

Abstract

There are two main purposes of this project. The first is to study the effect of behavioural changes (edge removal in this project) on the epidemic progress and network properties using network theory. We combined three epidemic models (SI, SIR and SIIR model) with four behaviour models. According to the sources and types of the awareness that leads to behavioural changes, the four behaviour models are global-disease based, global-belief based, local disease-based and local-belief based. We tested all the combination possibilities on different network structures by varying parameters of epidemic control strategies on our designed simulation. In order to reduce the bias, each single experiment has been run 150 times on theoretical networks and 300 times on real world networks.

The second purpose of this project is to design epidemic control strategy related to behavioural changes. We choose to view the effect of behavioural changes as edge removal in this project, although it could have other effects like reduced infection rate. Different from previous works, our strategies mainly focused on using local information to find critical edges for each node since the behavioural change is on the individual level. We tested various methods to find critical edges including using similarity, high degree, clustering coefficient, and local community detection algorithms. We compared these strategies with both the global strategies that need the whole network information to find critical edges and the local strategy that randomly removes edges to infected individuals.

Executive summary:

- I design a behaviour-disease model and create a simulation for doing experiments on it, see page 17-23.
- I design and implement various new local edge removal strategies based on existing network algorithms, see page 23-29.
- The research carried out and reported in this thesis consists of a comparison and evaluation of five local edge removal strategies and three control strategies, leading to following discoveries:
 - ✧ Some local community detect algorithm could be used reversely to find intercommunity edges only using each node's local information in network with community structure.
 - ✧ The higher the community strength, the easier to find intercommunity edges using local algorithm
 - ✧ Local edge removal strategies needs fewer network information and is also epidemic information independent, which is more feasible and realistic since epidemic and network information is too dynamic to know in real life.
 - ✧ “Local Modularity strategy” modified from Clauset's local community detection algorithm played as well as global edge betweenness algorithm in finding intercommunity edges and also faster in large network.
 - ✧ Reducing contact with people in other community was more effective than only reducing contact with infected people in own community with regard to control disease in global view, because communities can be separated.
 - ✧ Local strategy applied to each node during epidemic process could be as effective as being applied before epidemic start if each node removes enough percentage of edges. It also removes fewer edges totally.
 - ✧ If there is asymptomatic infectious state, increasing awareness spread rate is in a limited help because the awareness spread path might not be same as disease spread path.

Acknowledgements

First and foremost, I would like to show my deepest gratitude to my supervisor, Dr. Steve Gregory, who has provided me with invaluable advice, guidance and encouragement in every stage of finishing this project. Without his enlightening instruction, impressive kindness and patience, I could not have completed my thesis. His keen and vigorous academic observation enlightens me not only in this thesis but also in my future study.

I shall extend my thanks to Harriet Mills, Katy Turner, Caroline Colijn and Anders Johansson for all their kindness and help in the area of epidemic modelling.

Thanks to Harriet Mills for giving feedback on experiment results and for invaluable advice.

Last but not least, I would give my sincere appreciation for my families for their unconditional love and support.

Contents

1 INTRODUCTION.....	1
1.1 AIMS AND OBJECTIVES.....	1
1.2 STRUCTURE.....	2
2 BACKGROUND.....	3
2.1 NETWORK	3
2.1.1 <i>Real networks</i>	3
2.1.2 <i>Simulated networks</i>	3
2.1.3 <i>Idealized networks or theoretical networks</i>	4
2.1.4 <i>Social contact networks</i>	6
2.1.5 <i>Measure properties of network</i>	7
2.2 EPIDEMIC MODEL	9
2.2.1 <i>Deterministic differential equation model</i>	9
2.2.2 <i>Stochastic agent-based model</i>	10
2.2.3 <i>Measurement of epidemic</i>	10
2.3 BEHAVIOUR MODELS	10
2.3.1 <i>Cause of behavioural changes</i>	11
2.3.2 <i>Effect of behavioural changes</i>	11
2.4 SOME DISEASE CONTROL STRATEGIES IN EXISTING STUDIES	14
2.4.1 <i>Finding nodes connecting multiple communities</i>	14
2.4.2 <i>Finding critical edges in network</i>	15
2.4.3 <i>Intermittent social distancing strategy in [Valdez et al. 2010]</i>	15
2.4.4 <i>Rewire edge strategy in [Lagorio et al. 2011]</i>	15
3 METHODS.....	17
3.1 EXPERIMENT DESIGN	17
3.1.1 <i>Experiment one: global-belief-based</i>	18

3.1.2	<i>Experiment two: local- prevalence-based</i>	18
3.1.3	<i>Experiment three: global- prevalence-based</i>	18
3.1.4	<i>Experiment four: local- belief-based</i>	18
3.1.5	<i>Other experiments:</i>	19
3.2	NETWORKS USED IN EXPERIMENTS	19
3.2.1	<i>Random network</i>	19
3.2.2	<i>Exponential network</i>	20
3.2.3	<i>Scale-free network</i>	20
3.2.4	<i>Community network</i>	20
3.2.5	<i>Real network</i>	21
3.3	EPIDEMIC MODELS	22
3.3.1	<i>SIRR model</i>	22
3.3.2	<i>SIR model and SI model</i>	22
3.3.3	<i>Statistics of epidemic model</i>	23
3.3.4	<i>Awareness spread model</i>	23
3.4	EDGE REMOVAL STRATEGIES	23
3.4.1	<i>Local edge removal strategies</i>	24
3.4.2	<i>Strategies used to contrast</i>	28
3.5	THE SIMULATION ALGORITHM	29
4	RESULTS AND ANALYSIS.....	30
4.1	BEHAVIOURAL CHANGES HAPPEN BEFORE DISEASE SPREAD	30
4.1.1	<i>Infection ratio analysis in different network structure</i>	30
4.1.2	<i>Epidemic results on different remove fraction</i>	31
4.1.3	<i>Compare with global strategies</i>	35
4.2	ALL POPULATION AWARE IN DIFFERENT TIME STEPS	40
4.3	BEHAVIOURAL CHANGES HAPPEN DURING DISEASE SPREAD	41
4.3.1	<i>Compare with strategies applied before epidemic process</i>	41

4.3.2	<i>Compare with local “SI link” strategy</i>	42
4.4	THE EFFECT OF AWARENESS SPREADING	44
4.5	EFFECTS OF NETWORK PROPERTIES ANALYSIS.....	45
4.5.1	<i>Clustering coefficient</i>	45
4.5.2	<i>Degree distribution</i>	46
4.5.3	<i>Mixing parameter.....</i>	47
5	DISCUSSION	48
6	EVALUATION.....	50
7	FUTURE WORKS	51
7.1	DIFFERENT SPREADING RATE.....	51
7.2	REWIRE THE REMOVED EDGES	51
7.3	TEST ON OTHER EPIDEMIC NETWORKS	51
7.4	AWARENESS SPREAD IN DIFFERENT NETWORK	51
8	REFERENCES.....	52
9	APPENDIX: SOURCE CODE.....	55

1 Introduction

The most common infectious disease in human is usually transmitted by air during close proximity interactions. During the outbreak of influenza A (H1N1) in 2009, not only the disease itself but also public's sustained and widely influenced reaction has been studied. The phenomenon that most people wear face masks and stay away with public places mainly result from individual's fear of the high death rate through the whole world. It reminds us again how individual's behavioural change can make such a huge difference to the society. Thus individual's behavioural change again arise people's wide awareness to study how it can be linked to disease control since McNeil's first related study in 1976. Many researchers have tried to quantify and capture such behavioural changes in a systematic study both using mathematical method and simulation method. Our study is based on doing experiment on simulation, so we will not provide mathematical analysis such as mean-field. We focus on the influenza-like disease spreading by close proximity interactions, so all epidemic models, behavioural models and networks are related to this kind of disease. In this section, we will first describe the objectives of the project and then will give the structure of this dissertation.

1.1 Aims and objectives

The aim of our project is to study how human behavioural changes can affect epidemic spread and what kind of strategy related to behavioural changes can effectively stop disease spreading.

To achieve this aim, my project has the following objectives:

- Build a model combining behavioural changes and disease spreading.
- Build a simulation of the model.
- Design and test the performance of various strategies related to behavioural changes on simulation by varying different parameters.
- Evaluate the performance of various strategies

Our experiments are conducted for addressing following research questions:

- a) How behavioural changes will affect the epidemic size, duration and peak?
- b) How behavioural changes will affect the network structure?
- c) Whether it is possible for individual making behavioural change based on local information?
- d) How network structure and properties will affect disease spread and strategy performance?
- e) How awareness spread can affect the epidemic results?

1.2 Structure

This dissertation consists of 5 main chapters. Firstly, in “Background” section, we review the basic concept of epidemic models and the categories of behaviour models. Then various methods of integrating behavioural changes in epidemic models are shown. Next epidemic related networks are introduced along with the measurement methods of epidemic and network properties. By the end of background, related existing disease control strategies will be shown and be analysed to see where can be improved and what can be used in this project. Secondly, the “Methods” section explains how our simulation is developed and how our experiments are conducted. Next comes the “Results and analysis” section, where we will give a reasonable analysis according to experiments results. In fourth section, we summarize our key experiment findings and discuss its relationship to current epidemic and behaviour studies. An evaluation of this project is also explained in this section to determine whether objectives have been met. Finally, future work is discussed related to further improvements and further extensions that could be carried out based on our existing findings.

2 Background

2.1 Network

In this project network is used to model the structure of population where epidemic spreading on. The population is represented by a network where a node stands for an individual, and an edge between two nodes means that the two individuals interact in a way allowing disease transmission. The node degree is the number of the contacts of the individual. In network theory, there are various kinds of networks which differ in network properties (such as degree distribution). The networks related to epidemic study can be classified into theoretical networks, real networks and empirical networks depending on how networks are constructed [Matt and Ken 2005]. We will review these networks and explain how these networks can be used in our study. Also we will review how the network properties can affect epidemic spreading, which is important in design our disease control strategy.

2.1.1 Real networks

It is hard to find a real network related for a specific disease. Because having knowledge of every individual and relationship leading to infection is a time consuming and complex task. Although people use three main techniques to gather network information: infection tracing, complete contract tracing and diary-based studies. All these three techniques have disadvantages. The infection tracing can only record the interaction leading to disease transmission while contact tracing and diary-based studies rely entirely on individual providing correct and complete data [Matt and Ken 2005]. Unless we can find an existing dataset related to our project, it is less possible to use a real network in our project since we have no time and ability to gather such huge data ourselves.

2.1.2 Simulated networks

A simulated network is built by following the observed social characteristics. It is like creating a plausible topology. The properties of a network (like the degree distribution) can be set according to the real social characteristics. The process of disease transmitting in society can be modelled in such network to study the epidemic dynamics.

Some researchers have tried to simulate an airborne disease outbreak. [Matt and Ken 2005] reviewed that in order to study the spread of smallpox, a network with 2000 individuals was generated with a given age distribution and household size that agree with the value for the United States Census 2000. Several acting rules are required among the networks such as each community had connections with four other communities, households had 1 to 7 people per family, person-to-person transmission probabilities varies according to different places and so on. However, it is hard to proof that such simulation can fully represent the real situation of disease spreading in community. The unknown structure of network also help less in analysing what element of network is important in the epidemic dynamics.

This method is worth trying if a suitable dataset about the daily social communication of a city or town is available. It needs to know how people are connected and how they communicate each other. There are some existing datasets provided online by researchers in this area. But most of the datasets are small like a dataset of a club. There rarely has a dataset with a relationship of a town or a city. The only possible dataset about social communities are from virtual web communities such as “Facebook” which may not be as same network structure as the one within real population. So it is hard to find convincing datasets.

2.1.3 Idealized networks or theoretical networks

Since the real networks and simulated networks both need huge effort to obtain large scale and reliable data on real population. Many idealized networks have been constructed to theoretically study networks. These networks are based on the known features of real-transmission networks (like how individuals are distributed and how they are connected) but are simplified a lot to catch the key point that affecting the epidemic dynamic mostly. Here we introduce several idealized networks can be possible used in epidemic study.

2.1.3.1 Random network

In random network, each individual are randomly connected to mixed number of other individuals and the network has no clustering. Each individual has similar degree just like the random-mixing model. The disease spread rate is reduced in this model compared with random-mixing model because an infected individual reduces its neighbours' susceptible contacts and continues infects more susceptible neighbours to let susceptible status around it less and less. In random-mixing model, the number of susceptible around an individual is according to the whole population not the local environment, so random network limits the rate of disease spread locally [Matt and Ken 2005].

Instead of randomly choosing fixed number of contacts for each individual, the network can also be formed by connecting any two nodes with probability p . This network has a Poisson degree distribution and the growth rate in it is still reduced. Therefore, the epidemic dynamic in random network are same as an SIR epidemic in a randomly mixed population [Matt and Ken 2005].

2.1.3.2 Lattices

In lattices, individuals are regularly spaced on a grid (usually two dimensions) and each individual is connected to its contacts. The characteristic of this network is that contacts are all localized in space and the infection tends to spread locally because of clustering. Lattice models also show a reduction in early growth rate of the disease compared with random-mixing model. Growth rate are reduced even more rapidly than in the random networks, which is because the clustering makes fast saturation of the local environment [Matt and Ken 2005]. Lattice also shows a wave-like spread of infection. Both epidemic sizes and epidemic durations obey a power-law, which helps explain some observed behaviour of infection.

2.1.3.3 Small-world networks

In lattices, two random selected individuals have long path between them while in small world networks, there are some long-range edges connecting the individuals that are not neighbours. So small-world networks show the characteristic of both clustering and global connectivity [Watts and Strogatz 1998]. It is studied that a few long-range edges can significantly speed the population-level spread of infection. In this kind of network, the effects of both clustering of connection and long-range transmission of disease have been observed important in disease spread. Human social networks are thought to be small worlds [Matt and Ken 2005], thus this structure is important in modelling disease spreading in human society. Since features of small-world networks also present in social mixing network, it is suitable to do experiment on small-world networks to find disease control strategies. For example, it is said that the clustering phenomenon makes disease spread locally while the short long-range edges can enable rapid spread disease to reach wider area outside local regions [Matt and Ken 2005]. Therefore it is a good idea to consider what kind of behaviour changes can reduce the long-rang edges so that disease can be kept in small regions.

2.1.3.4 Spatial networks

Individuals are positioned within a given area and two individuals are connected with a specific probability decided by a connection kernel which usually decays with distance to make connection locally. Spatial network can be configured to generate lattices or small-world networks. The degree distribution of spatial networks is usually Poisson.

2.1.3.5 Scale-free networks

Scale-free networks [Barabási and Albert 1999] are used to study those highly connected individuals. The degree distribution in scale-free network is power-law, which means that some node have extreme large number of contacts compared with average nodes. The new joined individual tends to connect with those highly connected individuals, which is familiar with the phenomenon that individual wants to be friends with the most popular people. In this kind of network, the highly connected individuals play a vital role in the spread and maintenance of infection, where infection is concentrated among these highest degree individuals [Matt and Ken 2005]. Therefore, it seems valuable to study how to control the disease spread by controlling these highly connected individuals, which has been tried in our project.

2.1.3.6 Exponential random graph models

Exponential random graph models provide a method to construct network with a given set of properties. Since the probability of connection between two nodes is independent, the constructor can decide how exactly any two nodes are connected without breaking the network properties. [Matt and Ken 2005] reviewed that Markov Chain Monte Carlo can be used to create plausible networks only requiring knowing the network structure. Therefore, once the information of the social contact network structure where the influenza spreads on is known, a simulated network can be created using this method.

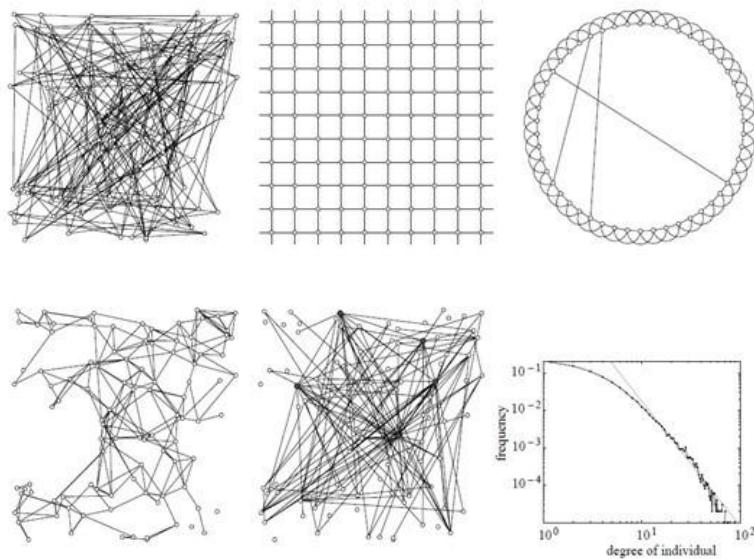


Figure 1: Five distinct network types from left to right: random, lattice, small world (top row), spatial and scale-free (bottom row). The bottom right-hand graph shows the power-law degree distribution of scale-free network. Source from [Matt and Ken 2005]

In summary, as shown in Figure 1, different network types have different network properties and suit for different study emphases. The random network have short path length but low clustering while the lattices have high clustering but long path length. Small-world networks

have clustering and short path length while scale free network particularly have some highly connected nodes. Therefore, lattices are useful to study how clustering stop disease spread and small world can be used to study whether reducing long-range edges (same as increase path length) can help control disease spread. When we want to control disease spread through highly connected individuals, studying scale-free network should be a better choice.

2.1.4 Social contact networks

Based on the previous review, we have known that many properties of network structure can affect the disease spreading, such as degree distribution, level of clustering, path length and so on. Thus disease show different dynamics in different theoretical networks such as random network, small-world network, lattices and scale-free networks. Since we tend to study flu like diseases, we may focus on the social contact networks where the influenza spreads on. In this section, we are going to introduce some networks used in previous papers to study influenza-like disease spreading. For the theoretical network, some researchers think scale-free network can stand for the contact network because it is consistent with the fact that some people or groups are highly contacted in real life. [Salathé and Jones 2010] studied how to control disease in networks with community structure. In their paper, they believe that social networks show significant community structure, with high clustering and modular. The structure of social contact network can also be revealed from real investigated data. After gathering the data of close proximity interaction in a high school, [Salathé et al. 2010] found that the contact network was a high-density network with typical small-world properties and a homogeneous distribution of both interaction time and interaction partners among subjects. In [Lagorio et al. 2011], they use one homogeneous network and one heterogeneous network to represent the social network respectively. For the homogeneous network, they use Erdős-Rényi networks with Poissonian degree distribution:

$$P(k) = \frac{e^{-\langle k \rangle} \langle k \rangle^k}{k!}.$$

For the heterogeneous network, they use finite scale-free networks with degree distribution:

$$P(k) = \frac{k^{-\lambda} \exp\left(-\frac{k}{K}\right)}{Li_\lambda\left(e^{-\frac{1}{K}}\right)},$$

where K is the degree cut-off. However, it is not clear that whether the real world contact network is scale-free or small-world or not. Since the real world contact network may be quite complicated and contain all of those properties, only one theoretical network not be able to represent it.

Alternatively, researchers will create an empirical social contact network using census data, traffic data, migration data or demographic data (shown in section 4.2). For example, an urban contact network model was built based on demographic information for the city of Vancouver, British Columbia [Meyers et al. 2005]. Another method of building empirical networks was shown in [Salathé and Jones 2010]. The social network website “Facebook” data was used to extract sub-graphs that individuals, who have friendships, live in the same dormitory residence, study in same class and have other actions that are likely to transmit an infection, will have many edges connecting each others. They use this assumption to construct several contact networks. [Salathé et al. 2010] also use wireless sensor network technology to obtain high-resolution data of close proximity interactions during a typical day at an American high school and thus construct a simulated network with weighted edges for an influenza-like disease. The problem is the accuracy of these kinds of empirical network is not known. Whether the data collected is convincing enough still need further verification.

Besides, [Bansal et al. 2007] have studied several empirical contact networks based on the real contact networks and find that the empirical human contact networks created in many research papers fit an exponential degree distribution. In the appendix of [Bansal et al. 2007], the algorithm of creating such an exponential degree network is provided. It first generates an uncorrelated random network and then rewrites the edges to show an exponential degree distribution. This is also a good choice despite those theoretical networks.

2.1.5 Measure properties of network

Individual's behavioural changes often lead to the changes in the properties of network. These changes usually play a vital role in controlling disease and our strategy also focus on controlling disease by changing these properties. Therefore, measuring these properties help us analyse and should be carried out frequently through the whole experiment.

2.1.5.1 Degree distribution

The probability $P(k)$ defines a node chosen at random will have degree k . The $P(k)$ distribution for all the nodes shows the degree distribution of the network. Degree distribution is an important way of measuring how network structure has been changed by individuals' behavioural changes. The average degree k of an undirected graph is defined as the number of all edges divided by the number of all nodes times two:

$$k = \sum_k kP(k) = \frac{2E}{N}$$

The degree distribution is critical in measuring the heterogeneity of the network that related to individuals' potential to become infected and causing infection. The higher the degree a node has, the more likely it will be infected and infect others. The moments of the degree distribution can help calculating the degree distribution variance and n th moment of $P(k)$ is defined as [Leon et al. 2011]:

$$\langle k^n \rangle = \sum_k k^n P(k) .$$

For the correlated network in our simulation, the probability $P(k' | k)$ of a node with degree k is dependent on the degree k' of its neighbours. Then the average of k' for all nodes with degree k can be calculated as [Leon et al. 2011]:

$$k_{nn} = \sum_k k' P(k' | k) .$$

k_{nn} is used to show the relationship between the degree of the node and its neighbours' degrees.

2.1.5.2 Distance

In a network, the distance between two nodes is the shortest path between them. The average distance $\langle d \rangle$ of all pairs of nodes can be defined as mean of the distances between all nodes [Leon et al. 2011]:

$$\langle d \rangle = \frac{1}{N(N + 1)} \sum_{i \neq j} d_{i,j} ,$$

where N is the number of nodes in the network and diameter of the network is $\max(d_{i,j})$.

The distance always shows how many steps are needed to spread disease from one node to another. The effect of small-world phenomenon is because of short average distance between all nodes which makes disease spread much more rapidly. So distance study has clear effect in controlling disease.

Shortest distance through a network also helps quantify the node betweenness centrality which is defined as the proportion of shortest path that pass through a single node [Leon et al. 2011]:

$$B_i = \frac{\# \text{shortest path through } i}{N(N - 1)},$$

where N is the number of nodes in the network. Edge betweenness centrality is defined similarly, which is the proportion of shortest path that pass through a single edge. It is widely known that central nodes are likely to be infected early and most of disease will spread through betweenness node. Thus controlling nodes with high betweenness plays an important role in disease control [Salathé and Jones 2010].

2.1.5.3 Clustering

The clustering coefficient ϕ represents the local density of a graph. It is defined as the probability that two neighbours of a node are also connected with each other and can be calculated as [Leon et al. 2011]:

$$B_i = \frac{3 \times \# \text{of triangles in the network}}{\# \text{of connected triples}}.$$

The clustering coefficient for each node in undirected graph can be calculated as:

$$C_i = \frac{2|e_{jk}|}{k_i(k_i - 1)},$$

where k_i is the number of neighbours of node i, e_{jk} defines all edges among neighbours of node i. Clustering always be considered important in disease spread for the rapid local depletion of susceptible individuals[Leon et al. 2011].

2.1.5.4 Community

The community is defined as each sub-graph having a high density of edges within it and lower density of edges between sub-graphs (see Figure 2). Community structure has been detected in studies of real world social networks representing the groups of people with highly similar demographics [Kitchovitch and Liò 2011]. Girvan and Newman [Girvan and Newman 2002] first introduce a community detection algorithm of “edge betweenness” and show that identifying community structure can help split the social and biological networks into meaningful clusters. [Salathé and Jones 2010] find that immunization interventions that change the structure of the contact network targeted in network with strong community structure are much more effective.

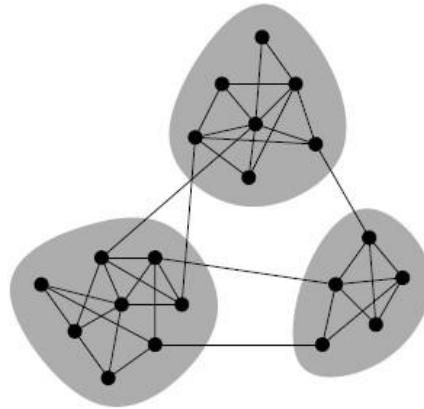


Figure 2: The nodes fall into sets of nodes (shaded) with many edges in set and few edges between sets.
(Source from [Newman 2006])

2.2 Epidemic model

The spread of disease is defined by an epidemic model. There have been a lot of epidemic models but most of them follow a similar pattern. Stemming from the first SIR model, most of the epidemic models contain several epidemic states (such as S, I, R states in SIR model) representing whether individual is infected or not. There are rates for people transferring from one state to another. The main difference between these models is that different models claim different numbers and names of state and define different methods to transfer node from one state to another. The most widely studied epidemic model is the SIR (susceptible-infected-recovered) model [Kermack and McKendrick 1927], which is considered as the foundation of almost all other disease models.

The epidemic models can be divided into deterministic and stochastic types. In deterministic model, the model is formulated by differential equations to express the changes of epidemic process with respect to time. The changes can be calculated using information of previous step. So the whole epidemic process is deterministic by inputting the initial status. A stochastic model contains random variable, which is used to estimate probability distribution of potential results.

2.2.1 Deterministic differential equation model

Below we first introduce the basic deterministic SIR model. As shown in Figure 3, individual is in susceptible state (S) at first. Once being infected, individual transfers to infective state (I). At last it becomes removed (R) which means recover from the disease and immune or dead. Infected individuals infect all states contacted with them at rate β per unit time while γ means average recover rate per unit time. This can be described by the following coupled non-linear differential equations [Bailey 1975]:

$$\frac{ds}{dt} = -\beta is, \quad \frac{di}{dt} = \beta is - \gamma i, \quad \frac{dr}{dt} = \gamma i ,$$

where s, i, r are the fraction of the population in each states and $s + i + r = 1$. The SIR model suit for infectious disease that confers lifelong immunity such as influenza, which is also our project focuses on. However, deterministic equation based model is more appropriate for representing expected results under mean-field approximation in mathematical study.

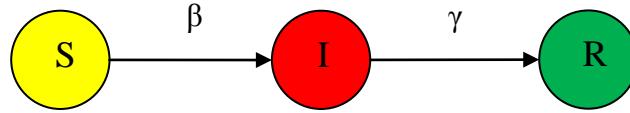


Figure 3: SIR model with the transitions of infection (rate β) and recovery (rate γ) (Source from [Funk et al. 2010])

2.2.2 Stochastic agent-based model

The above differential equation model assumes homogeneity and perfect mixing within compartments, which means the population mixes at random and each individual has same contacts per unit time and has same infect and recover rates. However in real life it is less possible. Thus agent-based mode is needed to capture heterogeneity among individuals and in network topology. In the agent-based model, each individual is separately represented and must be in different epidemic states discussed in deterministic model. Instead of using equation, each individual transmits between epidemic states by considering its neighbours' states in a specific probability.

Since our project focuses on individual's behavioural changes, the stochastic agent-based model is more suitable to study the changes in individual level and it can easily be run on different network topologies (e.g. random, scale-free etc.)

2.2.3 Measurement of epidemic

For the deterministic equation based model, the basic reproduction number denoted by R_0 is used to determine whether an epidemic occurs or die out, which is defined as the average number of secondary infections from a single infective individual in a fully susceptible population [Diekmann et al. 1990]. The infective individual makes βN contact per time step leading to new infections with duration of $1/\gamma$. So the basic reproduction number can be calculated by

$$R_0 = \frac{\beta N}{\gamma},$$

If $R_0 > 1$, the infection die out, otherwise there will be an epidemic in the population.

For the stochastic agent-based model, we can get the number of individuals in each epidemic state each time step. So we can calculate the ratio of infectious cases in each time step, the total infected cases at all (epidemic size), the highest infection rate and the epidemic duration as a measurement of epidemic.

2.3 Behaviour models

In the previous section, we have discussed epidemic models. Now in this section we will introduce different behaviour models and then we try to find a combination model of the SIR model talked above and the behaviour models. Our review in this section is based on the idea in [Funk et al. 2010]. There are various ways to model the behaviour change over time. [Funk et al. 2010] gave a classification of behaviour-disease model based on the causes and consequences of behavioural change. They focus on conceptual difference in the process to

combine behaviour element into disease model. In the following parts, we first discuss the classification of causes of behavioural changes and then give the summary of consequences of behavioural changes along with the integration of behaviour-disease model.

2.3.1 Cause of behavioural changes

With regard to the cause of behavioural changes,[Funk et al. 2010] classified all behaviour models based on the source and type of information that result in people's behavioural changes. The source of information can be global and local. If individuals base their behavioural choices on publicly information such as newspaper, TV news and other media, the source is global. Otherwise, they assume that the information is local and taken from social or spatial neighbourhood only. The global information is known useful for arising cluster to against disease spread (like the clustered occurrence of beliefs about vaccines) while local information is recognized to completely stop a disease from spreading in some circumstance.[Funk et al. 2009][Funk et al. 2010] The types of information can be separated into objective and subjective. If the information is related to disease prevalence, this kind of information is prevalence-based and subjective. Otherwise it is belief-based and has nothing to do with the disease prevalence [Funk et al. 2010]. For example, the decision about childhood vaccines is belonged to belief-based since the disease may not actually breaks out.

2.3.2 Effect of behavioural changes

[Funk et al. 2010] review three possible effects of behavioural change to disease spreading: disease state in epidemic model (such as S, I, R in SIR model), the infection or recovery rate and the contact network structure where the disease spread. The first type effect is mainly used in study how the decision to vaccinate affects the disease dynamics, because the change in state is assumed as a result of accepting vaccination. Since our project focuses more on the behavioural change rather than accepting vaccination, we mainly review the last two types of behavioural change which stand for the effect of people reducing contacts with neighbours when they view or believe the present of disease. With respect to the effect of model parameters like infection or recovery rate, it is mainly because people seek some treatment and have protective actions like wear face mask. For the effect of the network structure, it is always a consequence of people cutting edge of possible contagion by staying at home or avoiding contact with infected people. In order to distinguish with the effect of parameters change, many researchers let the individual forms a new edge after cutting one so that the structure can have a significant change. In the following part, we will review how these two effects have been integrated into epidemic models and try to find a way to form our own behaviour-disease model. We also review what kind behaviour related parameters have been created in models and how researchers studied them.

2.3.2.1 Regard behavioural change as change of parameter: Change of edges

Many researchers model the behavioural change as a spread of awareness. The awareness can be fear or the belief of presence of the disease. Such awareness may lead people hiding, isolating themselves to reduce contact with neighbours in network or flee to stay away from where the disease exists. Whether people accept such awareness is based on a possibility just like the idea that people is infected based on transmission probability. Therefore behavioural parameters can be integrated into epidemic model using the way of disease spreading. [Epstein et al. 2008][Funk et al. 2009][Funk et al. 2010]

S: Susceptible to pathogen and fear
 I_F : Infected with fear only
 I_P : Infected with pathogen only
 I_{PF} : Infected with pathogen and fear
 R_F : Removed from circulation due to fear
 R_{PF} : Removed from circulation due to fear and infected with pathogen
R: Recovered from pathogen and immune to fear

Figure 4: State space from [Epstein et al. 2008]

		Get scared	Not get scared
Get sick	$\alpha\beta$	$(1-\alpha)\beta$	
Not get sick	$\alpha(1-\beta)$	$(1-\alpha)(1-\beta)$	

Figure 5: Transmission Probabilities from [Epstein et al. 2008]

[Epstein et al. 2008] extends the basic SIR model to seven states shown in Figure 4. Each individual can be infected by disease (pathogen, see Figure 5) or awareness (fear, see Figure 5) and recovered from disease or fear. Let α denote the awareness transmission rate and β denote the disease transmission rate. Figure 5 shows the possibilities of individual in S state becoming state I_F , I_P , I_{PF} or remaining state S when he contacts an individual in state I_{PF} . If α is set to 0, a purely SIR model of disease is shown and if β is set to 0, a SIR model of awareness is got. Both the model will show a declining susceptible S-curve and an increasing disease infection S-curve. [Epstein et al. 2008] define two possible behavioural changed when individual is infected by awareness. The first one is that they hide themselves by cutting all edges around them in network for a specified time steps. The second response is that they move themselves to a new location in the network within specified distances which is considered as flee. In such a model, they need to study the effects of the parameters in the disease dynamics including the size of network, the degree distribution, movement and contact radii, the transmission rates of the disease and awareness, the duration of sickness, the distance fled and the duration of self-isolation.

They use experimental simulation to study the first behaviour and the second behaviour separately and also study the case combining these two behaviours. In their conclusion, the percentage of fled people has significant influence to the speed and size of disease. They find that even a small amount of flight dramatically increases the number of infected people and spreads the disease to further areas. However self-isolation does have effective reduce of the total number of infected people. This means that our project may not consider transferring people as a way to stop disease but do take individual's self-isolation into consideration.

2.3.2.2 Regard behavioural change as change of parameter: Infection rate

[Funk et al. 2009][Funk et al. 2010] similarly consider the awareness of disease spreads independently even in a different network from where disease spreads. But they consider the awareness fading parameter in the model which makes the model more practical. The behavioural changes are denoted as the changes in infection rate.

S ₋	Susceptible unaware
I ₋	Infected unaware
R ₋	Recovered unaware
S ₊	Susceptible aware
I ₊	Infected aware
R ₊	Recovered aware

Figure 6: Epidemic states from [Funk et al. 2010]

β	Infection rate from unaware infected to unaware susceptible
$\sigma_S \beta$	Infection rate from unaware infected to aware susceptible
$\sigma_I \beta$	Infection rate from aware infected to unaware susceptible
$\sigma_S \sigma_I \beta$	Infection rate from aware infected to aware susceptible

Figure 7: Infection rates from [Funk et al. 2010]

[Funk et al. 2009][Funk et al. 2010] define six states shown in Figure 6 to denote whether individual is aware or unaware. They assume the behavioural changes happen both in susceptible individuals and infected individuals. They denote the infection rate from unaware person to unaware person as β . If the infected individual takes some action to prevent the disease spreading, the reduction in infectivity is shown by a factor $0 < \sigma_s < 1$ multiple infection rate β . If the susceptible individuals are aware of the disease, they will reduce contact or take medication which reduces their infectiveness. Such reduce is denoted as a reduction factor $0 < \sigma_I < 1$. Thus the four resulted infection rate is shown in Figure 7. They also define a rate of infected individual becoming aware. The recovery rate is different between aware and unaware people. All these parameters help [Funk et al. 2009][Funk et al. 2010] find that reduced infectivity, shorter duration of infection of those aware and infected, reduced susceptibility of aware susceptible can inhibit the spread of disease. This influence is more effective if the network of awareness overlaps with the network of disease transmission. Therefore our project can try these effective methods in our models.

[Kiss et al. 2009] uses similar model but relates the awareness fading to the disease prevalence. Both the time and the decrease of disease prevalence will lead to awareness fade more quickly. They also get the conclusion that if individual have high awareness and seek treatment earlier, the disease prevalence will be reduced.

2.3.2.3 Regard behavioural changes as changes of network structure

Many researchers have found that the network structure can affect the disease spreading. [Huang and Li 2007][Salathé and Jones 2010] found that scale-free networks with strong community structure are helpful for reducing the disease spreading. Infection in large degree nodes can result in a larger-scale epidemic spreading in strong community structure but the prevalence scale is also more predictable in strong community structure. [Shaw and Schwartz 2008] design a model on an adaptive network. Based on susceptible-infective-recovered-susceptible (SIRS) epidemic model, if there is an edge connecting non-infected node to infected node, that non-infected node will cut the edge and add a new edge to another randomly selected non-infected node. They found that before rewiring all nodes show Poissonian degree distribution. But after rewiring, infected nodes tend to have low degrees and non-infected nodes have larger degrees since non-infected nodes tend to rewire to non-infected nodes. This also leads to the balance of the epidemic and disease-free states. Another worth learning parameter is the distance from nearest infective node. As shown in Figure 8(a), rewiring significantly decrease the number of nodes connected to an infected node compared to

un-rewired model. Figure 8(b) shows that the distribution of distance decays in the same way in an adaptive network and random network except that adaptive network has some nodes totally disconnected with the infected nodes.

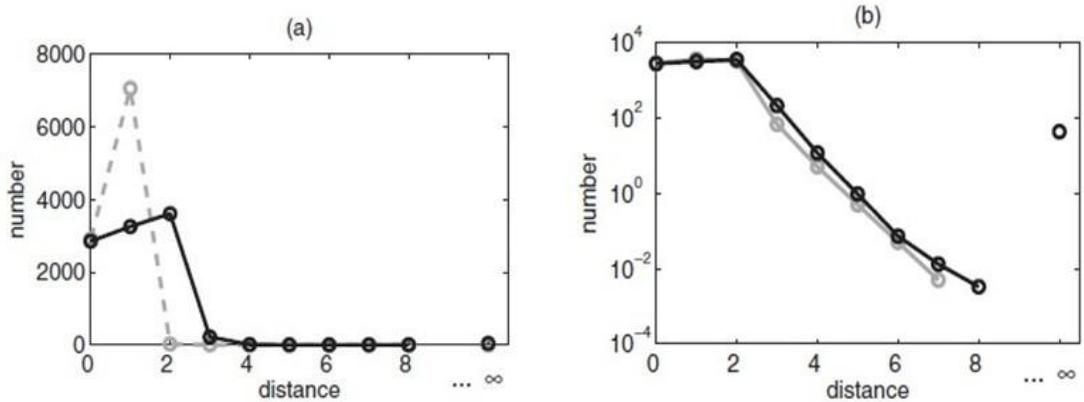


Figure 8: Distribution of distances from the nearest infected node. (a) Solid black line: with rewiring; dashed gray line: no rewiring (b) Black line: rewired case; and gray line: distribution for random graphs. Source from [Shaw and Schwartz 2008]

All these findings remind us that studying network structure is important in our project. It is worth study whether such adaptive network can be relates to a community structure since the infected node and non-infected nodes tend to form a community respectively. Also we can integrate parameter changes talked in above section into this adaptive network since node may show different rewire behaviour according to its local situation.

In summary, by reviewing these behaviour models we find that there are two types of behavioural changes: change the parameters in epidemic model or change the network structure. When modelling the behaviour spreading, most papers regard behaviour spread as a disease spread so that each individual accept a new behaviour based on a specific probability. Our project tends to study how the effect of removing edges which result from behavioural changes and lead to changes of network structure can affect epidemic process and design a strategy related to this kind of change to control epidemic.

2.4 Some disease control strategies in existing studies

In this section, we are going to review some strategies that can effectively control the spread of disease. These strategies may not necessarily related to behavioural changes but these strategies can give us some intuition about what kind of properties can be considered in controlling disease. Since there has few existing strategies about behavioural changes, we find some strategies related to immunization. The immunization strategies are aimed at finding key node and let that node immunization while our behavioural changes strategies can focus on finding key edges or the edges of key nodes and cut or rewire those edges. The immunization and behavioural changes strategies both have the idea of finding key features of network. We can learn what kind of feature is important and how to find it from those exiting immunization strategies so that we can apply them in our strategies. That is why we will introduce some immunization strategies below.

2.4.1 Finding nodes connecting multiple communities

In order to keep the infection of disease in one community only, the nodes connecting multiple communities show great role in controlling disease spread. One strategy of controlling disease spread is to immunize the individuals connecting multiple communities (so-called community

bridge)[Salathé and Jones 2010]. Despite of finding these nodes with betweenness centrality as community bridges, [Salathé and Jones 2010] design a community bridge finder (CBF) algorithm without the need to know the complete network structure. The general idea of CBF is to pick a random node n and follow a random path to arrive another node $n+1$. One step each time and Check if there is connection between node n and those visited nodes in previous steps. If there is no connection except the one between n and $n+1$, n is the potential community bridge. Then check whether the neighbours of node $n+1$ connect any visited nodes. If there is no connection, then the node n is the community bridge. Otherwise keep following random path and check again. The empirical and computationally generated networks tests shows CBF algorithm performed as well as traditional deterministic algorithm calculation shortest path to find betweenness central nodes. The CBF algorithm makes improvement because it does not need to know the whole global network structure which is often known little in real human contact process compared with traditional deterministic algorithm. But the efficiency of CBF is lower than traditional deterministic algorithm. So there still has some space to improve.

2.4.2 Finding critical edges in network

Paper [Marcelino and Kaiser 2009, 2012] have tested several edge removal strategies on real global flight network and found that cancelling fewer flights could result in both a larger slowdown of spreading and a smaller number of infected individuals compared to hub removal strategies. Removing critical edges means only fewer flight might be cancelled and fewer passengers are affected in contrast to the hub removal strategies which need to shut down the whole hub airports. They observed that removing the highest betweenness edges was the best at predicting critical edges that carried the most traffic loads leading to a large reduction in epidemic size. They also found that removing the edges ranked by structural Jaccard coefficient can delay the peak of epidemic mostly. They showed the critical role of highly ranked connections for the transmission of disease and demonstrated that the mechanism behind the good performance of these edge removal strategies display a community structure. The strategies tested are all global strategies which rank edges by using the whole network information while our project also concentrate on these highly ranked edges but only using local information to find them.

2.4.3 Intermittent social distancing strategy in [Valdez et al. 2010]

This strategy aims at SIR model in an adaptive network. Each time step an infected individual tries to spread the disease to a susceptible neighbour with probability β . If it fails, the susceptible will cut the edge with the infected individual with a probability σ for a period t . After that period the two nodes will be reconnected. This process repeats each time step until infected node transfer to R state. This strategy only requires each node using their local information. It was proofed to be useful (using percolation theory) in disease control for non-highly heterogeneous network by choosing a specific edge cut probability σ . In the real network test, this strategy still can stop some epidemic spreading for non-virulent disease. The main idea of this strategy is to create a susceptible cluster by cutting edges. Reconnection of removed edges make this strategy has less economic cost but still workable.

2.4.4 Rewire edge strategy in [Lagorio et al. 2011]

There are two strategies introduced in paper [Lagorio et al. 2011] according to whether susceptible nodes have information about the state of their neighbours. In strategy A, susceptible nodes have no state information of their neighbours until they have physical contact. At each time step, each infected node try to transmit the disease to its neighbours with some probability. If it failed, its susceptible neighbour will cut the edge between them and rewire to connect another randomly chosen susceptible node with probability w . In strategy B, each node

knows the state of its neighbours. Thus each susceptible node attached to an infected node will cut the edge between them and rewire to connect another randomly chosen susceptible node with probability w before physical contact. After proof using analytical approach and numerical simulation, they find that by increasing the rewiring probability w , the size of the epidemics can be reduced and the propagation can be stopped eventually with high rewiring probability w . Strategy A will not effective when the infection probability is high enough for physical contact happened earlier than rewiring. But strategy B always performs well both in homogeneous networks and finite heterogeneous networks. Since the rewiring probability is higher in scale free networks, they improve strategy B to rewire according to the degree of the infected nodes, which isolates the super-spreaders in scale-free networks. Although both strategies are shown workable in control epidemic spread. These strategies are not fully consistent with real social life. Because if node rewrites randomly, the possibility that it rewrites back when its previous neighbour is recovered is higher than rewiring to other unknown nodes. But they are considered equal in both strategies.

In summary, most strategies are based on two general ideas: finding key nodes or decide which edges to cut. Key nodes always have great roles in scale-free network or networks with community structures. These nodes have high degree or play a bridge role. Successfully finding them and do something with them always lead to effective control in disease spreading. When it comes to cut edges, most strategies will choose to cut edges with infected nodes and some may build new edges to susceptible nodes. These methods have been confirmed to be useful. But few of them think of cutting edges that connect further area (another community) to prevent disease spreading further, which our project focus on. Reducing these long range edges results in high level of clustering which may tend to keep infection in local area. Additionally, study in [Epstein et al. 2008] shows that totally moving susceptible nodes to another uninfected area may not be a suitable strategy since even a small amount of flight can dramatically increases the number of infected people and spreads the disease to further areas. This means that when designing rewire strategy, it is important to carefully control the number of edges that can be cut.

3 Methods

In this section, we are going to explain the methods of our experiments in detail. Our experiments have three main components: networks, epidemic models and edge removal strategies. Networks define where the disease spread while epidemic models determine how disease will spread. The edge removal strategies can be applied in different time steps of epidemic process to form our four main experiment models. Firstly, we are going to summarize experiment models. Then the detail of networks used and epidemic models will be discussed in section 3.2 and 3.3 respectively. Finally, algorithms of local edge removal strategies and disease control strategies will be described in section 3.4 and the whole simulation will be introduced in section 3.5.

3.1 Experiment design

According to source and type of behaviour (see section 2.3), the whole experiments can be separated into four parts (shown in table 1): strategies applied before disease start, strategies applied by infected nodes, strategies applied by all nodes in different steps and strategies applied by aware nodes. These four kinds of experiments are shown in Table 1 and a summary of all experiment parameters are shown in Table 2.

	Belief-based	Prevalence-based
Global	All nodes aware and strategies applied before disease start;	All nodes aware in different steps; Strategies applied by all nodes in different steps;
Local	Awareness spread; Strategies applied by aware nodes;	Only infected become aware, no awareness spread; Strategies applied by infected nodes.

Table 1: Behavioural change model

Networks	Edge removal strategies	Epidemic models	Variables	Output
Random; Exponential; Scale-free; Community; Real network;	Random; Similarity; CC; LFM; Local Modularity; High degree; Global: Edgebetweenness; InfoMap; Local: SI link;	SI model	Day; Remove fraction;	Percentage of infected nodes; Total removed edges;
			Remove fraction;	Epidemic size, duration, peak; Total removed edges;
		SIIR model	Remove fraction; Asymptomatic duration; Awareness spreading speed;	Epidemic size; Total removed edges;

Table 2: Experiment parameters

3.1.1 Experiment one: global-belief-based

This experiment can be viewed as whole population become aware of disease before disease actually starts spreading, which means that all nodes in network will apply edge removal strategies at beginning (regard as behavioural changes influenced by publicly available information). When each node applies strategy, there is a remove fraction limiting how many edges can be removed totally. Each strategy will find one candidate edge to remove each time and update before keep removing. If upper limit remove fraction is reached, no more edges will be removed for each node. On the contrary, if one node cannot find any candidate edge, even upper limit has not been reached the node will not remove any edge further.

The parameters of epidemic models are fixed for each network. As shown in table 2, we record the changes of infection ratio with spreading days in SI model to see how these strategies can affect the disease spreading. We also vary remove fraction to test that to what extent these strategies can control the disease in SIR model. The epidemic situations are measured by the percentage of infected node (SI model) and epidemic size, duration and peak (SIR model). Total removed edges are also recorded since one strategy leads to low epidemic size may not be the best control strategies if it totally removes more edges than other strategies in same circumstance.

All strategies are tried in this part except remove “SI link” strategy, because this experiment is aimed at comparing with global strategies which does not need the epidemic information. What we want to find out here is whether it is possible to get same control effect as global strategies only using local information.

3.1.2 Experiment two: local- prevalence-based

The setting of parameters, variables and outputs in this experiment are the same as experiment one (as shown in Table 2) except that edge removal strategies are applied during the disease spreading instead of at beginning. Additionally, global strategies will not be applied since this experiment is mainly for finding the effects of behavioural changes based on local information. Therefore “SI link” strategy will be used to compare with our local strategies in this experiment. When a node is infected, it will apply the strategies immediately before infecting other nodes.

3.1.3 Experiment three: global- prevalence-based

This experiment is a combination of experiment one and two. When the disease is spreading, the infected nodes will become aware and apply the edge removal strategies (just like experiment two). We set a global step which means that in this step, all nodes that have not been infected will apply the strategies (like experiment one). We vary this global step variable to see whether these edge removal strategies (when stands for belief-based and prevalence-based global behavioural changes) have different performances in belief and prevalence based global model.

We only use community network, SIR model and outputs of epidemic size and total removed edges in this experiment. The reason is that all strategies perform best in community network. For simplicity, SIR model and epidemic size are enough to conduct this experiment.

3.1.4 Experiment four: local- belief-based

In order to find out how local belief-based behavioural changes can affect disease spread, we use SIIR model to allow awareness spread. This means that not only infected nodes but also

aware nodes will apply edge removal strategies. In this case, infected nodes become aware and aware node will infect other unaware nodes. We also add an asymptomatic infected state before infectious state in SIR model to see whether awareness spread will help disease control in consideration of the fact that removal strategies may not always be applied in time before infected nodes start infecting. So beside remove fraction variable, asymptomatic duration and awareness transmission rate are varied as well.

3.1.5 Other experiments:

After conducting previous four experiments, it is surprising that one of our strategies (named “Local Modularity”) has extremely good control of disease in networks with community structure. So we continue to do three more experiments focusing on the effects of network properties including cluster coefficient, degree distribution and community strength on the performance of the “Local Modularity” strategy. Epidemic models and behavioural changes models are fixed in this part for simplicity. The details of experiments are shown in Table 3.

Network properties	Network	Model	Edge removal strategies	Outputs
Clustering coefficient	Exponential network			Epidemic size; Total removed edges;
Degree distribution	Scale-free network; Community network with power law degree distribution	SIR model; Global-Belief based behaviour model;	Local Modularity	The rate of removed edges are intercommunity edges;
Community strength	Community network			Epidemic size; Total removed edges;

Table 3: Parameters for experiments studying the effects on network property

3.2 Networks used in experiments

We use totally five types of networks that have been widely used in epidemic studies, including random network, exponential network, scale-free network, community network and real network. The properties of these networks have been discussed in section 2.1. In this section we mainly focus on the generation and data source of these networks. All networks used in our experiments are undirected and unweighted.

3.2.1 Random network

The random network in our experiment is generated using the Erdős-Renyi binomial model. There are 128 nodes in this network and each pair of nodes is connected with probability 1%, which results in total 859 edges. It is automatically generated using library of JUNG graph framework (`edu.uci.ics.jung.algorithms.generators.random.ErdosRenyiGenerator<V, E>`).

3.2.2 Exponential network

First we use JUNG to generate a random network with 168 nodes and 650 edges (each pair of nodes is connected with probability 0.05). Then we rewired it to exponential degree distribution by using greedy rewiring algorithm in [Bansal et al. 2007].

For the purpose of experiment in section 3.1.5, we also rewire the exponential networks to have different average clustering coefficients using algorithm in [Bansal et al. 2009]. We generated 5 exponential networks with same properties except the clustering coefficient. (See Table4)

Exponential network	Nodes	Edges	Mean degree	Coefficient of variation	Clustering coefficient
1	168	650	7.738	1.001	0.16
2					0.2
3					0.3
4					0.4
5					0.5

Table 4: Parameters of exponential networks

3.2.3 Scale-free network

To generate a scale-free network, we use an improved preferential attachment method in [Batagelj and Brandes 2005] to add a node each time and create edges between it and the existing nodes with the probability proportional to existing nodes' degrees. The generated scale-free network contains 128 nodes and 542 edges totally with mean degree to be 8.

3.2.4 Community network

To construct community networks, we use a program implementing the algorithm described in the paper [Lancichinetti and Fortunato 2009]. The algorithm is called Lancichinetti-Fortunato-Radicchi benchmark (denote as LFR) which can construct undirected and unweighted network with community structure and power-law degree distribution. When running the program, there is a mixing parameter to control the community strength. It stands for the ratio of external degree of total degree. The higher the mixing parameter, the lower the community strength will be. We generated five Girvan-Newman benchmark kind networks as shown in table5.

Community network	Mixing parameter	Community size/number	Total nodes	Total edges	Degree
1	0.05	32/4	128	1024	16
2	0.1				
3	0.2				
4	0.3				
5	0.4				

Table 5: Parameters of community networks

For the experiment to study the effect of power-law degree distribution in section 3.1.5, we also generate a community network with same degree distribution as a scale-free network

using Pajek which is a software of large network analysis and visualization developed by Vladimir Batagelj and Andrej Mrvar. The community network has totally 508 edges and 128 nodes. The degree distribution is shown in Figure 9:

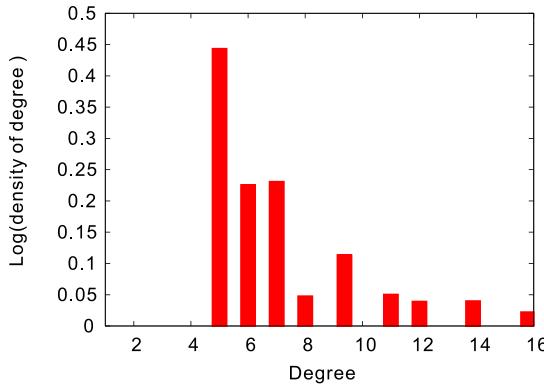


Figure 9: The degree distribution of community network and scale-free network

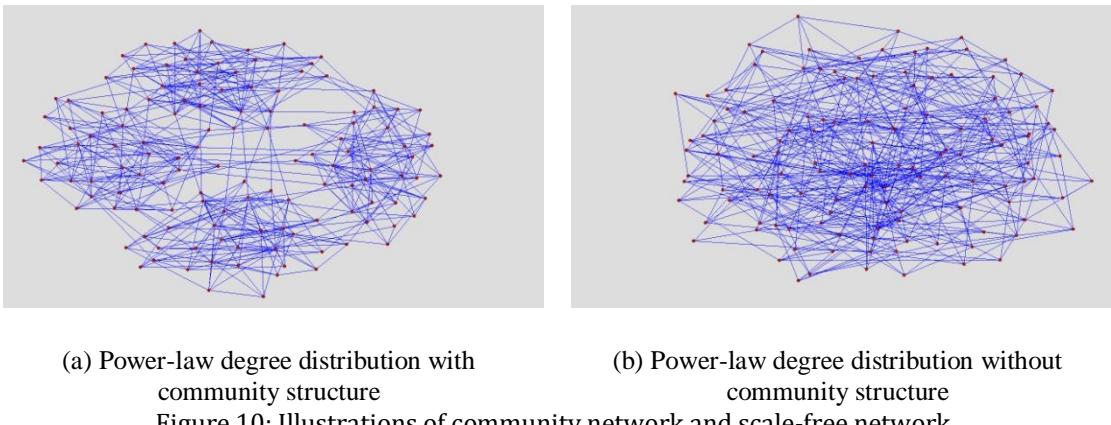


Figure 10: Illustrations of community network and scale-free network

The community network and scale-free network have same network parameters except that the community network contains four communities. (See Figure 10)

3.2.5 Real network

It is hard to find a real network related for a specific disease. Because having knowledge of every individual and relationship leading to infection is a time consuming and complex task. Since our experiments are aimed at testing the performance of edge removal strategies. We find three datasets used in real network tests: blogs, school and global flight network. The first two are social networks and the flight network is a real life network but is not social network.

1) “Blogs” network

“Blogs” is a network of blogs on the Window LiveTM Spaces platform from [Gregory 2009]. This network has strong community structure with totally 273 communities found by “InfoMap” community detection algorithm. There are totally 3982 nodes and 6803 edges in this network.

2) “School” network

“School” network is generated from a high school contact data of a day from [Salathé et al. 2010]. We only kept the contacts with duration longer than 5 second to form a network work with totally 786 nodes and 20345 edges. This dataset (have been mentioned above) is

collected using wireless sensor network technology and records high-resolution data of close proximity interactions among students, teachers and staffs during a typical day at an American high school. This dataset is collected particularly for influenza-like disease study so it is highly relevant to our project, although the data only records one day contact.

3) “Flight” network

We find a dataset of real global flight network from [Marcelino and Kaiser 2012]. This paper test alternative strategies using edge removal as cancelling targeted flight connection for controlling influenza spreading over airline network. They obtained scheduled flight data for one year provided by OAG Aviation Solutions (Luton, UK). 500 top airports which represent 95% of the global traffic are selected to form this dataset (totally 13038 edges). We choose this dataset for the reason that this comes from real data and our edge removal strategies also accord with the flight cancellation scenario. The difference between our strategies and the strategies in paper [Marcelino and Kaiser 2012] is that we only find removal edges using local information.

3.3 Epidemic models

Since our edge removal strategies care about the individual heterogeneity and different network topologies, stochastic agent based epidemic models are our best choices. In this section, we are going to introduce our threes epidemic models in detail, including SI model, SIR model and SIIR model.

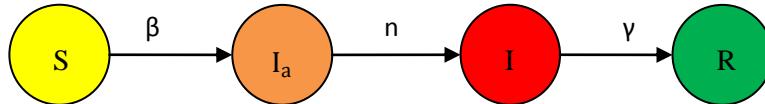


Figure 11: An illustration of epidemic model with parameters S (susceptible), I_a (asymptomatic infectious), I (symptomatic infectious), R (recovered), β (infection rate), n (asymptomatic duration), γ (recover rate)

3.3.1 SIIR model

Here we define SIIR model as shown in Figure 11: each node can be in one of four possible states: S (susceptible), I_a (asymptomatic infectious), I (symptomatic infectious), R (recovered). Initially all nodes are susceptible and one node will be chosen randomly to be in asymptomatic state. At each time step (one day), node in susceptible state can be infected by its neighbour in asymptomatic or symptomatic infectious state with probability β , which means that if one node has k neighbours in state I_a or I, the probability of being infected will be $1-(1-\beta)^k$. Once a node in state S is infected, it will transfer to state I_a . Each node in state I_a takes n days to become state I and then become recovered state R with probability γ . During a number of time steps, the initial infection can spread through the network and the simulation will be halted once no node is in state I_a or I.

3.3.2 SIR model and SI model

After defining SIIR model, both SIR model and SI model can be transferred from SIIR model. When we set $n=0$, state I_a will no longer exist and the model turns into SIR model. Similarly, if both n and γ are set to be 0, only state S and state I will be left. Thus SIIR model becomes SI model.

3.3.3 Statistics of epidemic model

The values of parameters used in these models vary according to different specific disease. Since we are not concentrate on one specific disease but a category of flu-like infectious disease which are transmitted from person to person by the respiratory or close-contact, we use the method introduced in paper [Salathé and Jones 2010] to decide our epidemic parameters. In epidemiology, R_0 (the basic reproduction number) represents the number of nodes that one node can infect on average over its infectious period. If $R_0 < 1$ the infection will die out otherwise infection will spread in a population. R_0 is widely used in mathematical models particularly in ordinary differential equations but cannot be used in our agent-based model. However, paper [Salathé and Jones 2010] mentioned that transmission rate β can be chosen according to the relationship that $R_0 \sim (\beta/\gamma) * d = 3$ where d is the mean network degree, and γ is the recovery rate. They choose $\gamma = 0.2$ (an average infectious period of 5 days) and $R_0 = 3$ because these parameter values reflect most widespread infectious diseases in previous studies. Therefore, we also set $R_0 = 3$ and let the infectious period to be 5 days in our simulation ($\gamma = 0.2$). So the transmission rate β in different networks can be calculated as following in Table 6.

Networks	Mean degree d	γ	R_0	$\beta = R_0 * \gamma / d$
Random	13.421	0.2	3	0.0447
Exponential	7.738			0.0775
Scale-free	8.46			0.07
Community	16			0.0375
Flight network	52			0.01
Blogs	3.42			0.1756
School	51.77			0.0116

Table 6: Epidemic parameters for different networks

3.3.4 Awareness spread model

In our model, only aware nodes will apply edge remove strategies. Nodes can be aware in different step by different means according to different experiment purpose (explained in section 3.1). For the global condition, all nodes become aware at same time influenced by public information. For the local, infected nodes become aware and the awareness can spread, which is same as disease spread (see Figure 12). Unaware node affected by its aware neighbour by probability w . If $w=0$, no awareness spreads.

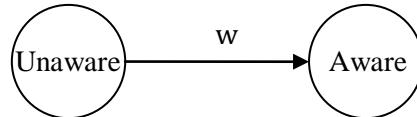


Figure 12: Awareness spread model

3.4 Edge removal strategies

Since behavioural changes are conducted by the individual who has limited knowledge of the whole network, the strategies used to select critical edges since each node can only use local information. We modified range of different measurements of network parameters including

similarity, degree, clustering coefficient, community strength and so on to form five candidate local edge removal strategies. In the following part, we will firstly provide all our modified local edge removal strategies and then give an introduction of those strategies we used to compare with.

3.4.1 Local edge removal strategies

When we apply these strategies, we try not to cause nodes with zero degree because in real life no individual can be totally isolated. Therefore, we will not remove an edge to the node only having one degree even if this edge is the best choice calculated by these strategies.

1) Similarity

We use Jaccard similarity coefficient proposed by Jaccard in paper [Jaccard 1901] to measure how similar neighbourhood connectivity structure of each node and its neighbours are. The Jaccard similarity coefficient are defined as $\sigma_{Jaccard} = \frac{|\tau_i \cap \tau_j|}{|\tau_i \cup \tau_j|}$, where τ_i equals to the degree of node i. This equation calculates the ratio of common neighbours between two nodes. If two vertice have exact same degrees, the similarity coefficient reaches maximum value of 1. Otherwise, a low coefficient reveals that the two vertice may belong to different network structure (e.g. community structure) and the edge connecting them will become a good target to remove resulting in a separate the two network structures. Therefore, each time for each node we firstly calculate the similarity between the node and each of its neighbours. Secondly we remove the edge leading to the neighbour that has the lowest similarity.

2) Clustering Coefficient (denote as CC)

Clustering Coefficient has been introduced in section 2.1 and local clustering coefficient of a node measures to how closely the neighbours are connected to each other which is viewed as a level of clustering. So we are going to find the critical edge for each node in purpose of increasing CC.

The steps for each node are described as following:

- Calculate the original CC for the node
- Calculate the changes of CC for removing each edge of the node respectively
- Choose the edge that can increase CC mostly to remove

3) LFM

We modify a local community detection algorithm called local fitness maximization (LFM) in paper [Lancichinetti et al. 2009] to try to find those intercommunity edges. Fitness of subgraph g is used to measure a community in paper [Lancichinetti et al. 2009] and is defined as:

$$f_g = \frac{k_{in}^g}{(k_{in}^g + k_{out}^g)^\alpha}$$

, where k_{in}^g and k_{out}^g are the total internal and external degrees of the nodes in community g and α is a positive value to control the size of the community. The community detection method starts from a node A and aims to find the largest fitness by adding new nodes or removing existing nodes in community, which means that the final detected community from node A will have lower f_g if a new node is added or a node of the community is removed. They also introduced a concept of node fitness f_g^A which is the fitness of node A in sub-graph g and is defined as the variation of sub-graph fitness with and without node A [Lancichinetti et al. 2009]:

$$f_g^A = f_{g+\{A\}} - f_{g-\{A\}}$$

In each iteration step, they find the neighbour with the largest fitness and add it into the sub-graph g . Then they recalculate the fitness of each node and removes the one has negative fitness. After each add and remove action, the fitness of each node will be updated and the algorithm will continue to iteration until all neighbours of sub-graph g have negative value. At the end, the generated sub-graph is the largest community from A.

When it comes to our modified strategy, we define an edge fitness based on the idea of node fitness in paper [Lancichinetti et al. 2009] which is variation of sub-graph fitness with and without the target edge. The fitness of edge e connecting node A and node n can be defined as:

$$f_g^e = f_{g+\{n\}} - f_{g-\{n\}}$$

, where sub-graph g consists of node A and its neighbours and $g - \{n\}$ here means node n is moved out of subgraph g and edge e is removed from original graph. Our idea of finding critical edges for each node is similar with paper [Lancichinetti et al. 2009], which is to remove the edge with lowest negative edge fitness value. After testing various α values, we find that $\alpha=2$ has the best performance in our experiment, so we set $\alpha=2$ in all experiments. When we apply the definition to our experiment, the fitness of sub-graph g which consists of node A and its neighbours can be calculated as:

$$f_g^A = \frac{2 * \text{internal degrees}}{(\text{internal degrees} + \text{external degrees})^2}$$

The fitness of each edge from node A can be calculated as:

$$f_g^{ei} = \frac{2 * I}{(2 * I + E)^2} - \frac{2 * I - 2 * i}{(2 * I + E - e - i - 1)^2}$$

, where ei stand for the fitness of edge connecting node A and node I, I is the total internal degrees of sub-graph g , E is the total external degrees of sub-graph g , i is internal degree of node I and e is the external degree of node i.

The LFM strategy for each target node consists of following steps:

- Consider the target node and its neighbours as a sub-graph g and calculate the sub-graph fitness f_g
- Calculate the edge fitness for each edge of target node
- Remove the edge with lowest fitness value
- If no edge has negative fitness value, no edge will be removed

Following is an example:

Suppose node A will apply the LFM strategy (see Figure 13). Nodes in dark blue colour are the neighbours of node A named from 1 to 6. They construct a sub-graph named g . The red lines are external degrees of sub-graph g while green lines are the internal degrees.

Calculate fitness of sub-graph g (consist of node A and its neighbours, see Figure 13):

$$f_g = \frac{22}{(22+10)^2} = 0.0215$$

Fitness for each neighbour: (f_g^{ei} stand for the fitness of edge connecting node A and node i)

$$f_g^{e1} = 0.0215 - \frac{22 - 2 * 4}{(22 + 10 - 2 - 4 - 1)^2} = -0.0009$$

$$f_g^{e2} = 0.0215 - \frac{22 - 2 * 3}{(22 + 10 - 3 - 1)^2} = 0.001$$

$$f_g^{e3} = 0.0215 - \frac{22 - 2 * 1}{(22 + 10 - 2 - 1 - 1)^2} = -0.004$$

$$f_g^{e4} = 0.0215 - \frac{22 - 2 * 2}{(22 + 10 - 1 - 2 - 1)^2} = -0.001$$

$$f_g^{e5} = 0.0215 - \frac{22 - 2 * 4}{(22 + 10 - 3 - 4 - 1)^2} = -0.0028$$

$$f_g^{e6} = 0.0215 - \frac{22 - 2 * 2}{(22 + 10 - 2 - 2 - 1)^2} = -0.00319$$

f_g^{e3} has the lowest fitness value, so the edge connecting node A and node 3 will be removed. Then go back to step a). Sub-graph fitness f_g will become 0.0255 and the edge fitnesses of e1, e2, e4, e5 and e6 will be updated. If there is a negative value, we can continue remove the edge of node A with lowest value. Otherwise no edge of node A will be removed any more.

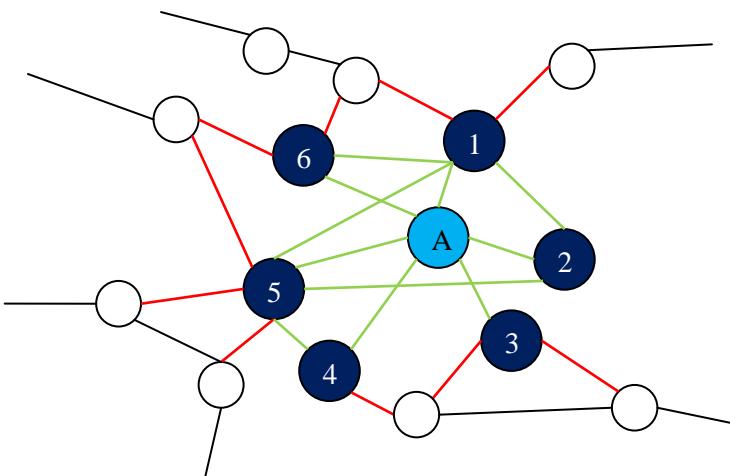


Figure 13: An illustration of sub-graph start from node A

4) Local Modularity (denote as LM)

This strategy is also modified from a local community detection algorithm in paper [Clauset 2005]. As shown in Figure 14, paper [Clauset 2005] defines a measure of local community C by taking the boundary of C into consideration. Boundary B of C makes up of those nodes in the subset of C that have at least one neighbour in U . It is expected that community structure tends to have sharp boundary with larger proportion of connections from boundary to local community C and fewer to the unknown part U that is outside of community.

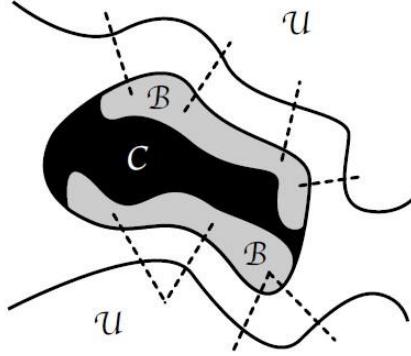


Figure 14: An illustration of Local Modularity (source from [Clauset 2005])

They measure a local community using local modularity R which is defined as [Clauset 2005]:

$$R = \frac{\sum_{ij} B_{ij} \delta(i, j)}{\sum_{ij} B_{ij}} = \frac{I}{T} ,$$

Where B_{ij} is 1 if nodes i and j are connected and either node is in B and $\delta(i, j)$ is 1 when either node $i \in B$ and node $j \in C$ or vice versa. T is the number of edges that with at least one endpoint in B while I is the number of edges in T with no endpoint in U .

This local modularity R lies from 0 to 1. If $R=1$ it means that a complete community is found which has no edges leading to unknown part U . It means that no intercommunity edge exists. Therefore, we tend to remove those edges that can increase R mostly by assuming that removed edges have high possibility to be intercommunity edges.

Suppose node A will apply “Local Modularity” strategy. The steps can be described as follows:

- a) Consider node A and its neighbours as local community C and find the B for C
- b) Calculate R of local community C
- c) For each neighbour i in B : Define E_i is the number of external degrees of node i which connecting nodes in U , I_i are defined as the number of internal degrees of node i that connecting nodes in C (except node A) and $\sum node j$ is defined as the nodes in C but not in B (except A) connecting node i and I_j is defined as number of the degrees of node j connecting other nodes in C but not in B (include node A). e is the number of edges that $\sum node j$ are connecting each other.
- d) Calculate $\Delta R_i = R - \frac{I - I_i - 1 + \sum I_j - e}{T - E - 1 + \sum I_j - e}$
- e) Choose the edge from node A to node i that ΔR_i is lowest negative value. If there is no negative value, no edge of node A will be removed otherwise iterate from step a).

Here we give an example of calculation (see Figure 15):

$$R = \frac{6}{12}$$

$$\Delta R_1 = \frac{6}{12} - \frac{6 - 2 - 1 + 1 - 0}{12 - 3 - 1 + 1 - 0} = \frac{6}{12} - \frac{4}{9} = 0.055$$

$$\Delta R_2 = \frac{6}{12} - \frac{6-2-1+1-0}{12-1-1+1-0} = \frac{6}{12} - \frac{4}{11} = 0.136$$

$$\Delta R_4 = \frac{6}{12} - \frac{6-0-1+0-0}{12-2-1+0-0} = \frac{6}{12} - \frac{5}{9} = -0.055$$

Node 3 is in C not in B , so ignore it. Since ΔR_4 is the lowest negative value, the edge from node A to node 4 will be removed.

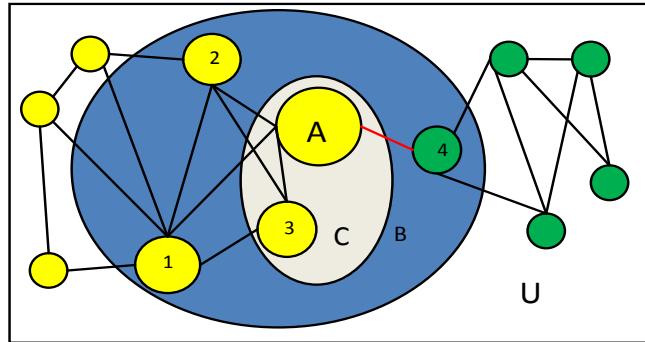


Figure 15: An illustration of applying “Local Modularity” in node A.

5) High degree

According to the study in [Marcelino and Kaiser 2012], hub removal strategy has been regarded as an effective way to stop influenza spreading over the airline network by shutting down the highly connected airport. Highly connected nodes are found by knowing the whole network information. We modified it so that it can be applied by individual node in a local way. Each time when a node needs to apply the “high degree” strategy, it firstly calculates the degrees of all of its neighbours. Then the node will remove edge to the neighbour with highest degree. High degree nodes are more likely to be infected and to spread diseases to more other nodes. Therefore, removing edge from high degree nodes is expected to keep them uninfected and reduce the opportunity of getting infected from them as well.

3.4.2 Strategies used to contrast

1) Global EdgeBetweenness strategy

The betweenness of an edge is defined as the times that it belongs to shortest paths between all pairs of nodes. It reveals how important an edge is in connecting the whole network. Edges with high betweenness value tend to have more important role to delay or stop the disease spreading. We implement the strategy based on the library from JUNG which is based on the algorithm in paper [Girvan and Newman 2002]. It calculates edge betweenness for all edges in the graph and returns desired numbers of edges with highest betweenness value. In our strategy, we remove these edges with high betweenness value to rewire the network. The running time is $O(kmn)$ where k is the number of edges to remove, m is the total number of edges, and n is the total number of nodes. The problem of this strategy is that it takes long running time in large network (e.g. global flight network with 500 nodes and 13038 edges) so we find another strategy to use in large network described in 2).

2) “InfoMap” strategy

“InfoMap” is a community detection algorithm of Martin Rosvall and Carl T which use the probability flow of random walks on a network to reveal community structure. We firstly run “InfoMap” strategy to find all communities and then regard all intercommunity edges as candidate removal edges. The above two strategies are all global for they need the information

of whole network. Although our edge removal strategies are only using local information, they still can be compared with these two global ones when all nodes in the network apply local edge removal strategies at same time. For time complexity reason, we use “global EdgeBetweenness” strategy in small networks and use “InfoMap” in large networks.

3) Remove SI link strategy

As reviewed in [Funk et al. 2010], individual’s behavioural changes may lead to susceptible nodes removing edge from infected nodes, which has been studied by a lot of researchers. [Shaw and Schwartz 2008][Bagnoli and Sguanci 2007]. Therefore, we also compare our local strategies with this kind of automatic changes by individuals. We implement it in a way that each node finds all of its infected neighbours and removes edges from these infected nodes randomly according to remove fraction.

3.5 The simulation algorithm

All simulations were implemented in Java using the JUNG (<http://jung.sourceforge.net/>) graph framework for the graph data model. For the whole experiment, we used a Monte Carlo approach and averaged results from 150 runs for each possible combination on theoretical networks and 300 times on real dataset. The disease starting node is chosen at random each run. The input parameters of simulation are summarized in table 7 below:

Parameter name	Explanation
Infection rate (β)	This is calculated according to specific network (Method has been discussed in section 3.3.3). Each iterative step, for each neighbour of infected nodes, a random number r in $(0,1]$ is generated; If $\beta \geq r$, the neighbor become infected.
Asymptomatic infectious duration (n)	If $n=0$: SIR model; If $n>0$: SIIR model; When a node get infected, it will be in state asymptomatic I for n step and then become symptomatic I state.
Recover rate (γ)	All nodes in state symptomatic I take d steps ($d=5 \cong \gamma=0.2$) to become state R. If $d \rightarrow \infty (\gamma = 0)$: SI model; Otherwise SIR or SIIR model.
Network data	All network (undirected and unweighted) data are stored in Pajek NET format.
Remove fraction (denote as rf)	Maximum remove limit for each node;
Strategy number	Define which strategy to use;
Maximum step	Simulation will stop if reach maximum step or there is no node in state I
Global step	-1: All nodes become aware and apply strategy before disease start; (-1, maximum step): infected nodes apply strategy first and at global step all nodes have not applied will apply strategy; > Maximum step: only aware nodes apply strategy.
Awareness spread rate (w)	Infected nodes become aware; Each iterative step, for each unaware neighbour of aware nodes, a random number r in $(0,1]$ is generated, if $w > r$ unaware neighbour become aware; $w = 0$: No awareness spread, only infected nodes become aware.

Table 7: Input parameters of simulation

4 Results and analysis

4.1 Behavioural changes happen before disease spread

4.1.1 Infection ratio analysis in different network structure

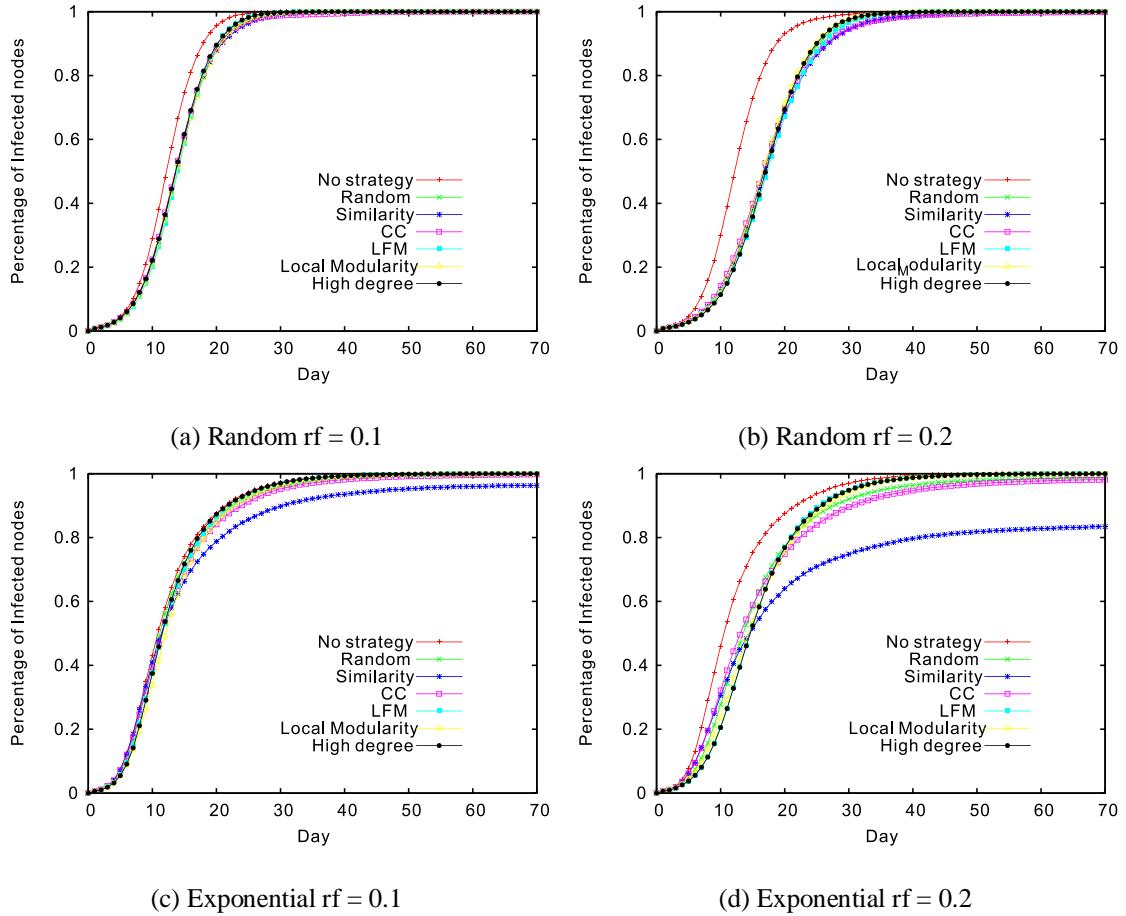


Figure 16: Ratio of infected nodes in SI model in Random and Exponential network with different remove fraction (rf)

As can be seen in Figure 16, our five candidate local edge removal strategies are applied before disease start in random and exponential network respectively. The disease is spreading in SI model which means that all nodes will be infected at last if the network has not been separated. In the random network (a), it shows that local edge removal strategy can delay the disease spread compared with no strategy applied, but random removal has similar performance with purposive removal. However, when it comes to exponential network (c), we can see that “Similarity” performances much better than other strategies since it leads to fewer nodes to be infected at last. When we compare (a) and (b) or (c) and (d), we can find that increasing remove fraction results in better control for all strategies, especially for “Similarity” in exponential network which separates the network. “CC” strategy shows better delay function in disease spread with more edges being removed but still infects all nodes at last. For the other strategies, their performances are just similar with random removal even when more edges are removed. Therefore, it can be known that “Similarity” and “CC” can be more effective when each node removes more edges.

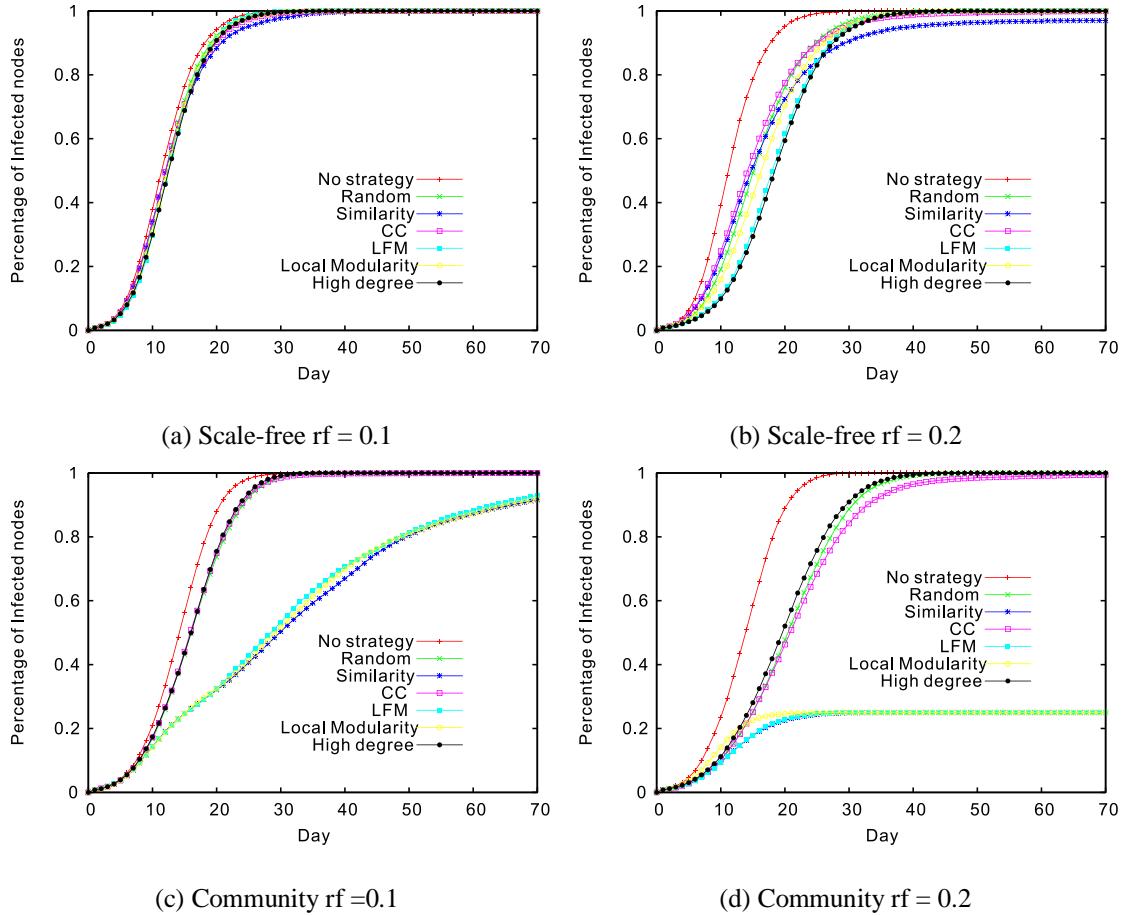


Figure 17: Ratio of infected nodes in SI model in Scale-free and community network with different remove fraction (rf)

Figure 17 has same compare structure as Figure 16 except that the test networks are replaced by scale-free network and community network. From (a) and (b), it is clear that increasing remove fraction for each node is useful in scale-free network, which is same as random and exponential network. The difference is that “LFM”, “Local Modularity” and “High degree” strategies present better delay in first 20 days of outbreaks even better than “Similarity”. When we do test in community network, totally different pictures are revealed in Figure 17 (c) and (d) compared with previous three kind of networks. Even only 10% edges are removed for each node; “Similarity”, “LFM” and “Local Modularity” display nearly twice delay in disease spread when 50% nodes are infected (30 days in contrast to 15 days for other strategies). What’s more, when remove fraction becomes 20%, these three strategies totally isolate the infection area, which only lead to approximately 20% nodes being infected at last.

On the whole, we can see that local edge removal strategies indeed helpful in delaying or controlling disease spread and some of them perform extremely well in network with community structure. Although in other network structures, they cannot isolate the infection area completely, delay function has been found in early steps of epidemic process. Therefore, we tend to study them in SIR model in next section hoping that disease can die out in early stage because of this delay phenomenon. Since it has been shown that more removed edges result in better performance, we also study the effect of remove fraction further in the following part.

4.1.2 Epidemic results on different remove fraction

In this part, we apply local edge removal strategies on four kind of networks using SIR model as epidemic model to see how remove fraction can influence epidemic process. Firstly in the

random network as revealed by Figure 18 (a) and (c), when remove fraction is enhanced, epidemic size and peak are decreased and disease will die out at 50% remove fraction. All strategies have similar performance in disease control except “Local Modularity” and this is due to its fewer total removed edges at same remove fraction compared with other strategies (as shown in (d)). Remove fraction is an up removal limit for each node. This means that if some strategies cannot find suitable edge to remove, they will not remove any more even up removal limit has not been reached. So we can see that “Local Modularity” strategy has stricter requirement in choosing candidate edges than others. With remove fraction increasing, for all strategies, the epidemic duration first ascends to be higher than no strategy situation and then continues to decrease to be much lower than no strategy situation. Purposive removal does not give improvement compared with completely random removal.

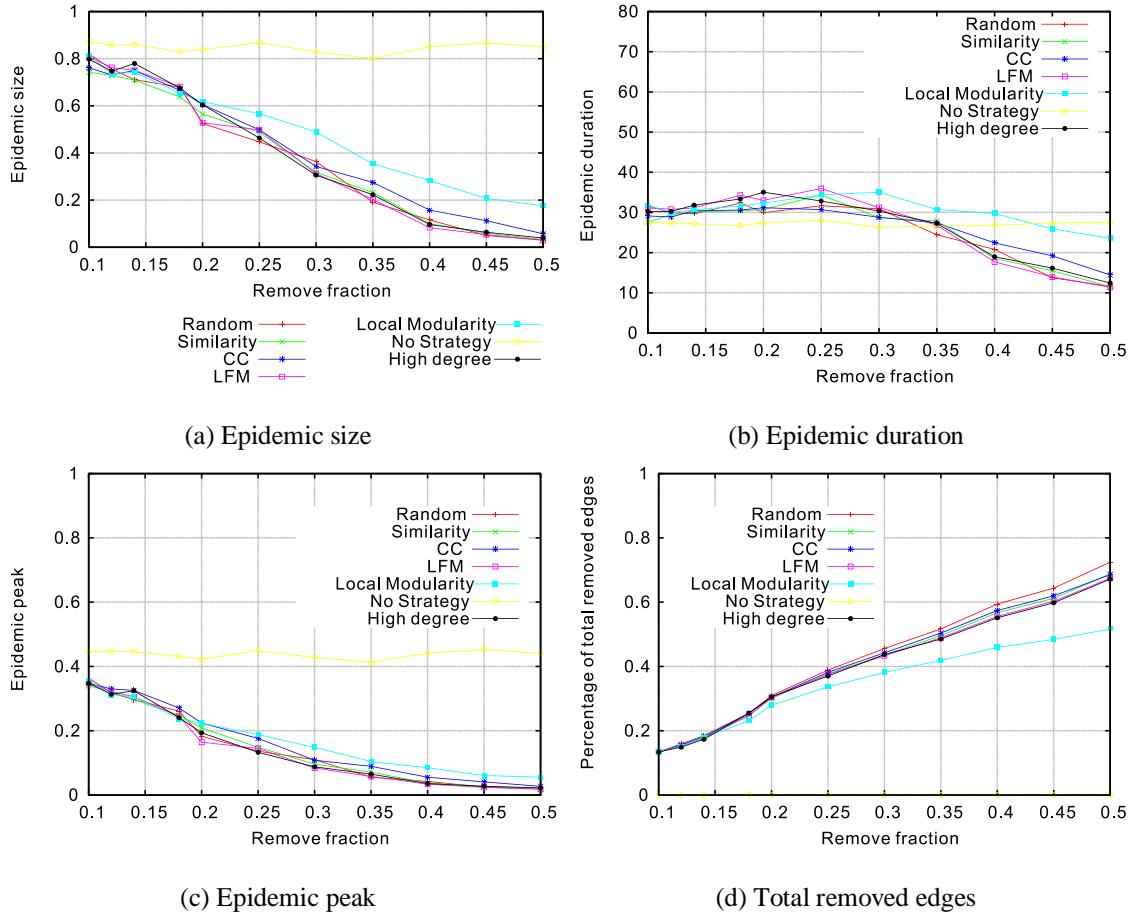


Figure 18: Random network in SIR model

The effect of remove fraction to epidemic size, peak, duration and total removed edges in exponential network resembles that in random network. The main different is that each strategy has different performance in exponential network. From Figure 19 (a) and (c), “LFM” and “High degree” have better performance than “similarity” and random removal when remove fraction is large enough and they totally remove fewer edges either shown in (d). Although “Local Modularity” and “CC” do not perform as well as other strategies, they also remove much fewer edges. Therefore it is hard to say whether “Local Modularity” and “CC” have bad or good control in epidemic. Additionally, it seems that strategies leading to lower epidemic size tend to have longer epidemic duration when remove fraction is not high enough, which is revealed from (a) and (b) when remove fraction is between 0.3 and 0.4.

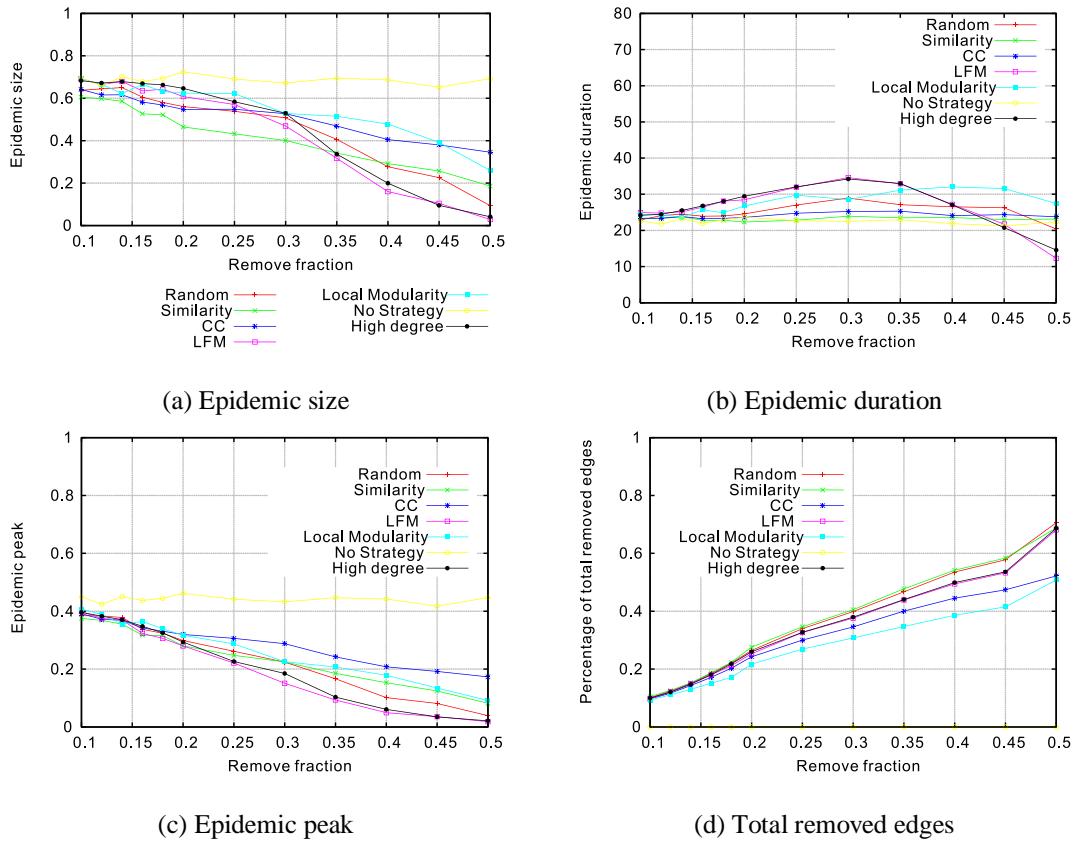


Figure 19: Exponential network in SIR model

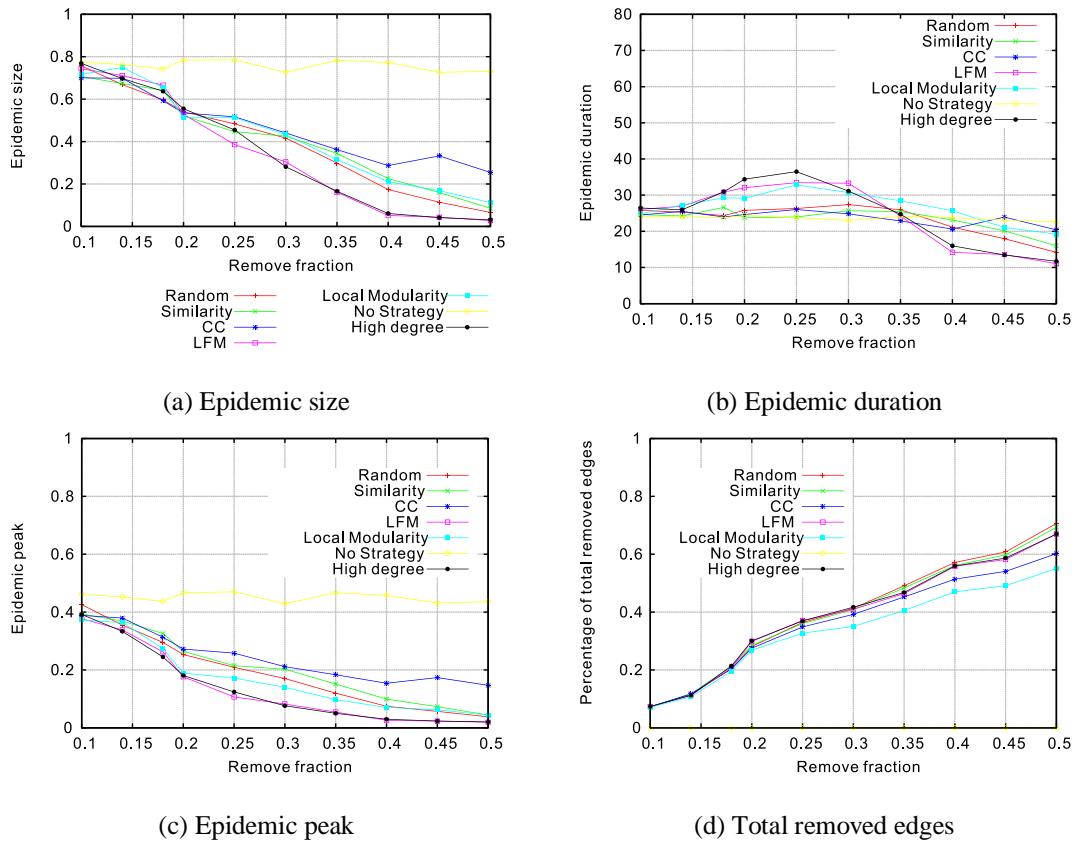


Figure 20: Scale-free network in SIR model

When it comes to scale-free network (see Figure 20), the situation is quite similar with the one in exponential network except that “Local Modularity” results in better epidemic control than “CC”, “Similarity” and “Random” removal for removing much fewer edges. “LFM” and “High degree” are still better than “Random” and “Similarity” since they remove same number of edges but have lower epidemic size and peak. “Local Modularity” also deserves study for it is the third best in disease control with fewest totally removed edges.

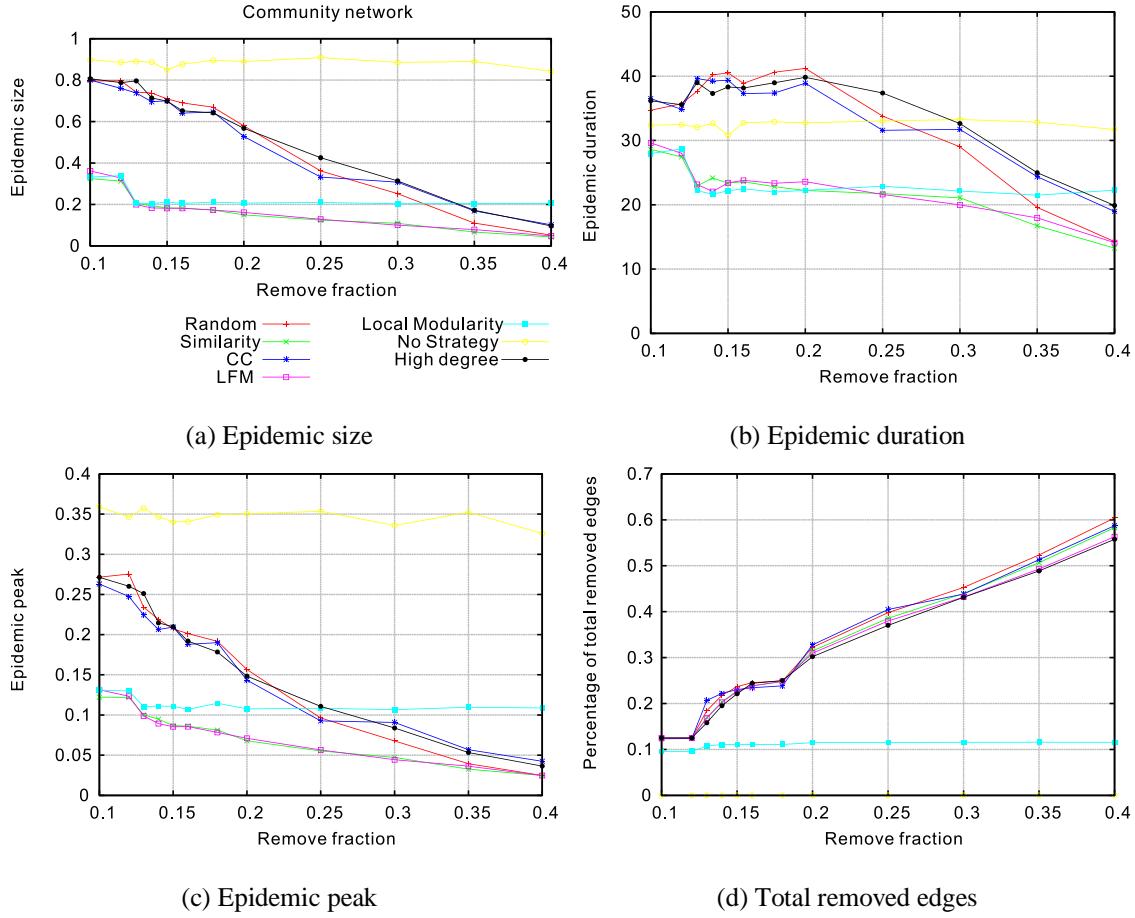


Figure 21: Community network in SIR model

As in section 4.1.1, when apply these local edge removal strategies in network with community structure, some strategies present extremely good effects in disease control. From Figure 21 (a), (b) and (c), we can see that “LFM”, “Local Modularity” and “Similarity” are better than any other strategies from low remove fraction. For these three strategies, from 10% to 13% remove fraction, epidemic size drops down quickly. Then from 13% remove fraction, epidemic size keeps stable for “Local Modularity” and decreases slowly for “LFM” and “Similarity” until disease dies out. We can see from (d) that “LFM” and “Similarity” removes equal edges with other strategies (except “Local Modularity”) but perform nearly twice better than other strategies (except “Local Modularity”) at 10% remove fraction. However, with remove fraction growing, this advantage becomes smaller until no advantage shown at 40% remove fraction. Same situation is also viewed in measure of epidemic duration and peak. The most special case come to “Local Modularity” strategy, which shows a threshold at 13% remove fraction. As shown in (d), even remove fraction keeps increasing, “Local Modularity” only removes little more edges between 10% and 13% remove fraction and then no more edges are removed further. That is also why epidemic size, peak and duration keep unchanged from 13% to 40% remove fraction. When compare “Local Modularity” with “LFM” and “Similarity” in (a) and (d), it can be seen that at 14% remove fraction, they lead to same epidemic size but “Local Modularity” removes approximately twice fewer edges than the other two. When remove fraction becomes 40%, although “LFM” and “Similarity” totally control the disease spread,

they remove six times edges as many as “Local Modularity”, which account for 60% of the total edges in network. Obviously, the more edges are removed the better the disease can be controlled since all nodes are isolated. So the reason that “LFM” and “Similarity” perform better than “Local Modularity” is because they remove too many edges. Therefore, we can conclude that “Local Modularity” should be the best strategy in community network because it removes fewer edges and reach a satisfying control result. Also it will not continue to remove more edges if all critical edges have been found, which means that it has a high accuracy compared with “LFM” and “Similarity” which find critical edges at the cost of much more removed edges.

In summary, except random network, local edge removal strategies could effectively delay and stop disease spread, especially in community network. “LFM” and “High degree” performed best in exponential and scale-free network when remove fraction was large enough while “Local Modularity”, “LFM” and “Similarity” showed best disease control in community network. The higher the remove fraction, the better the disease could be controlled. “Local Modularity” was different from all other strategies and it was more accurate in find critical edges. It separated the community by only removing fewer edges and would stop removing once community has been separated. Although it did not show best results in random, exponential and scale-free network, it might because it removed much fewer than other strategies at same circumstance. Therefore, “Local Modularity” was the best strategy so far for its excellent performance in community network.

4.1.3 Compare with global strategies

	Random	Exponential	Scale-free	Community
Similarity	259/859	180/650	155/542	169/1024
CC	261/859	*	*	*
LFM	*	*	*	169/1024
Local Modularity	*	*	*	109/1024
EdgeBetweenness	259/859	180/650	155/542	110/1024

Table 8: Total removed edges/total edges for each strategy in different network from Figure 22 (a) to (d)

Strategy	Random	Similarity	CC	LFM	Local Modularity	High degree	Edge Betweenness
Total removed edges	3390	3372	3324	3362	1548	3348	25%
	26%	25.8%	25.5%	25.7%	11.9%	25.7%	

Table 9: Total removed edges and percentage for each strategy in flight network

Table 8 shows the total removed edges for each strategy in different network from figure 20 (a) to (d). Only best strategies in section 4.1.1 are listed and the “Global EdgeBetweenness” strategy is set to remove same number of edges as best local edge removal strategy for comparison. Table 9 is the detail of total removed edges for each strategy in flight network. Since “Global EdgeBetweenness” is computing complexity demanding, we use the data from [Marcelino and Kaiser 2009] to make a comparison.

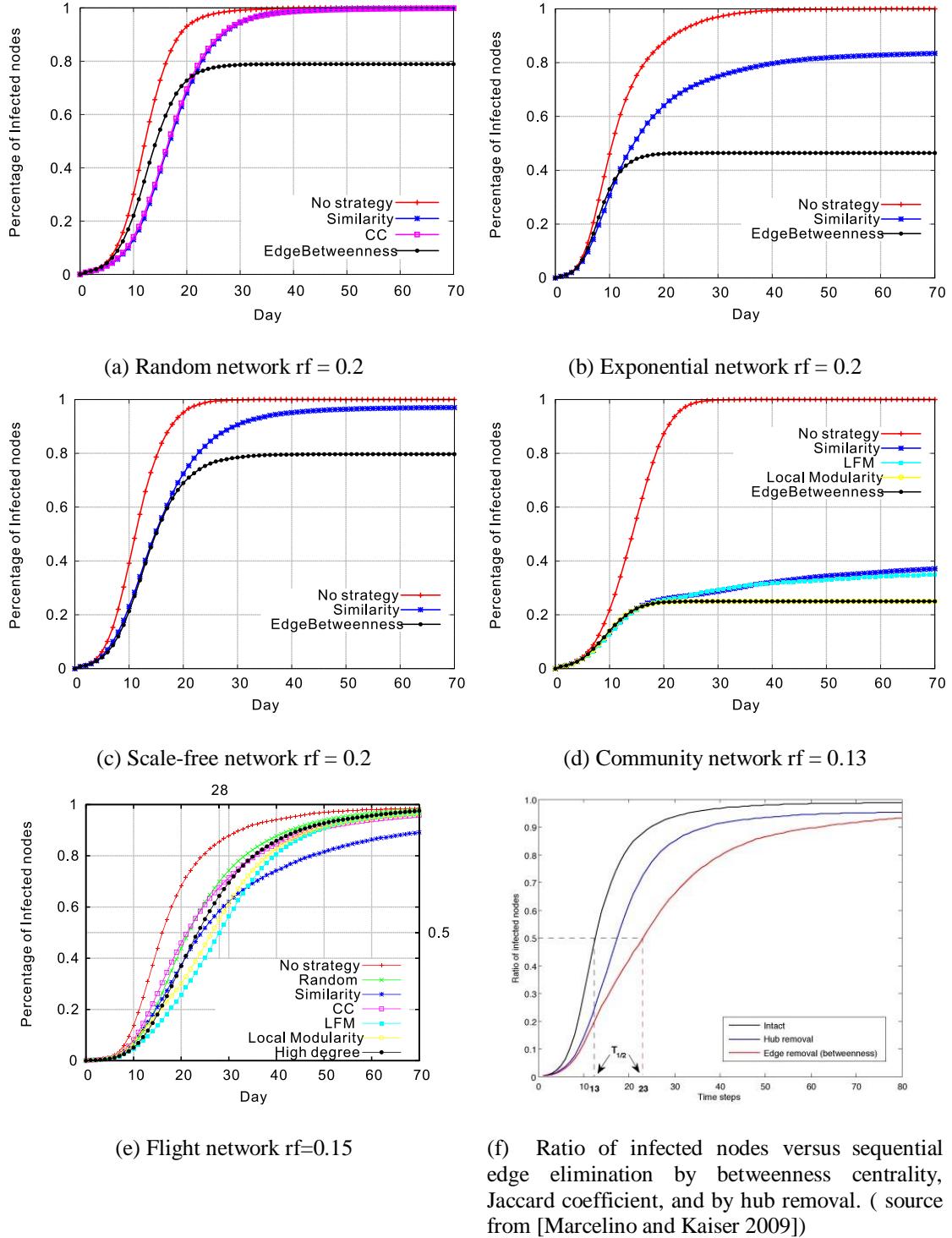
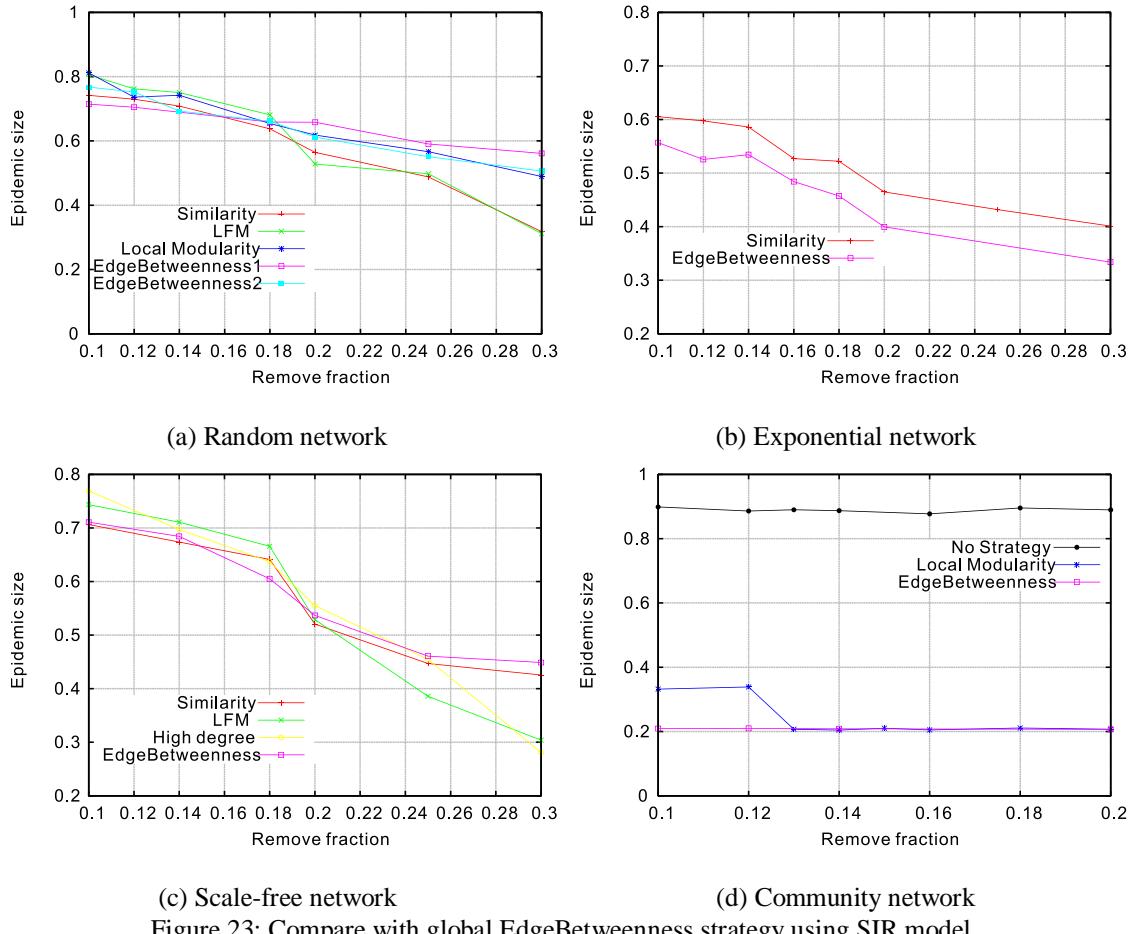


Figure 22: Compare with global EdgeBetweenness strategy using SI model in different network structure. rf stands for remove fraction.

As revealed by Figure 22 (a) to (c) and table 8, “EdgeBetweenness” leads to much fewer infected nodes at the end of epidemic than other strategies. It seems that “EdgeBetweenness” removes the most critical edges that connecting the different parts in the whole network so that disease has been kept in only one part of network. Even in networks without community structure, it still can separate the network while local edge removal strategy seems has no such ability. However, in the community network (d), we can find that “Local Modularity” performs as well as “EdgeBetweenness” with same number of removed edges. Although the local edge removal strategy have higher final infection ratio, in the early stage of epidemic process in (a)

and (b), local strategies delay the infection better than the global one and this will be studied further in flight network (see Figure 22 (e)).

Figure 22 (e) is the testing results in real global flight network. Paper [Marcelino and Kaiser 2009] found that removing high betweenness edges (cancel flight line) is better than removing hub nodes (shut down airport with highest centrality) in slowing down the disease spread (shown in (f)). “Edge betweenness” was regarded as superior at predicting the most critical edges. However, it is quite slow to calculate in larger network. From (f), the time step where half of the nodes are infected is 23 after 25% of edges are removed by “Edge betweenness” while in our study (Figure 22 (e)), the time step for “LFM” is 28 and “Local Modularity” is 26 applying the same condition. The final infection ratio is more than 90% for “Edge betweenness” in (e), which is same as “Similarity” in (f). This means that local edge removal strategy can delay disease spread at least as well as “global edge betweenness” only using local information and not be computationally as costly as “Edge betweenness”.



Since local edge removal strategies show better delay functions than “Global EdgeBetweenness” in early epidemic stage, we test them in SIR model further to see whether disease can die out early if we increase remove fraction for each node. As can be seen in Figure 23, we use SIR model epidemic model to compare the performance of local and global edge removal strategies in different remove fraction for each node. Only those good strategies in section 4.1.2 are used to compare with “Global EdgeBetweenness”. The details of total removed edges in different remove fraction are shown in table.10. In random network “EdgeBetweenness 1” is set to remove similar number of edges with “Similarity” and “LFM” strategies while “EdgeBetweenness 2” remove as many as “Local Modularity” algorithm. Similarly, in scale-free network, “EdgeBetweenness” is set to remove same amount as “LFM” and similar amount with “Similarity” and “High degree”. “Similarity” and “EdgeBetweenness” have equal total removed edges in exponential network, which is also same situation for “Local

Modularity” and “EdgeBetweenness” in community network. The detail of removed edges for exponential and community networks can be found in Figure 19.d and Figure 20.d respectively.

Remove fraction: (%)		0.1	0.12	0.14	0.18	0.2	0.25	0.3
Random	Similarity	13.4	15.5	18.4	24.8	30.1	37.8	43.5
	LFM	13.5	15.2	17.7	24.6	30.3	37.5	43.3
	EdgeBetweenness1	13.4	15.3	18	24.8	30.1	37.6	43.4
	Local Modularity	13.4	15.2	17.7	23.3	28	33.7	38
	EdgeBetweenness2	13.4	15.2	17.7	23.3	28	33.7	38
Scale-free	Similarity	7.3	*	11.7	21.3	28.6	36	40.8
	LFM	7.4	*	11.3	20.8	29.9	37.1	41
	High degree	7.4	*	11.3	21.4	30.1	37	41.7
	EdgeBetweenness	7.4	*	11.3	20.8	29.9	37.1	41
Exponential	Similarity	10.61	12.61	15.28	22.3	27.69	*	40.62
	EdgeBetweenness	10.61	12.61	15.28	22.3	27.69	*	40.62
Community	Local Modularity	9.77	9.77	10.93	11	11.52	*	*
	EdgeBetweenness	9.77	9.77	10.93	11	11.52	*	*

Table 10: The percentage of total removed edges for different strategies in different network structure in Figure 23. (show in %)

Figure 23 (a), (b) and (d) all demonstrate that “EdgeBetweenness” perform better than local strategies at low remove fraction, but when remove fraction large enough, local strategies will lead to same or lower epidemic size comparing “EdgeBetweenness”. On the other hand, “EdgeBetweenness” continues to have better epidemic control than “Similarity” no matter how remove fraction changes.

The result shown in section 4.1.2 is that “Local Modularity” is the best candidate local remove strategy in community network. So we are going to test it further with global strategy particularly in real network. Since “EdgeBetweenness” has high computational cost real world larger network, we use another global strategy called “InfoMap” to instead. Since it is a global strategy, it will not remove edges according to remove fraction. Firstly we run “InfoMap” to find all intercommunity edges and then get the epidemic result after removing all these edges as a reference. We vary the remove fraction for “Local Modularity” to compare with “InfoMap” both in SI and SIR model.

Firstly, we run the test in the two social networks: “Blogs” and “School” networks. From Figure 24 (a) and (b), “InfoMap” completely control the disease spread by removing all intercommunity edges which account for 18.6% of total edges. All other strategies perform better than random removal except that “CC” is incomparable because it removes fewest edges and has highest epidemic size. “LFM” and “Local Modularity” show smallest epidemic size but “Local Modularity” removes much fewer edges than “LFM”. So “Local Modularity” is still the best among all candidate local edge removal strategies. When compare it to global “InfoMap” strategy, in order to reach same epidemic size, “Local Modularity” needs to remove twice edges of “InfoMap” at 50% remove fraction. The “Blogs” network consists of 273 communities, so the size for each community is small. Thus when “InfoMap” isolates these communities, the disease dies out extremely quickly. Although “Local Modularity” cannot perform as well as “InfoMap”, but at 25% remove fraction where they remove same number of edges, the epidemic size for “Local Modularity” is also very low about 15%. When it comes to “School” network (see Figure 24 (c) and (d)) which shows scale-free property,

“InfoMap” plays approximately four times better than local edge removal strategies at same total removed edges (19% remove fraction). The performance of local edge removal strategies seems incomparable because the strategy leads to low epidemic size also removes more edges. The “school” network has low community strength since the intercommunity edges account for 32% of total edges. So the performance of local edge removal strategies is similar with that in theoretical scale-free network. (See Figure 20)

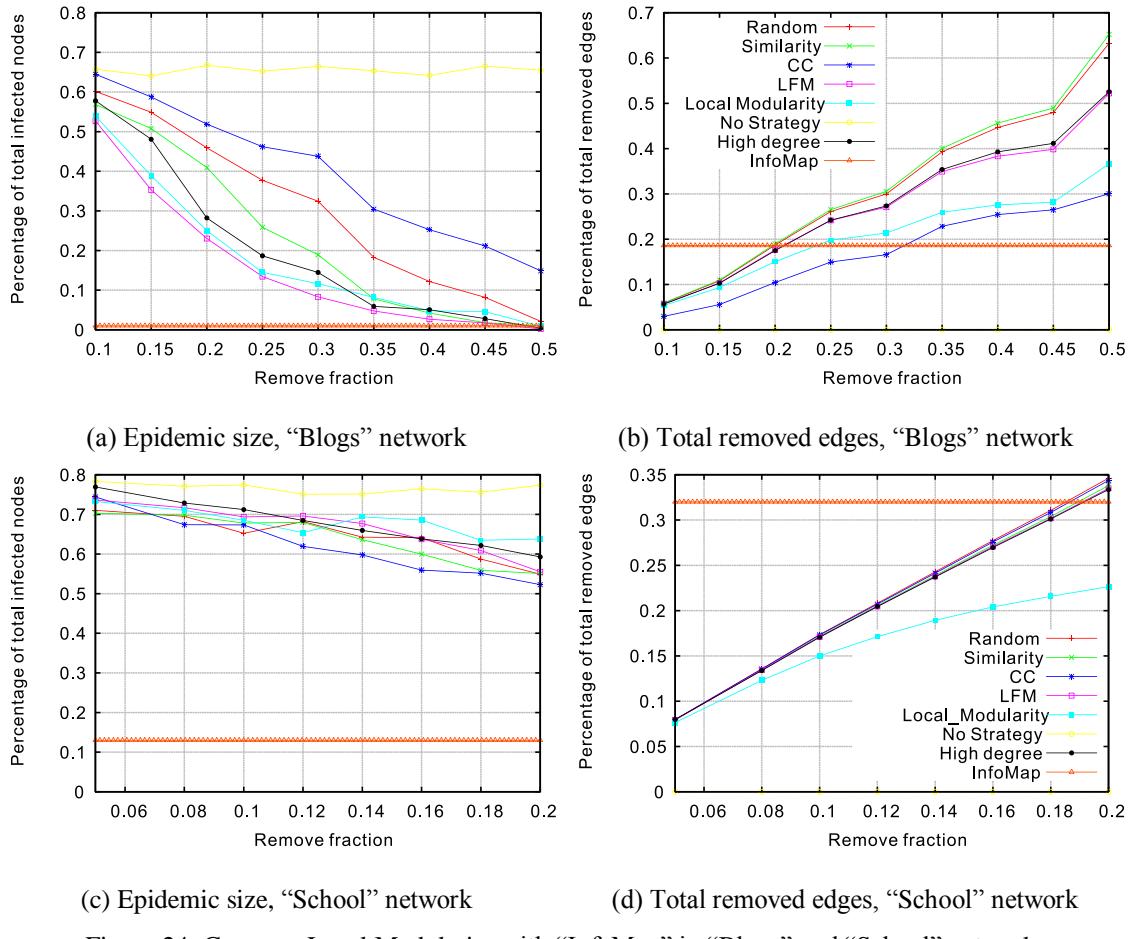


Figure 24: Compare Local Modularity with “InfoMap” in “Blogs” and “School” network

Secondly, we test it on real global flight network. Comparing “Local Modularity” with “InfoMap” in SI model (see Figure 25 a and b), “InfoMap” keeps the infection ratio at around 30% when disease spreads on the 70th day by removing lower than 15% edges while the best result of “Local Modularity” on 70th day is around four times worse by removing more than 17% edges. We can see that with the growth of remove fraction, the final number of infected nodes firstly drop and then remain unchanged. This is because that the total removed edges reach an upper limit when remove fraction is larger than 50% as shown in (b). Since “InfoMap” removes all intercommunity edges, we can assume that “Local Modularity” finds most of the intercommunity edges but not all of them. So the epidemic has been controlled but not as well as that in “InfoMap” due to that the communities in network have not been totally isolated. There is no recover in SI model, so all nodes will be infected if the network is connected. We just use SI model to test how well those strategies can isolate the communities. In order to accord with realistic epidemic process, we continue to compare them in SIR model below.

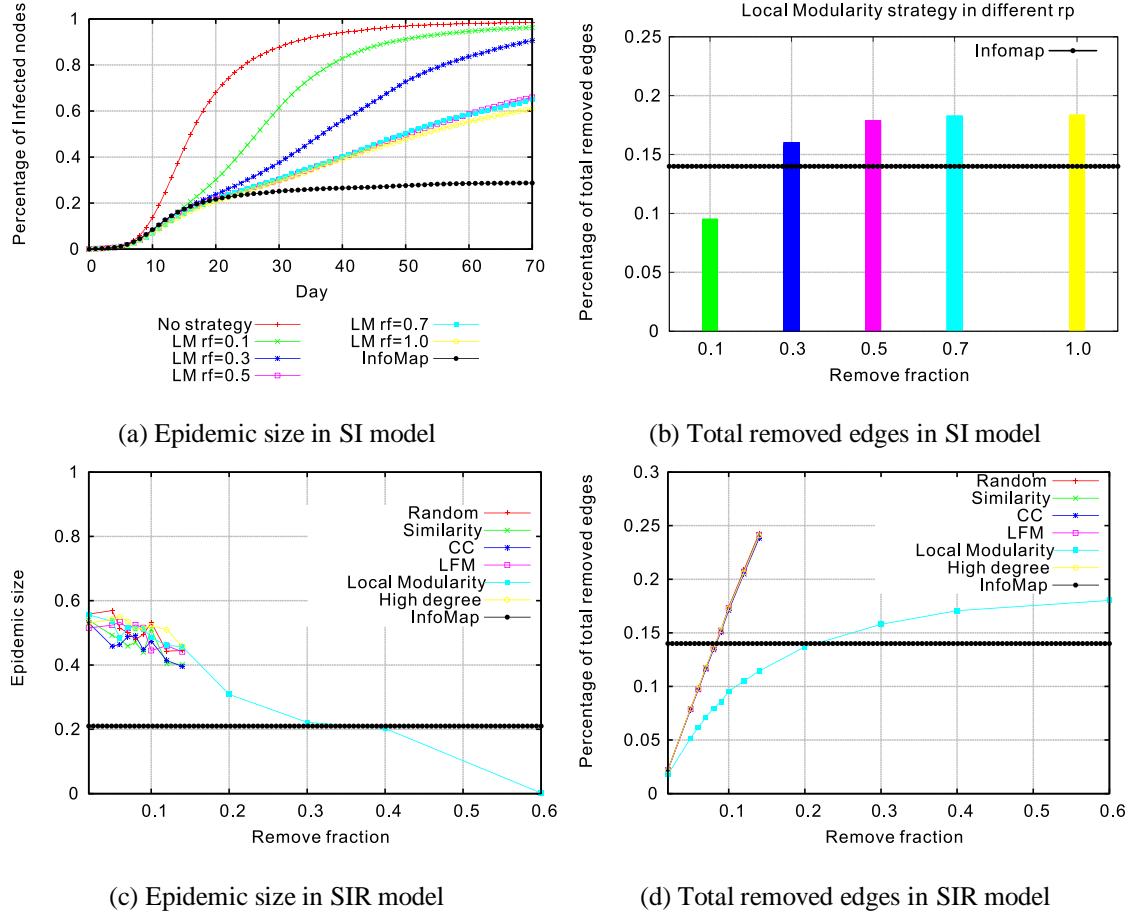


Figure 25: Compare Local Modularity with “InfoMap” in flight network

As in (b), in (d) we also first run “InfoMap” to remove all intercommunity edges and get the epidemic size in (c). When comparing “Local Modularity” with other local strategies, we can see that at 14% remove fraction, other local strategies already totally remove twice edges as many as “Local Modularity” but have similar epidemic size. Thus we only continue to increase remove fraction for “Local Modularity” to see whether it can reach same control results as “InfoMap”. With the increasing of remove fraction, the epidemic size for “Local Modularity” keeps decreasing and equals that in “InfoMap” between 30% and 40% remove fraction with only fewer than 2% more edges removing. Disease even dies out when “Local Modularity” removes 5% of total edges more than “InfoMap” at 60% remove fraction.

In conclusion, after comparing local strategies with global strategies, we found that local strategies could perform as well as global ones in SIR model in some networks if remove fraction was high enough. “Local Modularity” could even delay the early spreading of disease in community and flight network in SI model and reached a satisfying final infection ratio. Although “InfoMap” had perfect disease control in “real network”, “Local Modularity” showed a relative good control results in network with community structure either. Due to noisy or incomplete network information in real life, “Local Modularity” is more practice and robust to use in real life compared with “InfoMap” or “EdgeBetweenness”. Therefore, it is possible to only use local information and consume less computing cost to get the similar epidemic control as global strategies.

4.2 All population aware in different time steps

In the previous experiments, we only consider that the whole population becomes aware before epidemic is known (all strategies applied before disease start). Actually in real life, population may make behavioural changes after disease prevalence, so in this section, we vary the time

when strategies are applied to make a contrast.

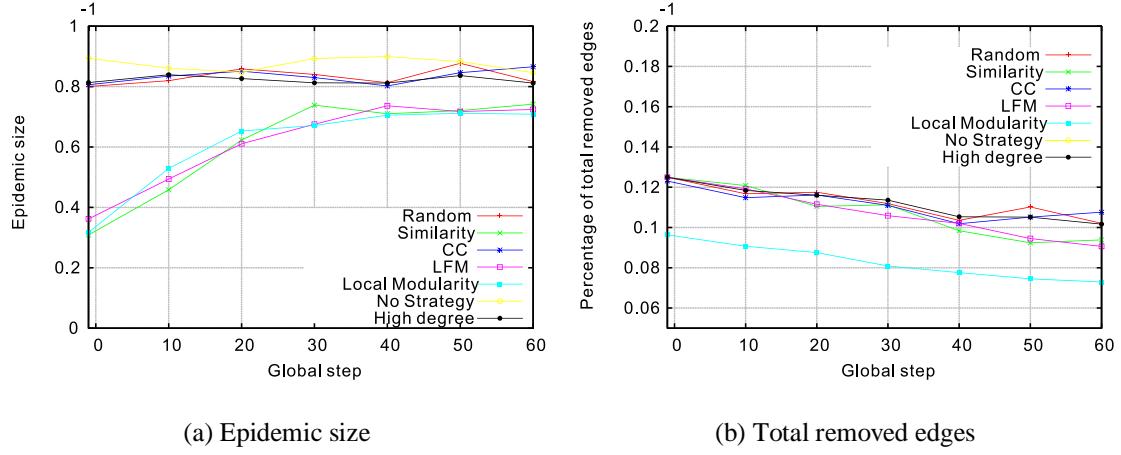


Figure 26: Community network in SIR model

As shown in Figure 26, the earlier the strategies are applied the lower the epidemic size will be for “Similarity”, “LFM” and “Local Modularity” strategies, but it will also cause more edges to be removed. For other strategies, although more edges are removed if all population becomes aware earlier, the performance does not show much difference in contrast to later applying. Therefore, it cannot conclude that it is better for epidemic control if all population conducts behavioural changes earlier, because more edges will be needed to remove also.

4.3 Behavioural changes happen during disease spread

4.3.1 Compare with strategies applied before epidemic process

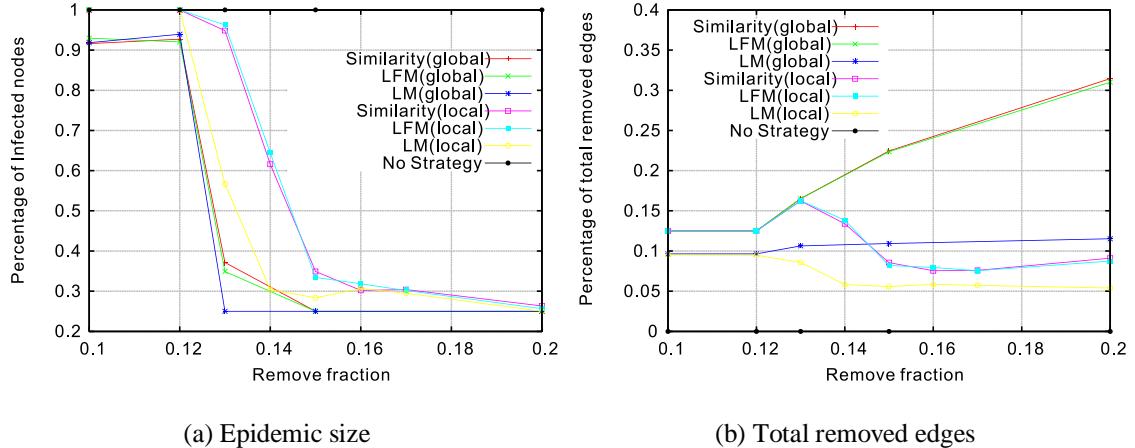


Figure 27: Compare global and local in community network using SI model (Local Modularity denote as LM)

After viewing the result in previous section, there is a strong possibility that local strategies may not need to be applied completely before disease start and still can result in good disease control. So we conduct an experiment to let node apply strategy only when it is infected (local) instead of before disease spreading (global). From Figure 27(a), we can find that strategies applied globally indeed have better performance than local ones at small remove fraction. However, when remove fraction increase, this superiority become descending and finally they get same performance at 20% remove fraction with local ones removing only half total removed edges of the global ones. It can be seen that strategies applied globally perform better in small remove fraction due to removing more edges and if remove fraction high enough, local

application can be equally effective with fewer edges removing. Although local application of strategies need higher remove fraction to get same performance as global ones, the total removed edges are not higher but even lower. So it seems that only individual in disease prevalence area to make behavioural changes would be enough for disease control if they could remove critical edges before infecting others. There may be no need for whole population to remove critical edges since disease may die out quickly if behavioural changes are effective enough.

4.3.2 Compare with local “SI link” strategy

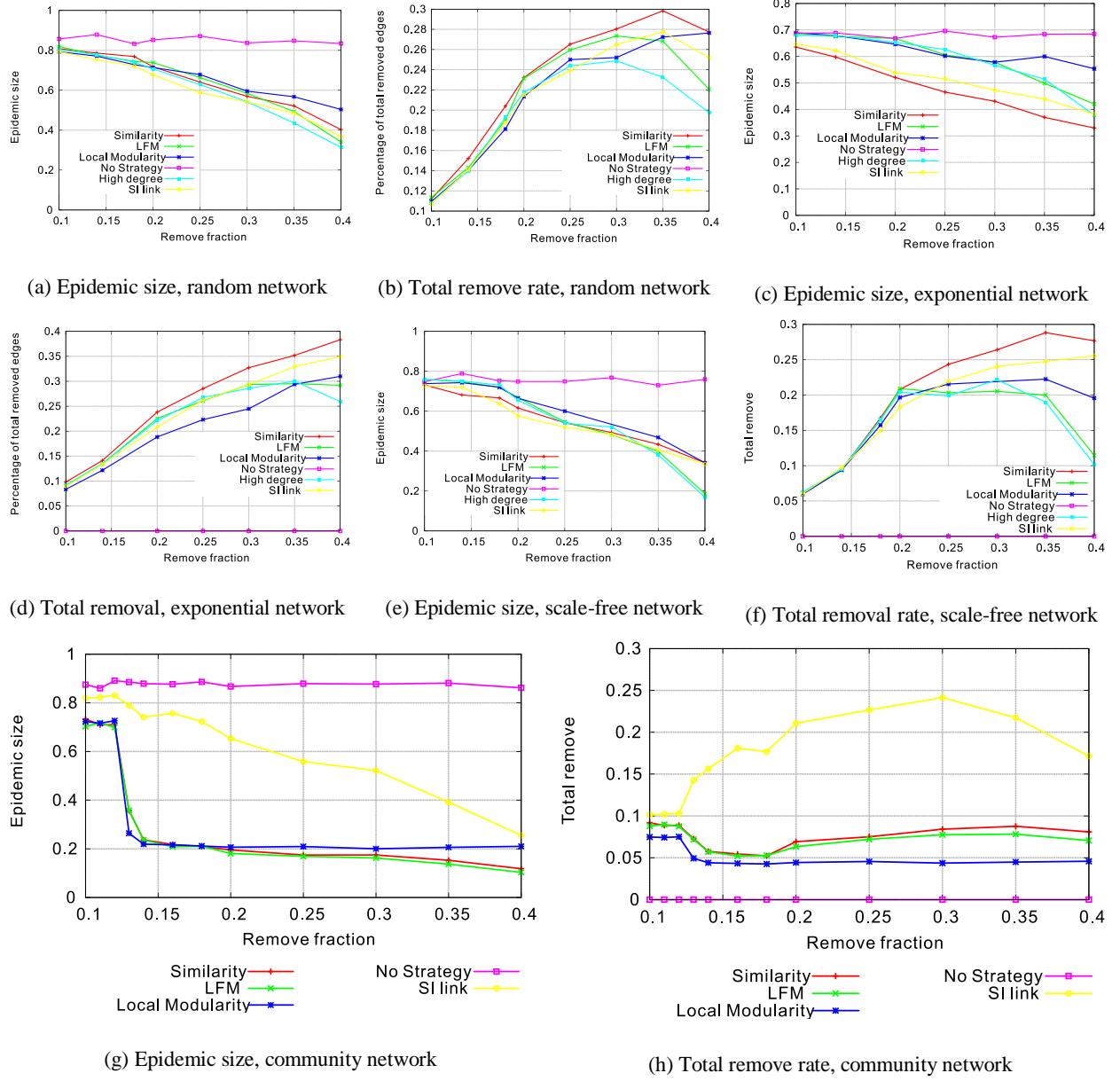


Figure 28: Compare with “SI link” in different network structure using SIR model

As knowing that local strategies applied with disease spread can be effective as well in previous experiments, we are going to compare it with the “SI link” strategy which is viewed as human instinct to reduce contact with infected people and can only be applied if neighbours’ epidemic states are revealed.

Firstly in random network as shown in Figure 28 (a) and (b), “SI link” has similar performance

with “High degree” strategy with similar number of total removed edges before 25% remove fraction and much more removal than “High degree” from 25% to 40% remove fraction. Secondly, in exponential network, “Local Modularity” is as effective as “High degree” and “LFM” if remove fraction is less than 30% but removes fewer edges totally. “Similarity” is better than “SI link” strategy due to more edge removal. Thirdly, in the scale-free network, “SI link” has same performance with “Similarity” by removing fewer edges. “High degree” and “LFM” have much better performance than other strategies when remove fraction is 40% by only removing fewer than half edges of other strategies. Finally, in community networks “Similarity”, “LFM” and “Local Modularity” have similar performance as tested in strategies applied globally in section 4.1.2 Figure 21(a), which are much better than “SI link”. Although “SI link” tend to perform better with remove fraction increasing, the total removed edges still much higher than other three strategies.

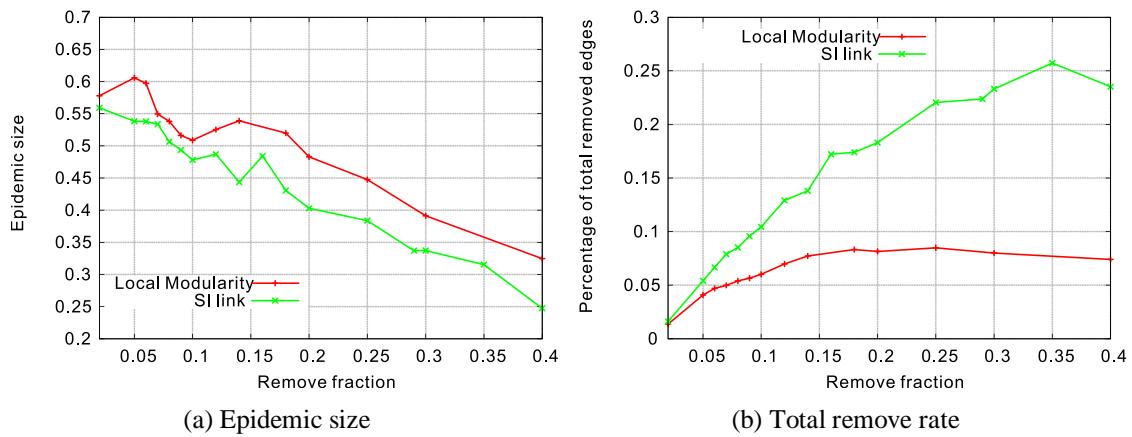


Figure 29: Compare “Local Modularity” and “SI link” in flight network in local spread model

Comparing “Local Modularity” with “SI link” in flight network (see Figure 29), we found that “SI link” has little lower epidemic size than “Local Modularity” at same remove fraction but the total removed edges are much higher. With the growth of remove fraction, “Local Modularity” keeps the total removed edges unchanged at last but “SI link” continue to remove more edges and even to remove twice edges of “Local Modularity” at 40% remove fraction. This is similar with the situation in section 4.1.3 where “Local Modularity” only removes the most critical edges while other strategies just remove as much as possible.

The overall results in this section indicated that strategies applied with disease spreading could be as effective as global one but removed much fewer edges. The difference from global experiment was that the number of total removed edges first increase and then decrease with remove fraction going up, which is because that disease dies out so quickly that only few nodes are influenced (In global experiment, all nodes will be influenced no matter there is a disease or not). “Local Modularity” was still the best strategy in community network but in other network structures it did not perform as well as it in global experiment. Because some of local strategies could suddenly die out the disease with remove fraction high enough, which leads to fewer nodes infected and thus fewer edges are removed. However, if the remove fraction is too high, it means that the disease dies out because infected nodes tend to remove all its edges instead of only critical edges are removed. So even some strategies could reach low epidemic size at high remove fraction, it does not mean that they are as well as those strategies reach same epidemic size with lower remove fraction. Like “Local Modularity” and “SI link” in (g), “SI link” reaches 20% epidemic size at 40% remove fraction while “Local Modularity” only need 13% remove fraction to get same performance. “Local Modularity” was showed much better than “SI link” because it removed all critical edges but “SI link” only isolated infected nodes. Therefore, according to all above results, we could say that some of local edge remove strategies are more effective than “SI link” because they are epidemic status independent and focus on critical edges rather than infected nodes.

4.4 The effect of awareness spreading

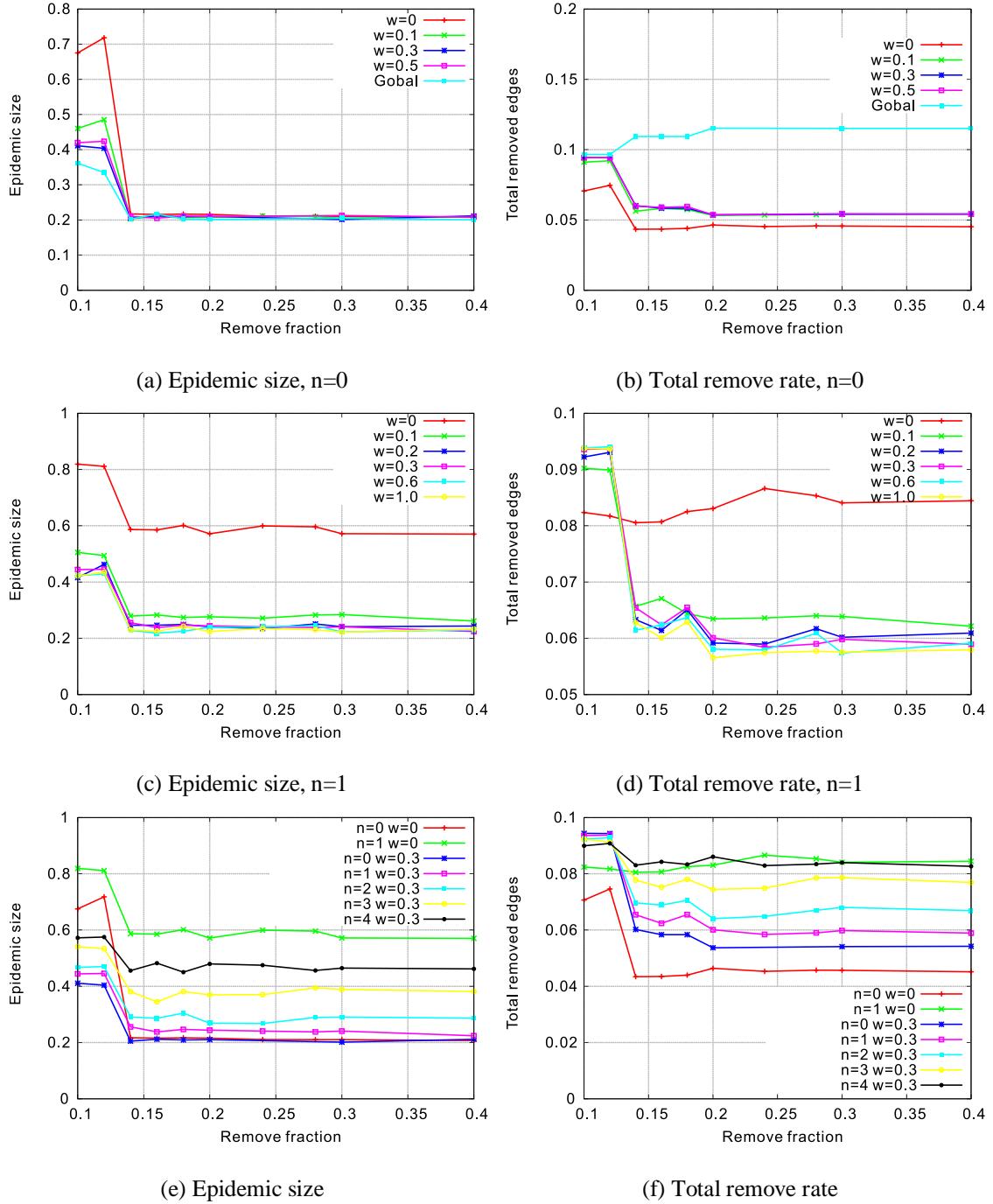


Figure 30: The effect of awareness spreading (using SIIR model), where n is the duration of asymptomatic infectious state and w is the awareness spreading rate.

By far, we have tried all population become aware based on publicly available information when there is a disease (Section 4.2) or there is a belief of disease (Section 4.1). We also have tried individual accepts behavioural changes based on local information when disease prevails (Section 4.3). Since our local edge removal strategy is epidemic independent and the earlier all population becomes aware the better the disease will be controlled, it is worth to try that individual takes behavioural changes not only by disease prevalence but also by awareness of disease. So in this section, we set all infected nodes become aware and awareness will spread simultaneously with disease spread. Once a node becomes aware, it will apply the edge removal strategy. In section 4.3, we set that all infected nodes will first apply edge removal strategy

and then start infecting other nodes, which may not quite realistic in real life since some of infections are asymptomatic. So we add an asymptomatic infectious state in SIR model and nodes in this state will not become aware but still be infectious.

Firstly, the duration of asymptomatic duration is set to be 0, which turns SIIR model into SIR model (see Figure 30 (a) and (b)). It can be seen that before the point of 14% remove fraction, the faster the awareness spreads the lower the epidemic size will be and the epidemic size tends to be as low as that in global test but only removes half edges of global one. There is a large different performance between awareness spread and no awareness spread, but once awareness spread, high spread speed does not show much better performance and removes same number of edges as low spread speed. However, when remove fraction is larger than 14%, epidemic size is same for all situations and the strategy of no awareness spread removes fewest edges.

When duration of asymptomatic infectious is 1 day long as shown in Figure 30 (c) and (d), there is a quit large difference in epidemic size between having awareness spread and no awareness spread. The epidemic size when there is awareness spread is only half of that in no awareness spread and the total removed edges are 2% fewer except when remove fraction is lower than 14%. As same as (a) there is no much difference in different awareness spread speed.

From (e) and (f), we can see that the longer the asymptomatic infectious state, the higher the epidemic size will be. But if there is an awareness spread, although the asymptomatic infectious state is long, it is better than no awareness spread in a short asymptomatic infectious state. Because black line and green line remove same number of edges but black line has lower epidemic size. When asymptomatic infectious state is 1 day, the awareness spread in 0.3 rate leads to three times better performance than no awareness spread and nearly as good as the performance in no asymptomatic infectious state. “Local Modularity” performs best in no asymptomatic infectious state and no awareness spread for it has the best epidemic control with lowest removed edges.

In summary, “Local Modularity” performed best when infected nodes became aware immediately after being infected (asymptomatic infectious state). If nodes could apply remove strategies in time before infecting other nodes, there is no need for awareness spreading when remove fraction is large enough for “Local Modularity” to find all intercommunity edges. But in low remove fraction, higher awareness spread could help to get similar good performance as all population applying strategy before disease start with much fewer removed edges, because it is applied locally. However, when there is asymptomatic infectious state, the longer the state, the worse the disease control will be. Increasing awareness spread speed can help improve the performance of epidemic control, but if the asymptomatic infectious state is too long, it will be too late to make behavioural changes.

4.5 Effects of network properties analysis

To understand the underlying mechanism of these results we test the local edge removal strategies on three different network properties: clustering coefficient, power-law degree distribution and community strength to see which one has the significant influence to the performance of these strategies.

4.5.1 Clustering coefficient

We test “Local Modularity” on five exponential networks with different clustering coefficient (from 0.1 to 0.5, see Figure 31 (a) and (b)). “Local Modularity” removes similar number of edges in different clustering coefficients and the results of epidemic size are also similar in small remove fraction less than 30%. From 30% remove fraction to 50% remove fraction, although the network with higher clustering coefficient tends to show smaller epidemic size,

the ability of epidemic control for “Local Modularity” is not proportional to clustering coefficient since it performs better in clustering coefficient value of 0.3 than 0.4. So clustering coefficient affect a little to the performance of local edge removal strategies.

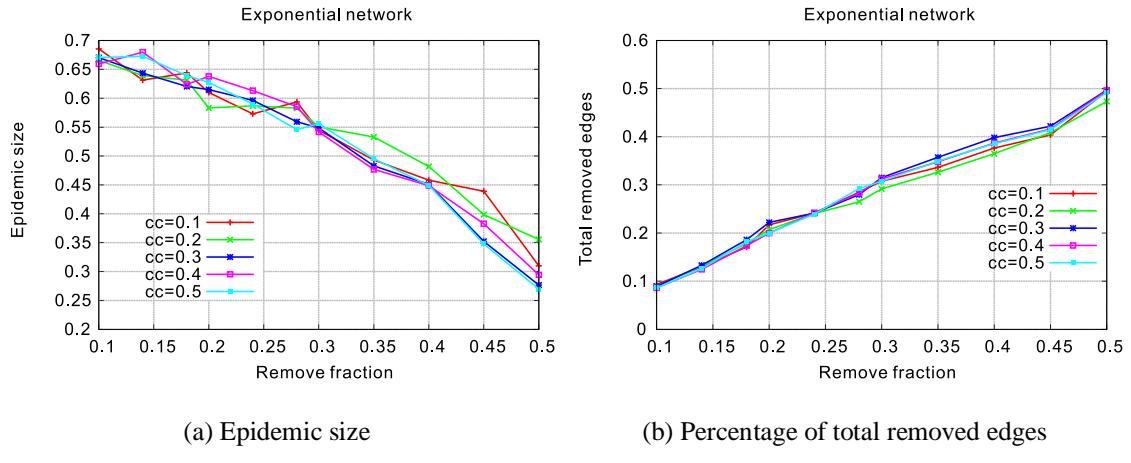


Figure 31: Exponential networks in different clustering coefficient

4.5.2 Degree distribution

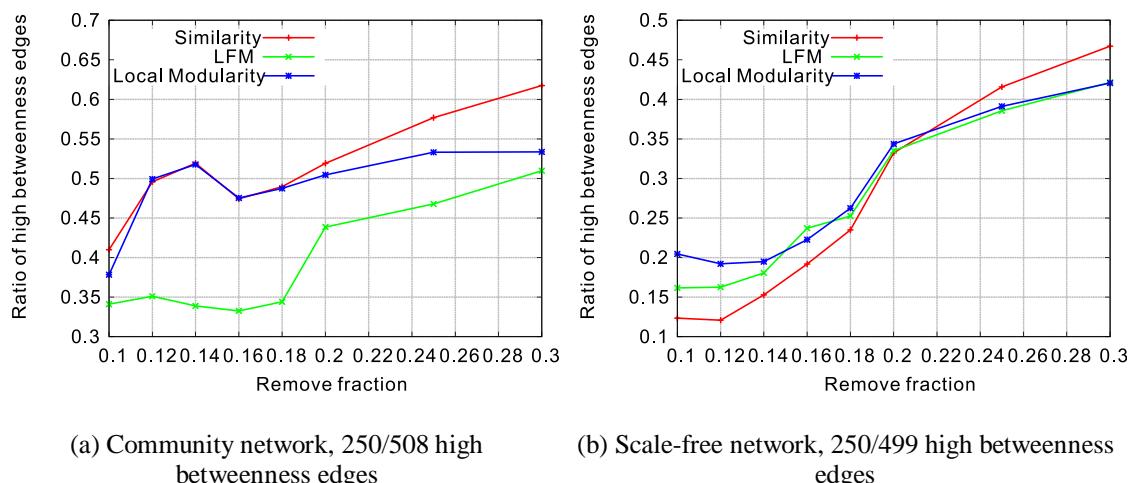


Figure 32: Show the percentage of removed edges that are high betweenness edges in different remove fractions. Scale-free and community network both with power law degree distribution

As observed in previous results, “Local Modularity”, “Similarity” and “LFM” prevents disease spreading most effectively in community network. Since our generated community network has power law degree distribution. We are going to test whether such good result is due to community structure or power law degree distribution. Firstly we calculated the 250 highest betweenness edges in community network and 250 in scale-free network. Then we record that for each strategy at each remove fraction, what percentage of total removed edges are among these high betweenness edges.

From Figure 32 (a) in community network, more than 50% of total removed edges for “Local Modularity” and “Similarity” at 12% remove fraction are critical edges compared with about 20% for “Local Modularity” in scale-free network. With remove fraction growing, this percent drops a little first and then increases a lot for “Similarity” and a little for “Local Modularity” while for “LFM”, the increase becomes rapidly after 18% remove fraction. The drop is because total removed edges increase a lot but the found critical edges only increase a little. However, in scale-free network Figure 32 (b), although the rate of critical edges increases with

removing fraction, accuracy in it is always smaller than that in community network at each remove fraction, especially at low remove fraction. To get the nearly 50% accuracy, the strategies in scale-free network needs approximately three times remove fraction more than that in community network (for “Similarity”). Therefore, power law degree distribution seems help little for local edge removal strategies to find the most critical edges.

4.5.3 Mixing parameter

From above results, we may guess that only community structure affect the local edge removal strategies mostly. So we vary the mixing parameter which stands for the rate of intercommunity edges among all edges to do test.

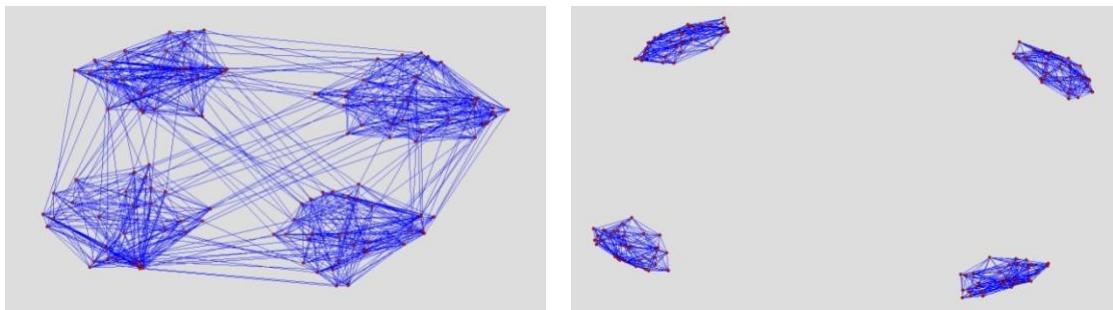


Figure 33: An illustration of community network before and after applying “Local Modularity”

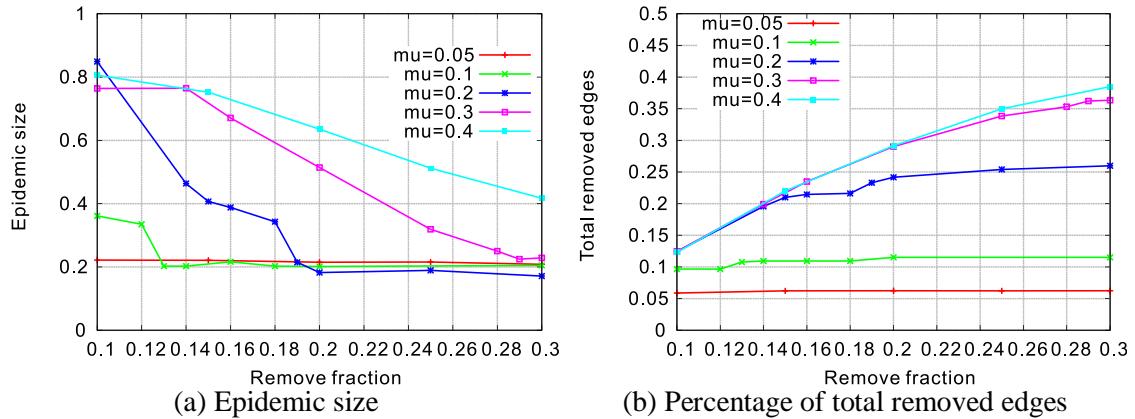


Figure 34: Local Modularity applied in community networks with different μ

In order to get 20% epidemic size, the smaller the mixing parameter (μ), the lower the remove fraction is needed and also the smaller the total removed edges will be (see Figure 34 (a), (b)). This means that high community strength and clear community structure help “Local Modularity” to find critical edges at the cost of lower remove fraction and also fewer total removed edges.

On the whole, we found that it was community structure rather than clustering coefficient or power law degree distribution that was critical in the performance of local edge removal strategies. The higher the community strength, the easier and more accurate would be to find all intercommunity edges. Figure 33 shows the community network with $\mu=0.1$ before and after applying the “Local Modularity” algorithm when remove fraction is 0.3. It can be found that all intercommunity edges have been removed accurately so that epidemic could be kept only in one community.

5 Discussion

In this section, we will summarize the observed results and provide our most important findings.

Firstly, we found that it is possible to only use local information to find most critical edges in the whole networks. These critical edges were usually the intercommunity edges in the networks with community structure or the highest betweenness edges in our study. The local edge removal strategies could both slow down the disease spreading and result in much fewer infected individuals as well as the global edge removal strategies if the remove fraction for each node is large enough, but the total number of removed edges were smaller. One of the advantages of local edge removal strategy was that each individual could find its own critical edges in the network by only knowing neighbours' information and it was found fast in computing compared with global edge betweenness strategy, which is quite suitable for guiding individual's behavioural changes since that individual is usually not available to whole population information and has no ability to conduct complex computing. Another advantage was that local edge removal strategies could be applied during epidemic process only by infected nodes, which means that only individual in infection area make behavioural changes promptly is enough for disease control. Strategies being applied locally could be as effective as being applied before epidemic start. Although remove fraction was larger, but the total removed edges were much fewer. Based on our review of global strategies in background section, we know that global strategy is hard to be applied in real life for the unknown topology of network where disease spread and for the huge computational complexity of some global strategies (like global edge betweenness) in large network. Also individuals tend to remove edges individually in real life. Therefore, it is quite meaningful and practical in real life to find such a disease control strategy only using local information.

Secondly, some local community detection algorithm could be used reversely to find edges of each node which were not in its local community. In our testing, we found that “Local Modularity” modified from [Clauset 2005] was the most effective algorithm in finding intercommunity edges. Because it could find all intercommunity edges in community network at the cost of removing small number of total removed edges compared with other local edge removal strategies and nearly remove as many edges as global strategy to get the same control results. The most important thing was that it will not continue to remove more edges if all intercommunity edges have been found even though remove fraction for each node continue to increase. “Similarity” and “LFM” also had good delay and control of epidemic in community networks although they removed more edges compared to “Local Modularity”. In random network, all strategies showed same performance as random removal, which means that it is hard to find critical edges in random network since all edges are generated randomly. “LFM” and “High degree” were found relatively good in reducing total infected individuals and slowing down the epidemic in scale-free network but the advantages were not so distinct. In the exponential network, “Similarity” was better than others but also removing relatively more edges. “Local Modularity” always removed the fewest edges in all network structure but only performed best in community structure, so it is hard to judge its performance in other network structures. After all, our results demonstrated that “Local Modularity” was the best local edge removal strategy to slow down and control epidemic, especially in the network with community structure.

Thirdly, like the results in paper [Marcelino and Kaiser 2009], our results also showed that community structure was important in disease control since network could be separated more easily than in other network structures. Additionally, we found that local edge removal strategies performed best in network with community structure and the higher the community strength, the easier to find intercommunity edges. “Local Modularity” performed well in finding intercommunity edges was because of community structure rather than clustering

coefficient or power-law degree distribution.

Fourthly, our local edge removal strategies performed better than the strategy that randomly removes edges to infected individuals in all kinds of networks especially community network, because we focused on removing most critical edges rather than randomly removal. So reducing contact with people in other community was found more helpful than only reducing contact with infected people in own community with regard to disease control in a global view since communities could be separated by purposive removal. Additionally, from the realistic view, it is usually hard to distinguish infected nodes at early epidemic stage, so the strategy removes from infected node is less feasible than local edge removal strategy which is independent of epidemic status.

Finally, we did some exploration in awareness spreading, although it was not systematic enough, it could be viewed as an inspiration in awareness study for future works. It is known that individual's behavioural changes are the results of awareness of epidemic information. We regarded the spread of behavioural changes as a spread of awareness among people, which is in a separate way from the disease spread. This kind of awareness could be gained globally from publicly information or locally from neighbours. Thus awareness could spread among neighbours just like disease. Since local edge removal strategy is independent of disease prevalence, it could be applied any time due to being affected by awareness. In our first conclusion, we knew that local edge removal strategy applied by infected nodes could be equally effective as being applied by all nodes before epidemic start but needed higher remove fraction. However, if we allow awareness spreading, we found that locally applying reach a similar epidemic size as globally applying at same remove fraction. Although the total removed edges were more than no awareness spreading one but were still much smaller than that in globally applying. It seems that by allowing awareness spreading, we converted the locally applying to a globally applying in a small range area rather than the whole population. Another finding about awareness was that when the infected individual cannot make behavioural changes promptly before it infected others due to asymptomatic infection, increasing awareness spreading rate helped but only limited to small asymptomatic delay. For disease with long asymptomatic duration, allowing awareness spreading helped little in disease control unless the spreading rate was set to 1. Otherwise, the path of awareness spreading might not be same as disease spreading, which means it was too late to make behavioural changes.

Our current study test different strategies and different remove fractions in different network structures using four kinds of disease-behaviour models. It means that we have a large number of scenarios to test. Therefore, several simplifications have been conducted and could be improved in future studies. Firstly, we run each experiment 150 times on theoretical networks and 300 on real network. Since the epidemic start is randomly chosen each time and the real network has number of total nodes more than 300, it would be better to repeat more times to see the results in flight network. Also for theoretical networks, we only use one instance for each theoretical network structure while it would be better to use different instances at each run. Secondly, when testing the effect of remove fraction, we choose the value of remove fraction roughly. It might be more accurate if we can increase the density of testing value, but it will cost much more running time at the same time.

In conclusion, our results showed that it is possible to guide individual's behavioural changes to slow down and control disease spreading efficiently in network with community structure and potentially other networks. The effect of behavioural change to epidemic is worth study further since we have observed quite optimistic performances of local edge removal strategies.

6 Evaluation

The first objective of this project is to build a disease-behaviour model and simulate it. This was completed by designing four types of behavioural change models and three types of epidemic models. The models have taken all causes of behavioural changes into consideration and allowed different combinations between behaviour and disease models. This objective has been archived successfully as the implemented simulation enables all kinds of experiment combination by varying test parameters and records all results in txt files for analysis later. We also implemented some network generation and rewiring algorithms discussed in methods section. Since our project is not aimed at software development, our code is only for doing experiment and there is no user interface. But the node and edge were all implemented as class so that it could be extended further for other network studies.

Based on the success complete of first objective, the second objective is to design epidemic control strategies related to behavioural changes. We have designed five local edge removal strategies according to existing resources and also implement three existing strategies that are used as contracts. In order to test their performance, we measured epidemic results, total removed edges and proportion of critical edges in removed edges in different network structures. Our experiment has showed obvious different epidemic control results for different strategies and found that “Local Modularity” had superior performance in disease control. We have successfully proofed that it is possible to use local information to detect critical edges. All research questions at beginning have been addressed to some level using experiment results and some even lead to further research questions (such as the effect of awareness spread), which means that we have reached the second objective to find desired strategies related to behavioural changes with good control of epidemic. Although the values selected in experiments were not intensive enough due to too many combinations, the general trend in each plot we got was enough for evaluating strategies’ performance. What we have done was to try various possibilities as much as possible and focused on good results to do relatively further testing. If we have more times, we could concentrate on the best strategy to do more detailed experiment and much further analysis about underlying mechanism. Since we have not got the real network dataset that could represent where the disease spread in real life, we only used the real global flight network to do experiments. It is better to do test on real life network but it is hard to find suitable data which has been discussed in section 2.1.1.

The third objective is to analysis designed strategies. We have visualized all results data into plots and compared them with existing global and local strategies. We also used a lot of plots to show how behavioural changes affect disease spread and network structure. For each plot, we evaluated the results according to the context and explained each phenomenon considering practical significance. So the third objective can be considered success.

The key contribution of our project is to build a new model to study behavioural changes and explore a new direction to help epidemic control (find critical edges locally), which can be used not only in epidemic studies but also in network theory studies.

7 Future works

The good results of local edge removal strategies highlighted the potential further exploitation in the study of behavioural changes.

7.1 Different spreading rate

For the simplicity, our experiment only set the transmission rate to be calculated under the condition when the basic reproduction ratio $R_0=3$. Since different disease has different epidemic parameter. Our project can be extended to study a particular disease or explore the distribution under all kinds of epidemic possibilities. It is interesting to see the local edge removal strategies' performance under various circumstances.

7.2 Rewire the removed edges

As review in background, many existing studies of behaviour-disease relationships not only remove the edges but also add new connections. There are two kinds' consequences of behavioural changes: change parameter and change network structure. We only studied the change of edges and have not considered the change of network structure which is usually results from rewire behaviour. The rewire can be integrated in local edge removal strategies easily by replacing removing to rewire. Studying the effect of local edge removal strategies on network structure might be useful in understanding the underlying mechanism.

7.3 Test on other epidemic networks

Since the limit of epidemic data source, we only tested the local edge removal strategy in theoretical networks, real global flight network and two empirical social networks. It is worth being tried in real epidemic network since it is not sure whether the epidemic network has community structure or not. Also it is worth finding out whether such local edge removal strategies is realistic in real life and whether or to what extent individual will make behavioural changes based on such strategies.

7.4 Awareness spread in different network

We only study the awareness spread in early stage. It can be extended further like building more realistic and advanced awareness spread model and combine it with local edge removal strategies. Additionally, the awareness can spread in a different network from where disease spreading. Because sometimes individual gets awareness from a virtual network (like internet) and actually contacts others in real life social network. The awareness can be gathered from virtual network and the local edge removal strategies will be applied in real life network. Then the behaviour-disease model will be more complex and thus more dynamic results will be observed, which might be quite interesting.

8 References

- [Bagnoli and Sguanci 2007] Bagnoli, F., Liò, P. and Sguanci, L., 2007. “Risk perception in epidemic modeling”, *Phys. Rev. E* **76**, 061904–7. (doi:10.1103/PhysRevE.76.061904)
- [Bailey 1975] N. T. J. Bailey, 1975. The Mathematical Theory of Infectious Diseases and its Applications, *Hafner Press, New York*.
- [Bansal et al. 2007] Bansal, S., Grenfell, B. T. and Meyers, L. A., 2007. “When individual behaviour matters: homogeneous and network models in epidemiology”, *J. R. Soc. Interface* **4**, 879-891. (doi:10.1098/rsif.2007.1100)
- [Bansal et al. 2009] Bansal, S., Khandelwal, S. and Meyers, L. A., 2009. “Exploring biological network structure with clustered random networks”, *BMC Bioinformatics* **10**, 405. (doi:10.1186/1471-2105-10-405)
- [Barabási and Albert 1999] Barabási, A.L. and Albert, R., 1999. “Emergence of scaling in random networks”, *Science* **286**, 509-512. (doi:10.1126/science.286.5439.509)
- [Batagelj and Brandes 2005] Batagelj, V. and Brandes, U., 2005. “Efficient generation of large random networks”, *Pre.* **71** (3), 036113-+.
- [Burgess et al. 2011] Simon Burgess, Eleanor Sanderson and Marcela Umaña-Aponte, 2011. “School ties: An analysis of homophily in an adolescent friendship network”, *CMPO Working Paper Series* No. 11/267.
- [Clauset 2005] Clauset, A., 2005. “Finding local community structure in networks”, *Phys. Rev. E* **72**, 026132. (doi:10.1103/PhysRevE.72.026132)
- [Diekmann et al. 1990] Diekmann O, Heesterbeek JA, Metz JA, 1990. “On the definition and the computation of the basic reproduction ratio R_0 in models for infectious diseases in heterogeneous populations”, *J Math Biol* **28**, 365–82.
- [Epstein et al. 2008] Epstein, J. M., Parker, J., Cummings, D. and Hammond, R. A., 2008. “Coupled contagion dynamics of fear and disease: mathematical and computational explorations”, *PLoS ONE* **3**, e3955. (doi:10.1371/journal.pone.0003955)
- [Funk et al. 2009] Funk, S., Gilad, E., Watkins, C. and Jansen, V. A. A., 2009. “The spread of awareness and its impact on epidemic outbreaks”, *P. Natl. Acad. Sci. USA* **106**, 6872-6877. (doi:10.1073/pnas.0810762106)
- [Funk et al. 2010] Funk, S., Gilad, E. and Jansen, V. A. A., 2010. “Endemic disease, awareness, and local behavioural response”, *J. Theor. Biol.* **264**, 501-509. (doi:10.1016/j.jtbi.2010.02.032)
- [Funk et al. 2010] Funk, S., Salathé M. and Jansen, V. A. A., 2010. “Modelling the influence of human behaviour on the spread of infectious diseases: a review”, *J. R. Soc. Interface* **7**, 1247-1256. (doi:10.1098/rsif.2010.0142)
- [Girvan and Newman 2002] Girvan, M. and Newman, M. E. J., 2002. “Community structure in social and biological networks”, *P. Natl. Acad. Sci. USA* **99** (12), 7821-7826.

- [Huang and Li 2007] Wei Huang and Chunguang Li., 2007. “Epidemic spreading in scale-free networks with community structure”, *J. Stat. Mech-Theory E.* (doi:10.1088/1742-5468/2007/01/P01014)
- [Jaccard 1901] Jaccard, P., 1901. “Étude comparative de la distribution florale dans une portion des Alpes et des Jur”, *Bulletin de la Société Vaudoise des Sciences Naturelles* **37**, 547-579.
- [Kermack and McKendrick 1927] Kermack, W. O. and McKendrick, A. G., 1927. “A contribution to the mathematical theory of epidemics”, *Proc. R. Soc. Lond.A* **115**, 700-721. (doi:10.1098/rspa.1927.0118)
- [Kiss et al. 2009] I.Z. Kiss et al., 2009. “The impact of information transmission on epidemic outbreaks”, *Math.Biosci.* (doi:10.1016/j.mbs.2009.11.009)
- [Kitchovitch and Liò 2011] Kitchovitch S, Liò P, 2011. “Community Structure in Social Networks: Applications for Epidemiological Modelling”, *PLoS ONE* **6**(7), e22220. (doi:10.1371/journal.pone.0022220)
- [Lagorio et al. 2011] Lagorio, C., M. Dickison, et al., 2011. “Quarantine-generated phase transition in epidemic spreading”, *Phys. Rev. E* **83** (2), 026102. (doi:10.1103/PhysRevE.83.026102)
- [Lancichinetti and Fortunato 2009] Lancichinetti, A. and Fortunato, S., 2009. “Benchmarks for testing community detection algorithms on directed and weighted graphs with overlapping communities”, *Phys. Rev. E* **80** (1), 016118. (doi:10.1103/PhysRevE.80.016118)
- [Lancichinetti et al. 2009] Lancichinetti, A.; Fortunato, S. and Kertesz, J., 2009. “Detecting the overlapping and hierarchical community structure of complex networks”, *New J. Phys.* **11**, 033015, cite arxiv:0802.1218 [physics.soc-ph] (doi:10.1088/1367-2630/11/3/033015)
- [Leon et al. 2011] Leon Danon, Ashley P. Ford, Thomas House, et al., 2011. “Networks and the Epidemiology of Infectious Disease”, *Interdiscip.Perspect.Infect.Dis.*, **vol.** 2011, 284909. (doi:10.1155/2011/284909)
- [Marcelino and Kaiser 2009] Marcelino J, Kaiser M., 2009. “Reducing influenza spreading over the airline network”, *PLoS Currents Influenza*.**1**, RRN1005. (doi:10.1371/currents.RRN1005.)
- [Marcelino and Kaiser 2012] Marcelino, J. and Kaiser, M., 2012. “Critical paths in a metapopulation model of H1N1: Efficiently delaying influenza spreading through flight cancellation”, *PLoS Currents Influenza.* **4**, e4f8c9a2e1fca8. (doi:10.1371/4f8c9a2e1fca8)
- [Matt and Ken 2005] Matt J Keeling and Ken T.D Eames, 2005. “Networks and epidemic models”, *J. R. Soc. Interface* **2**(4), 295-307. (doi:10.1098/rsif.2005.0051)
- [Meyers et al. 2005] Meyers, L. A., Pourbohloul, B., Newman, M. E., Skowronski, D. M. and Brunham, R. C., 2005. “Network theory and SARS: predicting outbreak diversity”, *J. Theor. Biol.* **232**, 71-81. (doi:10.1016/j.jtbi.2004.07.026)
- [Newman 2002] Newman, M. E. J., 2002. “Spread of epidemic disease on networks”, *Phys. Rev. E* **66**, 016128. (doi:10.1103/PhysRevE.66.016128)

- [Newman 2006] Newman, M. E. J., 2006. “Modularity and community structure in networks”, *P. Natl. Acad. Sci.***103** (23), 8577-8582. (doi:10.1073/pnas.0601602103)
- [Salathé and Jones 2010] Salathé M. and Jones, J. H., 2010. “Dynamics and Control of Diseases in Networks with Community Structure”, *PLoS Comput.Biol.***6**(4), e1000736. (doi:10.1371/journal.pcbi.1000736)
- [Salathé et al. 2010] Salathé M., M. Kazandjieva, et al., 2010. “A high-resolution human contact network for infectious disease transmission”, *P. Natl. Acad. Sci.***107**(51), 22020-22025. (doi:10.1073/pnas.1009094108)
- [Shaw and Schwartz 2008] Shaw, L. B. and Schwartz, I. B., 2008. “Fluctuating epidemics on adaptive networks”, *Phys. Rev. E* **77**, 066101. (doi:10.1103/PhysRevE.77.066101)
- [Gregory 2009] Gregory, S. 2009. “Finding Overlapping Communities Using Disjoint Community Detection Algorithms”, *Springer Berlin Heidelberg*, pp. 47-61. (doi: 10.1007/978-3-642-01206-8_5)
- [Valdez et al. 2010] Valdez, L. D., P. A. Macri, et al., 2012. “Intermittent social distancing strategy for epidemic control”, *Phys.Rev. E***85**(3), 036108. (doi: 10.1103/PhysRevE.85.036108)
- [Watts and Strogatz 1998] Watts, D. J. and Strogatz, S. H., 1998. “Collective dynamics of “small-world” networks”, *Nature* **393**, 440-442. (doi: 10.1038/30918)

9 Appendix: Source code

- LFM strategy function

```

public void LFM(Graph<node, edge> g, node a)
{
    int total;
    double max = 0.0;
    int remove = 0;
    node tmp;
    Iterator<node> collection;
    total = g.getNeighborCount(a);
    if(total<=1)
        return;
    //contain the edges of each neighbors which is connecting between neighbors
    ArrayList<node> neighbors = new ArrayList<node>(g.getNeighbors(a));
    //int[][][0] for the number of internal degree, int[][][1] for external degree
    int[][] connection = new int[total][2];
    for(int i=0; i<total; i++)
    {
        connection[i][0] = 0;
        connection[i][1] = 0;
        tmp = neighbors.get(i);
        collection = g.getNeighbors(tmp).iterator();
        while(collection.hasNext())
        {
            node n = collection.next();
            if(neighbors.contains(n))
                connection[i][0]++;
            else
                connection[i][1]++;
        }
        connection[i][1]--;// count for node a
        connection[i][0]++ ;//count for node a
    }
    double f;
    double f_node;
    double in=0.0;
    double out=0.0;
    for(int i=0; i<total; i++)
    {
        in =(double)connection[i][0] + in;
        out =(double)connection[i][1] + out;
    }
    in = in +(double)total;
    double t = in + out;
    double t2 = t*t;
    f=in/t2; //calculate origin fitness

    max =f;
    for(int i=0; i<total; i++)
    {
        //calculate change of fitness for each node, find the one increase mostly
        double t3 = (in+out-(double)connection[i][0]-(double)connection[i][1]-1);
        double t4 = t3*t3;
        f_node = (in-2*(double)connection[i][0])/t4;
        if(f_node > max)
        {
            remove =i;
        }
    }
}

```

```

        max = f_node;
    }
}
if(max==f || g.degree(neighbors.get(remove))<=1)
    return;
int id = g.findEdge(a, neighbors.get(remove)).getID();
if(g.removeEdge(g.findEdge(a, neighbors.get(remove)))==true)
{
    remove_edges.add(id);
    removalEdges++;
}
● Local Modularity function
public void Local_modularity(Graph<node, edge> g, node a)
{
    int total;
    double max = 0.0;
    int remove = -1;
    node tmp;
    Iterator<node> collection;
    total = g.getNeighborCount(a);
    if(total <=1)
        return;
    ArrayList<node> neighbors = new ArrayList<node>(g.getNeighbors(a));
    ArrayList<node> c = new ArrayList<node>();//nodes in area c not in B
    ArrayList<node> n_c = new ArrayList<node>();
    //contain the edges of each neighbors which is connecting between neighbors
    //int[][][0] for the internal degree, int[][][1] for external degree
    int[][] connection = new int[total][2];
    int[] inner_connection = new int[total];
    //calculate internal edges and external edges for each neighbour
    for(int i=0; i<total; i++)
    {
        connection[i][0] = 0;
        connection[i][1] = 0;
        tmp = neighbors.get(i);
        collection = g.getNeighbors(tmp).iterator();
        while(collection.hasNext())
        {
            node n = collection.next();
            if(neighbors.contains(n))
                connection[i][0]++;
            else
                connection[i][1]++;
        }
        connection[i][1]--;//exclude edge to node a to get external degree
    }
    double f;
    double f_node;
    double in=0.0;
    double out=0.0;
    int k=0;
    for(int i=0; i<total; i++)
    {
        // only count those in B AND c, exclude the one in C but not in B
        if(connection[i][1]>0)
        {
            k++;
            in =(double)connection[i][0] + in;
            out =(double)connection[i][1] + out;
        }
    }
}

```

```

        else
        {
            if(connection[i][0]>0)
                c.add(neighbors.get(i));
            in = (double)connection[i][0] + in;
        }
    }
    in = in/2;
    in = in +(double)k;// add edge to node a
    //exclude the edges with two endpoint in C
    int delete=0;
    for(int y=0; y<c.size();y++)
    {
        node n_inner= c.get(y);
        inner_connection[neighbors.indexOf(n_inner)]=1;
        for(int y2=0;y2<c.size();y2++)
        {
            if(y!=y2 && g.findEdge(n_inner, c.get(y2))!=null)
            {
                inner_connection[neighbors.indexOf(n_inner)]++;
                delete++;
            }
        }
    }
    in = in-delete/2;
    f = in/(in+out); //original Local Modularity

    max = f;
    int flag = 0;
    for(int i=0; i<total; i++)
    {
        flag = 0;
        n_c.clear();
        if(connection[i][1]>0)
        {
            node t = neighbors.get(i);
            for(node n : c)
            {
                if(g.getNeighbors(n).contains(t))
                {
                    flag = 1;
                    n_c.add(n); //store the nodes in C connecting node i
                }
            }
        }
        if(flag==0) //changed modularity
            f_node= (in-(double)connection[i][0]-1)/
                (in+out-1-(double)connection[i][1]);
        else //changed modularity, update the nodes in area c
        {
            double in2 = in-(double)connection[i][0]-1;
            double out2      =      out-(double)connection[i][1]      +
            (double)connection[i][0];
            for(node n : n_c)
            {
                int index = neighbors.indexOf(n);
                //add edges of nodes in c connecting node I to internal edges
                in2 = in2 + (double)inner_connection[index];
            }
            //the edge between n_c has been added twice, reduce one time
            delete=0;
        }
    }
}

```

```

        for(node n : n_c)
        {
            for(node m : n_c)
            {
                if(!n.equals(m)&&g.findEdge(n, m)!=null)
                {
                    delete++;
                }
            }
            in2 = in2-delete;
            f_node = in2/(in2+out2); // new modularity
        }
        // find the one increase Local Modularity most
        if(f_node > max )
        {
            remove =i;
            max = f_node;
        }
    }
if(remove == -1 || g.degree(neighbors.get(remove))<=1)
    return;
int id = g.findEdge(a, neighbors.get(remove)).getID();
if(g.removeEdge(g.findEdge(a, neighbors.get(remove)))==true)
{
    remove_edges.add(id);
    removalEdges++;
}
}

```