diversification of subjects we also discovered that the inclusion of additional resources within a tweet and the determination of the optimal size of the HT, are factors that must also be taken into account in order to get the attention of other users.

The first of these strategies was the incorporation of additional content to the tweet; this can be expressed under the form of image, video or URL. In Fig. 5 we can observe that, the kind of resource included in the tweet, impacted the number of RTs. In this way, in order of relevance, it is possible to observe that the highest impact was obtained



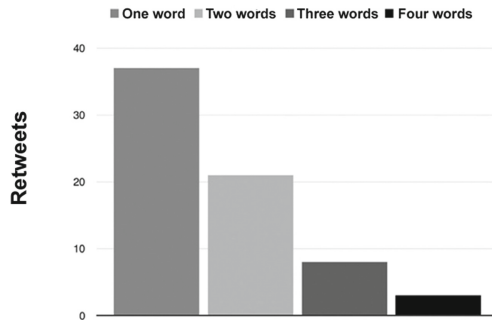
**Fig. 4.** Users that they followed *Root* after a mention or a RT

when a URL is attached, then when an image is associated, and finally when attaching a video. It calls our attention the fact that when a video is included the impact is minimum compared with the other alternatives. We can affirm that we agree with Zarella [16], where it was affirmed that the inclusion of images or URLs does more attractive a tweet for the users.



**Fig. 5.** Including additional content to the tweet

The use of HTs is another way to reach more Twitter users, but choosing an appropriate length for them is crucial for its success. In Fig. 6 we can see that HTs formed by a single word were more successful than those formed by two to four concatenated words.



**Fig. 6.** Including a HT to tweet

In this sense, we disagree with Weng et al. [12] who affirmed that to include HTs it is not important, demonstrating that using a HT of proper size creates more interest in the user. Based on these preliminary it is possible to deduce that the simplicity is a key element. This is shown in Fig. 6.

In order to validate our results, we applied this same experiment to a set of twelve randomly selected, but assuring to have a representation of all our linguistic categories, during a period of one month. The results are shown in Table 2. What we could observe is that the results obtained here correspond with the results obtained



for our *Root*. Users belonging to the “Unknown” or “Ordinary” categories tend to remain in these categories since their tweets are not very interesting or diverse. On the other hand, users belonging to the “Outstanding” or “Famous” linguistic values, who usually are diverse in the way they tweet, in the same sense as stated in this work, tend to get more users as time passes.

**Table 2.** Growth in number of followers

User	Linguistics values	New followers
a <sub>1</sub>	Unknown	4
a <sub>2</sub>	Unknown	-3
a <sub>3</sub>	Ordinary	75
a <sub>4</sub>	Ordinary	-14
a <sub>5</sub>	Outstanding 1	383
a <sub>6</sub>	Outstanding 1	865
a <sub>7</sub>	Outstanding 2	5,371
a <sub>8</sub>	Outstanding 2	49,692
a <sub>9</sub>	Outstanding 3	26,940
a <sub>10</sub>	Outstanding 3	7,727
a <sub>11</sub>	Famous	100,424
a <sub>12</sub>	Famous	532,222

In fact, another observation that we could make is that users tend to remain in the same linguistic category. In this sense, users categorized as “Famous” have an exponential growth in the number of followers; this is a logic consequence since, as more followers spread their tweets, it is easier to gain new followers. For example, the account to @BarackObama in a period of two months had 1,417,820 new followers and tweets with at least 400 RTs. And although some users stopped following him, the number of new followers was bigger than the number of users who stopped following him.

## 6 Conclusions and Future Work

As results of our work we concluded that it was not easy to determine the factors that allow a Twitter user to get more followers. Firstly, we could observe that users tend to keep steady, and it is not very frequent that a user changes of linguistic category. However, we could demonstrate that there exist means through which a user wishing to get more followers, can get it.

The use of several tools, such as HTs, mentions, and included resources, shown to be a good way of getting more users interested in one’s timeline. The length of the HT is another factor that has an influence. As shorter and simpler HTs are used, higher is their probability of success. Another factor to be taken into account is to get users of the “Outstanding” or “Famous” linguistic categories to get interested in one’s tweets. If one of these users gets interested enough in what we tweet, he can retweet, and then our

probabilities of success are higher. Finally, diversifying the subjects of discussion helps a lot also to get more followers.

The proposed strategies can be used on any type of account, not mattering if it is personal or corporative. Our research proposes guidelines to continue with the study of the influence of the users and criteria of growth.

The review field of study provides an opportunity for future research. We considered to focus in analyzing exclusive the behavior of “Outstanding 3” and “Famous” users and thus define new metrics and could be implemented in a model to describe the behavior of influencers.

## References

1. Lee, C., Kwak, H., Park, H., Moon, S.: Finding influentials based on the temporal order of information adoption in twitter. In: Proceedings of the 19th International Conference on World Wide Web, pp. 1137–1138. ACM, April 2010. Maxwell, J.C.: A Treatise on Electricity and Magnetism, 3rd ed., vol. 2. Oxford, Clarendon, pp. 68–73 (1892)
2. Quercia, D., Ellis, J., Capra, L., Crowcroft, J.: In the mood for being influential on twitter. In: 2011 IEEE Third International Conference on Privacy, Security, Risk and Trust (PASSAT) and 2011 IEEE Third International Conference on Social Computing (SocialCom), pp. 307–314. IEEE, October 2011
3. Suh, B., Hong, L., Piroli, P., Chi, E.H.: Want to be retweeted? large scale analytics on factors impacting retweet in twitter network. In: 2010 IEEE second international conference on Social computing (socialcom), pp. 177–184. IEEE, August 2010. Foster, I., Kesselman, C.: The Grid: Blueprint for a New Computing Infrastructure. Morgan Kaufmann, San Francisco (1999)
4. Boyd, D., Golder, S., Lotan, G.: Tweet, tweet, retweet: conversational aspects of retweeting on twitter. In: 2010 43rd Hawaii International Conference on System Sciences (HICSS), pp. 1–10. IEEE, January 2010. Foster, I., Kesselman, C., Nick, J., Tuecke, S.: The physiology of the grid: an open grid services architecture for distributed systems integration. Technical report, Global Grid Forum (2002)
5. Leskovec, J., Adamic, L.A., Huberman, B.A.: The dynamics of viral marketing. *ACM Trans.Web (TWEB)* **1**, 5 (2007)
6. Cha, M., Haddadi, H., Benevenuto, F., Gummadi, P.K.: Measuring User Influence in Twitter: The Million Follower Fallacy. *ICWSM* **10**(10–17), 30 (2010)
7. Romero, D.M., Galuba, W., Asur, S., Huberman, Bernardo A.: Influence and passivity in social media. In: Gunopulos, D., Hofmann, T., Malerba, D., Vazirgiannis, M. (eds.) ECML PKDD 2011. LNCS (LNAI), vol. 6913, pp. 18–33. Springer, Heidelberg (2011). doi: [10.1007/978-3-642-23808-6\\_2](https://doi.org/10.1007/978-3-642-23808-6_2)
8. Metaxas, P., Mustafaraj, E., Wong, K., Zeng, L., O’Keefe, M., Finn, S.: What do retweets indicate? results from user survey and meta-review of research. In: Ninth International AAAI Conference on Web and Social Media, April 2015
9. Bakshy, E., Hofman, J.M., Mason, W.A., Watts, D.J.: Everyone’s an influencer: quantifying influence on twitter. In: Proceedings of the Fourth ACM International Conference on Web Search and Data Mining, pp. 65–74. ACM, February 2011
10. Anagnostopoulos, A., Kumar, R., Mahdian, M.: Influence and correlation in social networks. In: Proceedings of the 14th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, pp. 7–15. ACM, August 2008

11. Crandall, D., Cosley, D., Huttenlocher, D., Kleinberg, J., Suri, S.: Feedback effects between similarity and social influence in online communities. In: Proceedings of the 14th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, pp. 160–168. ACM, August 2008
12. Weng, J., Lim, E. P., Jiang, J., He, Q.: Twitterrank: finding topic-sensitive influential twitterers. In: Proceedings of the Third ACM International Conference on Web Search and Data Mining, pp. 261–270. ACM, February 2010
13. Rogers, E.M.: Diffusion of Innovations. Simon and Schuster, New York (2010)
14. Miller, G.R., Burgoon, M.: Persuasion research: review and commentary. *Commun. Yearb.* **2**, 29–47 (1978)
15. Lee, K., Palsetia, D., Narayanan, R., Patwary, M.M.A., Agrawal, A., Choudhary, A.: Twitter trending topic classification. In: 2011 IEEE 11th International Conference on Data Mining Workshops (ICDMW), pp. 251–258. IEEE, December 2011
16. Zarrella, D.: The science of retweets. Accessed 15 December 2009