

Simulated News Spreading on Social Network Sites—Analysis from Network Evolving and Node Measurement

Bing Zhang

1. Introduction

Hundreds years ago, the news is spread by talking from one person to another. At that time the spread speed and distance is largely limited. As the traditional media rise, there are many way for news to spread like TV or newspaper. Nowadays, an emerging channel for news to spread is online social network site (SNS) like Twitter or Facebook where users can broadcast information to their friends and others who care about it. Among various micro blogging systems, Twitter is the most popular service by far. In US and the world, there are more than 175 million twitter users [1]. The news received from twitter is more instantaneous and fast spreading. We saw SNS played important role in many social events like Arab spring or Olympic game. In this paper, we will first make a literature review about twitter like social network and its mechanism. Then we will investigate the question: first of all, giving the same network condition, what will influence the spread the news. Secondly, what will happen when the network is growing? Finally, we will introduce a complementary HubNet activity and the limit of the model.

2. Literature Review

Like other social network websites, Twitter is a website where users register, fill in their personal profiles, choose some of the other users who are already in the website to follow, and then broadcast their thoughts or moods in a limit of 140 words. When the users you are following say something new, they will appear in your timeline at the homepage of Twitter. In this model, we use the scale free network to simulate the initial condition of the network. According to Dechun

Liu's report[1], in the past, Tong yang Yu collected 4036 users of a SNS website and their relationships to form an undirected graph.[6] He studied some basic attribute of the graph, including the degree distribution, clustering, and network core, finding that the network of SNS users is likely to be a scale-free network. Though Twitter-like website forms a directed graph, the main attribute is the same.

Many researchers have investigated the relationship among centrality, cluster coefficient and network. They provided different kinds of calculation model for discrete graphs. We use the basic approach to access the centrality. For the local cluster coefficient, we take Stonedahl's paper as reference [4]. In this following section, we will discuss the correlation among the number of people received new, the centrality of news creator and its local cluster coefficient.

3. The Model

In this model, every node represents a person in the SNS. A yellow square represents the news. Initially, we randomly pick a people as news creator. The news spread follow the below rules:

- a. Every people held the news will twitter the news to all their neighbors
- b. When a person receives the news, he may be interested to the news. In this case we say this
people hold the news
- c. If a person has totally different opinion with the news creator, he will read but ignore the news

The network can grow as the news being spread on the network. The growth follow the below rule:

- a. If the person holds the news, he will follow the creator of that piece of news

We visualize the above rule using different color as shown in Fig. 1. Red color is one group of people with same opinion. If we want to investigate the influence of different opinion, we can split the entire graph to form two groups—red and blue. The blue will ignore the news. Pink nodes represent those people hold the news. The green represent those people with same opinion but are not interested in the news. The grey represent those people with different opinion but received the news. Both grey and green nodes read the news. However, they won't further retwitter it.

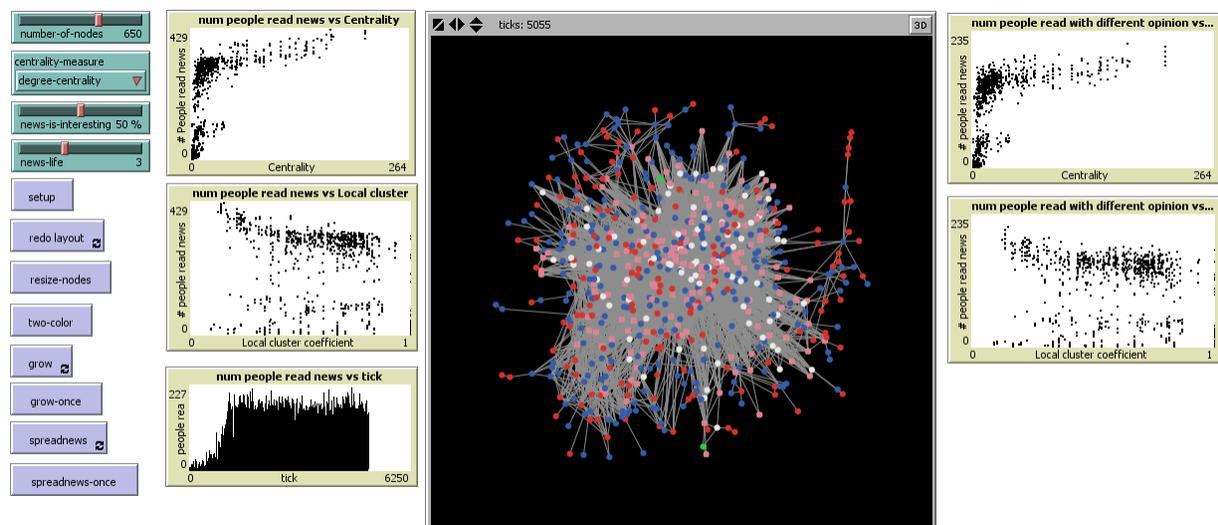


Fig. 1 overview of SNS simulation model

The Fig.1 is an overview of the running model. We use two different kinds of metrics to describe the characteristic of a node in network—centrality and local cluster coefficient. According to Forrest Stonedahl, the fraction of neighbors of the node whose neighbors are also neighbors of the target node, normalized by the highest clustering-coefficient in the network. The lower the clustering coefficient of a node, the less overlap there is among its neighbors, encouraging wider adoption more quickly [4]. As described in paper “Methods to Determine Node Centrality and

Clustering in Graphs with Uncertain Structure” [5], the centrality can be used to describe how important a node is in a graph. There are four different types of centrality. Here we mainly use the degree centrality and closeness centrality to analyze the graph.

4. Simulated Result Analysis

4.1. Single color graph

We first let the network grow for 200 cycles, and then perform news spread function. The result is shown in Fig. 2. We found that the number of people received the news is proportional to the centrality.

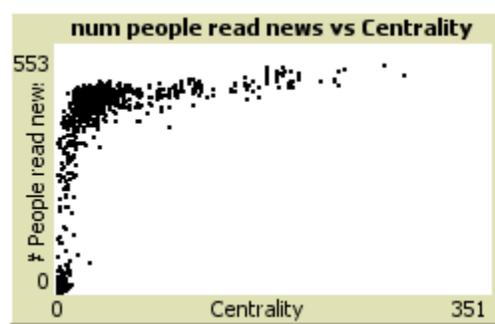


Fig. 2 Number of people received news versus centrality

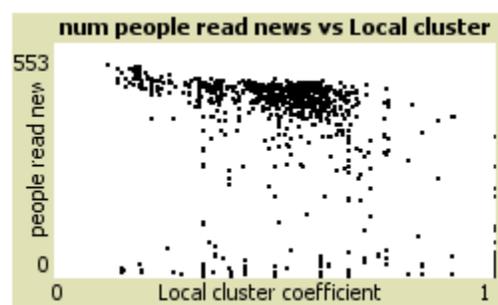


Fig. 3 Number of people received news versus cluster coefficient

The cluster measurement is shown in Fig. 3. The number of people received the news is inverse proportional to the local cluster coefficient.

4.2. Two color graph

The simulation result of two color graph is shown in Fig.5 and Fig.6.

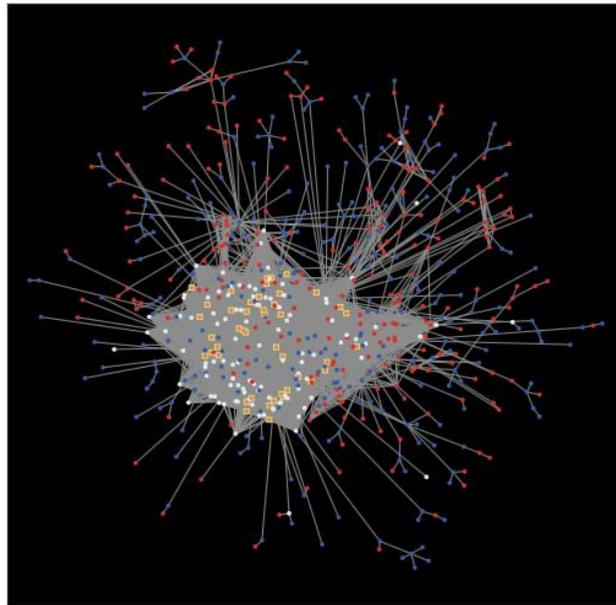


Fig. 4 Network for two colors

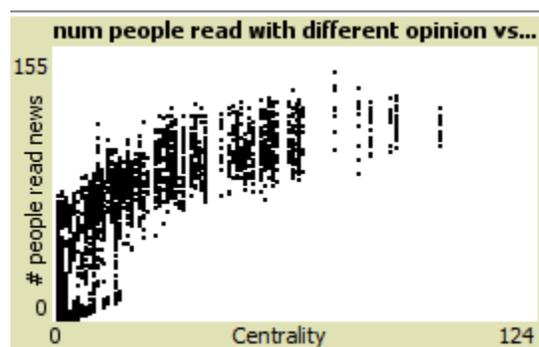


Fig. 5 Number of people received news versus centrality

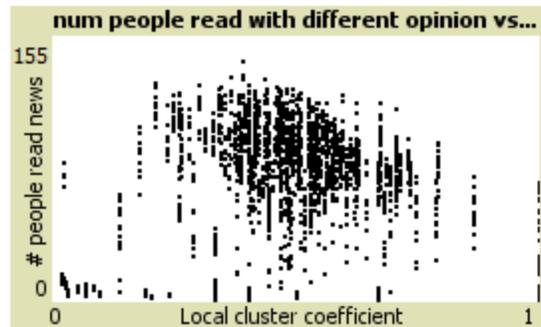


Fig. 6 Number of people received news versus cluster coefficient

The result is similar to one color graph that the number of people received the news is proportional to the centrality and inverse proportional to the local cluster coefficient. However, we found that the absolute value of the number of people received the news is smaller compared with single color graph. The reason for this phenomenon is that the blue node clock the news spread along some path so that increases the difficulty for news being spread.

4.3. Network Evolving

We also observe the number of people received the news among different cycles. The result is shown in Fig. 7.

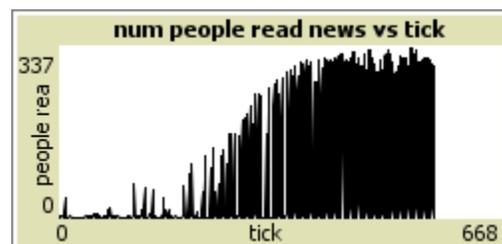


Fig.7 Number of people received news versus tick

We can conclude that as the network growing, with the same news life cycle, more people can receive the news. This is because when the network grows, the entire network becomes more

connected. There is one interesting found that after certain cycles, the number seems saturated. This saturation comes from the news life cycle, which limits the depth the news can spread. Under the condition that the total number of nodes is much more than the saturated number, if we increase the new life cycle, more people will receive the news.

4.4. Influence from Network Evolving

After waiting for the network to grow for a long time, we measure the relationship between number of people received the news and the centrality. The result is shown in Fig. 8. Although as a whole this graph still tells us the proportional relationship between number of people received the news and centrality, it is more saturated, especially in lower x-axis region. This comes from the quite connected graph. Some node even has low centrality. Because the graph is quite connected, the news still get spread.



Fig.8 correlation after long time network growth

5. Comparison with Reference Pattern

We consider validating our simulation data with reference pattern. In Liu's paper [1], they got the result of rumor spread with time. This model is related to our simulation. From her result as

shown in Fig. 9, we found the number of rumor is proportional to the time. This result matches out result shown in Fig. 7.

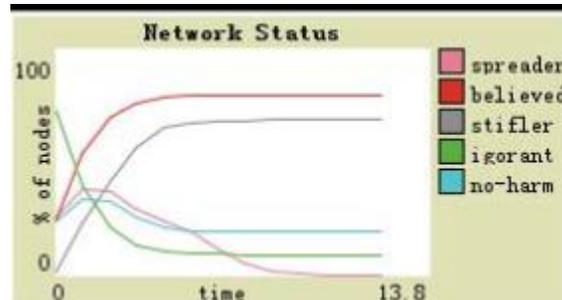


Fig. 9 Liu's result

In the paper “Want to be Retweeted? Large Scale Analytics on Factors Impacting Retweet in Twitter Network” [3], the author provide the statistics about the real twitter data. In Fig. 10, on the Xaxis, we put users into buckets according to an interval of around 100 followers, ranging from 0 to roughly 5000. On the Y-axis, we plot the retweet rate of users in that particular bucket. It shows a very strong linear relationship between the number of followers (x-axis) and retweet rate (yaxis). In other words, intuitively, the larger is the audience, the more likely the tweet gets retweeted. This result is the same as our result for the centrality.

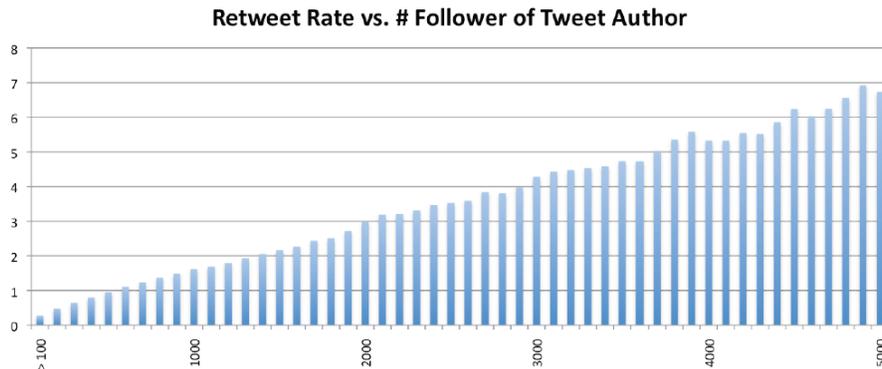


Fig. 10 Real date for retwitter

Joseph J. Pfeiffer stated the methods to measure the centrality and cluster coefficient in his paper [5]. He illustrated that the correlation between centrality and local cluster coefficient is inverse relationship. This conclusion further validates our result from simulation.

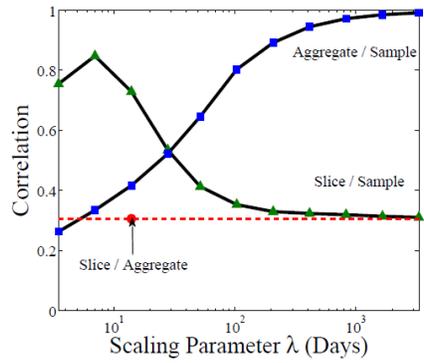


Fig. 11 correlation between aggregation and slice

6. Complementary HubNet Activities

As a complement to the NetLogo Social Network simulation model, I designed a HubNet activity that allows a group of people in classroom participate. The activity is shown in Fig. 12.

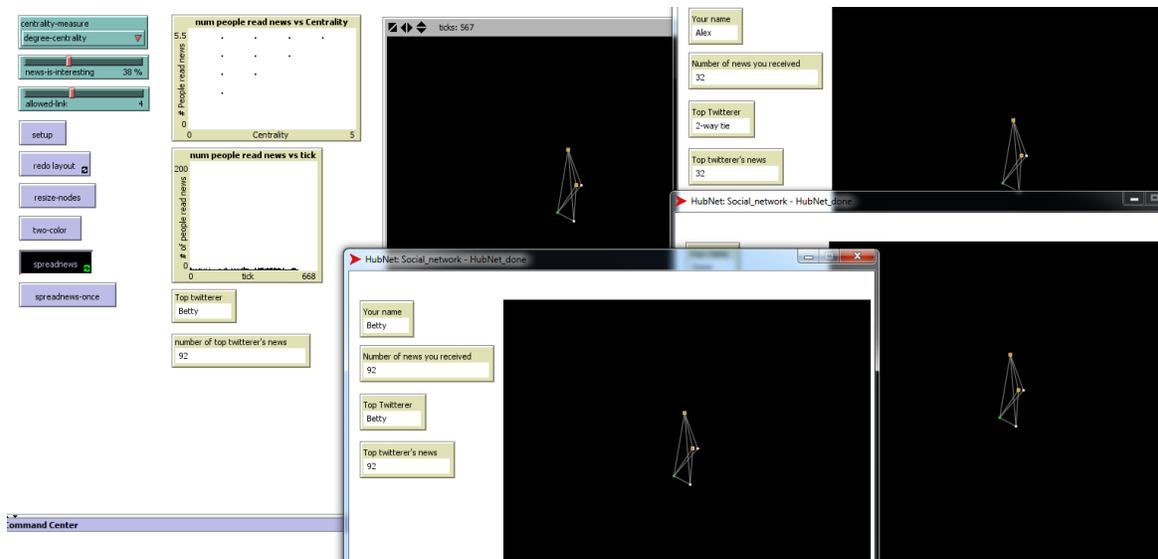


Fig. 12 HubNet activity

In this activity, each participate student is a node in the network. They can use their mouse in client to choose the people they want to follow (connect). They can relick that people to remove the following relationship (disconnect). In this case the network will be a truly dynamic network where everyone is trying to optimize their own network. The purpose of this activity is to receive as many news as possible.

Fig. 12 is a demo running of the HubNet activity. In the screen there are three clients and one server. Because the news creation and spread rule is the same as NetLogo model, which means people do not know who will create the next news. It is impossible for student to connect to some kind of stable news creator. Besides, the server can limit the total number of links one student can form. This constrain avoid the case that students may want to link to every other students and eventually the network will become fully connected.

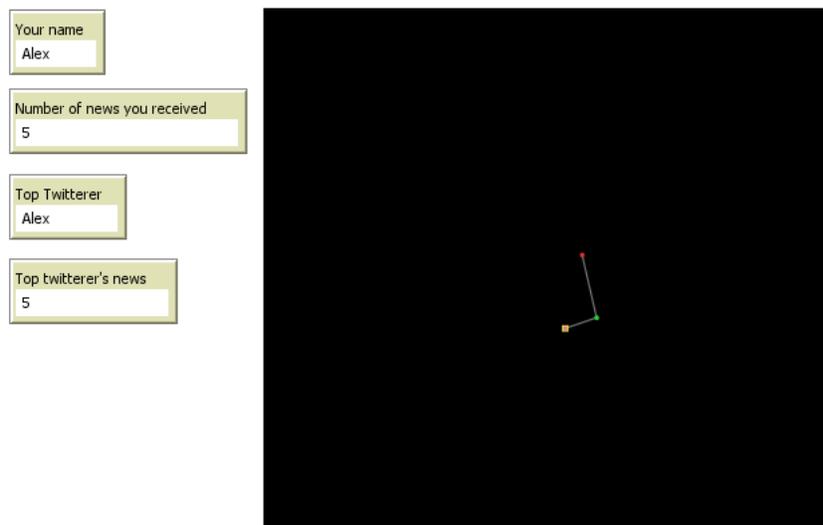


Fig. 13 Client window

One possible scenario for this activity is that, the network will form some cluster. To receive the maximum amount of news, students try to connect to different cluster. However, this behavior will break the established cluster. The interesting result is the network is under a dynamically changing condition.

7. Limit and Future Work

The main limitation in this simulation model is that we didn't introduce the real SNS data. All the network topology comes from idea situation and algorithm growth. Although our model is still representative. It is different from real situation.

8. Conclusion and inspiration

In this paper we built a Social Network Site model and investigate the characteristic of nodes that will influence the spread of news. We found that the number of people received the news is proportional to the centrality and inverse proportional to the local cluster. We validate our result by comparing with other reference and previous works.

This result explained why celebrity's twitter can be seen by many people. First of all, they have many followers, which mean they have a compared large centrality. Secondly, their followers are quite diverse. So their local cluster coefficient is small. This result also gives us some inspirations. To expand our network, we should try to make connection with more people, and more important thing is that we need to meet different kinds of people from different regions with different backgrounds.

Besides, as we see from network growth simulation, as the network become more connected, the number of people received news is larger. As the time goes by, when more and more people have

access to the internet, our society can be more closely connected and the speed of news spread will be much faster.

Reference

- [1] Dechun Liu; Xi Chen, "Rumor Propagation in Online Social Networks Like Twitter -- A Simulation Study," Multimedia Information Networking and Security (MINES), 2011 Third International Conference on , vol., no., pp.278,282, 4-6 Nov. 2011
- [2] Barabási, A.L., R. Albert, and H. Jeong, Mean-field theory for scalefree random networks. *Physica A: Statistical Mechanics and its Applications*, 1999. 272(1-2): p. 173-187.
- [3] Suh, B.; Lichan Hong; Pirolli, P.; Chi, Ed H., "Want to be Retweeted? Large Scale Analytics on Factors Impacting Retweet in Twitter Network," Social Computing (SocialCom), 2010 IEEE Second International Conference on , vol., no., pp.177,184, 20-22 Aug. 2010
- [4] Stonedahl, Forrest, Rand, William M. and Wilensky, Uri, Evolving Viral Marketing Strategies (July 2010). GECCO 2010, Portland, Oregon, USA, July 7-11, 2010 ; Robert H. Smith School Research Paper No. RHS 06-133.
- [5] Joseph J. Pfeiffer , Jennifer Neville, Methods to Determine Node Centrality and Clustering in Graphs with Uncertain Structure, Fifth International AAAI Conference on Weblogs and Social Media. 9 pages, 4 Figures, 2 Apr 2011
- [6] Suh, B., et al. Want to be retweeted? Large scale analytics on factors impacting retweet in Twitter network. 2010. IEEE.