

Information Propagation in Social Networks with Overlapping Community Structure

Narisa Zhao, Xiaojun Liu

Institute of Systems Engineering, Dalian University of Technology
Dalian, 116024 -PRC.

[e-mail: liuxiaojun@mail.dlut.edu.cn]

*Corresponding author: Xiaojun Liu

*Received September 15, 2016; revised August 7, 2017; accepted August 13, 2017;
published December 31, 2017*

Abstract

Many real networks exhibit overlapping community structures. Recent studies have been performed that analyze the impact of overlapping community structure on information propagation, but few of them concerned with individual behaviors. From this point of view, we propose a Markov process model to evaluate the performance of information propagation in social networks with overlapping community structures. In addition, many individual social behaviors are combined in the model. For example, individuals may exhibit selfish behaviors, such as individual and social selfishness, and people may discard the information after they have used it. The accuracy of the model is verified by simulation. Furthermore, the numerical results show that both overlapping community structure of the network and individual behaviors have a significant impact on the outbreak size and propagation speed of the information. Additionally, the overlapping community structure of the social network can reduce the impact of selfishness on information propagation.

Keywords: Social network, Information propagation, Overlapping community, Individual behaviors

This research was supported by the General Program of Natural Science Foundation of China (Grant No. 61471083), Key Program of Natural Science Foundation of China (Grant No. 71431002), and Humanities and Social Sciences Research Program of the Ministry of Education of China (14YJA630044).

1. Introduction

With the development of network and mobile phone technology, online social media such as Twitter, Facebook, and Wechat has become one of the most important ways for people to share content and communicate with friends. Users of social media create online social networks of friends. As a result of their popularity, information propagates more quickly and widely in social networks than real society. They are exploited to propagate news, adopt technology and promote products. Given the importance of social networks in our society, they have attracted much attention from researchers in recent years; for instance, many researches have analyzed information propagation in microblogs [1,2], Facebook, and Twitter[3].

The propagation mechanism is an important research problem for information propagation in social networks. In order to gain a better understanding of the propagation mechanism, many researches have been devoted to the modeling problem. Some researches have focused on information dynamic evolution [4,5]; these models describe the propagation process and then predict the propagation speed and range of information. Many popular epidemic dynamic models evolved from disease infection models. They assume that information propagates from an infected individual to a susceptible individual through the link between them [6]. The simplest models are the susceptible-infected-susceptible (SIS) model [7] and the susceptible-infected-remove (SIR) model [8]. The epidemic dynamic model has been widely used to analyze information propagation in social networks. For instance, some studies developed a model for rumor propagation in social networks [9], and some use epidemic dynamic routing in mobile social networks [10]. In this paper, we also assume information propagation in social networks in an epidemic dynamic way.

In addition to propagation mechanisms, the network structure is also very important for the process of information propagation. Many studies have proved that network structures (power-law degree distribution, small-world property, hierarchical clustering effect, etc.) are important properties that influence information propagation in social networks [11-13]. For example, studies [14-16] have found that epidemics propagate on scale-free networks more easily than on random and small-world networks; research [17] indicated that information propagates more slowly in social networks with a higher clustering coefficient. Furthermore, many studies have shown that individuals can form different communities based on social ties (similar interests, occupation, age, etc.) [10]. Individuals are densely connected in the same community and are sparsely connected in different communities. Many algorithms have been developed to uncover community structure in social networks [18]. Moreover, theoretical network models have been proposed to describe community structure in various forms [19]. Community structure, which is also a special structure of social networks, plays an important role in information propagation [20].

Previous studies mostly focus on information propagation in social networks with non-overlapping communities. However, users may belong to more than one community. For instance, users may participate in many groups simultaneously, for they have different scientific activities or personal lives (school, hobby, family) in social networks [21]. This implies that social networks have overlapping community structures. Furthermore, research [22] established that overlapping community structures pervasively existed in empirical networks. Recently, there have been an increasing number of studies investigating the overlapping communities of networks. Most are focused on detecting overlapping

communities in the networks [21]. A small number of studies examined epidemic propagation in networks with overlapping communities. For example, research [23] studied contagion diffusion in network characterized by a Watts-Strogatz ring lattice with an overlapping community structure. Research [7] studied this problem on a scale-free network model. But their level of overlapping communities is fixed. Research [24] investigated the epidemic process on a scale-free network model with different levels of overlapping communities. They all illustrate the important role of overlapping communities in the process of epidemic propagation.

However, most studies on overlapping communities focus on epidemic diffusion and only consider the topology of the network and do not study information transmitting in social networks based on individual social behaviors. Many studies have shown that users exhibit selfish behavior [10,25]. In particular, selfish behaviors can be divided into two classes, individual selfishness and social selfishness. Individual selfishness means that users may not be willing to forward information to other users when they receive information, while social selfishness means users are more willing to forward information to users in the same community than users in different communities [10]. Research [25] explored the impact of selfish behaviors on epidemic routing in mobile social networks with overlapping communities. However, their model only allows for networks with two communities. It is necessary to extend this study to more realistic situations. On the other hand, they only consider the impact of selfish behaviors, but our work considers more behaviors. As research [10] showed, users may discard information after receiving when they think it is no longer of any use. Additionally, users may not be interested in the information at present but may become interested in it in the future. These social behaviors may have a certain impact on information propagation.

In this paper, we try to explore the performance of information propagation in social networks with overlapping community structures and the above social behaviors. For this purpose, we first propose a theoretical framework based on a Markov process model considering both overlapping community structure of network and social behaviors. We divide the community into overlapping and non-overlapping parts and present it using ODE equations that combined these behaviors. Then we check the accuracy of the model through simulation. Moreover, experimental results show that all of the above behaviors and overlapping community structures of networks have certain impacts on the outbreak size and propagation speed of the information. Furthermore, the impact of selfishness interacts with the community overlap level. With an increase in the overlap level, the impact of selfishness can be reduced.

The rest of the paper is organized as follows. In section 2, we give an overview of related works. Section 3 presents the information propagation model with overlapping communities. Simulations and experimental results are shown in section 4. Finally, we conclude this paper briefly in section 5.

2. Related Work

Many studies have investigated the spread of epidemics in networks. Recent studies have focused on the impact of community structure on propagation. Research [26] found that the degree of community has a significant influence on epidemic spreading efficiency. Research [27] used empirical and simulated networks to investigate the spread of disease. Their results demonstrated that community structure strongly affects disease dynamics. Research [10] presented a theoretical framework to evaluate the impact of social behaviors on the

performance of information propagation in mobile social networks with multiple communities. Research [28] illustrated that the connection mixing style significantly impacted epidemic spreading in weighted networks and that the increase in the number of communities amplifies the effect of mixing style. Research [29] demonstrated that the propagation performance of a meme can be predicted by its early community concentration. They found that the more communities a meme permeates, the more it spreads. Research [30] showed that strong communities can facilitate global diffusion by enhancing local, intercommunity spreading. But the communities in these works are all assumed to be non-overlapping and few of them consider social behaviors.

Several works studied information propagation in networks with overlapping community structures. Research [23] studied diffusion in networks with overlapping community structures. Through simulation of the SIR model, they found that contagions spread fast in networks with overlapping community structures, and short paths exist in these networks. Research [7] proposed a network model with overlapping community structure, and the degree distribution of the network follows a power law. By simulating the SIS epidemic spreading process in the generated networks, they found that overlapping community structure has a significant impact on the infection prevalence and spread velocity in the early stages of the infection process. Research [25] proposed a Markov process model to describe the performance of epidemic routing in mobile social networks with overlapping communities, allowing networks with two communities. They explored the impact of selfish behaviors based on simulation and found that selfishness makes the performance worse. Research [31] classified vertices as overlapping and non-overlapping. By simulating epidemic models on both synthetic and real-world networks, they found that overlapping vertices play a vital role in the spread of epidemics across communities. Research [24] proposed a rewiring algorithm that can change the community structure from overlapping to non-overlapping while maintaining the degree distribution of the network. They experimented on the SIS epidemic process in the generated scale-free networks and real-world networks. Results showed that epidemics spread faster on networks with a higher level of overlapping communities, and the effect interacts with the average degree's effect. However, few works have addressed the impact of individual behaviors on information propagation in social networks with overlapping communities. Thus, we believe this work is worthy of investigation.

3. Theoretical Framework

3.1 Network Model

We assume that the nodes in the network can be divided into four types: uninterested, interested, satisfied and discarded. Uninterested nodes denote nodes that are not interested in the information, while interested nodes mean that the nodes are interested in the information but do not get it. Similarly, satisfied nodes denote that the nodes get the information, and discarded nodes mean that the nodes already have the information, discard it and would not propagate it any more.

Nodes in the network communicate with each other by social media, such as blog sharing communities, microblog services, and WeChat. Individuals have an opportunity to share information with others. For example, they follow one person or review one message, and then they have an opportunity to send the information to each other. We assume that the occurrence of communication contact between two nodes follows a Poisson process. Therefore, the inter-meeting time between two nodes conforms to an exponential distribution with the

parameter denoted by λ . Because our goal is to evaluate the impact of selfish behaviors, we assume that nodes have the same communication contact model. In fact, our model can be easily extended to cases with heterogeneous communication contact models.

The information propagation process is shown in Fig. 1; a satisfied node in the network can send the information to the interested node with certain probability when they have communication contact. Similar to the work [32], we assume that an uninterested node changes to interested and that satisfied nodes discard the information according to exponential distribution with parameter ρ and μ , respectively.

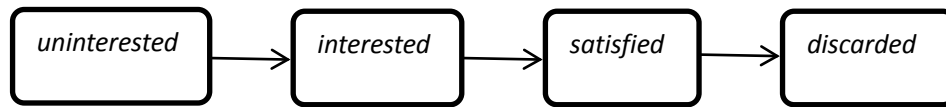


Fig. 1. Information propagation process

In our work, the network has a finite number N of individuals. At the beginning, only one satisfied node has the information. The network is divided into K communities due to some interests, and any two communities may overlap. That is to say, one node belongs to at most two communities, and there are C_K^2 overlap parts at most. Fig. 2 shows an example of a network with 3 communities. Due to the selfish nature, a satisfied node in j community helps an interested node in i community with probability p_{ji} . $1 - p_{ii}$ denotes individual selfishness, and $1 - p_{ji}$ denotes social selfishness.

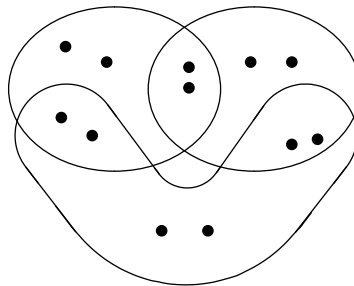


Fig. 2. Example of network with overlapping communities

3.2 Information Propagation Process

Now, we model the information propagation process based on the network model described in Section 3.1. We begin to give some definitions as follows. Let $X_i(t)$ denote the number of satisfied nodes in community i at time t and $MX_{ij}(t)$ denote the number of satisfied nodes belonging to community i and j simultaneously ($i, j \in [1, K]$, $MX_{ii}(t) = 0$, $MX_{ji}(t) = MX_{ij}(t)$). Similarly, $Y_i(t)$ and $MY_{ij}(t)$ denote the number of interested nodes; $Z_i(t)$ and $MZ_{ij}(t)$ denote the number of uninterested nodes; $D_i(t)$ and $MD_{ij}(t)$ denote the number of discarded nodes. In the above definitions, the nodes in community i contain nodes that belong to both communities i and j . A community can be divided into two components. One component is the non-overlapping part, in which nodes belong to one community only. Another component is overlapping parts including $K - 1$ overlapping parts, in which nodes

belong to two communities.

As shown above, interested nodes may become satisfied nodes. On the other hand, the satisfied nodes may discard information. Thus, for satisfied nodes in community i , given a small time interval Δt , we can obtain

$$X_i(t + \Delta t) = X_i(t) + \sum_{k \in \{Y_i(t) - \sum_j MY_{ij}(t)\}} \varphi 1_k(t, t + \Delta t) + \sum_j \sum_{k \in \{MY_{ij}(t)\}} \varphi 2_k(t, t + \Delta t) - \sum_{k \in \{X_i(t)\}} \sigma_k(t, t + \Delta t), \quad (1)$$

Symbol $\{Y_i(t) - \sum_j MY_{ij}(t)\}$ and $\{MY_{ij}(t)\}$ denote the set of interested nodes in non-overlapping parts and overlapping parts of community i at time t , respectively. $\{X_i(t)\}$ denotes the set of satisfied nodes. Additionally, $\varphi 1_k(t, t + \Delta t)$ denotes whether node k in the non-overlapping parts receives information in time interval $[t, t + \Delta t]$, while $\varphi 2_k(t, t + \Delta t)$ denotes whether node k in the overlapping parts receives information. If $\varphi 1_k(t, t + \Delta t) = 1$, we say node k in the non-overlapping parts receives information in time interval $[t, t + \Delta t]$, but node k does not receive information if $\varphi 1_k(t, t + \Delta t) = 0$. Similar means of $\varphi 2_k(t, t + \Delta t)$.

For an interested node k in community i , it may receive information from satisfied nodes in any community. First, when node k is in the non-overlapping part, there are $X_j(t) - MX_{ij}(t)$ satisfied nodes in community j , which are not in the same community as k (the set of these nodes marked as A), and every one of them may forward information toward node k . Because nodes encounter each other according to exponential distribution with parameter λ and taking into account that k receives information from a node in A with probability p_{ji} , k does not receive information from the nodes in A with probability $(1 - p_{ji}(1 - e^{-\lambda\Delta t}))^{X_j(t) - MX_{ij}(t)}$. Avoiding repeating calculation of overlapping parts, we can obtain

$$P(\varphi 1_k(t, t + \Delta t) = 1) = 1 - \prod_{j=1}^K (1 - p_{ji}(1 - e^{-\lambda\Delta t}))^{X_j(t) - MX_{ij}(t)} (1 - p_{gq}(1 - e^{-\lambda\Delta t}))^{-\sum_{g,q \neq i} MX_{gq}(t)} \quad (2)$$

Second, when node k in the overlapping part belongs to both community i and j , there are $X_g(t) - MX_{gi}(t) - MX_{gj}(t)$ satisfied nodes in community g , which are not in the same community as k (the set of these nodes marked as B). Therefore, node k does not receive information from the set of B with probability $(1 - p_{gq}(1 - e^{-\lambda\Delta t}))^{X_g(t) - MX_{gi}(t) - MX_{gj}(t)}$, $q = i$ or j , $p_{gq} = \max\{p_{gi}, p_{gj}\}$. Similarly, avoiding repeating calculation of overlapping parts, we have

$$P(\varphi 2_k(t, t + \Delta t) = 1) = 1 - \prod_{g=1}^K (1 - p_{gq}(1 - e^{-\lambda\Delta t}))^{X_g(t) - MX_{gi}(t) - MX_{gj}(t)} * (1 - p_{pq}(1 - e^{-\lambda\Delta t}))^{-\sum_{p,q \neq i,j} MX_{pq}(t)} (1 - p_{ii}(1 - e^{-\lambda\Delta t}))^{MX_{ij}(t)} \quad (3)$$

Additionally, $\sigma_k(t, t + \Delta t)$ denotes whether node k discards information in time interval $[t, t + \Delta t]$. Because the satisfied nodes discard information according to exponential distribution with parameter μ , we have

$$P(\sigma_k(t, t + \Delta t) = 1) = 1 - e^{-\mu\Delta t} \quad (4)$$

Combining with Eqs.(1)-(4), we can obtain the expectation as follows,

$$\begin{aligned}
 E(X_i(t + \Delta t)) &= E(X_i(t)) + (E(Y_i(t)) - \sum_j E(MY_{ij}(t)))E(\varphi 1_k(t, t + \Delta t)) \\
 &\quad + \sum_j E(MY_{ij}(t))E(\varphi 2_k(t, t + \Delta t)) - E(X_i(t))E(\sigma_k(t, t + \Delta t))
 \end{aligned} \tag{5}$$

Further, we have

$$\begin{aligned}
 E(\dot{X}_i(t)) &= \lim_{\Delta t \rightarrow 0} \frac{E(X_i(t + \Delta t)) - E(X_i(t))}{\Delta t} \\
 &= (E(Y_i(t)) - \sum_j E(MY_{ij}(t))) \lim_{\Delta t \rightarrow 0} \frac{E(\varphi 1_k(t, t + \Delta t))}{\Delta t} + \sum_j E(MY_{ij}(t)) \lim_{\Delta t \rightarrow 0} \frac{E(\varphi 2_k(t, t + \Delta t))}{\Delta t} \\
 &\quad - E(X_i(t)) \lim_{\Delta t \rightarrow 0} \frac{E(\sigma_k(t, t + \Delta t))}{\Delta t}
 \end{aligned} \tag{6}$$

By exchanging the order of expectation and exponential distribution [33,34], we can get

$$\begin{aligned}
 &\lim_{\Delta t \rightarrow 0} \frac{E(\varphi 1_k(t, t + \Delta t))}{\Delta t} \\
 &= \lim_{\Delta t \rightarrow 0} \frac{E(1 - \prod_{j=1}^K (1 - p_{ji}(1 - e^{-\lambda \Delta t}))^{X_j(t) - MX_{ij}(t)} (1 - p_{gq}(1 - e^{-\lambda \Delta t}))^{-\sum_{g,q \neq i} MX_{gq}(t)})}{\Delta t} \\
 &= \lambda (\sum_{j=1}^K p_{ji}(E(X_j(t)) - E(MX_{ij}(t))) - p_{gq} \sum_{g,q \neq i} E(MX_{gq}(t))),
 \end{aligned} \tag{7}$$

And the following equation

$$\begin{aligned}
 &\lim_{\Delta t \rightarrow 0} \frac{E(\varphi 2_k(t, t + \Delta t))}{\Delta t} \\
 &= \lambda \sum_j E(MY_{ij}(t)) \cdot (\sum_{g=1}^K p_{gq}(E(X_g(t)) - E(MX_{gi}(t)) - E(MX_{gj}(t))) \\
 &\quad - \sum_{p,q \neq i,j} p_{pq} E(MX_{pq}(t)) + p_{ii} E(MX_{ij}(t)))
 \end{aligned} \tag{8}$$

Then, we get the following ODE equation

$$\begin{aligned}
 E(\dot{X}_i(t)) &= (E(Y_i(t)) - \sum_j E(MY_{ij}(t))) \lambda (\sum_{j=1}^K p_{ji}(E(X_j(t)) - E(MX_{ij}(t))) - p_{gq} \\
 &\quad \cdot \sum_{g,q \neq i} E(MX_{gq}(t))) + \lambda \sum_j E(MY_{ij}(t)) \cdot (\sum_{g=1}^K p_{gq}(E(X_g(t)) - E(MX_{gi}(t)) - E(MX_{gj}(t))) \\
 &\quad - \sum_{p,q \neq i,j} p_{pq} E(MX_{pq}(t)) + p_{ii} E(MX_{ij}(t))) - \mu_i E(X_i(t))
 \end{aligned} \tag{9}$$

Similarly, for the satisfied nodes belonging to both community i and j , we can obtain the equation as follows,

$$MX_{ij}(t, t + \Delta t) = MX_{ij}(t) + \sum_{k \in \{MY_{ij}(t)\}} \varphi 2_k(t, t + \Delta t) - \sum_{k \in \{MX_{ij}(t)\}} \sigma_k(t, t + \Delta t) \tag{10}$$

Then, we get the following ODE equation

$$\begin{aligned}
 E(M\dot{X}_{ij}(t)) &= \lambda E(MY_{ij}(t)) (\sum_{g=1}^K p_{gq}(E(X_g(t)) - E(MX_{gi}(t)) - E(MX_{gj}(t))) \\
 &\quad - \sum_{p,q \neq i,j} p_{pq} E(MX_{pq}(t)) + p_{ii} E(MX_{ij}(t))) - \mu_i E(MX_{ij}(t)),
 \end{aligned} \tag{11}$$

For the interested nodes in community i , we have

$$Y_i(t + \Delta t) = Y_i(t) - \sum_{k \in \{Y_i(t) - \sum_j MY_{ij}(t)\}} \varphi 1_k(t, t + \Delta t) - \sum_j \sum_{k \in \{MY_{ij}(t)\}} \varphi 2_k(t, t + \Delta t) + \sum_{k \in Z_i(t)} \nu_k(t, t + \Delta t) \quad (12)$$

Symbol $\nu_k(t, t + \Delta t)$ denotes whether uninterested node k becomes an interested node in time interval $[t, t + \Delta t]$. Note that the process changes according to an exponential distribution with parameter ρ ; therefore,

$$P(\nu_k(t, t + \Delta t)) = 1 - e^{-\rho_i \Delta t} \quad (13)$$

Similar to Eq.(9), we can get

$$\begin{aligned} E(\dot{Y}_i(t)) = & -\lambda(E(Y_i(t)) - \sum_j E(MY_{ij}(t))) \cdot (\sum_{j=1}^K p_{ji}(E(X_j(t)) - E(MX_{ij}(t)))) \\ & - p_{gq} \sum_{g,q \neq i} E(MX_{gq}(t)) - \lambda \sum_j E(MY_{ij}(t)) \cdot (\sum_{g=1}^K p_{gq}(E(X_g(t)) - E(MX_{gi}(t))) \\ & - E(MX_{gj}(t))) - \sum_{p,q \neq i,j} p_{pq} E(MX_{pq}(t)) + p_{ii} E(MX_{ij}(t)) + \rho_i E(Z_i(t)) \end{aligned} \quad (14)$$

Using similar methods, we also obtain the following ODE equations

$$\begin{aligned} E(\dot{MY}_{ij}(t)) = & -\lambda E(MY_{ij}(t)) (\sum_{g=1}^K p_{gq}(E(X_g(t)) - E(MX_{gi}(t)) - E(MX_{gj}(t))) \\ & - \sum_{p,q \neq i,j} p_{pq} E(MX_{pq}(t)) + p_{ii} E(MX_{ij}(t))) + \rho_i E(MZ_{ij}(t)), \end{aligned} \quad (15)$$

$$E(\dot{Z}_i(t)) = -\rho_i E(Z_i(t)), \quad E(\dot{MZ}_{ij}(t)) = -\rho_i E(MZ_{ij}(t)), \quad (16)$$

$$E(\dot{D}_i(t)) = \mu_i E(X_i(t)), \quad E(\dot{MD}_{ij}(t)) = \mu_i E(MX_{ij}(t)). \quad (17)$$

Combined with Eqs. (9), (11), (14), (15), (16) and (17), we can get $N + C_K^2$ unknowns and $N + C_K^2$ equations. Therefore, we can get the value of these unknowns easily by the Matlab ODE suite.

4. Simulation and Experimental Results

In this section, we provide a performance evaluation. We try to explore how the network structure—overlapping communities and the social behaviors impact information propagation. Our main objective is to know the maximum number of nodes that have ever received the information and the propagation speed of information. For example, the businessman hopes to maximize the number of people who have ever received the advertisements. Thus, we use the metric $R(t)$ to evaluate the performance of the information propagation.

$$R(t) = \sum_{j=1}^K E(X_j(T)) + \sum_{j=1}^K E(D_j(T)) - \sum_{i=1}^K \sum_{j=i}^K E(MX_{ij}(T)) - \sum_{i=1}^K \sum_{j=i}^K E(MD_{ij}(T)) \quad (18)$$

The symbol $R(t)$ denotes the total number of nodes that have ever received the information at time t , including the satisfied nodes and discarded nodes. Obviously, the bigger $R(t)$ is, the wider the information propagation range will be.

4.1 Simulation result

In this subsection, we check the accuracy of our theoretical model by conducting simulations using Matlab simulator. We set $N=1800$ nodes, including $T1=300$ interested nodes and $T2=1500$ uninterested nodes at time 0. Without loss of generality, we evenly divided these nodes into 5 communities. That is, $K=5$, and there are $C_K^2=10$ overlapping parts.

We define community overlap level P , which is a probability denoting the percentage of nodes in the overlapping portion of the total nodes. We can obtain that every overlapping part has $M1=P \times T1 / C_K^2$ interested nodes and $M2=P \times T2 / C_K^2$ uninterested nodes. Similarly, we can obtain that every community has $S1=(1-P) \times T1 / N + (N-1) \times M1$ interested nodes and $S2=(1-P) \times T2 / N + (N-1) \times M2$ uninterested nodes.

In simulation, we set the parameter $P=0.5$. Therefore, every overlapping part has 90 nodes, and every community has 540 nodes (including interested nodes and uninterested nodes). We assume that only one node has the information at time 0, and this node is in the non-overlapping parts of the first community. Furthermore, we assume that nodes in one community send the information to nodes in other communities with probability 0.2 and nodes in the same community with probability 0.8. Additionally, we set $\lambda = 3.71 \times 10^{-6}$ [10], $\mu = \mu_i = 0.0002$, $\rho = \rho_i = 0.0001$, $1 \leq i \leq K$. Let maximal lifetime T increase from 1s to 50,000s, and we run simulations 10 times; the results are shown in Fig. 3.

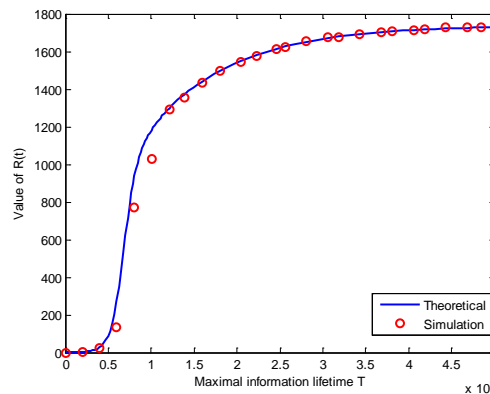


Fig. 3. Comparison between simulation and theoretical results

As shown in Fig. 3, the average deviation between the theoretical result and the simulation is very small, about 4.46%. This demonstrates the accuracy of our theoretical model. For this reason, we will only carry out the theoretical case for the performance analysis.

4.2 Impact of overlapping community

First, we will explore the impact of the overlapping communities on information propagation. From this subsection, we set total nodes $N=3000$, include 1 satisfied node, 499 interested nodes and 2500 uninterested nodes. Other settings are the same as that in the simulation. Fig. 4 a) shows the final outbreak size of information with a maximal information lifetime value increase from 0 to 10,000 and community overlap level $P=0.2$ and $P=0.8$. Fig. 4 b) shows the final outbreak size of information with value of P increase from 0 to 1 and a varying maximal information lifetime.

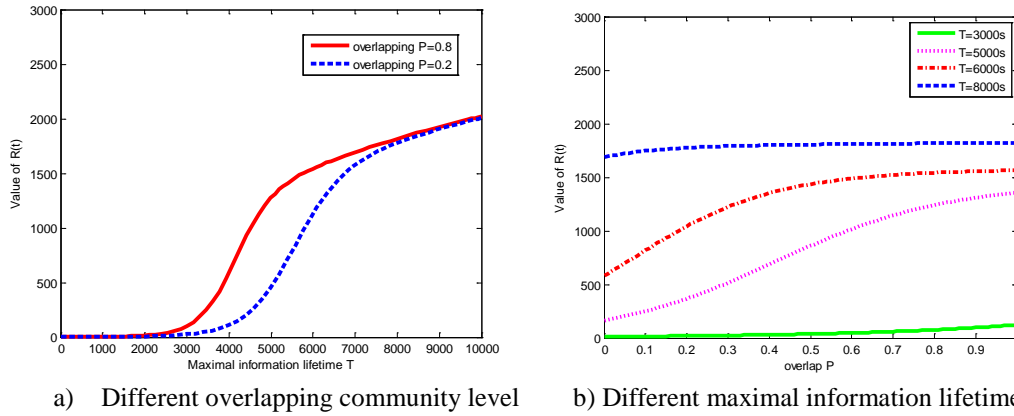


Fig. 4. Impact of the community overlap level

From the results shown in Fig. 4, we can know that the community overlap level impacts information propagation. We can see from Fig. 4 a) that the information propagates faster when the community overlap level P is bigger, and the outbreak size is almost the same when the maximal information lifetime is long enough. However, by comparing the trends of the four curves in Fig. 4 b), we find that there is some inconsistency in the influence of community overlap level with different maximal information lifetimes. The increase in speed of outbreak size with a higher maximal information lifetime is larger than that with a lower maximal information lifetime. But when $T = 8000s$, the value of $R(t)$ remains almost unchanged with an increase in the community overlap level P . Thus, the community overlap level affects the information propagation speed but does not impact the information outbreak size when the information lifetime is long enough. The overlapping community level impacts the information propagation speed because nodes in the overlapping parts have greater probability of sending the information to more nodes.

4.3 Impact of selfishness

Many works proved that selfishness influences information propagation. We will explore the impact of individual selfishness on networks with overlapping communities. We fix the community overlap level as $P=0.4$ and other settings are the same as those in the overlapping community analysis. Let individual selfishness vary from 0 to 1. We obtain Fig. 5 a). This shows the final outbreak size of information with an increase in individual selfishness from 0 to 1 and a varying maximal information lifetime. In order to analyze the case with different levels of individual selfishness, we assume social selfishness $p_1 = 1 - p_{ij} = 1, 1 \leq i, j \leq K, i \neq j$, which indicates information propagation within the community. Additionally, we fix $T = 10,000s$, and other settings are the same as those in the overlapping community analysis; we get Fig. 5 b). This shows the final outbreak size of information with value of P increase from 0 to 1 as well as varying individual selfishness $p_2 = 1 - p_{ii} = 0.2, 0.5, 0.7, 0.9$.

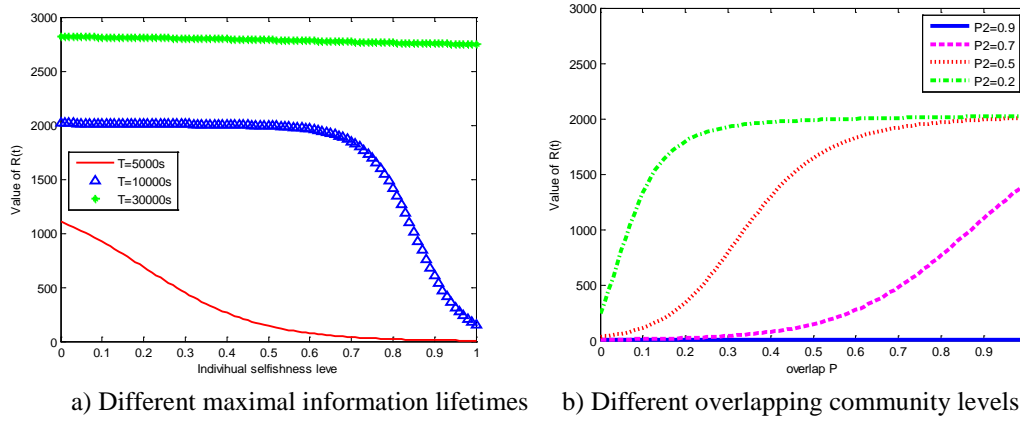


Fig. 5. Impact of individual selfishness

By analyzing the results shown in **Fig. 5 a)**, we know that the outbreak size decreases with the increase in individual selfishness. But there is some difference in the three curves. When the maximal information lifetime is $T=5000s$, the value of $R(t)$ changes quickly. When the maximal information lifetime is $T=10,000s$, the value of $R(t)$ changes slowly. However, when the maximal information lifetime is $T=30,000s$, the value of $R(t)$ is almost unchanged. Thus, individual selfishness impacts the outbreak size of information when the maximal information lifetime is not long enough.

From **Fig. 5 b)** we know that the outbreak size increases with the growth in community overlap level P for different levels of individual selfishness. Therefore, the overlapping community structure reduces the impact of individual selfishness. However, there is little inconsistency in the influence of community overlap level P for different levels of individual selfishness. When the individual selfishness is $p_2=0.9$, the value of $R(t)$ is almost unchanged. Thus, extreme selfishness influences information propagation severely. When individual selfishness is $p_2=0.7$, the value of $R(t)$ increases all the time with the growth in community overlap level P . Additionally, when individual selfishness is lower, the value of $R(t)$ increases at the beginning, while the value is almost unchanged in the later stages with the growth in the community overlap level, and the value of $R(t)$ is no more than 2000.

Now, we begin to explore the impact of social selfishness on information propagation in networks with overlapping communities. Similarly, we fix the community overlap level as $P=0.4$, and the other settings are the same as those in the overlapping community analysis. Let individual selfishness vary from 0 to 1. We obtain **Fig. 6 a)**. It shows the final outbreak size of information propagation with an increase in social selfishness from 0 to 1 and varying maximal information lifetime. In order to analyze the case with different levels of social selfishness, we assume individual selfishness $p_2 = 1 - p_{ii} = 0, 1 \leq i, j \leq K, i \neq j$, which means nodes in the same community show altruism toward each other, and social selfishness $p_1 = 1 - p_{ij} = 0.5, 0.7, 0.9$. Additionally, we fix $T = 5000s$, and the other settings are the same as those in the overlapping community analysis. We can obtain **Fig. 6 b)** for the outbreak size of information with an increase in community overlap level P from 0 to 1.

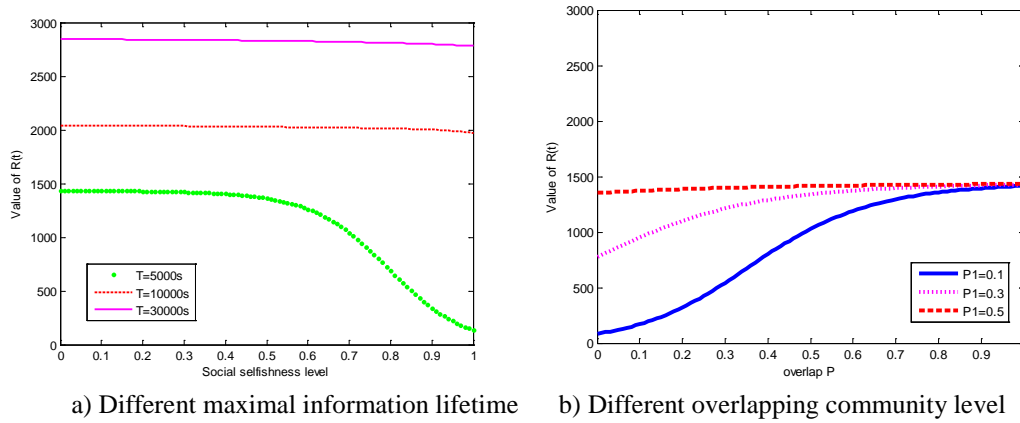


Fig. 6. Impact of social selfishness

As shown in **Fig. 6 a)**, outbreak size decreases with the increase in social selfishness when the maximal information lifetime is $T=5000s$. But, when the maximal information lifetime is $T=10,000s$ or $T=30,000s$, the value of $R(t)$ is almost unchanged. Thus, social selfishness has an impact on information propagation, but the impact on the outbreak size of information is weak when the maximal information lifetime is long enough.

From the results shown in **Fig. 6 b)**, we know that outbreak size increases with the growth in community overlap level P for different levels of social selfishness. The increase in speed of outbreak size with greater social selfishness is lower than that with less social selfishness. Additionally, the value of $R(t)$ remains unchanged with the growth in community overlap level when social selfishness is 0.5 and the value of $R(t)$ is no more than 1500.

The results illustrated in **Fig. 5** and **Fig. 6** can be explained by the impact of selfishness on information propagation in networks with overlapping communities and by the fact that an overlapping community structure can reduce the influence of selfishness; more nodes are in the overlapping parts and more nodes send information with higher probability. But the impact of the overlapping community structure on the information propagation is weak when the social selfishness level is small. The impact of selfishness plays a decisive role in information propagation.

4.4 Impact of the changed probability

As described above, the satisfied nodes discard information according to exponential distribution with parameter μ_i in community i . Now, we begin to explore the impact of these parameters. For simplicity, we set $\mu = \mu_i$ and maximal information lifetime $T = 50,000s$; the other settings are the same as those in the overlapping community analysis. **Fig. 7** shows the final outbreak size of information with an increase in the value of μ from 0.0001 to 0.005 and varying community overlap level P .

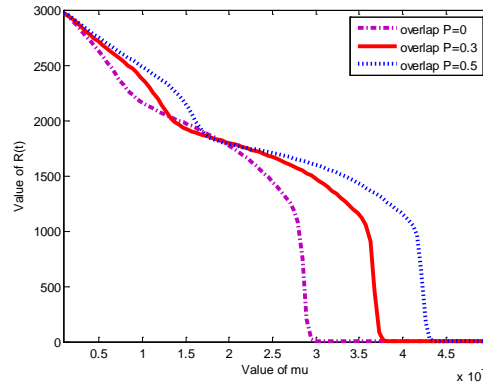


Fig. 7. Impact of the parameter μ

From the results shown in Fig. 7, we can see that the outbreak size decreases with the growth in parameter μ . Furthermore, we find that $R(t)$ equals about 1 as μ reaches a certain value. This means that none of the nodes can receive information other than the one received information at time 0. Additionally, the speed at which $R(t)$ reaches about 1 is slower when the community overlap level P is large. We know that the larger the μ , the higher the probability that the nodes discard the information. The information may disappear from the network by adjusting the parameters, and the overlapping community structure can reduce the influence of discarding.

In order to analyze the impact of the probability of uninterested nodes becoming interested nodes, we assume $\mu = \mu_i = 0.001$, $\rho = \rho_i, 1 \leq i \leq K$. We obtain Fig. 8 for the final outbreak size of information with a value of ρ increase from 0.0001 to 0.005 and varying community overlap level P .

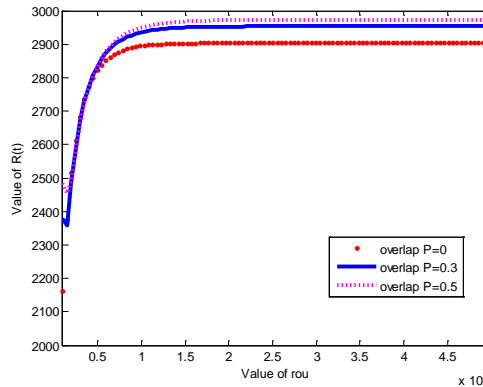


Fig. 8. Impact of parameter ρ

As shown in Fig. 8, the outbreak size increases with the growth of parameter ρ at the beginning and becomes steady as ρ reaches about 0.001. Furthermore, we find that the increase in speed of $R(t)$ with different community overlap level P is almost the same, but the final value of $R(t)$ is little changed. This result shows that uninterested nodes become interested nodes with larger probability, the information propagation will perform better, but

the variance is small. Additionally, $R(t)$ is large when the community overlap level P is large. This demonstrates that the overlapping community structure can help information propagation but with little correlation with parameter ρ .

5. Conclusion

Considering that many real social networks exhibit overlapping community structures in which nodes may belong to more than one community, in this paper, we presented a Markov process model to analyze information propagation in social networks with this structure and combined many individual social behaviors as well. The simulation results show that our model is very accurate. The experiments results show that the individual social behaviors have a certain impact on the performance of information propagation in networks with overlapping communities, and the impact of social behavior interacts with the impact of overlapping communities. First, the information propagates faster when the overlapping community level is bigger. Second, individual selfishness and social selfishness impact the outbreak size and propagation speed of information propagation in networks with overlapping communities, and the overlapping community structure can reduce the influence of selfishness. Third, people may discard information after they use it; when the discard probability is large enough, the information may disappear from the network, and the overlapping community structure can reduce the influence of discarded behavior. Finally, uninterested nodes become interested nodes with large probability, the information propagation will perform better, and the outbreak size is somewhat larger when the community overlap level is large. The result of our work may provide new ideas for investigating information propagation in online social media.

Some limitations of our study should be noted, which can be addressed in future studies. First, our study investigates information propagation only theoretically. In the future, we will collect a large number of data from real networks to conduct empirical studies and perfect our model. Second, we assume that one node belongs to at most two communities, but one node may belong to many communities in the real world. We plan to extend the model in future work. Nevertheless, our model is comparable to those in earlier studies on overlapping communities.

References

- [1] Y. Li, M. Qian, D. Jin, et al. "Revealing the efficiency of information diffusion in online social networks of microblog," *Information Sciences*, vol. 293, pp.383-389, 2015. [Article \(CrossRef Link\)](#)
- [2] C. Chen, P. Zhu, L. Chen, et al. "Modeling and Evaluating Information Diffusion for Spam Detection in Micro-blogging Networks," *KSII Transactions on Internet & Information Systems*, vol. 9, no.8, pp.3005-3027, 2015. [Article \(CrossRef Link\)](#)
- [3] F. Buccafurri, G. Lax, S. Nicolazzo, et al. "A model to support design and development of multiple-social-network applications," *Information Sciences*, vol. 331, pp.99-119, 2016. [Article \(CrossRef Link\)](#)
- [4] P. A. Grabowicz, J. J. Ramasco, V. M. Eguíluz. "Dynamics in online social networks," *Dynamics on and of Complex Networks*, Springer, vol. 2, pp. 3-17, 2013. [Article \(CrossRef Link\)](#)
- [5] R. Kumar, J. Novak, A. Tomkins, "Structure and evolution of online social networks," *Link Mining: Models, Algorithms, and Applications*, Springer, pp.337-357, 2010. [Article \(CrossRef Link\)](#)
- [6] C. Liu, Z. K. Zhang. "Information spreading on dynamic social networks," *Communications in*

- Nonlinear Science and Numerical Simulation*, vol. 19, no.4, pp. 896-904, 2014.
[Article \(CrossRef Link\)](#)
- [7] J. Chen, H. Zhang, Z. H. Guan, et al., “Epidemic spreading on networks with overlapping community structure,” *Physica A: Statistical Mechanics and its Applications*, vol. 391, no. 4, pp.1848-1854, 2012. [Article \(CrossRef Link\)](#)
- [8] X. Meng, Z. Li, X. Wang. “Dynamics of a novel nonlinear SIR model with double epidemic hypothesis and impulsive effects,” *Nonlinear Dynamics*, vol. 59, no. 3, pp. 503-513, 2010.
[Article \(CrossRef Link\)](#)
- [9] L. Zhu, H. Zhao, H. Wang. “Complex dynamic behavior of a rumor propagation model with spatial-temporal diffusion terms,” *Information Sciences*, vol. 349, pp.119-136, 2016.
[Article \(CrossRef Link\)](#)
- [10] Y. Wu, S. Deng, H. Huang. “Information propagation through opportunistic communication in mobile social networks,” *Mobile Networks and Applications*, vol. 17, no. 6, pp.773-781, 2012.
[Article \(CrossRef Link\)](#)
- [11] M. Draief. “Epidemic processes on complex networks,” *Physica A: Statistical Mechanics and its Applications*, vol. 363, no.1, pp.120-131, 2006. [Article \(CrossRef Link\)](#)
- [12] J. Z. Wang, Z. R. Liu, J. Xu. “Epidemic spreading on uncorrelated heterogenous networks with non-uniform transmission,” *Physica A: Statistical Mechanics and its Applications*, vol. 382, no. 2, pp.715-721, 2007. [Article \(CrossRef Link\)](#)
- [13] C. Peng, X. Jin, M. Shi. “Epidemic threshold and immunization on generalized networks,” *Physica A: Statistical Mechanics and its Applications*, vol. 389, no. 3, pp.549-560, 2010.
[Article \(CrossRef Link\)](#)
- [14] B. R. Pastorsatorras, A. Vespignani, “Epidemic spreading in scale-free networks,” *Physical review letters*, vol. 86, pp.3200–3203, 2001. [Article \(CrossRef Link\)](#)
- [15] R. Pastorsatorras, A. Vespignani, “Epidemic dynamics and endemic states in complex networks,” *Physical Review E*, vol. 63, pp.066117, 2001. [Article \(CrossRef Link\)](#)
- [16] C. Moore, M. E. J. Newman, “Epidemics and percolation in small-world networks,” *Physical Review E*, vol. 61, pp.5678–5682, 2000. [Article \(CrossRef Link\)](#)
- [17] A. Grabowski, R. A. Kosiński, “Epidemic spreading in a hierarchical social network,” *Physical Review E*. vol. 70, pp.031908, 2004. [Article \(CrossRef Link\)](#)
- [18] S. Fortunato. “Community detection in graphs,” *Physics reports*, vol. 486, no. 3, pp.75-174, 2010.
[Article \(CrossRef Link\)](#)
- [19] L. Botha, S. Kroon. “A community-based model of online social networks,” //*The 4th SNA-KDD Workshop on Social Network Mining and Analysis*. 2010. [Article \(CrossRef Link\)](#)
- [20] X. Chu, J. Guan, Z. Zhang, et al. “Epidemic spreading in weighted scale-free networks with community structure,” *Journal of Statistical Mechanics: Theory and Experiment*, vol. 2009, no. 7, pp.P07043, 2009. [Article \(CrossRef Link\)](#)
- [21] G. Palla, I. Derényi, I. Farkas, et al. “Uncovering the overlapping community structure of complex networks in nature and society,” *Nature*, vol. 435, no. 7043, pp.814-818, 2005.
[Article \(CrossRef Link\)](#)
- [22] Y. Y. Ahn, J. P. Bagrow, S. Lehmann. “Link communities reveal multiscale complexity in networks,” *Nature*, vol. 466, no. 7307, pp.761-764, 2010. [Article \(CrossRef Link\)](#)
- [23] F. Reid, N. Hurley, “Diffusion in networks with overlapping community structure,” in *Proc. of Data Mining Workshops (ICDMW), 2011 IEEE 11th International Conference on*, pp.969-978, 2011. [Article \(CrossRef Link\)](#)
- [24] J. Shang, L. Liu, X. Li, et al. “Epidemic spreading on complex networks with overlapping and non-overlapping community structure,” *Physica A: Statistical Mechanics and its Applications*, vol. 419, pp.171-182, 2015. [Article \(CrossRef Link\)](#)
- [25] Y. Wu, S. Deng, H. Huang. “Performance Analysis of Epidemic Routing in Delay Tolerant Networks with Overlapping Communities and Selfish Nodes,” *International Journal of Computers Communications & Control*, vol. 8, no. 5, pp.744-753, 2013. [Article \(CrossRef Link\)](#)

- [26] X. Wu, Z. Liu. "How community structure influences epidemic spread in social networks," *Physica A: Statistical Mechanics and its Applications*, vol. 387, no. 2, pp.623-630, 2008. [Article \(CrossRef Link\)](#)
- [27] M. Salathé, J. H. Jones. "Dynamics and control of diseases in networks with community structure," *PLoS Comput Biol*, vol. 6, no. 4, pp.e1000736, 2010. [Article \(CrossRef Link\)](#)
- [28] Y. Min, X. Jin, Y. Ge, et al. "The role of community mixing styles in shaping epidemic behaviors in weighted networks," *PloS one*, vol. 8, no. 2, pp.e57100, 2013. [Article \(CrossRef Link\)](#)
- [29] L. Weng, F. Menczer, Y. Y. Ahn. "Virality prediction and community structure in social networks," *Scientific reports*, vol. 3, pp.2522, 2013. [Article \(CrossRef Link\)](#)
- [30] A. Nematzadeh, E. Ferrara, A. Flammini, et al. "Optimal network modularity for information diffusion," *Physical review letters*, vol. 113, no. 8, pp.088701, 2014. [Article \(CrossRef Link\)](#)
- [31] J. Shang, L. Liu, F. Xie, et al., "How overlapping community structure affects epidemic spreading in complex networks," *Computer Software and Applications Conference Workshops (COMPSACW)*, pp.240-245, 2014. [Article \(CrossRef Link\)](#)
- [32] J. Shang, L. Liu, F. Xie, et al. "Performance analysis of information propagation in DTN-like scale-free mobile social network," *KSII Transactions on Internet & Information Systems*, vol. 8, no. 11, pp.3984-3996, 2014. [Article \(CrossRef Link\)](#)
- [33] R. Bakhshi, L. Cloth, F. Wan, et al. "Mean-Field Analysis for the Evaluation of Gossip Protocols," *ACM Sigmetrics Performance Evaluation Review*, vol. 36, no. 3, pp.31-39, 2009. [Article \(CrossRef Link\)](#)
- [34] R. Pastor-Satorras, A. Vespignani. "Epidemic spreading in scale-free networks," *Physical Review Letters*, vol. 86, no. 14, pp.3200-3203, 2001. [Article \(CrossRef Link\)](#)



Narisa Zhao is a professor of System engineering institution at Dalian University of Technology, China. He received M.S. degree and Ph.D. degree in operational research and system engineering from Dalian University of Technology. His research interests include complex networks, text mining and e-commerce.



Xiaojun Liu is a Ph.D. student of System engineering institution at Dalian University of Technology, China. She received M.S. degree in Institute of mathematical sciences from Inner Mongolia University in 2014. Her research interests include complex networks and text mining.