Novel Aspects Coming From the Directionality of Online Relationships: A Case Study of Twitter

Haewoon Kwak, Changhyun Lee, Hosung Park, Hyunwoo Chun and Sue Moon
World Class University (WCU) program of Web Science Technology
College of Information Science & Technology, KAIST, Korea

The online relationship in Twitter, known as follow, is directed. People can follow any other person without an approval. In this article we show novel aspects of Twitter that come from the directionality in relationships: topological characteristics of the directed network, word-of-mouth information spreading via retweet, and online relationship dissolution. We wrap up the article with future directions in our information diffusion study.

DOI: 10.1145/1980002.1980007 http://doi.acm.org/10.1145/1980002.1980007

1. INTRODUCTION

Online social networking service (OSN), one of the fastest growing web services, has become a significant source of recorded human activities. For instance, on Facebook 30 billion pieces of contents including web links and photos are newly uploaded each month, and on Twitter users generate one billion short messages per week as of 2011. OSNs have become the medium for people to share daily lives by exchanging short messages, uploading photos, posting brief memos, etc. The friend relationship lays the foundation for other systems to build upon for recommendation engines, cooperation-based security, online search, and other personalization functions. A good knowledge of friend relationships is essential to the success of such systems.

Most OSNs, such as Myspace and Facebook, offer a means to establish explicit online relationships called 'friends'. These relationships are commonly bidirectional. If user A is a friend of user B, user B is also a friend of user A. Researchers have found very social characteristics, such as high clustering, from the friend network based on reciprocal relationships there. Meanwhile, unlike most OSNs, the online relationship in Twitter, known as *follow*, has the directionality. A user can follow any other user and the user being followed need not follow back; the relationship of following and being followed requires no reciprocation. This one-sided relationship is quite unique and deserves scrutiny.

In this article we investigate the following three problems that we believe are affected by Twitter's unique features. First, we investigate if the topological characteristics are similar

to or different from other known OSNs. Then we report on how people spread information via word-of-mouth on Twitter. Other OSNs also allow information sharing, but esoteric features such as retweets and mentions help users go one step further and not only share but *spread* information easily. Lastly, we report on the pervasiveness of unfollow. When relationships are bidirectional, breaking an existing relationship has emotional cost [Tong et al. 2008]. As the follow relationship is one-way, unfollow is likely to have little of such cost and we examine its pervasiveness. This article is a brief summary of our ongoing research on Twitter.

2. STRUCTURAL PROPERTIES OF TWITTER'S FOLLOW NETWORK

First, we show how the directionality of follow relationships affects the entire network structure. We have crawled the near-complete Twitter follow network and obtained 41.7 million user profiles and 1.47 billion follow relations among them as of 2010[1].

Low reciprocity It is well known that "reciprocal interaction pervades every relation of primitive life" [Thurnwald 1932]. Nevertheless, the reciprocity of declared online relationships is not easy to measure because they are typically bidirectional and require an additional effort to gauge reciprocity. A previous study measures reciprocity in guestbook activities and reports a high reciprocity of 77% on Cyworld guestbook [Chun et al. 2008].

Calculating reciprocity in a directed network is straight-forward. In Twitter we find a low level of reciprocity; 77.9% of user pairs with the follow relationship are one-way and only 22.1% are reciprocal.

Non-power-law follower distribution Power-law has been observed in numerous and diverse networks and is understood as a natural law of organization. We fit the follower degree distribution of Twitter to a power-law distribution and obtain the exponent of 2.276, where a power-law distribution is supposed to have the exponent between 2 and 3. However, our distribution is not a straight line in the log-log plot of the Complimentary Cumulative Density Function (CCDF), and there are far more users with followers $> 10^5$ than the power-law distribution predicts. A similar tail behavior in a degree distribution has been reported from Cyworld in [Ahn et al. 2007], but not from other social networks. What is common between Twitter and Cyworld is the high visibility of celebrities. Fans can locate celebrities easily and establish online relationships in both services.

Short average diameter of 4.1 Ever since Stanley Milgram's 'six degrees of separation' experiment [Milgram 1967], the concept of degrees of separation has become one of the key metrics to summarize the structure of a social network. Recently, Leskovec and Horvitz report on the MSN messenger network of 180 million users that the median and the 90% degrees of separation are 6.0 and 7.8, respectively [Leskovec and Horvitz 2008]. In Twitter, the information pathway is exactly the reverse path of follow relationships because a tweet or information flows from a user to his followers. We calculate the median and mode of the path lengths between all pairs of users and they are both 4; and the average path length is 4.12. We note that a piece of information crosses 5 or fewer hops between 93.5% of user pairs, taking fewer hops than in other known social networks. These measured lengths are quite short for a network of the Twitter size. It is an interesting phenomenon that bespeaks

[1] We make our dataset publicly available at `http://an.kaist.ac.kr/traces/WWW2010.html`

for the strength of Twitter's role as an information spreading medium.

3. INFORMATION DIFFUSION IN TWITTER

Twitter users created a tweet-relaying convention called retweet, which now is a functional feature supported by Twitter. It is an effective means for a user to relay a followee's tweet to his followers. It is considered the feature that has made Twitter a new medium of information dissemination.

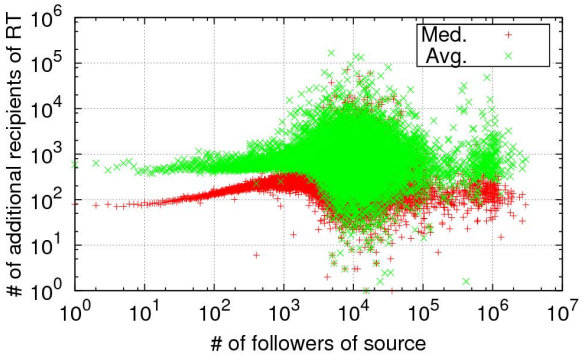

Fig. 1. Average and median numbers of additional recipients of the tweet via retweets

As a quantitative measure for the impact of retweets in information diffusion, we count the number of additional recipients who are not immediate followers of the original tweet writer. Figure 1 from [Kwak et al. 2010] displays its average and median per tweet against the number of followers of the original tweet writer. The median lies almost always below the average, indicating that many tweets have a very large number of additional recipients. Up to about $1,000$ followers, the average number of additional recipients is not affected by the number of followers of the tweet source. In other words, no matter how few followers a user has, the tweet is likely to reach a certain number of audience, once the user's tweet starts spreading by retweets. This shows the power of retweeting; the mechanism of retweet has given every user the power to spread information. Individual users have the power to dictate which information is important and should spread by retweet as in the form of word-of-mouth, which collectively determines the importance of the original tweet. In a way we are witnessing the emergence of collective intelligence in the form of individual and independent decisions of retweets.

4. RELATIONSHIP DISSOLUTION IN TWITTER

'Friend' relationships in most OSNs are known to remain rigid regardless of the actual relationship [Tong et al. 2008]. In contrast Twitter users can easily stop following (*unfollow*) because there needs no confirmation from the followee to do so. Unfollow, thus, is a verifiable action of breaking an online relationship. We collected daily snapshots of the follow relationships of 1.2 million Korean-speaking users over the course of 51 days as well as their tweets. Since Twitter does not offer an explicit record of the unfollow actions,

we compare two consecutive snapshots and infer the act of unfollow among users. We have found that unfollow is prevalent in Twitter, and the reciprocity of the relationship, the duration of the relationships, the followees' informativeness, and the overlap of relationships are critical to one's decision to unfollow others [Kwak et al. 2011]. We followed up with interviews with 22 users to understand the motivations behind the unfollow behavior. Our survey respondents answered that they unfollowed those who left many tweets within a short time, tweeted about topics uninteresting to them, or babble about mundane life, where the order of the answers is in decreasing importance.

5. FUTURE DIRECTIONS

We are currently analyzing temporal and topological characteristics of information diffusion on Twitter. We are tracking the TwitPic[2] URLs that point to user-uploaded photos. The preliminary result shows that a TwitPic URL spreads fast. This observation agrees with the fast diffusion of tweets mentioning trending topics in [Kwak et al. 2010]. We also observe the shape of diffusion trees are quite similar irrespective of contents types. Our study of non-textual information diffusion complements the existing studies of textual information diffusion in Twitter. In a way the relationship dissolution is a form of information filtering. Combined with studies of information diffusion, it will shed light on how we consume information online.

Acknowledgement

This research was supported by WCU (World Class University) program under the National Research Foundation of Korea and funded by the Ministry of Eduation, Science and Technology of Korea (Project No: R31-30007)

REFERENCES

AHN, Y.-Y., HAN, S., KWAK, H., MOON, S., AND JEONG, H. 2007. Analysis of topological characteristics of huge online social networking services. In *WWW '07: Proceedings of the 16th international conference on World Wide Web*. ACM Press, New York, NY, USA, 835–844.

CHUN, H., KWAK, H., EOM, Y.-H., AHN, Y.-Y., MOON, S., AND JEONG, H. 2008. Comparison of online social relations in volume vs interaction: a case study of Cyworld. In *Proc. of the 8th ACM SIGCOMM Internet Measurement Conference*. ACM, New York, NY, USA, 57–70.

KWAK, H., CHUN, H., AND MOON, S. 2011. Fragile online relationship: A first look at unfollow dynamics in Twitter. In *CHI '11: Proccedings of the 29th international conference on Human Factors in Computing Systems*.

KWAK, H., LEE, C., PARK, H., AND MOON, S. 2010. What is Twitter, a social network or a news media? In *WWW '10: Proceedings of the 19th international conference on World wide web*. ACM, New York, NY, USA, 591–600.

LESKOVEC, J. AND HORVITZ, E. 2008. Planetary-scale views on a large instant-messaging network. In *Proc. of the 17th international conference on World Wide Web*. ACM, New York, NY, USA, 915–924.

MILGRAM, S. 1967. The small world problem. *Psychology today 2,* 1, 60–67.

THURNWALD, R. 1932. *Economics in Primitive Communities*. Oxford University Press.

TONG, S. T., VAN DER HEIDE, B., LANGWELL, L., AND WALTHER, J. B. 2008. Too much of a good thing? the relationship between number of friends and interpersonal impressions on facebook. *Journal of Computer-Mediated Communication 13,* 3, 531–549.

[2]http://twitpic.com