Analyzing Patterns of Information Cascades based on Users' Influence and Posting Behaviors

Geerajit Rattanaritnont
The University of Tokyo
4-6-1 Komaba, Meguro-ku
Tokyo, Japan 153-8505
aomi@tkl.iis.u-tokyo.ac.jp

Masashi Toyoda
The University of Tokyo
4-6-1 Komaba, Meguro-ku
Tokyo, Japan 153-8505
toyoda@tkl.iis.u-tokyo.ac.jp

Masaru Kitsuregawa
The University of Tokyo
4-6-1 Komaba, Meguro-ku
Tokyo, Japan 153-8505
kitsure@tkl.iis.u-tokyo.ac.jp

ABSTRACT

Nowadays people can share useful information on social networking sites such as Facebook and Twitter. The information is spread over the networks when it is forwarded or copied repeatedly from friends to friends. This phenomenon is so called "information cascade", and has been studied long time since it sometimes has an impact on the real world. Various social activities tends to have different ways of cascade on the social networks. Our focus in this study is on characterizing the cascade patterns according to users' influence and posting behaviors in various topics. The cascade patterns could be useful for various organizations to consider the strategy of public relations activities. We explore four measures which are cascade ratio, tweet ratio, time of tweet, and exposure curve. Our results show that hashtags in different topics have different cascade patterns in term of these measures. However, some hashtags even in the same topic have different cascade patterns. We discover that such kind of hidden relationship between topics can be surprisingly revealed by using only our four measures rather than considering tweet contents. Finally, our results also show that cascade ratio and time of tweet are the most effective measures to distinguish cascade patterns in different topics.

Categories and Subject Descriptors

H.4 [Information Systems Applications]: Miscellaneous; J.4 [Computer Applications]: Social and Behavioral Sciences—Sociology

General Terms

Experimentation, Measurement

Keywords

Information cascade, Information diffusion, Social network, Microblog, Twitter

TempWeb '12, April 16-17 2012, Lyon, France

1. INTRODUCTION

Nowadays people can share useful information on social networking sites such as Facebook and Twitter. The information is spread over the networks when it is forwarded or copied repeatedly from friends to friends. This phenomenon is so called "information cascade", and has been studied long time [1, 2, 3, 6, 7, 9, 10, 11, 12, 14, 15, 16, 17, 18] since it sometimes has an impact on the real world.

Various social activities tends to have different ways of cascade on the social networks. For example, just after the Great East Japan Earthquake in 2011, Japanese Twitter users performed an energy saving activity so called "Operation Yashima". Since its name was taken from a famous Japanese animation program, this activity was quickly and widely spread over Twitter in such emergency situation. On the other hand, Fukushima Daiichi Nuclear Power Plant faced failures according to the earthquake, and it caused a lot of serious problems that cannot be solved immediately. Then these problems were continually talked and discussed for a long time involving experts.

Activities in the social networks seems to have typical patterns of information cascades. Our focus in this study is on characterizing the cascade patterns according to users' influence and posting behaviors in various topics. The cascade patterns could be useful for examining how activities affect people in the social networks, and how they are similar to the past typical activities. Such knowledge is important for various organizations to consider the strategy of public relations activities learning from past lessons.

In this paper, we conduct a research on Twitter to understand patterns of information cascade and behaviors of participating users in various topics such as earthquake and political topics. We investigate whether different topics have different cascade patterns or not by exploring four measures, which are cascade ratio, tweet ratio, time of tweet, and exposure curve. The cascade ratio determines how much people can influence their friends, the tweet ratio determines how much people talk in each topic, the time of tweet determines how long a topic is still popular in the network, and lastly the exposure curve determines how easy people are influenced by their friends. We consider Twitter hashtags as representatives of topics and perform experiments on a real Twitter dataset.

Table 1: Examples of hashtags in each topic

Topic	Total	Examples
Earthquake	54	jishin, genpatsu, prayforjapan, save_fukushima, save_miyagi, nicojishin, 84ma
Media	49	nicovideo, nhk, news, fujitv, cnn, aljazeera, r_blog
Politics	102	bahrain, iranelection, wiunion, teaparty, gaddafi, humanrights, weinergate
Entertainment	85	madoka_magica, akb48, atakowa, tigerbunny, anohana, beiberfact, jwave
Sports	20	hanshin, f1jp, dragons, sbhawks, cwc2011, canucks, soccer
Idiom	41	nowplaying, shoutout, followme, justsaying, pickone, followfriday, whatif

The Twitter dataset used in this paper is crawled from March 11, 2011 to July 11, 2011. It consists of 260 thousand users and 783 million tweets. We select top 500 frequently used hashtags from the dataset and categorize them according to topics. We firstly study the pattern of hashtag cascades in each topic by using statistical approach. We then further analyze the relationship between cascade patterns and topics by using clustering algorithm. Our results show that hashtags in different topics have different cascade patterns in term of our four measures. For example, the earthquake topic has low cascade ratio, low tweet ratio, short lifespan, and high persistence, while the political topic has high cascade ratio and high persistence. However, some hashtags even in the same topic have different cascade patterns. For instance, the earthquake hashtags can be divided into the hashtags directly related to the Great East Japan Earthquake, the media-related hashtags, and the political-related hashtags or the hashtags about the nuclear power plant. We discover that such kind of hidden relationship between topics can be surprisingly revealed by using only four measures rather than considering tweet contents. Finally, among four measures we explored, our results also show that cascade ratio and time of tweet are the most effective measures to distinguish cascade patterns in different topics.

The rest of this paper is organized as follows. Section 2 introduces related work on information diffusion in online blogging and social networking services. Section 3 explains the dataset. In Section 4, we describe four measures of users' influence and posting behaviors, and investigate the characteristics of information diffusion over six major topics. Then we conduct further analysis by using clustering algorithm in Section 5. Finally, we conclude this paper and future work in Section 6.

2. RELATED WORK

Information diffusion in online blogging services has been studied for a decade [6, 1, 11, 10]. Gruhl et al. [6] studied the dynamics of information propagation in weblogs. They investigated characteristics of long-running topics due to outside world events or within the community. Adar et al. [1] developed a tool to visualize the flow of individual URLs over a blog network. Leskovec et al. [11] also studied information propagation in weblogs. The proposed models that simulate the spread of information in blogspace and verified them in the real datasets.

Instead of blogsphere, researchers are also interested in information diffusion on other networks especially upcoming social networks [12, 16, 18, 9, 7, 14, 3]. Liben-Nowell et al. [12] traced the spread of information at individual level and found that information reach people in a narrow deep pattern, continuing for several hundred steps. Similarly, Sun et al. [16] conducted an analysis on information diffusion in Facebook and discovered that large cascade begins with a substantial number of users who initiate short chains.

In most recent years, as Twitter becomes one of the most popular micro-blogging services and allows us to obtain its data via Twitter API, it gains much interest from many researchers [4, 5, 8, 13, 15, 2, 17, 19, 20]. Romero et al. [15] studied information spread in Twitter and showed that controversial political topics are particularly persistent with repeated exposures comparing to other topics. Moreover, rather than understanding how information itself is spread, Bakshy et al. [2] exploited information cascade to identify influencers in Twitter. Scellato et al. [17] also extracted geographic information from information dissemination process and utilized it to improve caching of mulitimedia files in a Content Delivery Network.

Although various measures are studied to explain the patterns of information cascade, there are possibly more standard measures to distinguish them in different topics, for instance, earthquake and political topics. Besides, it is still unclear which measure are the most effective. We thus explore four simple measures, which are cascade ratio, tweet ratio, time of tweet, and exposure curve, to express the cascade patterns and finally verify the effective of each measure in our experiments.

3. TWITTER DATASET

We crawled the Twitter dataset from Twitter API from March 11, 2011 when the Great East Japan Earthquake took place to July 11, 2011. Our data collection consists of user profiles, timestamp and tweet contents including retweets. We started crawling from famous Japanese users. We firstly got timelines of these users, then repeatedly expanded the set of users by tracing retweets and mentions in their timelines. We then obtained 260 million users as active users and 783 million tweets. Instead of friend-follower relationships, we consider interactions such as mentions and retweets among users because they are stronger than friend-follower relationships. When a user A has at least one retweet from a user B or A has at least one mention to B, A has a directed link to B. In this case, we call B as a outgoing neighborhood of A. We extracted 31 million links by considering only active users.

To study information cascade according to different topics, we treat a hashtag as the representative of a topic. We select top 500 frequently used hashtags from the dataset and manually categorize them according to topics. Then we examine how the use of hashtags spread over the user

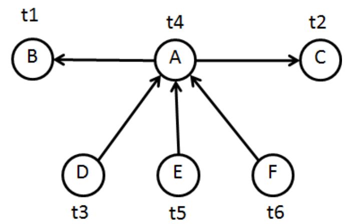

Figure 1: An example of hashtag cascade

interaction network. Moreover, to provide meaningful distributions in the rest of this study, we focus only on hashtags that have at least 1,000 participating users. We found that the majority belong to six major topics, which are earthquake, politics, media, entertainment, sports, and idiom. The number of tweets containing one of our interested hashtags are in range between 20 thousand to 1 million. Table1 shows examples of hashtags in each topic. The earthquake topic is mainly about the Great East Japan Earthquake, e.g., "jishin" (earthquake) and "genpatsu" (nuclear power plant). The political topic is related to political issues and events all over the world. Many of them refer to the uprising events in the Middle East, e.g., "bahrain" and "iranelection". The media topic is represented by communication channels including television networks, news channels, and video sharing websites, e.g, "nhk" and "cnn". The entertainment topic refers to television programs, movies and artists especially Japanese animations, e.g., "madoka_magica" and "tigerbunny". The sports topic corresponds to sports teams and tournaments. Most of them are Japanese baseball teams, e.g., "hanshin" and "dragons". Finally, the idiom topic is a popular phrase used as Twitter culture, e.g., "nowplaying" and "followme" . Although it is still unclear that the idiom topic should be really treated as the topic or not, we include this in our work because it was studied by Romero et al. [15].

4. MEASURES OF USERS' INFLUENCE AND POSTING BEHAVIORS

4.1 Cascade Ratio

The cascade ratio measures how much a user can influence his/her friends. We consider that a user A directly influenced a user B with respect to a given hashtag h, if B has a link to A, and B's first post of h followed A's post of h. It can be implied that B observed A's post and decided to post the same hashtag as A. The cascade ratio of a user u with respect to h is then defined as the fraction of users influenced by u within the all users who posted h:

$$cr(u,h) = \frac{C(u,h)}{U(h)} \tag{1}$$

where $C(u,h)$ is the number of users who linked to u and posted h after u, and $U(h)$ is a number of all users who posted h.

In Fig.1, there are 6 users who posted a hashtag at timings $t1$ to $t6$. The user A influenced the user D and E, and A's cascade ratio is $2/6$. Moreover, the user A followed the user B and C, and posted the hashtag after them. In this case, we consider both B and C influenced A, and B's cascade ratio is $1/6$ as same as C's cascade ratio.

Fig.2 shows point-wise average cascade ratio distributions. x is cascade ratio and y is the number of occurrences of cascade ratios normalized by total number of users using a given hashtag. The plot is in log-log coordinate and calculated as a cumulative distribution function, where y or $P(x)$ is the probability at a value greater than or equal to x. The red line is the point-wise average distribution of a particular topic, the blue line is the point-wise average distribution of all hashtags, and the green line is 90% confidence interval. In addition to the point-wise average distributions, we calculate the 90% bootstrap confidence intervals to test a null hypothesis. Our null hypothesis is that the particular topic has no difference in cascade ratio from a set of all hashtags. If 90% confidence interval do not contain average distribution of a topic, we can reject the null hypothesis and conclude by 90% confidence level that the topic has statistically significant difference in cascade ratio from the population. Otherwise, we cannot conclude by 90% confidence level that the topic has no difference in cascade ratio from the population.

According to Fig.2, The earthquake, media, sports, and idiom topics have relatively low cascade ratio. People participating in these topics used hashtags independently not because of seeing from their friends' tweets. On the contrary, the political topic has relatively high cascade ratio. When people posted political hashtags, many of their friends started to post the same hashtags after them.

4.2 Tweet Ratio

Tweet ratio shows how much people talk about a topic. It is the proportion of how many times a user uses a hashtag comparing to all tweets of the same hashtag. The tweet ratio tr of a user u posting a hashtag h is then simply defined as below:

$$tr(u,h) = \frac{T(u,h)}{\sum_u T(u,h)} \tag{2}$$

where $T(u,h)$ is the number of tweets containing the hashtag h posted by the user u.

Fig.3 illustrates point-wise average tweet ratio distributions. x is tweet ratio and y is the number of occurrences of tweet ratios normalized by total number of users using a given hashtag. Each line is plotted in log-log coordinate and calculated as a cumulative distribution function, where y or $P(x)$ is the probability at a value greater than or equal to x. The red line is the point-wise average distribution of a particular topic, the blue line is the point-wise average distribution of all hashtags, and the green line is the 90% confidence interval.

The earthquake, media, and idiom topics have relatively low tweet ratio. People in these topics repeated to use same hashtags very few times. On the other hand, the political topic has relatively high tweet ratio. People repetitively posted same hashtags about the political topic many times.

4.3 Time of Tweet

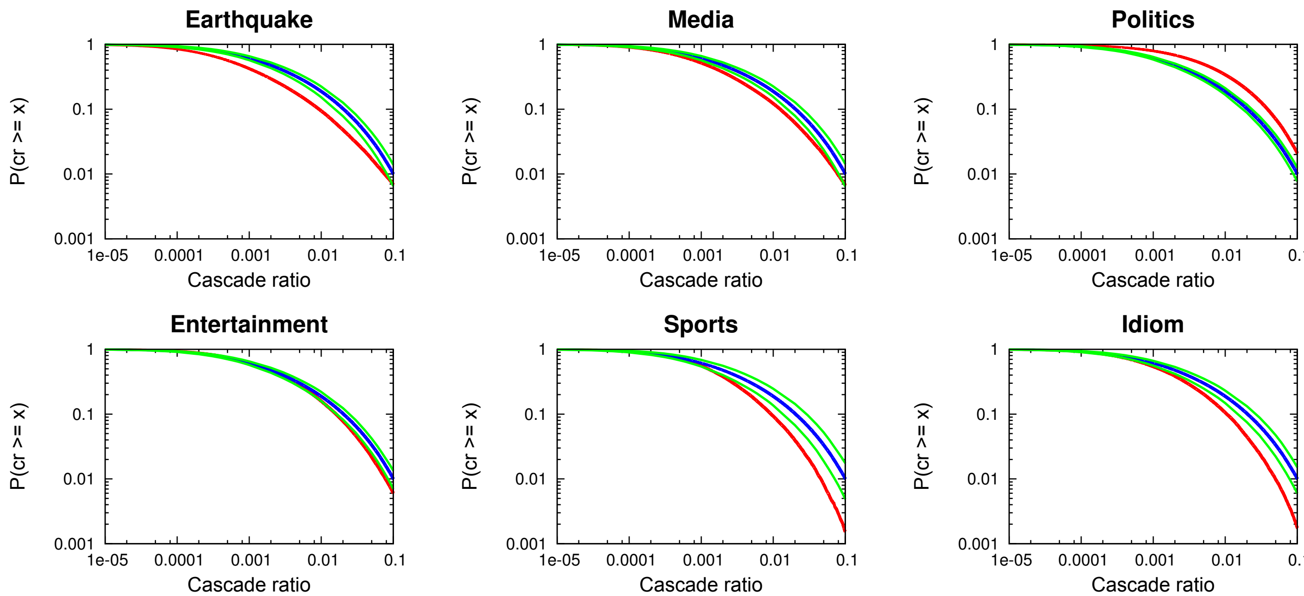

Figure 2: Point-wise average cascade ratio distributions of each topic

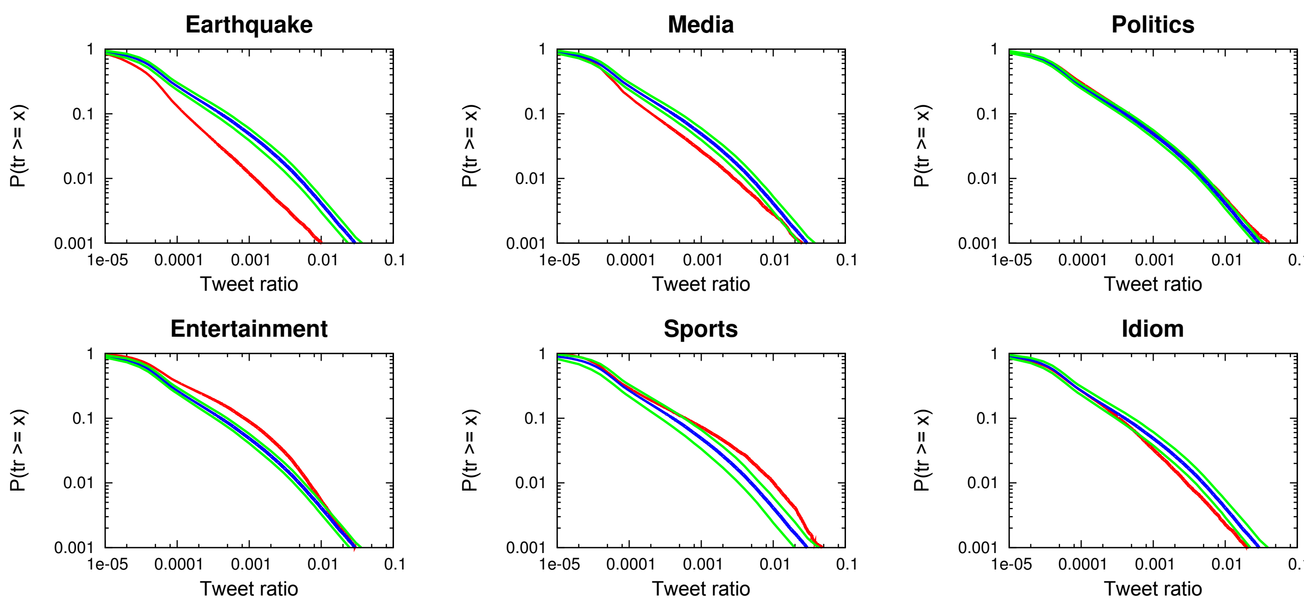

Figure 3: Point-wise average tweet ratio distributions of each topic

Time of tweet demonstrates how long a topic is popular in the network. It is time of each usage of a hashtag from its first appearance. The time ti of a tweet tw containing a hashtag h is then straightforwardly defined as the difference in time between tw and the first tweet of h.

Fig.4 shows point-wise average time distributions. x is time of tweet in hour(s) and y is the number of occurrences of time normalized by total number of tweets comprising a given hashtag. Each line is plotted as a cumulative distribution function, where y or $P(x)$ is the probability at a value greater than or equal to x. The red line is the point-wise average distribution of a particular topic, the blue line is the point-wise average distribution of all hashtags, and the green line is the 90% confidence interval.

The earthquake topic falls down at first period. A large number of tweets were posted soon after the topics were raised to Twitter and gradually decreased when time passed. We can imply that people talked very much about the Great East Japan Earthquake during that time and in turn rarely said about it when the situation was back to normal. Conversely, the entertainment and sports topics lay in a diagonal. The number of tweets did not change according to time. People continually talked about these topics during the period of time.

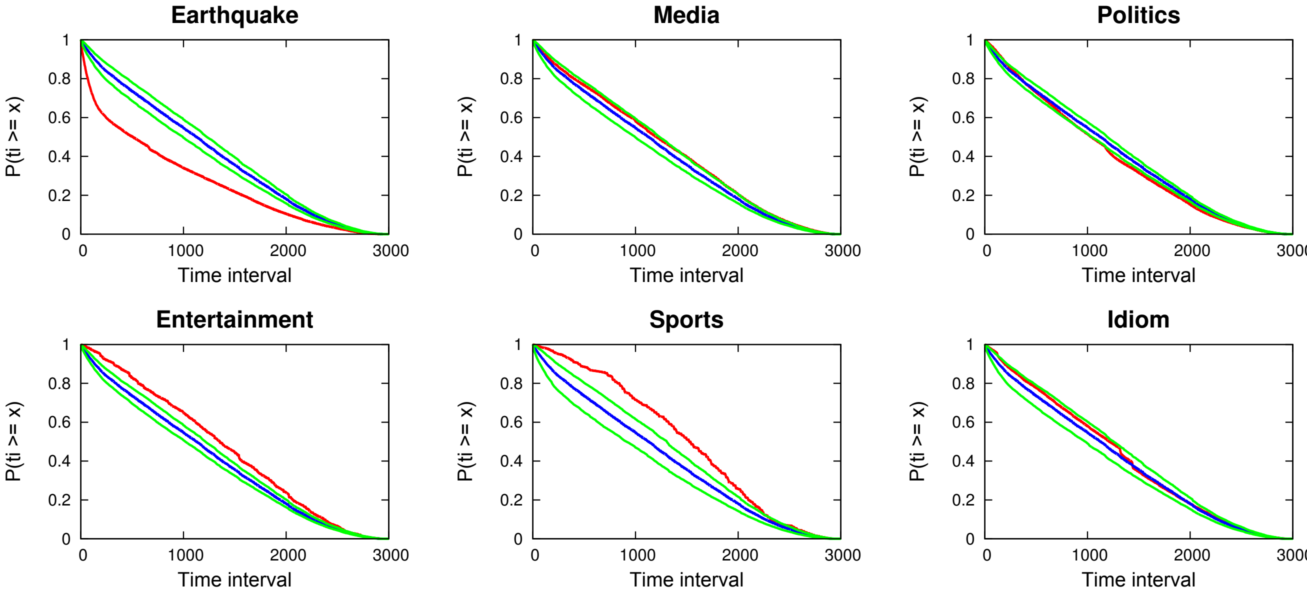

Figure 4: Point-wise average time distributions of each topic

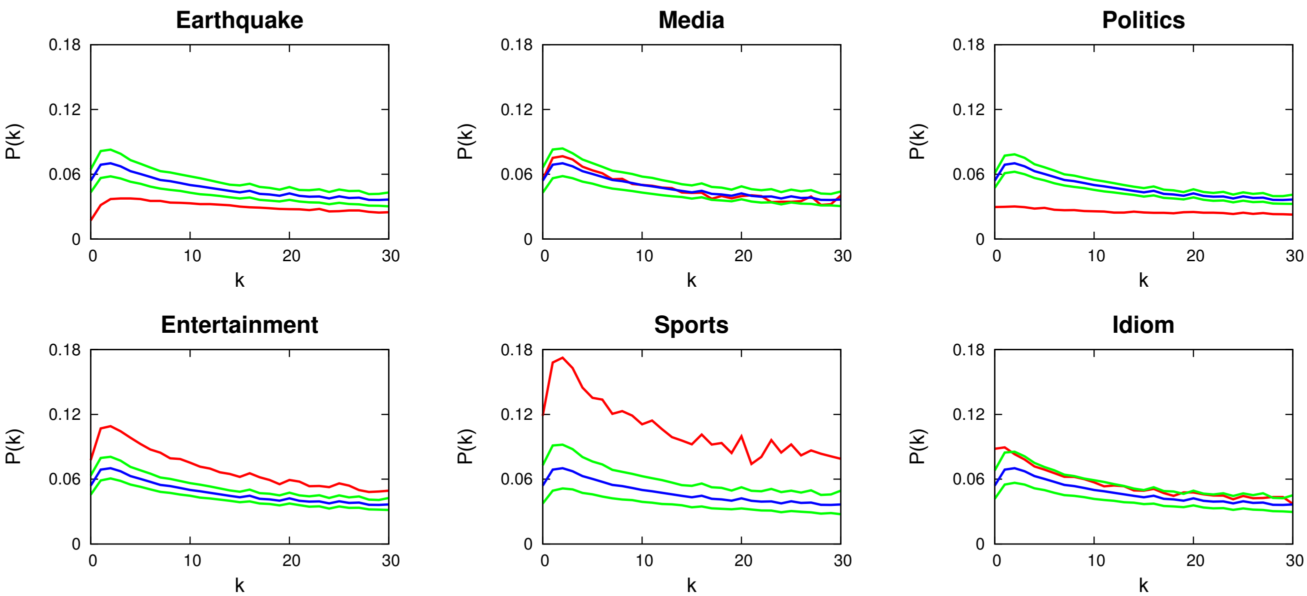

Figure 5: Point-wise average exposure curves of each topic

4.4 Exposure Curve

The last measure is exposure curve proposed by Romero et al. [15]. It determines how easy people are influenced by their friends. The exposure curve $P(k)$ is defined as below:

$$P(k) = \frac{I(k)}{E(k)} \tag{3}$$

where $I(k)$ is the number of users who started to post the hashtag h right after their k outgoing neighborhoods and $E(k)$ is the number of users who have k outgoing neighborhoods posting the hashtag before them at some time.

Fig.5 depicts point-wise average exposure curves. x is k neighborhoods who used a hashtag before a user and y the probability $P(k)$ that a user u will use the given hashtag h right after his/her k friends. The red line is the point-wise average exposure curve of a particular topic, the blue line is the point-wise average exposure curve of all hashtags, and the green line is the 90% confidence interval.

The peaks of the curves, are at $k = 4$ for the earthquake topic and $k = 2$ for the entertainment and sports topics. That means the maximum probability that people will start to post a hashtag about the earthquake topic is when four neighborhoods used that hashtag before them as well as two neighborhoods in case of the entertainment and sports topics. Besides, since the political topic has no peak, we can say that the number of neighborhoods who used a given hashtag do not affect people participating in this topic to start to use the same hashtag. Nevertheless, we here focus on shape of the curve rather than identifying whether the curve is higher or lower than the average. The curve $P(k)$ of the earthquake and political topics do not change as k increases. These two topics are thus high persistent. In turn, the curve $P(k)$ of the entertainment and sports topics fall down rapidly after the peaks. The probability that a user will start to use a hashtag decreases as k increases. We can say that these two topics are low persistent.

4.5 Patterns of Topic-Sensitive Hashtag Cascades

By using cascade ratio, tweet ratio, time of tweet, and exposure curve, we summarize patterns of hashtag cascades according to six major topics as in Table 4. "H" means high, "L" means low, and - means No statistically significant difference from the population.

The earthquake topic has low cascade ratio, low tweet ratio, short lifespan, and high persistence. The media and idiom topics have same patterns, which are low cascade ratio and low tweet ratio. The political topic has high cascade ratio and high persistence. The entertainment and sports topics have similar patterns, which are high tweet ratio, long lifespan, and low persistence, and the sports topic additionally has low cascade ratio.

5. RELATIONSHIPS BETWEEN CASCADE PATTERNS AND TOPICS

In this section, we further investigate the relationship between cascade patterns and popular topics in Twitter and examine the effectiveness of each measure we described in earlier section. We perform k-means clustering based on the distributions of cascade ratio, tweet ratio, time of tweet, and exposure curve. Each hashtag is represented as a vector of values captured from n points in each distribution. For each hashtag, we select n=93 points proportional to the log scale.

We use Euclidean distance as a distance measure and randomly assign each hashtag to a cluster at initialization. Considering six major topics in our study, we vary the number of clusters as $k = 6, 7, 8$. Since k-means algorithm provides different results depending on the initialization, we perform five trials for each k and evaluate clustering results by using normalized mutual information (NMI). Instead of other evaluation measures such as purity and F measure, it can be used to compare clustering quality with different numbers of clusters. For each trial, we compute NMI to evaluate clustering results. We then pick up the trial that provides the highest NMI at each k. Since those results when $k = 6, 7, 8$ have the same trend, we then choose the result of $k = 6$ to consider throughout this study.

Additionally, we are able to investigate the effectiveness of each measure on the clustering results by using NMI. We perform clustering by relying on all of four measures, and leaving one measure out at each experiment. Fig.6 demonstrates the average NMI of five trials in each approach when $k = 6$. We can see that NMI decreases when cascade ratio

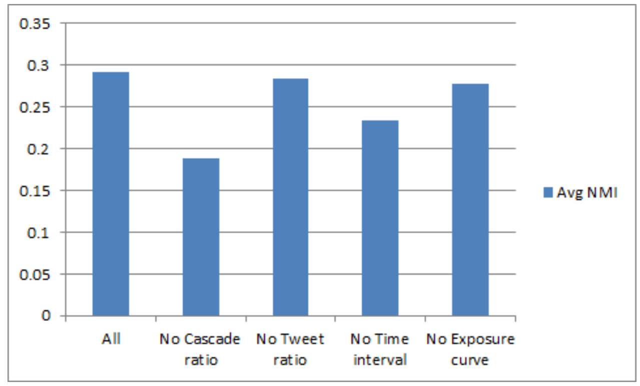

Figure 6: Average NMI of each approach when $k = 6$

Table 3: Clustering result when $k = 6$

No. of hashtags	c0	c1	c2	c3	c4	c5
Earthquake	25	9	1	5	8	0
Media	1	20	1	12	10	2
Politics	0	4	47	2	26	15
Entertainment	0	10	5	39	5	6
Sports	0	2	0	17	0	1
Idiom	1	16	1	7	10	0

or time of tweet are not used. Therefore, cascade ratio and time of tweet are said to be the most effective measures to characterize hashtag cascade, while tweet ratio and exposure curve even proposed in the existing work are not effective as we expect. According to Table 3, we can obtain the same result by using only cascade ratio and time of tweet.

Table 3 and Table 4 illustrate clustering result and cascade patterns of each cluster when $k = 6$ respectively. We can conclude that hashtags from the same topic or the topics having similar patterns of cascade are assigned into the same cluster. For example, the majority of the earthquake topic are assigned into cluster 0. Moreover, the cascade pattern of this cluster in Table 4 is the same as the pattern of the earthquake topic in Table 3. In the same way, because the media and sports topics have same cascade patterns, the majority of these two topics are put together into cluster 1.

However, some of them even from the same topic have different behaviors and thus put into other clusters. For example, the hashtags in the earthquake topic are mainly divided into cluster 0, 1, and 4. The hashtags in cluster 0 are directly related to the Great East Japan Earthquake such as "jishin", "save_miyagi", and "84ma" (Operation Yashima). On the other hand, the earthquake hashtags in cluster 1, which the majority of the media topic are assigned to, are hashtags such as "iwakamiyasumi" (a journalist who spread information about nuclear power plant after the accident at Fukushima Daiichi Nuclear Power Plant) and "nicojishin". We can see that they are somehow related to the media topic. Likewise, the earthquake hashtags in cluster 4, which its major members are the political topic, are hashtags such as "save_fukusima" and "cnic" (Citizen's Nuclear Information Center). Because they are about the nuclear power plant which needs the Japanese government to concern and

Table 2: Patterns of hashtag cascades in each topic

Topic	Cascade ratio	Tweet ratio	Time of tweet	Exposure curve
Earthquake	L	L	L	L
Media	L	L	-	-
Politics	H	-	-	L
Entertainment	-	H	H	H
Sports	L	H	H	H
Idiom	L	L	-	-

Table 4: Patterns of hashtag cascades in each cluster when $k = 6$

Cluster	Cascade ratio	Tweet ratio	Time of tweet	Exposure curve	Major Topics
Cluster 0	L	L	L	L	Earthquake
Cluster 1	L	L	-	-	Media, Idiom
Cluster 2	H	H	-	L	Politics
Cluster 3	L	H	H	H	Sports, Entertainment
Cluster 4	-	L	-	L	Media, Idiom
Cluster 5	H	H	L	L	Politics

take actions on, their cascade patterns are closely related to political topic.

In the same way as the media hashtags, they are primarily split into cluster 1, 3, and 4. The hashtags in cluster 1 are Japanese television media such as "fujitv", "nhk", and "tvasahi", while the media hashtags in cluster 3 are Japanese Internet media such as "r_blog" (Rakuten blog), "ameblo" (Ameba blog), and "2chmatome". Furthermore, the media hashtags in cluster 4, which its major members are again the political topic, are hashtags such as "aljazeera", "wikileaks", and "alarabiya". Since these kind of media mainly serve political news, their cascade patterns are closely related to political topic too.

Lastly, the entertainment and sports hashtags are largely assigned into the same cluster, cluster 3. The entertainment hashtags here are Japanese animations and artists such as "tigerbunny" and "akb48" respectively, while the sports hashtags are Japanese baseball teams such as "hanshin" and "dragons". It is probably that both of them are hobbies, gain much interest from their fans and thus share common behaviors.

Due to the above analysis, it is interesting that we can discover hidden relationship between topics by using only four measures rather than seeing tweet contents.

6. CONCLUSIONS

We studied the patterns of information cascade in six popular topics in Twitter, which are earthquake, media, politics, entertainment, sports, and idiom. We found that different topics mostly have different patterns of hashtag cascades in term of cascade ratio, tweet ratio, time of tweet, and exposure curve. For example, the earthquake topic has low cascade ratio, low tweet ratio, short lifespan, and high persistence, while the political topic has high cascade ratio and high persistence. However, some hashtags even in the same topic have different cascade patterns. For instance, the earthquake hashtags can be divided into the hashtags directly related to the Great East Japan Earthquake, the media-related hashtags, and the political-related hashtags or the hashtags about the nuclear power plant. We discover that such kind of hidden relationship between topics can be surprisingly revealed by using only four measures rather than considering tweet contents.

Finally, as future work, we need to explore other useful characteristics such as expert level of individual users and verify which measures are the most appropriate to explain patterns of hashtag cascades in different topics. Moreover, we need to investigate other clustering algorithms and other similarities whether they still provide the same results or not.

7. REFERENCES

[1] E. Adar, L.A. Adamic, Tracking Information Epidemics in Blogspace. In 2005 IEEE/WIC/ACM International Conference on Web Intelligence (WI2005), pp. 207–214, 2005.

[2] E. Bakshy, J.M. Hofman, W.A. Mason, D.J. Watts, Everyone's an Influencer: Quantifying Influence on Twitter. In 4th ACM International Conference on Web Search and Data Mining (WSDM2011), pp. 65–74, 2011.

[3] E. Bakshy, B. Karrer, L.A. Adamic, Social Influence and the Diffusion of User-Created Content. In 10th ACM Conference on Electronic Commerce (EC2009), pp. 325–334, 2009.

[4] C. Castillo, M. Mendoza, B. Poblete, Information Credibility on Twitter. In 20th International Conference on World Wide Web (WWW2011), pp. 675–684, 2011.

[5] M. Cha, H. Haddadi, F. Benevenuto, K.P. Gummadi, Measuring User Influence in Twitter: The Million Follower Fallacy. In 4th International AAAI Conference on Weblogs and Social Media (ICWSM2010), pp. 10–17, 2010.

[6] D. Gruhl, R. Guha, D. Liben-Nowell, A. Tomkins, Information Diffusion Through Blogspace. In 13th International Conference on World Wide Web (WWW2004), pp. 491–501, 2004.

[7] D. Kempe, J. Kleinberg, E. Tardos, Maximizing the Spread of Influence through a Social Network. In 9th ACM SIGKDD Knowledge Discovery and Data

Mining (KDD2003), pp. 137–146, 2003.
[8] H. Kwak, C. Lee, H. Park, S. Moon, What is Twitter, a Social Network or a News Media?. In 19th International Conference on World Wide Web (WWW2010), pp. 591–600, 2010.
[9] J. Leskovec, L.A. Adamic, B.A. Huberman, The Dynamics of Viral Marketing, ACM Transaction on the Web, 1(1):5, May 2007.
[10] J. Leskovec, L. Backstrom, J. Kleinberg, Meme-tracking and the Dynamics of the news Cycle. In 15th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD2009), pp. 497–506, 2009.
[11] J. Leskovec, M. McGlohon, C. Faloutsos, N. Glance, M. Hurst, Patterns of Cascading Behavior in Large Blog Graphs. In 7th SIAM International Conference on Data Mining (SDM2007), pp. 551–556, 2007.
[12] D. Liben-Nowell, J. Kleinberg, Tracing Information Flow on a Global Scale Using Internet Chain-Letter Data. In the National Academy of Sciences, 105(12):4633–4638, March 25, 2008.
[13] B. Meeder, B. Karrer, A. Sayedi, R. Ravi, C. Borgs, J. Chayes, We Know Who You Followed Last Summer: Inferring Social Link Creation Times in Twitter. In 20th International Conference on World Wide Web (WWW2011), pp. 517–526, 2011.
[14] M.E.J. Newman, S. Forrest, J. Balthrop, Email Networks and the Spread of Computer Viruses. Physical Review E, 66(035101), 2002.
[15] D.M. Romero, B. Meeder, J. Kleinberg, Differences in the Mechanics of Information Diffusion across Topics: Idioms, Political Hashtags, and Complex Contagion on Twitter. In 20th International Conference on World Wide Web (WWW2011), pp. 695–704, 2011.
[16] E. Sun, I. Rosenn, C. Marlow, T. Lento, Gesundheit! Modeling Contagion through Facebook News Feed. In 3rd International AAAI Conference on Weblogs and Social Media (ICWSM2009), pp. 146–153, 2009.
[17] S. Scellato, C. Mascolo, M. Musolesi, J. Crowcroft, Track Globally, Deliver Locally: Improving Content Delivery Networks by Tracking Geographic Social Cascades. In 20th International Conference on World Wide Web (WWW2011), pp. 457–466, 2011.
[18] D.J. Watts, A Simple Model of Global Cascades on Random Networks. In the National Academy of Sciences, 99(9):5766–5771, April 30, 2002.
[19] J. Weng, E.-P. Lim, J. Jiang, Q. He, TwitterRank: Finding Topic-Sensitive Influential Twitterers. In 3rd ACM International Conference on Web Search and Data Mining (WSDM2010), pp. 261–170, 2010.
[20] S. Wu, J.M. Hofman, W.A. Mason, D.J. Watts, Who Says What to Whom on Twitter. In 20th International Conference on World Wide Web (WWW2011), pp. 705–714, 2011.