Physica A 392 (2013) 1090–1099

Contents lists available at SciVerse ScienceDirect

Physica A

journal homepage: www.elsevier.com/locate/physa

Assembly effect of groups in online social networks[☆]

W. Fan [a,*], K.H. Yeung [a], K.Y. Wong [b]

[a] *Department of Electronic Engineering, City University of Hong Kong, Hong Kong*
[b] *Macao Polytechnic Institute, Macao*

ARTICLE INFO

Article history:
Received 22 May 2012
Available online 23 November 2012

Keywords:
Online social networks
Virus spreading
Groups

ABSTRACT

Due to the popularity and growth of online social networks, security in these networks becomes a critical problem. Previous works have proved that a virus can spread effectively in social networks. In this paper, groups in social networks are studied. We notice that groups on social network services sites can assemble people with similar characteristics, which may promote virus propagation in these networks. After our analysis, it is found that the use of groups can shorten the distance among users, and hence it would cause faster virus spread. We propose a virus propagation model and simulate it in a group network to show the assembly effect of groups. Our result shows that even with only one random attack, a virus can still spread rapidly, and the direct contact among group members is the reason for fast spreading.

1. Introduction

Online social networks (OSNs) have become popular nowadays. According to Ref. [1], Facebook is the largest online social network site in the world. Facebook has more than 800 million active users and more than 50% of these users log on in any given day [2]. Some people even spend hours checking friends' recent statuses and playing games on the social network sites every day. But in recent years, it has been reported that OSN sites such as Facebook and MySpace were attacked by hackers [3,4]. Researchers also find that it is possible to create a malicious application on the Facebook platform to collect data for attacking purposes [1]. In our earlier papers [5,6], we propose virus propagation models to study virus propagation in Facebook. In these models, a virus spreads by sharing information among friends, and the propagation is found to be fast in networks having scale-free property. In our study, we also find that users sometimes would share different information with different people by using "groups". Members of a group usually care more about some particular things than other users. For instance, users who like a product may create a group and they will pay more attention to the information about this product than other people. This will accelerate information spreading, but also accelerate virus propagation. To analyze the effect of groups on virus spreading, an extended Susceptible-Infected-Susceptible (SIS) model is proposed and simulations on groups are performed in this paper.

For better study of virus propagation in OSNs' groups, we need to investigate the structure of groups first because network topologies have an influence on virus and information spreading. In a previous study [7], Centola evaluated the effects of network structure on behavior diffusion and found that the behavior spread faster in clustered-lattice networks than random networks. Some other studies found that a virus spreads faster in a network with higher average degree [8–10].

[☆] This work was supported by Macau Science and Technology Development Fund numbered 039/2010/A and City University Strategic Research Grant No. 7002734.

[*] Correspondence to: Department of Electronic Engineering, City University of Hong Kong, 83 Tat Chee Avenue, Kowloon, Hong Kong. Tel.: +852 27889869.
E-mail address: fanwei.fw@gmail.com (W. Fan).

0378-4371/$ – see front matter © 2012 Elsevier B.V. All rights reserved.
doi:10.1016/j.physa.2012.11.017

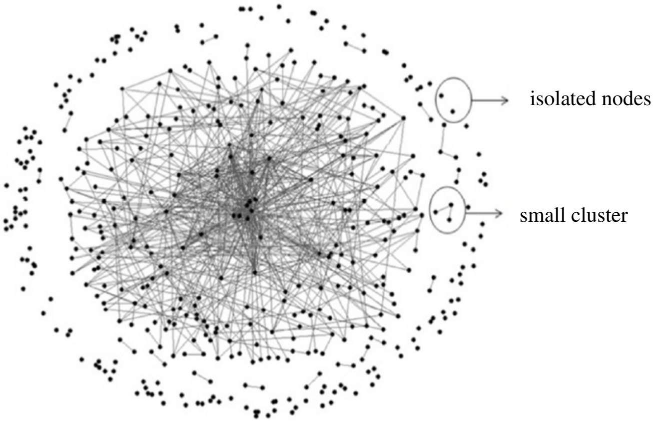

Fig. 1. An example of a group network. This group is randomly chosen from the groups in Facebook, showing $N = 445$.

Therefore, understanding the structure of OSNs is helpful for the study of virus spreading in these networks. Although it has been found that some OSNs have scale-free topology, such as MySpace, orkut, cyworld, and so on [11–13], the structure of groups is still unknown. Therefore, it is necessary to find out whether groups in OSNs have a scale-free topology or not.

In this paper, we will study the Facebook case. We designed a program to crawl over some groups on Facebook to investigate their structures. In Section 2, friendship in groups and factors which would affect groups' structure will be discussed. In Section 3, we will calculate the average distance among group users and compare the participation percentages of events in groups with events outside groups. We will see that groups have the assembly effect on users. In Section 4, we will propose an extended SIS model for virus propagation, which bases on information sharing and considers the latency period of infection. Virus behavior will be simulated in a group network. Moreover, we will investigate under what condition can the virus cause more infected users in the network. It is found that by attacking several random selected users or only one intentionally selected user, hackers can infect users much faster.

2. Group networks of OSNs

In OSNs, a group for a theme can be created to provide a communication platform for users interested on that theme to share information [14]. In the case of Facebook, groups can be created by anyone. The one who creates a group (creator) can control who can see their group. For the users who would like to join a group, they must be approved by the group's administrators, or added by other members. A user can join several groups that he/she is interested in, and he/she can also leave the groups if necessary. Members of a group can post photos, chat with others, share documents and schedule events. These activities are all available to all members in the group, and a member will receive notifications if there are new posts in the group in the default setting. It is reported that a user is connected to 80 community pages, groups and events on average [2]. Some members of these groups even become friends on the Internet although they do not know each other in real life.

In this section we discuss the collection of the data of users and their connections obtained from groups on Facebook. Due to Facebook's privacy control, we cannot obtain the completed member list if the size of a group is over 6000, so we only collect user connections of those groups with less than 6000 members. On the other hand, as some users set up some special settings to protect their privacy, we cannot find out these users' degrees or even their identities. Nonetheless, after studying 32 groups, we find that these users (with privacy control) account for 5%–15% of the whole group population. They are removed in our data collection. The data and characteristics of groups described in this section were collected in May 2010.

2.1. Group's scale-free topology

We plot an example of a group with $N = 445$ members on Facebook in Fig. 1. Nodes in this figure are users in this group, and an edge between two nodes means these two users are friends. In this paper, the term "friend" means that two users have established a relationship on a social networking site so that they will appear in each other's list of friends. In this group, 37% of the users are isolated individuals, which means that none of their friends is a member of this group. We place these nodes in the outer ring of the network shown in Fig. 1, whereas the nodes with high degrees are placed in the center of the circle. In this network, there is a large connected cluster which includes most nodes. But we also noticed that there are some nodes, composing small clusters which have only two or three connected nodes. These small isolated clusters are disconnected from most nodes, but only connected to other several nodes in the same cluster. So if there is an epidemic breaking out in the main connected network, the nodes of small clusters have more chances to survive.

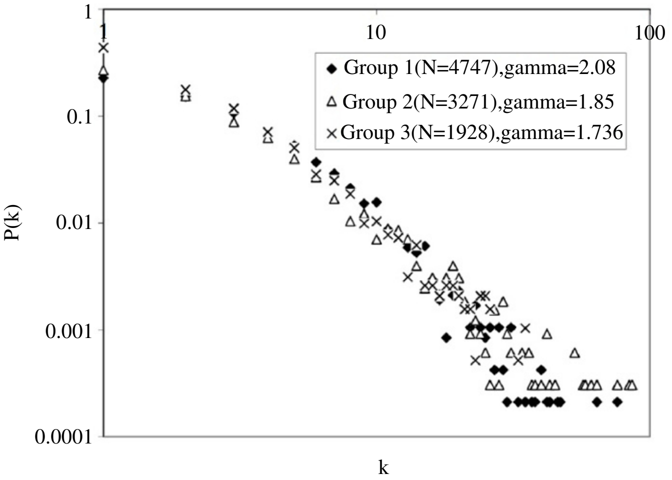

Fig. 2. Degree distributions of three selected groups of Faceboook. In these networks, a user's degree is defined as the number of his/her friends in the group.

Then we will study the degree distributions of group networks. In a group network, the nodes are the members in the group. The degree of a node in the group is defined as the number of the user's friends in the same group. For example, a user of Facebook has 10 friends on Facebook, so the degree of this user on the Facebook network is $k_{\text{Facebook}} = 10$. Suppose that the user has 2 of the 10 friends join a common group, Group A. So the degree of this user in the group network is $k_{\text{Group A}} = 2$, where $k_{\text{Group A}} \leq k_{\text{Facebook}}$. Our data of more than 30 groups in Facebook show that all these groups are scale-free networks, whose degree distribution follows power-law. In other words, the probability of any node in the network that has a degree of k is $P(k) \sim k^{-\gamma}$, asymptotically. As many researches have pointed out that OSNs are scale-free networks, our results are reasonable considering members of a group are also a part of the whole Facebook network. For clarity, we plot the degree distributions of three groups in Fig. 2. By regression, we can get the power-law exponent γ for each group. In our dataset, the exponent is in the region of $1 < \gamma < 3$.

2.2. Factors affecting group's topology

Though groups have similarities in topology, they have differences. In this part we will analyze two factors which would affect group's topology: size and purpose of the group. There are groups of different sizes on Facebook; for example, groups with only several members, or groups with thousands of members, and in Facebook, groups are created for different purposes. For example, a group can be created for the alumni of a university, and alumni can publish their latest events. Another group can be created for an artist, and the members of the group can discuss the news from this artist. In the following we will show the effect of these two factors respectively. While discussing one factor, the other factor will be kept as constant as possible.

Firstly we will study the influence of group size. Groups selected in this study are of one specific simple theme: they are created for the same Facebook application. We randomly choose an application named "Restaurant City" and search for groups for this application. The purpose of these groups is to get to know more users who like the same application and add them as temporary friends. To show the influence of group size, we select twenty such groups (out of hundreds) with different sizes. We plot the correlation of average degree of groups, power-law exponent and percentage of isolated nodes with group size in Fig. 3.

As Fig. 3 shows, while the size of the group increases the percentage of isolated nodes has a tendency to decrease. However, the average degree of a group network increases. It is because for these groups, larger size implies that members can get to know more people. When a user open the page of a group, randomly selected members of this group will be shown on top of the page. And one can view the list of members of the group, as well. So users have more opportunities to add new friends in a group with more members. If more pairs of members become new friends, the average degree of this group would be larger and insolated nodes would be less.

Secondly we will study the influence of group purpose. We observed that all those groups we studied above are for a particular application and one specific purpose—adding temporary friends. As applications on the Facebook platform are distinct from each other, groups for a different application may have a different purpose. To show how the purpose of the group affects the group's structure, we choose several groups of another application named "Country Story". The purpose of these groups is helping members add long-term friends. We label the new chosen application as B, while the application "Restaurant City" as A. Both applications A and B are popular on Facebook. Application A attracts more users than application B does, and A was available on the Facebook platform much earlier. We select two groups for each application. Then we

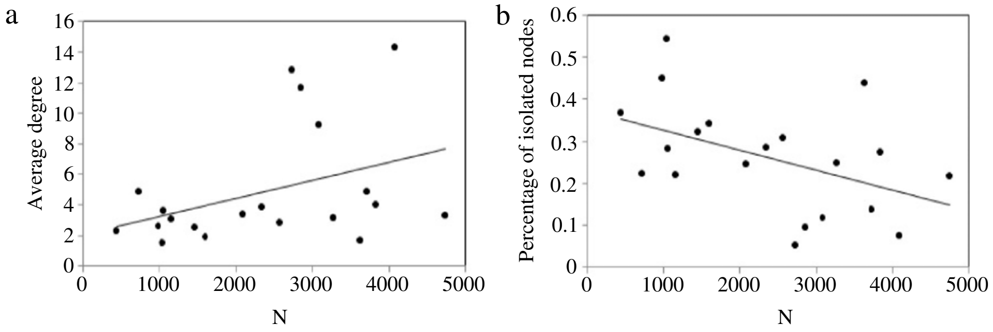

Fig. 3. (a) Correlation of average degree with group size. (b) Correlation of percentage isolated nodes with group size.

Table 1
Parameters of groups for different applications.

Group	Size N	Percentage of isolated nodes (%)	Average node degree
Group #1 (A)	3271	25	3.15
Group #2 (A)	1458	32	2.55
Group #3 (B)	2087	19	6.87
Group #4 (B)	646	19	4.84

compare their average degrees and percentage of isolated nodes, as listed in Table 1. It clearly shows that members in groups for application B are closer than those for application A, although application A is more popular and the groups of application A have larger sizes. In this paper, the term "close" means that there are lots of connections in a network and the percentage of isolated users is small. The reason for this result is that the purpose of group #3 and group #4 is to make members becoming long-term friends, while that of group #1 and group #2 is to add short-term friends.

We make several observations in this section. Firstly, it is found that groups of Facebook are scale-free networks, which is in accordance with the OSNs' scale-free topology. Since members of a group are approximately randomly taken from OSN, they can be seen as small social networks in OSN and help us understand the structure of the whole network. Secondly, we found that group networks are not connected networks as there are a considerable number of isolated individuals and small clusters. In the second part of this section we study two factors which would affect the structural parameters of a group. Our third observation is that groups of larger sizes tend to have a larger average degree and smaller insolated part, which implies that users of a larger group would have more interactions. Finally, the purpose of a group can also have an effect on the structure of the group: if a group is created to help members become friends, the connections in this group would be more than normal groups.

3. Comparison of network with groups and network without groups

As mentioned above, groups provide a platform for users of a common interest to share information. Therefore, it is more likely that information will spread faster and more widely in a network with groups than without groups. To verify this, we will study the assembly effect of groups in this section from two aspects: distance of groups and the participation percentage of events. Data in this section was collected in August 2011.

3.1. Distance of group network

While spreading information or a virus in a network, the information or virus must be passed from one user to another. Considering the case of a network without groups, a user can only disseminate information to his/her direct friends. Therefore, if a user, i, wants to share information with another user, j, where i and j are not friends, the information must go through the shortest path between them (from i to i's friend, to other users, to j's friend to j). Therefore, supposing that the members in a group can only contact others with the traditional friend-to-friend method, the average distance of a group is the average number of steps needed to share information between any two members. Let d_{ij} denote the distance between users i and j. Then the average distance d_i of a user i to other members in a group is

$$d_i = \left(\sum_{j \neq i} d_{ij}\right) \Big/ (N-1). \quad (1)$$

Then the average distance D of the group can be obtained as $D = \left(\sum_i d_i\right)/N$.

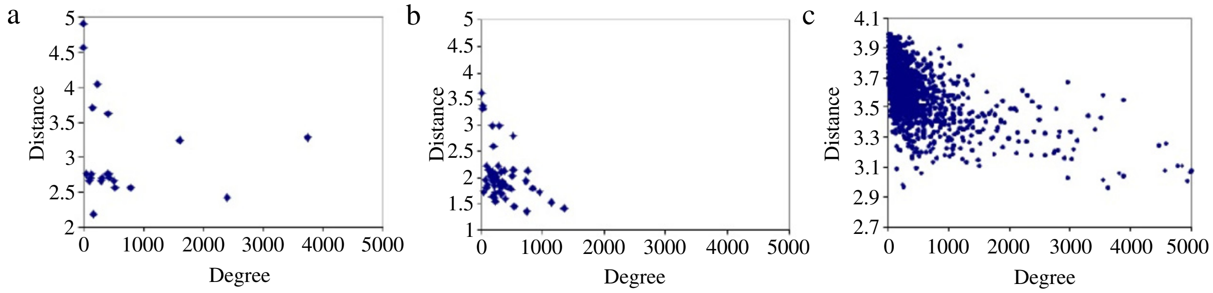

Fig. 4. Correlation between distance and degree of three selected groups. (a) Group #1 with 37 members, average distance D : 3.15, average degree: 587; (b) group #2 with 83 members, average distance D : 1.98, average degree: 347; (c) group #3 with 2345 members, average distance D : 3.57, average degree: 493.

Table 2
Participation and negligence percentage of events.

		Going	Maybe	Decline	Total	Participation (%)	Negligence (%)
Events created in groups	Event #1	6	4	8	62	9.7	70.9
	Event #2	11	1	5	71	15.5	76.1
	Event #3	675	59	112	1,509	44.7	43.9
	Event #4	322	29	29	1,210	26.6	68.6
	Event #5	334	216	838	1,844	18.1	24.7
Public events	Event #6	82,102	10,538	31,522	4,822,124	17.0	74.2
	Event #7	28	56	262	3,168	0.9	89.1
	Event #8	1,652	387	1,461	18,073	9.1	80.6
	Event #9	334	190	217	4,142	8.1	82.1
	Event #10	325	311	756	5,669	5.7	74.4

Fig. 4 shows the relation between distances d_i and degrees k_i for three selected groups. As shown in the figures, users with large degrees likely have small distances, and similarly, users with small degrees likely have large distances. Because a small distance means faster information spreading, these results show that members with large degree are key users to spread a virus in the case of virus spreading. Note that the average distances D we obtain from the three groups are all larger than 1.

However, in a network with groups, users can contact others using different methods. For example, users can view the member list of a group to learn new members and send direct messages to them, or create an event and invite all other members to attend it. It means that users can contact others with a distance of 1. Since the average distance of a group D is always larger than 1, the use of groups provides a faster way to spread information or a virus among group members.

3.2. Participation percentage of events

Any user of Facebook can create events and invite friends to join. If an event is created on the page of a group, all the members of that group will be automatically invited even they are not friends of the event creator, and everyone can see and join this event if it is public. Users who are invited would get notifications about this event, and users can select “going”, “maybe”, “decline” options or just ignore this event. In this part we will collect the data about the participation percentage of events, which would provide an insight into the probability that users respond to the information that they care about. Table 2 records the number of users that take different options of ten events. These events are randomly chosen from Facebook. Events #1–5 were created in groups and most guests are members of the corresponding groups. In addition, the topics of events #1–5 are closely related with their groups’ theme. Two of these five groups are for alumni of universities; two are for some digital products; one is for an application on Facebook. But the guests of events #6–10 are not members of specific groups. It is clearly shown that, in the events #1–5 the percentages of users who choose the “Going” option are much higher than those in events #6–10 on average, and the users who take no option in events #1–5 are lower. This is expected because the topics of the first 5 events are related with the themes of their groups, which their guests care about. On the contrary, the guests of the latter 5 events are widely chosen from the Facebook network, so the topics of these events can only attract a small portion of users.

In this section we studied the difference between a network with groups and a network without groups. We found that in groups users can contact other members with a distance 1, which is a much faster way than connecting others using the traditional friend-to-friend way. This feature will help information or virus spread faster and more widely. We also found that if an event is created in a group and its topic attracts most of the guests, the percentage of participation is higher than that in an event without groups (a public event which is not created in a group). This is because groups in OSNs can gather users with a common interest and most of the events in groups are what members care about. Information or a virus can also spread faster if this feature is made use of. Our study in this section verifies that groups have the assembly effect on information dissemination.

4. Model for virus propagation in group networks

We have investigated groups' scale-free topology and assembly effect, and found that groups can assemble users and provide a faster method for information spreading. We will then study how these features of groups would affect virus propagation in OSNs. Recently it has been reported that most existing viruses in OSNs are spread through Uniform Resource Locators (URLs) which direct users to malicious websites or malicious applications in Facebook. When a user receives the links and clicks them, his/her computer will be infected and malicious links will be sent to his/her friends. Sometimes, these malicious URLs are shortened with a URL shortening service [15] to confuse people, making them click the malicious links more easily. In this section we will propose a virus propagation model which is based on the viral form of malicious URLs.

We will apply our model in group networks, in which all members are supposed to have the common interest. The data in Section 3.2 shows that the participation ratio of a group event is much higher than that of a public event. We assume that if a hacker posts a malicious URL with a title which seems attractive to the members of a group, the probability that members click this URL is consistent with the participation percentage of events. In our simulations, data that we obtained in Section 3 will be used.

Our model is as follows.

Suppose that a group has N nodes and they are numbered from 1 to N. Before the propagation, all nodes are susceptible. The first step of this model is to select a number of initial infected users who will send out a malicious URL in the beginning. We assume that a virus spreads from step t_0 and the number of initial infected users is I_0 at time step t_0.

In the traditional Susceptible-Infected-Susceptible (SIS) epidemic model, a susceptible user will become infected with probability β at each time step after he/she receives malicious URLs from friends. These infected nodes would turn susceptible again with probability 1 in the next time step. Based on the analysis in Refs. [16,17], we denote the fraction of infected nodes with degree k at step t by $y_k(t)(0 \leq y_k(t) \leq P(k))$ and the fraction of susceptible nodes with degree k at step t by $x_k(t)$, where $x_k(t) + y_k(t) = P(k)$. The time evolution of our model is given by

$$\partial_t y_k(t) = -y_k(t) + \beta k[P(k) - y_k(t)]\Theta(y(t)). \tag{2}$$

The variable $\Theta(y(t))$ is the probability that an edge connects to an infected user and we have

$$\Theta(y(t)) = \left[\sum_k k y_k(t)\right] \Big/ \langle k\rangle . \tag{3}$$

The denominator $\langle k\rangle$ is the average degree of the network. Eq. (2) means that the fraction of infected nodes with degree k is proportional to the infection probability, number of friends, and $\Theta(y(t))$. We can evaluate the fraction of infected nodes using

$$\partial_t y(t) = \sum_k \partial_t y_k(t) = -y(t) + \beta\left[\langle k\rangle - \sum_k k y_k(t)\right]\Theta(y(t)). \tag{4}$$

However, not all the users stay online when they receive the malicious URLs, there may be a period before they see and decide whether to click the links, like the latency of epidemics. Therefore, we extend the SIS model to consider this behavior. We assume that p_1 of users who receive links will see them immediately while $1 - p_1$ of users will see these links one more time step later. At time step t the fraction of users who have received malicious URLs but have not seen them yet is $z(t)$. Users who see the URLs will click them with a probability β then become infected. Eq. (4) can be rewritten as

$$\partial_t y(t) = -y(t) + \beta z(t) + p_1\beta\left[\langle k\rangle - \sum_k k y_k(t) - \sum_k k z_k(t)\right]\Theta(y(t)). \tag{5}$$

The model we proposed is for virus spreading in a group, or it can be seen as the message dissemination in a group. We suppose that there is a group with all the members interested in its theme while users outside this group are not interested in it, so we do not consider the paths outside the group in this section. In our simulations, we recorded the number of infected users at every time step to study the virus evolution. In the first two parts of this section, simulations are based on the traditional friends' network in a group. In the third part, the virus spreading in a network with distance of 1 are considered.

4.1. Virus propagation with extended model

The density of infected users $y(t)$ and the density of users who receive malicious URLs but have not seen them in time $z(t)$ are shown in Fig. 5. It can be seen that both $y(t)$ and $z(t)$ rise rapidly, and their values increase while the infection probability increases. The curves of density of infected users show the density of new infected users at time step t in the

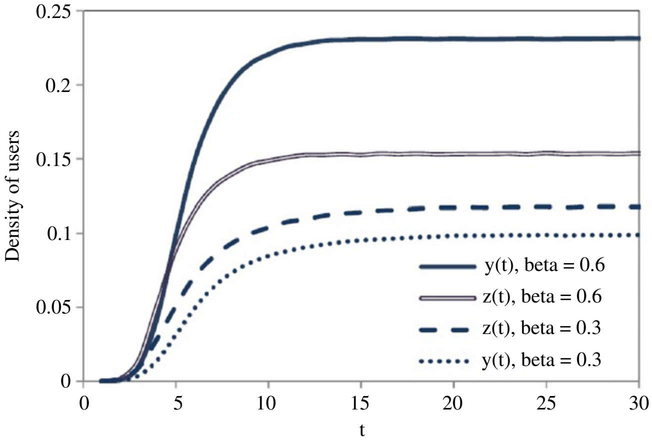

Fig. 5. Virus propagation with extended SIS model. Data is averaged over 1000 runs. In this simulation, $I_0 = 1$.

extended SIS model, not the accumulative infected users since propagation begins. In the beginning of virus propagation, we have $z(0) = 0$. We label the initial infected user by ini and $\Theta = k_{\text{ini}}/\langle k\rangle N$ if $I_0 = 1$, then we obtain

$$y(1) = \frac{p_1\beta}{N}k_{\text{ini}}\left(1 - \frac{k_{\text{ini}}}{N\langle k\rangle}\right). \tag{6}$$

In common cases, we have $k_{\text{ini}} \ll N\langle k\rangle$, so

$$y(1) \approx \frac{p_1\beta}{N}k_{\text{ini}}. \tag{7}$$

And we can obtain the $z(1)$

$$z(1) = \frac{1-p_1}{p_1\beta}y(1) \approx \frac{1-p_1}{N}k_{\text{ini}}. \tag{8}$$

In our simulation $p_1 = 0.5$, so the value of $z(t)$ is larger than $y(t)$ in the beginning of propagation. But after a few steps, they will become smaller or stay larger due to the different values of infection probability.

In this part our result shows that if the value of infection probability β increases, the percentage of infected users will also increase. We assume that if hackers post a malicious URL in a group, the title of this URL can attract most members of this group. The infection probability would be consistent with the participation ratio of a group event, which is higher than a public event. Therefore, the virus will spread faster in a network with groups than in a network without groups.

4.2. Virus propagation under different initial conditions

In the above simulations, we have $I_0 = 1$, and the initial infected user is randomly selected. Thus we have about half of the possibility to choose a node with no friend and a node in a small isolated cluster. If the initially infected user is in such a small isolated cluster, the malicious URL will not be transferred to most nodes of the network. However, in normal cases, hackers prefer to attack an important user. In the following, we will intentionally choose a user ini to be the initial infected user. We select ini from the users whose degrees k_{ini} are larger than a specific value. We will consider cases that use the initial infected user with $k_{\text{ini}} > 0, 1, 3, 5$, respectively. We will also consider the case that uses the node having a maximum degree $k_{\max} = 199$ as ini for comparison. The results are shown in Fig. 6. As can be seen, the virus spreads faster if we attack the user with a higher degree. This can also be partly explained with Eq. (7): if the value of k_{ini} increases, the virus will spread faster in the beginning.

Then we will assume $I_0 > 1$ to find out how the numbers of initial infected users affect the virus's spread. In this case, we have $\Theta = (k_{\text{ini}\,1} + k_{\text{ini}\,2} + \cdots + k_{\text{ini}\,I_0})/\langle k\rangle N$ in the beginning of the attack, which yields

$$y(1) \approx \frac{p_1\beta}{N}(k_{\text{ini}\,1} + k_{\text{ini}\,2} + \cdots + k_{\text{ini}\,I_0}) \tag{9}$$

while I_0 is a small number. So a larger I_0 would result in the growth of the spreading speed. Fig. 7 plots the virus spreading processes with different I_0. We can see that as the I_0 increases, the virus spreads faster, and the number of infected users at each time step also increases. But the difference is not significant if we have $I_0 \geq 3$.

In this part we have studied the effect of the initial condition on virus spreading. It is found that a network can be easily hacked by attacking several random selected users, or by attacking only one well-connected user. In both cases, the virus can spread faster.

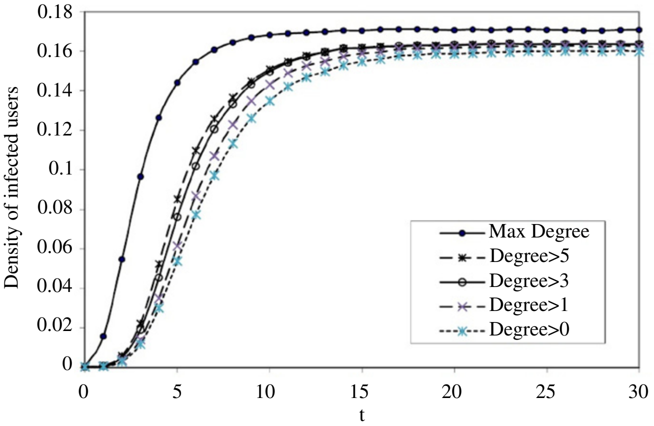

Fig. 6. Virus propagation processes with initial infected user of special degree. Data is averaged over 1000 runs. In this simulation, $I_0 = 1$.

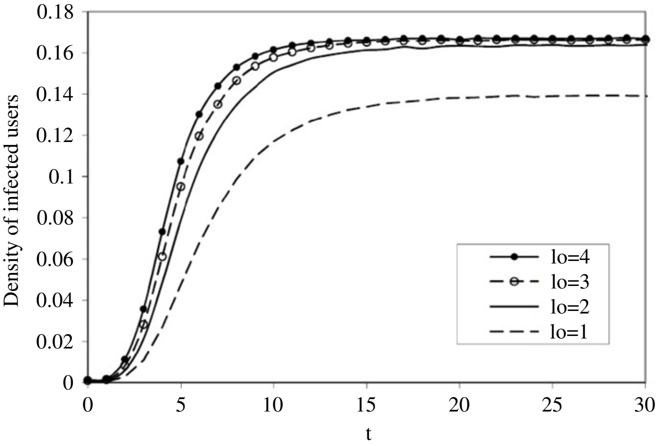

Fig. 7. Virus propagation processes under different initial conditions. Data is averaged over 1000 runs. In this simulation, $I_0 = 1, 2, 3, 4$.

4.3. Virus propagation in networks with different connections

As mentioned in Section 3, users of groups can contact all the members in groups with distance 1. In this case, $\langle k \rangle = N - 1$ and $P(\langle k \rangle) = 1$, so the value of Θ is $I_0(N-1)/\langle k \rangle N \approx I_0/\langle k \rangle$. Consequently, at the beginning of the infection, we have

$$y(1) \approx p_1 \beta I_0 \left(1 - \frac{I_0}{N-1}\right). \tag{10}$$

In common cases, $I_0 \ll N$, Eq. (10) can be written as

$$y(1) \approx p_1 \beta I_0. \tag{11}$$

Comparing with Eq. (7), the term on the right-hand side in Eq. (11) is larger. So users in a network with a distance of 1 will be more easily attacked. We compare these two situations in Fig. 8. We assumed that if one user is infected, all the members of this group will receive malicious links. As shown in the figure, the virus spreads much faster and broader than the traditional way. There are oscillations on the curve of the group with distance of 1. The reason is that in the beginning of the propagation, the number of infected users rises rapidly, causing the number of susceptible users to decrease quickly. So the number of infected users for the next time step will decrease, which will make the number of susceptible users increase conversely. As a result, the curve has oscillations in the beginning.

Based on the results obtained in this section, we can conclude that there are two plans if someone wants to attack members of a group via the traditional friend-to-friend way. The first plan is that: first create a new account, join an existing group, and then add some members in the group as friends and send malicious URL to them. The second plan is to attack several members in this group, and these members can be selected randomly. Besides these two methods, it will be much easier to spread a virus by exploiting the assembly effect of group: to join the group and send out a malicious URL which is disguised as some information that members of this group may be interested in.

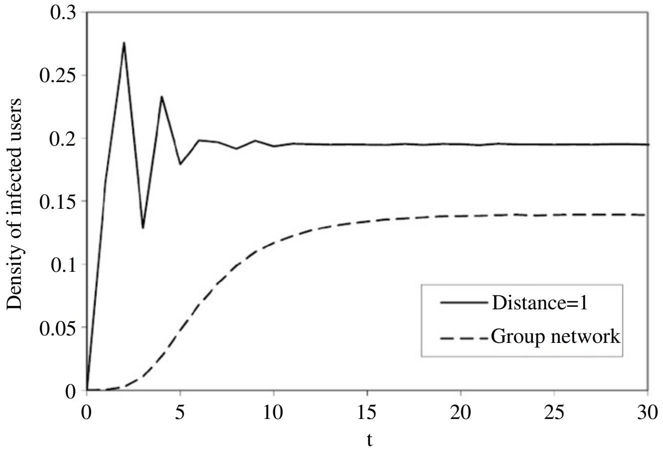

Fig. 8. Virus propagation processes in different ways. Data is averaged over 1000 runs. In this simulation, $I_0 = 1$.

However, it is difficult for users to protect themselves from this kind of attack. It may help if one does not install applications or add strangers as friends. However, if the friends of a user are infected, he/she will receive their sharing information, including those malicious ones. On the other hand, it is impossible to ignore all the posts to prevent from receiving any virus or from being attacked; however, doing this will lose the fun of using online social networks. It is necessary to be careful when enjoying online social networks, but at the same time the service providers of OSN should take more responsibility. If they can be more considerate when releasing new products or they can fix bugs in a timely manner, attacks on online social networks can be relieved.

5. Conclusion

In this paper, groups in Facebook were studied. We found that the group networks have power-law degree distributions, which is in good agreement with the scale-free topology of OSNs. Moreover, it is found that the sizes of groups can affect the average degree, and members of a bigger group are more active. We collected the data of groups for different applications, and learned that the purpose of a group determines how well the group is connected.

To verify the assembly effect of groups in virus propagation, we study the virus spreading with different network connections and infection probabilities. In a group, any member can contact all other members with a distance of 1, even if these members are not his/her friends. In this case, information or a virus would spread faster than in the traditional friend-to-friend network. Another aspect of the assembly effect of group is the high participation percentage of group events, the values of which were used in our simulations. Based on the results obtained in Section 3 we realize that the interaction among users can be dangerous if it is utilized by hackers. We proposed an extended SIS virus propagation model in Section 4, based on the behaviors of sending malicious messages, and we used it to model the virus spreading in a group. Our result showed that the virus could spread easily, by attacking several randomly selected users, or attack only one intentionally selected user. We demonstrated that even in a traditional friend-to-friend network, a higher infection probability would help a virus spread faster. Our study in this paper found that groups make users of a social network closer. It will help information spreading, but at the same time, it will also give a chance for hackers to spread a virus effectively.

References

[1] C. Patsakis, A. Asthenidis, A. Chatzidimitriou, Social networks as an attack platform: Facebook case study, in: Networks, 2009, ICN'09, Eighth International Conference, 2009, pp. 245–247.

[2] http://www.facebook.com/press/info.php?statistics#/press/info.php?statistics.

[3] Facebook users suffer viral surge. http://news.bbc.co.uk/2/hi/technology/7918839.stm.

[4] Malicious code targeting social networking sites users. http://www.us-cert.gov/current/archive/2009/03/04/archive.html#malicious_code_targeting_social_networking.

[5] W. Fan, K.H. Yeung, Virus propagation modeling in Facebook, in: Asonam, 2010 International Conference on Advances in Social Networks Analysis and Mining, 2010, pp. 331–335.

[6] W. Fan, K.H. Yeung, Online social networks—paradise of computer viruses, Physica A 390 (2011) 189–197.

[7] D. Centola, The spread of behavior in an online social network experiment, Science 329 (5996) (2010) 1194–1197.

[8] C.C. Zou, D. Towsley, W. Gong, Email virus propagation modeling and analysis, Tech. Rep. TR-03-CSE-04, Umass ECE Dept., May 2003.

[9] C.C. Zou, D. Towsley, W. Gong, Email worm modeling and defense, in: Proc. 13th Int. Conf. Computer Communications and Networks, ICCCN'04, October 2004, pp. 409–414.

[10] C.C. Zou, D. Towsley, W. Gong, Modeling and simulation study of the propagation and defense of Internet e-mail worms, IEEE Transactions on Dependable and Secure Computing 4 (2) (2007) 105–118.

[11] A. Mislove, M. Marcon, K.P. Gummadi, P. Druschel, S. Bhattacharjee, Measurement and analysis of online social networks, in: Proceedings of the 7th ACM SIGCOMM Conference on Internet Measurement, IMC'07, 2007.

[12] R. Kumar, J. Novak, A. Tomkins, Structure and evolution of online social networks. in: Proceedings of the 12th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, KDD'06, 2006.

[13] Y. Ahn, S. Han, H. Kwak, S. Moon, H. Jeong, Analysis of topological characteristics of huge online social networking services, in: Proceedings of the 16th International Conference on the World Wide Web, WWW'07, 2007.

[14] http://www.facebook.com/groups.

[15] A yet another malicious Facebook app: father crashes and dies.
http://www.securelist.com/en/blog/6115/Yet_another_malicious_Facebook_app_Father_crashes_and_dies.

[16] R. Pastor-Satorras, A. Vespignani, Epidemic dynamics and endemic states in complex networks, Physical Review Letters 86 (2001) 3200 [MEDLINE]; Physical Review E 63 (2001) 066117.

[17] R.M. May, A.L. Lloyd, Infection dynamics on scale-free networks, Physical Review E 64 (2001) 066112.