$\mathcal{A}\mathcal{P}$
ijpam.eu

ANALYTICAL SOLUTION FOR THE SPREAD OF EPIDEMIC DISEASES IN COMMUNITY CLUSTERED NETWORK

Chang Phang[1][§], YongHong Wu[2], Benchawan Wiwatanapataphee[3]

[1]Department of Mathematics and Statistics
Universiti Tun Hussein Onn Malaysia
Batu Pahat, 86400, Johor, MALAYSIA

[2]Department of Mathematics and Statistics
Curtin University
Perth, WA 6845, AUSTRALIA

[3]Department of Mathematics
Mahidol University
Bangkok, 10400, THAILAND

Abstract: We present a bond percolation model for community clustered networks with an arbitrarily specified joint degree distribution. Our model is based on the Probability Generating Function (PGF) method for multitype networks, but incorporate the free-excess degree distribution, which makes it applicable for clustered networks. In the context of contact network epidemiology, our model serves as a special case of community clustered networks which are more appropriate for modelling the disease transmission in community networks with clustering effects. Beyond the percolation threshold, we are able to obtain the probability that a randomly chosen community-i node leads to the giant component. The probability refers to the probability that an individual in a community will be affected from the infective disease. Besides that, we also establish method to calculate the size of the giant component and the average small-component size (excluding the giant component). When the clustering ef-

[§]Correspondence author

fect is taken into account through the free-excess degree distribution, the model shows that the clustering effect will decrease the size of the giant component. In short, our model enables one to carry out numerical calculations to simulate the disease transmission in community networks with different community structure effects and clustering effects.

AMS Subject Classification: 92D30
Key Words: epidemics, community clustered networks, probability generating function

1. Introduction

It has long been recognized that two of the key features of social networks are community structure and clustering effect. The former one emphasizes that the links are dense in a community but sparse between communities, while the later one refers to the relative number of triangles in a network. For the community structure effect, the links between communities will make the network less heterogeneous and result in larger epidemic prevalence in the exponential degree distribution networks [1]. For other types of degree distributions such as the power-law degree distribution, the authors of [2] developed an algorithm to obtain a social network model with a multiple-community structure with adjustable clustering coefficients and adjustable degree of community. They showed that the heterogeneous network is less efficient than the homogeneous network in spreading of epidemic. Different to [2], we study the bond percolation model of community clustered networks with an arbitrary joint degree distribution by using the PGF formalism.

In a series of papers [3, 4, 5], M.E.J. Newman used PGF for random graphs with arbitrary distributions of vertex degree. With the mathematics of generating functions, the author managed to calculate exactly some statistical properties of such graphs in the limit of large numbers of vertices, including the mean of component size and the giant component. Using the combination of mapping to percolation models and the generating function method, Newman established the analytic expressions for the size of epidemic outbreaks and the mean degree of individuals affected in an epidemic. Following Newman's work, many researchers consider more features to improve the percolation models based on the generating function method. For example, in [6], the author developed a model to represent heterogeneous populations so as to study the mixing patterns. Apart from that there are many other models, including percolation models for random directed networks [7], models for two compet-

ing disease spreading over the same network at the same time [8].Despite all these efforts, for the PGF formalism type of models, the impact of community structure on epidemic spreading has not been well considered. In this paper, we intend to fill this gap by proposing a PGF model with community structure and investigating epidemic spreading upon these community structure networks.

There are some other models which use other mathematical tools or simulation to predict the transmission of infectious diseases in social networks. Among that including [9, 10, 11]. In [9], the authors proposed a growth model to create networks with overlapping community structure and investigating epidemic spreading upon these networks by numerical simulation, while in [10], the authors proposed a model which the links in and between communities are dynamical to study the epidemic spreading in communities. Apart from that, a model consists of two types of nodes with different behavior patterns, i.e. active nodes and passive nodes was proposed to study the effects of social impact on the epidemic spreading in complex networks based on SIS epidemic model in [11] and simulations are done for the model in ER random network, BA scale-free network, structured scale-free network , and WS small-world network. Different from these models, our model having specific joint degree distribution in and between the communities. For this purpose, we use PGF formalism to generate the community structure model with the specific degree distribution we desire for each node. The biggest advantage is the PGF formalism allows us to get the analytical solution.

Apart from that, one of the areas which get considerable attention from researchers is clustered networks. In [7], the author introduced a class of random clustered networks and showed that the clustered networks had small component sizes and bigger epidemic threshold in comparison to the same preferential mixing unclustered networks.We intend to study the clustering effect in our community structure model.

In this paper, we present a bond percolation model of community clustered networks with an arbitrary joint degree distribution (i.e. the degree distribution can be specified arbitrarily). Our model is based on the PGF formalism for multitype networks introduced by Antoine Allard et.al. [12]. Their multitype network model is the extension of the PGF formalism which investigated by M.E.J. Newman. In addition, we incorporate the free-excess degree distribution, which was introduced in [13] to make it applicable for clustered networks. We focus on complex networks with arbitrary joint degree distribution. In the context of contact network epidemiology, our model serves as a special case of community clustered networks which are more suitable for modelling the disease transmission in community networks with the clustering effect. For certain

clustering coefficient and beyond the percolation threshold, we obtain the probability that a randomly chosen community-i node leads to the giant component. In the context of contact network epidemiology, this probability refers to the probability that an individual in a community is affected by the infectious disease. In addition, we derive formulae to calculate the size (i.e. fraction) of the giant component and the average small-component size (excluding the giant component) in the community structure network. If the disease transmission rate between each pair of communities is the same in both direction, the size of the giant component in each community-i is equivalent to the probability that a randomly chosen community-i node leads to the giant component. When the clustering effect is taken into account through the free-excess degree distribution , our model shows that the clustering effect leads to the reduce of the size of the giant component.

The rest of the paper is organized as follows. In Section 2, we discuss some assumptions to be used in our community clustered networks. In Section 3, we present the PGF formamlism for the proposed community clustered networks. In Sections 4 and 5, we focus on the calculation of the outbreak size distribution and the percolation threshold, followed by numerical simulations in Section 6. Some conclusions are given in Section 7.

2. Community Clustered Networks

We discuss a model with 2 communities which can be generalized to multi-communities. Throughout the discussion, we assume that

(a) there exist realizable degree sequences which lead to simple graphs (i.e. networks) with no self loop,

(b) the inter-community edges will be redistributed according to the three ways as explained in Section 3.1.3,

(c) the isolated nodes in each community will remain isolated,

(d) the highest degree is 8 for model 1, and 6 for model 2,

(e) it is possible for two nodes connecting to a mutual node to connect themselves, thereby forming a triangle. It is represented by the clustering coefficient, C,

(f) One link is counted as two edges.

Assumption (b) implies that power-law distribution will not be applied here. In other words, nodes with higher degree do not necessarily have more inter-community edges. In assumption (c), we assign some isolated nodes to each

community. In a finite time, the isolated nodes remain isolated when we randomly connect the inter-community edges to the nodes with different degrees in each community. This can serve as a check point for our computer program when we use it to find distribution for the size of small component. There is no contact for every isolated node, namely there is no transmissibility of disease. Hence, the number of isolated nodes in each community will remain the same. Apart from that, we set the highest degree at 8. In other words, the network is not a highly right skewed network and does not have super infection nodes.

In some of our analysis, we will further assume that one of the communities has a low vaccination rate and thus a higher rate of disease transmissibility among their members.BEsides that, the rate of disease transmissibility along the inter-community edges is low, as sick individuals will travel less. For our models, we assume that the exact number of nodes having degree k, denoted by n_k, is known. Hence, we can write the exact generating function for the probability distribution in the form of a finite polynomial.

3. Formalism

We now present a formalism that describes the bond percolation model of community networks. It is based on the PGF formalism for multitype networks introduced by Antoine Allard et.al. [12].

3.1. Degree Distribution

First, we assume that the arbitrarily specified degree distribution will produce a realizable degree sequence. Let $P_{i=1}(k_1, k_2, ..., k_M)$ and $P_{i=2}(k_1, k_2, ..., k_M)$ be the probability degree distributions that a randomly chosen community-i node is connected to k_1 nodes, k_2 nodes, k_3 nodes and so on. Since we deal with community structure networks, among the edges of a node, there are some edges which may connect to nodes in other community.

For the sake of simplicity, we discuss the PGF formalism for the model with two communities. In this case, we have two ways to represent the degree distribution, namely number of nodes in each community or number of nodes connecting community-i node to community-j.

3.1.1. Degree Distribution for Number of
Nodes in Each Community

In the first way, we need to have degree distribution for the number of nodes in each community. As an example, we consider the following degree distribution

$$p_1(\mathbf{k}) = \{10, 10, 10, 45, 205, 135, 85\},$$

$$p_2(\mathbf{k}) = \{10, 10, 5, 85, 110, 185, 135\},$$

where $\mathbf{k} = \{k_l\}_{l=0}^6 = \{l\}_{l=0}^6 = (0, 1, 2, 3, 4, 5, 6)$ and the value of $p_i(k_l)$ is the number of nodes with degree l in the community-i. In the above example, there are 10 nodes with 0 degree (i.e. isolated nodes) in community 1, 10 nodes with 1 degree, 10 nodes with 2 degree, 45 nodes with 3 degree and so on in community 1. Apart from that, we can get the information about the number of edges in each community, namely there are 2170 and 2450 edges respectively in communities 1 and 2 in our model example and the total edges is 4620. Note that here one link is counted as two edges. If we divide $P_i(\mathbf{k})$ by N_i where N_i denotes the number of nodes in community-i, we will have a probability degree distribution.

$$P_1(\mathbf{k}) = \frac{1}{500}\{10, 10, 10, 45, 205, 135, 85\},$$

$$P_2(\mathbf{k}) = \frac{1}{540}\{10, 10, 5, 85, 110, 185, 135\}.$$

Without considering the rate of transmissibility of diseases, we have the following PGF

$$G_i(\mathbf{x}) = \sum_{k=0}^{\infty} P_i(k) \prod_{l=1}^{M} x_l^{k_l} \tag{1}$$

where $i = 1, 2$; M denotes the number of communities which is 2 in this example and $x_l^{k_l}$ denotes node in community-l with degree k_l.

3.1.2. Degree Distribution for Number of Nodes Connecting
Community-i to Community-j

The second way of representing the degree distribution is by considering the number of nodes connecting community-i to community-j. We have the following degree distribution:

$$p_1(\mathbf{k}_1) = \{10, 10, 20, 80, 200, 120, 60\},$$

$$p_1(\mathbf{k}_2) = \{420, 50, 20, 10, 0, 0, 0\},$$
$$p_2(\mathbf{k}_1) = \{460, 50, 20, 10, 0, 0, 0\},$$
$$p_2(\mathbf{k}_2) = \{10, 10, 10, 100, 140, 180, 90\},$$

where $p_i(\mathbf{k}_i)$ is as defined before, while $p_i(\mathbf{k}_{jl})$ denotes the number of nodes in community-i with $\mathbf{k}_{jl}$ edges linking to community j in which $\mathbf{k}_{jl} = l$ for $l = 0, 1, 2, \ldots$.

For the above example, $p_1(\mathbf{k}_1)$ shows that there are 10 nodes with 0 degree (i.e. isolated nodes) in community 1, 10 nodes with 1 degree, 20 nodes with 2 degree and so on in community 1; The data of $p_1(\mathbf{k}_2)$ shows that there are 420 nodes in community 1 with no edge linking to community 2, 50 nodes in community 1 having one edge linking to community 2 and so on. The same will apply to the nodes in community 2. We can also get the information about the number of inter-community edges and the number of intra-community edges in each community. The number of intra-community edge is 2170 in community-1 and 2450 in community-2.For the two communities model, $\mathbf{k} = \mathbf{k}_1 + \mathbf{k}_2$, hence the total edges must be the same as that by the first way mentioned in Section 3.1.1 (where it is 4620). The number of inter-community edges for communities 1 and 2 must be the same, in our model, we have 120 inter-community edges.

We can rewrite the degree distribution above as probability degree distribution as follows

$$P_1(\mathbf{k}_1) = \frac{1}{500}\{10, 10, 20, 80, 200, 120, 60\},$$

$$P_1(\mathbf{k}_2) = \frac{1}{500}\{420, 50, 20, 10, 0, 0, 0\},$$

$$P_2(\mathbf{k}_1) = \frac{1}{540}\{460, 50, 20, 10, 0, 0, 0\},$$

$$P_2(\mathbf{k}_2) = \frac{1}{540}\{10, 10, 10, 100, 140, 180, 90\}.$$

Without considering the rate of transmissibility of diseases, let G_{ij} be the generating function where the subscript ij represents the chosen node in community-i connected to node in community-j, and let $P_i(k_{uv})$ be the probability that a randomly chosen community-i node with u edges connects to the nodes in community v. Then, we have the following PGF

$$\begin{aligned}G_{11}(\mathbf{x}) &= P_1(k_{01}) + P_1(k_{11})x_1 + P_1(k_{21})x_1^2 + P_1(k_{31})x_1^3 + P_1(k_{41})x_1^4 \\ &\quad + P_1(k_{51})x_1^5 + P_1(k_{61})x_1^6 \\ &= \frac{10}{500} + \frac{10}{500}x_1 + \frac{20}{500}x_1^2 + \frac{80}{500}x_1^3 + \frac{200}{500}x_1^4 + \frac{120}{500}x_1^5 + \frac{60}{500}x_1^6\end{aligned}$$

$$G_{12}(\mathbf{x}) = \frac{420}{500} + \frac{50}{500}x_2 + \frac{20}{500}x_2^2 + \frac{10}{500}x_2^3$$

$$G_{21}(\mathbf{x}) = \frac{460}{540} + \frac{50}{540}x_1 + \frac{20}{540}x_1^2 + \frac{10}{540}x_1^3$$

$$G_{22}(\mathbf{x}) = \frac{10}{540} + \frac{10}{540}x_2 + \frac{10}{540}x_2^2 + \frac{100}{540}x_2^3 + \frac{140}{540}x_2^4 + \frac{180}{540}x_2^5 + \frac{90}{540}x_2^6$$

and $\frac{dG_{11}(\mathbf{1})}{dx_1} = \frac{41}{10}$, $\frac{dG_{12}(\mathbf{1})}{dx_2} = \frac{6}{25}$, $\frac{dG_{21}(\mathbf{1})}{dx_1} = \frac{2}{9}$, $\frac{dG_{22}(\mathbf{1})}{dx_2} = \frac{233}{54}$,
where $\frac{dG_{ij}(\mathbf{1})}{d\mathbf{x}}$ denotes the average number of edges connecting nodes in community-i to nodes in community-j, and we represent it as z_{ij}.

Definition 1. Let z_{ij} be the average number of edges connecting nodes in community-i to nodes in community-j, then

$$z_{ij} = \sum_{k_1=0}^{\infty} \cdots \sum_{k_M=0}^{\infty} k_j P_i(k_1, k_2, \cdots, k_M) \equiv \sum_{k=0}^{\infty} k_j P_i(\mathbf{k}).$$

In matrix form, for two communities, we have

$$\mathbf{z} = \left(\begin{array}{cc} z_{11} & z_{12} \\ z_{21} & z_{22} \end{array} \right).$$

Let the community-i nodes occupy a fraction w_i of the network, and define

$$\mathbf{w} = \left(\begin{array}{cc} w_1 & 0 \\ 0 & w_2 \end{array} \right).$$

We have $\mathbf{wz} = (\mathbf{wz})^T$, $tr(\mathbf{w}) = 1$ and $w_1 = \frac{z_{21}}{z_{12}+z_{21}}$, $w_2 = \frac{z_{12}}{z_{12}+z_{21}}$. In our model, we have $w_1 = \frac{z_{21}}{z_{12}+z_{21}} = \frac{2/9}{6/25+2/9} = \frac{25}{52}$, $w_2 = \frac{z_{12}}{z_{12}+z_{21}} = \frac{6/25}{6/25+2/9} = \frac{27}{52}$. This means that we have $\frac{25}{52}$ fraction of nodes in community 1 and $\frac{27}{52}$ fraction of nodes in community 2. The followings two ways can be used to determine how strong the community structure is:

(a) Let Z be the number of edges connecting nodes in community-i to nodes in community-j, then we will have $Z = \frac{n}{2m}z$ where n is the number of nodes and m is the number of edges. Hence we can use the information to obtain the modularity which is given by $Q = Tr(\mathbf{Z}) - \|\mathbf{Z}^2\|$.

(b) In [14], the degree of community σ , is given by $\sigma = p/q$ where p is the probability for the event that there exist links within the community and q is the probability for the event that there exist links between the communities. In this work, we redefine $\sigma = \frac{tr(z)}{\sum_{i \neq j} z_{ij}}$ where $\sigma >> 1$ implies strong community. It is easy to show that our model has $Q = 0.446$ or $\sigma = 18.02$.

3.1.3. Algorithm for Redistributing the Intercommunity Links

There are a few possible ways to consider the distribution of intercommunity links including: the links are randomly attached to the nodes in each community, the links are equally attached to the nodes in each community, the links are preferably attached to the nodes with higher degrees in each community.

Consider the two communities model, where the links are equally attached to the nodes in each community. Let n be the number of intercommunity links, and let M be the highest degree of the nodes in the respectively community. If $n \bmod M = R$, then we have $YM + R = n$. Assuming that P and Q are the number of nodes with degree 1 and degree M, then we have the number of nodes with degree $M + 1$ is $Y + R$, the number of nodes with degree M is $P - R$ where P is the number of nodes with degree 1, the number of nodes with degree 1 is $Q - Y$ where Q is the number of nodes with degree M.

3.2. The Occupied Degree Distribution

For the discussion of the PGF formalism involving rate of disease transmission, we should consider the occupied degree distribution. For the two communities model, we define a bond occupation probability matrix as

$$\mathbf{T} = \begin{pmatrix} T_{11} & T_{12} \\ T_{21} & T_{22} \end{pmatrix}.$$

Definition 2. The probability that a randomly chosen degree-$\mathbf{k}$ node has $\tilde{\mathbf{k}}$ occupied edges is

$$P_i(\tilde{\mathbf{k}}|\mathbf{k}) = \prod_{i=1}^{M} \binom{k_l}{\tilde{k}_l} (T_{il})^{\tilde{k}_l} (1 - T_{il})^{k_l - \tilde{k}_l}.$$

The occupied degree distribution, $\tilde{P}_i(\tilde{\mathbf{k}})$, is

$$\tilde{P}_i(\tilde{\mathbf{k}}) = \sum_{k=\tilde{k}}^{\infty} P_i(\tilde{\mathbf{k}}|\mathbf{k}) P_i(\mathbf{k}) = \sum_{k=\tilde{k}}^{\infty} P_i(\mathbf{k}) \prod_{l=1}^{M} \binom{k_l}{\tilde{k}_l} (T_{il})^{\tilde{k}_l} (1 - T_{il})^{k_l - \tilde{k}_l}.$$

Hence, the PGF is

$$G_i(\mathbf{x}; \mathbf{T}) = \sum_{\tilde{k}=0}^{\infty} \tilde{P}_i(\tilde{\mathbf{k}}) \prod_{l=1}^{M} x_l^{\tilde{k}_l} = \sum_{k=0}^{\infty} P_i(\mathbf{k}) \prod_{l=1}^{M} \sum_{\tilde{k}_l=0}^{k_l} \binom{k_l}{\tilde{k}_l} (x_l T_{il})^{\tilde{k}_l} (1 - T_{il})^{k_l - \tilde{k}_l}$$

$$G_i(\mathbf{x}; \mathbf{T}) = \sum_{k=0}^{\infty} P_i(\mathbf{k}) \prod_{l=1}^{M} [1 - (x_l - 1)T_{il}]^{k_l},$$

where $G_i(1; \mathbf{T}) = 1$ if $P_i(\mathbf{k})$ is properly normalized.

From Definition 2, we can get the average occupied degree connecting nodes in community-i to nodes in community-j as $\tilde{z}_{ij} = \frac{\partial G_i(\mathbf{1};\mathbf{T})}{\partial x_j} = T_{ij} \sum_{k=0}^{\infty} k_j P_i(\mathbf{k})$ $= T_{ij} z_{ij}$. For example, $\frac{\partial G_1(\mathbf{x};\mathbf{T})}{\partial x_2}|_{x_1=1, x_2=1}$ gives $T_{12} z_{12}$.

For two communities (we assume each node has at most degree 6 as for in model 2) , we have the following generating functions when we consider the rate of transmissibility,

$$G_1(\mathbf{x}; \mathbf{T}) = P_1(k_0) \left[\frac{L_{11}}{L_1}(1 - (x_1 - 1)T_{11})\frac{L_{12}}{L_1}(1 - (x_2 - 1)T_{12}) \right]^{k_0}$$

$$+ P_1(k_1) \left[\frac{L_{11}}{L_1}(1 - (x_1 - 1)T_{11})\frac{L_{12}}{L_1}(1 - (x_2 - 1)T_{12}) \right]^{k_1}$$

$$+ \ldots + P_1(k_6) \left[\frac{L_{11}}{L_1}(1 - (x_1 - 1)T_{11})\frac{L_{12}}{L_1}(1 - (x_2 - 1)T_{12}) \right]^{k_6},$$

$$G_2(\mathbf{x}; \mathbf{T}) = P_2(k_0) \left[\frac{L_{21}}{L_2}(1 - (x_1 - 1)T_{21})\frac{L_{22}}{L_2}(1 - (x_2 - 1)T_{22}) \right]^{k_0}$$

$$+ P_2(k_1) \left[\frac{L_{21}}{L_2}(1 - (x_1 - 1)T_{21})\frac{L_{22}}{L_2}(1 - (x_2 - 1)T_{22}) \right]^{k_1}$$

$$+ \ldots + P_2(k_6) \left[\frac{L_{21}}{L_2}(1 - (x_1 - 1)T_{21})\frac{L_{22}}{L_2}(1 - (x_2 - 1)T_{22}) \right]^{k_6},$$

where L_1 and L_2 are the total number of edges in communities 1 and 2 respectively and L_{ij} is the total number of edges which link nodes in community-i to nodes in community-j. Note that $G_1(\mathbf{1}; \mathbf{T}) = 1$ and $G_2(\mathbf{1}; \mathbf{T}) = 1$.

3.3. The Occupied Excess Degree Distribution

Definition 3. The occupied excess degree distribution is given by

$$\tilde{Q}_{ij}(\tilde{\mathbf{k}}) = \frac{1}{z_{ji}} \sum_{k=\tilde{k}}^{\infty} (k_i + 1)P_j(\mathbf{k} + \boldsymbol{\delta}_i) \prod_{l=1}^{M} \binom{k_l}{\tilde{k}_l}(T_{jl})^{\tilde{k}_l}(1 - T_{jl})^{k_l - \tilde{k}_l}.$$

Hence, the PGF is given by

$$F_{ij}(\mathbf{x}; \mathbf{T}) = \sum_{\tilde{k}=0}^{\infty} \tilde{Q}_{ij}(\tilde{\mathbf{k}}) \prod_{l=1}^{M} x_l^{\tilde{k}_l} = \frac{1}{z_{ji}} \sum_{k=0}^{\infty} k_i P_j(\mathbf{k}) \prod_{l=1}^{M} [1 - (x_l - 1)T_{jl}]^{k_l - \delta_{il}},$$

where the ij represents the chosen edges connecting community-i and community-j.

For our 2-community model, we can obtain $F_{ij}(\mathbf{x}; \mathbf{T})$ by

$$F_{ij}(\mathbf{x}; \mathbf{T}) = \frac{1}{\tilde{z}_{ji}} \frac{\partial G_j(\mathbf{x}; \mathbf{T})}{\partial x_i}. \tag{2}$$

For two communities networks with nodes of at most degree 6, using (2), after some algebra work, we have

$$\begin{aligned}
F_{11}(x_1, x_2; \mathbf{T}) = {}& p_0 + p_{01}x_2 + p_{02}x_2^2 + p_{03}x_2^3 + p_{04}x_2^4 + p_{05}x_2^5 \\
& + p_{10}x_1 + p_{11}x_1x_2 + p_{12}x_1x_2^2 + p_{13}x_1x_2^3 + p_{14}x_1x_2^4 \\
& + p_{20}x_1^2 + p_{21}x_1^2x_2 + p_{22}x_1^2x_2^2 + p_{23}x_1^2x_2^3 + p_{30}x_1^3 \\
& + p_{31}x_1^3x_2 + p_{32}x_1^3x_2^2 + p_{40}x_1^4 + p_{41}x_1^4x_2 + p_{50}x_1^5,
\end{aligned}$$

where the parameter p_{uv} denotes the probability that a node with excess degree v is reached by following a randomly chosen edge with excess degree u. Similar formulae can be established for $F_{12}(x_1, x_2; \mathbf{T})$, $F_{21}(x_1, x_2; \mathbf{T})$ and $F_{22}(x_1, x_2; \mathbf{T})$ where the parameter p in the above equation will be replaced by q,r,and s respectively.

3.4. The Occupied Free-Excess Degree Distribution

In order to consider the clustered network $(C > 0$), we apply the free-excess degree distribution concept introduced in [12]. Analogously to the excess degree, we follow one of the edges of node v_0 to reach a neighbour v_1 having degree $d(v_1) = i + 1$(i is the excess degree). We are interested in calculating the probability that the node has k neighbours that are not connected back to v_0 (via a triangle), which is given by $\binom{i}{k}(1 - C)^k C^{i-k}$

Definition 4. The free excess degree distribution is given by

$$\begin{aligned}
G_c(x) &= \sum_{k=0}^{\infty} e_k x^k = \sum_{k=0}^{\infty} \sum_{i=0}^{\infty} q_i \binom{i}{k} C^i \left(\frac{1-C}{C}\right)^k x^k \\
&= \sum_{i=0}^{\infty} q_i C^i \sum_{k=0}^{\infty} \binom{i}{k} \left(\frac{1-C}{C} x\right)^k = \sum_{i=0}^{\infty} q_i C^i \left(1 + \frac{1-C}{C} x\right)^i \\
&= \sum_{i=0}^{\infty} q_i \left[C + (1-C)x\right]^i = G_1\left[C + (1-C)x\right]
\end{aligned}$$

$$e_k = \sum_{i=0}^{\infty} q_i \binom{i}{k} (1-C)^k C^{i-k} = \sum_{i=0}^{\infty} q_i \binom{i}{k} C^i \left(\frac{1-C}{C} \right)^k$$

By Definition 4, we use the generating function associated with the occupied excess degree distribution as given in Definition 3. Thus, we have the following relationship

$$Fc_{ij}(\mathbf{x}; \mathbf{T}) = F_{ij}(C + (1-C)\mathbf{x}; \mathbf{T}) \tag{3}$$

4. Outbreak Size Distribution

In this section, we discuss an iteration method to obtain the probability that a randomly chosen community-i node leads to the giant component. Although the discussion applies to the model without considering the clustering coefficient C, for the model with clustering coefficient C, we will have Fc_{ij} instead of F_{ij} as in (3).

Let $H_{ij}(\mathbf{x}; \mathbf{T})$ be the generating function for the size distribution of the component reached by following an $i \to j$ edge.

$$H_{ij}(\mathbf{x}; \mathbf{T}) = x_j F_{ij}(\mathbf{H}_j(\mathbf{x}; \mathbf{T}); \mathbf{T}) \tag{4}$$

The solution for the equation (4) can be found by seeking the stable fixed point of the mapping

$$H_{ij}^{(n)}(\mathbf{x}; \mathbf{T}) = x_j F_{ij}(\mathbf{H}_{jM}^{(n-1)}(\mathbf{x}; \mathbf{T}); \mathbf{T}) \tag{5}$$

as $i = 1 \ldots M$ and $n \to \infty$ for initial conditions $H_{ij}^0(\mathbf{x}; \mathbf{T}) = x_j$. The equation says that when we follow an $i \to j$ edge, we find at least one node at the other end (the factor of x_1, x_2 on the RHS), plus some other clusters of nodes (each represented by $H_{ij}(\mathbf{x}; \mathbf{T})$) which are reachable by following other edges attached to that node. The number of the clusters is distributed according to the coefficients of x_1, x_2 in $F_{ij}(\mathbf{x}; \mathbf{T}) = \frac{1}{z_{ij}} \frac{\partial G_j(\mathbf{x}; \mathbf{T})}{\partial x_i}$, and hence the appearance of $F_{ij}(\mathbf{x}; \mathbf{T})$.

For our model, using (4) and (5), we have the following iteration process

$$H_{11}^{(n)}(x_1, x_2; \mathbf{T}) = x_1 F_{11}(H_{11}^{(n-1)}(x_1, x_2; \mathbf{T}), H_{12}^{(n-1)}(x_1, x_2; \mathbf{T}); \mathbf{T})$$

$$H_{12}^{(n)}(x_1, x_2; \mathbf{T}) = x_2 F_{12}(H_{21}^{(n-1)}(x_1, x_2; \mathbf{T}), H_{22}^{(n-1)}(x_1, x_2; \mathbf{T}); \mathbf{T})$$

$$H_{21}^{(n)}(x_1, x_2; \mathbf{T}) = x_1 F_{21}(H_{11}^{(n-1)}(x_1, x_2; \mathbf{T}), H_{12}^{(n-1)}(x_1, x_2; \mathbf{T}); \mathbf{T})$$

$$H_{22}^{(n)}(x_1, x_2; \mathbf{T}) = x_2 F_{22}(H_{21}^{(n-1)}(x_1, x_2; \mathbf{T}), H_{22}^{(n-1)}(x_1, x_2; \mathbf{T}); \mathbf{T})$$

where the initial condition is given by $H_{11}^{(0)}(x_1, x_2; \mathbf{T}) = p_0 x_1$, $H_{12}^{(0)}(x_1, x_2; \mathbf{T}) = q_0 x_2$, $H_{21}^{(0)}(x_1, x_2; \mathbf{T}) = r_0 x_1$, and $H_{22}^{(0)}(x_1, x_2; \mathbf{T}) = s_0 x_2$. For the first iteration in our model 2, we have

$$
\begin{aligned}
H_{11}^{(1)}(x_1, x_2; \mathbf{T}) =& x_1 F_{11}(H_{11}^{(0)}(x_1, x_2; \mathbf{T}), H_{12}^{(0)}(x_1, x_2; \mathbf{T}); \mathbf{T}) \\
=& x_1 F_{11}(p_0 x_1, q_0 x_2; \mathbf{T}) \\
=& x_1 \{ p_0 + p_{01}(q_0 x_2) + p_{02}(q_0 x_2)^2 + p_{03}(q_0 x_2)^3 \\
& + p_{04}(q_0 x_2)^4 + p_{05}(q_0 x_2)^5 \\
& + p_{10}(p_0 x_1) + p_{11}(p_0 x_1)(q_0 x_2) + p_{12}(p_0 x_1)(q_0 x_2)^2 \\
& + p_{13}(p_0 x_1)(q_0 x_2)^3 + p_{14}(p_0 x_1)(q_0 x_2)^4 \\
& + p_{20}(p_0 x_1)^2 + p_{21}(p_0 x_1)^2(q_0 x_2) + p_{22}(p_0 x_1)^2(q_0 x_2)^2 \\
& + p_{23}(p_0 x_1)^2(q_0 x_2)^3 \\
& + p_{30}(p_0 x_1)^3 + p_{31}(p_0 x_1)^3(q_0 x_2) + p_{32}(p_0 x_1)^3(q_0 x_2)^2 \\
& + p_{40}(p_0 x_1)^4 + p_{41}(p_0 x_1)^4(q_0 x_2) + p_{50}(p_0 x_1)^5 \} \\
=& \, p_0 x_1 + p_{01} q_0 x_1 x_2 + p_{10} p_0 x_1^2 + O(x_1, x_2)^3.
\end{aligned}
$$

Similarly, we have

$$
\begin{aligned}
H_{11}^{(1)}(x_1, x_2; \mathbf{T}) &= q_0 x_2 + q_{10} r_0 x_1 x_2 + q_{01} s_0 x_2^2 + O(x_1, x_2)^3, \\
H_{21}^{(1)}(x_1, x_2; \mathbf{T}) &= r_0 x_1 + r_{01} q_0 x_1 x_2 + r_{10} p_0 x_1^2 + O(x_1, x_2)^3, \\
H_{22}^{(1)}(x_1, x_2; \mathbf{T}) &= s_0 x_2 + s_{10} r_0 x_1 x_2 + s_{01} s_0 x_2^2 + O(x_1, x_2)^3.
\end{aligned}
$$

For the second iteration, we have

$$
\begin{aligned}
H_{11}^{(2)}(x_1, x_2; \mathbf{T}) =& x_1 F_{11}(H_{11}^{(1)}(x_1, x_2; \mathbf{T}), H_{12}^{(1)}(x_1, x_2; \mathbf{T}); \mathbf{T}) \\
=& \, x_1 F_{11}(p_0 x_1 + p_{01} q_0 x_1 x_2 + p_{10} p_0 x_1^2, q_0 x_2 \\
& + q_{10} r_0 x_1 x_2 + q_{01} s_0 x_2^2; \mathbf{T}).
\end{aligned}
$$

After some algebra work, we obtain

$$
\begin{aligned}
H_{11}^{(2)}(x_1, x_2; \mathbf{T}) =& \, p_0 x_1 + p_{01} q_0 x_1 x_2 + p_{10} p_0 x_1^2 + (p_{01} q_{01} s_0 + p_{02} q_0^2) x_1 x_2^2 \\
& + (p_{10} p_{01} q_0^2 + p_{11} p_0 q_0 + p_{01} q_{10} r_0) x_1^2 x_2 \\
& + (p_{10}^2 p_0 + p_{20} p_0^2) x_1^3 + O(x_1, x_2)^4.
\end{aligned}
$$

Similarly, we have

$$H_{12}^{(2)}(x_1, x_2; \mathbf{T}) = x_2 F_{12}(H_{21}^{(1)}(x_1, x_2; \mathbf{T}), H_{22}^{(1)}(x_1, x_2; \mathbf{T}); \mathbf{T}),$$
$$H_{21}^{(2)}(x_1, x_2; \mathbf{T}) = x_1 F_{21}(H_{11}^{(1)}(x_1, x_2; \mathbf{T}), H_{12}^{(1)}(x_1, x_2; \mathbf{T}); \mathbf{T}),$$
$$H_{22}^{(2)}(x_1, x_2; \mathbf{T}) = x_2 F_{22}(H_{21}^{(1)}(x_1, x_2; \mathbf{T}), H_{22}^{(1)}(x_1, x_2; \mathbf{T}); \mathbf{T}),$$

and the iteration process continues until obtaining $H_{ij}^{(4)}(x_1, x_2; \mathbf{T})$ for model 2.

Denoting $K_i(\mathbf{x}; \mathbf{T})$ as the generating function for the size distribution of the whole component, we have the

$$K_i(\mathbf{x}; \mathbf{T}) = x_i G_i(\mathbf{H}_i(\mathbf{x}; \mathbf{T}); \mathbf{T}). \tag{6}$$

Now we can have the following definition.

Definition 5. Let $K_i(\mathbf{x}; \mathbf{T})$ be the total number of nodes reachable from a randomly chosen vertex in community-i , i.e. the size of the component to which such a node belongs, is generated by

$$K_i(\mathbf{x}; \mathbf{T}) = x_i G_i(\mathbf{H}_i(\mathbf{x}; \mathbf{T}); \mathbf{T}).$$

Since type-i nodes occupy a fraction w_i of the network, the size distribution of the component reached from a randomly chosen node is generated by

$$K(\mathbf{x}; \mathbf{T}) = \sum_{i=1}^{M} w_i K_i(\mathbf{x}; \mathbf{T}) = \sum_{i=1}^{M} w_i x_i G_i(\mathbf{H}_i(\mathbf{x}; \mathbf{T}); \mathbf{T}). \tag{7}$$

For our 2-communities model, using (7), we have

$$K_1(x_1, x_2; \mathbf{T}) = x_1 G_1(H_{11}^{(n)}(x_1, x_2; \mathbf{T}), H_{12}^{(n)}(x_1, x_2; \mathbf{T}); \mathbf{T}),$$
$$K_2(x_1, x_2; \mathbf{T}) = x_2 G_2(H_{21}^{(n)}(x_1, x_2; \mathbf{T}), H_{22}^{(n)}(x_1, x_2; \mathbf{T}); \mathbf{T}).$$

Above the percolation threshold, we will find the giant component. Hence the probability that a randomly chosen community-i node leads to the giant component is $P_i = 1 - K_i(\mathbf{1}; \mathbf{T})$. According to [12], for symmentric transmissibility $\mathbf{T} = \mathbf{T}^T$, the size of the giant component, S, is equal to P.

5. Percolation Threshold

Using the moment property of the PGF, the average number of community-i nodes in the small component reached from a randomly chosen node is obtained by differentiating $H_{ij}^{(n)}(\mathbf{x}; \mathbf{T})$ with respect to x_i . Hence, the average number of community-i nodes in the small component, $\langle s_i \rangle$, is given by

$$\langle s_i \rangle = w_i + \sum_{l=1}^{M} w_l \sum_{j=1}^{M} \tilde{z}_{lj} \alpha_{lj}^{(i)}, \tag{8}$$

where $\alpha_{lj}^{(i)} \equiv \frac{\partial H_{lj}(\mathbf{x};\mathbf{T})}{\partial x_i}|_{\mathbf{x=1}}$. $\alpha_{lj}^{(i)}$ is well approximated by the solution of $\alpha_{lj}^{(i)} = \delta_{ij} + \sum_{n=1}^{M} T_{jn} \beta_{lj}^{(n)} \alpha_{jn}^{(i)}$ and where the average excess degree is given by $B_{lj}^{(n)} = \frac{1}{T_{jn}} \frac{\partial F_{lj}(\mathbf{x};\mathbf{T})}{\partial x_n}|_{\mathbf{x=1}}$.

Define $[A_{ij}]_{\mu v} = T_{ij} \beta_{\mu v}^{(j)} \delta_{iv}$. If there are only two communities, A can be obtained by

$$A = \begin{pmatrix} T_{11}\beta_{11}^{(1)} & 0 & T_{12}\beta_{11}^{(2)} & 0 \\ T_{11}\beta_{21}^{(1)} & 0 & T_{12}\beta_{21}^{(2)} & 0 \\ 0 & T_{21}\beta_{12}^{(1)} & 0 & T_{22}\beta_{12}^{(2)} \\ 0 & T_{21}\beta_{22}^{(1)} & 0 & T_{22}\beta_{22}^{(2)} \end{pmatrix}, \tag{9}$$

with the average of s given by

$$\langle s \rangle = \sum_{i=1}^{M} \langle s_i \rangle \propto \frac{1}{det(I - A)}. \tag{10}$$

Hence, the phase transition occurs when $det(I - A) = 0$.

6. Numerical Simulations

Model 1: We summarize the model 1 as follows : There are two communities (i.e. $M = 2$) with probability degree distribution: $P_i(k_1, k_2) \equiv P_i(\mathbf{k})$ as shown below:

$$P_1(\mathbf{k}_1) = \frac{1}{5700} \{10, 2910, 2120, 480, 10, 70, 60, 20, 20\},$$

$$P_1(\mathbf{k}_2) = \frac{1}{5700} \{5300, 300, 100, 0, 0, 0, 0, 0, 0\},$$

$$P_2(\mathbf{k}_1) = \frac{1}{5900}\{5500, 300, 100, 0, 0, 0, 0, 0, 0\},$$

$$P_2(\mathbf{k}_2) = \frac{1}{5900}\{10, 3010, 2210, 450, 40, 80, 60, 20, 20\}.$$

After redistributing the intercommunity links, we obtain

$$P_1(\mathbf{k}) = \frac{1}{5700}\{10, 2910 - Y, 2120, 480, 10, 70, 60, 20, 20 - Y, Y + R\},$$

$$P_2(\mathbf{k}) = \frac{1}{5900}\{10, 3010 - Y, 2210, 450, 40, 80, 60, 20, 20 - Y, Y + R\},$$

where $Y = 62$ and $R = 4$.

We have average edges $z = \begin{pmatrix} \frac{482}{285} & \frac{5}{57} \\ \frac{5}{59} & \frac{100}{59} \end{pmatrix}$ and the distribution of nodes among the communities is $w = \begin{pmatrix} \frac{57}{116} & 0 \\ 0 & \frac{59}{116} \end{pmatrix}$. More details are shown in Table 1.

	Community 1	Community 2
Total nodes	5700	5900
Total inter community edges	9640	10000
Total intra community edges	500	500

Table 1: Number of edges in each community in model 1

The modularity for this model is $\sigma = 19.634$ or $Q = 0.451$. First, we study the effect of transmissibility rate. In this case, we have the same transmissibility rate for all the links, namely either inter or intra community edges within the same community have the same transmissibility rate. For example, $T = 0.1$ means that $T = \begin{pmatrix} 0.1 & 0.1 \\ 0.1 & 0.1 \end{pmatrix}$. Figure 1 shows the result obtained.

Secondly, we study the effect of clustering in this mode. Using (10), the epidemic threshold is determined to be 0.599. Above the epidemic threshold, clustering will reduce the probability that a randomly chosen community-i node leads to the giant component. In this case we use $T = 0.64$. Figure 2 shows the result obtained.

Thirdly, by using (8), we determine the average number of community-i nodes in the small component, $\langle s_i \rangle$, reached from a randomly chosen node for different clustering coefficient, C, in Figure 3.
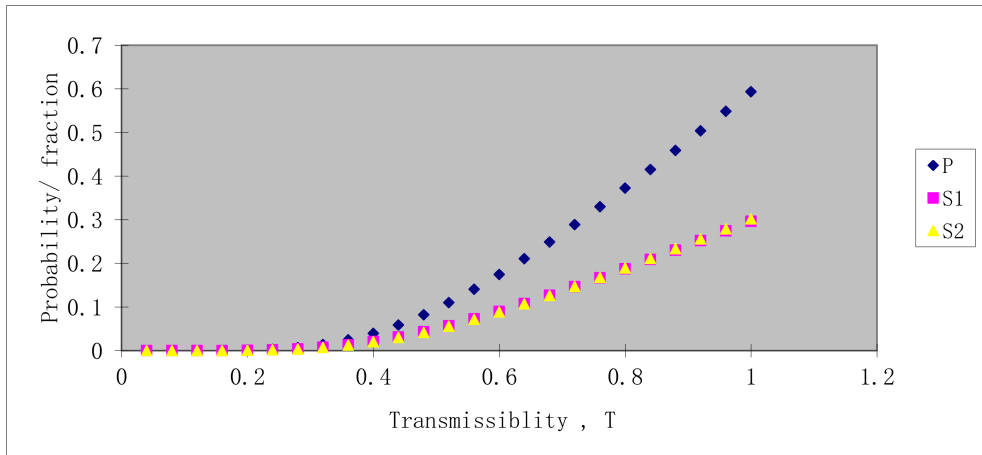
Figure 1: Diagram showing the probability, P, that a randomly chosen community-i node leads to the giant component versus tranmissiblity rate, T, for model 1. The diagram also shows the fraction of giant component in communities 1 and 2 ($S1$ and $S2$) versus tranmissiblity rate, T. The epidemic threshold in this example is 0.599.

Model 2: We summarize the model 2 as follows : There are two communities (i.e. $M = 2$) with probability degree distribution: $P_i(k_1, k_2) \equiv P_i(\mathbf{k})$ as shown below :

$$P_1(\mathbf{k}_1) = \frac{1}{500}\{10, 10, 20, 80, 200, 120, 60\},$$

$$P_1(\mathbf{k}_2) = \frac{1}{500}\{420, 50, 20, 10, 0, 0, 0\},$$

$$P_2(\mathbf{k}_1) = \frac{1}{540}\{460, 50, 20, 10, 0, 0, 0\},$$

$$P_2(\mathbf{k}_2) = \frac{1}{540}\{10, 10, 10, 100, 140, 180, 90\}.$$

For analysis, we distribute the inter-community edges randomly across nodes in each community, we get

$$P_1(\mathbf{k}) = \frac{1}{500}\{10, 10, 10, 45, 205, 135, 85\},$$

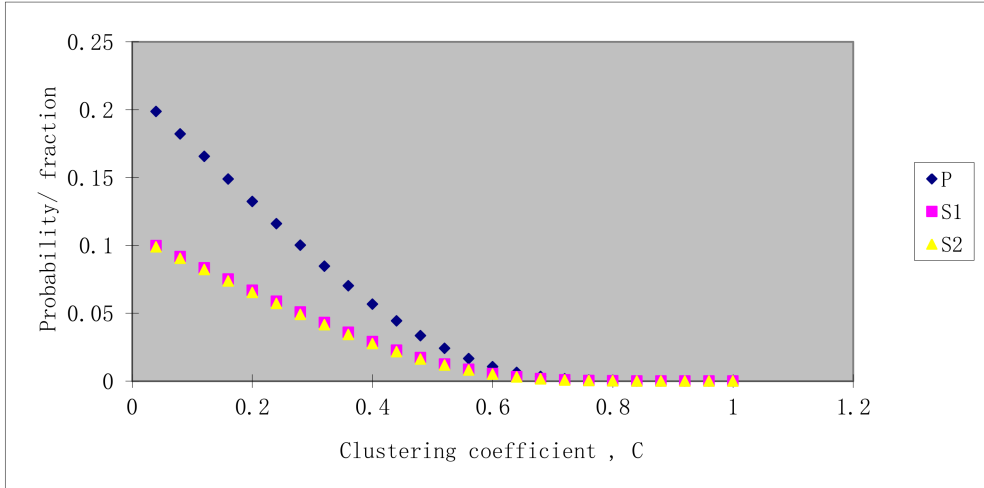$$P_2(\mathbf{k}) = \frac{1}{540}\{10, 10, 5, 85, 110, 185, 135\}.$$

Figure 2: The probability, P, that a randomly chosen community-i node leads to the giant component versus clustering coefficients, C, for model 1 when $T = 0.64$. $S1$ and $S2$ are the fraction of giant component in communities 1 and 2 respectively
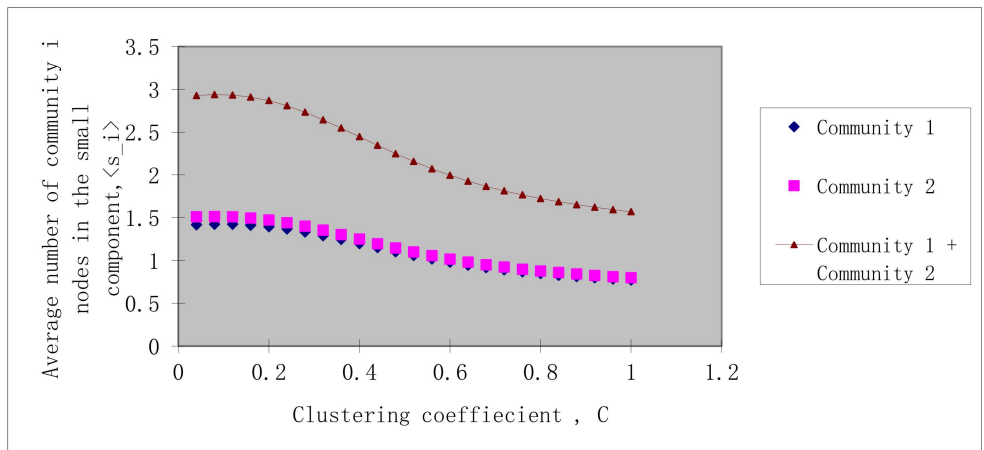

Figure 3: A diagram showing the average number of community-i nodes in the small component, $\langle s_i \rangle$, reached from a randomly chosen node for different clustering coefficient, C, for model 1 when $T = 0.64$.

We have average edges $z = \begin{pmatrix} \frac{41}{10} & \frac{6}{25} \\ \frac{2}{9} & \frac{233}{54} \end{pmatrix}$ and the distribution of nodes among the community is $w = \begin{pmatrix} \frac{25}{52} & 0 \\ 0 & \frac{27}{52} \end{pmatrix}$. More details are shown in Table 2.

	Community 1	Community 2
Total nodes	500	540
Total inter community edges	2050	2330
Total intra community edges	120	120

Table 2: Number of edges in each community in model 2

The modularity for this model is $\sigma = 18.20$ or $Q = 0.446$. For transmissibility, $T = \begin{pmatrix} \frac{3}{10} & \frac{1}{5} \\ \frac{1}{5} & \frac{6r}{10} \end{pmatrix}$. Using (10), we obtain $det(I - A) = -0.0464 + 0.1002r$. When $r = 0.4633$, phase transition occurs, namely the giant component first appears. If $r = 1$, that is when $T = \begin{pmatrix} \frac{3}{10} & \frac{1}{5} \\ \frac{1}{5} & \frac{6}{10} \end{pmatrix}$, there exists a giant component. (i.e. outbreak of disease)

We present the results in Figure 4 and Figure 5.

7. Concluding Remarks

In this paper, we focus on complex networks with an arbitrary joint degree distribution. We study community clustered networks which are more suitable for modelling the disease transmission in community networks with clustering effects. With certain clustering coefficient and beyond the percolation threshold, we obtain the probability that a randomly chosen community-i node leads to the giant component. In the context of contact network epidemiology, this probability refers to the probability that an individual in a community is affected by the infectious disease. We have also derived formulae to calculate the size of the giant component and the average small-component size (excluding the giant component). Taking into account the clustering effect through the free-excess degree distribution, the model shows that the clustering effect will decrease the size of the giant component.
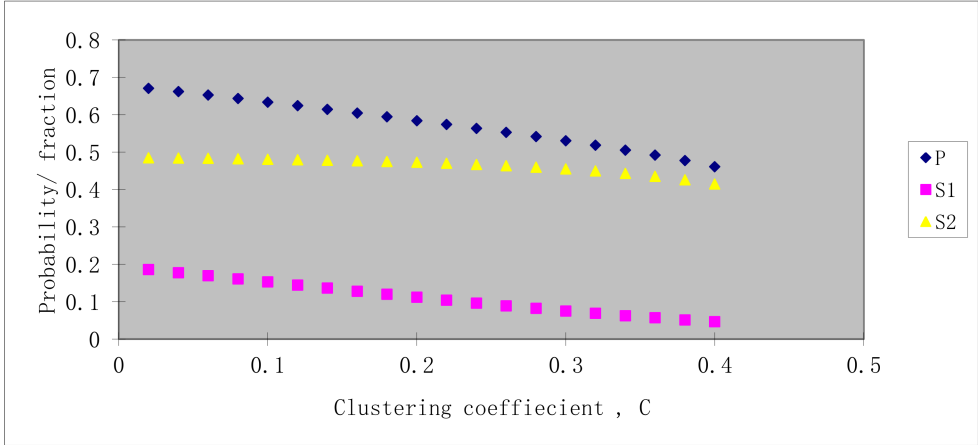
Figure 4: The probability,P, that a randomly chosen community-i node leads to the giant component versus the clustering coefficient,C, for model 2. $S1$ and $S2$ are the fraction of the giant component in communities 1 and 2 respectively.
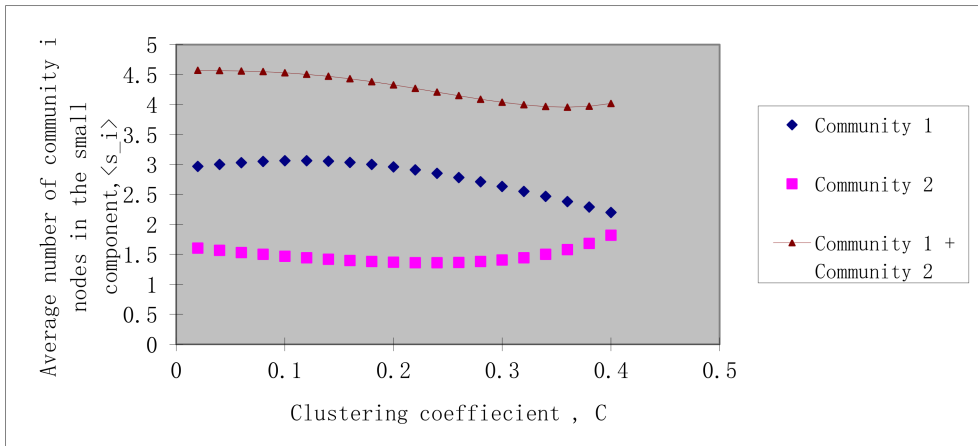

Figure 5: A diagram showing the average number of community-i nodes in the small component, $\langle s_i \rangle$, reached from a randomly chosen node for different clustering coefficient, C, for model 2.

Acknowldgements

References

[1] D.H. Zanette, Dynamics of rumor propagation on small-world networks, *Phys. Rev. E*, **65**, 041908, (2002),1-9. DOI: 10.1103/PhysRevE.65.041908.

[2] X.Y. Wu and Z.H. Liu, How community structure influences epidemic spread in social networks, *Physica A*, **387**, (2008),623-630. DOI:10.1016/j.physa.2007.09.039.

[3] M.E.J. Newman, S.H. Strogatz, and D.J. Watts, Random graphs with arbitrary degree distribution and their application. *Phys. Rev. E*, **64**, 026118, (2001), 1-17.DOI: 10.1103/PhysRevE.64.026118.

[4] M.E.J. Newman. Spread of epidemic disease on networks. *Phys. Rev. E*, **66**, 016128, (2002), 1-11. DOI: 10.1103/PhysRevE.66.016128.

[5] C. Moore and M.E.J. Newman. Epidemics and percolation in small world networks. *Phys. Rev. E*, **61**, (2000), 5678-5682. DOI:10.1103/PhysRevE.61.5678.

[6] A. Vazquez. Spreading dynamics on heterogeneous populations : multitype network approach. *Phys. Rev. E*, **74**, 066114, (2006), 1-11. DOI: 10.1103/PhysRevE.74.066114.

[7] M. Boguna and M.A. Serrano. Generalized percolation in random directed networks. *Phys. Rev. E*, **72**, 016106, (2005), 1-7. DOI: 10.1103/PhysRevE.72.016106.

[8] B. Karrer and M.E.J. Newman. Competing epidemic on complex networks. *Phys. Rev. E* ,**84**, 036106, (2011),1-12. DOI: 10.1103/PhysRevE.84.036106.

[9] J.C. Chen, H.L. Zhang, Z.H. Guan, and T. Li. Epidemic spreading on networks with overlapping community structure. *Physica A*, **391**, (2012),1848-1854. DOI:10.1016/j.physa.2011.10.011.

[10] J. Zhou and Z.H. Liu. Epidemic spreading in communities with mobile agents. *Physica A*, **388**, (2009), 1228-1236. DOI:10.1016/j.physa.2008.12.014.

[11] S.J. Ni, W.G.Weng, and H. Zhang. Modeling the effects of social impact on epidemic spreading in complex networks. *Physica A*, **390**, (2011),4528-4534. DOI: 10.1016/j.physa.2011.07.042.

[12] A. Allard, P.A. Noel, and L.J. Dube. Heterogeneous bond percolation on multitype networks with an application to epidemic dynamics. *Phys. Rev. E* , **79**, 036113, (2009), 1-9. DOI: 10.1103/PhysRevE.79.036113

[13] Y. Berchenko, Y.A.Randrup, M. Teicher, and L. Stone. Emergence and size of the giant component in clustered random graphs with a given degree distribution. *Phys. Rev. Lett.*, **102**, 138701, (2009),1-4. DOI: 10.1103/PhysRevLett.102.138701.

[14] Z.H. Liu and B.B. Hu. Epidemic spreading in community networks. *Europhysics Letters*, **72**, (2005), 315-321. DOI: 10.1209/epl/i2004-10550-5.