Information Diffusion on Twitter: everyone has its chance, but all chances are not equal

Cazabet Remy	Nargis Pervin	Fujio Toriumi	Hideaki Takeda
National Institute of Informatics	National University of Singapore	The university of Tokyo	National Institute of Informatics
Tokyo, Japan	Singapore, Singapore	Tokyo, Japan	Tokyo, Japan
remy.cazabet@gmail.com	nargisp@comp.nus.edu.sg	tori@sys.t.u-tokyo.ac.jp	takeda@nii.ac.jp

***Abstract*—Twitter is a web 2.0 social network which attracted much attention recently for its usage as an alternative media for information diffusion. From the recent events in arab countries, to natural disaster such as earthquakes or tsunamis, Twitter has proven to be a credible alternative to traditional means of information diffusion. Relatively few works have been done on this question of information diffusion, and in particular on the relative importance of different kind of users on this question. In this paper, we show that all users are not equal on the aspect of information diffusion. By investigating thoroughly the retweet chain lengths of users on a large dataset, we found that the number of followers of users plays an important role in their capacity to propagate information. From our observations we propose a very simple model, which is accurate enough to generate realistic length of retweet chains on the network. We consequently show, by studying a Twitter dataset centered on the Japanese Earthquake and Tsunami in March 2011, that such a crisis impact greatly the propagation of information. Finally, we use our results to discuss on the means of improving information diffusion to reach targeted users.**

I. Introduction

Web 2.0 social networks platforms, such as Facebook, Twitter, Google Plus or Flickr, have attracted lot of attention in the recent years. As more and more users join them, new usages develop, and their role in society increase. Twitter is one of the most influential platform, as highlighted by recent events. Two examples are now especially famous: first, the key role played by Twitter during the so-called arab spring (see, for example, [**?**]), during which insurgents used it as a major way of organization and as an information source. Secondly, in an unrelated topic, during the 2011 earthquake and tsunami in Japan, during which users relayed information ranging from tsunami alert to rescue request, or more simply shared local information.

One key feature of Twitter, which makes it maybe the most efficient web 2.0 social network platform to share information in time of crisis, is that it is particularly suited for information diffusion. Unlike other platforms, there is no reciprocity needed in the relations, and therefore the most influential users can gather tens of thousands of so called followers, people who register to be informed of their publications. On the opposite side, as in others small world networks, most users have very few relations, ensuring a more convincing power law distribution of in-degree than, for instance, Facebook. The consequence is that different users can play very different roles in the network, in particular in terms of information diffusion.

In some previous works that we will present in the next section, authors have investigated the factors that could influence the success of the diffusion of a piece of information. Most works done to predict the success of tweets based on their content have resulted in relatively poor results. If their is for sure a relation between the content of a tweet and it's probability to be retweeted, it seems nearly impossible to predict the success of a future tweet based on its content, or to derive a model which could generate realistic tweet diffusion based on what is said on the network.

More successes have been achieved by relying on the properties of the network itself, and in particular the properties of the user. The relation between the number of followers of a user (to which we will refer further as its in-degree) and the average length of its tweets has been shown several time.

The goal of this paper was to push further the study of this relation, which is, until now, the more reliable way of predicting the success of a tweet.

Our paper is organized as follows: after reviewing briefly related works in the next section, we will present the experiments we have conduced in order to describe more precisely the role of the number of followers in the probability to produce seminal tweets. We will then use these findings to propose a simple model, which allows us to generate, starting from the sequence of unique tweets posted by users in the network, chains of randomly generated lengths following the proposed model. We then check that the generated retweet chain length match real data, not only on the aggregated level, but also on the aspect of the relative influence of most popular and less popular users in the generation of long chains of retweet. We will continue by showing that these retweeting behaviors can be strongly disrupted by an event such as a large earthquake. Finally, in the last section, we will discuss how our findings can help to understand how it could be possible to optimize the probability of sharing a given information on the network.

II. Related works

The question of who are the key influential users, and why, has already attracted a lot of attention. One of the most influential work made on this topic is presented in Bakshy et al. [**?**]. The authors, after observing that the distribution of size of retweet chains were roughly following a power law, were interested in understanding which factors could explain the success of popular tweets. They found a correlation between the number of followers and the probability of producing a widely shared tweet. They also found, by asking people to classify tweets' content, that tweets with content classified as "interesting" were more likely to spread. However, one important finding was that predicting which particular content or which particular user might produce a widely propagated tweet was mostly unreliable. They concluded that, if one wants to optimize the diffusion of a particular information, one must use a large number of sources users, relying on average effects.

In Suh et al. [**?**], several factors that could be correlated to the retweet success of tweets have been investigated. First, the authors have studied the correlation between the content of the tweet and its number of retweets. However, considering either the hashtags included in the tweets or the URLs did not show any convincing correlation. It is therefore likely that predicting the number of retweets based on the content of the message would be very difficult, as such success can be highly influenced by a large number of factors, and in particular factors external to Twitter. The same tweet, for instance about the Fukushima power plant, published just before or just after the 2011 earthquake might result in very different spreading. In a second attempt, the authors studied how the properties of the tweet's author might impact the diffusion of his tweets. They found that the past number of published tweets had barely any impact. On the contrary, they found a very strong and regular correlation between the number of followers of a user and the average number of retweets of their tweets. They found that two other factors, the age of the tweeter account and the number of followees, had some kind of correlation with the average retweet count. However, these relations are not linear nor monotonic. Furthermore, one can note that both of them are not independent with the number of followers: the number of followers tend to increase with the age of the account, and users with more followees tend to have, in average, more followers, as it has been observed, for example, in Krishnamurthy et al. [**?**].

Kwak et al. [**?**] have also investigated the relation between the number of followers of tweets' authors and their retweets. In their case, they studied more particularly the number of retweets made by users who were not direct followers of the original posters. Interestingly, they argue that users with less than 1000 followers tend to have the same average number of retweets. However, they do not display only the average retweet count, along with the median. And this median tends to increase with the number of followers. Therefore, we can assume that even though the average number of retweet do not change, the distribution of the retweet count of all tweets by the user changes with the number of followers.

Some other works have studied the influence of users, in particular Cha et al. [**?**], who found that users with high in-degrees are usually influential in term of new information, but some of them are not influential in term of retweeting. They also investigate the relation between influential users and different topics.

III. Dataset description

Tweet Data: We used a Twitter dataset collected during the great Tohoku earthquake in Japan in March 2011 and described thoroughly in [**?**]. The dataset covers a period of 20 days (from 5^{th} March, 2011 to 24^{th} March, 2011), and consists of 362,435,649 tweets posted by 2,711,473 users in Japan. This dataset is remarkable by its completeness: the authors have checked that 80% to 90% of all published tweets by these users were present in this dataset.

Fig. 1 shows the normalized retweet count for 20 days of period. The first two major peaks represent the two big earthquakes on 11^{th} and 12^{th} March. After the disaster, retweet count progressively returns to its normal average values.

Follower Network Data: In Twitter, follower network depicts the social relationship among the users. Follower information has been collected by crawling Twitter API in May, 2013 for the active users who have been mentioned more than 20 times in the dataset. Follower network dataset consists of 300,104 users and 73,446,260 relationships information. Degree distribution has been shown in Fig. 2 by plotting cumulative fraction of users against the number of followers/followees of user.

We can note that this follower network is not complete, and collected after the events. However, in this work, we are not interested in the particular follower/followees relationships, but simply on the number of followers of each user, which is probably less sensitive. Nevertheless, we did not take simply the number of followers of the users from Twitter, bust instead counted how many followers in our dataset were following each particular user, to increase accuracy. For example, some users are very popular outside of Japan, and therefore have a large total number of followers, but are not followed by many of our selected users.

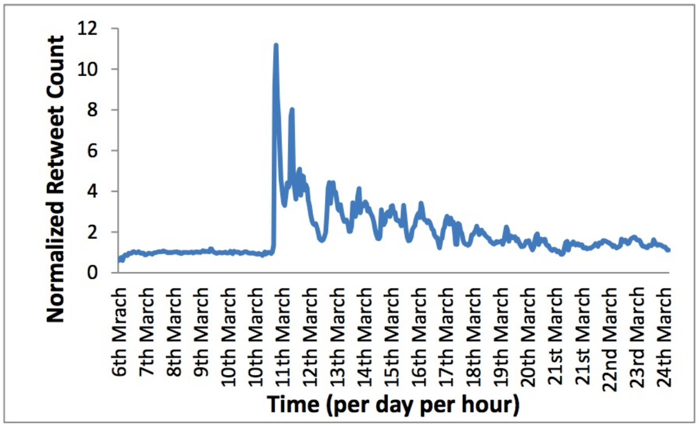


Fig. 1. Tweet Distribution over days (Normalized)

IV. From observations to a simple model

In the first part of this section, we will present some findings on the role played by the number of followers of a user in the propagation of its tweets. In the second part, we will validate that these findings are reasonably accurate by using them as base in a model for the generation of tweets diffusion.

1) Retweet Chain Length distribution: We first began to check, on our dataset, the correlation between the number of followers of a user and the average length of the retweet

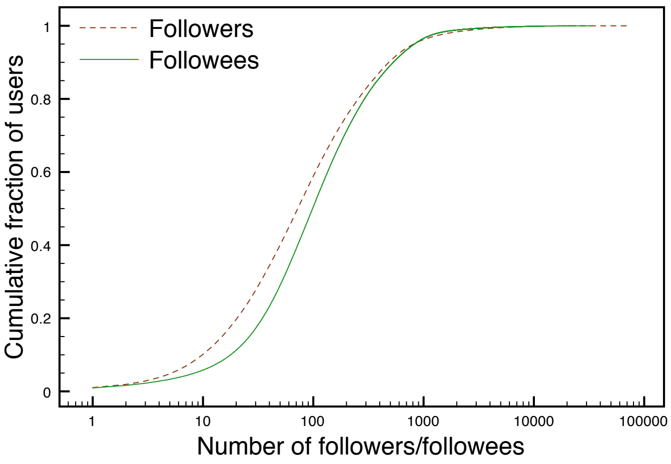


Fig. 2. Cumulative fraction of users by degree

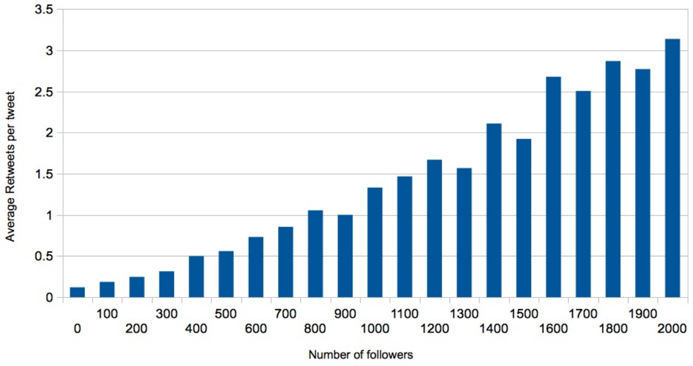


Fig. 3. Average number of retweets per tweet vs. the number of followers

chains of the tweets he posts. This relation has already been discussed in other papers and have been validated on our dataset. We observed the same correlation as depicted in fig. 3. Most works use a dataset collected by randomly choosing a relatively small portion of the tweets over a large number of users for long period of time. On the contrary, our dataset is more close to completeness, for a shorter period and with less users. To identify tweets belonging to the retweet chains, we first recognize all original tweets not containing the “RT @” mention. In the second step, for each of these original tweets we counted all tweets of the form “RT @” in which the section which appears at the right of the last RT mention is identical to the original tweet. This process might have some flaw as some retweets might not be formatted in the exact pattern that we are searching for (as identified by Boyd et al.[**?**]), but as the retweet functionality have been now implemented in most Twitter clients including the official one, a higher percentage of retweets are correctly formatted. Furthermore, the Japanese language needs far less characters than English on average for the same content, so it is probably less common to modify the original tweet for concision reasons.

The corrleation in Fig. 3 is found on average retweet chain length. However, it is obvious that there is a great disparity in the length of retweet chains. Taking the average will loose the insight of the distribution of retweet chain length. Moreover, it has been shown in [**?**] that the overall distribution of retweet chain length follows a power law. As a consequence, computing an average value is mostly meaningless. Therefore, we searched to characterize the relation between followers and information diffusion, not through the mean diffusion, but through the evolution of the parameters of the power law representing this distribution.

In this paper, we observe several time the power law nature of some distributions. As it has been thoroughly discussed by Clauset et al. [**?**], with the recent growth in interest for power law distribution some works claim to find power law distributions where they are not actually the best model for the observed distribution due to lack of rigor in the evaluation. To mitigate this issue in this paper we use maximum-likelihood to fit the curve instead of least-squares fitting. Further, using a goodness of fit test based on Kolmogorov-Smirnov statistic we checked whether power-law is in fact a good fit. Finally, we verify that a log-normal distribution is not an obvious better descriptor, as it is sometime the case. The power law nature of these distributions are not claimed as result of our work, but are used to represent the distribution and its evolution in convenient way. Statistically, the power law distribution is a very good approximation of the distribution. It could probably be possible to improve the accuracy by finely tuning a cut-off for it. But this cut-off will most likely depend on the size of the particular network, while we were more interested in a general relation between the in-degree of users and their probability to generate retweet chains of a given size.

A. *Influence of the number of followers*

We have validated (fig. 3) that the number of followers of a user has an influence on the average retweet chains of tweets. Therefore, we know that they do not have the same probability distribution of being retweeted. In this section, we study how this distribution change with the number of followers. To do so, we computed the distribution on the whole network, but only for users of a given number of followers. We restricted our analysis to the first five days of our dataset, from 5 to 9 of March, 2011 -days preceding the earthquake- as we thought that users might change their behaviors after this event. Note that we kept one day, the 10 of March, in order to check that our finding were also valid on this day.

In order to have enough data for statistically significant results, in particular for users with large number of followers (FC) who are less common, we compute the distribution for all possible i with all users u such as $FC(u) = i, i*5 < FC(u) < (i+1)*5, i*10 < FC(u) < (i+1)*10, i*50 < FC(u) < (i+1)*50, i*100 < FC(u)*100, i*500 < FC(u)*500$ and $i*1000 < FC(u)*1000$. Then, for each of these results, if we have a minimum of 10,000 chains with a strength greater than 1, that is, retweeted at least one time by a user different than the original one, we checked the goodness-of-fit of the power law distribution, and, for most of them we found high values. We eliminated results for which the goodness of fit was below 0.1 – the threshold recommended in [**?**]. We found that the power law distribution was more accurate when considering only users of similar number of followers than when aggregating all tweets of all users. We also observed that the power law was not always the best model for users with many followers, in particular more than 2000 followers.

For these users, there seems to be a cut-off effect, which means that the number of very long chains (more than 1000 retweets) is fewer than expected with a power law. However, these events being really rare, for simplicity reasons we decided to neglect this effect.

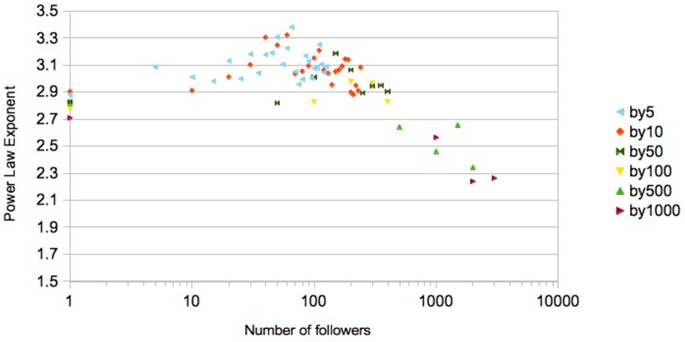


Fig. 4. Parameter α of the distribution of retweet chains length according to the number of followers of users

In the following step, we investigated how the parameters of the power law, α and x_{min}, are estimated by the follower count of the users. In Fig. 4 all values of α are plotted. x_{min} do not vary much, being remarkably small, usually between 2 and 3, which means that the power law is valid even to characterize the probability of small chains. We can observe that the parameter α changes in two phases. Before reaching 100 followers, the value of α is stable or even slightly increasing. But starting from approximately 100 followers and more, the more followers a user have, the higher the value of α. This means that the more followers a user have, the higher its probability of being widely retweeted.

We can interpret this result in the following manner: for users with less that 100 followers, this property does not strongly affect the probability of being retweeted. Above 100 followers, more followers means more widely propagated tweets. From these observations we computed two simple approximations of the relation between α and FC. Under a hundred followers the relation is linear, while it is a logarithmic relation above 100. These models are plotted on figures 5 and 6 with their parameters. Though other fitting models could be adopted, we choose these ones for their simplicity.

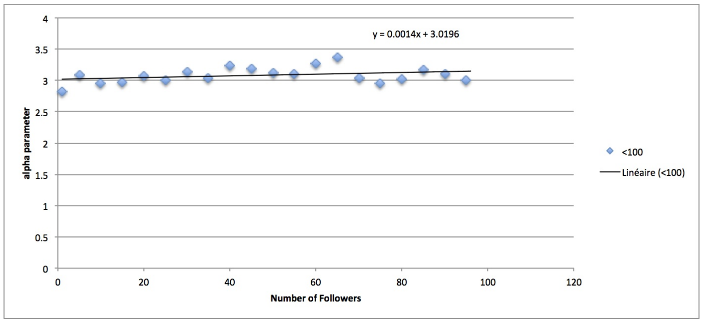


Fig. 5. Relation between α and FC with $FC < 100$

B. *Validating the model of retweet chain length distribution*

Given a Twitter dataset, we want to propose a simple model that can be used to simulate the propagation of tweets on it without having to consider the content of the tweets or the user-specific behaviors as these properties are highly contextual. Of course, the consequence is that our model will be valid only at the global level and not for the individual users performances. But works such as the one by Bakshy et al.[**?**] have shown that such a precision is not plausible.

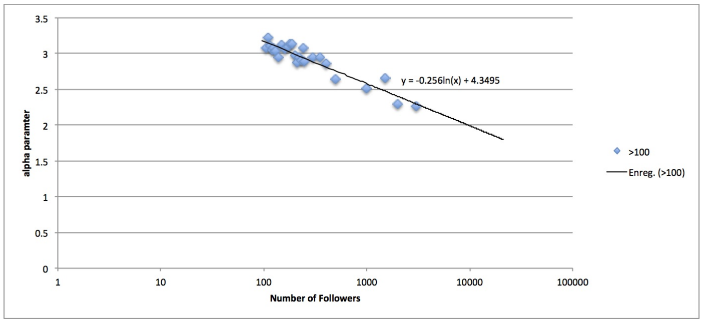


Fig. 6. Relation between α and FC with $FC > 100$

Two papers have already proposed models of tweet diffusion. We will begin by presenting these works. In Yang et al. [**?**], the proposed model is based on factor graph model. More precisely, they train their model using 22 features, such as messages' contents, views from followed users, time delay between views of the tweet etc. For each individual tweet posted, the authors attempts to predict the probability of getting retweeted by poster's immediate followers and other followers until the end of the tree. This method is therefore strongly linked to the studied network and to the previous observations. The authors do not give the correspondence of the global properties between the results produced by their model and the observed data.

Another model have been proposed in Yang et al. [**?**]. The idea is relatively similar: the authors use the Cox proportional hazard regression model to quantify the degree to which a number of features of both users and tweets predict the speed of diffusion. Namely, they use the fact that a tweet uses link or mention, the number of posts of a user, its rate of being mentioned by others, and other parameters which are not specified. Then, they classify tweets in their dataset by topics, and study, for each topic, which factors are the more influential. They found that the important factors vary greatly from topic to topic. As in the previous model, the results obtained are specific to the dataset.

We must also notice a strong possible bias in these models. In both cases, the model is tuned from the data by looking at the retweet chains. These retweet chains are deduced from the content of the tweet. For instance, if a tweet is of the form "$RT@\ u_1\ :\ \ content$", they will assume that the poster is retweeting the user u_1. Similarly, if the tweet is of the form "$RT@\ u_1\ :\ RT@\ u_2\ :\ content$", they assume that the tweet was initially posted by u_2, and that the current user saw it from u_1. However, in reality, this assumption is mostly inaccurate for mainly two reasons:

- For character limitation reasons, users tend to keep only one user in the "trace" and therefore, it is not possible to know the real propagation of a tweet from the content of a tweet itself. When using the official retweet option, only the name of the original author of the tweet appears.

- When a tweet becomes popular, users do not need to follow someone who retweeted it to see it: the tweet can appear on other medias or even as a recommended tweet on Twitter's platform. If users retweet it for this reason, they will only cite the original author.

The consequence of these two problems is that we cannot rely on the content of the tweet to study the propagation of tweets. This might, in particular, conduce us to overestimate the impact of the original author in the diffusion.

In the previous section, by studying the data we have observed that the distribution of the length of retweet chain could be represented as a power law, with an α parameter defined as a function of the in-degree of the user. In this section, we show that this model, while being very simple, reproduce accurately the global properties of retweets.

In order to do this validation, we adopted the following procedure: for each original tweet published in our dataset on March 10 (the day prior to the earthquake, which was not included in the data used for studying the actual distributions) by a user for whom we know the number of followers, we randomly generate a retweet chain length according to our model. We also count how many times this particular tweet has been actually retweeted. When this operation has been done for all tweets published during this day, we compared the total number of retweets of each length generated by our model to the real numbers. Fig. 8 shows that the correspondence is very good, which means that, despite the simplicity of the model it fits accurately the data.

However, this result could have been achieved without taking into account the variations in retweetability due to the number of followers. We conducted another test, in which we compare the relative proportions of tweets of a given retweet chain length with given FC.

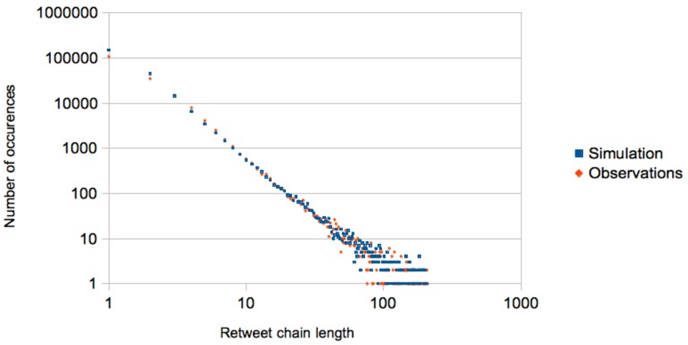


Fig. 8. Correspondence between retweet chains generated by the model and real retweet chains

More precisely, we sorted retweet chains in four categories:

- Short chains: between 1 and 10 retweets
- Medium chains: between 10 and 50 retweets
- Long chains: between 50 and 500 retweets
- Very long chains: more than 500 retweets

We also categorize our users in four:

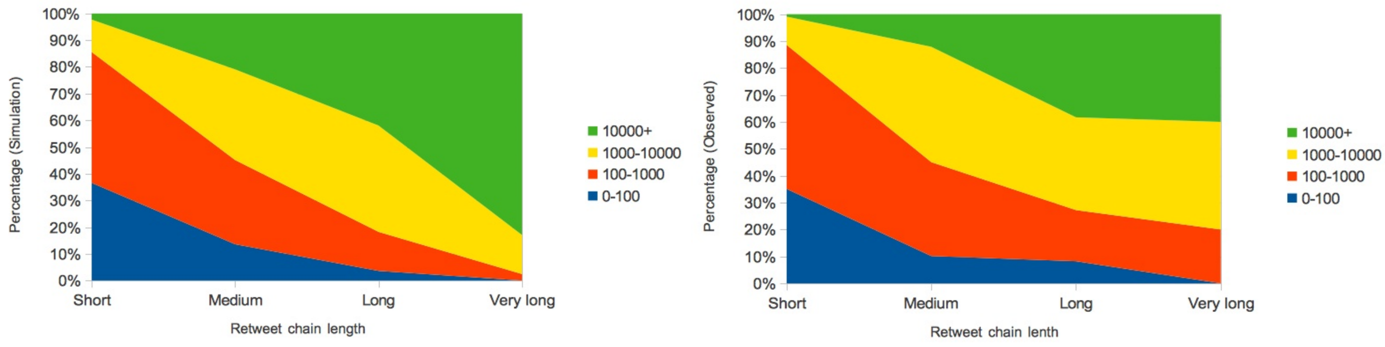
- Few followers: between 0 and 100 followers
- Medium number of followers: between 100 and 1000 followers
- Many followers: between 1000 and 10000 followers
- Super hubs: more than 10000 followers

In figure 7, we can see that the share of all tweets of a given retweet chain length due to users of a given number of followers follows a very similar pattern in our simulations and in the real network. Interestingly, we see that most short retweets are due to users with few followers. The longer the chains, the higher the proportion of users with a large number of followers. This profile depends on two factors: the difference in probability of having long chains according to the number of followers, and the relative proportion of tweets made by these users. We can however notice that in our simulation, we tend to overestimate the role of users with many followers in very long chains. This might be explained by a cutoff effect on the power law distribution of retweet chain length of these users. As we stated before, the power law model was not perfectly satisfactory for users with very high in-degree, because of a cutoff effect.

C. Effect of the Earthquake

Several previous works have studied the effect of major crisis on Twitter. Sakaki et al. [?] have shown that such events have an impact so clear on the social network that it is possible to know not only the occurrence of an event such as an earthquake or a typhoon, but also their scale, their precise time of occurrence and location. Earle et al. [?] have shown that the increase in the number of tweets during an earthquake has also been used to detect this earthquake. Several other works [?] [?] have also used the same phenomenon, the increase in the number of tweets, to capture information about earthquakes. However, here, we are interested in the effect that such disaster can have on the retweet behavior. We have observed that there are more tweets published during the crisis, but the question arises: is it because there are more unique information published and the retweet behavior is unchanged? or is it because people tend to retweet more?, or a combination of both? By studying the distribution of retweet chains during the period of crisis, we can answer to this question.

We ran, for each day, the same study on the evolution of the scaling parameter α than we have done for the pre-earthquake period. We found that for each day, the pattern of slow increase until $FC = 100$ and then logarithmic decrease was still present. However, we also observed a strong variation of the α values, depending on the day. Figure 9 illustrates this evolution. Each point represents the average value of α for all values of FC in the range 50-100. As we can see, as soon as the earthquake happened, the α parameter decreases strongly. In the following days, this value continuously increase, until coming back to normal values. This mean that for the same user posting the same tweet, the probability of being retweeted is more important after the earthquake than before. We consequently checked which was the change in behaviors inducing the more change in the number of tweets published: a raise in retweets or a raise in original tweets. We found that, on the day



(a) Simulated data

(b) Observations

Fig. 7. Comparison of the proportion of retweets chains of a given length vs. users with a given number of followers. We can observe that our simulation reproduces the properties of the real data, that is, an increasing proportion of tweets made by users with more followers as the length of retweet chains increases.

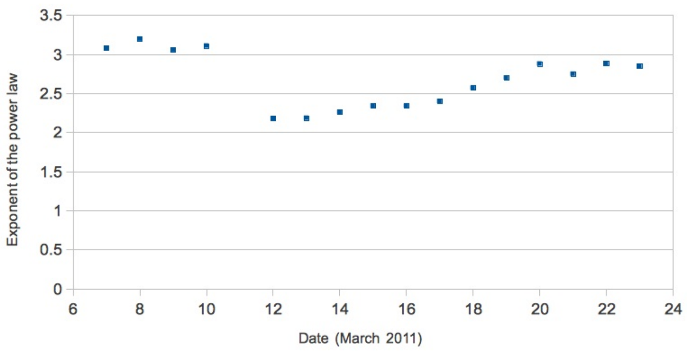


Fig. 9. Evolution of the scaling parameter α according to the date. Immediately after the earthquake, the value decreased strongly, which means that a user with the same number of followers before and after the earthquake has a higher probability of being widely retweeted after the earthquake. The value increased slowly in the following days, but cannot reach the pre-event level.

of the event, the most influential change was the raise in the number of original tweet published, accounting for roughly two thirds of the additional tweets. However, during the following days, the number of original tweets came back to the same levels as before the earthquake, or sometimes even lower. This can be explained by power shortage in some areas after the quake, altogether with a change in behaviors. The percentage of total tweets being retweet, on the contrary, stayed lower than average for several days after the quake, coming back slowly to normal. This comfort our observations. As a conclusion, we can say that, after such a crisis event, there is an important change in the retweet behaviors.

V. Discussion and usage of the model

We have seen in the previous section that generating retweet chain lengths according to our observations gives realistic results. Therefore, we can consider using this model to study further the propagation of information on Twitter. When using such a model to study tweet propagation, we always have to keep in mind that our findings are not applicable to any particular tweet. Probably, what will make a tweet successful in the end will be the content of the tweet, the context, and the intrinsic characteristics of the users who publish and see the tweet.

One problem of particular interest for us was the question of the propagation of information in the time of crisis. We know that Twitter has been widely used as a way to propagate information during crisis, in particular when traditional sources of information are not accessible, for catastrophe reasons or even control by another entity. Therefore, one question that we can ask is: how could we ensure the diffusion of an important information to a wide range of users? Or more generally, how users with many followers are better at diffusing information than less popular users? As we said earlier, the particular properties of a tweet are certainly more important than the degree of its poster. However, the impact of this degree on the probability of diffusion of a tweet is not disputable. One could argue that causes and consequences might have been inverted, and that users who publish many popular tweets have many followers because they tend to publish interesting tweets. If that was true, the assumption that the same tweet published by two users with different in-degrees would have different probability of being retweeted would be wrong. However, while it is likely that these users became popular because they tend to publish more interesting tweets than other users, their current status of "celebrity" actually give them a higher power of propagation due to their large audience. An example of this can be found in [**?**], in which the role of users who are so called amplifier of a tweet is investigated. When these popular users retweet an information published by another less popular user, the further propagation of the tweet is increased strongly. Hence, the same information, tweeted by users of different popularity, achieve different success. It therefore makes sense to study the relative impact of users according to their audience.

The question we want to answer here concerns the probability to publish highly retweeted tweets. For a particular information to reach many people on Twitter, the only possibility is to produce a tweet that is propagated by many. We therefore want to know who are the persons who are the more likely of publishing such information. We know that any user can, of course, publish such a tweet, but we also saw that the probability was much higher for users with more followers. We therefore investigate the probability of producing a tweet of

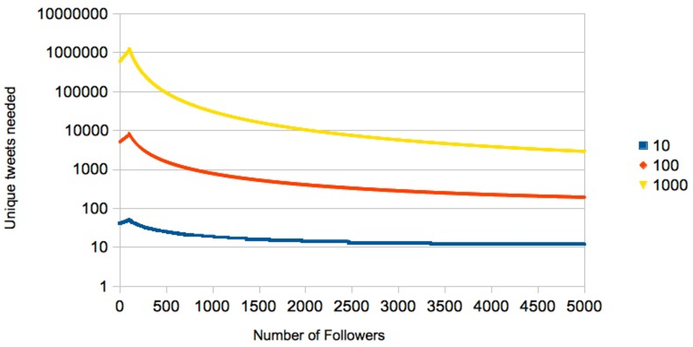


Fig. 10. Number of unique tweets needed to have one tweet of a given length (10,100 or 1000). The longest the chain, the strongest the difference between popular and less popular users.

retweet chain length 10, 100 or 1000 according to the degree of the user. We use the parameters of our model described previously to compute one out of how many tweets might reach the objective length. Figure 10 presents the results of this study. We can observe that the difficulty to produce a tweet of a given length vary strongly with the popularity of users. This difference is far more important for long chains than for short chains. When looking at the frequency of tweets of retweet chain length 10, the difference between users of different degree stays always smaller than a factor 10. On the contrary, for tweets of length 1000, the difference between a user followed by 100, and a user followed by 5000 is of a factor more than 100. This information might be relevant if, for example, one wants to diffuse an information on the network. If the targeted users can be identified, and represent only a small subset of the network, then it is not primordial to have influential users publishing this information. It might be more pertinent to incentive users who belong to this subset, and these users will share it to a short distance in the network, corresponding to the targeted users. On the contrary, if one wants to reach a very broad audience, and think that anyone in the network might have access to this information, then having nodes with a lot of followers publish the information might help, as these users are more likely to generate long chains, that is, to make the information travel further in the network from the original publisher.

These observations are not contrary to the findings presented in [**?**], but might complete them. In this paper, the authors investigated roughly the same problem, and came to the conclusion that the impact of users with a lot of followers was not statistically so much greater than users with a few followers. For someone who would like to diffuse a given information on Twitter they therefore recommend to concentrate more on users with few followers, as it is usually easier to make such users diffuse information. However, our results suggest that this is true if we consider the average length of retweet chains. But due to differences in the inclination of the power law (α parameter), this recommendation is to be tempered if one considers that long chains are more important than small ones. One information retweeted one thousand times might give a better diffusion than one hundred pieces of information retweeted only ten times.

VI. Conclusion and perspectives

In this paper, we have provided new insights into the propagation of information in Twitter. In particular, we have shown that the distribution of the probability of producing a tweet of a given retweet chain length can be represented as a power law for users of a given in-degree. We have also shown that the parameters of these power laws were different for users of different in-degrees. More notably, we have shown that the parameter α decreases as the number of followers increases with a logarithmic correlation.

Moreover, we have shown that this characterization was pertinent enough to constitute a simple model, allowing us to generate realistic lengths of retweet chains, reproducing some key features of the distribution of long retweet chains. One future work will be to enhance this model by actually generating this retweet chain. On this dataset or another one with the same information, as we know the follower network and the tweets published by all users, we would therefore be able to play alternative versions of the diffusion of information on this network, during a studied period. By comparing the diffusion which actually took place with the generated one, on aspects such as individual information received by each user, or total amount of people informed of a particular piece of information, such as hashtags or URL, we will be able to get more information on how information is diffused on Twitter, on what can be modeled by global properties and what depends on exceptional and unpredictable behaviors or particular users.

Acknowledgment

We thank Genta Kaneyama (Cookpad Inc.) for assistance in collecting data from Twitter.

References