4-2010

Finding influentials based on the temporal order of information adoption in twitter

Changhyun LEE

Haewoon KWAK
Singapore Management University, hkwak@smu.edu.sg

Hosung PARK

Sue MOON

Citation

1

Finding Influentials Based on the Temporal Order of Information Adoption in Twitter*

Changhyun Lee, Haewoon Kwak, Hosung Park, and Sue Moon

Department of Computer Science, KAIST, Korea
{chlee, haewoon, hosung}@an.kaist.ac.kr, sbmoon@kaist.edu

ABSTRACT

Twitter offers an explicit mechanism to facilitate information diffusion and has emerged as a new medium for communication. Many approaches to find influentials have been proposed, but they do not consider the temporal order of information adoption. In this work, we propose a novel method to find influentials by considering both the link structure and the temporal order of information adoption in Twitter. Our method finds distinct influentials who are not discovered by other methods.

Categories and Subject Descriptors: J.4 [Computer Applications]: Social and Behavioral Sciences

General Terms: Algorithms, Measurement

Keywords: Social Networks, Twitter, Information Diffusion, Influentials, Ranking

1. INTRODUCTION

Twitter is a microblogging service that has emerged as a new medium for communication recently. Different from most online social networking sites, the relationship of following between users can be unidirectional; a user does not have to follow those who follow him. In Twitter, a user receives all the messages from those he follows, and this unique mechanism of following and subscription of tweets make Twitter a medium of information diffusion.

Many approaches to find influentials have been proposed so far. The simplest approach is to count the number of followers. Another technique is to mine the link structure, such as PageRank. In our previous work [1], we apply the PageRank algorithm and its extensions to Twitter. However, these approaches do not consider the temporal order of information adoption, whereas web pages link only to already existing web pages in world wide web, and they, thus, inherently weigh old information rather more than new information. The order of information adoption is critical to assess the influence. In the theory of diffusion of innovation [2], Rogers divides individuals into five categories by the temporal order of information adoption, and the fast two categories are very social and have the highest degree of opinion leadership.

In this work, we propose a novel method to find influentials by considering both the link structure and the order of information adoption in Twitter. We find influentials by existing approaches including PageRank and demonstrate significant discrepancy among

them. To develop a ranking method that reflects Twitter's explicit mechanism to facilitate information diffusion, we first analyze the information diffusion patterns in Twitter. We discover that information diffusion mostly happens in the early period. Users with many followers are not always the best information spreader when the order of information adoption is taken into account. Therefore, to assess the influence of a user, we propose a measure called *effective readers* who are newly exposed to information. Based on this measure, we develop a simple method to find influentials in Twitter. Our method, unlike others, takes into consideration the temporal order of information adoption.

2. INFORMATION DIFFUSION PATTERNS IN TWITTER

What is a proper way to rank influentials in Twitter? How should we define influentials in Twitter? To answer these questions we conduct a detailed study of information diffusion patterns.

For our experiments, we have crawled profiles of all users on Twitter from June 3rd to September 25th, 2009. Among 41 million users we collected profiles of, there exist 1.47 billion directed relations of following and being followed. We then record the top 10 trending topics every 5 minutes and collect tweets mentioning trending topics. In total we have collected 4,262 unique trending topics and their 223 million tweets for four months.

To analyze how information diffuses over time in Twitter, we first identify the tweets that belong to the same thread of discussion within a trending topic. Kwak *et al.* show that a trending topic becomes active and inactive repeatedly [1]. In this work we rely on the temporal proximity to cluster tweets of the same context and leave natural language processing of tweets for semantic verification for future work.

Within each thread of trending topics above, we calculate the cumulative number of users who have received a tweet regarding each topic and see how this number changes over time. A user may or may not read all the tweets one receives, and we label these users *potential readers*. We plot the information spreading over time in terms of the cumulative number of potential readers. A few selected topics are shown in Figure 1.

The cumulative number of potential readers increases fast in the early stage, and its growth slows down over time. This behavior indicates that information diffusion mostly happens in the early period in Twitter and early tweets spread better. We then plot the growth of the cumulative number of writers over time as a red dotted line in Figure 1. Comparing the growth of potential readers with that of writers, we find that the growth rate of potential readers slows down even when the number of writers increases steadily as most clearly shown in 'harry potter'. Influence of writers is time-dependent, and late writers have less influence than early ones.
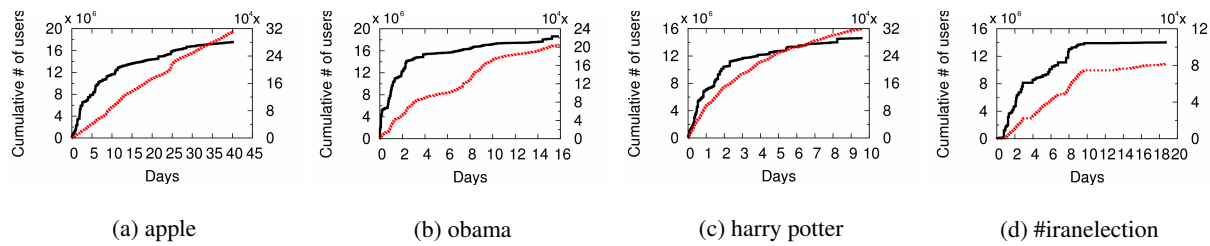
(a) apple　　　　　(b) obama　　　　　(c) harry potter　　　　　(d) #iranelection

Figure 1: The growth of potential readers (above black solid line) and writers (below red dotted line) over time. The tics of left y-axis is the number of potential readers, and that of right y-axis is the number of writers.

Rank	Ranking by effective readers		Ranking by # of followers		Ranking by PageRank		Ranking by # of retweeted tweets	
	Identity	Remark	Identity	Remark	Identity	Remark	Identity	Remark
1	Pete Cashmore	news on social media	ashton kutcher	actor	ashton kutcher	actor	Pete Cashmore	news on social media
2	NPR Politics news	news on politics	Britney Spears	musician	Barack Obama	president of U.S.	BNO News	news
3	E! Online	news on entertainment	Ellen DeGeneres	show host	CNN Breaking News	news	TweetMeme	news on Twitter
4	Brandon Mendelson	author	CNN Breaking News	news	Ellen DeGeneres	show host	oxfordgirl	journalist
5	CNN Breaking News	news	Oprah Winfrey	show host	Britney Spears	musician	CNN Breaking News	news
6	The New York Times	news	Twitter	subject of this paper	Oprah Winfrey	show host	Michael Arrington	news on technology
7	SiliconAlleyInsider	news on technology	Barack Obama	president of U.S.	THE_REAL_SHAQ	sports star	Fabolous	musician
8	Guardian Tech	news on technology	Ryan Seacrest	show host	John Mayer	musician	The New York Times	news
9	Michael Arrington	news on technology	THE_REAL_SHAQ	sports star	Twitter	subject of this paper	lil duval	comedian
10	Ana Marie Cox	magazine editor	Kim Kardashian	model	Ryan Seacrest	show host	Iran	about Iran

Figure 2: Top 10 users ranked by effective readers, the number of followers, PageRank, the number of retweeted tweets

Not all followers of a user hear about a certain topic from the user; they could hear it first from others they follow. We define the effective reader of a writer as the user who has been exposed to the thread for the first time through the writer's tweet. To compare the number of followers and the number of effective readers of each user, for 80% of users only 20% of their followers turn out to be effective readers. That is, having many followers does not always make a user influential in information diffusion. Our findings underline the importance of the temporal order in information adoption.

3. FINDING INFLUENTIALS WITH EFFECTIVE READERS

In the previous section we have shown that effective readers can be far fewer than potential readers. The concept of effective readers reflects the importance of timeliness of information adoption as argued in previous literature [2]. Therefore we find influentials based on the number of effective readers of a user.

In our method, a user can be in either of two states: the user has already read a tweet of a trending topic (state 1) or not read yet (state 0). We assume that a user reads all tweets of followings in chronological order.

- Initialization: All users begin in state 0.

$$\forall u \in \mathbf{U}, S(u) = 0 \quad (1)$$

where $\mathbf{U}$ is the set of all users and $S(u)$ is the state of user u.

- Information diffusion: A user changes to state 1 if followings write a tweet of a trending topic. Once a state of a user changes to state 1, that of the user never returns to state 0.

Effective readers, ER, of tweet w written by user u are defined as:

$$ER_0(w) = \{v | v \in follower(u) \text{ and } S(v) = 0\} \quad (2)$$

where $follower(u)$ is the set of user u's followers. The influence of a user u is defined as the total count of the effective readers for all tweets of user u.

$$IF_0(u) = \sum_{w \in \mathbf{T}(u)} \|ER_0(w)\| \quad (3)$$

where $\mathbf{T}(u)$ is all tweets written by user u.

We present the top 10 influentials by our method and others in Figure 2 (Retweeted tweets are tweets cited by other users). Different from other methods, most of the influential users in our model are news media. We claim that news media has significant influence in spreading information to effective readers. Quantitative comparison of our model with ranking by the number of followers shows there is only 34% of overlaps for the top $1,000$ influentials. The overlaps with other algorithms such as PageRank are even lower, showing the uniqueness of our method.

4. FUTURE WORK

While our model finds distinct influentials who are not discovered by other methods, a few more considerations can improve the model to reflect human nature in the real world. First, a user may not read all the tweets of his followings. The more followings a user has, the less likely one pays attention to all followings. Second, not all users read or remember the same even under the same condition. People forget information as time passes. In the future, we plan to extend our method with the above considerations.

5. REFERENCES

[1] H. Kwak, C. Lee, H. Park, and S. Moon. What is Twitter, a Social Network or a News Media? In *WWW '10: Proceedings of the 19th international conference on World Wide Web*, 2010.

[2] E. M. Rogers. *Diffusion of Innovations, 5th Edition*. Free Press, 5 edition, August 2003.