

Abstract

The primary aim of this project is to study the effects of network properties on the process of disease spread; especially the effects coming from community structure. Therefore, we will simulate two of the most typical infection disease transmission models SIS and SIR on various network structures and then find out the data relations.

In our project, two main groups of experiments have been done, which tested the effects of the average degree and the community strength to three parameters of epidemic evaluation: epidemic growth rate, duration time and its final size. Additionally, different from previous work on this subject, there are three new properties: overlap rate, population flowing rate and three different community structured networks which are also tested in this project. To generate the network with these different properties, we use two different methods. Moreover, for different community structure construction, the algorithms will have little variation as described in detail in this paper. Furthermore, to decrease the bias of unstable results, every single experiment has been down 500 times to obtain the average value.

Based on this project, we found that the average degree and community strength play a significant role in network-based studies of disease spread. However, the influence coming from other variables such as ‘overlap rate under strong community structured network’ and ‘population flow rate’ are also considerable. In terms of the effects of different community structure with the same community strength, the hierarchical structured network expresses a notable ability to control the epidemic process.

contents

1.	Introduction.....	1
2	Background knowledge	2
2.1	Network.....	2
2.1.1	Introduction.....	2
2.1.2	Social network theory	2
2.1.3	Properties of social network.....	3
2.1.4	Typical real social network	7
2.1.5	Algorithm for artificial network.....	10
2.1.6	Network evolution	16
2.2	Epidemic model.....	17
2.2.1	Introduction.....	17
2.2.2	Types of epidemic models.....	17
2.2.3	Simple deterministic models.....	18
2.2.4	Stochastic epidemic models.....	23
2.3	Chapter summary	26
3.	Previous work on effects of network structure to epidemic process.....	28
4.	Methods	29
4.1	Questions	29
4.2	Experiment design summary	29
4.3	Infection disease data choose	30
4.4	Network generation description	30
4.4.1	Experiment one	30
4.4.2	Experiment two.....	40
4.5	Data structure for network.....	45
4.6	Disease transmission	46
5.	Results and analysis	48
5.1	Epidemic growing rate analysis.....	48
5.1.1	Average degree under given degree distribution.....	48
5.1.2	Community strength (measured by mixing rate in LFR algorithm).....	49
5.1.3	Overlap rate under given community strength.....	50
5.1.4	Population flow rate	51
5.1.5	Different community structure with similar evaluated parameter.....	52
5.2	Epidemic duration days and final size analysis	52
5.2.1	Average degree under given degree distribution.....	53
5.2.2	Community strength (measured by mixing rate in LFR algorithm).....	54
5.2.3	Overlap rate under given community strength.....	55
5.2.4	Population flow rate	56
5.2.5	Different community structure with similar evaluated parameter.....	57
6.	Conclusion and discussion.....	59

7.	Future work.....	61
8.	Bibliography	62
9	Appendix: Source code.....	67

1. Introduction

In understanding the behavior of diseases, particularly infectious diseases, it is possible to arrive at an effective way to control an epidemic process [77]. In recent time, researchers have constructed epidemic models based on network theory that represents contacted individuals (includes by a direct and indirect connection), thus allowing for the possibility to simulate and express the spread process. Hence, network theory is considered indispensable to analyse the disease spread process [40]. Nowadays, most studies about the effects of network structure within infectious disease spread focus on how the properties of network, like degree distribution , edge weights and clustering coefficient, impact on the parameters of an epidemic, such as the epidemic growth rate, and the final size of the epidemic. Meanwhile, community structure is becoming a new study area for how it affects disease spread.

- ***Objectives***

In our project, the aim is to understand how the network structure affects the disease spread process. In detail, we implement two levels of experiments:

1. Spread standard SIS and SIR throughout static network with different properties.
2. Simulate standard SIS and SIR model on the network along with community evolution events, which relate to the concept of dynamic network.

Based on the simulation results, five questions should be answered:

1. How does the average degree affect epidemic growth rate, epidemic duration time and its final size?
2. How does the community strength affect epidemic growth rate, epidemic duration time and its final size?
3. How does the overlap rate in network with overlapped community affect epidemic growth rate, epidemic duration time and its final size?
4. Is the population flow rate a factor which affects epidemic growth rate, epidemic duration time and final size of epidemic?
5. Is there a variance in how different community structures affect the epidemic growth rate, the duration time of the epidemic and the final size of the epidemic?

- ***Structure***

This report consists of five main chapters. Firstly, the basic background knowledge describes the theory that will be used in this project, which is composed of two scientific fields: network theory and epidemic modeling. Next, there will be a briefly summary of the methods and results of previous work. In Chapter 4, the details of experiments which have been done in the project will be proposed. Based on the experiments statistics and data, a reasonable analysis will be provided in Chapter 5. Finally, the project conclusion and future work will be described.

2 Background knowledge

2.1 Network

2.1.1 Introduction

‘Network’ is an efficient way of expressing system structure and relationship, and is all around us, including tangible objects in the Euclidean space, such as the Internet, subway or electric power grids, and neural networks [10]. Alternatively, it can be described as an abstract space containing collaborative relations between individuals or companies [10]. In academic areas, network is used as a basic environment or framework for many other studies, such as biology, computer science, economics and sociology. It is a young and active science based on graph theory. To be exact, network theory uses the graphs as a representation tool for the connection or relations between discrete objects. In recent times, a lot of models based on network theory have been implemented, such as information networks, social networks and biological networks. In the case of this study, we will focus on the social network, which is defined as a structure made up of individuals tied by one or more interdependency, such as common interests, friendship, relationships of beliefs or sexual relationships [38].

2.1.2 Social network theory

In terms of social network theory, the social aspect can be considered as a set of people or organizations with connections or interactions between them [78, 86]. This theory aims at analyzing social relationship represented by network structure that consists of nodes and edges. These relationships include business between companies [52, 60], friendships between individuals [61], intermarriages between families [69] and so on. In a social network graph, the nodes are individuals or actors within the networks, while the edges are the relationships between them.

Different from traditional sociological studies, which focus on the effects or functions of individual actors that consist of the network, the social network theory produces an alternative view, where the relationships between the individuals are more important than the attributes of individuals themselves [55]. This approach has turned out to be useful for explaining many real-world phenomena like news and rumours, diffusion of innovations [21], markets exchange relations [81]. Similarly, it has also been used to analyse the spread of both infectious diseases and health-related behaviours studied within the sphere of medical sociology [89].

In the real world, social network is based on graph theory and belongs to complex network structure field, without trivial topological characters. In other words, this kind of network cannot stand by purely regular or random graphs. During the last few years, scientists have tried to find out how social networks have both the properties shared with non-social networks, such as the World Wide Web, biological networks [23, 68], and the distinctive properties which make social

networks different. In previous areas research has explored the small-world effect [88], and the trend of a network to evolve into a scale-free state that is characterised by heavy tail degree distribution [1]. While the latter part focuses on network clustering or transitivity. Stated by Newman's research [65], the clustering as a result of some social mechanisms is not significant in non-social networks. In addition, community structure has been found and taken into consideration as a symbolised property in social network studies.

2.1.3 Properties of social network

To analyse the network, firstly, its properties need to be considered. This is because these properties will dramatically affect the whole structure of the networks. In other words, to present these properties is a kind of way to represent the network. Hence, we can allow these properties to be variable parameters to describe or represent different networks. This subsection will illustrate the features that appear in common networks of many different types.

2.1.3.1 Small-world effect

In real networks, even though most member nodes are not neighbours with each other, they can be reached from every other node only by a few steps or hops. This is called small-world effect, which has been experimentally proved by Stanley Milgram in 1960s, and more rigorously proved in a mathematical way by Poll and Kochen [74].

Table 2.1: Basic statistics for a number of real networks in different types

	network	nodes	edges	Average Degree	Average path (L)
social	Film actors	449,913	25,516,482	113.43	3.48
	Company directors	7,673	55,392	14.44	4.60
	Email message	59,912	86,300	1.44	4.95
information	WWW nd.edu	269,504	1,497,135	5.55	11.27
	Roget's Thesaurus	1,022	5,103	4.99	4.87
	WWW Altavista	203,549,046	2,130,000,00	10.46	16.18
technologi	Internet	10,697	31,992	5.98	3.31
	power grid	4,941	6,594	2.67	18.99
	train routes	587	19,603	66.79	2.16
biological	metabolic network	765	3,686	9.64	2.56
	protein interactions	2,115	2,240	2.12	6.80

The properties measured are: total number of vertices; total number of edges; average degree and average path L .

To give the exact mathematical description, we use L to denote the mean distance between nodes or vertex in an undirected network:

$$L = \frac{1}{\frac{1}{2}n(n+1)} \sum_{i \geq j} d_{ij} \quad (2.1)$$

where d_{ij} stands for the distance from node i to node j . If a network has small-world feature, the value of L will grow in speed of $\log(N)$, where N is the population size. Recently, the “small-world effect” has been given a more exact meaning: if the value L scales slower with the network size N for a fixed mean degree, this network has the attribute of the small-world effect [68].

To give some real examples to illustrate this phenomenon, in table 2.1, a number of features of different types of real networks have been shown. As we can observe, the value of L in all cases is much smaller than the number of total nodes in a graph, which means every node in the network can be reached from every other node through quite a short path.

This property is very significant and useful when compared with many other networks, including random graph, whose L value increases exponentially with population size N . For instance, if one is doing the research on information spreading across a network, the small-world effects is one of the main reasons that leads to a much faster diffusion speed.

2.1.3.2 Degree distribution

“Degree” is a very important characteristic in graph or network theory, because it affects a lot of other features that lead to an impact on the structure of network. First of all, a definition of “degree” of a graph: the degree is a characteristic for the vertices in a network, which stands for the number of the edges incidentally or connected to a specific vertex.

The most basic topological characterization is “degree distributions”, which is denoted as p_k which stands for the proportion of vertices within the network having degree k or, equivalently, the probability that a node chosen uniformly at random has degree k [10]. Mathematically, if there are N nodes or vertices in a network and N_k stands for the number of them who have degree k , we get the degree distribution:

$$p_k = \frac{N_k}{N} \quad (2.2)$$

Another way to represent the same information, degree distribution, is to present the degree in the form of a cumulative distribution function:

$$p_k = \sum_{k'=k}^{\infty} p_{k'} \quad (2.3)$$

Both of the two ways have advantages and disadvantages [68]. For the first method, it can give a clear and direct visualisation of the distribution, which have no interaction between any neighbouring points in the plot [68]. As a comparison, the second method can present all the original data. Additionally, the cumulative way also decreases the noise.

After giving the definition, some typical degree distribution will be exhibited. The simplest version of degree distribution in a random graph studied by Erdos and Renyi [28, 29], has every edge present or removed in an equal probability. Therefore, the degree distribution is Poisson or binomial in a finite network [68]. By contrast, the networks in the real-world are much more complex than the random one. As showed in figure 2.1 (b), degree distribution represents far away from binomial distribution and, furthermore, is highly right-skewed, which means the shape of the distribution has a long right tail making the value much greater than the mean value. One of them is called power-law distribution that most types of network utilise.

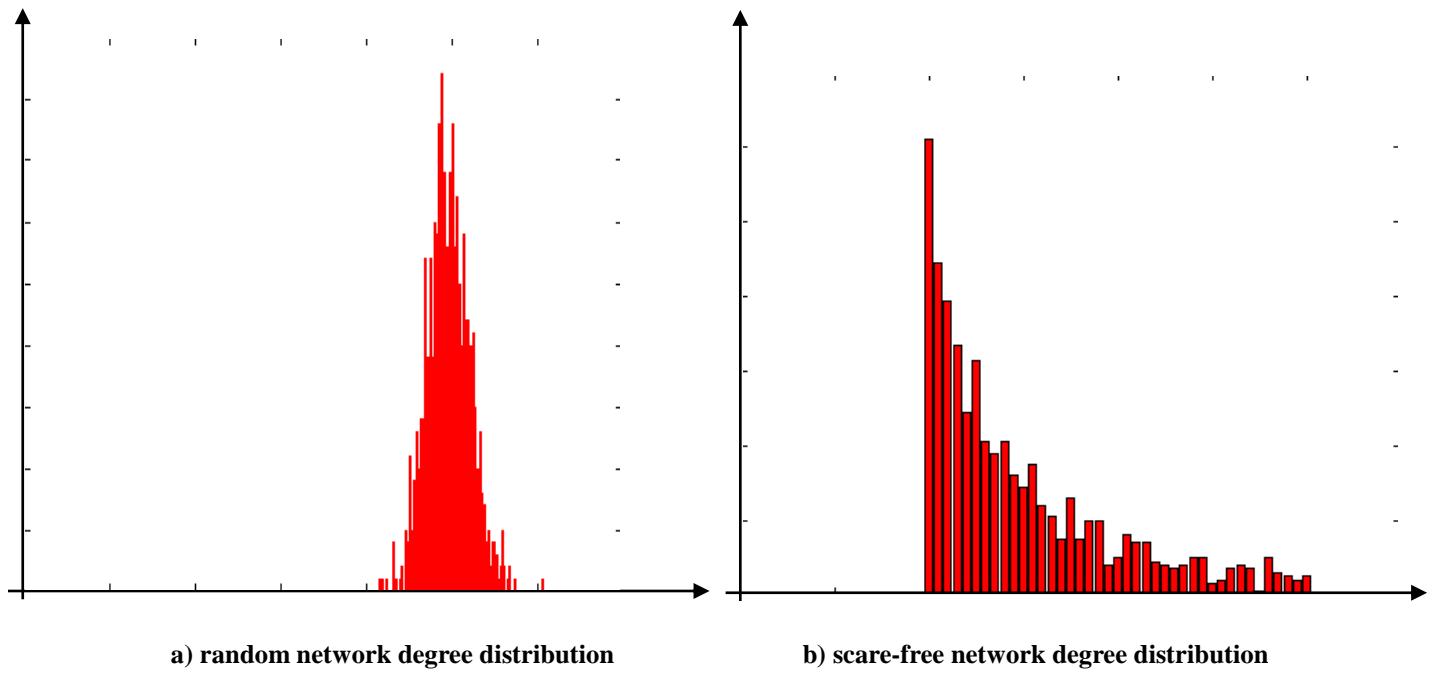


Figure2.1

2.1.3.3 Clustering coefficient

One more significant property for a complex network is clustering, also known as transitivity. It is a typical property of acquaintance networks. In graph or network theory, a clustering coefficient can be used to present the likelihood of nodes in a network which tend to cluster together [86]. In terms of social networks, the clustering represents a phenomenon that two of your friends are also likely to be friends with each other. Evidence from research done by Watts and Strogatz suggests that in most real-world networks, especially in social networks, nodes have the tendency to get together and create tightly-woven groups [88]. In addition, this property in real-world networks is more obvious and significant than in a random network [39].

To consider the clustering coefficient in a geometrical way, it means the presence of a high number of triangles [10]. Mathematically, there are two points of view to illustrate it: the global and the local.

- ***Global clustering coefficient***

The first attempt to measure it was by Luce and Perry [22]:

$$C = \frac{3 * \text{number of triangles in the network}}{\text{number of connected triples of vertices}} \quad (2.4)$$

Here a “triangle” means three vertices that are tied by two undirected edges. A triangle has three triplets. We give an example in fig 2.2:

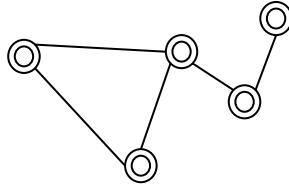


Fig 2.2 this simple network has six triples and one triangle, hence the global clustering coefficient:

$3 * 1/6 = \frac{1}{2}$. For every individual vertices (from left to right), their clustering coefficient is: $1, 1, \frac{1}{3}, 0, 0,$

respectively. And the average value is $c = \frac{7}{15}$

- ***Local clustering coefficient***

An alternative way to define the clustering coefficient is to refer to the local value proposed by Watts and Strogatz [88]. A quantity C_i (the clustering coefficient of node i) representing how likely $a_{jk} = 1$ for two neighbors j and k of node i . Its value equals the ratio between the number of edges e_i in sub graph of node i and $k_i(k_i-1)/2$, where k_i donates the degree of node i , the formulation is given below in 2.5 [87, 88] and 2.6 [88]:

$$C_i = \frac{2e_i}{k_i(k_i-1)} = \frac{\sum_{j,k} a_{ij}a_{jk}a_{ki}}{k_i(k_i-1)} \quad (2.5)$$

$$C_i = \frac{\text{number of triangles include vertex } i}{\text{number of connected triples include vertex } i} \quad (2.6)$$

When a vertex has degree 0 or 1, we put $C_i = 0$. Furthermore, we denote the average clustering coefficient for the network in function 2.7, and give an example as in fig 2.2:

$$C = \frac{1}{n} \sum_i C_i \quad (2.7)$$

2.1.3.4 Community structure

The last property we will introduce here is community structure, which is widely detected in most social networks [30, 86]. It is a concept that individuals in a network tend to concentrate in a few subgroup. This trend allows the groups of vertices connected more densely internally than connections between groups. Moreover, based on the notion of community, we elicit the definition of “clique”. A clique is a complete sub graph of three or more nodes. It is a strong situation of community [10].

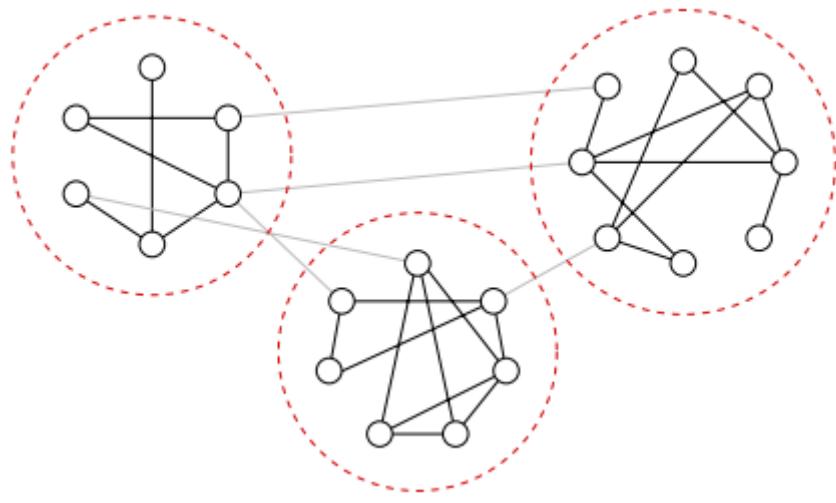


Fig 2.3 in graph $G(n,m)$, a community is a sub graph $G'(n',m')$, whose nodes are connected tightly.

(Source form Girvan, M., and M. E. J. Newman, 2004, ref. [66])

To find communities in network is quite useful to identify substructures or components in a network that provide an insight of network function affects each other [33]. Additionally, this property will affect a lot of activities on the network, such as disease spread [59] or rumours propaganda.

2.1.4 Typical real social network

From now on, we have concentrated on the basic knowledge of the network about its properties and where and why we will use these. In this subsection, some real network structure will be presented, which can show the network concept in a visual way. Especially, the property of community structure is the focus.

Firstly, we show a network of collaborations of scientists working for the Santa Fe institute studied by Girvan and Newman in Fig 2.4. It has 271 vertices which represent each scientist in residence

here from 1999to 2000, the edges are drawn between co-operators. As we can see from the figure, the network has an obvious group-layout. It has a strong density of connections within the research group, whereas few connections between each groups.

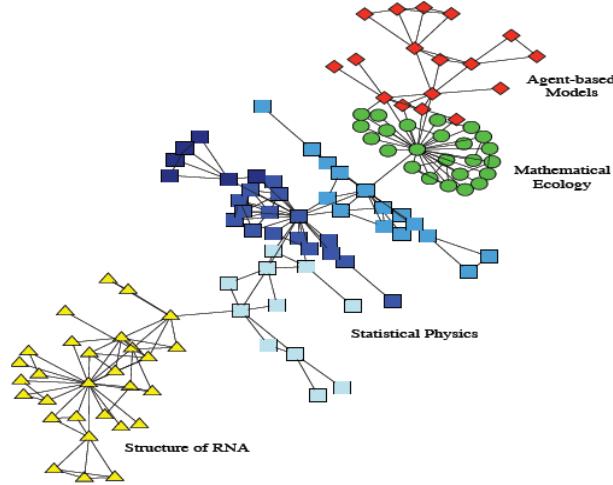


Fig 2.4 Collaboration network between scientists
 (Source form Girvan, M., and M. E. J. Newman, 2002, ref. [33])

Next, Fig 2.5 shows one of the most famous networks for study. These are data collected from the members of a university karate club over two years by Wayne Zachary. During these study years, there was a big disagreement between the club's instructor and administrator, which finally lead to the instructor's leaving and taking about a half of the original members to start a new club. As showed in the picture, the whole network is divided into two main aggregations. One group is almost around node 1 and another is around vertex 33 and 34. But different from the first community structure, firstly, inside the community, there is one hub node; secondly, between the two groups, a number of connections exist between the side (not hub) nodes.

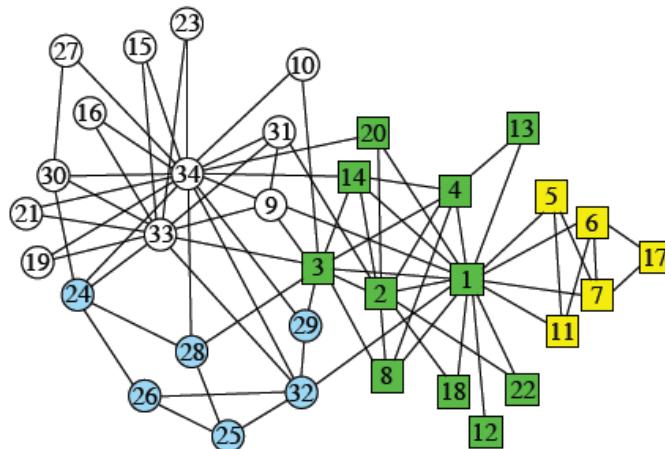


Fig 2.5 Zachary's karate club, a standard benchmark in community detection.
 (Source form Zachary W. 1977, ref. [92])

The next example comes from World Wide Web. In the network, we consider the webpage as nodes and hyperlinks as edges. Seeing from fig. 2.6, the degree distribution is accounted as power-law distribution. Also, similar to the previous network structure, there are hub nodes for every subgroup. But the difference is that groups tend to be connected by hub nodes, not side nodes.

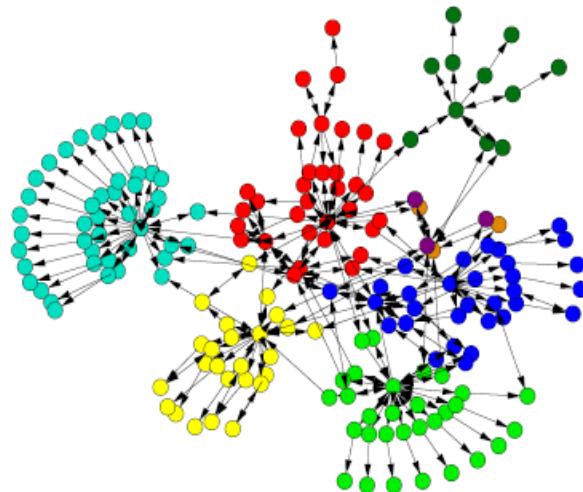


Fig 2.6 Community structure in technological networks
 (source from Newman and Girvan, 2004, ref. [66]).

Furthermore, a network with overlap community structure will be illustrated. In fig 2.7, we show a network for word association derived from “bright” [64]. This kind of structure is suitable for some real-world networks consisting of individuals who belong to more than one subgroup [35].

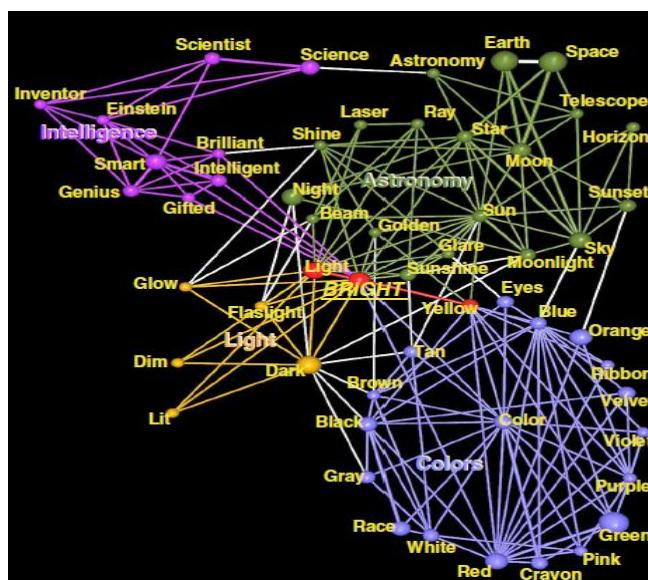


Fig 2.7 Overlapping communities in a network of word association.
 (source from Palla et al., 2005, ref. [37])

Finally, we exhibit a hierarchical community structure in Fig 2.8. As we can see, hierarchical-community network is not a network with simple overlapped communities. They express the phenomena that subgroups exist inside group. Closely woven relationships in small communities can make a big community.

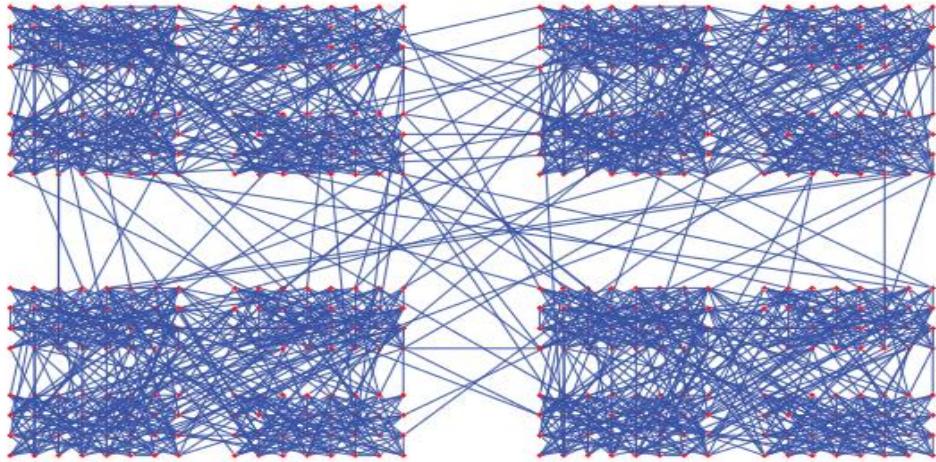


Fig 2.8 a network with a hierarchical structure.
(source from Andrea et al., 2009, ref. [45])

In addition, to focus on disease spread, Bearman and his group provide four typical community structures: core infection model, inverse core model, bridge between disjoint groups model and spanning tree [8]. These models are similar to the examples shown above but have more specific background.

2.1.5 Algorithm for artificial network

To implement the study based on network theory essentially means we need a well-defined network with clear structure and, meanwhile, all the properties of the network should be controlled. Ideally, we can find the real network instances to launch experiments, but it is hard to control every parameter. Hence, some artificial networks with the specific characteristics appear. In this section, some typical artificial network will be constructed using different methods.

2.1.5.1 Random graph

Random network was the first artificial network was made in 1959 by Erdős and Rényi which was named ER random graph. It is based on the probabilistic theory. In their papers, the network was generated with N nodes and K edges which denote as $G_{N,K}$. To construct the network, the first step is initialing N disconnected nodes. Next, they connect pairs of randomly selected nodes until the number of links equals K [27]. Using this way, the given network is only one result of many possible outcomes.

An alternative procedure to generate random model is by connecting each couple of nodes with a

probability p , where $0 < p < 1$. To compare with the first method, it has probability $p^k(1-p)^{N(N-1)/2-K}$ to get K edges in graph [11,28]. And in addition, it is an easier analytical calculation than the previous method.

2.1.5.2 Small-world network

Small-world network is a most-studied model to simulate the real-world. It has small-world property mentioned in section 2.1.3.1. To construct a small-world network, we can build it based on a low dimensional regular lattice, which consisted by L vertices with connections between the nearest k neighbors. And then adding or rewiring links using probability p for every node in turns to make a low density of short path that connect the remote parts in the lattice. When $p = 0$, the graph is a regular lattice, whereas $p = 1$, it becomes almost a random graph [68]. The process is shown in fig 2.9:

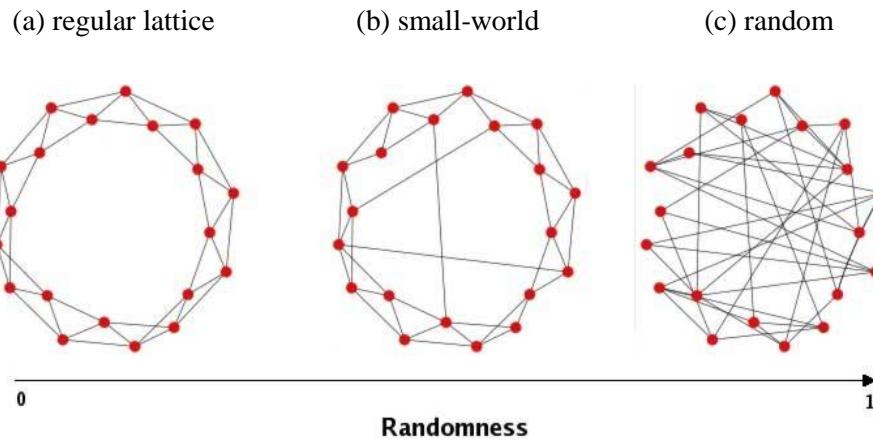


Fig 2.9 (a) an example of a regular lattice with $L = 20$, $k = 2$. (b) an example of small-world network with three shortcuts. (c) an example of random derived from rewiring the (a) graph.
 (Source form Watts & Strogatz, 1998, ref. [88])

- **Clustering coefficient value**

For a small-world network, many properties can be calculated in mathematical way easily. First of all, we express the clustering coefficient value for it. Barrat and Weight [6] proposed that:

$$\mathcal{C} = \frac{3(k-1)}{2(2k-1)} (1-p)^3 \quad (2.8)$$

- **Degree distribution**

Next main aspect is degree distribution. Even though it is not good to represent the real-world networks' feature, not the same goal as scale-free network, we should compute this parameter when describing the network model. The equation provided below:

$$p_j = \begin{cases} \sum_{n=0}^{\min(j-k)} \binom{n}{k} (1-p)^n p^{k-n} \frac{(pk)^{j-k-n}}{(j-k-n)!} e^{-pk} & j \geq k \\ 0 & j < k \end{cases} \quad (2.9)$$

- **Average path length**

Finally we mention the average path length, which is another significant property for small-world network except high clustering. We use ℓ to denote this parameter. Newman and his group give the final solution of ℓ [67]:

$$\ell = \frac{\varepsilon}{2k\sqrt{1+2\varepsilon/L}} \tanh^{-1} \frac{1}{\sqrt{1+2\varepsilon/L}} \quad (2.10)$$

where $\varepsilon = 1/(kp)$.

2.1.5.3 scale-free network

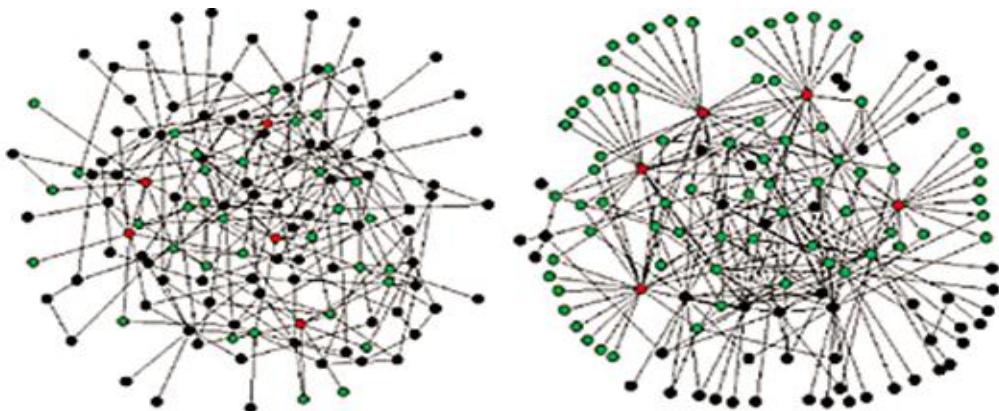
Another typical artificial model is called scale-free network. Following the definition of scale-free network whose degree distribution follows a power law, it is a network contains some nodes that have a tremendous number of connections, whereas most nodes have only few [7].

Mathematically, we use P_k to denote the fraction of nodes in the network having k connections to other nodes, then, the distribution follows:

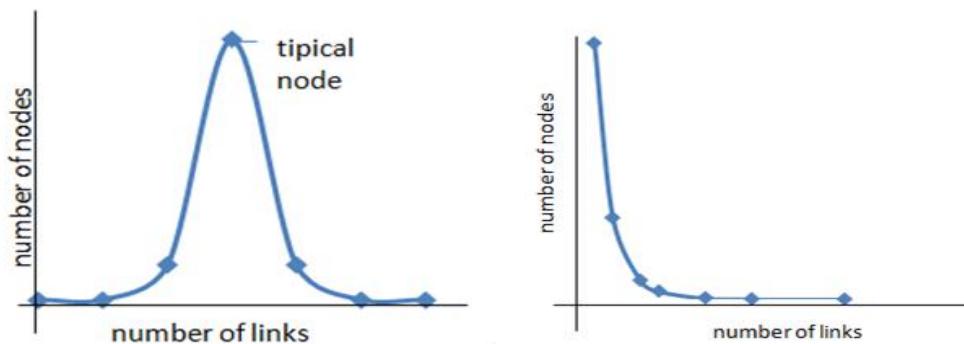
$$p_k = ck^{-\lambda} \quad (2.11)$$

Usually, c is constant and $c \approx (\lambda - 1)m^{\lambda-1}$, and m stands for the lower boundary for the connectivity of a node [18]. In addition, some evidences derived from percolation study in scale-free network proved that the parameter λ is typically in the range $2 < \lambda < 3$ [18].

From now on, the scale-free property have been found in a lot of real networks, including the world wide web [2], metabolic networks [41], transport system, and some social network responsible for spread of disease [25]. Two simple examples compare the random and scale-free networks described in the following to show this model in concise way. The fig2.10 shows their structure with random network on left and scale-free one on right includes hubs (red nodes) with huge number of links. Correspondingly, their degree distribution is expressed in fig2.11.

**Fig 2.10 (a) random network****(b) scale-free network**

(Source from Mike, 2007, ref. [56])

**Fig 2.11 (a) bell-like degree distribution
for random network****(b) power-law degree distribution
scale-free network**

2.1.5.4 Network with community structure: LFR algorithm

Until now, we have constructed some basic network which is usually used in graph and network properties analysis. To do the research with other science areas like biology or sociology, more realistic network needs to be simulated. In this part, we will introduce an algorithm that can construct undirected and unweighted graph with nodes assigned into communities, which is called Lancichinetti-Fortunato-Radicchi benchmark, denote as LFR.

The LFR benchmark is a planted partition model, where the nodes degrees and groups' size follow a power law with exponent t_1 and t_2 , respectively [42]. To realize one LFR benchmark, every input parameter should be mentioned first:

Table2.2: Basic statistics for a number of real networks in different types

parameter	describtion
N	number of total nodes in network
K	average degree
maxk	maximum degree
μ	mixing parameter, which represent the ratio between the external degree and the total degree of the node. When $\mu \rightarrow 0$, the overall network has strong community structure, meanwhile, when $\mu \rightarrow 1$, the community strength decreases.
t1	minus exponent for the degree sequence
t2	minus exponent for the community size distribution
minc	minimum for the community sizes
maxc	maximum for the community sizes
on	number of overlapping nodes
om	number of memberships of the overlapping nodes

A realization LFR network, with 1000 nodes and other parameters $k = 6$, $\text{maxk} = 15$, $\mu = 0.1$, $\text{minc} = 20$, $\text{maxc} = 50$, corresponding to the network shown in Fig 2.12.

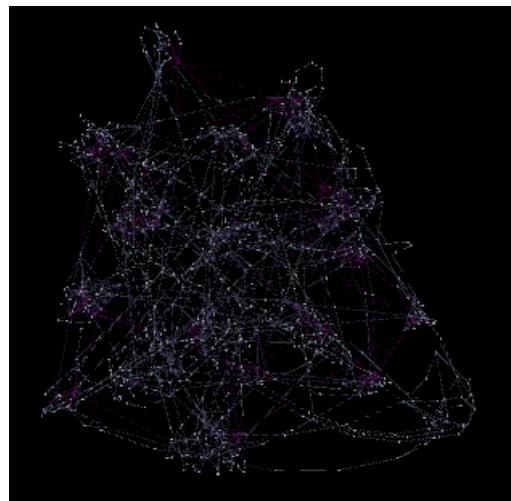


Fig 2.12 network generate by LFR algorithm, with parameter: N=1000,k = 6, maxk = 15, mu = 0.1, minc = 20, maxc = 50.

To launch the process of LFR benchmark includes the following steps [44]:

1. Firstly, all of nodes will be assigned one membership, which according to, in general, power-law distribution. Then, N degrees $\{k_i\}$ built up with power-law distribution with exponent t_1 will be random allocate to N nodes
2. Next, for each node, a fraction $1 - \mu$ of its edges will connect within the same community nodes and a fraction μ edges will link to other nodes of the whole network, where μ is the mixing parameter listed in table 2.

3. Then, assign the communities size $\{s_\omega\}$ by building random numbers from another power-law distribution with exponent t_2 . Obviously, the sum of the node memberships equals the sum of the community sizes $\sum_i v_i = \sum_\omega s_\omega$. Then, every node should choose a community belonging within the limitation: $s_{max} = \max\{s_\omega\} \leq N$ and $v_{max} = \max\{v_i\} \leq n_c$, where N is the total nodes in network and n_c is the number of communities. We use Fig 2.13 to illustrate this process:

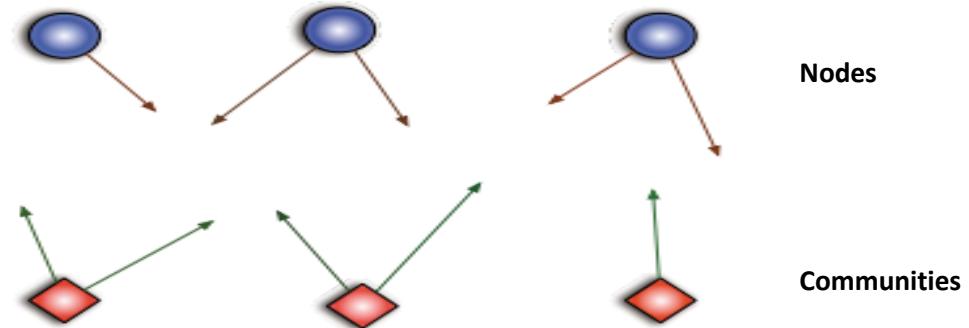


Fig 2.13 process of the bipartite network for allocate nodes to their communities [47]

This picture shows a bipartite network that has two groups: one is N nodes (top) and another is n_c communities (bottom). Each community has s_ω edges, meanwhile each node has as many links as its own memberships v_i . Then, iterate the rewiring process for the bipartite network until the constriction $\sum_{i \rightarrow \omega} s_\omega \geq k_i^{(in)}, \forall i$ is satisfied.

4. Next, each community will be generated one by one. Every communities have degree set $\{k_i(\omega)\}$, which can construct by the configuration model [63].
5. Finally, we will build the links between different communities with degree sequence $\{k_i^{(out)}\}$, where $k_i^{(out)} = \mu_i k_i$.

The LFR benchmark is based on Girvan and Newman (GN) benchmark and adds more complex properties of network with only linear computation time [46]. To know more detail about the algorithm, you can download the source code and read more article from fortunate's homepage: <http://sites.google.com/site/santofortunato/>

2.1.6 Network evolution

After experiencing some typical static network structure, we should consider a more complicated situation: community evolution and dynamic. Even though the analysis of community evolution is still in the infancy stage, many scientists have already recognized its importance. Usually the network structures are relatively fixed in time, but they are inherently changeable [17]. Because the edges within the network can be cut, created and rewired and the nodes can update their states [17].

Moreover, to consider a more complicated structure, a multi-mode network that composed by multiple heterogeneous social actors [82], co-affecting and changing ability of the network is expressed more significantly. Therefore, the community evolution should be accounted for as an important factor when analysing some specific network, such as co-disease spreading between multi-actors, cluster of frequently interacting authors in the blogosphere [49], and mobile subscriber network [90].

From Fortunato's viewpoint [34], the events which happen in the lifetime for communities are birth, growth, contraction, merger, split and death, which showed in fig 2.14. In other words, all these events can affect the evolution process of a community.

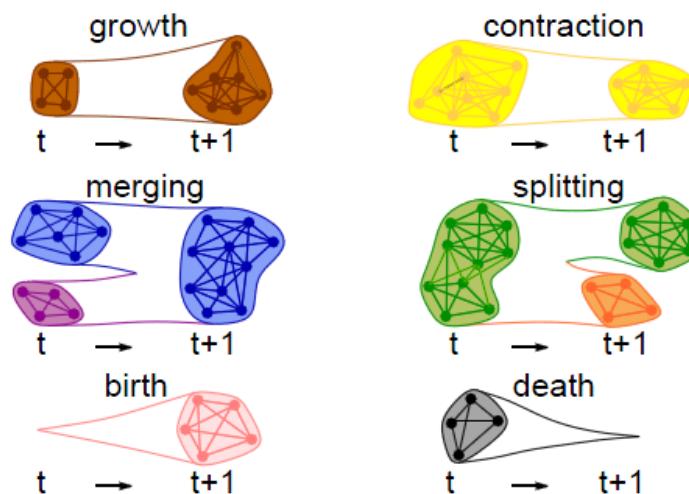


Fig 2.14 Possible change for community evolution.

(source from Palla et al., 2007, ref. [70])

To simulate the evolution process is to implement these events on a constructed network. Some simple way like cutting the links between nodes, membership switching has been attempted by scientists. For instance, Duan and his team created random weighted directed graphs with embedded community structure. Then change the community structure within four different time stages [24]. Furthermore, Tang described an approach for generating a multi-mode network. He created a set of latent communities as the basic network, and then randomly changed a proportion of community memberships at each stage.

2.2 Epidemic model

2.2.1 Introduction

In terms of epidemiology, the traditional way of studying disease spreading is based on Newtonian physics [91], which means that when we consider the effects of the disease only in a straightforward way, that is, all the connection or interactions are either regarded as second-order processes [91] or are ignored. Furthermore, in the past, the epidemiology study is on the population (whole or global) level [72], which has a big limitation to express the process of disease spread. More recently, a new viewpoint that considers the infectious disease process in nonlinear way [73] point out that the traditional thinking of disease spread processes is seen as insufficient. By contrast, the modern epidemiology is based on the individual level to simulate the disease process. To be explicit, in a complex system, everything or every individual is variant or changeable which can be marked by different states. Furthermore, points or individuals in a system are interdependent and affect each other which cannot be interpreted by static and linear models [75]. Nowadays, in order to illustrate the mechanisms of infectious disease spread, some nonlinear mathematical modeling is proposed, which allows us to describe the complex disease spread process and the inter-effects between individuals and, moreover, they have been proved as very useful tools to investigate the way diseases outbreak and spread to predict the future course in order to control an epidemic.

2.2.2 Types of epidemic models

2.2.2.1 Deterministic compartmental Epidemic Models

A deterministic compartmental model (DCM) can be defined as a model composed by individuals, which are classified into different compartments (states) depending on the specific disease which are under study [85]. Furthermore, the deterministic model is often expressed as differential equations; this is because the individuals whose change rate from one compartment to another are denoted as derivatives. And a solution of the DCM is a function of time, and, moreover, is generally uniquely dependent on the initial data [51].

To simulate a deterministic model, all the changes of compartments must be calculated using only the history data [13]. In other words, the process of disease spread must be deterministic.

2.2.2.2 Stochastic Epidemic Models

But the real life epidemic process has various random characters. Consequently, we create stochastic models which are used to calculate the probability of the final events [85]. The events include, for example, the probability distribution of final epidemic size, the probability distribution of the disease vanishing time, and the associated mean [85]. In other words, the probability distribution for these events is the solution of a stochastic model.

To make it clear, the aim of stochastic modeling is to describe a stochastic epidemic process. And a stochastic process is a collection of random variables:

$$\{X_t(s)/t \in T, s \in S\} \quad (2.12)$$

Where T stands for time, either discrete or continuous:

$$T = \{0, 1, 2, 3, \dots\} \text{ or } T = [0, \infty) \quad (2.13)$$

And S represents a sample space or the events mentioned before.

The stochastic models, in contrast to the deterministic version, simulates the process using the idea of uncertain modeling. The study of stochastic epidemic models is based on probability theory [51]. Furthermore, stochastic models usually used to expose the inherent features of demographics or environment variability.

2.2.3 Simple deterministic models

Some deterministic models, proposed in 1927, by W. O. Kermack and A. G. Mckendeick [43], are well-known epidemic models which can illustrate the typical disease spreading process. According to the different properties of diseases that concern immunity or without that, models have different structure. The terminology SIR model is used to represent a disease which considers immunity against re-infection. Furthermore, the SIS is used to indicate a process that an individual is infected from the susceptible class to the infective class and then recover to the susceptible class again, which means this kind of disease cannot get immunity. Additionally, the SEIR and SEIS stand for the models which include an exposed period between susceptive state and infective state. They can be considered as variation versions of the basic SIR and SIS model.

This part will introduce the simplest model: SIS and SIR without considering the birth and death and environment factors. And then, we will incorporate the demographic element into models.

2.2.3.1 The SIS Epidemic Model

In SIS epidemic models, all of the individuals are classified into two classes (according to the disease situations), susceptible or infectious, denoted by S and I , respectively. Normally, the SIS model is used for sexually transmitted diseases [85].

The mechanism of simple SIS model:

As showed in fig2.15, at initial time, all the individuals are marked in a state or classified in a class randomly, either S or I state. If a susceptible individual has an infectious neighbor, it has β probabilities to be infected. Then after a successful infection process, it turns into infectious state, but does not have any immunity to the disease. Therefore, after the time of recovery, the infectious individuals return to a susceptible state. As shown in the picture, the infected individuals have

γ rate to recover into susceptive class per unit time.

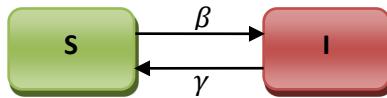


Fig2.15 mechanism of compartmental SIS model
(source from Fred Brauer, 2000, ref. [14])

The mechanism of SIS model with births and deaths:

Taking the demography factors into account, we get the SIS with births and deaths model as below. The diagram 2.16 shows its mechanism. Solid arrows stand for infection or recovery while dotted arrows stand for births and deaths.

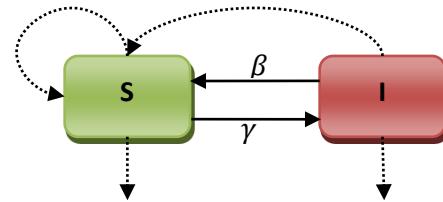


Fig2.16 mechanism of compartmental SIS model with births and deaths
(source from Linda J. S. Allen, 2000, ref. [50])

Additionally, depending on the mechanism of this model, we will get the differential equations for SIS model with a pre-assumption that $S(t) + I(t) = N$ is the total population size.

$$\begin{aligned} \frac{dS}{dt} &= -\beta SI + \mu(I + R); \\ \frac{dI}{dt} &= \beta SI - (\mu + \gamma)I; \end{aligned} \quad (2.13)$$

The new symbol μ stands for the average death rate. And in this model, we assume that the birth rate equals to the death rate to illustrate it in a simple way, hence, $dN/dt = 0$.

Furthermore, the solution of this model is also determined by the breakout threshold called R_0 . Let $S(t)$ and $I(t)$ be a solution to this model, then:

- ◊ If $R_0 < 1$, then $\lim_{t \rightarrow +\infty}(I(t), R(t)) = (N, 0)$, the infection dies out, (disease free equilibrium).
- ◊ If $R_0 > 1$, then $\lim_{t \rightarrow +\infty}(S(t), I(t)) = (\frac{N}{R_0}, N(1 - \frac{1}{R_0}))$, there is an epidemic. (endemic equilibrium).

- **Epidemic break out threshold R_0**

In epidemiology, R_0 is called the basic reproduction number. It is a number of an epidemiological meaning that how many people will be secondary infected by a single infective individual enter into a wholly susceptible population [14].

theorem1: the $\frac{\beta S(0)}{\gamma}$ is a threshold in this case, denoted by R_0 .

If $R_0 < 1$, the infection dies out, while if $R_0 > 1$ there is an epidemic.

Proof:

To begin, we know the $S(t)$ and $I(t)$ should be non-negative. Hence, either $S(t)$ or $I(t)$ reaches to zero the process is consider be terminated. As be experimented by [14], the number of infective people increases as long as $S > \gamma/\beta$; but since the susceptive number decreases through the whole period, I will decrease finally and approach to zero. To sum up, if $S(0) < \gamma/\beta$ or $R_0 < 1$, I decreases to zero which means there is no epidemic, while if $S(0) > \gamma/\beta$ or $R_0 > 1$, I first increase to a maximum level and then decrease to zero, which stands for epidemic.

2.2.3.2 The SIR Epidemic Model

In a simple way, the SIR model can be derived from the SIS model by adding R class with considering the infected people after recovers will have the immunity against the disease. Hence, the SIR model have three compartmental that Susceptible - Infected – Recovered and also can be described as Susceptible - Infected – Removed. In the second situation, we allow the people in the population die from it [32]. In another words, the whole population size is changeable. Examples of diseases suitable for the first class include influenza and chicken pox. The famous diseases like HIV or the Bubonic Plague that lead most people die from contracting it belong to the second category.

The mechanism of SIR model:

To keep things simple, we consider the SIR model without births and deaths situation. In the SIR epidemic model, we classify all the population in to three classes, labeled S , I , and R , which represent susceptive, infective and recovered, respectively. Moreover, let $S(t)$, $I(t)$, and $R(t)$ denote the number of individuals who are in the state S , I , and R in the time t . At initial time, all the individuals are classified in a class, S or I , randomly. If a susceptible individual has an infectious neighbor, it has β probabilities to be infected. Then after a successful infection process, turn into infectious state; every individual in the infectious class has γ probabilities to recover, which depends on the specific disease usually. Then, after recover from infected state, the individual turn into recovered class, which means that it has the immunity to the disease and will not be infected again. The flow of this model is shown in Fig 2.17.

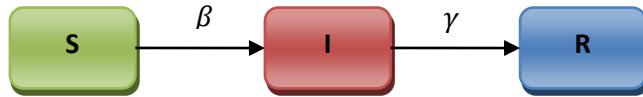


Fig 2.17 flow chart for the SIR model
 (source from Fred Brauer, 2000, ref. [14])

Same as SIS model, fixed the population size, $N=S(t)+I(t)+R(t)$, then the model can be described in the following differential equations:

$$\begin{aligned}\frac{dS}{dt} &= -\beta SI; \\ \frac{dI}{dt} &= \beta SI - \gamma I; \quad (2.14) \\ \frac{dR}{dt} &= \gamma I;\end{aligned}$$

Additionally, this model is based on four assumptions:

- 1) Every individual in the population has an equal probability to contact the disease with rate β , which is called the disease infection rate. Hence, an infected individual is able to transmit the disease with βN people per unit time.
- 2) The infectious individual has the γ probability to leave this class, γ is usually depends on the specific disease that under study.
- 3) For the second and third equations, we assume that all the individuals leaving from the susceptible class will turn into the infected class. And the number of leaving from the susceptible class is equal to the number of entering into the removed class per unit time.
- 4) The population size is considered as fixed.

Looking at the assumption (2), we need give it a full mathematical explanation to make it more reasonable for epidemiological meaning. We use $C(s)$ stand for the number of the individuals who are in the infectious class for s time units continually. If a rate γ to leave from the infectious class per unit time, then:

$$\frac{dc}{dt} = -\gamma C; \quad (2.15)$$

Hence, the solution of this differential equation is:

$$C(s) = C(0)e^{-\gamma s} \quad (2.16)$$

which shows that the portion of the infected individuals remaining in the infectious class for s time units is $e^{-\gamma s}$. Therefore, we can explain the assumption (2) in a way with epidemiological meaning that the length of the disease infective period is accord to an exponential distribution with mean $\int_0^\infty e^{-\gamma s} ds = \frac{1}{\gamma}$.

For the assumption (4), it can be also explain as that the rate of infection and recovery process is significant faster than the birth and deaths time scale, which allow us to ignored the demography factors.

The mechanism of SIR model with births and deaths:

The previous subsection describes the basic SIR model which considers the simplest situation. In that model, we have dropped births and deaths factors to maintain the population size stable. But for diseases that are prevalent in some region or in term of endemic must be considered in another way. Some reasons: firstly, with small size of the total population, the effects of vaccine or quarantine is significant to weaken the epidemic [14]; in addition, a flow of new infected or susceptible individuals enter into the population need to be think about [14]. Finally, a larger proportion (when considering that the whole population size is small) of people will be dead from the disease, which should allow the death be taken into account.

The main process is as same as the simple SIR model, only incorporate the shift or enter and departure path for births or deaths. The diagram 2.18 shows its mechanism. Solid arrows stand for infection or recovery while dotted arrows stand for births and deaths.

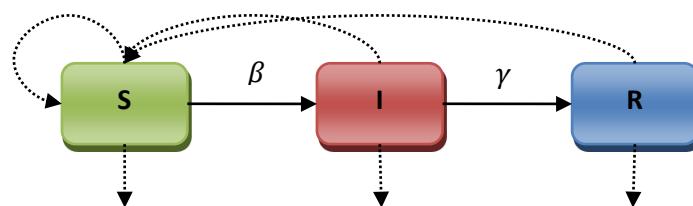


Fig 2.18 SIR compartmental model with births and deaths.

(source from Linda J. S. Allen, 2000, ref. [50])

Using the differential equations to describe its mechanical with assumption that the total population size $N=S(t)+I(t)+R(t)$:

$$\begin{aligned}\frac{ds}{dt} &= -\beta SI + \mu(I + R); \\ \frac{dI}{dt} &= \beta SI - (\mu + \gamma)I; \quad (2.17) \\ \frac{dR}{dt} &= \gamma I - \mu R;\end{aligned}$$

Similar with SIS model, we only add the third equation to the previous solutions.

In this case, $R_0 = \frac{\beta}{\mu+\gamma}$, where $1/(\mu + \gamma)$ stands for the length of the infectious period [14].

According to theorem1, the solutions to these differential equations are:

- ✧ If $R_0 < 1$, then $\lim_{t \rightarrow +\infty}(S(t), I(t), R(t)) = (N, 0, 0)$, the infection dies out, this point is defined as disease free equilibrium.
- ✧ If $R_0 > 1$, then $\lim_{t \rightarrow +\infty}(S(t), I(t), R(t)) = (\frac{N}{R_0}, \frac{\mu N}{\mu+\gamma}(1 - \frac{1}{R_0}), \frac{\gamma N}{\mu+\gamma}(1 - \frac{1}{R_0}))$, there is an epidemic. This point is called endemic equilibrium.

Additionally, we give out second theorem here:

Theorem 2: Assume $\mu = 0$. if $R_0 \frac{S(0)}{N} > 1$, the number of individuals in infected class will increase (epidemic). On the other hand, if $R_0 \frac{S(0)}{N} \leq 1$, the infective number decrease monotonically (no epidemic).

Here we define the $R_0 S(0)/N$ as the initial replacement number. It gives the meaning of the average number of secondary infections produced by an infected individual at the start of the epidemic [36, 37]. As we can see from the theorem 2, if the replacement number is smaller than one, the epidemic will vanish ultimately. By contrast, if the data is bigger than one, the population will suffer an epidemic outbreak.

2.2.4 Stochastic epidemic models

From now on, we have described the most famous deterministic compartment models (SIR and SIS). But in some situations, the deterministic model is not sufficient. There are some reasons why the stochastic idea has been proposed. Firstly, for example, real epidemics in our life can either end up because of a limited number of individuals are infected, or go extinct with a good isolation measures that making a large proportion of the population have been protected, which help to stop the disease a lot. It is only stochastic models that can express this kind of characters and the probability of each event taking place [16]. Additionally, stochastic models are often used when we want to show the inherent features of the demographics or environment variability, especially

when the population size or initial number of infective individuals is very small [51]. Finally, Demiris points out that stochastic model are more intuitively logical to define, because that they instinctively describe the contact processes between individuals [19].

The previous subsection has introduced some simple deterministic models. The aim of this part will provide the stochastic version that directly relate to their deterministic counterparts. There are many techniques for construct the stochastic, such as discrete time Markov chain (DTMC), continuous time Markov chain (CTMC), and stochastic differential equation (SDE). To be simple, we will only derive stochastic SIR and SIS using the first method.

2.2.4.1 Assumption and Terminology for stochastic model

The stochastic model will be described in the following part:

Assumption1: the total population size N is constant. To make it simple, we decide the birth rate equals to the death rate.

Assumption2: same as the deterministic model, $S(t) + I(t) + R(t) = N$. In other words, all the individuals should be classified in the three groups (remove R class in SIS model).

Then, we give some notations:

- $S(t)$, $I(t)$, $R(t)$ represent discrete random variables for the number of susceptive, infected, and recovered individuals at time t , respectively, where $t \in \{0, \Delta t, 2\Delta t, \dots\}$:

$$S(t), I(t), R(t) \in \{0, 1, 2, \dots, N\} \quad (2.18)$$

- β : contact rate; γ : recover rate; μ : average death rate.
- $P_i(t)$ stands for the probability of the infected number equals i at time t :

$$P_i(t) = \text{Prob}\{I(t) = i\} \quad (2.19)$$

Therefore, $i = 0, 1, 2, \dots, N$, and $\sum_{i=0}^N p_i(t) = 1$

2.2.4.2 SIS model

In a DTMC SIS model, because of the constant total population size, using the assumption 2 that $S(t) = N - I(t)$, we can consider there is only one independent variable, $I(t)$. Then the question change to that we should express the stochastic process of $I(t)$ from $t = 0$ to $t = \infty$.

First, we make one more assumption: the changing process of infected number $I(t)$ has the Markov property. Therefore the number at time $t + \Delta t$ only depends on one previous time step t :

$$\text{Prob}\{I(t+\Delta t)/I(t), I(t-\Delta t), \dots, I(0)\} = \text{Prob}\{I(t+\Delta t)/I(t)\} \quad (2.20)$$

Then, in order to complete the model, we provide $P_{ji}(t+\Delta t, t)$ to stand for the probability of the number of infected individuals which is j at time $t+\Delta t$ with a precondition that the infected number is i at time t :

$$P_{ji}(t+\Delta t, t) = \text{Prob}\{I(t+\Delta t) = j | I(t) = i\} \quad (2.21)$$

Next, to let the model be described, we decrease the time step Δt to small enough so that changes of the number of infected individuals is at most one during Δt : $j = i+1$ or $j = i-1$ or $j = i$.

From now on, we can illustrate the stochastic SIS epidemic model in fig 2.19:

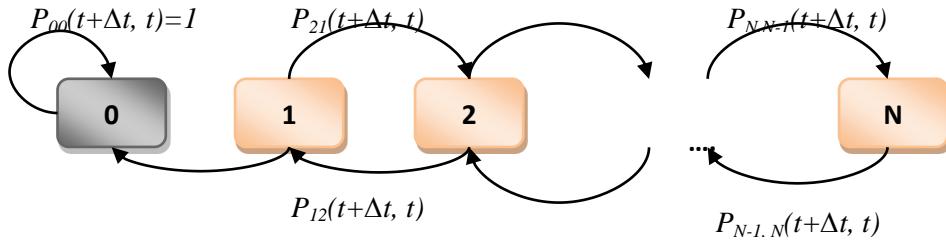


Fig 2.19 stochastic SIS epidemic model.

(source from Linda J. S. Allen, 2000, ref. [50])

As we can see from the fig 2.19, the change of infected number, $\{I(t)\}_{t=0}^{\infty}$, can be described as a Markov process. Some features of this model can be easily explained by Markov chain theory [4, 83]. This picture says the final state must be 0, which means, in DTMC SIS model, the result will reach to disease-free equilibrium, except when the basic reproduction number is significant large [50]. But depends on the parameter and initial conditions, the duration of the epidemic can be very long. In another words, the convergence rate of the disease-free equilibrium can be very low.

Now, we give the mathematically description for stochastic SIS model. The transition probabilities between states satisfy:

$$P_{ji}(\Delta t) = \begin{cases} \frac{\beta i(N-i)}{N} \Delta t, & j = i+1 \\ (\mu + \gamma)i \Delta t, & j = i-1 \\ 1 - [\frac{\beta i(N-i)}{N} + (\mu + \gamma)i] \Delta t, & j = i \\ 0 & j \neq i+1, i-1 \end{cases} \quad (2.22)$$

To make it simple, we substitute $\frac{\beta i(N-i)}{N}$ as $b(i)$, which stands for the transition probability for a new infection. While substitute $(\mu + \gamma)i$ as $d(i)$, which means a death or a recovery [50]. Then, let $p(t) = (p_0(t), p_1(t), \dots, p_N(t))^T$ represents the probability vector for $I(t)$. Moreover, we use a matrix to explain the transition probability:

$$\begin{pmatrix} 1 & d(1) \Delta t & 0 & & 0 & 0 \\ 0 & 1 - [b(1)+d(1)] \Delta t & d(2) \Delta t & \cdots & 0 & 0 \\ 0 & b(1) \Delta t & 1 - [b(2)+d(2)] \Delta t & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & d(N-1) \Delta t & 0 \\ 0 & 0 & 0 & \cdots & 1 - [b(N-1)+d(N-1)] \Delta t & d(N) \Delta t \\ 0 & 0 & 0 & \cdots & b(N-1) \Delta t & 1 - d(N) \Delta t \end{pmatrix}$$

Now, the DTMC SIS model is completely constructed. Therefore, we can simulate the change process of $I(t)$ through the epidemic time. At any time $t \in [0, \infty)$, we get:

$$p(t+\Delta t) = P(\Delta t) p(t) = P^{n+1}(\Delta t) p(0) \quad (2.23)$$

Then, we should only obtain an initial probability vector $p(0)$.

2.2.4.3 SIR model

Similar with SIS model, in DTMC SIR model, we consider two random variables $\{S(t), I(t)\}$ [82]. Then, our aim is to trace the change of S and I number in the Markov process.

Let $p_{(s,i)}(t) = \text{Prob}\{S(t)=s, I(t)=i\}$, where $s, i = 0, 1, 2, \dots, N$. Next, we define the transition probability under an assumption that Δt is small enough that at most one change will happen during time step t :

$$p_{(s+k,i+j),(s,i)}(\Delta t) = \text{Prob}\{(\Delta S, \Delta I) = (k, j) / (S(t), I(t)) = (s, i)\} =$$

$$\begin{cases} \beta i (N-i) \Delta t / N, & (k, j) = (-1, 1) \\ \gamma i \Delta t, & (k, j) = (0, -1) \\ \mu i \Delta t, & (k, j) = (1, -1) \\ \mu (N-s-i) \Delta t, & (k, j) = (1, 0) \\ 1 - [\frac{\beta i (N-i)}{N} + \gamma i + \mu (N-s)] \Delta t, & (k, j) = (0, 0) \\ 0, & \text{otherwise} \end{cases} \quad (2.24)$$

Then, same with SIS model, giving the initial situation $p(0)$, we can calculate the whole path of the change of infected and recovered numbers.

2.3 Chapter summary

This chapter has collected and introduced two aspects of knowledge which will be used in our project: network structure especially community structure and epidemic model. Both of them have well developed and various models which have been proposed.

In the theory of network part, we described four main properties of a social network that was studied widely and affected the structure of the network dramatically. In addition, these properties can be changed as input parameters in order to change the network structure. Then, we showed various real social networks, especially focus on the different types of community structure. Next, four network generation algorithm were proposed: random graph generation, small-world network, scale-free network and network with community structure. These algorithms are some of the main flow of artificial network generation methods. Finally, to give one more step to the research, the concepts of dynamic and evolution of community structure in network were proposed, which open a new area for network study.

Next, we use one section to illustrate the knowledge system of epidemic model. Two different ways for constructing the model: deterministic and stochastic methods. The deterministic method is the basic theory for epidemic science, whilst stochastic concept can be used to simulate the real situation of epidemic process. In addition, for network based epidemic processing, stochastic concept will be adopted. Next, we describe SIR and SIS model in these two ways that depending on whether the disease have or have no immunity state, respectively. Additionally, we can add some classes or states when it is necessary for describing some specific diseases. For example, for diseases like HIV, we can add a class called “exposed” which stands for a latent period. This model is called SEIR model. And also MSIR model, using for measles, adds a maternally-derived immunity class for babies are not born into the susceptible class but has immunity to the disease for the first few months. Hence, these two models are the basic version of other derived epidemic model.

3. Previous work on effects of network structure to epidemic process

Based on the background knowledge of network and epidemic model, this chapter will collect and summarize some previous work in this area, which can help us to clear our question and object. In addition, we can learn from their work and improve some part to design our experiment.

Until now, a lot of studies about the effects of network structure to infectious disease spreading have been launched, which combine and use the two parts knowledge in background section. Most of them focus on how some properties of network, like degree distribution and clustering coefficient, impact parameters of an epidemic (eg. the basic reproductive ratio R_0 , the size of the epidemic and so on). Meanwhile, some researchers find out that the community structure is a main factor to affect disease spreading process.

- ***Methods used***

A typical simulation for disease spreading based on network consists of the following sequence: network generation, a disease model with specific limitation (eg. Recover rate, infection rate) running on the network. During one iteration, individuals' states will change between totally assumed situations. Moreover, births and deaths may occur.

Among these researches, different ways of network generation are mentioned. In paper [15], the epidemics process was launched on a random graph with tunable clustering. More complex, in Mills work of TB and HIV analysis [59], they use a method described by Read and Keeling in [47], which is characterized by a locality parameter D. In addition, the network with homogeneous nodes and network with heterogeneous nodes appear in [57].

- ***Result and achievement from now on in this area***

From different experiments, some result got. In paper [83], Miller point out that clustering is a dominant element controlling the epidemic growth rate. In addition, he mentioned that the effects of heterogeneity and edge weights to disease spreading are minimal when the network structure strongly clustered. Furthermore, similar work did in [2], Salathe et al. simulated the process of disease spreading on network with strongly community structure, and they conclude that community structure has a major impact on epidemic process. Additionally, they found that immunization interventions for individuals bridging communities are better than simply using for highly connected individuals. Moreover, Britton and his group use empirical and investigate the variation of epidemic threshold and probability of a large outbreak with the clustering in the network and find that as the clustering increases, the epidemic decreases [15]. More researches have been done in [26, 62, 80]

4. Methods

To launch the experiments, in the same way as the previous work, two main parts include: social network generation and epidemic dynamic model simulation. In this part, firstly, we will make our question through reification. Then, detail methods will be proposed.

4.1 Questions

Based on the background knowledge of network and epidemic model, using the previous work as reference, this chapter will answer the question of how the network structure affects the epidemic process, especially the community structure's effects. In more detail, this question is divided into five sub-questions below. Moreover, to define the objects more clearly, they are also illustrated in fig 4.1:

1. How does the average degree (the most basic property of the network) affect the epidemic growth rate, epidemic duration time and epidemic final size?
2. How does the community strength (the new popular topic appears recently) affect the epidemic growth rate, the duration time and its final size?
3. How does the overlap rate in the network with overlapped community (new area which focuses on deep community structure) affect the epidemic growth rate, the duration time of the epidemic and the final size of the epidemic?
4. The effects of the population flow rate (refer to the dynamic network or network evolution) to the epidemic growth rate, the duration time and the final size of the epidemic.
5. Different community structure's effect (different community structure with the same generation parameter) to epidemic growth rate, epidemic duration time and its final size.

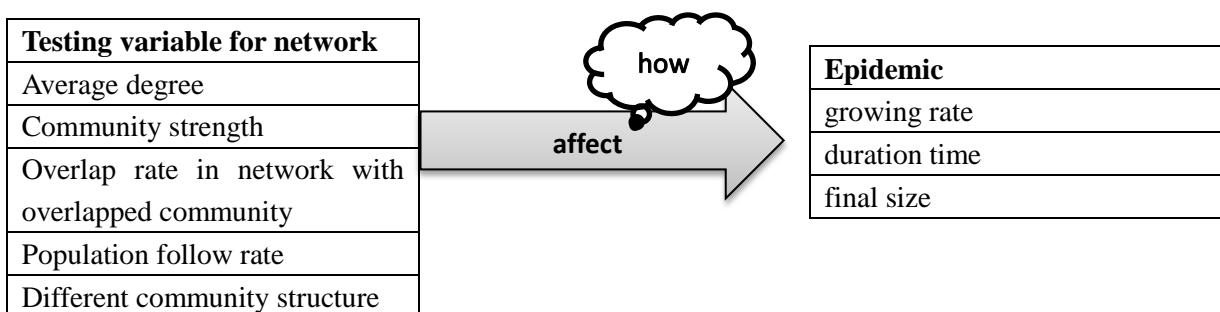


Fig 4.1 schematic diagram of experiment objective

4.2 Experiment design summary

To solve these questions, as shown in table 4.1, the whole process of the experiment is presented below. The table should read from right to left as an equation. Firstly, we should fix the disease parameter: infection rate and disease durations^[1], which are used in the epidemic dynamic model.

[1] disease durations: the value depends on disease property and varies a little in different individuals. Mathematically, durations $\tau = 1/\gamma$, where γ is the recovery rate.

Then, two main groups of experiments are conducted. The first one focuses on the effects of the properties of the network, including average degree, community strength, community overlap rate and probability of individuals' movement, which aims to solve the first four questions. Another's aim is to find how the different community structure of a network can affect the epidemic process. Finally, the aim is to calculate the value of growth rate, durations and final size of the epidemic. In addition, in all of these networks, both the network with and without weighted edges were tested.

Table 4.1 experiments structure

Output	Network Data (variable data)								Disease Data (fixed)
<ul style="list-style-type: none"> • epidemic growing rate • epidemic duration time • epidemic final size 	LFR Algorithms to construct network				Network structure				
	Average Degree	Community strength	Overlap rate	Probability of move (dynamic)					
	Value: 5-35 	Value: 0-1 	Value: 0-1 	Value: 0-1 	Random network	overlap-community	Bridge-community	Hierarchical structure	Every group of data comes from a specific disease

4.3 Infection disease data choose

After analysis revealing information about some infectious disease, a new parameter D is used to classify them. Where

$$D = \text{infection rate} * \text{duration time} . \quad (4.1)$$

When D tends to 0, it means an individual will recover more easily, and vice versa. In our experiments, we choose two typical values of D for express real disease. One is D equals 0.9 that close to 1, where infection rate equals 0.3 and duration days equals 3 (influenza, ref [80]), and another is D equals 0.1 that tend to 0, where infection rate equals 0.01 and the duration in days equals 10.

4.4 Network generation description

4.4.1 Experiment one

In the first group of experiments, four independent experiments are included, which will check the effects of the average degree, community strength, and population flow rate and overlap rate, respectively. For each one, one parameter as variable is considered, which is called the testing value. All the other valuables were fixed. Then, during each experiment, we gave each fixed value six values and arranged them into six groups. Additionally, all the networks constructed in this part are using the method described by Fortunato in [44].

For every generated network, we have a series of statistics value used to evaluate it. To be exact,

their modularity, using the algorithm proposed by Blondel etc in [9], average Clustering Coefficient, the algorithm mentioned in [48], and average path described in [12] are calculated.

4.4.1.1 Average Degree as testing value

In these experiments, we generated six groups of networks with different fixed corresponding parameters list below, in table 4.2:

Table 4.2 input parameter (1)

Group number	Mixing parameter	Population flowing rate	Population	Min community	Max community
1	0	0	500/1000	10/20	50/100
2	0.1	0	500/1000	10/20	50/100
3	0.2	0	500/1000	10/20	50/100
4	0.3	0	500/1000	10/20	50/100
5	0.6	0	500/1000	10/20	50/100
6	0.9	0	500/1000	10/20	50/100

As shown in the table, all the networks here have 500 or 1,000 individuals with different community strength range from 0 to 0.9. Moreover, one thing should be emphasized that all the networks generated here are without overlapped community structure, which is another testing variable, and will be tested in other experiments. In terms of the testing value: average degree will be examined from 5 to 30 for each testing group. In total, 30 networks are constructed here. The fig 4.2- fig 4.6 illustrates the examples of the networks and their main statistic properties:

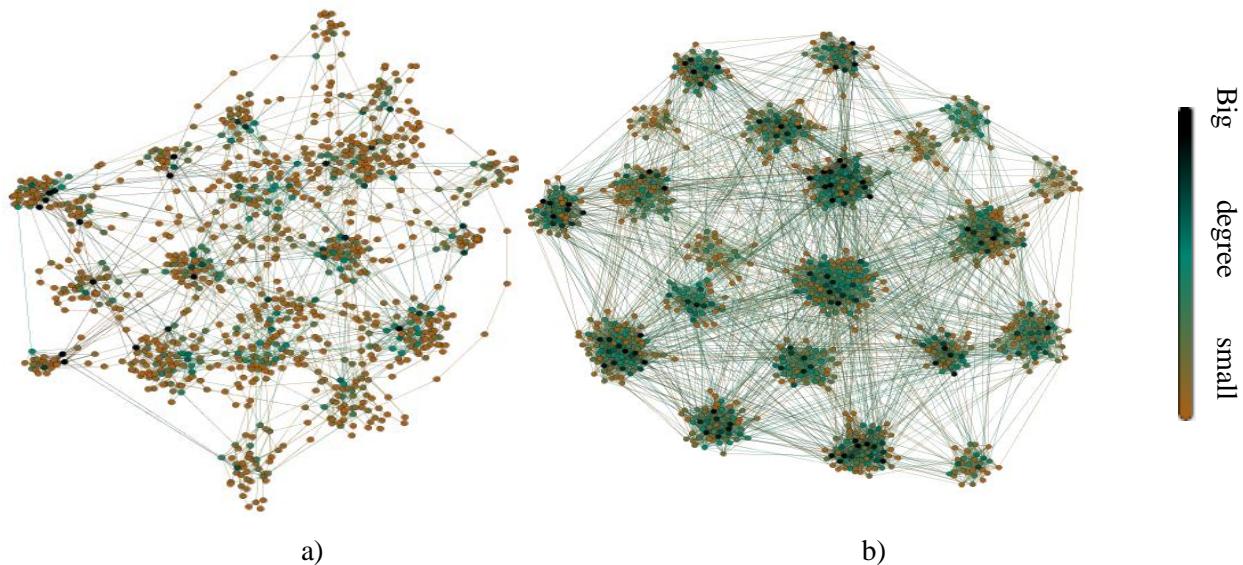


fig 4.2 a) network has 1000 nodes with average degree equals 5,community strength equals 0.1, the color bar shows the degree for each node, **b)** network whose average degree equals 20, and all the other parameter same as network a).

As shown in fig 4.2 a), it has an average degree 5 with a power-law degree distribution. And the colour bar represents the degree of each node. The network has strong community structure with the mixing parameter setting as 0.1. In addition, its average path length is 5.028, which is longer than the network in 4.2 b) whose average path is 3.098 with average degree 20. Their detail statistical information is list in table 4.3 and fig 4.3-4.4:

Table 4.3: statistics for generated network (1)

	Network in 4.2 a)	Network in 4.2 b)
modularity	0.74	0.841
Avg. Clustering Coefficient	0.197	0.379
Avg. path length	5.028	3.098

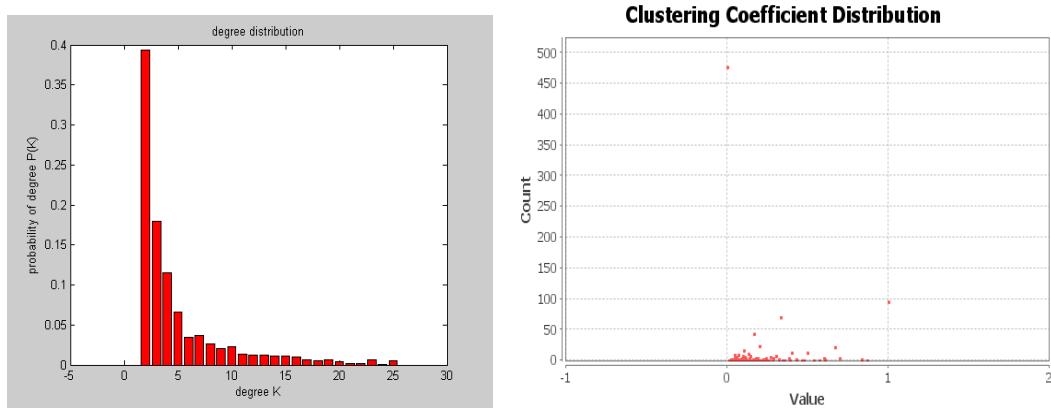
**a)****b)**

fig 4.3 the degree distribution (left) and clustering coefficient distribution (right) for network in fig 4.2 (a)

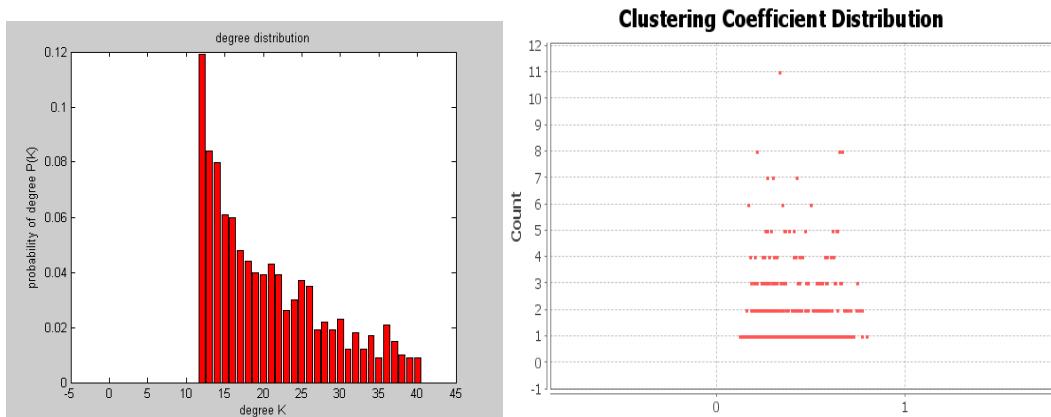
**a)****b)**

fig 4.4 the degree distribution (left) and clustering coefficient distribution (right) for network in fig 4.2 (b)

By contrast, the networks constructed in fig 4.5 have weak community structure, where the mixing parameter is 0.9 for both of the two. The network left has average degree 5 corresponding to network in fig4.2 a). Meanwhile, the network on the right has average degree 20, which has more density relationship between nodes.

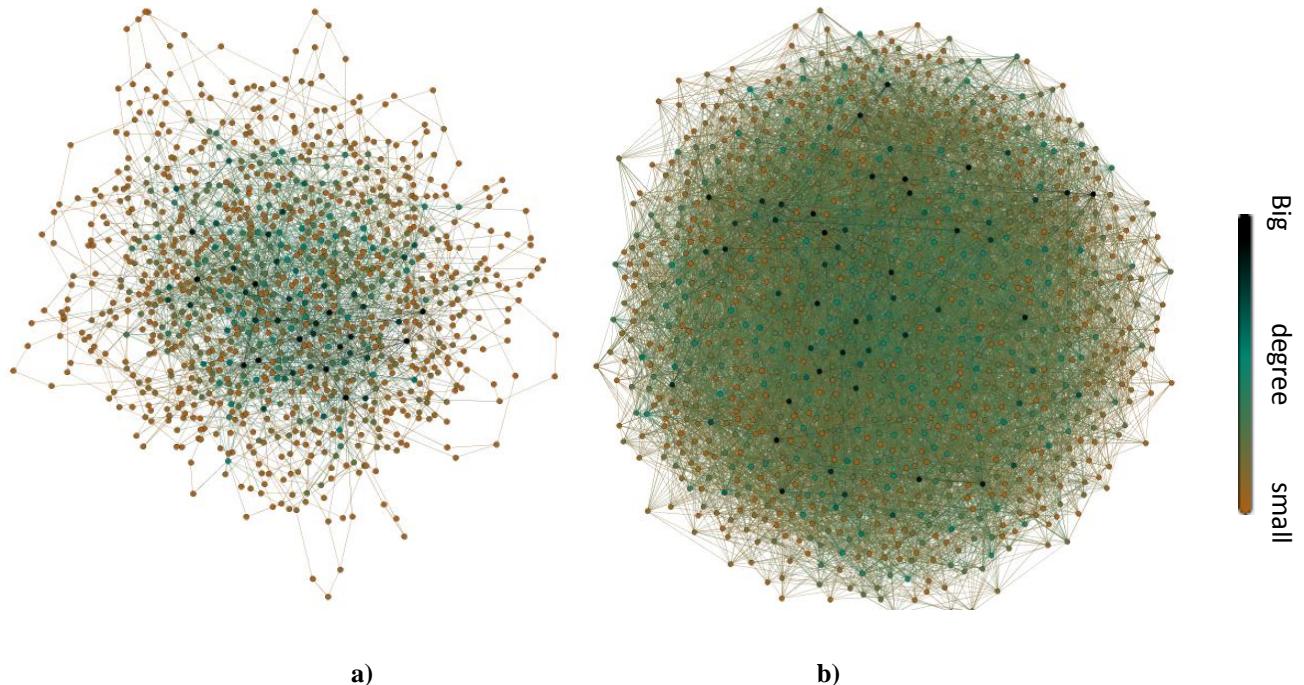


fig 4.5 a) network has 1000 nodes with average degree equals 5,community strength equals 0.9, the color bar shows the degree for each node, **b)** network whose average degree equals 20, and all the other parameter same as network a).

In addition, these two networks' detailed statistical information is listed in table 4.4 and fig 4.6- fig 4.7. As we can see that the value of average clustering coefficient is much smaller than the network generated in fig 4.2, which has a very strong community strength. Hence, we can speculate that the community strength and clustering coefficient have positive correlations. In other words, if it is difficult to calculate the modularity of a network, which is often used to evaluate the strength of community structure, the clustering coefficient can be an auxiliary parameter to estimate community structure.

Table 4.4: statistics for generated network (2)

	Network in 4.5 a)	Network in 4.5 b)
modularity	0.313	0.194
Avg. Clustering Coefficient	0.017	0.024
Avg. path length	4.128	2.62

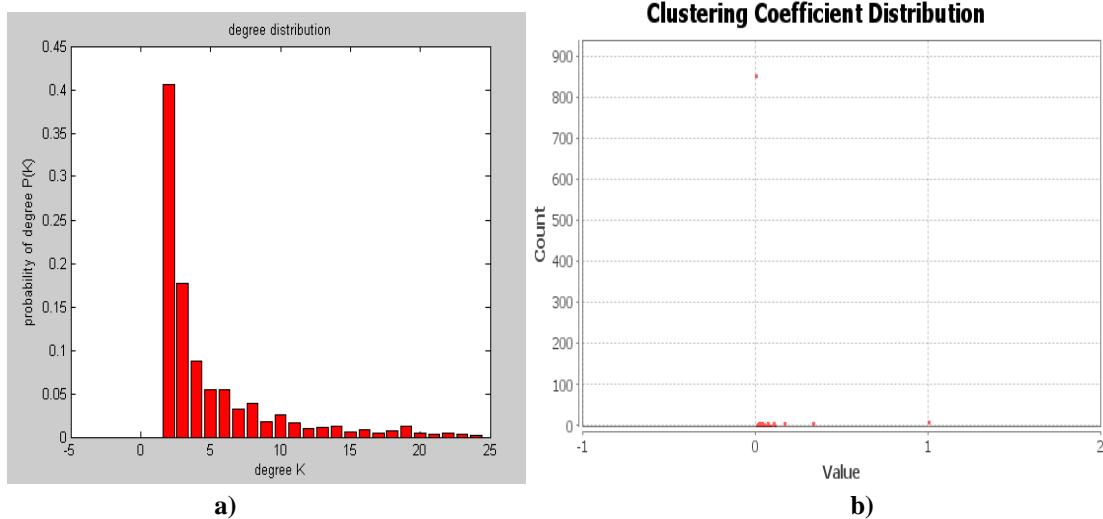


fig 4.6 the degree distribution (left) and clustering coefficient distribution (right) for network in fig 4.5 (a)

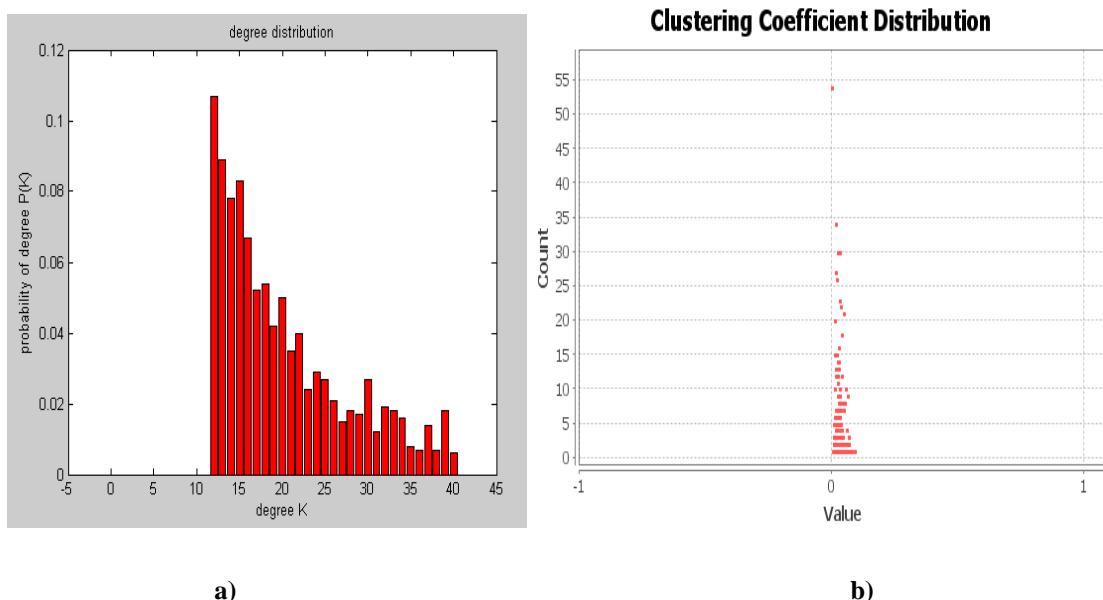


fig 4.7 the degree distribution (left) and clustering coefficient distribution (right) for network in fig 4.5 (b)

4.4.1.2 Community strength as testing value

Same as section 4.4.1.1, we also generated six groups of networks with different fixed corresponding parameters in this section. But the community strength changed as the testing variable this time, the detail input parameter shown in table 4.5

Table 4.5 input parameter (2)

Group number	Average degree	Population flowing rate	Population	Min community	Max community
1	5	0	500/1000	10/20	50/100
2	10	0	500/1000	10/20	50/100
3	10	0.1	500/1000	10/20	50/100
4	20	0	500/1000	10/20	50/100
5	30	0	500/1000	10/20	50/100
6	30	0.1	500/1000	10/20	50/100

To find out the correlation between community strength and some of epidemic parameters is not a very clear task. This is because until now there is no unified parameter to evaluate the community strength. In standard practice, scientist use modularity as the standard method when quantifying community structure in network. But for artificial network, like LFR benchmark, they use a value called mixing parameter to express the relations. First, we should give the exact definition of modularity, which is more widely used:

Modularity is a benefit function used to decide a division of a network into modules or communities. Good divisions will get high scores and vice versa. The modularity is a similar conception of community that has internal connections of density and sparse connections between different modules [76].

Next, we find out that mixing parameter has a power-law relation to the modularity and another auxiliary parameter clustering coefficient, which strengthens the viewpoint that when mixing value tends to 0, strong community or modularity exists in the network, whilst when the value tends to 1, the network will change to a random structure. The data of networks with average degree of 5 provided in table 4.6 and fig.4.8 give evidence to this argument.

Table 4.6 community strength statistic of network

Mixing parameter	0	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
modularity	0.805	0.694	0.572	0.466	0.358	0.344	0.368	0.321	0.342	0.313
Avg. Clustering Coefficient	0.257	0.197	0.119	0.073	0.036	0.023	0.016	0.015	0.013	0.017
Avg. path length	3.645	5.028	4.635	4.394	4.281	4.126	4.368	4.159	4.183	4.128

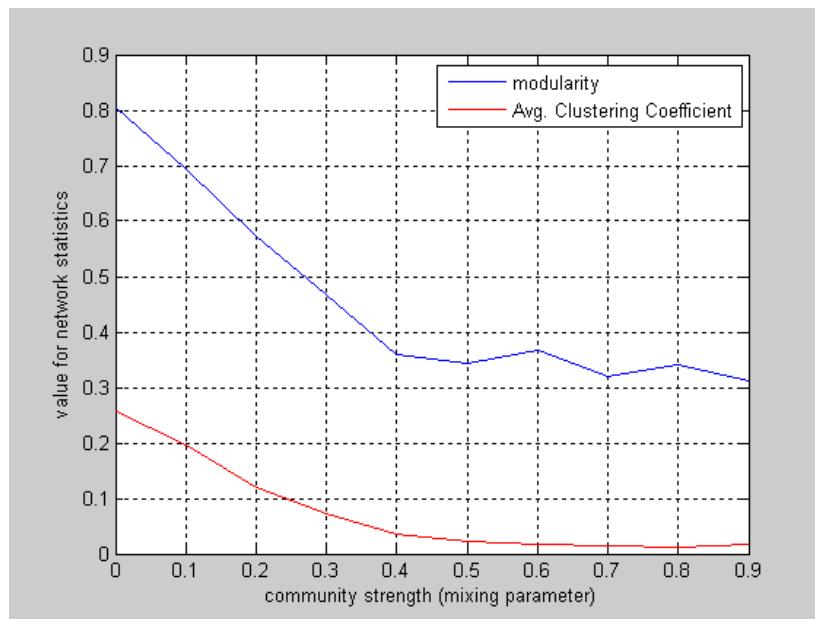


Fig.4.8 the relations between mixing parameter in LFR algorithm and modularity

In fig 4.9 a), an example of the network with average degree of 10 and strong community strength.is drawn. Different communities are labeled by different colours. As shown in the picture, 19 communities consist of the whole network. In addition, the corresponding modularity, average clustering coefficient and average path length are 0.837, 0.234 and 3.824, respectively. By contrast, fig 4.9 b) shows a network with less strength of community structure, with modularity of 0.737, average clustering coefficient of 0.165 and average path length of 3.469.

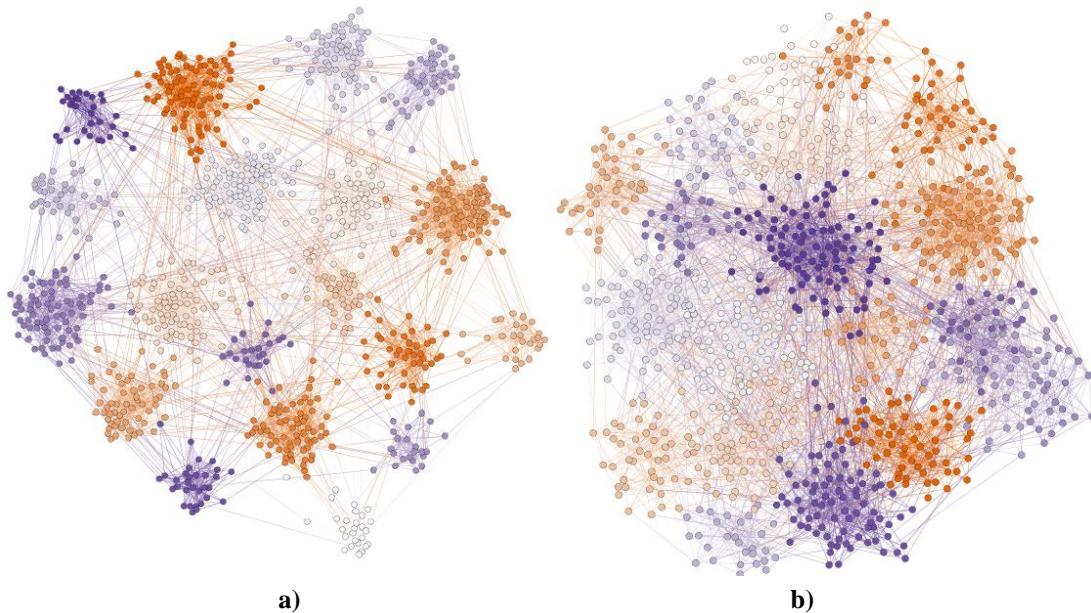


Fig 4.9 network with clear community structure denote by different colors

4.4.1.3 Overlap rate as testing value

In this part, we generated network with overlapped communities using the parameter ‘-on’ and ‘-om’ in LFR algorithm to control the ratio of overlap. To be simple, all the overlapped nodes we assumed that they belong to at most two communities, where we set ‘-om’ always equals 2. Furthermore, the value of ‘-on’ which denote the overlapped people equals the total population multiply the overlap ratio. We tested the overlap rate from 0 to 1 to observe effects. As in previous experiments, all the other fixed parameters are classified into six groups listed in table 4.7:

Table 4.7 input parameter (3)

Group number	Average degree	Community strength	Population flowing rate	Population	Min community	Max community
1	5	0.1	0	500/1000	10/20	50/100
2	10	0.3	0	500/1000	10/20	50/100
3	15	0.9	0	500/1000	10/20	50/100
4	20	0.1	0	500/1000	10/20	50/100
5	25	0.3	0	500/1000	10/20	50/100
6	30	0.9	0	500/1000	10/20	50/100

Fig 4.10 gives three examples generated in this part of all the network, which here have strong community structure with mixing parameter of 0.1. From left to right, their overlap rates are 0.1, 0.2 and 0.4 respectively. Among the experiments, the group 3 and 6 almost useless. This is because these two networks have very weak community structure. In addition, their statistical information list in table 4.8

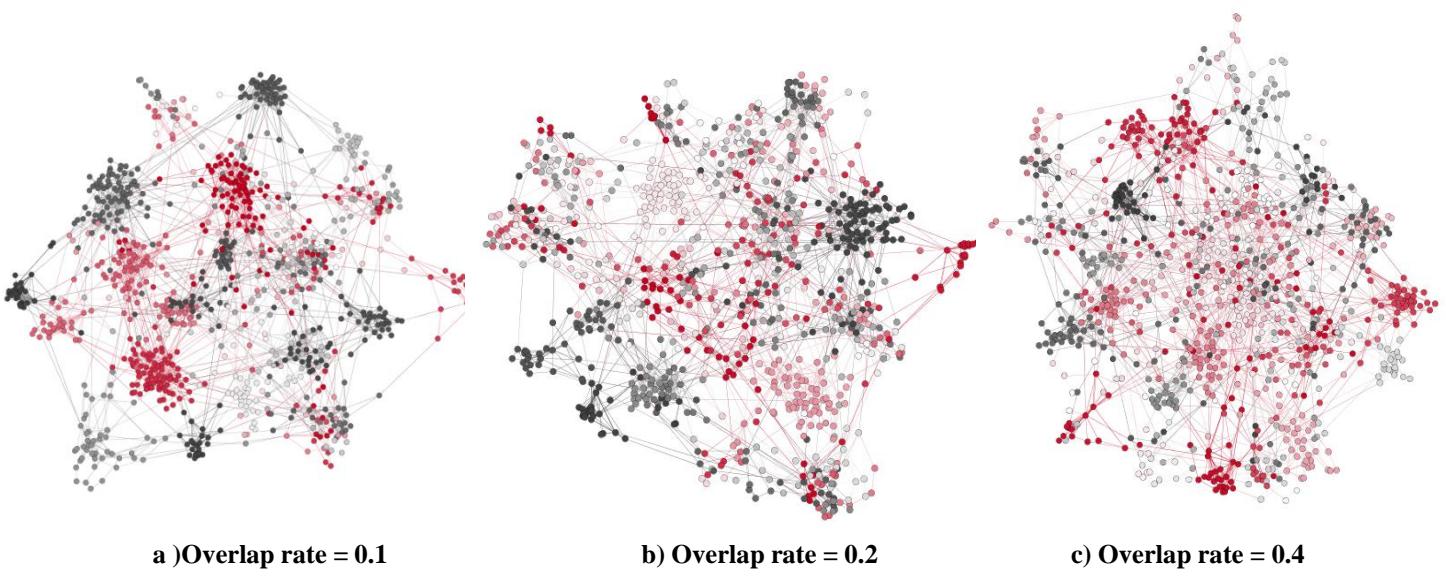


Fig 4.10 network with 1000 nodes and strong community structure. From left to right, the communities overlap rate increase from 0.1 to 0.4.

Table 4.8 statistics for generated network (3)

	Network in 4.10 a)	Network in 4.10 b)	Network in 4.10 c)
modularity	0.313	0.194	0.087
Avg. Clustering Coefficient	0.017	0.024	0.019
Avg. path length	4.128	2.62	1.98

4.4.1.4 Probability of movements for individuals

In this sub-group of experiments, dynamic network are generated. As same as previous experiments, we divide fixed variables into six groups, where the data list in table 4.9. To generate dynamic network, the first step is to build a static network based on given parameters. Second, for every evolution step, every individual have specific probability to move to another communities. In the experiments, we use two methods to simulate the network evolution process. The first is called random movement evolution and another is based on the common population flow phenomena during epidemic outbreak which includes community merging, contraction, splitting, birth and death.

Table 4.9 input parameter (4)

Group number	Average degree	Community strength	Community overlap rate	Population	Min community	Max community
1	5	0.1	0.1	500/1000	10/20	50/100
2	10	0.3	0.1	500/1000	10/20	50/100
3	15	0.9	0.1	500/1000	10/20	50/100
4	20	0.1	0.1	500/1000	10/20	50/100
5	25	0.3	0.1	500/1000	10/20	50/100
6	30	0.9	0.1	500/1000	10/20	50/100

- **Random movement evolution**

To achieve network dynamic, we consider the whole evolution process as a Markov process^[1]. First of all, we give out some assumptions the experiments based on here:

1. We only consider the movements between communities. In other words, the movements inside community will be thought has little effects to the disease spreading.
2. In terms of individual, we assume that for each time step, they have same probability to move to every other community. Mathematically, we use $P(A_{n+1}^i | B_n^i)$ denotes the probability of node i belong to community A at time step n+1 under condition that it belong to community

[1]A stochastic process with the Markov property is one for which conditional on the present state of the system, its future and past are independent.^[53]

B at time step n . And for $\forall \{n | 0 < n < \infty\}$ and $i, P(\omega_{j_1}^i | \omega_{j_2}^i) = p$, where ω represent the community.

Next, for every time step and individual, he or she has probability p to move to another random chosen community X and with probability δ connecting to each member in X . Furthermore, the old relations with previous community members will be cut with probability σ .

- **Directional movement**

Different from random movement, the directional movement use different moving probability p for different individuals. In this situation, we assume that the probability p for people in epidemic outbreak area [1] higher than the normal area without epidemic. In addition, if one overlapped community outbreak with an epidemic, the individuals in the overlapped area tend to move from the community without epidemic. To implement the community splitting, during every evolution step, for each individual who have two memberships A and B, and the place A have an epidemic, they delete the membership A with probability p , which simulate the disappear of overlapped communities. Fig 4.11 illustrates this process. Moreover, if both community A and B have an epidemic, the people tend to move to other communities that same as process in random movement.

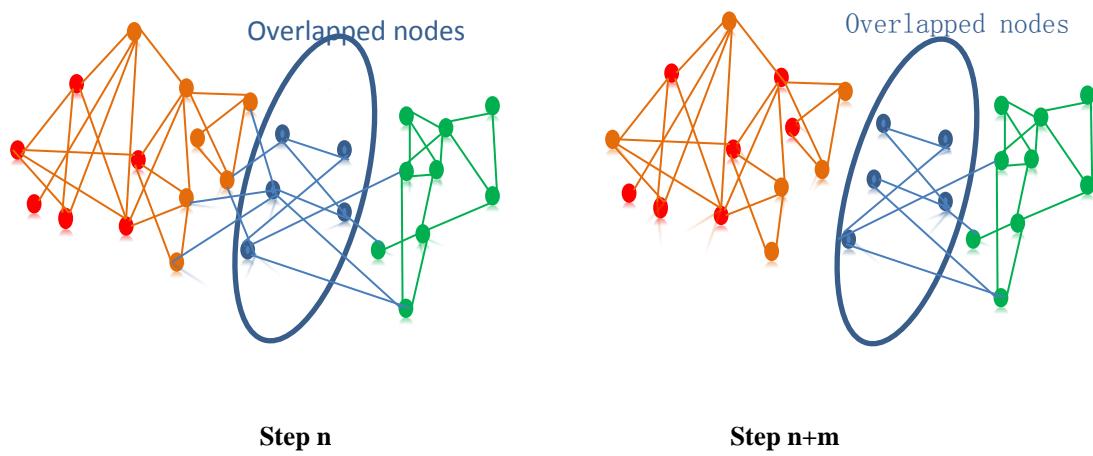


Fig 4.11 The evolution of network connection when the epidemic outbreak in the orange area, where red nodes donate infection individuals.

[1]When the proportion of infection people to the whole population of the specific community, we consider this area outbreak an epidemic.

4.4.2 Experiment two

Until now, the experiments which have been done can answer the question from 1 to 4 in section 4.1. To answer question 5, we will use another way for network generation. In this part, every testing group of networks have same main parameters, but with different network structure. These experiments here are focus on the effects of several different community structures: bridge community, overlapped community and hierarchical community.

In this part, all the networks are established based on small-world network, which was widely proved in real social network. To analyze community structure's impact, basically, we need propose a comparison network without community structure. Here, the most basic small-world network will be built up firstly. Next, the bridge communities, overlapped communities and hierarchical communities will be set up in turn.

4.4.2.1 Parameters

To implement the experiments, the main parameters for every group of networks should be same. Simply, the total population, average degree and rewiring rate were set as same number for four networks in one testing group. In total, we designed five groups of experiments with parameters list in table 4.10:

Table 4.10 input parameter (4)

Group number	Average degree	rewiring rate	Population
1	6	0.1	1000
2	10	0.1	1000
3	20	0.1	1000
4	10	0.3	1000
5	20	0.3	1000

4.4.2.2 Small-world network generation

To generate the small-world network, the method was described in section 2.1.5.2. In fig 4.13, we gives an example of small-work network with 1000 nodes, 6 initial neighborhoods and rewiring rate of 0.1 and the corresponding degree distribution. In addition, same as experiments in Experiment one, all the statistical information for every constructed network were recorded. In this case, the network in fig 4.13 has average path length of 6.207, average degree of 6 and average clustering coefficient of 0.44.

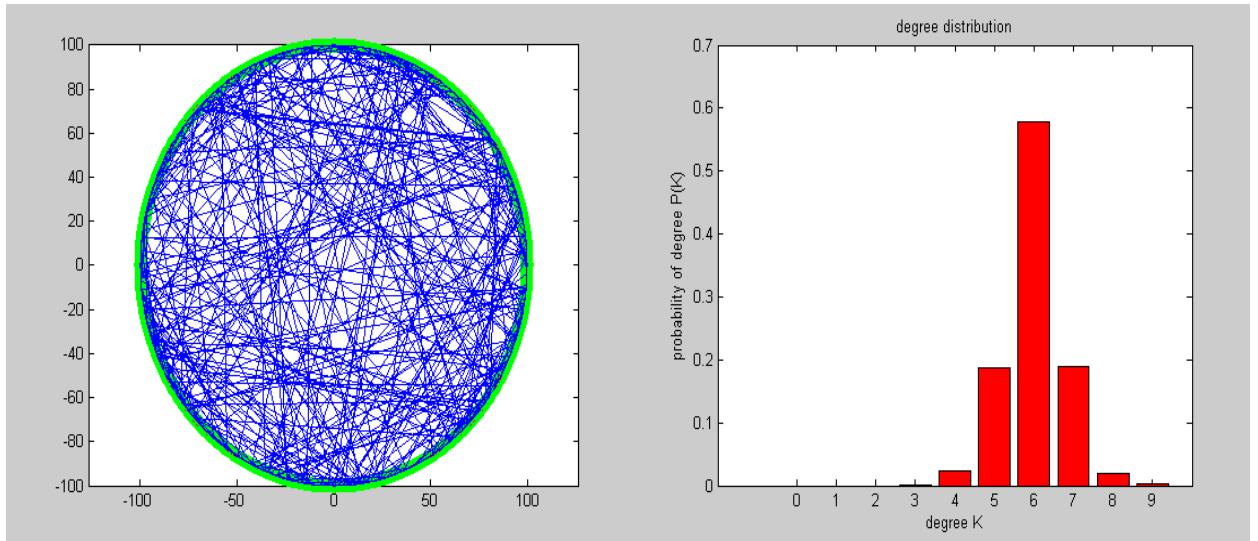


fig 4.13 small-work network with 1000 nodes and its corresponding degree distribution

4.4.2.3 Bridge disjoint community structured network

Bridge disjoint community structure was described in paper [8], which was considered as one of the most common social structure for disease transmission analysis. To construct the network, the following steps include:

1. Equally allocate individuals into m communities. For every community, using parameter k (nearest k neighbors) and rewiring rate p to construct m small-world networks.
2. For each node, it has probability p' to link to every node from other communities. In order to make the average degree unchanged, for building one new edge, we should cut an internal edge. To be more detail, firstly, from step one, we can calculate the internal average degree k^{in} for every community and the probability of the connection between every two nodes k/N , where N stands for the total population in this community. Then, we set $p' = \varphi * (k/N)$, φ representing the ratio of links presence between nodes from different communities and nodes from the same one. In our experiments, the φ always was set as 0.2, which make the entire network has strong community structure.

Fig 4.14 gives an example of bridge disjoint network with 1000 nodes and 10 communities. Every community has small-world property with k setting to 6 and p equals 0.1 initially. Furthermore, its degree distribution and clustering coefficient distribution list in fig 4.15. Other statistic like average path length and average clustering coefficient is 5.601 and 0.391, respectively.

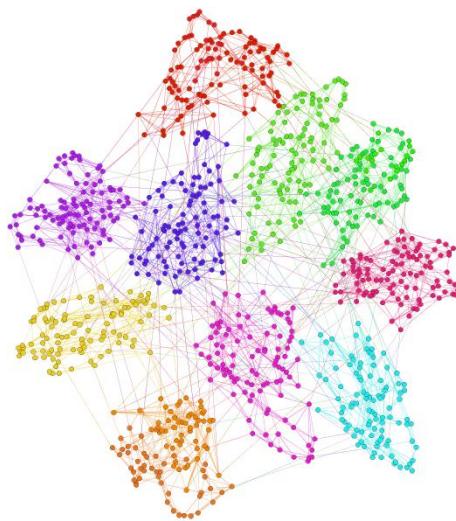


Fig 4.14 bridge disjoint network with highly community structure

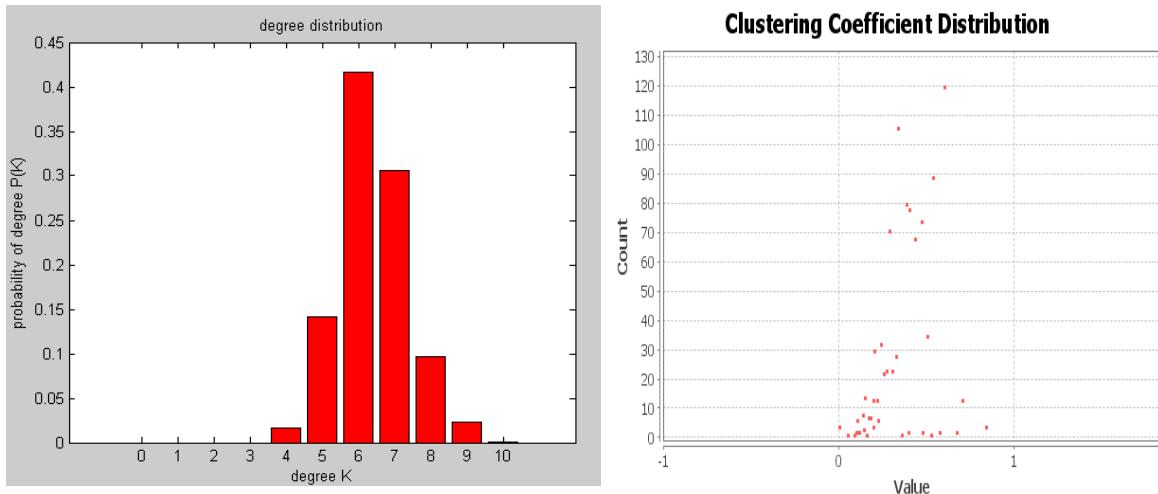


Fig 4.15 statistic information for bridge disjoint network in fig 4.14

4.4.2.4 Overlapped community structured network

Different from the method in experiment one, the overlapped communities construct here use a more simple way and consist of small-world networks. To be simple, we only build two communities with overlapped part. The following three steps describe the generation process:

1. Equally allocate individuals into 2 communities, donated as O_1 and O_2 . Then, randomly choose x percent individuals from each community to make up a new small group or community C. In our experiments, for a network with 1000 nodes, x was set as 5.
2. For each community O_1 , O_2 and C, we constructed as small-world network with parameters $k_1 = k_2 > k_c$, where k_1, k_2, k_c represent k nearest neighbors for O_1 , O_2 and C, respectively.
3. Build up connections between overlapped part C and original communities O_1 , O_2 . For every

pair of nodes n_i and n_c , one from original group and one from part C, they have $p = \frac{k_1}{\frac{N}{2}(1-0.05)}$

probability to make up a link. To control the effects of average degree's change minimal, for every new edges born, an old edge of node n_i will be deleted.

In fig 4.16, an overlapped community structured network was presented. It has 1000 individuals with 100 nodes in overlapped part. This kind of model can represent the situation of social activities at urban fringe. For example, the red part represents people in a city, while blue part represents rural populations. And the overlapped area can illustrate some trade or business between them. Its statistics are provided in fig 4.17. Additionally, its average path length is 5.501 and average clustering coefficient is 0.363

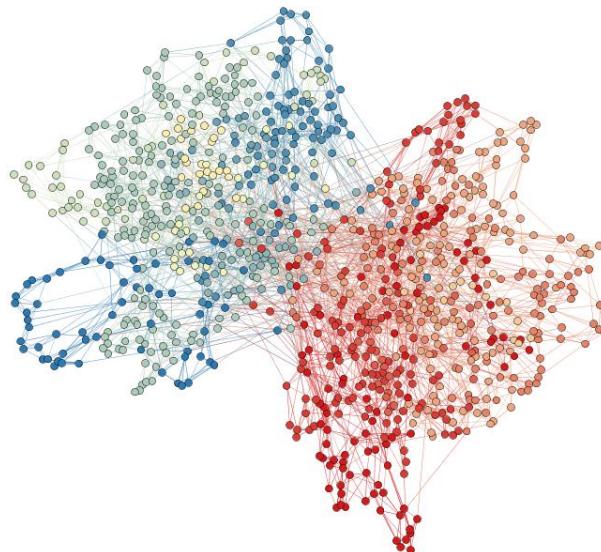


fig 4.16 overlapped community structured network with 1000 nodes and 100 nodes belong to overlapped part

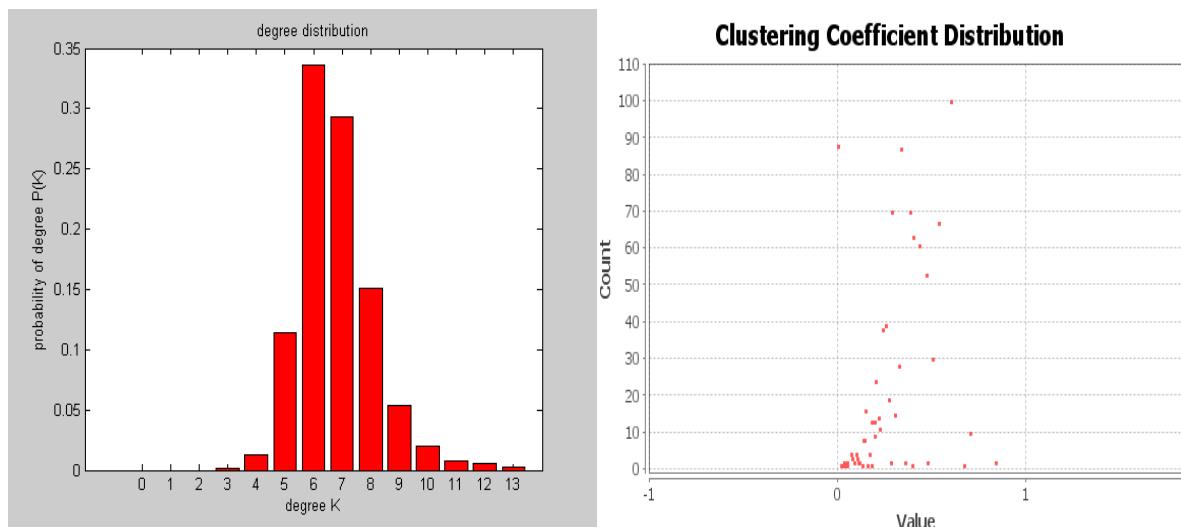


Fig 4.17 statistic information for overlapped network in fig 4.16

4.4.2.5 Hierarchical community structured network

Finally, hierarchical community structured network will be construct. It can be considered as a specific instance of overlapped network, but with more requirements of the structure shape. In paper [20], Clauset, Moore and Newman proposed an algorithm for hierarchical community network generation. And the source code can be downloaded from <http://tuvalu.santafe.edu/~aarong/hierarchy/>.

In our experiments, we use another way for network generation. A network has 3 hierarchies include 10 basic communities will be as an example constructed here. The detail process summarized below:

1. Assign all the individuals into 10 communities and calculate the size of every community.
2. Group the 10 communities into second level community. In addition, we assume that every second level community has at least two basic communities. Then, we should calculate and record the size for each second level community.
3. After give all the individuals their two memberships, basic membership and second level membership, we begin to construct the network. Firstly, small-world network is constructed within each basic community. Secondly, for every second level community, we use probability $p = \varphi * (\frac{k}{\text{basic community size}})$ to build a link between every pair of nodes come from different basic communities. Where k is the nearest k neighbors during basic network generation and φ is the ratio of connection probability between every pair of vertices within basic community and within second level community.
4. Finally, to connect every second level community, we give probability $p' = \varphi' * p$ to build up edges between them.

Fig 4.18 shows the hierarchical network has 10 basic communities and 3 second level communities. It is generated by setting k as 6, p equals 0.5 and p' equals 0.5 respectively. In terms of statistic information, its average path length is 5.314 and average clustering coefficient is 0.341

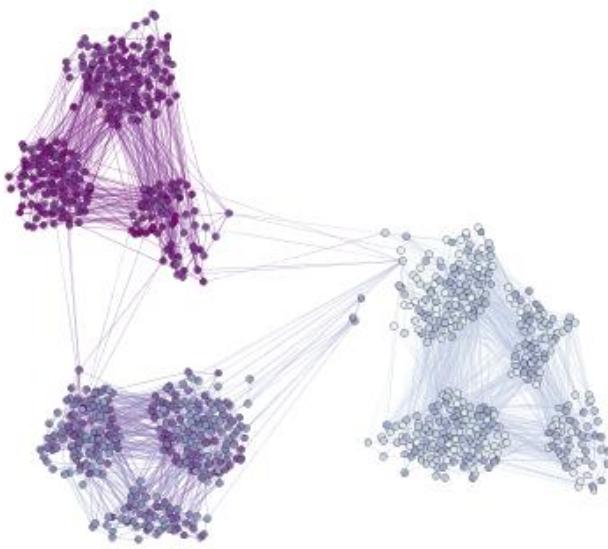


fig 4.18 hierarchical network with 1000 nodes which belong to 10 basic communities and 3 second level communities

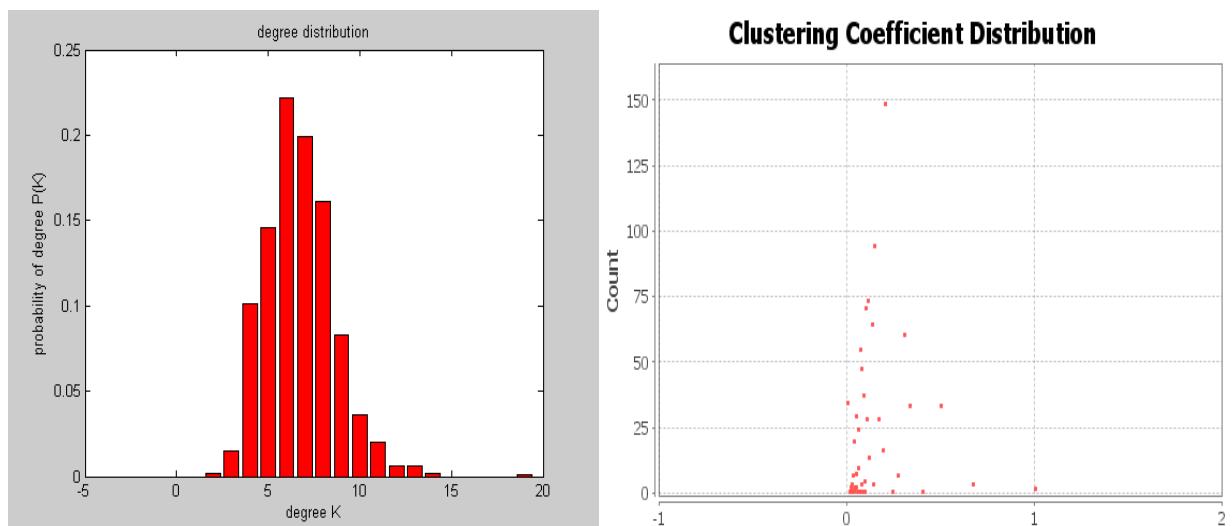


Fig 4.19 statistic information for hierarchical network in fig 4.18

4.5 Data structure for network

After arriving at a clear position on the experiments design, we need to use program language to implement these. The algorithm for network generation was introduced in the beginning of 4.4. In this subsection, the main data structure used will be described.

- ***Adjacency matrix***

Simply, we use adjacency matrix to describe and store the information of network. In mathematics, for a finite graph G has n vertices, we use $n \times n$ matrix with the value of position $V_{ij} = 1$ to

represent there is an edge between vertices i and j , whilst $V_{ij} = 0$ means no links between them.

In the algorithm to generate networks is a process to construct an adjacency matrix and the finally information of the network would be stored by the matrix. Furthermore, the process of network evolution or changes is also expressed by matrix transformation.

- *Adjacency list*

To inquire the information from matrix is time consuming, hence, when simulated the epidemic process, we use adjacency list instead of adjacency matrix. Before this step, all the information will be translate from adjacency matrix to adjacency list.

- *Weights list*

For weighted network, we added one more data structure to save information of the edges' weight. It is an one-dimensional array has m length, where m is the number of total edges in graph G . In the experiments, we generate m random numbers range from 0 to 1 following a power-law distribution. These weights express how close the relationship is between the two connected nodes i and j , which is in order to affect the probability of infection events.

- *State list*

State list is an array storing the information of each node in network for every step. Hence, it will change with every time step to record the state. Moreover, we used state matrix to be an accumulative way for recording the history of the changing process. To be exact for every new time step, the state matrix will add one more line to save the content of state list. In other words, the state matrix is a history document for entire event process; meanwhile state list is a note for the current time.

4.6 Disease transmission

Based on daily contact network, the disease transmission will happen. During the transmission, each individual is in one of two states and one of three states for SIS and SIR model respectively. Finally, all the result data include epidemic growing rate, final size and epidemic duration days, we collected from simulation of SIR model. At each (daily) time step the state transform for each individual is determined stochastically, based on the infection rate of specific disease and relevant recovery rates.

As mentioned before, all the transmission processes were implemented on both un-weighted and weighted network, which will affect the probability of infection for one individual. Firstly, we assumed that the infection rate for each contact one day is β . In other words, it also means that the probability of not be infected after one contact is $1 - \beta$. In terms of un-weighted network, if one individual has k connections with other infection people, the probability of not be infected is $(1 - \beta)^k$. In contrast, it has $1 - (1 - \beta)^k$ possibility to turn to infected state. In weighted network, the infection rate equals βW , where W is a vector of weights $W = \{w_1 w_2 \dots \dots w_k\}$. Then,

the probability of be infected is $1 - \prod_k(1 - \beta w_i)$.

In fig 4.20, one process of epidemic on dynamic network will be shown as an example. The network has 500 individuals in total and 17 communities. In the initial time step, the disease breaks out in two communities. Then, the transmission process runs with infection rate of 0.02 and duration days of 10.

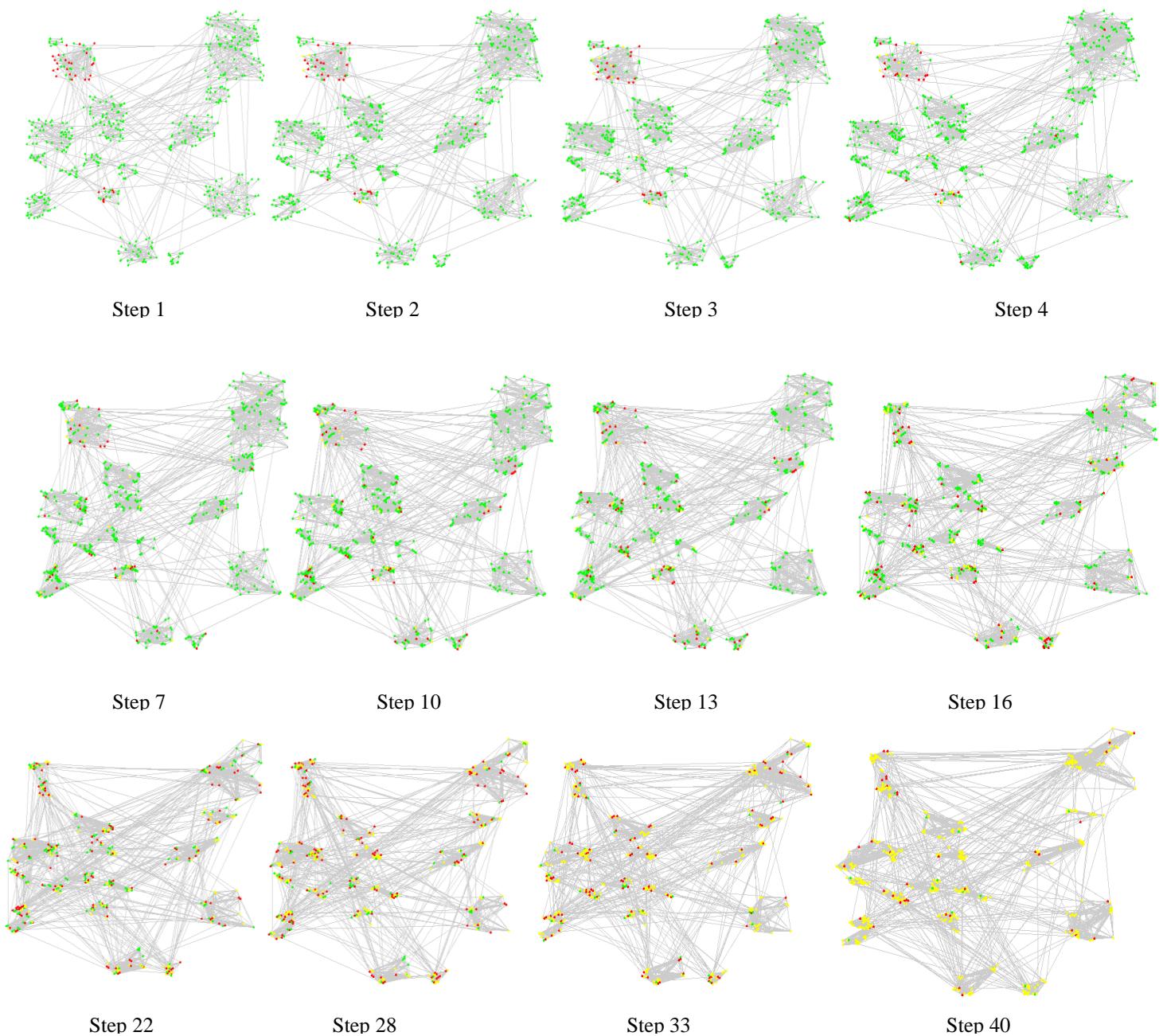


fig 4.20, one process of epidemic spread on dynamic network

5. Results and analysis

This chapter describes details of the results obtained from experiments introduced in previous chapters. To express the questions and answers more clearly, we use the right side values in fig 4.1: epidemic growth rate, epidemic duration time and its final size as the clue for this chapter. Additionally, we will analyze how the five input variables affect them. In order to remove the bias of a one time running result, every result is the average value of 500 implementations. Finally, we should mention that every group of experiments have been done with disease parameter $D=0.1$ and $D=0.9$. They present same trend influence but with different strength. Here we choose the relative evident results to illustrate.

5.1 Epidemic growing rate analysis

In the experiments, we tested five factors to find out whether they can affect the epidemic growth rate and how they implement these effects. We find that (for given degree distribution), the average degree, community strength and population flow rate have clear correlations to the epidemic growth rate. Furthermore, under the same community strength, the overlap rate of communities has positive correlation to the growth rate. By contrast, different community structure with similar setting parameter has a limited impact. Edge weights that measure closeness have impact only when there are correlations between different edges.

5.1.1 Average degree under given degree distribution

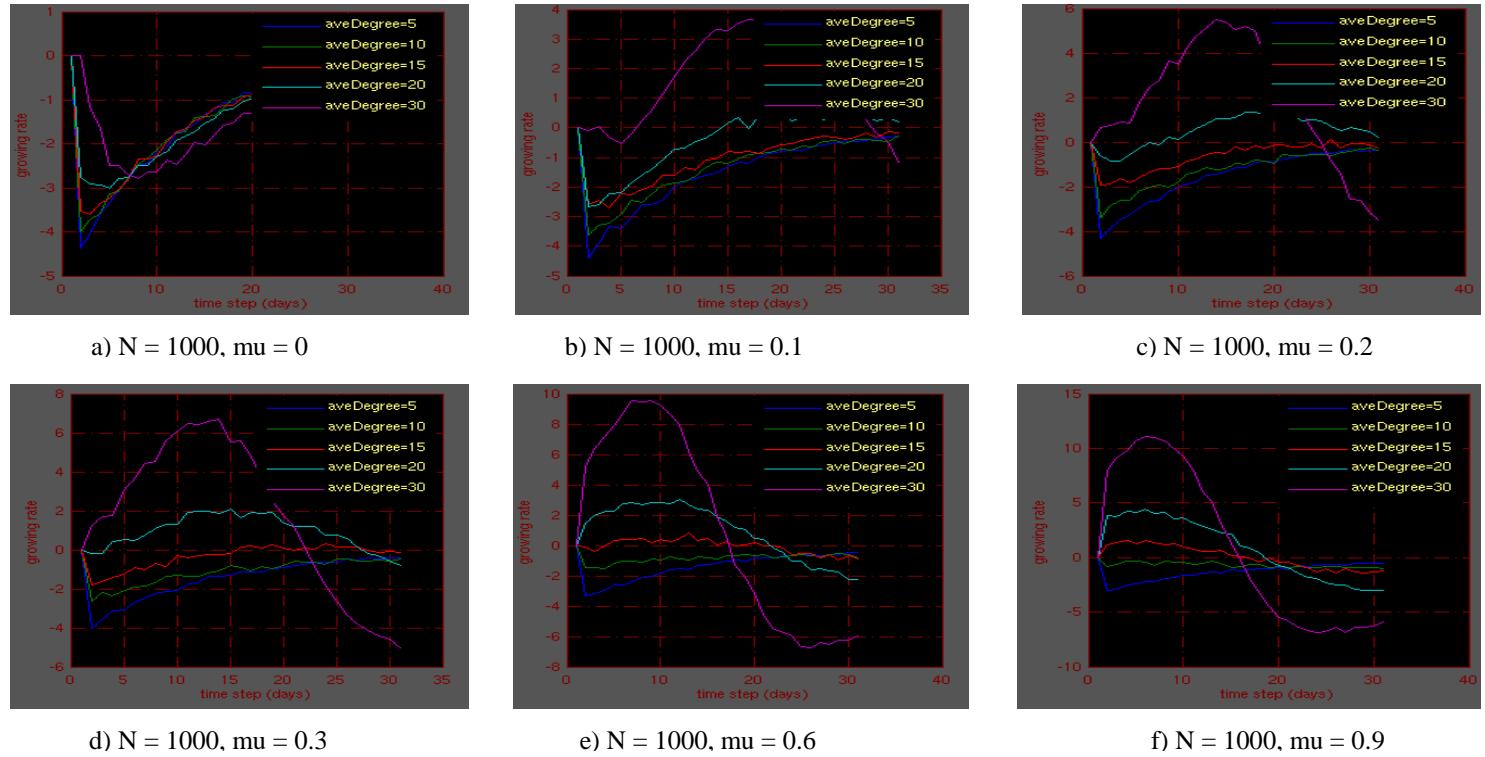


fig 5.1 effects of average degree to growing rate (D=0.1)

Due to the huge number of experiments conducted, it is observed that the average degree of the network can significantly affect the epidemic growth rate. As shown in fig 5.1, under different groups of fixed variables, it can be seen that all of these express the same trend of effects. A bigger average degree corresponds to a higher growth rate and a smaller value refers to a lower growth rate. Take fig 5.1 b) for instance, the bigger average degree which leads to a bigger amplitude, which result in slower convergence speed than smaller one. Meanwhile, these results also illustrate that disease didn't spread through the entire network under smaller average degree. In addition, comparing fig a) to f), it can be seen that the effects of average degree are stronger in a weak community structured network than the effects in network with a strong community structure.

5.1.2 Community strength (measured by mixing rate in LFR algorithm)

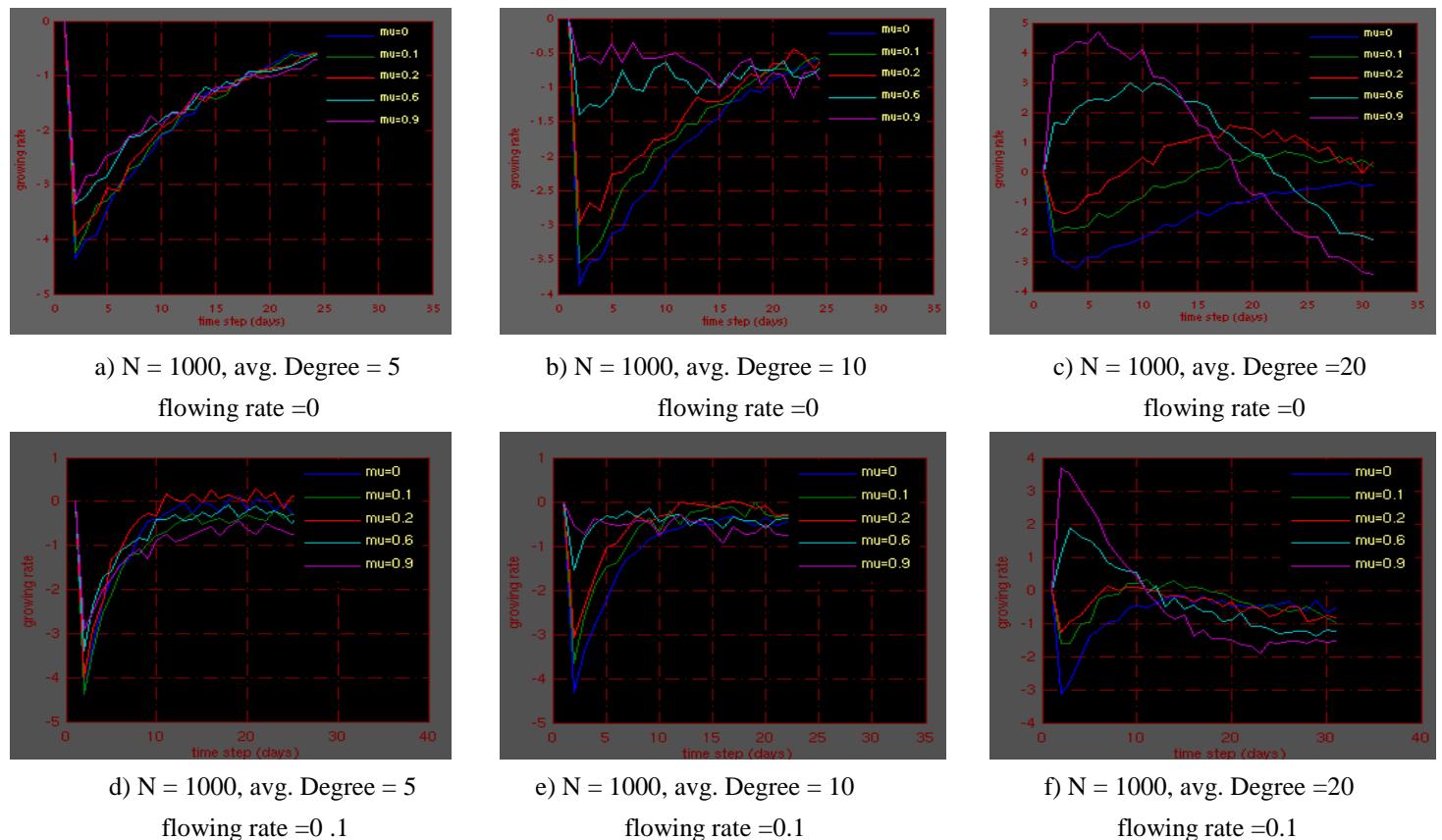


fig 5.2 effects of community strength to growing rate (D=0.1)

Similar to the effects of average degree, the community strength is another main factor to affect epidemic growth rate, but in contrast, it has a negative correlation to growth rate. As it can be seen from fig 5.2, a bigger mixing value results in a higher growth rate. This means the community structure will restrict the disease spreading process. Take fig c) for example, these networks have 1000 population with an average of 20. The purple line increases quickly at the

beginning whilst the green line has a downward trend for the same period. In addition, we can see that the epidemic illustrated by the green and blue line didn't break out. They have been limited within the community and the number of infectious individuals continues to decrease throughout the whole period. By contrast, the epidemic process represented by purple, light blue and red line experienced from outbreak to disappear. Moreover, there is a comparison between the first three pictures. It shows that community strength has little effect to a sparse network as generated in fig a). Meanwhile, comparing the figures in first row with that in second row, we find that the moving events will decrease the effects of community structure.

5.1.3 Overlap rate under given community strength

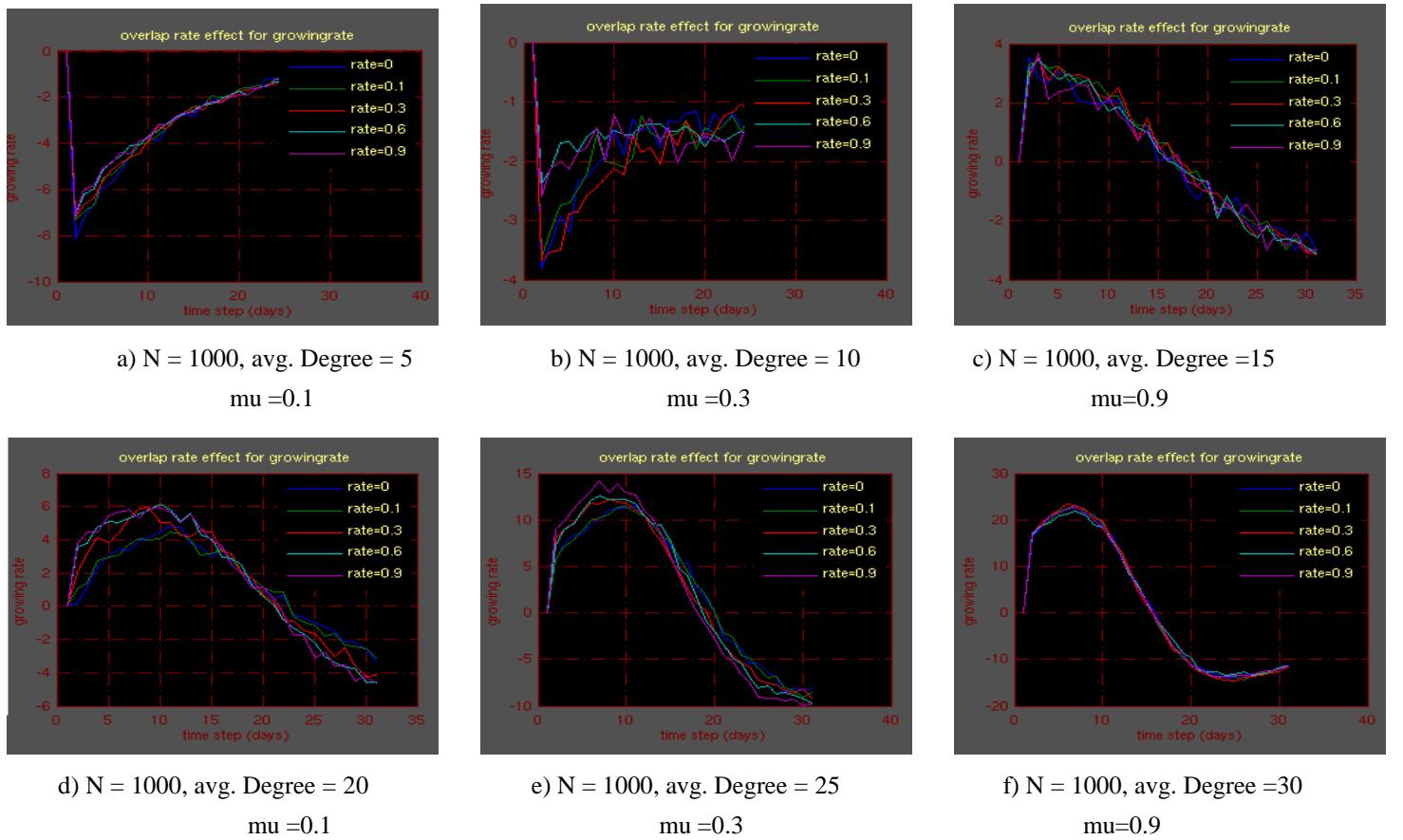


fig 5.3 effects of overlap rate to growing rate (D=0.1)

In terms of overlap rate, we only test the simplest situation that the overlapped nodes at most have two memberships. Observed from fig 5.3, the overlap structure will help disease spread, which is illustrated clearly by fig b), d) and e). But its effect is not as dramatic as the first two factors: average degree and community strength. In fig a), a small average degree network is illustrated, all the lines conveying the same trend. This means under this kind of situation, even though community strength is strong, the overlap strength has no effects on the growth rate. Furthermore,

its effect will also be lost when the community strength becomes weaker, which is presented in c) and f). In addition, compared with picture b) and e), we can see the effect of overlap rate to epidemic growth rate which will depend on average degree. In other words, the variable overlap rate is an assistant parameter which affects the growth rate.

5.1.4 Population flow rate

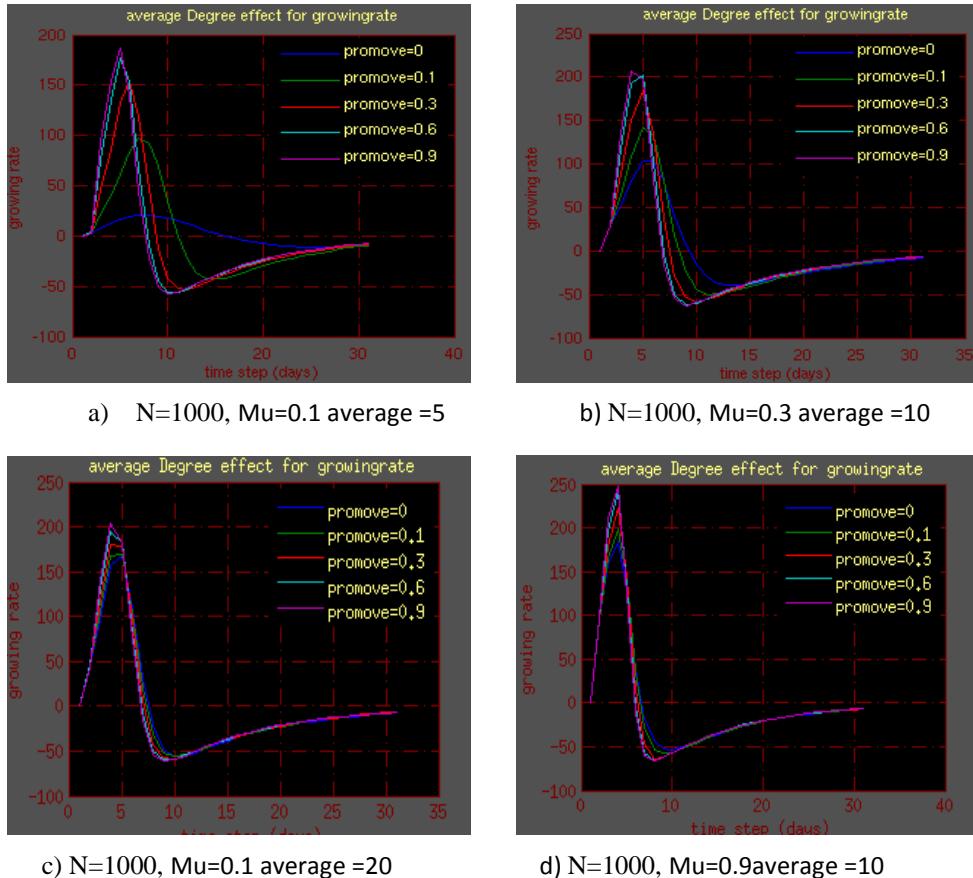


fig 5.4 effects of population flowing rate to growing rate (D=0.9)

We tested probability of movement of individuals from 0 to 0.9 under different network situations: a strong community structured and a relatively weak one. The results are shown in fig 5.4, where we can see that this input parameter has positive correlation to the growth rate obviously in a) and b). Taking picture a) for example, under a strong community structured network with relative low density relations, the moving events can increase the epidemic growing speed more than three times of static network. But along with the average degree increase or the community structure becoming weaker, the effects of population flow rate will almost certainly rapidly decrease.

5.1.5 Different community structure with similar evaluated parameter

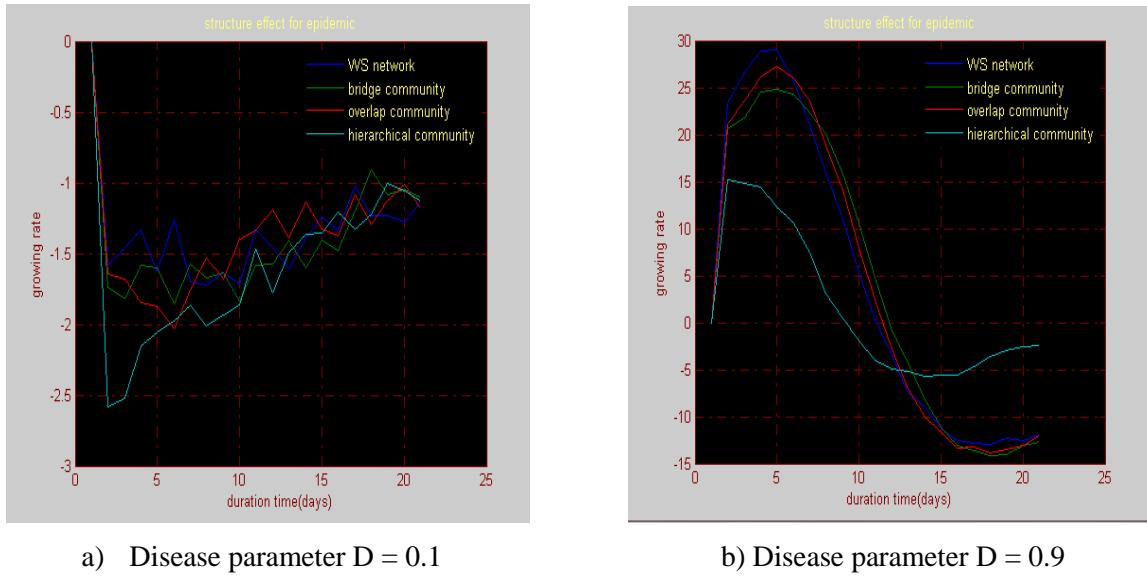


Fig 5.5 effects of different community structure to growing rate

From the results illustrated in fig 5.5, it can be seen that the network that has a stronger community structure which will control the speed of disease spread. In the left picture, the epidemic didn't break out whilst there was an epidemic process experienced on each network in the right picture. As we can see from the right one, one more piece of evidence proved that the overlapped structure positively affects the epidemic growth rate, which is described by the red line which has a more precipitous shape compared to the green line. In addition, the most significant result in this experiment is that the hierarchical structured network will dramatically constrict the disease spread. The big gap between the light blue line and other colored lines illustrates the phenomenon which reveals that the disease will be trapped in the dense part of the network, which decreases the opportunity for the disease to escape out.

5.2 Epidemic duration days and final size analysis

Next, we are going to analyse the value of the epidemic duration time and its final size. In terms of the duration, it is possible to calculate this by considering the day from epidemic break out to the day of no infected individuals existing in the monitored network. Same as epidemic growth rate, it is a result of a combination of all of the input parameters, especially, the population flow rate which can significantly change other parameters' effect in the final result.

Meanwhile, the result of the epidemic final size has been determined, which is another indicator for epidemic analysis. In the following description, the results of experiments will provide a focus on different network properties and give analyses to the results.

5.2.1 Average degree under given degree distribution

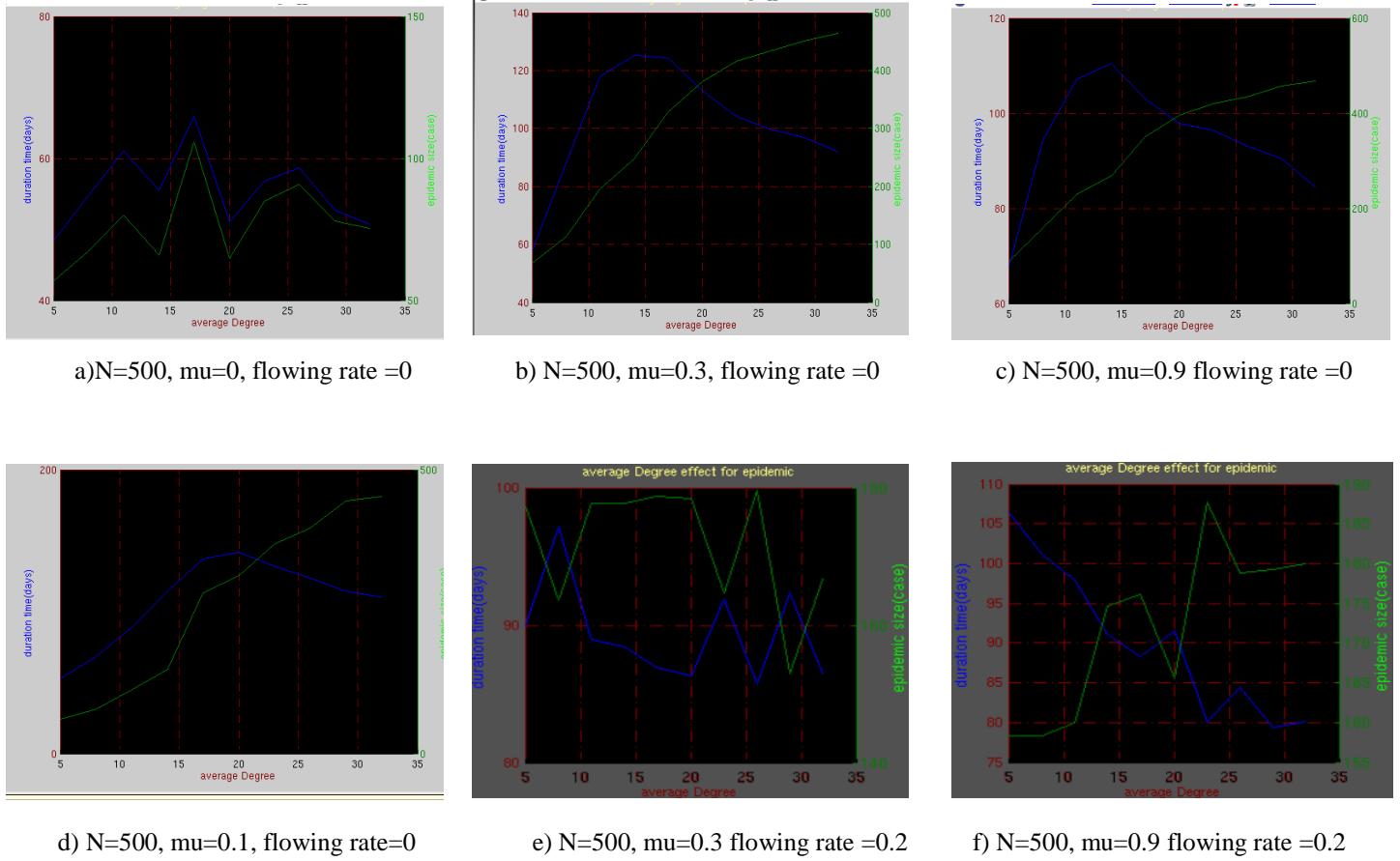


fig 5.6 average degree's effects to epidemic durations and final size

- *durations*

In these simulations, the degrees follow a power-law distribution with exponent parameter $t = 1$. As shown in fig 5.6 b) ~ d), the duration time (blue line) increases along with the average degree increase, where the average degree at a relatively smaller level. By contrast, the durations go downward after the average degree over some threshold. For example, in picture b), we can see that the growing trend of durations stopped when the average degree equals 18. Furthermore, compared with picture b) and e) or c) and f), it can be seen that under the same structured network, the population flow rate will interfere the average degree's effects.

- *final size*

In terms of epidemic final size, the result can be analysed by dividing into two groups: one is picture b), c) and d) compared with picture a) and e). Firstly, as picture b), c), d) shows that, the epidemic final size has a positive correlation with the average degree with given degree distribution. Their correlation values are 0.89, 0.96 and 0.81 for b) c) and d) respectively. By

contrast, picture e) and f) describes a lost effect of average degree to final size. It is probably because of its higher population flowing rate than other simulations.

In conclusion, firstly, it could be said that the effects of average degree to the duration time will depend on other network work properties such as community strength or population flow rate. In other words, its influence is not a linear function. Secondly, when analysing average degree' influence to the epidemic final size, the population flowing rate should be co-considered. To be exact, only when the population flowing rate is very low, will the average degree give positive effects to the final size. But it will lose its power when the population flow rate becomes higher.

5.2.2 Community strength (measured by mixing rate in LFR algorithm)

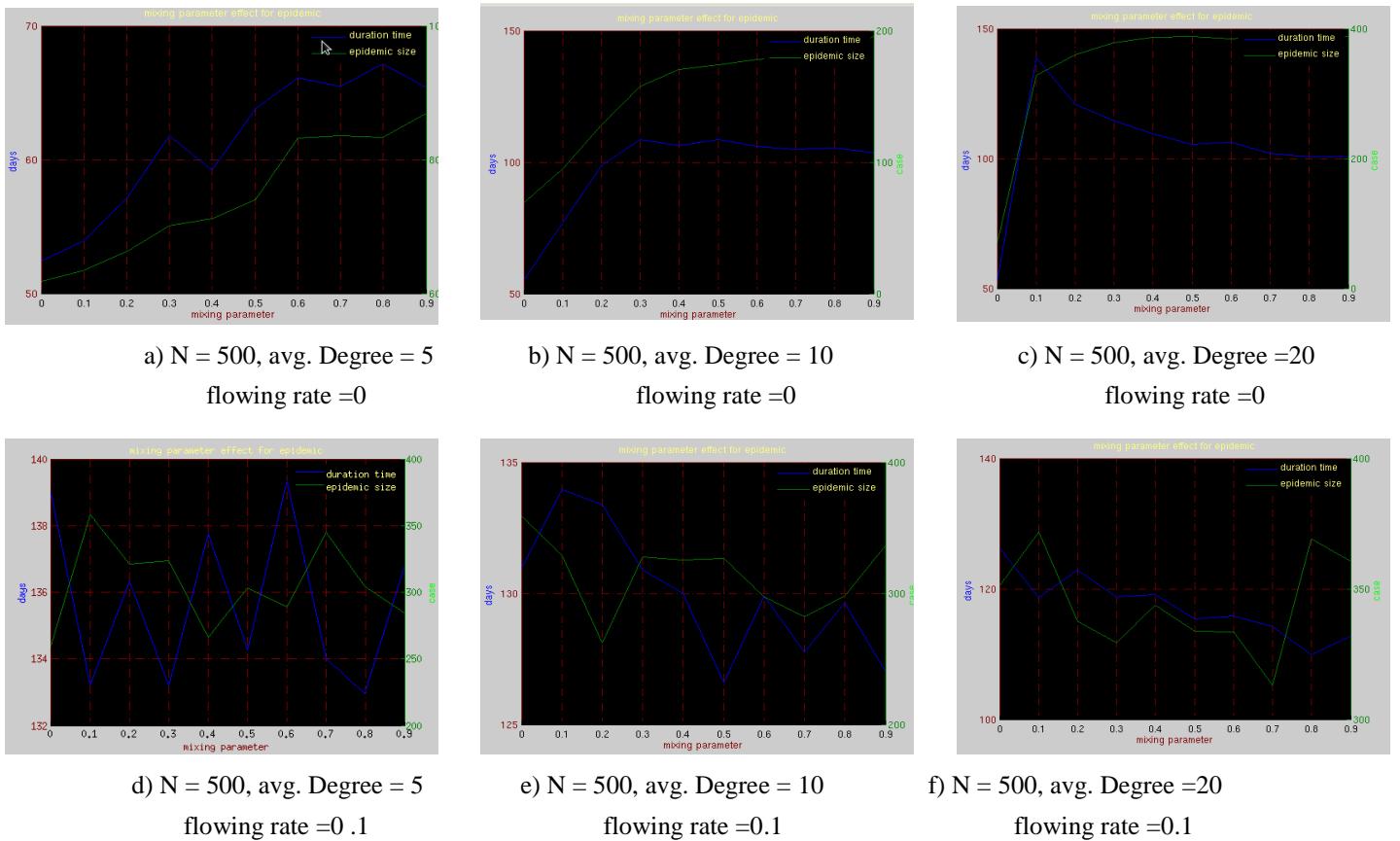


fig 5.7 effects of community strength to epidemic durations and final size

- *durations*

In the fig 5.7 a) ~ c), all the pictures reveal that the community strength has negative correlations with the durations. The trend increase of duration time stopped when the epidemic size tended to become stable. Moreover, as can be seen in figure c), when the epidemic breaks out in the entire network, the duration time will decrease with the community strength decrease. In brief, because of the community structure's ability of epidemic constriction, it will limit the number of

infectious individuals during the epidemic process. Therefore, stronger community structure will lead to shorter duration time. On the other hand, if the total number of infectious individuals is the same, this is because a weaker community structure leads to a higher speed in disease spread, and it will finish the epidemic process quicker than in the case of a strong community structured network.

By contrast, the figures from d) to f) present the downtrend of duration time with the community strength becoming weaker and weaker. This is because the population flow brings a lot of uncertainties. One of the probably reason is the movements event which lead to a quicker spread of the disease, which results in a relatively shorter duration time.

- *final size*

In terms of the correlation of community strength and epidemic final size, it is obvious that community structure can be one factor to reduce the final size. But when the dynamic network is considered, the effects come from this value to the final size will be lost.

5.2.3 Overlap rate under given community strength

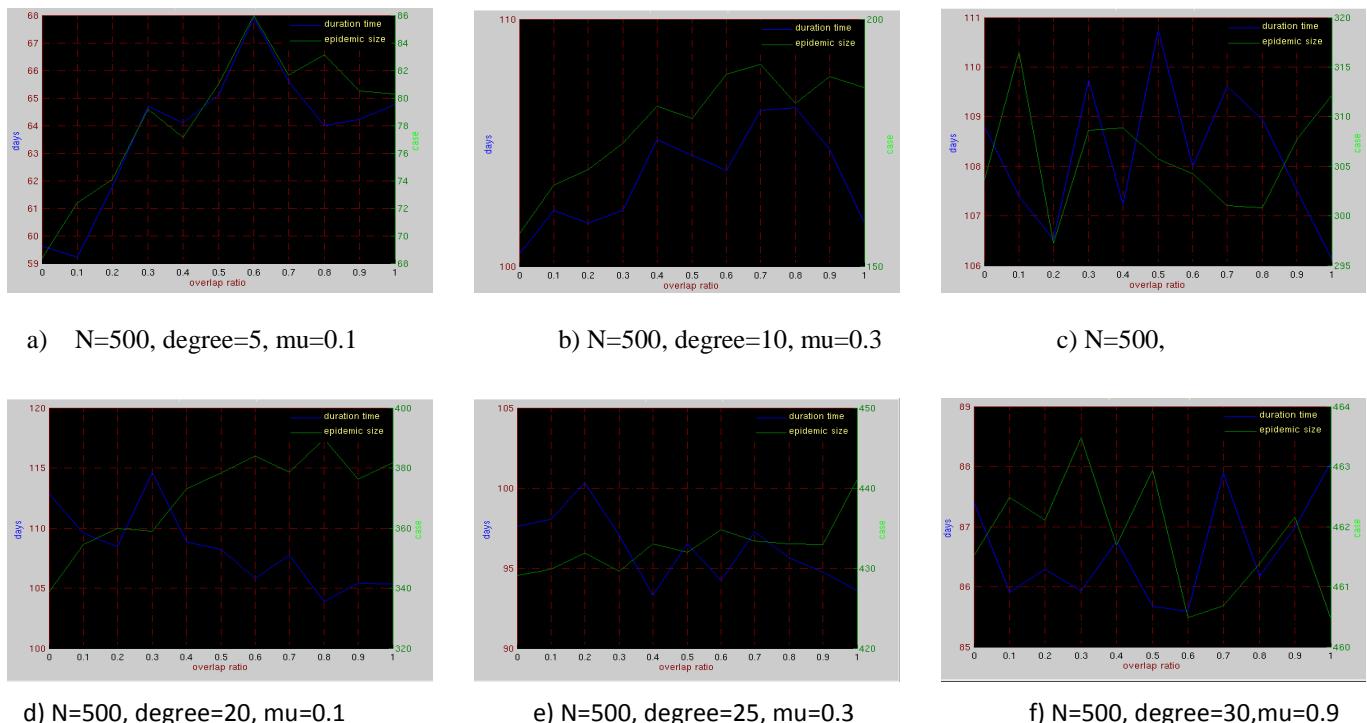


fig 5.8 effects of Overlap rate under given community strength to epidemic durations and final size in dynamic network

- *durations*

To analyze the effects which come from overlap rate, the basic community strength should firstly

be fixed. In addition, the overlap rate is a property relating to community structure. Hence, as we can see from fig c) and f), the duration time experiences a random change with a different overlap rate. In graph a) and b), the duration time shows a weak positive correlation with overlap rate with correlation value equal to 0.678 and 0.562, respectively.

In comparison, the duration line in fig. d) and e) represent downward trend with overlap ratio increase. Therefore, we can think the overlap rate is secondary parameter of community strength and average degree to affect durations.

- *final size*

Obtained from the pictures a), b), d) and e), we can see positive relations between the final size and overlap rate, especially clearly expressed by a strong community structured network a) and d). In statistics, the correlation values are 0.78, 0.87, 0.76 and 0.77 for a), b), d) and e).

5.2.4 Population flow rate

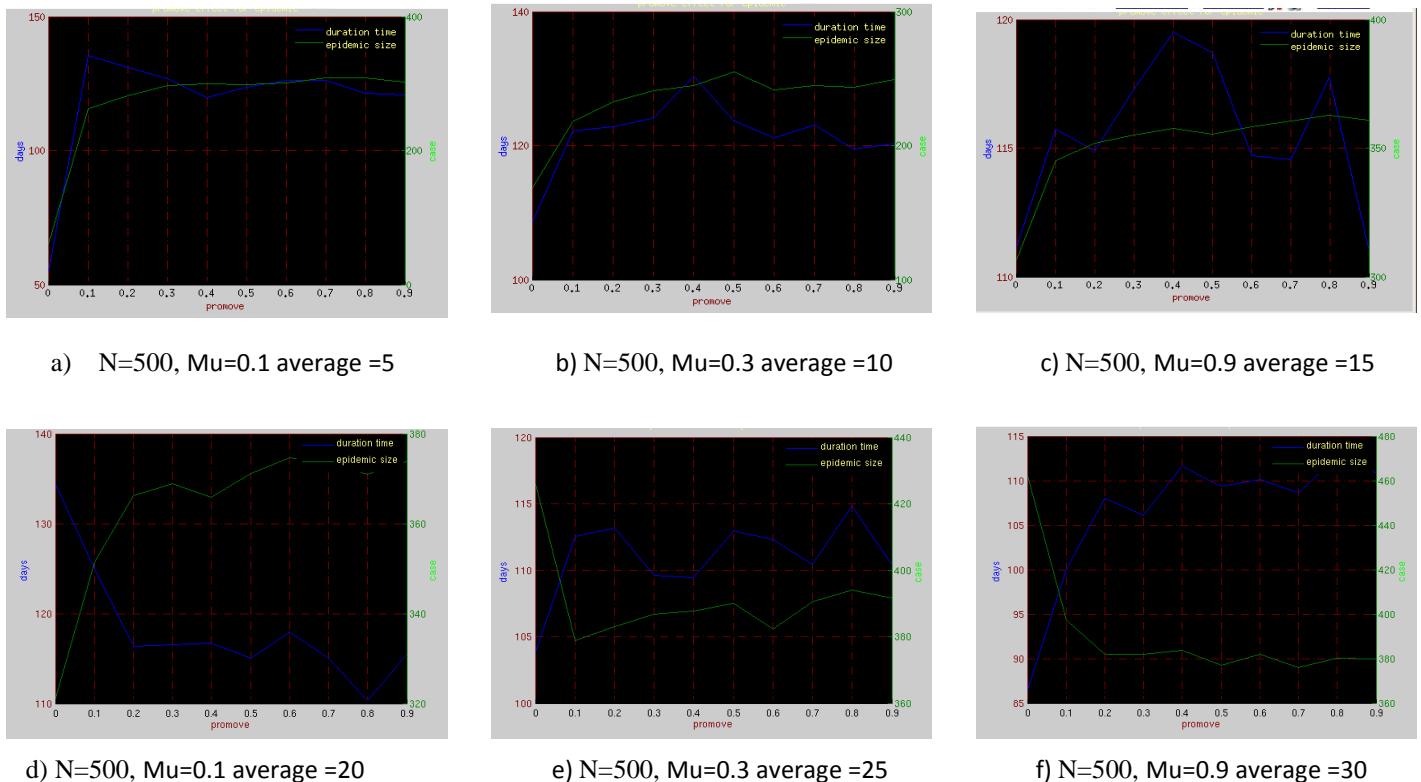


fig 5.9 effects of Population flowing rate to epidemic durations and final size in dynamic network

- *durations*

As can be seen from fig 5.9 a) to f), all the blue lines (illustrating duration time) except picture d) go upward with moving events happening and hovering around at a relatively high value. Compared to a

number of results, it is thought that the population rate gives uncertain factors to the duration time. In other words, the effects of flow rate depend on different network situation. Additionally, based on lot of experiments, for a determined network, the effects of population moving events can be predicted to levels of certainty.

- *final size*

From fig 5.9 a) to d), we can conclude that the population movements will promote the total number of infectious people. It is significantly different to consider the epidemic process in a static and dynamic network. But looking at the result from e) and f), the green line experiences an opposite behavior to the previous results. This probably is because that some other factors play a more important role in the very high density network.

5.2.5 Different community structure with similar evaluated parameter

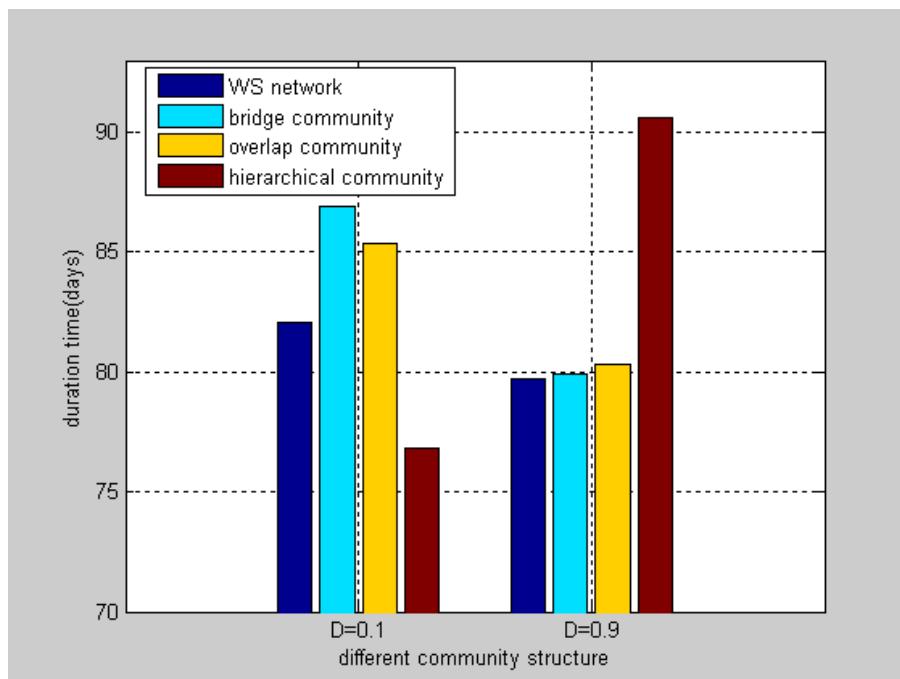


Fig.5.10 Effects of different community structured network to duration time

- *durations*

The results in fig 5.10 come from 1000 nodes networks with average degree 6 presented in table 4.10 group 1. As we can see that the process with disease parameter $D = 0.1$ and the parameter $D= 0.9$ are listed, which correspond to the process in fig 5.5 a) and b) respectively. When $D = 0.1$, the epidemic didn't break out and didn't spread through the entire network. Under this situation, because this specific hierarchical community structure restricts the process of disease spread, the total number of infectious individuals is much less than a simulation on other networks. This results in shorter durations for a hierarchical community structured network. By contrast, when D

$\lambda=0.9$, tends to produce a complete epidemic process. In other words, the disease spread throughout the entire network is quick. At this time, the duration of hierarchical structured network is expressed dramatically longer than others. This is because the speed of the disease spread is very slow in this kind of network structure, which leads to a long duration for the entire process. In conclusion, the hierarchical structure plays an important role in affecting epidemic process.

- *final size*

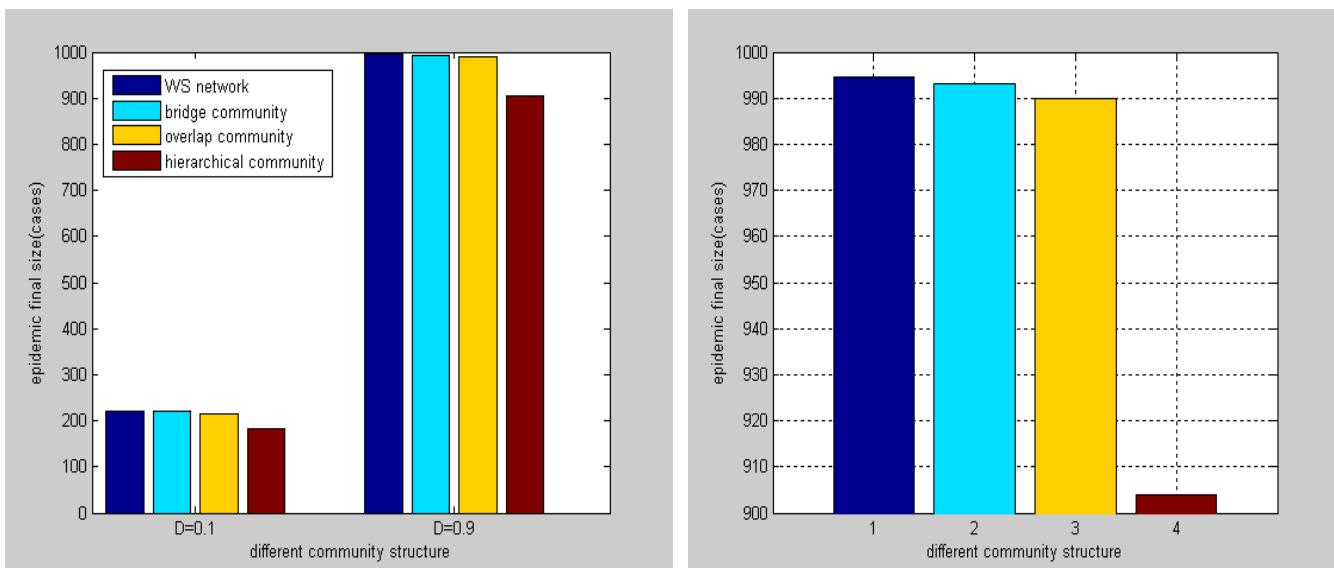


Fig.5.11 Effects of different community structured network to epidemic size

As we can see from the fig 5.11, the process under $D=0.1$ and $D = 0.9$ corresponds to the same one in fig 5.10 and the right side of the picture is the zoom-in of bar with $D=0.9$ in the left picture. Observed from these results, community structure can help disease control, especially in a hierarchical structured network. Even though D has a very high value 0.9, the hierarchical structure can protect some individuals from disease. To summarize, community structure can restrict epidemic size when an epidemic breaks out, especially a hierarchical structure which expresses significant effects within the epidemic process.

6. Conclusion and discussion

Based on the experimental results, a lot of potential relations are found between the network parameters and the epidemic evolutionary trend. The previous chapter gave the detailed results for every experiment which was conducted, and analysed the possible reasons. In this chapter, we the clear and evident answers will be concluded for the five questions proposed in section 4.1 to list what we achieved in this research.

Firstly, we found that the average degree for a given degree distribution and the community strength plays a significant role in affecting the epidemic growth rate. Both of them have obvious correlations with the growth rate. Average degree has a positive correlation whilst community strength has negative correlation.

Secondly, for the network with strong community structure, the overlap ratio is a considerable factor to change epidemic growth rate, which can cause disease to spread to a greater extent.

Thirdly, the population flow rate should be considered as an indispensable factor in how it affects the epidemic process in some specific network situation. For instance, in a sparse network with strong community structure, the flow rate is the most influential factor.

Fourthly, we find that a hierarchical structured network can constrict the spread process dramatically.

Fifthly, the final size of the epidemic is not so easy to predict by just one factor. The result is calculated depending on a number of network parameters. High population flow rate possibly leads to other factors to have less impact from their effects. In addition, if the other fixed parameters are weak, one parameter such as average degree and overlap rate will play a dominant role and have obvious correlation with epidemic final size. For example, if a static network has weak community structure, the final size of the epidemic will mainly depend on the average degree that the bigger average degree will lead to a bigger final size.

Finally, duration time relies on epidemic growth rate and its final size. Hence its value is difficult to calculate.

According to this conclusion, we can deduce two assumptions:

1. Epidemic growth rate can be considered as a function of average degree, community strength (represented by mixing value μ in table 2.2), population flow rate overlap rate and time t for a specific given disease, which we can describe as:

epidemic growth rate

$$= F(\text{average degree}, \text{community strength}, \text{overlap rate}, \text{population flowing rate})$$

Additionally, all of the variables have positive correlations with the growth rate. The relations are described in fig 6.1 a), where w_1 to w_4 represent the weights for different input factors.

- Epidemic final size is the nonlinear function of a number of network input parameters. The model is illustrated in fig 6.1 b), where x_1 to x_n represents network factors. Moreover, these factors have effects on each other. In other words, they combine to give the conclusive result of the final size of the epidemic.

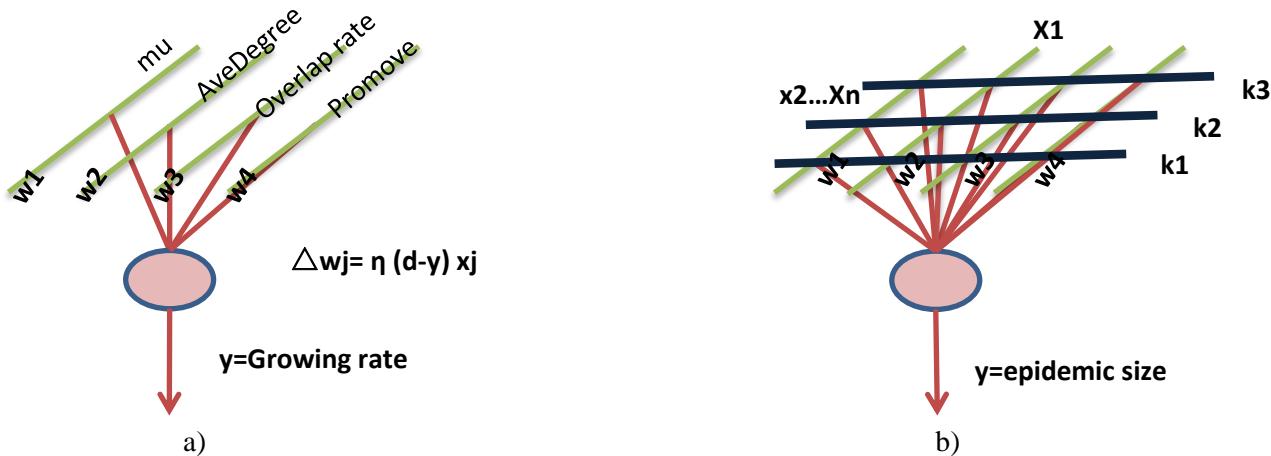


fig 6.1 models for the effects of network parameters on the epidemic variables

7. Future work

In chapter 6, we proposed two models for epidemic growth rate and its final size calculation. But the exactly value cannot be given. Hence, we need to find out weights for each input parameters in order to express the exact formulation.

Depends on a lot of observation, it was found that the trends of every simulation of these processes present the property of simple harmonic oscillation. Therefore, we can propose this function by the following formula:

$$y = A \times \sin(\omega x + \varphi) \quad (7.1)$$

Where A is the inverse function of x . ω and φ can be considered as the function of average degree, community strength, overlap rate and population flow rate. In fig 7.1, we use function 7.1 with different A , ω and φ to fit the trend of growth rate affected by a number of network parameters. In the following part, we will illustrate the idea of how to obtain the exact function in order to predict epidemic growth rate or duration times and its final size.

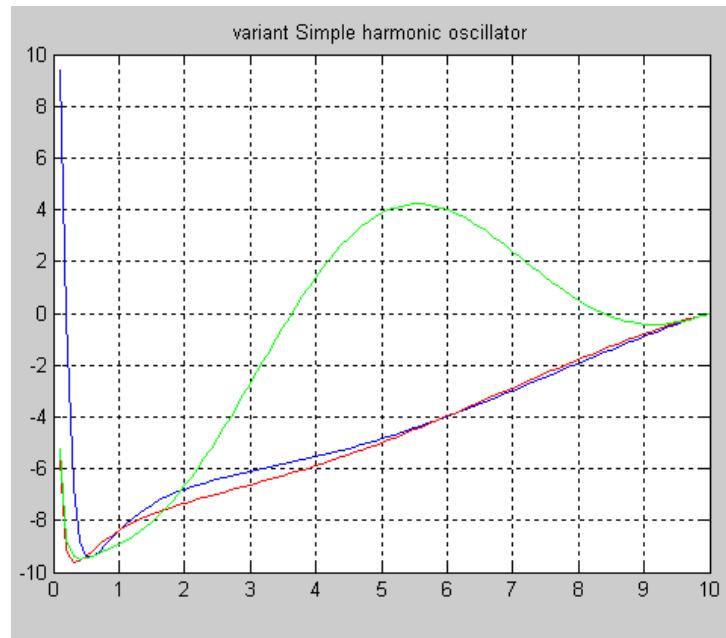


fig 7.1 fitting function to simulate the changing trend of growing rate

As shown in fig 7.1, it is an example of the first step of the future work for obtaining the equation for predicting the growth rate. After finding out the function that has a similar property with experiments on data relations, the concept of non-liner data regression will be used to fit the data trend. In addition, in the computer science area, the knowledge of machine learning, such as supervised learning can be used to help search the appropriate weights of these input variables.

8. Bibliography

- [1] Albert R.and.Barabasi A.L (1999) “Emergence of scaling in random networks” *Science*,286:509–512
- [2] Albert, R., Jeong, H., and Barabasi, A.-L.(1999), “Diameter of the world-wide web”, *Nature* 401:130-131.
- [3] Allen, E. J (1999) “Stochastic differential equations and persistence time for two interacting populations”. *Dyn. Contin. Discrete Impulsive Syst.*, 5: 271–281
- [4] Allen, L. J. S (2003) “An Introduction to Stochastic Processes with Applications to Biology” *Prentice Hall, Upper Saddle River, NJ*
- [5] Amaral LA, Scala A, Barthelemy M, Stanley HE.(2000), “Classes of small-world networks”, *Proc Natl Acad Sci U S A*. 2000 Oct 10;97(21):11149–11152
- [6] Barrat, A. and Weigt, M. (2000), “On the properties of small-world networks”, *Eur. Phys. J. B* 13:547-560.
- [7] BARABASI ALBERT-LASZLO and BONABEAU E.(2003), “scale-free networks”, *scientific American* 288: 60-69.
- [8] Bearman P. S. , Moody J. and Stovel K. (2004), “Chains of Affection: The Structure of Adolescent Romantic and Sexual Networks” *AJS Volume 110 Number 1*: 44–91
- [9] Blondel V. D., Guillaume Jean-Loup, Lambiotte R., Lefebvre E., “Fast unfolding of communities in large networks” *Statistical Mechanics: Theory and Experiment* 2008 (10), P1000
- [10] Boccaletta S., Latorab. V., Morenod, M. Chavezf M. , Hwanga D.-U. (2006) “Complex networks: Structure and dynamics” *Physics Reports* 424 pp: 175 – 308
- [11] Bollobás B., Random Graphs, Academic Press, London, 1985
- [12] Brandes U. (2001): “A Faster Algorithm for Betweenness Centrality”, *Mathematical Sociology* 25(2):163-177,
- [13] Brauer, F.and Castillo-Chávez, C. (2001). Mathematical Models in Population Biology and Epidemiology. NY: Springer.
- [14] Brauer F. (2000) “Compartmental Models in Epidemiology” in Brauer Fred and Wu Jianhong (ed) *Mathematical epidemiology*
- [15] Britton T., Deijfen M., Lageras N. A. and Lindholm M. (2008), “Epidemics on Random Graphs with Tunable Clustering” *J.Appl.Prob.* 45: 743-756
- [16] Britton T. (2009) “Stochastic epidemic models: A survey” *Mathematical Statistics Stockholm University*
- [17] Bryden J., Funk S., Geard N., Bullock S. and Jansen A. A. V. (2010), “Stability in flux: community structure in dynamic networks” *The Royal Society* published online 1 December 2010 doi: 10.1098/rsif.2010.0524
- [18] Cohen, R., Havlin, S. and ben-Avraham, D. (2005) "Structural properties of scale-free networks", in Handbook of Graphs and Networks: From the Genome to the Internet (eds S. Bornholdt and H. G. Schuster), Wiley-VCH Verlag GmbH & Co. KGaA, Weinheim, FRG. doi: 10.1002/3527602755.ch4
- [19] Demiris N. (2004) “Bayesian Inference for Stochastic Epidemic Models using Markov chain Monte Carlo Methods” the University of Nottingham
- [20] Clauset A., Moore C., and Newman M. E. J. (2008) “Hierarchical structure and the prediction of

- missing links in networks" *Nature* 453, 98-101
- [21] Dixon, R. B. (1928). The building of culture. New York: Scribner
- [22] Donetti, L., and Munoz M. A.. (2004), *J. Stat. Mech.* P10012.
- [23] Dorogovtsev S. N. and Mendes J. F. F. (2003) "Evolution of Networks Biological Nets to the Internet and WWW" NewYork: Oxford University Press
- [24] Duan D., Li Y., Jin Y., and Lu Z. (2009), "Community mining on dynamic weighted directed graphs," in Proceeding of the 1st ACM international workshop on Complex networks meet information & knowledge management. New York, NY, USA: ACM, pp. 11–18.
- [25] Dunham, Bigley J. and Berlin F. B. (2004), "Scale-free network models in epidemiology" derived from <http://academic.reed.edu/epi/papers/scalefree.ppt> (accessed 23 April 2011)
- [26] Eames K.T.D.(2007), "Modelling disease spread through random and regular contacts in clustered populations" *Theoretical population biology* 73: 104-111
- [27] Erdős P., Rényi A., (1959) *Publications of the Mathematical*. Debrecen 6
- [28] Erdos, P. and Renyi, A. (1960), "On the evolution of random graphs", *Publications of the Mathematical Institute of the Hungarian Academy of Sciences* 5: 17-61.
- [29] [Erdos, P. and Renyi, A (1961), "On the strength of connectedness of a random graph", *Acta Mathematica Scientia Hungaria* 12: 261-267.
- [30] Ferguson N, Cummings D, Cauchemez S, Fraser C, Riley S. (2005) "Strategies for containing an emerging influenza pandemic in Southeast Asia." *Nature* 437: 209–214.
- [31] Folino F., Pizzuti C. (2010), "Multiobjective Evolutionary Community Detection for Dynamic Networks" derived from <http://www.sciweavers.org/publications/multiobjective-evolutionary-community-detection-dynamic-networks> (accessed 20 April 2011)
- [32] GALLOP R. J. (1999) *Modeling General Epidemics: SIR MODEL*, Proceedings of the 12th Annual NorthEast SAS Users Group Conference.
- [33] Girvan M. and Newman M. E. J. (2002). "Community structure in social and biological networks". *Proc. Natl. Acad. Sci. USA* 99 (12): 7821–7826.
- [34] Greene D., Doyle D., Cunningham P. (2010), "Tracking the Evolution of Communities in Dynamic Social Networks" 2010 International Conference on Advances in Social Networks Analysis and Mining, Volume: 2010, Issue: 2010, Publisher: IEEE, Pages: 1-13
- [35] Havemann Frank, Heinz Michae, Struck Alexander and Jochen Glaser (2011) "Identification of overlapping communities and their hierarchy by locally calculating community-changing resolution levels" *journal of Statistical Mechanics: Theory and Experiment* 11: 1742-5468
- [36] Hethcote H. W. (1976), "Qualitative analyses of communicable disease models" *Math. Biosci.* 28:335–356
- [37] Hethcote H. W. (2000), "The mathematics of infectious diseases" *SIAM Rev.*, 42: 599–653
- [38] Hofmann R. A. (2011), "career opportunities in sport: social networking", derivrd from: www.icsspe.org/documente/Annette_Hofmann - Social Networking.pdf (accessed 20 April 2011)
- [39] Holland P. W. and Leinhardt S. (1971). "Transitivity in structural models of small groups". *Comparative Group Studies* 2: 107–124.
- [40] House T., Davies G., Danon L., Keeling J. M.(2009), "A motif-Based Approach to Network Epidemic"

Bulletin of Mathematical Biology 71:1693-1706

- [41] Jeong, H., Mason, S., Barabasi, A.-L., and Oltvai, Z. N. (2001), "Lethality and centrality in protein networks", *Nature* 411: 41-42.
- [42] Jingnan Liu "Advances in Computation and Intelligence" 5th International Symposium, ISICA
- [43] Kermack W. O.,McKendrick A. G. "Contributions to the Mathematical Theory of Epidemics. III. Further Studies of the Problem of Endemicity," Proceedings of the Royal Society of London. Series A, Vol. 141, (1933), pp. 94-122.
- [44] Lancichinetti A., Fortunato S., and Radicchi F. (2008): "Benchmark graphs for testing community detection algorithms"
- [45] Lancichinetti A., Fortunato S. and Kertész J., (2009), "Detecting the overlapping and hierarchical community structure in complex networks" *New J. Phys.* 11
- [46] Lancichinetti A., Fortunato S.,(2009) "Community detection algorithms: a comparative analysis" *Physical Review E* 80, 056117
- [47] Lancichinetti A., Fortunato S., (2009): "Benchmarks for testing community detection algorithms on directed and weighted graphs with overlapping communities" *Physical Review* Volume: 80, Issue: 1, Publisher: APS, Pages: 9
- [48] Latapy M. (2008), "Main-memory Triangle Computations for very large graphs", *Theoretical Computer Science (TCS)* 407 (1-3), p 458-473
- [49] Lin Y-R., Chi Y., Zhu S., Sundaram H., and Tseng B. L., "Facetnet: a framework for analyzing communities and their evolutions in dynamic networks," in Proceeding of the 17th international conference on World Wide Web (WWW'08), 2008, pp. 685–694.
- [50] Linda J. S. Allen (2000) "An Introduction to Stochastic Epidemic Models" in Brauer Fred and Wu Jianhong (ed) Mathematical epidemiology
- [51] Linda J. S. Allen (2008) "AN INTRODUCTION TO STOCHASTIC EPIDEMIC MODELS-PART I" derived from <http://www.springerlink.com/index/c62251111356477h.pdf> (accessed 18 April 2011)
- [52] Mariolis, P. (1975), "Interlocking directorates and control of corporations: The theory of bank control", *Social Science Quarterly* 56:425-439.
- [53] Markov process. (2011). In *Encyclopædia Britannica*. Retrieved from <http://www.britannica.com/EBchecked/topic/365797/Markov-process>
- [54] McDowell W., Reyes-Gonzalez L., Veloso F. M., (2010), "The Evolution of Community Structure in a Coauthorship Network," forthcoming in Systemics, Cybernetics and Informatics.
- [55] Mike (2005) "SOCIAL NETWORK SCIENCE" derived from: <http://www.istheory.yorku.ca/socialnetworktheory.htm> (accessed 18 September 2011)
- [56] Mike, 2007, "Disrupting Networks" <http://netwar.wordpress.com/2007/09/17/disrupting-networks/> (accessed 10 May 2011)
- [57] Miller J. C. (2009), "Spread of infectious disease through clustered populations". *J Roy Soc Interface* 6: 1121–1134.
- [58] Mills C. E., Robins J. M., Lipsitch M. (2004), "Transmissibility of 1918 pandemic influenza." *Nature* 432: 904–906.
- [59] Mills H. L., Cohen T, and Colijn C. (2010), "Modelling the performance of isoniazid preventive therapy for reducing TB in HIV endemic settings: the effects of network structure" university of

Bristol

- [60] Mizruchi M. S (1982), "The American Corporate Network", 1904-1974, Sage, Beverley Hills.
- [61] Moreno Y, Gomez J. B. and Pacheco A. F. (2002), "Instability of scale-free networks under node-breaking avalanches", *Europhys. Lett.*, 58 (4), pp. 630–636
- [62] Morris M. (1993) "Epidemiology and Social Networks - Modeling Structured Diffusion" *Sociol Method Res* 22: 99–126.
- [63] Molloy M. and Reed B.,(1995): "Random Struct. Algorithms" 6, 161
- [64] Nelson, D. L., McEvoy C. L. , and Schreiber T. A. (1998), The university of south orida word association, rhyme, and word fragment norms.
- [65] Newman M.E.J.(2001). "The structure of scientific collaboration networks" PNAS, 98:404–409.,
- [66] Newman, M. E. J., and Girvan M. (2004)," Finding and evaluating community structure in network" *Phys. Rev. E* 69(2), 026113.
- [67] Newman M. E. J., Moore C., and Watts D. J. (2000) "Mean-Field Solution of the Small-World Network Model" *PHYSICAL REVIEW LETTERS* volume 84, number 14
- [68] Newman M. E. J. (2003) "The structure and function of complex networks" *SIAM Review*,45:167–256
- [69] Padgett, J. F. and Ansell, C. K., Robust action and the rise of the Medici, 1400{1434, *Am. J. Sociol.* 98, 1259{1319 (1993).
- [70] Palla G., Barab ási A.-L. and Vicsek T. (2007), "Quantifying social group evolution", *Nature* 446,664.
- [71] Palla, G., Der_enyi I., Farkas I., and Vicsek T. (2005), "Uncovering the overlapping community structure of complex networks in nature and society", *Nature* 435, 814.
- [72] Pearce N. (1996), "Traditional Epidemiology, Modern Epidemiology, and Public Health", *American Journal of public health*, Vol.86, No. 5
- [73] Pool R (1989), "Is it chaos, or is it just noise?" *Science*; 243: 25-28.
- [74] Pool, I. de S. and Kochen, M (1978), "Contacts and influence", *Social Networks* 1:1-48
- [75] PHILIPPE P. and MANSI O. (1998), "NONLINEARITY IN THE EPIDEMIOLOGY OF COMPLEX HEALTH AND DISEASE PROCESSES" *Theoretical Medicine and Bioethics* 19: 591–607.
- [76] Reichardt J. and Bornholdt S. (2006). "Statistical mechanics of community detection". *Physical Review E* 74 (1): 016110.
- [77] Salathé M., Jones J. H. (2010) "Dynamics and control of diseases in networks with community structure". *PLoS Comp Biol* vol.6 (4).
- [78] Scott J. (2000), "Social Network Analysis: A Handbook", Sage Publications, London, 2nd ed.
- [79] Scripps J., Tan, P.N., Esfahanian, A. H. (2007) "Node Roles and Community Structure in Networks." In: Proceedings of the Ninth WebKDD and the First SNA-KDD 2007 Workshop on Web Mining and Social Network Analysis (WebKDD/SNA-KDD), pp. 26–35
- [80] Smieszek T., Fiebig L., Scholz R. W. (2009) "Models of epidemics: when contact repetition and clustering should be included." *Theor Biol Med Model* 6: 15.
- [81] Swedberg R (2003) "Principles of Economic Sociology" Princeton University Press
- [82] Tang L., Liu H. , Zhang J. , and NazeriZ.(2008), "Community evolution in dynamic multi-mode networks." In *KDD '08: Proceeding of the 14th ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 677–685.

- [83] Taylor, H. M., Karlin, S. (1998), "An Introduction to Stochastic Modeling" 3rd ed. Academic, San Diego
- [84] Tzekina I., DanthiK. and Rockmore D. (2008) , "Evolution of community structure in the world trade web" *Eur. Phys. J. B* 63, 541
- [85] Wang X. (2010), "Review of Classical Epidemic Models" derived from:
<http://legacy.samsi.info/200910/ug/may.../epidemic%20modeling%20review.pdf> (accessed 18 April 2011)
- [86] Wasserman, S. and Faust, K (1994), Social Network Analysis, Cambridge University Press, Cambridge.
- [87] Watts D.J.,(1999) Small Worlds: The Dynamics of Networks between Order and Randomness, Princeton University Press, Princeton, NJ,.
- [88] Watts D. J.and Strofatz S. H. (1998), "Collective dynamic of "small-world" network." *Nature* 393:440-442
- [89] William B.S.(2002). *The Word as Scalpel: A History of Medical Sociology*. New York, NY: Oxford University Press.
- [90] Wu B., Ye Q., and Yang S., "Group CRM: a new telecom CRM framework from social network perspective," in Proceedings of the 1st ACM International Workshop on Complex Networks in Information and Knowledge Management (CNIKM), Hong Kong, China, 2009.
- [91] Wulff H. R.(1999), "The concept of disease: from Newton to Aristotle". *Lancet* 354: 50
- [92] Zachary W. (1977). An information flow model for conflict and fission in small groups. *Journal of Anthropological Research*, 33, 452-473

9 Appendix: Source code

• Network generation

1. Bridge community structured network

Function

```
bridge_community(community_number,bridge_number,node_in_community,K,p)
%%%generate communities based on small-world network, information store
in adjacency matrix
for i=1:community_number
prefix=zeros(node_in_community,node_in_community*(i-1));
suffix=zeros(node_in_community,node_in_community*(community_number-i));
communities_network=[communities_network;prefix,wsNetwork(nodes_in_com
munity,K,p),suffix];
end

% create bridge community network
%random crate connections between communities
for i=1:bridge_number
.....
communities_network(x,y)=1;
communities_network(y,x)=1;
end
% end for create bridge community network
```

2. overlapped community structured network

```
Function overlap_community(nodes_basic,nodes_overlap,K1,K2,p)
%generate 3 communities
% A and B are the basic networks and O is the part for the overlap
network_A=wsNetwork(nodes_basic,K1,p);
network_B=wsNetwork(nodes_basic,K1,p);
network_O=wsNetwork(nodes_overlap,K2,p);

%construct fundamental matrix with 2*nodes_basic+nodes_overlap
communities_network=[network_A,append1,append2;append1,network_B,append
2;append3,append3,network_O];

%initial node membership
community_belong=sort(randperm(2*nodes_basic+nodes_overlap));
community_belong=[community_belong,sort(randperm(nodes_overlap))];
```

```

.....
community_belong=[community_belong;community_belong1,community_belong2,
community_belong3,community_belong4];

%the possibility for the pair of nodes in the basic network is K1/nodes_basic
%we set the possibility p1 for the nodes in overlap connect to both of
for i=1:nodes_overlap
    for j=1:nodes_basic
        out=rand(1,2);
        if out(1)<p1
            communities_network(j,2*nodes_basic+i)=1;
            communities_network(2*nodes_basic+i,j)=1;
        end
        if out(2)<p1
            communities_network(nodes_basic+j,2*nodes_basic+i)=1;
            communities_network(2*nodes_basic+i,nodes_basic+j)=1;
        end
    end
end
%end for construct network

```

3. Hierarchical community structured network

Function

```

hierarchical_community(total_nodes,first_level_communities,second_level
_communities,K,p)
%this function generate a 3 level hierarchical structure
%%%give every node the membership (basic level and second level)
%randomly assigns nodes into basic communities
first_community=randint(1,total_nodes,[1,first_level_communities]);
%calculate basic community size
for i=1:first_level_communities
    first_community_size(i)=length(find(first_community==i));
end

%assign basic communities into second level communities
second_community=fix(first_level_communities/second_level_communities);

%calculate the statistics of the second level community
for i=1:second_level_communities-1
    secondCommunity(i).number=second_community;

```

```

secondCommunity(i).size=sum(first_community_size(start:start+secondCommunity(i).number-1));
start=start+secondCommunity(i).number;
end
%%% end of give every node the membership (basic level and second level)

% generate adjacency matrix for basic communities
for i=1:first_level_communities
    f_communities_array(i).A=wsNetwork(first_community_size(i),K,p);
end

% generate adjacency matrix for second level communities
before_size=0;
for j=1:second_level_communities

suffix=zeros(first_community_size(before_size+1),(secondCommunity(j).size-first_community_size(before_size+1)));

s_communities_array(j).A=[f_communities_array(before_size+1).A,suffix];
prefix_row=0;
for i=2:secondCommunity(j).number
    prefix_row=prefix_row+first_community_size(before_size+i-1);
    prefix=zeros(first_community_size(before_size+i),prefix_row);

suffix=zeros(first_community_size(before_size+i),(secondCommunity(j).size-prefix_row-first_community_size(before_size+i)));

s_communities_array(j).A=[s_communities_array(j).A;prefix,f_communities_array(before_size+i).A,suffix];
end
before_size=before_size+secondCommunity(j).number;
end

.....
%using probability p to connect every pair of nodes between second level
communities
head=1;
tail=secondCommunity(1).number;
before_size=0;
for i=1:second_level_communities

```

.....

```

for j=1:secondCommunity(i).number-1

accumulation_size2=accumulation_size1+first_community_size(before_size+j)
    for m=j+1:secondCommunity(i).number
        for n=1:first_community_size(before_size+j)
            for l=1:first_community_size(before_size+m)
% set value for adjacency matrix
s_communities_array(i).A(accumulation_size1+n,accumulation_size2+l)=1;
s_communities_array(i).A(accumulation_size2+l,accumulation_size1+n)=1;
        end
    end
accumulation_size2=accumulation_size2+first_community_size(before_size+m)
end
accumulation_size1=accumulation_size1+first_community_size(before_size+j)
end
before_size=before_size+secondCommunity(i).number;
end

```

• Disease spread

```

function S=SIRmodel(network,probinf,probrec, numberofinfecteds)
%network is pairs list for network information
%probinf: probability per infectious contact, per time step that you are
infected
%probrec: probability per time step of recovery

%define the states of the nodes
n=length(unique(network(:,1)));%number of total nodes
statevector=zeros(1,n);
a = randperm(n);
f = a(1:numberofinfecteds);
statevector(f)=1;
%in the statevector, 0 is susceptible, 1 is infected and 2 is recovered

%run the model
for time=1:300 %for each timestep
    statevectornew=statevector;
    for node=1:n
        nodestate=statevector(node); %gets state of node
        if nodestate==0 %it is susceptible

```

```
%calculate infected partners and infection probability
nodespartners=network(network(:,1)==node,2);
lineweights=network(network(:,1)==node,3); %weighted network
infectedpartners=nodespartners(statevector(nodespartners)==1);

kinfected=length(infectedpartners);
if isempty(kinfected)==0 %if have infected partners
    probnodeinf=1;
    for pp=1:kinfected
        w=probinf*lineweights(nodespartners==infectedpartners(pp));
        probnodeinf=probnodeinf*(1-w);
    end
    probnodeinf=1-probnodeinf;
else
    probnodeinf=0;
end
r_inf=rand(1);
if r_inf<probnodeinf
    statevectornew(node)=1;
else
    statevectornew(node)=0;
end
elseif nodestate==1 %node is infected
    r_rec=rand(1);
    if r_rec<probrec
        statevectornew(node)=2;
    else
        statevectornew(node)=1;
    end
else %(nodestate==2) %elseif nodestate==3 - dead, remove from network
    %nothing happens in this state
end
end %end looking at nodes that time step
statevector=statevectornew; %update so looking at this next time step
%put in here any network changes/updates (network evolution)
%remove any nodes that have died, remove their links from the network
%and birth in new nodes (susceptible)
end %end timesteps
```