Computer and Information Science; Vol. 10, No. 3; 2017
ISSN 1913-8989 E-ISSN 1913-8997
Published by Canadian Center of Science and Education

A Successful Advertising Strategy over Twitter

Kyota Okubo[1] & Kazumasa Oida[1]

[1] Graduate School of Engineering, Fukuoka Institute of Technology, Fukuoka, Japan

Correspondence: Kazumasa Oida, Graduate School of Engineering, Fukuoka Institute of Technology, 3-30-1, Wajiro-Higashi, Fukuoka, Japan. Tel: 8192-606-3782. E-mail: oida@fit.ac.jp

Received: April 22, 2017 Accepted: May 20, 2017 Online Published: July 10, 2017

doi:10.5539/cis.v10n3p10 URL: https://doi.org/10.5539/cis.v10n3p10

Abstract

Large information cascades over online social networks have attracted a great deal of attention. The life times of most cascades are quite short, whereas recent advertising campaigns sometimes generate long-lived ones by employing effective dissemination strategies (instant-win, reminders, etc.). This paper reports one such campaign, YOGUR STAND, on the Twitter network in Japan. The data analysis shows that the campaign popularity has two interesting features. (1) It shows elastic behavior in that the impact of the destructive earthquake on the popularity was quite temporary. (2) It exhibits stationary behavior in that the campaign account gained approximately 2,000 new followers every day. The analysis also demonstrates that there were communities in the campaign participants. The campaign was successful because about 2.4 million Twitter users received the campaign retweets every day and 10-15% of them received the retweets for the first time.

Keywords: viral marketing, advertising campaign, information cascade, Twitter community

1. Introduction

Nowadays, large-scale information cascades on information networks, especially over online social networks (OSNs), have attracted a great deal of attention in many fields. In the case of content delivery networks (CDNs), for example, content virality information is crucial for efficient utilization of cloud resources (Krishnaswamy, Krishnan, Lopez, Willis, & Qamar, 2015). Accordingly, numbers of methods that predict the growth of the cascades have been intensively devised (Cheng, Adamic, Dow, Kleinberg, & Leskovec, 2014; Galuba, Aberer, Chakraborty, Despotovic, & Kellerer, 2010; Zhao, Erdogdu, He, Rajaraman, & Leskovec, 2015; Gao, Ma, & Chen, 2015).

The information diffusion process on the social network is considered more complex than those of infectious diseases and computer viruses. The size and the growth rate of an information cascade depend not only on the attractiveness inherent in the content and the structure of the underlying social network but also on user communities present in the social network because they suppress or accelerate the spread of cascades (Weng, Menczer, & Ahn, 2013). In addition, viral marketing (Klopper, 2002), a marketing strategy that focuses on spreading product information from person to person through social networks, and various paid services (e.g., Twitter promotion services) make the information diffusion processes even more complex.

Meanwhile, current advertising campaigns adopt various dissemination strategies over OSNs. For example, some strategies attempt to draw attention from online users by paying rewards to users who contributed to the dissemination or by incorporating games that contributors can enjoy. In this paper, we report the YOGUR STAND campaign, which adopts a worked-out strategy for disseminating the campaign web page over the Twitter network in Japan. The aim of this paper is (1) to report the details of the campaign so that our work plays an important role in creating effective advertising strategies and (2) to discuss the reason why a long-lived diffusion phenomenon occurred over the Twitter network.

The campaign strategy has the reminder and instant-win effects, by which community members were stimulated to retweet many times. The campaign was successful because during the campaign period, the YOGUR STAND account constantly acquired thousands of new followers every day. Furthermore, destructive earthquakes during the campaign could not hamper its success. Our data analysis demonstrates that 2.4 million Twitter users saw the campaign retweets every day and 10-15% of them saw the retweets for the first time. Since the life times of most information cascades are quite short (Goel, Watts, & Goldstein, 2012), the majority of large cascades in future might belong to this advertising campaign type.

The paper is organized as follows. Section 2 presents the related work. Section 3 describes the campaign strategies. Section 4 analyzes Twitter data to demonstrate that the campaign popularity presents both elastic and stationary behaviors. Section 5 provides various statistical results on campaign followers and user communities. Section 6 infers the cascade structure, which indicates how the YOGUR STAND web page spread with time, and Section 7 concludes the paper.

2. Related Work

Large-scale information cascades over online social networks have been intensively studied from a variety of viewpoints; e.g., the impact of communities on the growth of cascades (Weng et al., 2013; Forestier, Bergier, Bouanan, Ribault, Zacharewicz, Vallespir, & Faucher, 2015; Bakshy, Rosenn, Marlow, Adamic, 2012; Guo, Shaabani, Bhatnagar, & Shakarian, 2015), various features that large-scale cascades possess in common (Cheng et al., 2014; Myers, & Leskovec, 2014; Dow, Adamic, & Friggeri, 2013; Guille, Hacid, Favre, & Zighed, 2013; Cheng, Adamic, Kleinberg, & Leskovec, 2016; Adamic, Lento, Adar, & Ng, 2016), and the cascade models for predicting the future growth (Galuba et al., 2010; Zhao et al., 2015; Gao et al., 2015; Cheung, She, Junus, & Cao, 2016; Mishra, Rizoiu, & Xie, 2016; Elsharkawy, Hassan, Nabhan, & Roushdy, 2016; Rong, Zhu, & Cheng, 2016; Krishnan, Butler, Tandon, Leskovec, & Ramakrishnan, 2016).

More recently, many authors attempt to explain information dissemination processes based on human behavioral patterns. The authors in (Chen, Chen, & Agarwal, 2017) focus on user reactions after receiving positive feedback such as likes and retweets, which the authors call social incentives. In (Hoang & Lim, 2016), the authors point out three behavior factors (user virality, user susceptibility, and item virality) and propose new models that describe the relationships between them. Note that these approaches might need more sample data for evaluation because large information cascades seldom emerge.

Recent studies on viral marketing, on the other hand, generally formulate optimization problems, which maximize influence, revenue, sales campaigns, etc., based on information diffusion conditions, which are mostly on user behavior. Compared with previous work, current studies assume more detailed and realistic human behavior and incorporate more adoption factors into the campaign strategies. Some studies assume users may not have a complete knowledge about the products (Robles, Chica, & Cordn, 2016; Cordasco, Gargano, Rescigno, & Vaccaro, 2016), and some assume that inference recommendations from stores may have a non-negligible effect on the purchase decision (Hung, Shuai, Yang, Huang, Lee, Pei, & Chen, 2016). Some authors consider even more realistic models by introducing the timeliness of the product adoption (Lamba & Pfeffer, 2016; Khan, 2016). The authors in (Zhu, Peng, Chen, Zheng, & Zhou, 2016) devise a model that promote the sale of products and services using the current location. Meanwhile, our work is new in that we introduce a campaign strategy that has the reminder and instant-win effects.

The majority of authors in this research area provide simulation results, while some papers report analytical data on real dissemination phenomena. In (Dow et al., 2013), the authors investigate how a former President Barack Obama's photo generates cascades of reshares after his election victory. In (Chiu & Hsu, 2017), the authors report the information diffusion processes on the Sunflower Student Movement occurred in Taiwan. While, this paper deals with a successful advertising strategy using collected Twitter data. These practical reports are valuable in that large information cascades are not frequent evens. We think these reports play an important role in connecting theoretical work and real phenomena.

3. Advertising Strategy

This section explains a sale promotion campaign made by Coca-Cola Japan. The campaign was conducted over the Twitter network for about one month. The aim of the campaign is to increase the awareness of new yogurt product YOGUR STAND by increasing the number of Twitter users who receive retweet messages about the campaign. To do this, the campaign employed an incentive-based marketing strategy; a total of 1,000 users get two bottles of yogurt.

Figure 1 illustrates how to get the incentive. First of all, users who want the incentive must follow the campaign account (@YOGURSTAND). If user f in the figure follows the campaign account, the user receives a tweet message from the account every day at just after midnight. If user f retweets the tweet, the account quickly returns a reply including an 11-second slot machine video, indicating whether user f wins or not. A follower receives the lottery result only once a day no matter how many times the follower retweets. According to the interview on the YOGUR STAND campaign, about 100,000 people retweeted the campaign and they sent retweets five times on average.

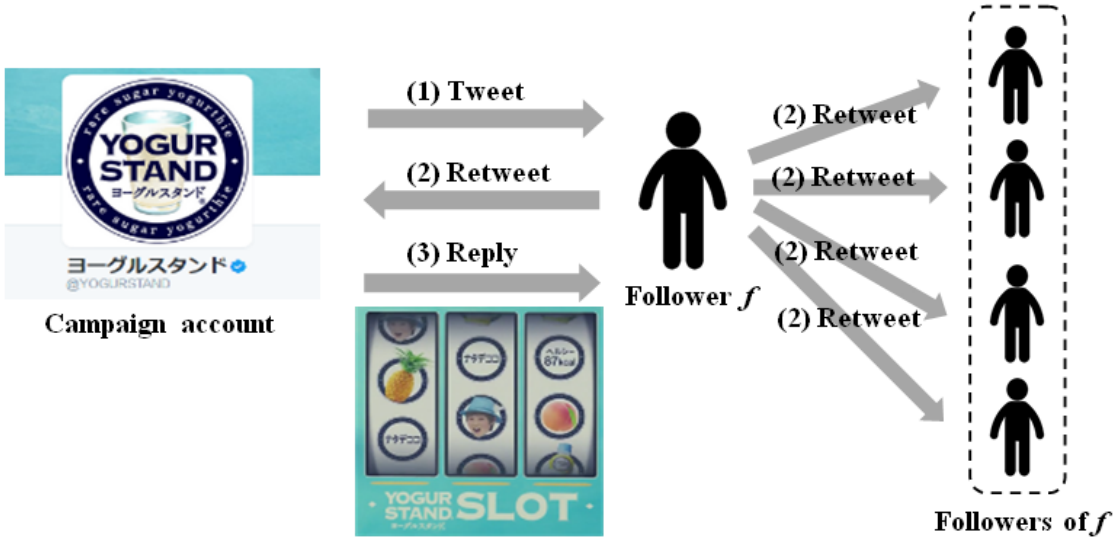

Figure 1. (1) Participants (followers of campaign account @YOGURSTAND) receives a tweet from the account every day. (2) If a participant *f* makes a retweet, the retweet (including the campaign web page URL) is shared among the followers of *f* and (3) a video notifying the lottery result soon arrives

As shown in Figure 1, a retweet from follower *f* is also transmitted to all followers of *f*. As a result, all these followers notice the new yogurt product. Accordingly, a key point of the success of this advertising campaign is to increase the number of campaign followers *f* (which we call participants). To get new participants, (re)tweets include the URL of the campaign web page that explains how to get the incentive. Generally, a user becomes a participant with a higher probability as the user more often receives the retweets. It is important to note that there exists a positive feedback mechanism: as the number of participants grows, more users see the retweet, so that more users become participants.

The campaign attempts to raise the number of retweets based on two effects: reminder and instant-win effects. Once a user becomes a follower of the account, the user receives a tweet message every day, which "reminds" the user of the campaign and stimulates the user to retweet; as a result, the number of retweets increases. After a participant makes a retweet, the campaign account always quickly returns a slot machine video notifying the lottery result (see Figure 1), and this "instant-win" approach inclines participants to retweet again.

4. Popularity Dynamics

4.1 Popular Contents

We investigated the most shared contents (web pages, photos, etc.) on the Twitter network in Japan by collecting Twitter messages written in Japanese using the Streaming APIs, which are provided by Twitter. We then calculated the ten most popular URLs per day. The popularity of a URL is evaluated based on the number of collected messages that contains the URL. The Streaming APIs allow us to collect a small percentage of all Twitter messages and the percentage may depend on time. Therefore, a popularity value is a measurement obtained from sample messages. We select the most popular contents through the Streaming APIs and then use the REST APIs when more precise quantitative analysis is required. Appendix A details how we obtained Twitter data through these APIs.

Figure 2 illustrates the popularity changes of three popular URLs: the YOGUR STAND campaign web page, the YOGUR STAND response video, and a web page on the earthquake. The collected data indicate that long-lived URLs are quite rare. We gathered 3,556,290 URLs in two months and only twelve of them were ranked in the top ten URLs ten times or more. The campaign web page was one of the twelve URLs. Figure 2 demonstrates a stable behavior of the campaign web page popularity. The campaign popularity sharply increased and entered a stationary state in two days. Then the popularity suddenly dropped on April 14 due to the occurrence of the Kumamoto earthquake. The popularity, however, began to shift to a higher level on April 22 (partly because people began to forget about the disaster). After the occurrence of the earthquake, a large number of web pages reporting the disaster appeared all at once, but they soon disappeared (see one such report in Figure 2).

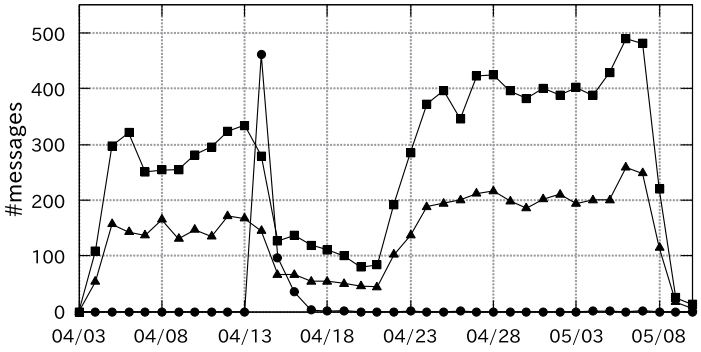

Figure 2. The popularity changes of the campaign web page (■), the response video from the campaign account (▲), and a web page on the earthquake (●). The campaign period is from April 4 to May 8

4.2 Campaign Effects

To investigate the popularity of the campaign web page from a different viewpoint, we collected "all" retweet messages that contain the web page URL using the REST APIs and extracted user IDs from them. Figure 3 compares the numbers of users who retweeted the campaign URL for the first time (shortly called new adopters) and who retweeted the URL more than once (shortly called repeaters). From the figure, the total number of users who retweeted the URL per day is roughly 14,000. Thus, the yogurt product information propagated from the YOGUR STAND account to the 14,000 Twitter users and their followers every day.

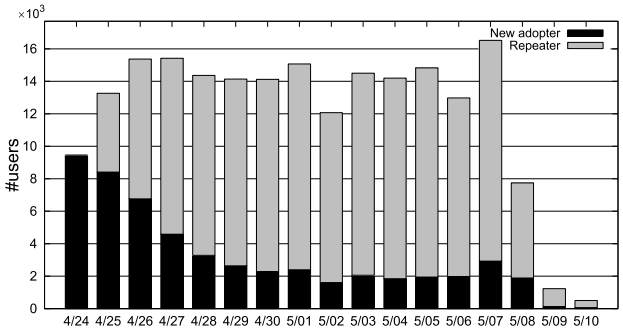

Figure 3. The numbers of users who retweeted for the first time (new adopters) and who retweeted more than once (repeaters) since April 24

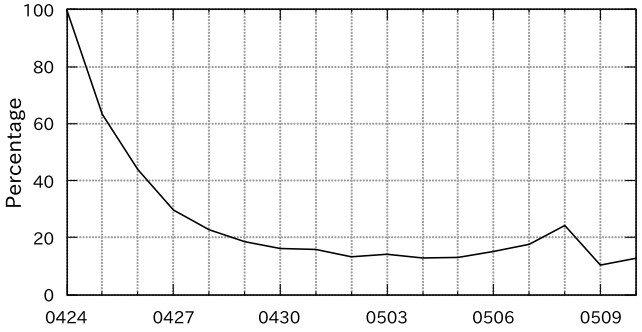

Figure 4. The percentage of users who retweeted for the first time since April 24

The campaign may not be considered successful if nearly all retweets were made by the same persons because the web page spreads more widely when new adopters retweet. Figure 3 exhibits that the majority of users who

made retweets were repeaters; however, new adopters constantly appeared. It can be seen from the figure that the number of new adopters converges to roughly 2,000 with time. This result is noteworthy because successful viral marketing over social networks is considered to yield an exponential growth of new adopters instead of the constant rate emergence (Klopper, 2002). Figure 4 exhibits the percentage of new adopters, which is obtained from Figure 3. The percentage seems to converge to 10-15% even though it somewhat rises on May 8 (the campaign expiration date).

5. Community Analysis

5.1 Sample Set

To infer statistical characteristics on users who retweeted the campaign account tweet (participants), we collected a sample of participants (called participant set G) by the following procedure:

1) Select a user u who retweeted the URL on April 24 and initialize the set as $G = \{u\}$.

2) If a follower f of a user in G retweeted the URL on April 24 or later, add f to set G.

3) Repeat 2) until there is no user to be added to the set.

Note that all members in G are participants and they are connected based on the followee-follower relationship. The number of members in G (denoted by $|G|$) is 16,747, which is 1/6 of the total participants reported in the interview on the YOGUR STAND campaign. Participants in set G were collected using the GET followers/ids API. The API did not always correctly respond. Due to the errors described in Appendix B, the API did not return follower IDs for 1,599 users.

5.2 Sample Statistics

Figure 5 shows two histograms that are derived from sample set G. Figure 5 (Left) exhibits the histogram of the number of followers per user. The histogram has a heavy tail because it roughly agrees with a log-normal distribution. The mean, maximum, and minimum of the number of followers are 505.8, 167,209, and 0, respectively. From a survey performed in 2016, this mean number roughly agrees with the mean number of followers per Twitter user in Japan (which is 426).

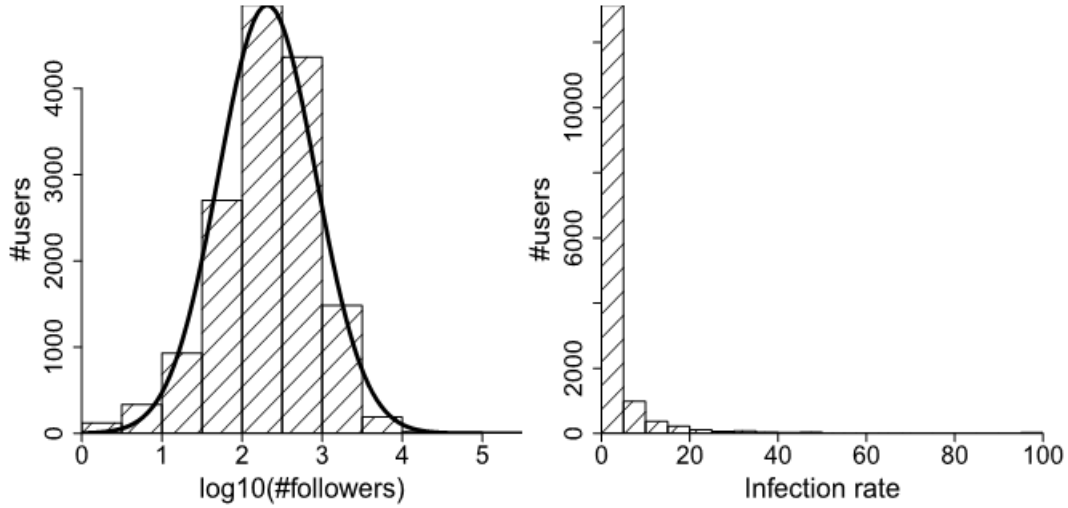

Figure 5. Left: the histogram of the number of followers per user, excluding users who have no followers (10%) and the probability density function of a log-normal distribution. Right: the histogram of the infection rate per user

Let us estimate how many Twitter accounts received the campaign URL every day. From Figures 3 and 5, 14,000 users retweet the URL and the mean number of followers is 505.8. Accordingly, about 2.4 million people receive the URL every day because $14{,}000+14{,}000\times505.8\times0.342=2{,}435{,}770$, where 0.342 represents the non-overlapping follower rate. From Figure 3, there are 2,000 new adopters every day. Then, the same calculation indicates that about 0.35 million people received the URL for the first time every day.

The term infection rate is used hereafter to indicate the percentage of the number of followers who are participants. Figure 5 (Right) illustrates the histogram of the infection rate per user. It can be seen that the

percentage is quite small. The mean of the histogram is 3.017%. On the other hand, the infection rate of all users in *G* is 0.435%, which we call the infection rate of set *G*. This small percentage implies that infection is a rare event (even though followers repeatedly heard retweets from their followees). Note, however, that even this low percentage generated thousands of new adopters.

The box plot in Figure 6(a) demonstrates that the mean infection rate per user monotonically decreases with the number of followers per user (we should ignore the rate at 100000-999999 because it is derived from two samples). This result is natural because according to the procedure of creating set *G*, a user in *G* has at least one follower who is a member of *G* with a high probability; therefore, the mean infection rate rises as the number of followers decreases.

Figures 5 (Left) and 6(a) explain why the infection rate of *G* (0.435%) is smaller than the mean infection rate per user (3.017%). The mean infection rate per user depends on the range of the number of followers that most users in *G* have. From Figure 5 (Left), most users have 100-999 followers, so the mean infection rate per user should be close to the mean at 100-999 in Figure 6(a). In other words, if most users had 1000-9999 followers, the two infection rates would roughly coincide.

Another finding from Figure 6(a) is that the mean infection rate does not fall at a constant rate. Figure 6(b) illustrates the phenomenon more clearly. In the figure, the amount of decrease in the mean infection rate (shortly called decay) between 1-2 and 3-9 is greater than the decay between 31-99 and 100-315. In general, the mean infection rate falls more slowly as more users in *G* follow other users in *G*. In the case of Figure 6(b), the decay is small especially over 31-999. This inhomogeneity of the decay indicates that participants follow other participants who have 31-999 followers with a higher probability. We therefore conjecture that there would be communities consisting of members who mostly have 31-999 followers.

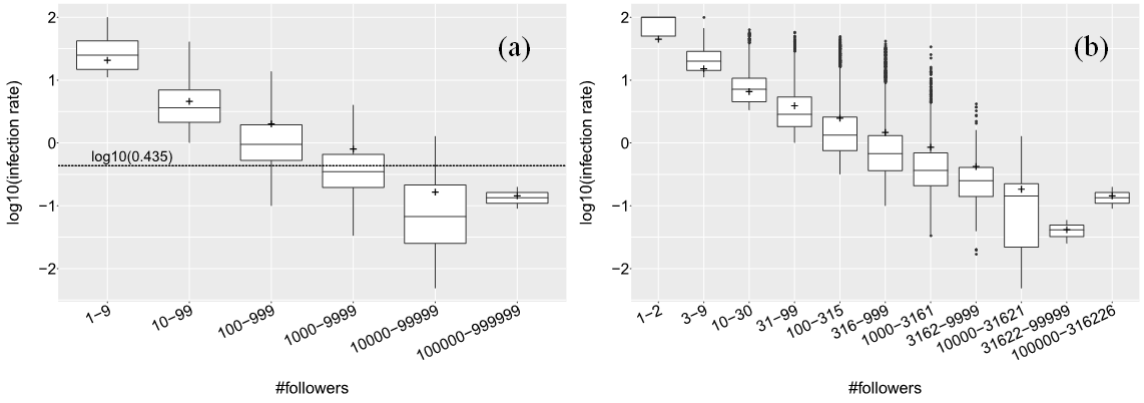

Figure 6. The box plots of the infection rate per user measured at many ranges of the number of followers per user. The range granularities are (a) 10 and (b) $10^{0.5}$. The mean infection rate (denoted by "+") decreases with the number of followers. The infection rate of *G* is 0.435%

5.3 Community Detection

Community detection is the division of the nodes into groups such that nodes in each group are densely connected among themselves. Currently, many community detection algorithms have been developed (Yang et al., 2016). To demonstrate that the inhomogeneity of the decay in Figure 6(b) is due to the existence of communities, we first detect communities in set *G* by using two state-of-the-art detection algorithms in programming language R: Spinglass (Reichardt & Bornholdt, 2006) and Walktrap (Pons & Latapy, 2005).

Table 1. The five largest non-overlapping communities detected by the two detection algorithms.

	Spinglass		Walktrap	
	#nodes	Mean #followers	#nodes	Mean #followers
1	1606	423	1775	445
2	1561	1319	682	2193
3	964	580	508	427
4	825	334	463	239
5	776	387	217	530

Table 1 shows the five largest non-overlapping communities in the followee-follower network of users in *G*. As shown in the table, the largest communities generated by the two algorithms are roughly the same. However, they output somewhat different results. Figure 7 exhibits the relationship between the histogram of the number of followers and the decay. The figure verifies the above-mentioned conjecture because it shows that the majority of community members have $10^{1.5}$-$10^{3.0}$ followers. It also clarifies the effect of communities: a greater number of followers of community members leads to a smaller decay. Note from Figure 7 that this effect is roughly independent of the detection algorithms and the number of communities.

The above result suggests that the decay could sense the "existence" of communities (although it could not derive communities from a graph). The community detection algorithms may output community data after spending a huge amount of time for a large network graph. (The computation times in our case were short (less than 13 minutes) because the size of *G* is small (|*G*|=16,747)). Meanwhile, the computation time to calculate the decay in Figure 6(b) is negligibly small because the horizontal axis is a log-scale of the number of followers. Thus, it provides the information on the existence of communities with a low computational cost. This outcome suggests that before directly extracting communities from a large graph, we should calculate the decay beforehand to confirm whether communities certainly exist or to search for the area in which communities surely exist.

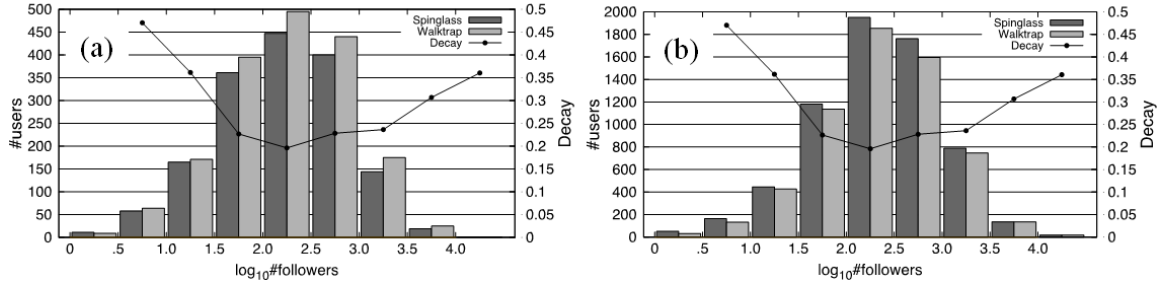

Figure 7. The histograms of the number of followers (a) for the largest community and (b) for the largest communities whose total number of nodes are 0.4|*G*|. A greater number of followers leads to a smaller decay

6. Cascade Structure

In order to understand the information cascade (the propagation structure of new adopters) of the campaign web page, we investigate from whom each participant received the web page URL for the first time. We infer the complete campaign cascade (shortly the original cascade) using a sample set (participant set *G*). Figure 8 exemplifies users in set *G* connected based on the followee-follower relationship. This participant network may not be an information cascade because a follower of a user may retweet the URL earlier than the user. (Note that in a cascade, a user must retweet under the influence of a retweet from another.) This phenomenon often occurs mainly due to the following two reasons. (1) The followee-follower relationship on Twitter highly dynamically changes (Myers & Leskovec, 2014). (2) The REST APIs sometimes did not answer our questions (see Appendix B).

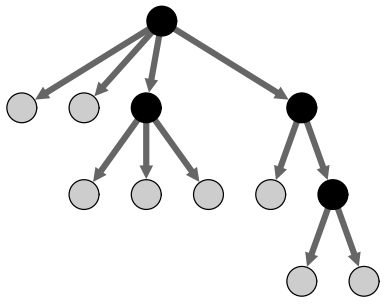

Figure 8. Users in *G* (not in *G*) correspond to black (gray) nodes and arrows indicate that (re)tweets flow from users to their followers. All black nodes are linked based on the followee-follower relationship

We first point out that if a user in *G* has two or more followees who are participants, we consider that the user is influenced by the earliest retweet. If it is difficult to find out by whom a user was influenced, we consider that a distinct cascade originates from the user. As a result of this, the participant network produces numbers of cascades and these are fractions of the original one. We use these fractions to estimate the structure of the original.

We derived 761 fractional cascades from the participant network. Figure 9(Left) shows the histogram of the number of nodes (users) per fraction. It can be seen from the figure that most cascades consist of less than 10 nodes and that a few cascades consist of more than 100 nodes. Figure 9(Right) exhibits the histogram of the Wiener index, which is expressed as the average distance of all node pairs in the cascade (Goel, Anderson, Hofman, & Watts, 2016) and the mathematical expression is in Appendix C. From Figure 9(Right), the maximum, minimum, and mean of the histogram are 4.8, 1.0, and 1.45, respectively.

In general, information cascades grow larger via combinations of the two structures: broadcast and viral (Goel et al., 2016), where broadcast indicates that a single node infects a large number of other nodes and viral indicates cascades grow through multiple generations with any one node infects only a few other nodes. The Wiener index can infer whether the cascade structure is more close to viral diffusion or not in such a way that a greater (smaller) Wiener index indicates the structure is more close to viral (broadcast) (Goel et al., 2016). Since all indexes are less than 4.8, we conjecture that the original cascade is close to the broadcast structure. Let us see some of large fractional cascades. As shown in Figure 10, large fractional cascades include broadcast nodes.

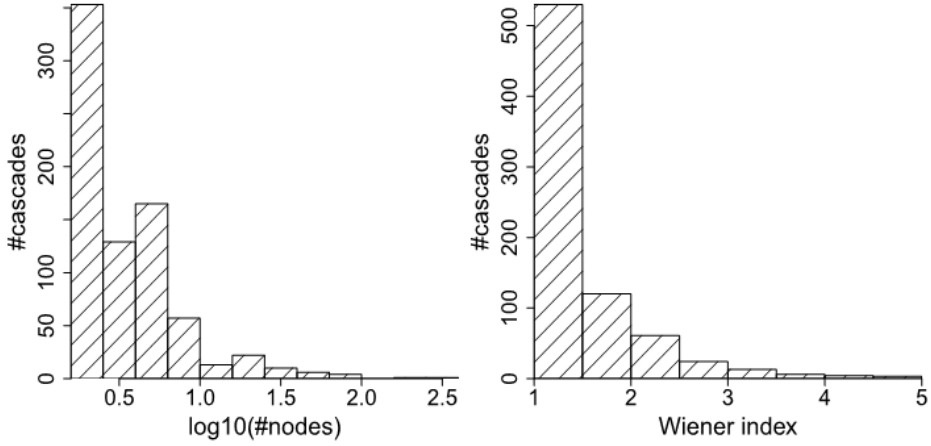

Figure 9. Left: the histogram of the number of nodes per fractional cascade. Right: the histogram of the Wiener index

7. Conclusions

The destructive Kumamoto earthquake occurred during the YOGUR STAND campaign period. The campaign web page popularity diminished by 60% just after the occurrence of the earthquake. However, after ten days, the popularity returned to the original level. The earthquake could not produce a long-lived popular content. In contrast, the campaign web page was one of the most popular contents on the Twitter network in Japan during the campaign period.

The campaign strategy is characterized by the reminder effect (i.e., a tweet made by the campaign account once a day reminds the account followers of the campaign) and the instant-win effect (i.e., a video notifying the lottery result is rapidly returned). Both effects incline a person to retweet many times. The campaign account constantly acquired about 14,000 participants and 2,000 new adopters every day. Furthermore, the yogurt product information propagated from the campaign account to about 2.4 million Twitter users every day and 0.35 million users of them received the information for the first time. According to the analysis made with sample participants, there were communities in the participant network and the propagation structure of the campaign information belonged to the broadcast type.

To propose more stable and successful advertising strategies on online social networks, our future work is to elucidate the mechanism that yields this long-lived diffusion phenomenon through computer simulations and data analysis. We would especially like to focus on how Twitter communities participated in spreading the web page over the social network.

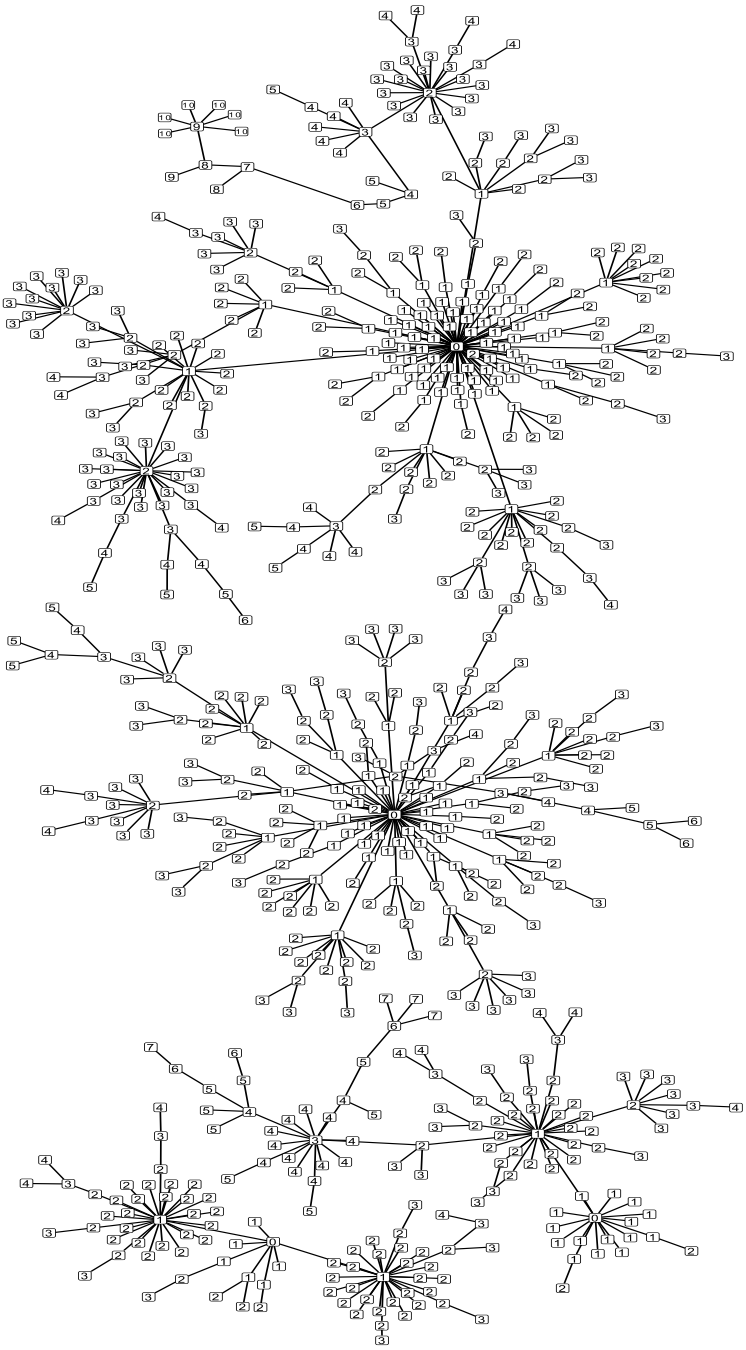

Figure 10. The largest four fractional cascades. The number in each node indicates the number of hops from the original adopter

References

Adamic, L. A., Lento, T. M., Adar, E., & Ng, P. C. (2016). Information evolution in social networks. In *Proceedings of the ninth ACM international conference on web search and data mining* (pp. 473–482). New York, NY, USA: ACM. https://doi.org/10.1145/2835776.2835827

Bakshy, E., Rosenn, I., Marlow, C., & Adamic, L. (2012). The role of social networks in information diffusion. In *Proceedings of the 21st international conference on World Wide Web* (pp. 519–528). New York, NY, USA: ACM. https://doi.org/10.1145/2187836.2187907

Chen, G., Chen, B. C., & Agarwal, D. (2017). Social incentive optimization in online social networks. In *Proceedings of the tenth ACM international conference on web search and data mining* (pp. 547–556). New

York, NY, USA: ACM. https://doi.org/10.1145/3018661.3018700

Cheng, J., Adamic, L. A., Dow, P. A., Kleinberg, J. M., & Leskovec, J. (2014). Can cascades be predicted? In *Proceedings of the 23rd international conference on World Wide Web* (pp. 925-936). https://doi.org/10.1145/2566486.2567997

Cheng, J., Adamic, L. A., Kleinberg, J. M., & Leskovec, J. (2016). Do cascades recur? In *Proceedings of the 25th International Conference on World Wide Web* (pp. 671-681). https://doi.org/10.1145/2872427.2882993

Cheung, M., She, J., Junus, A., & Cao, L. (2016, December). Prediction of virality timing using cascades in social media. *ACM Transactions on Multimedia Computing, Communications, and Applications (TOMM), 13* (1), 2:1–2:23. https://doi.org/10.1145/2978771

Chiu, S. I., & Hsu, K.-W. (2017). Information diffusion on facebook: A case study of the sunflower student movement in Taiwan. In *Proceedings of the 11th international conference on ubiquitous information management and communication* (pp. 48:1–48:8). New York, NY, USA: ACM. https://doi.org/10.1145/3022227.3022274

Cordasco, G., Gargano, L., Rescigno, A. A., & Vaccaro, U. (2016). Brief announcement: Active information spread in networks. In *Proceedings of the 2016 ACM symposium on principles of distributed computing* (pp. 435–437). NewYork, NY, USA: ACM. https://doi.org/10.1145/2933057.2933069

Dow, P. A., Adamic, L. A., & Friggeri, A. (2013). The anatomy of large facebook cascades. In E. Kiciman, N. B. Ellison, B. Hogan, P. Resnick, & I. Soboroff (Eds.), *ICWSM*. The AAAI Press.

Elsharkawy, S., Hassan, G., Nabhan, T., & Roushdy, M. (2016). Towards feature selection for cascade growth prediction on twitter. In *Proceedings of the 10th international conference on informatics and systems* (pp. 166–172). NewYork, NY, USA: ACM. https://doi.org/10.1145/2908446.2908463

Forestier, M., Bergier, J.-Y., Bouanan, Y., Ribault, J., Zacharewicz, G., Vallespir, B., & Faucher, C. (2015). Generating multidimensional social network to simulate the propagation of information. In *Proceedings of the 2015 IEEE/ACM international conference on advances in social networks analysis and mining 2015* (pp. 1324–1331). NewYork, NY, USA: ACM. https://doi.org/10.1145/2808797.2808870

Galuba, W., Aberer, K., Chakraborty, D., Despotovic, Z., & Kellerer, W. (2010). Outtweeting the twitterers - predicting information cascades in microblogs. In *Proceedings of the 3rd conference on online social networks* (pp. 3–3). Berkeley, CA, USA: USENIX Association.

Gao, S., Ma, J., & Chen, Z. (2015). Modeling and predicting retweeting dynamics on microblogging platforms. In *Proceedings of the 8th ACM international conference on web search and data mining* (pp. 107–116). New York, NY, USA: ACM. https://doi.org/10.1145/2684822.2685303

Goel, S., Anderson, A., Hofman, J., & Watts, D. J. (2016). The structural virality of online diffusion. *Management Science, 62* (1), 180-196. http://doi.org/10.1287/mnsc.2015.2158

Goel, S., Watts, D. J., & Goldstein, D. G. (2012). The structure of online diffusion networks. In *Proceedings of the 13th ACM conference on electronic commerce* (pp. 623–638). New York, NY, USA: ACM. https://doi.org/10.1145/2229012.2229058

Guille, A., Hacid, H., Favre, C., & Zighed, D. A. (2013, July). Information diffusion in online social networks: A survey. *SIGMOD Rec., 42* (2), 17–28. https://doi.org/10.1145/2503792.2503797

Guo, R., Shaabani, E., Bhatnagar, A., & Shakarian, P. (2015). Toward order-of-magnitude cascade prediction. In *Proceedings of the 2015 IEEE/ACM international conference on advances in social networks analysis and mining 2015* (pp. 1610–1613). NewYork, NY, USA: ACM. https://doi.org/10.1145/2808797.2809358

Hoang, T.-A., & Lim, E.-P. (2016). Tracking virality and susceptibility in social media. In *Proceedings of the 25th ACM international on conference on information and knowledge management* (pp. 1059–1068). New York, NY, USA: ACM. https://doi.org/10.1145/2983323.2983800

Hung, H.-J., Shuai, H. H., Yang, D. N., Huang, L. H., Lee, W. C., Pei, J., & Chen, M. S. (2016). When social influence meets item inference. In *Proceedings of the 22nd ACM SIGKDD international conference on knowledge discovery and data mining* (pp. 915-924). New York, NY, USA: ACM. https://doi.org/10.1145/2939672.2939758

Khan, A. (2016). Towards time-discounted influence maximization. In *Proceedings of the 25th ACM international on conference on information and knowledge management* (pp. 1873–1876). New York, NY, USA: ACM. https://doi.org/10.1145/2983323.2983862

Klopper, H. (2002). Viral marketing: a powerful, but dangerous marketing tool. *SA journal of information management, 4* (2). https://doi.org/10.4102/sajim.v4i2.159

Krishnan, S., Butler, P., Tandon, R., Leskovec, J., & Ramakrishnan, N. (2016). Seeing the forest for the trees: new approaches to forecasting cascades. In *Proceedings of the 8th ACM conference on web science* (pp. 249–258). New York, NY, USA: ACM. https://doi.org/10.1145/2908131.2908155

Krishnaswamy, D., Krishnan, R., Lopez, D., Willis, P., & Qamar, A. (2015, Jan). An open NFV and cloud architectural framework for managing application virality behaviour. In *2015 12th annual IEEE consumer communications and networking conference (CCNC)* (pp. 746-754). https://doi.org/10.1109/CCNC.2015.7158071

Lamba, H., & Pfeffer, J. (2016). Maximizing the spread of positive influence by deadline. In *Proceedings of the 25th international conference companion on World Wide Web* (pp. 67–68). https://doi.org/10.1145/2872518.2889412

Mishra, S., Rizoiu, M., & Xie, L. (2016). Feature driven and point process approaches for popularity prediction. In *Proceedings of the 25th ACM International on Conference on Information and Knowledge Management* (pp. 1069-1078). http://doi.org/10.1145/2983323.2983812

Myers, S. A., & Leskovec, J. (2014). The bursty dynamics of the twitter information network. In *Proceedings of the 23rd international conference on World Wide Web* (pp. 913–924). New York, NY, USA: ACM. https://doi.org/10.1145/2566486.2568043

Pons, P., & Latapy, M. (2005, December). Computing communities in large networks using random walks (long version). *ArXiv Physics e-prints*.

Reichardt, J., & Bornholdt, S. (2006, July). Statistical mechanics of community detection. *74* (1) https://doi.org/10.1103/PhysRevE.74.016110

Robles, J. F., Chica, M., & Cordn, O. (2016, July). Incorporating awareness and genetic-based viral marketing strategies to a consumer behavior model. In *2016 IEEE congress on evolutionary computation (CEC)* (pp. 5178-5185). https://doi.org/10.1109/CEC.2016.7748346

Rong, Y., Zhu, Q., & Cheng, H. (2016). A model-free approach to infer the diffusion network from event cascade. In *Proceedings of the 25th ACM international on conference on information and knowledge management* (pp. 1653–1662). New York, NY, USA: ACM. https://doi.org/10.1145/2983323.2983718

Weng, L., Menczer, F., & Ahn, Y. (2013). Virality prediction and community structure in social networks. *Scientific Reports 3, 2522*. https://doi.org/10.1038/srep02522

Yang, Z., Algesheimer, R., & Tessone, C. J. (2016). A comparative analysis of community detection algorithms on artificial networks. *Scientific Reports, 6*, 30750. https://doi.org/10.1038/srep30750

Zhao, Q., Erdogdu, M. A., He, H. Y., Rajaraman, A., & Leskovec, J. (2015). Seismic: A selfexciting point process model for predicting tweet popularity. In *Proceedings of the 21th ACM SIGKDD international conference on knowledge discovery and data mining* (pp. 1513–1522). New York, NY, USA: ACM. https://doi.org/10.1145/2783258.2783401

Zhu, W.-Y., Peng, W.-C., Chen, L.-J., Zheng, K., & Zhou, X. (2016, November). Exploiting viral marketing for location promotion in location-based social networks. *ACM Trans. Knowl. Discov. Data, 11* (2), 25:1–25:28. https://doi.org/10.1145/3001938

Appendix A

Twitter APIs

The Streaming APIs offer several streaming endpoints. We use GET statuses/sample to collect a small random sample of all public (re)tweets. Figure A1 shows our PHP program that gathers only (re)tweet messages through the APIs. In the figure, $status indicates data obtained through the Streaming APIs. The data is transformed to the JSON format to check whether it is a (re)tweet or not. If the data is a (re)tweet, it is appended to file ./json/data.json.

```
<?php
class SampleConsumer extends OauthPhirehose
{
   public function enqueueStatus($status)
   {
      $data = json_decode($status, true);
            if (is_array($data) && isset($data['user']['screen_name'])){
      $exfile = './json/data.json';
      file_put_contents($exfile,$status."\n", FILE_APPEND);
      }
   }
}
```

Figure A1. A PHP program that collects arriving public tweet messages

Meanwhile, the REST APIs provide programmatic access to read and write Twitter data. The Search API in the REST APIs returns tweet messages that match our query. The Search API searches against tweets published in the past seven days. Figure A2 exemplifies how we make queries. In the figure, 'q' possesses the search conditions, which specify that messages should include the URL and should be sent during the specified period. Variables 'lang,' 'count,' and 'include_entities' are used to determine language, the number of tweets returned per query, and whether the entities are included or not, respectively. The entities provide structured data including URLs, media, hashtags, etc. Received data are appended to file ./json/data.json. If there are more than 100 tweets, $tweets_arr['search_metadata']['next_results'] points the next results and they are requested at parse_str($next_results,$params). The search results are returned at the most 180 times per 15 minutes.

```
<?php
$params = array(
  'q'    => ' http://bit.ly/2dbfG2R since:2016-10-23 until:2016-10-24',
  'lang' => 'ja' ,
  'count' =>    100,
  'include_entities'=> 'true',
);
$request_number = 90000000;
$tweet_texts = array();
for ($i = 0; $i < $request_number; $i++) {
      $tweets_obj = $connection->get('search/tweets', $params);
        file_put_contents("./json/data.json",$tweets_obj, FILE_APPEND);
      $tweets_arr = json_decode($tweets_obj, true);
      for ($j = 0; $j < count($tweets_arr['statuses']); $j++) {
            $tweet_texts[] = $tweets_arr['statuses'][$j]['text'];
      }
      $next_results   =   preg_replace('/^\?/', '',$tweets_arr['search_metadata']['next_results']);
      if (!$next_results) {
            break;
      }
      if ($i%179==0){
            sleep(901);
      }
      parse_str($next_results, $params);
}
```

Figure A2. A PHP program that creates queries

Appendix B

Error Responses

We could not obtain follower IDs for 1,599 users in *G* because of the two error reasons. They are "Sorry, that page does not exist," which appeared 1,035 times, and "Not authorized," which occurred 564 times.

Appendix C

Wiener Index

The Wiener index *v* of a cascade is the average distance of all node pairs in the cascade and can be written as:

$$v = \frac{1}{n(n-1)} \sum_{i=1}^{n} \sum_{j=1}^{n} d_{ij} \quad (1)$$

where $n(>1)$ denotes the number of nodes in the cascade and d_{ij} is the shortest hop count from node *i* to node *j*.