Visualization of diffusion behavior pattern of influencers by genre on SNS

Chisae Iwashina
Ochanomizu University
Tokyo, Japan
chisae@itolab.is.ocha.ac.jp

Mitsuo Yoshida
Toyohashi University of Technology
Aichi, Japan
yoshida@cs.tut.ac.jp

Takayuki Itoh
Ochanomizu University
Tokyo, Japan
itot@is.ocha.ac.jp

***Abstract*—A large number of organizations and individuals use SNS (Social Networking Services) for the purpose of spreading accouncements nowadays. Among SNSs, Twitter is particularly well-used to spread accouncements and messages easily. Various commercial organizations including entertainers, fashion brands, and enterprises use Twitter to announce news and other information. Highlighted users who tweet actively on Twitter are called "influential users." Many users around influential users often spread their messages by the "retweet" function. This study visualizes the relationship between the tweets by the influential users and the surrounding user groups applying dendrogram and heatmap. From the visualization results, we found groups of users surrounding the influential users, numbers of retweets, and intervals between original tweets and their retweets depend on contens of tweets and aspects of surrounding users.**

***Index Terms*—SNS, diffusion, visualization**

I. Introduction

Social Networking Services (SNS) has developed as a casual communication and information gathering tool. It has also been used to diffuse information to both domestic and overseas. Twitter has been especially useful for many users all over the world to read, retrieve, send, and spread highly immediate information.

Many organizations (e.g., companies) and individuals spread information using Twitter thanks to such advantages.

In this study, we focused on "spread" of information on Twitter and visualized the behavior patterns of spread tweets and Twitter users who spread the information. This poster calls influential persons and advertising accounts on Twitter as "key persons", and a Twitter user who spreads the tweets of key persons as "retweeter". The poster shows examples visualizing the relationships between key persons and behavior patterns of retweeters, and discusses the characteristics of key persons.

II. Related work

A. Visualization of SNS data

There have been many studies on the visualization of communication behavior on SNS. Twitter is especially a convenient SNS because Twitter API makes us easy to collect various data such as attributes of users, friendships, contents of remarks, time and position when they tweet. Recently many studies and applications for visualization of SNS data have been presented. For example, it is possible to visualize the movement of people while natural disasters and analyze reactions of people on SNS before and after the natural disasters such as hurricane and tornado. Chae et al. [1] presented a single view display for simultaneously analyzing the spatial and temporal characteristics of tweets realizing a highly comprehensive visualization.

There have also been many studies using data extracted from Weibo, a famous Chinese microblog service [2] [3]. In these researchers, Weibo's important users, information diffusion paths, and interactions between communities are identified are visualized simultaneously.

B. Diffusion on SNS

We focus on the retweet function in this research, but there is a study recommending tweet's spreading partner focusing on the mentions function which is one of the features of Twitter. The mention function is what you see with the symbol "@ " and is a function used when sending information to a specific person. Wang et al. [4] use tweet spreading ability of the recommended user as a feature quantity by its own model, and finally defines the spreading factor and the spreading range of tweet so that it can be used as a user we propose a method to recommend.

III. Processing flow

This section presents the processing flow of our study as the following five steps. We selected key persons from various industries, collected their tweets, and visualized the diffusion of these retweets. We also examined how the characteristics of the retweeters differ depending on the key person.

A. Selection of key person

Key persons in this study correspond to "influential" users on Twitter. Although it is possible to quantify the influence on Twitter, our current study manually selects key persons according to our subjectivity. We select one key person who has an official account of Japan or Japanese (with or without the official mark) for each genre. We target users posting tweets at a certain time interval due to limitations of on data collection using Twitter API[1].

Japan Society of the Promotion of Science under Grant-in-Aid fir Scientific Research

[1]https://developer.twitter.com/en/docs/tweets/search/api-reference/get-search-tweets.html

B. Selection of tweets

We selected a certain number of tweets which are particularly important among past tweets by key persons. Although it is possible to quantify the importance of tweets, our current study selects tweets according to our subjectivity.

C. Collection of retweet data

This study collects retweets of the tweets selected by the previous step described in Section 3.2. We can search for retweets of an original tweet by appending "RT" to the text and requesting via Twitter Search API. Here, we need to set the date and time finely so that we can collect all the retweets in chronological order, due to a limitation that Twitter API can collect only 100 pieces of retweets per inquiry. Collected data items include Tweet id, Text, Created at (Retweet execution time), and User ID.

Note that it is impossible to restore the data spreading route from the retweet data that can be collected by the Twitter API like the Figure 1. For example, if the user A is a former tweet sender, the user B who directly browses the tweet and executes the retweet and the user C — D who browses the retweet of the user B and executes the retweet are considered. However, on the data that can be gathered by the Twitter API, in both of the retweets likewise, only the fact that the user B — C — D retweeted the utterance of the former tweet user A is not described. Therefore, in this research, retweeter is a user who retweeted the former tweet and does not consider the retweet from the retweet of anyone.

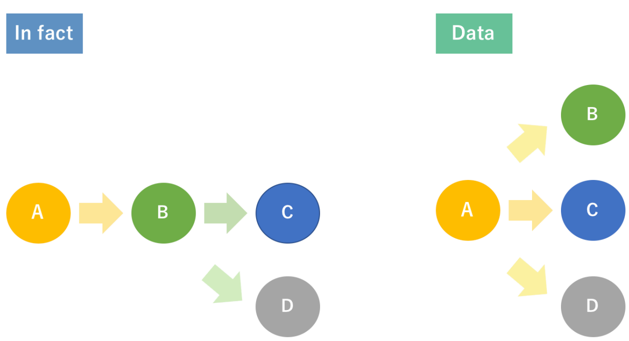

Fig. 1. Retweet data image diagram

D. Data extraction and Sampling of the repeaters

We extract User ID and Created at (the ID of the retweeter and the execution time of the retweet) from the retweet data previously collected as described in Section 3.3. This study calculates the difference between the origination time of a tweet and the retweet time. We perform these tasks for each selected tweet of each key person. If the same retweeter retweets the same tweet twice, we left only those with the smallest time difference from those retweets [2].

Next, we integrate the preselected tweets for each tweet for one key person and form a matrix which assigns tweets and retweeters to rows and columns. We call this matrix as "retweet matrix"(Figure 2).

	Tweet1	Tweet2	Tweet3	Tweet4	Tweet5
Retweeter1	2	2	0	0	0
Retweeter2	10	0	1	23	0
Retweeter3	3	0	0	0	10
	4	2	0	2	0

Fig. 2. Retweet matrix image diagram

Here, we calculate the value v_{ji} for the i-th row and the j-th column calculated from the time difference d_{ij} of the i-th tweet retweeted by the j-th retweeter as follows:

$$v_{ji} = \begin{cases} d_{max} - d_{ij} & (when\ there\ is\ a\ retweet\ execution) \\ -D & (when\ there\ is\ not\ a\ retweet\ execution) \end{cases}$$

Note that $-D$ is a negative constant (-1 in the current implementation), and d_{max} is the maximum value of the time difference of all retweets.

Finally, we may sample the retweeters if a too large number of retweeters are extracted.

E. Visualization (Selection of differences between retweeters)

Then, we construct two dendrograms consisting of original tweets and retweeters from the time difference d_{ij} of the i-th tweet retweeted by the j-th retweeter, and identify the orders of the original tweets and retweeters. This study visualizes the retweet matrix by a heat map with the dendrograms at once applying a Python library "seaborn" [3](Figure 3).

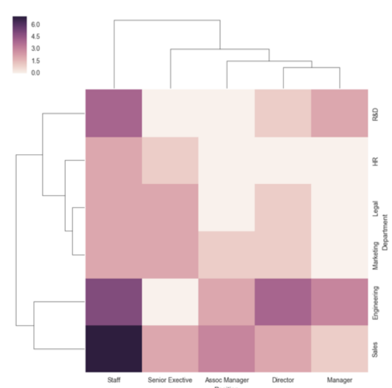

Fig. 3. Visualization image diagram

[2]It seems the retweet history is recorded every time the retweet button was pressed on the Web browsers or mobile applications since retweeting can be canceled by one click.

[3]http://seaborn.pydata.org

IV. Experiments

In this analysis, we select one key person from each of the following three genres.

- actor
- talent
- information distribution

We selected one key person from five genres (actor, talent, promotion account, sports player, and politician) respectively in this experiment. Then, we selected five tweets per key person manually, based on the number of retweets, the number of quoted retweets, and the presence or absence of images and moving pictures. We selected tweets less than one month before the data collection time due to the limitation of Twitter API.

To construct sampled retweeter datasets, we decided to extract retweeters who retweeted two, three, or four tweets from the selected five tweets of the key persons. If the number of selected retweeters exceeds 500, we randomly selected 500 as representative retweeters to be visualized.

This poster introduces and discusses the visualization results on the genres of actors. Here, we normalized the values in the retweet matrix by scaling to [0.0,1.0] as a pre-processing of visualization. Darker colors are assigned when the values are close to 1.0 corresponding to the small time differences. In other words, you can understand the speed difference of retweet reaction with the shade of color.

The retweeters are assigned to the vertical axis while the original tweets are assigned to the horizontal axis in the heat map. Their orders are specified from the orders in the dendrograms(Figure 4).

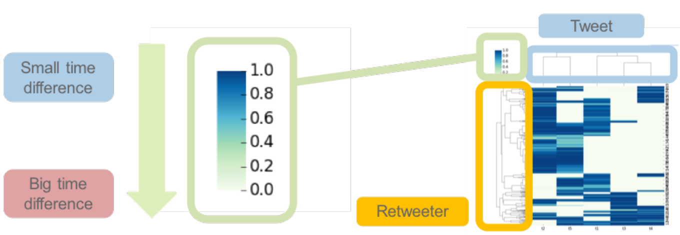

Fig. 4. How to read figures

A. Actor

TABLE I
CHARACTERISTICS OF ACTOR'S TWEETS.

no.	characteristic	retweeters(about)	tweet time(JST)
t1	image, other entertainers	5,000	3:36
t2	image, other entertainers multiple	13,000	21:19
t3	quoted retweet	2,000	22:41
t4	quoted retweet	3,000	19:17
t5	image, other entertainers multiple	12,000	23:32

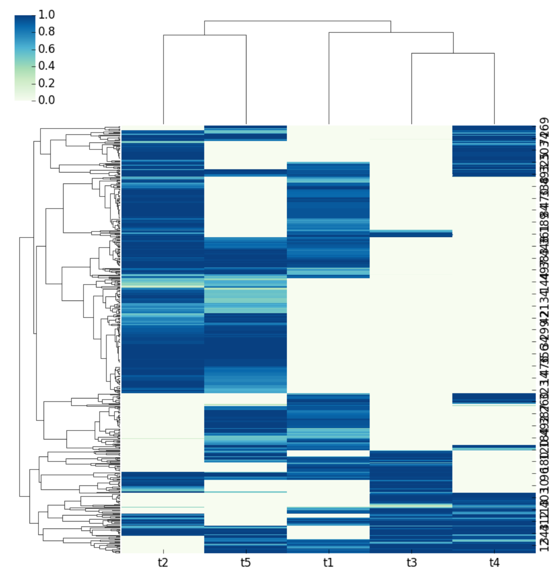

Fig. 5. Actor's retweeters. This key person has many enthusiastic fans who use Twitter mainly to cheer this key person.

Table I shows the statistics of the tweets. A large number of retweeters repeated retweets with all tweets. Also, a large number of retweeters reacted to all five tweets.

Figure 5 shows the visualization result. We found that retweeters who had a fast reaction to a tweet were also faster to react other tweets. At the same time, we could observe a blue belt continues beside along the horizontal axis. Looking at the visualization result along the vertical axis, t2 and t5 had similar retweets, and actually, some retweeters had equivalent time differences for t2 and t5. We found both t2 and t5 attached photos taking other entertainers in addition to the key person. It seems the retweeters surrounding the other celebrities in the photos responded in addition to the retweeters surrounding the key person, and the time difference was similar. Conversely, both t3 and t4 were quoted retweets, so it seems that few users responded as a whole. Meanwhile, we found that many retweeters had small time difference many reactions. We suppose this key person has many enthusiastic fans who use Twitter mainly to cheer this key person.

B. Talent

TABLE II
CHARACTERISTICS OF TALENT'S TWEETS.

no.	characteristic	retweeters(about)	tweet time(JST)
t1	movie	160	17:50
t2	quoted retweet	110	21:57
t3	nothing special	130	16:20
t4	image, other entertainers multiple	100	19:40
t5	image, co-star	530	22:04

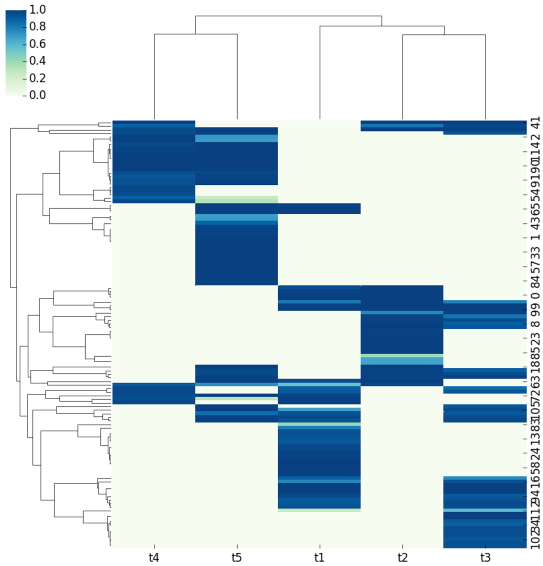

Fig. 6. Talent's retweeters. It seems that the content of the tweet which attention is drawn by the drama fan's retweeter and the retweeter for other purposes is different.

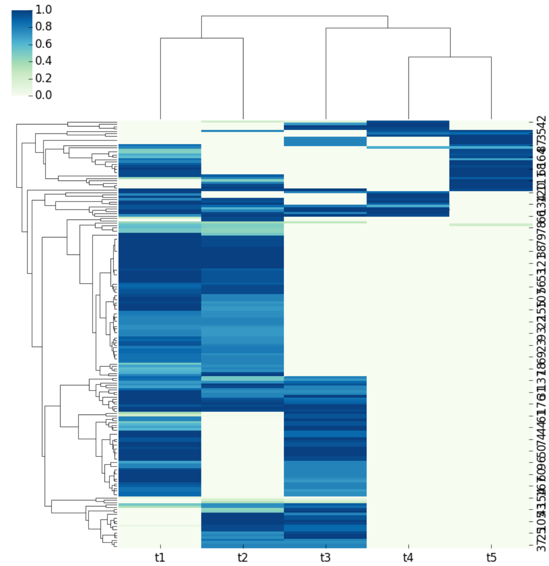

Fig. 7. Information distribution's retweeters. It seems that there are fewer users who respond to many tweets

Table II shows the statistics of the tweets.

Figure 6 shows the visualization result. There are a certain number of users who are reacting to multiple tweets, although not as much as the actor's retweeters. Many of the retweeters have a small time difference overall. Also, tweets with movies and images tend to have more retweets regardless of tweet time. Tweets on the drama starred by key persons who were broadcast at the collection time had a large number of respondents and the time difference was small. Tweets about this drama are t1, t3, t5, and few users retweet both these tweets and t2. It seems that the content of the tweet which attention is drawn by the drama fan's retweeter and the retweeter for other purposes is different.

C. Information distribution

TABLE III
CHARACTERISTICS OF INFORMATION DISTRIBUTION'S TWEETS.

no.	characteristic	retweeters(about)	tweet time(JST)
t1	image, food	1,300	0:55
t2	image, food	600	22:45
t3	image, beauty	450	20:30
t4	image, fashion	100	7:45
t5	image, game	280	8:15

Table III shows the statistics of the tweets.

Figure 7 shows the visualization result. Looking at the visualization results along the horizontal axis, similar colors are seen for each tweet. From this, it is considered that the behavior pattern of the individual twitter is the same for all tweets as a whole. Many of the retweeters are common to t1, t2. Since both of the contents of the tweet are food, it is inferred from the visualized image that it is a group of researchers who are interested in the same theme. In addition, although the tweet time is late at t1, t2, there are many thin colors (retweets with the big time difference), and it can be said that there is a retweeter that responds late to tweets at later times. On the other hand, t1, t2 on food and t3 on beauty, the layers of retweeters are greatly different, and it is considered that there are few retweeters who are interested in both food and beauty. Regarding t4, t5, the number of retweeters is extremely small. As the common point between these two tweets was that the tweet time was early in the morning, the primary reason for the small number of retweeters is probably the tweet time.

As inferred from the above visualization results, many of the information distribution accounts aim to deliver information to a wide range of users rather than targeting specific users, so their contents are diverse. As far as it is not a topic, it seems that there are fewer users who respond to many tweets.

V. CONCLUSION

This poster introduced our study on visualization of tweets group by key person and its retweeter groups. We found that the number of retweeters and time differences of retweets depend on the content of tweets. We could observe the regular retweeter surrounding key persons with the visualization using dendrogram and heat map.

As future work, we will analyze the dendrogram construction results in detail, and discuss the relationships between key persons and their retweeters while also exploring the field of interests of the retweeters. Also, we would like to develop a new visualization method to represent the relationship between key persons and retweeters more effectively.

ACKNOWLEDGMENT

This work has been partially supported by Japan Society of the Promotion of Science under Grant-in-Aid fir Scientific Research.

REFERENCES

[1] J. Chae, D. Thom, Y. Jang, S. Kim, T. Ertl, D. S. Ebert, ”Public behavior response analysis in disaster events utilizing visual analytics of microblog data”, Computers & Graphics Volume 38, pp. 51-60, 2014.

[2] S. Chen, S. Chen, Z. Wang, J. Liang, X. Yuan, N. Cao, Y. Wu, ”D-Map: Visual Analysis of Ego-centric Information Diffusion Patterns in Social Media”, IEEE Conference on Visual Analytics Science and Technology (VAST), 2016.

[3] S. Chen, S. Chen, L. Lin, X. Yuan, J. Liang, X. Zhang, ”E-Map: A Visual Analytics Approach for Exploring Significant Event Evolutions in Social Media”, IEEE Conference on Visual Analytics Science and Technology (VAST), 2017.

[4] B. Wang, C. Wang, J. Bu, C. Chen, W. V. Zhang, D. Cai, X. He, ”Whom to Mention: Expand the Diffusion of Tweets by @ Recommendation on Micro-blogging Systems”, the International World Wide Web Conference Committee (IW3C2), 2013.