



Stochastic Dynamics of Epidemic Networks

Haleh Ashki
May 2012
Advisor: Prof.Beerli

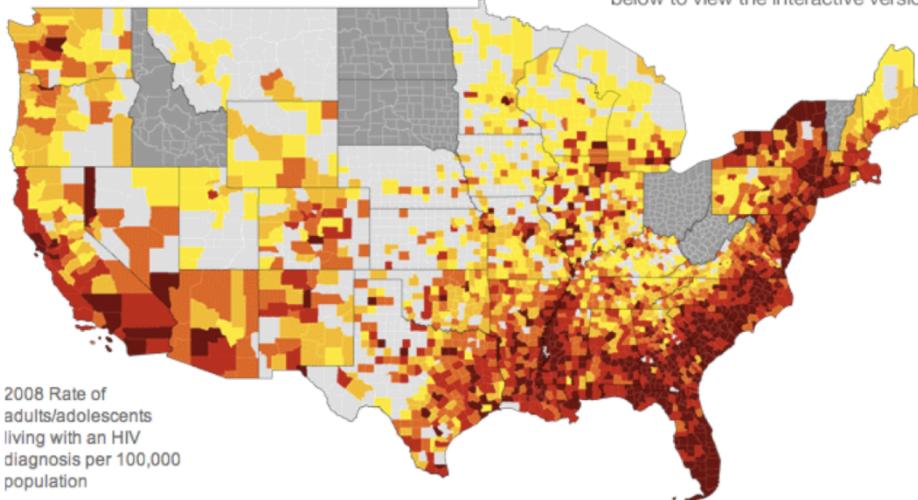




Persons Living with an HIV Diagnosis / by County / 2008



Click on this icon or on the map below to view the interactive version



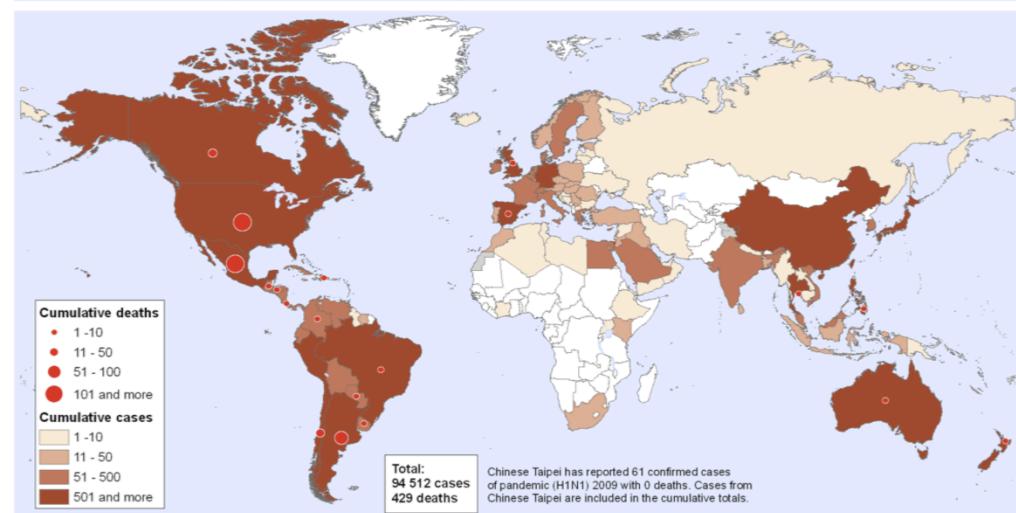
AK

HIV 2006

Pandemic (H1N1) 2009,

Number of laboratory confirmed cases as reported to WHO

Status as of 06 July 2009
09:00 GMT



The boundaries and names shown and the designations used on this map do not imply the expression of any opinion whatsoever on the part of the World Health Organization concerning the legal status of any country, territory, city or area or of its authorities, or concerning the delimitation of its frontiers or boundaries. Dotted lines on maps represent approximate border lines for which there may not yet be full agreement.

Map produced: 06 July 2009 09:00 GMT

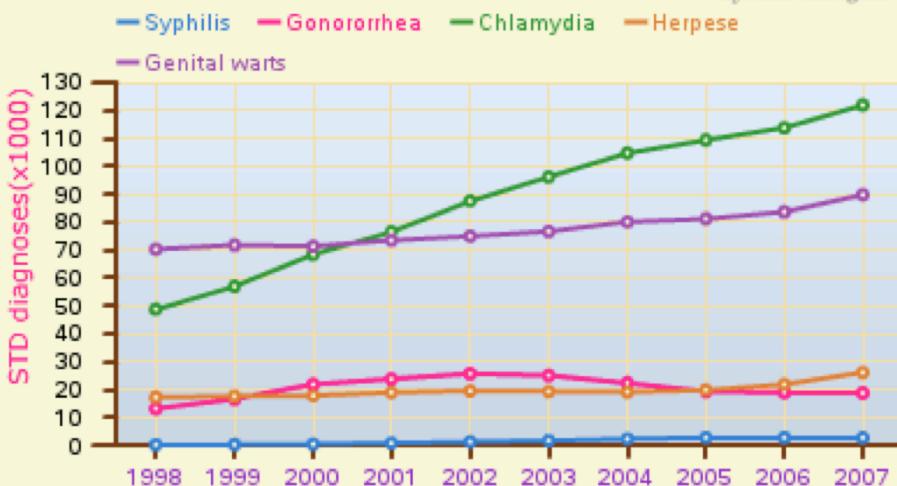
Data Source: World Health Organization
Map Production: Public Health Information and Geographic Information Systems (GIS)
World Health Organization

World Health Organization
© WHO 2009. All rights reserved

HINI 2009

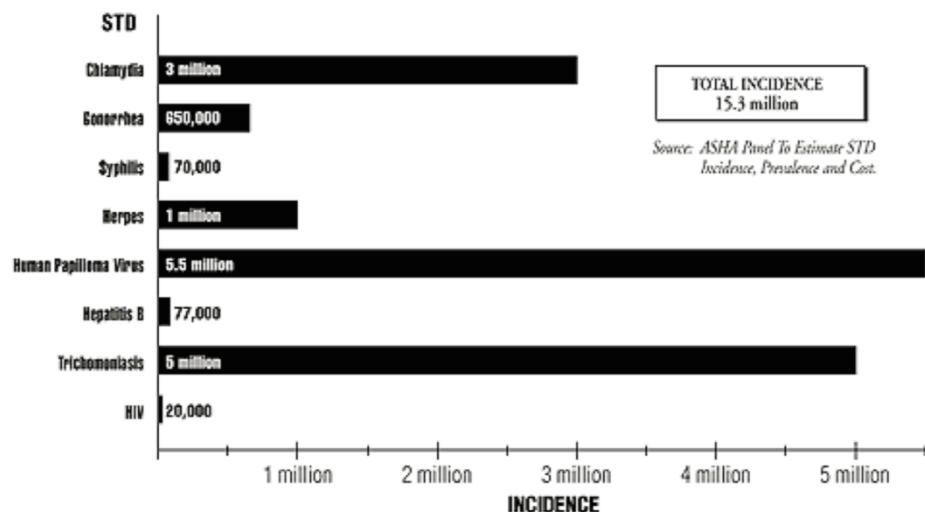
STI trend breakdown

WhyChurch.org.uk



Source: HPA Selected STI diagnoses made at GUM clinics

Figure 1. Estimated Annual New Cases of STDs



Outline

- General overview of epidemiology
- Important factors in epidemic models
- Introduction to networks
- Epidemic models
 - Network epidemiology
 - Dynamics of epidemic models
 - Static network
 - Dynamic network
- Intervention methods
- Model used in this research
 - Result
- Future work

Epidemiology

Epidemiology is the study of the **distribution** and determinants of health-related states or events (including **disease**). The aim of epidemiology is to identify the biological, behavioral, and environmental causes of health outcomes or diseases and apply this knowledge to the development of **effective disease interventions** to bring diseases under their **epidemic thresholds**.

Parameters and definitions in epidemiology

Outbreak vs Epidemic:

An **outbreak** means that a disease **emerges**, but by chance or because the transmission probability is low, **dies out** before spreading to the population at large.

In an **epidemic**, on the other hand, the infection escapes the initial group of cases into the community and results in **population-wide** incidence of the disease.

- **R₀**

The basic reproduction ratio, which is the average number of additional infectious caused by a newly infected individual

- **Epidemic threshold**

Is the separation point where a disease dies out or where it grows exponentially, it is the time point at which $R_0 = 1$.

Four concepts in epidemic models

States in
epidemic
diseases

Heterogeneity

Modeling of
epidemic
diseases

Population
structure

Four concepts in epidemic models

States in
epidemic
diseases

Heterogeneity

Modeling of
epidemic
diseases

Population
structure

To model the progress of an epidemic in a large population, individuals have been classified in three or more groups based on the behavior of disease.

SIR Model



Susceptible :individuals who are not yet infected or are susceptible to disease.

Infected :individuals who have been infected with the disease and can spread the disease.

Recovered: individuals who have been recovered from the disease and are not able to be infected again

Some other important epidemic models

- Infected individuals recover with no immunity to the disease

$$S \rightarrow I \rightarrow S$$

- Having another phase called latent or exposed phase, during which an individual is infected but not infectious.

$$S \rightarrow E \rightarrow I \rightarrow R$$

$$S \rightarrow E \rightarrow I \rightarrow S$$

- Babies are born with immunity to the disease protected by maternal antibodies, like as measles

$$M \rightarrow S \rightarrow I \rightarrow R$$

Formulating SIR model by Ordinary Differential Equations

$$\frac{dS}{dt} = -\beta SI,$$

$$\frac{dI}{dt} = \beta SI - \gamma I,$$

$$\frac{dR}{dt} = \gamma I.$$

β is the transition rate and γ is the recovery rate.

Four concepts in epidemic models

States in
epidemic
diseases

Heterogeneity

Modeling of
epidemic
diseases

Population
structure

Heterogeneity with respect to:

- Contact rate.
- Disease transition and recovery rate.
- Network structure.

To develop better models, fit with data, we have to take heterogeneity into account.

Four concepts in epidemic models

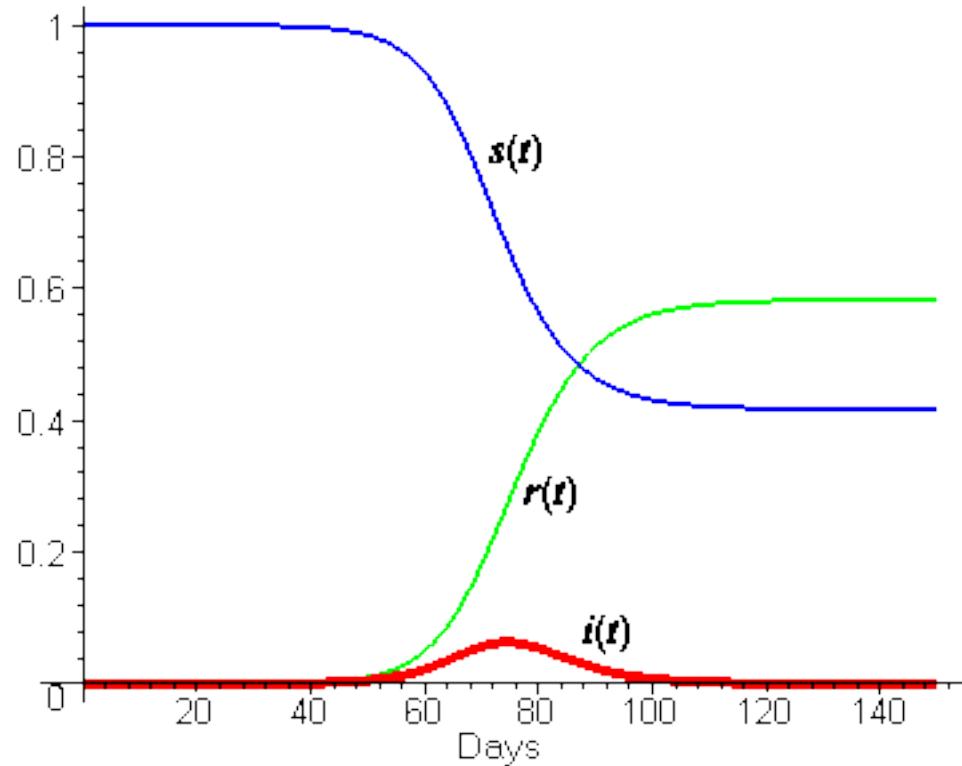
States in
epidemic
diseases

Heterogeneity

Modeling of
epidemic
diseases

Population
structure

- Epidemic models have been studies: **Deterministically and Stochastically**
In deterministic individuals in the population are assigned to different subgroups or **compartments** (SIR). The transition rates from one class to another are mathematically expressed as derivatives, and is been formulated using **ODEs**.



Four concepts in epidemic models

States in
epidemic
diseases

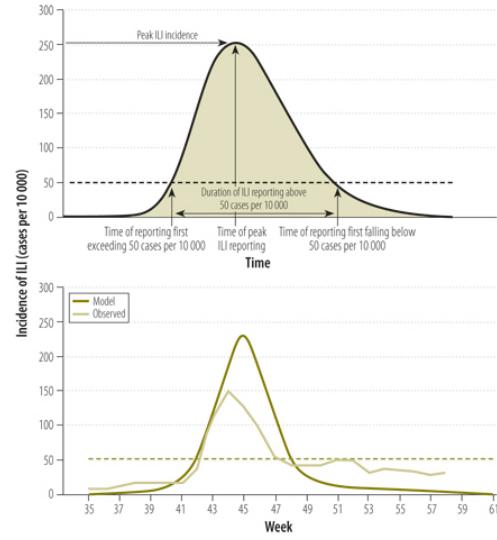
Heterogeneity

Modeling of
epidemic
diseases

Population
structure

Why that is important

The R₀ and outbreak size estimated by mathematical epidemiologist for Severe Acute Respiratory Syndrome (SARS) in 2002 and 2003 was much higher than the epidemic threshold, predicting a worldwide pandemic. Despite of this estimate, SARS has not emerged as a global pandemic.



Where the discrepancy comes from

Their models assumed **fully mixed populations** in which all individuals are equally likely to become infected, but this assumption does not hold in real population structure.

Role of population structure in epidemiology

Network epidemiology uses the network as a representation of population structure.

Representing the network using

- **Graph theory**

$$G = (V, E)$$

V: vertices or nodes and E: edges or lines

- **Adjacency matrix**

$$P(i, j)$$

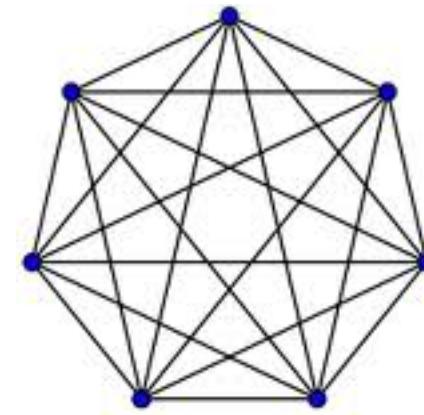
Number of edges between node i and node j.

Adjacency matrix representing a population is symmetric and sparse.

Some important networks to model population

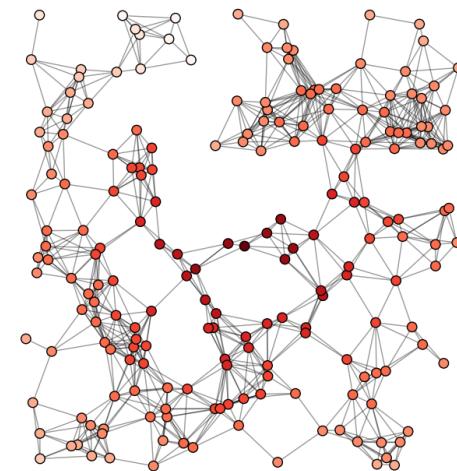
Complete Graph:

Is a simple undirected graph in which every pair of distinct vertices is connected by a unique edge.



Geometric Graph:

Nodes are connected if the Euclidean distance between them is less than an arbitrary number. This graph is good for modeling a **population with sub groups**.

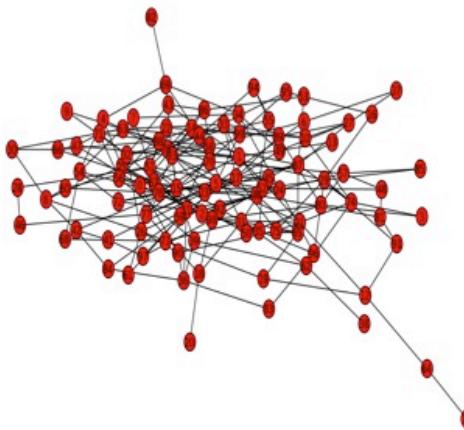


- Random Graph:

starting with a set of n vertices and adding edges between them at random.

- Picking the degree of a node from a probability distribution. This model is known as Configuration model as well.
- Creating a new edge to connect nodes based on some arbitrary probability p .

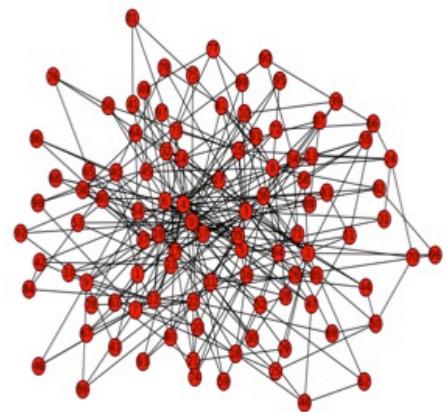
Some well-known random graphs that have been used in this research



Erdős-Rényi random graph



Watts-Strogatz graph



Barábasi-Albert graph

100 nodes. $P=0.05$, probability for edge creation. node degree = 3.



Network Epidemiology

Static networks

Probability Generating Function(PGF)

PGF is a polynomial whose coefficients are the probabilities. PGF is a powerful method in network theory to calculate properties of a random graph with degree distribution function.

$$G_0(x) = \sum_{k=0}^{\infty} P_k x^k$$

$G_0(x)$ is a generating function for the probability distribution of vertex degree k and P_k is the probability that a randomly chosen vertex on the graph has degree k .

Some properties of PGF

$$G_0(1) = \sum_k P(X = k) = 1.$$

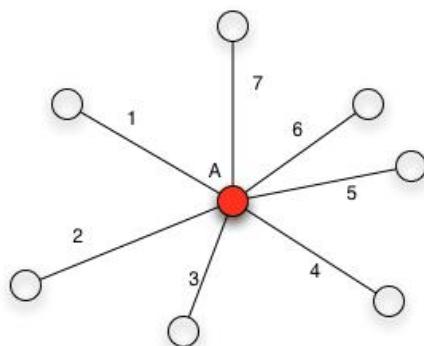
$$G'_0(x) = \sum_{k=0}^{\infty} k P_k x^{k-1}.$$

$$G'_0(1) = \sum_k k P(X = k) = E(X) = \langle k \rangle.$$

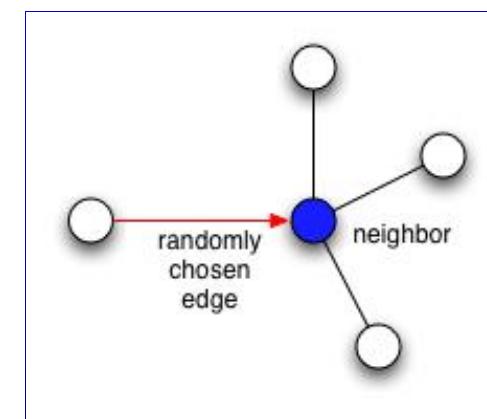
$G'_0(1)$ is the mean of the probability distribution.

Excess degree

Excess degree of a node is the number of neighbors excluding the one we reach that node by, which is the degree minus one.



$$G_0(x)$$



$$G_1(x)$$

The generating function of excess degree

Probability to find that node

Discount the edge we arrived along

$$G_1(x) = \frac{\sum_k kp_k x^{k-1}}{\sum_k kp_k} = \frac{G'_0(x)}{G'_0(1)}$$

normalization

The average of the excess degree : $G'_1(1)$

$$\begin{aligned}\sum_{k=0}^{\infty} (kq_k) &= \frac{\sum_{k=0}^{\infty} (k(k+1)p_{k+1})}{\sum_k kp_k} \\&= \frac{\sum_{k=0}^{\infty} ((k-1)kp_k)}{\sum_k kp_k} \\&= \frac{<k^2> - <k>}{<k>} \\&= \frac{G''_0(1)}{G'_0(1)} = G'_1(1)\end{aligned}$$

Probability distribution function and PGF of some graphs

Poisson distributed graph

$$P_k = \frac{Z^k e^{-z}}{k!}$$

$$G_0(x) = e^{z(x-1)}$$

Power law distributed graph

$$P_k = C k^{-\tau} e^{\frac{-k}{\kappa}}$$

$$G_0(x) = \frac{Li_{\tau}(xe^{\frac{-1}{\kappa}})}{Li_{\tau}(e^{\frac{-1}{\kappa}})}$$

Predicting reproduction ratio R0 and average outbreak size $\langle s \rangle$

Newman and his colleagues as well as some other groups have used the network properties and the PGF equations to calculate the R0 and average outbreak size.(Karrer and Newman; 2011)

$$R0 = T \sum_{k=0}^{\infty} (kq_k) = T \frac{\langle k^2 \rangle - \langle k \rangle}{\langle k \rangle} \quad T: \text{transmissibility rate}$$

In PGF format

$$R0 = TG'_1(1) = T \frac{G''_0(1)}{G'_0(1)}$$

Using the epidemic threshold definition $R_0 = 1$

$$T_c = \frac{\langle k \rangle}{\langle k^2 \rangle - \langle k \rangle}$$

$$T_c = \frac{1}{G'_1(1)} = \frac{G'_0(1)}{G''_0(1)}$$

The epidemic thresholds, T_c , is the minimum transmissibility, T required for an outbreak to become a large-scale epidemic.

Average outbreak size

$$\langle s \rangle = 1 + \frac{TG'_0(1)}{1 - TG'_1(1)}$$

Newman (2002)

Input	Output
Network structure	R_0
Disease transition rate	Average outbreak size

What is **missed**

Dynamics of the system.

disease incidence at any times or, in other words, calculating the number of susceptible, infected and recovered individuals **at any time**

Dynamics of SIR model Volz (2007)

$$PI = \frac{M_{SI}}{M_S} \quad , \quad PS = \frac{M_{SS}}{M_S}$$

PI is the probability that a susceptible node in graph has an infectious neighbor.

PS is the probability that susceptible node in graph has a susceptible neighbor.

Dynamics of number of infected and susceptible individuals

$$\frac{dS}{dt} = -rPI\theta g'(\theta)$$

$$\frac{dI}{dt} = -rPI\theta g'(\theta) - \mu I$$

θ is the fraction of degree of nodes that remain susceptible at time t.

r is the constant infectious rate and μ is the constant recovery rate.

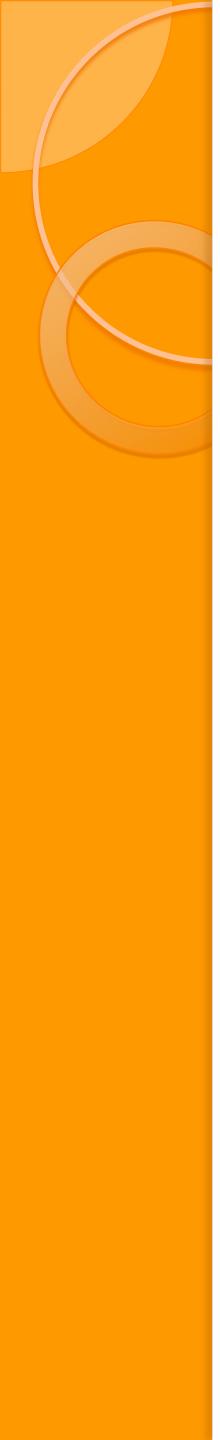
Epidemic threshold

The outbreak turns into an epidemic when

$$\frac{dPI}{dt} = 0 = \epsilon \left(r \frac{g''(\theta)}{g'(\theta)} - r - \mu \right)$$

The epidemic threshold in term of τ is:

$$\tau = \frac{g'(1)}{g''(1)}$$



Network Epidemiology

Dynamic networks

Dynamic networks

Volz and Newman methods consider **static not dynamic networks.**

In dynamic networks pattern of contact change in terms of

- The connections
- Number of contacts
- Duration of contacts

Dynamic network with contact exchange rate

Volz et all (2007,2009) designed **dynamic network method** by using Neighbor Exchange method(NE).

ρ : swap rate, exchange between neighbors. Yet the degree of node stay constant.

$$R0 = \frac{\tau\rho}{\mu} + \tau \frac{\mu + \rho}{\mu} \frac{g''(1)}{g'(1)}$$

Dynamic network with degree change rate

Volz et al(2011) created a dynamic network method, in which the degree of node change was based on a probability distribution.

A node is assigned degree κ using $\rho(\kappa)$, create edge at rate $\kappa \eta$ and existing edge break at η .

$$R0 = \frac{\beta}{\beta + \eta + \gamma} \frac{\eta + \gamma}{\gamma} \frac{\langle k^2 \rangle}{\langle k \rangle}$$

Intervention methods

The aim of epidemiology is to develop **effective disease interventions** to bring diseases under the **epidemic thresholds**.

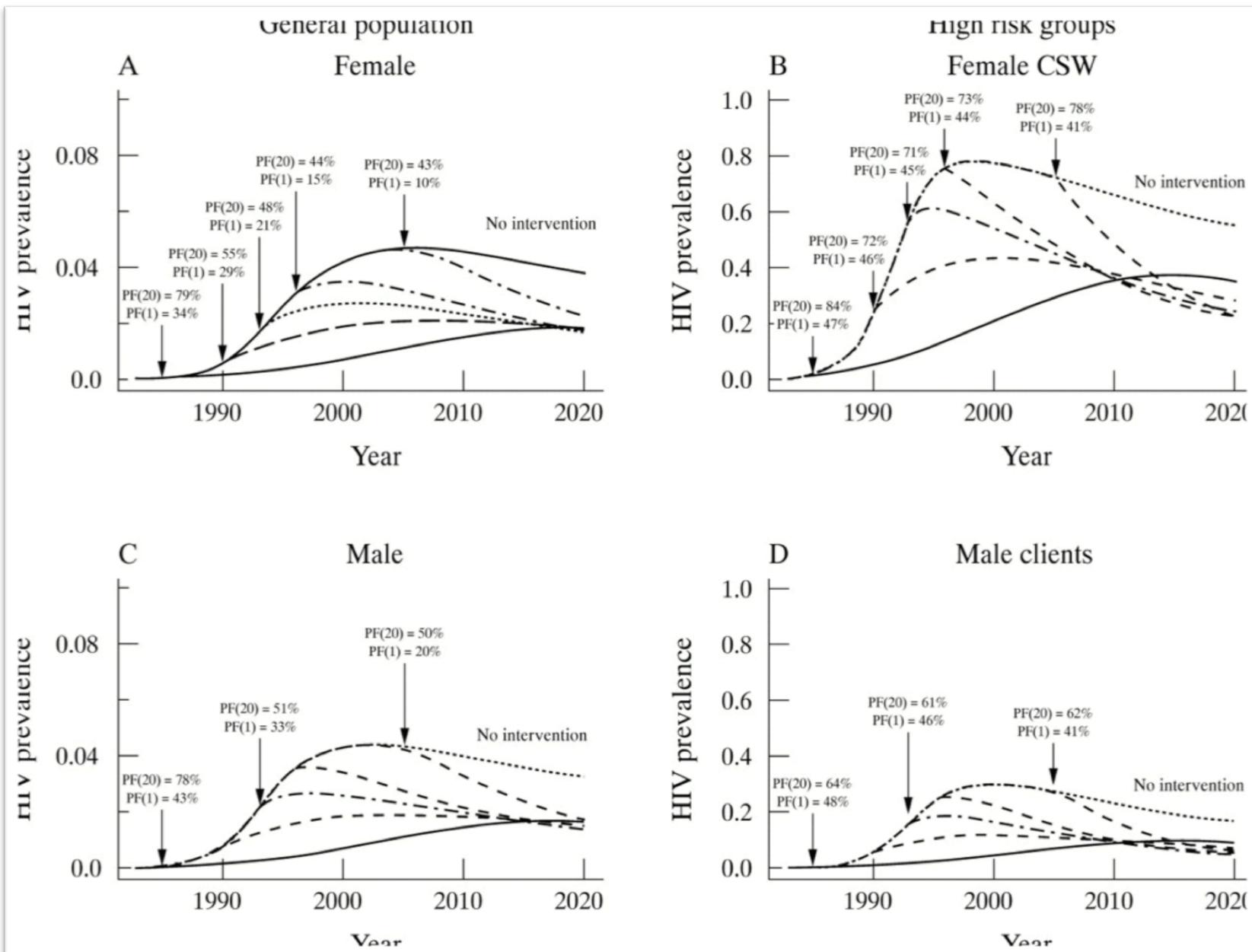
Disease Prevention

- Epidemiology goal is **improving** public health. One way is to **control** the disease by preventing occurrence of disease, reduce the propagation and consequence of disease.
- Different **intervention methods** apply in during an emerging epidemic to reduce the impact of disease in society.
- It's crucial to know **when, where** and **how large** the intervention has to be. As well as the **cost** of each intervention methods and intervention **feasibility**.

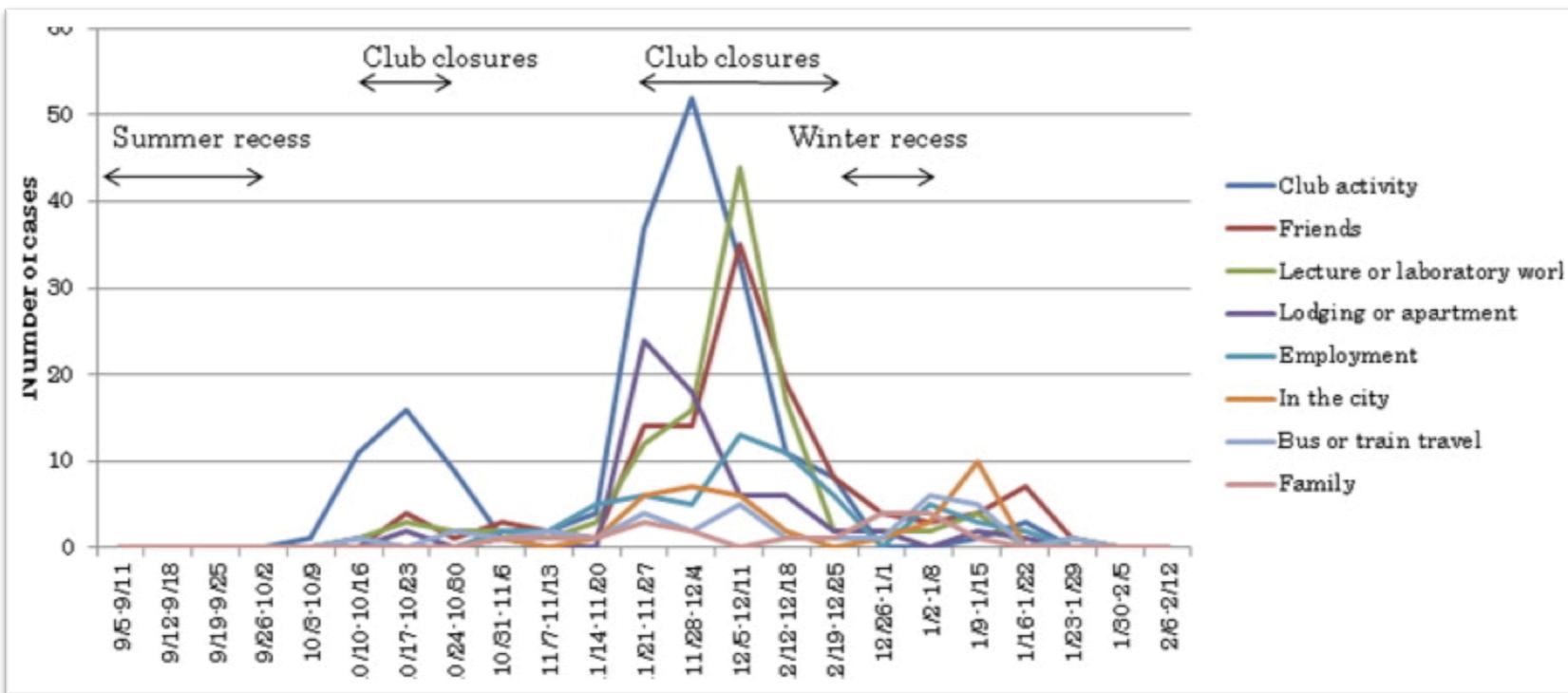
Intervention methods

- Reduce connectivity: Closing public spaces.
- Increase immunity:
 - Flu:Vaccination(Flu shot), Using mask, ...
 - HIV: drug users: free syringe
 - HPV:Vaccination

Impact of intervention method



Example of number of infected cases in different public places during an epidemic



Impact of intervention methods

The impact of intervention methods on estimated R_0 and $\langle S \rangle$ is been tested by different group and methods (Meyers et al 2004).

To test the impact of intervention methods, we have to be able to test that by applying

- At different time during epidemic.
- By different rate of connection property:
degree, connections, durations
- By different transition and recovery rate.

Model comparison

	Dynamics of epidemic	Dynamic Transition and Recovery rate	Dynamic Network	Dynamic Intervention methods	R0 and outbreak size estimation
Newman et al.	No	No	No	No	Yes
Volz et al.	Yes	No	Yes	No	Yes



Model I have used in this study

Markov chain

A Markov chain is a stochastic process with the Markov property. The Markov property states that at any time the probability of being in a state depends only on what the previous state was.

$$P(X_n \mid X_{n-1}, X_{n-2}, \dots) = P(X_n \mid X_{n-1})$$

- **State vector**

X_n represents the state of the Markov chain at time n.

- **Transition matrix**

Describes the process

$$Q(j \mid i) = Q_{i,j}$$

Continuous time Markov process

Time is continuous, but the state variable is discrete.

The evolution of a continuous-time Markov process is given by the first-order differential equation

$$\dot{X} = QX$$

These equations known as **forward Kolmogorov differential equation**. Then the Kolmogorov equations are a means for determining features of the stochastic process.

Important factor to propagation of infectious disease

- Number of connections among individual.
- Transition rate(immunity) which is unique per individual.
- Recovery rate or recovery period.

Transition Matrix

Is an $N \times N$ **adjacency** matrix representing the connection among individual.

The **transition rate** β_i added as an coefficient in each row of matrix.

Therefore this matrix is a **weighted adjacency matrix** representing both **connectivity and transition rate**.

State Vectors

- There are three state vectors:
- $X_S(t)$ is a probability vector of size n , indicates the probability of each individual being in epidemic state, S at time t .
- $X_I(t)$ is a probability vector of size n , indicates the probability of each individual being in epidemic state, I at time t .
- $X_R(t)$ is a probability vector of size n , indicates the probability of each individual being in epidemic state, R at time t .

Continuous time Markov chain of SIR model

There are three Markov processes for each compartment of the epidemic SIR model happening in parallel

$$\frac{d(XI)}{dt} = P \times XS - XR,$$

$$\frac{d(XS)}{dt} = -P \times XS,$$

$$\frac{d(XR)}{dt} = L \cdot XI$$

Simulation

(code is been implemented in matlab and some of the networks and adjacency matrices are created in NetworkX, python package)

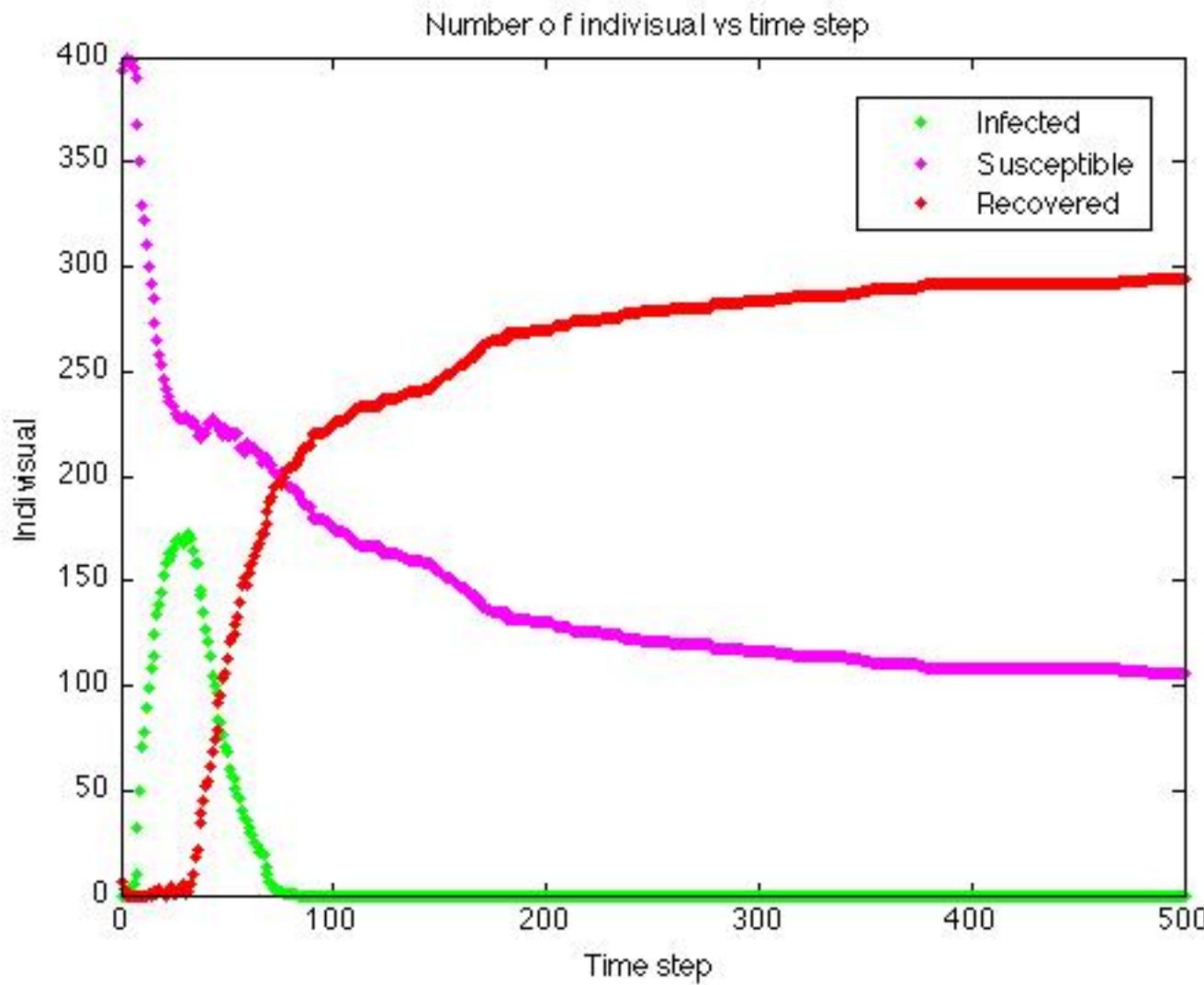
Input:

1. Network structure as an adjacency matrix.
2. Vector of independent random variables for transition rate.
3. Recovery rate.

Output:

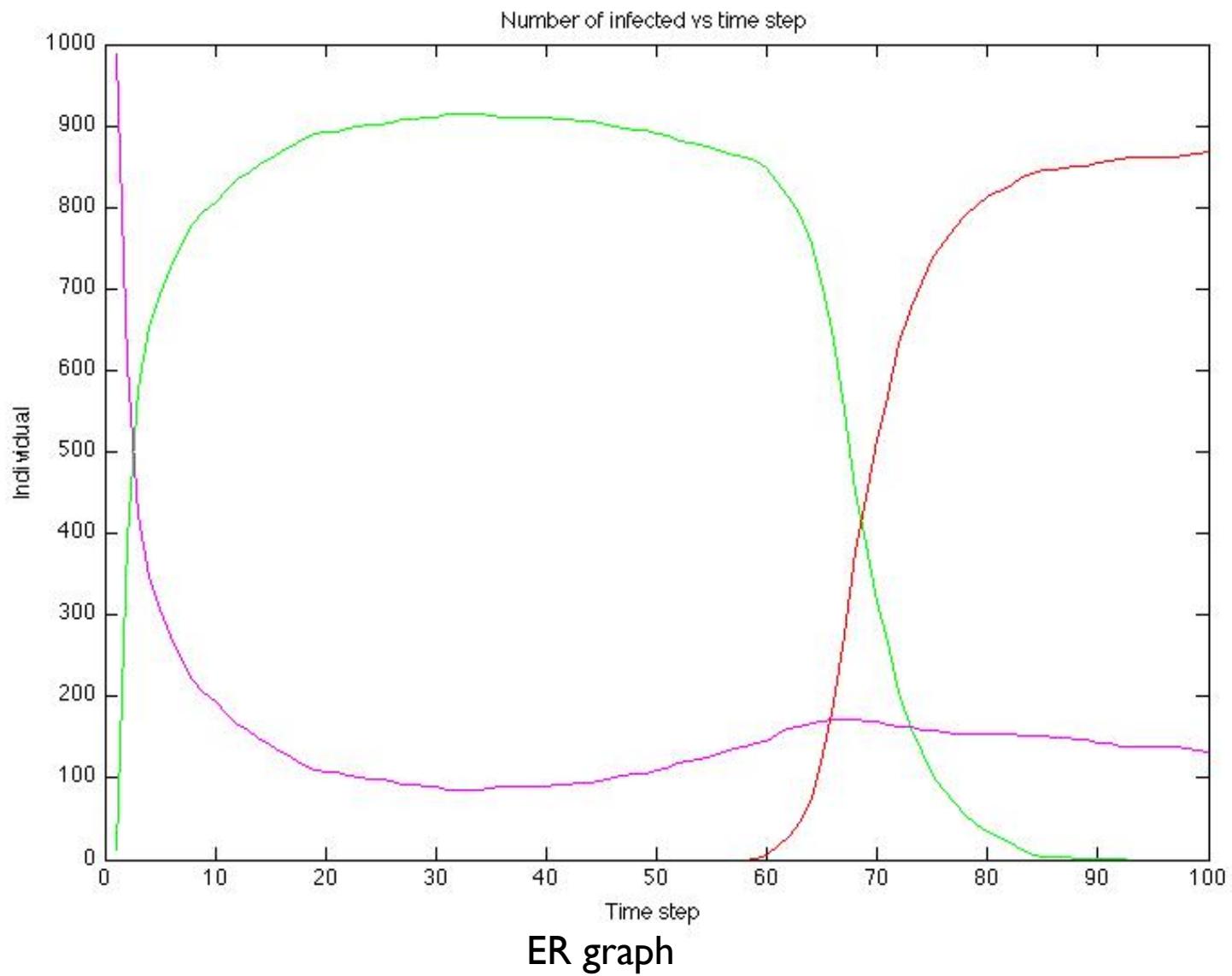
Number of Susceptible, Infected and Recovered over time.

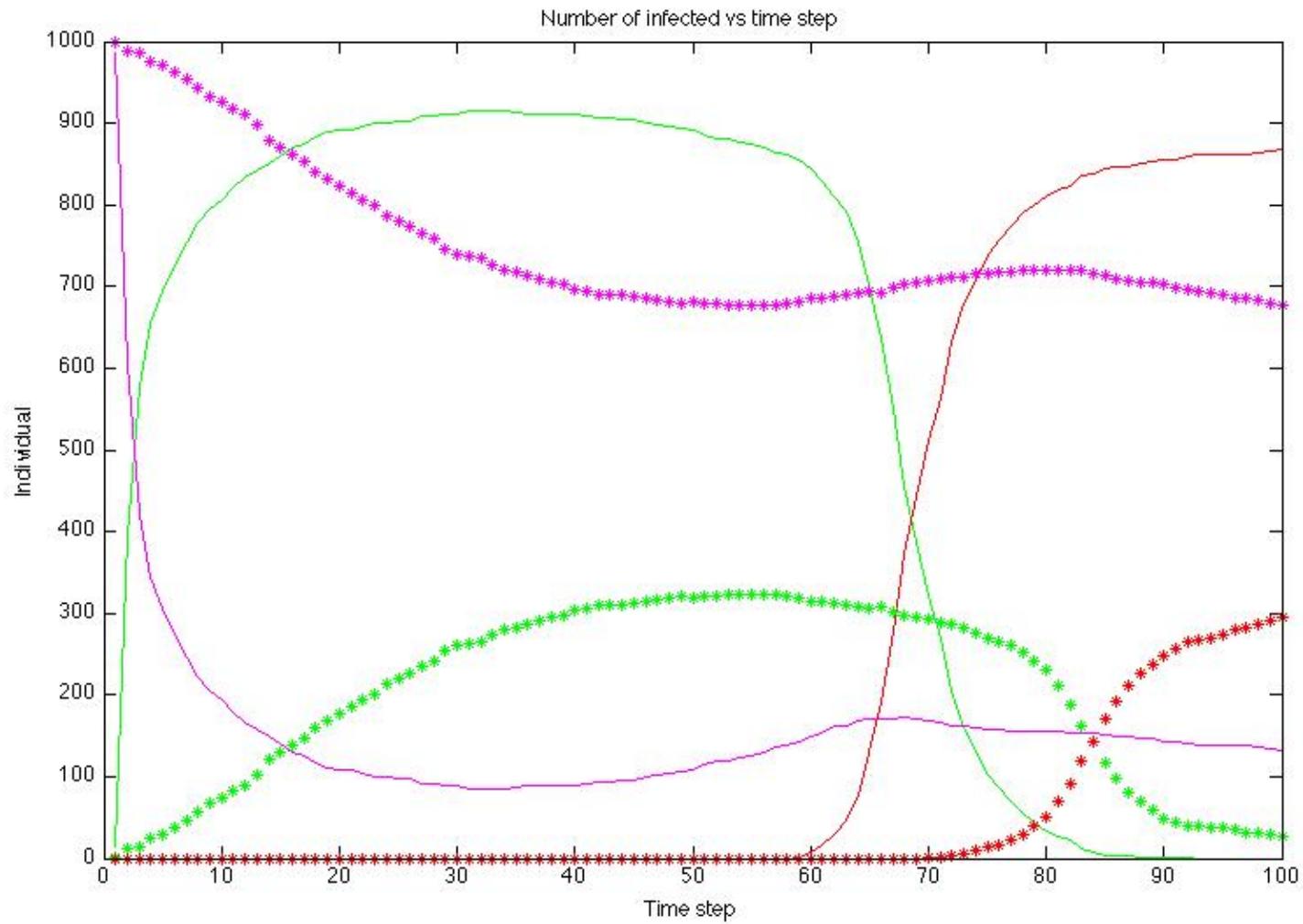
Each individual is labeled as susceptible, infected or recovered based on the highest state probability at that particular time step, and that is repeated over the whole time domain



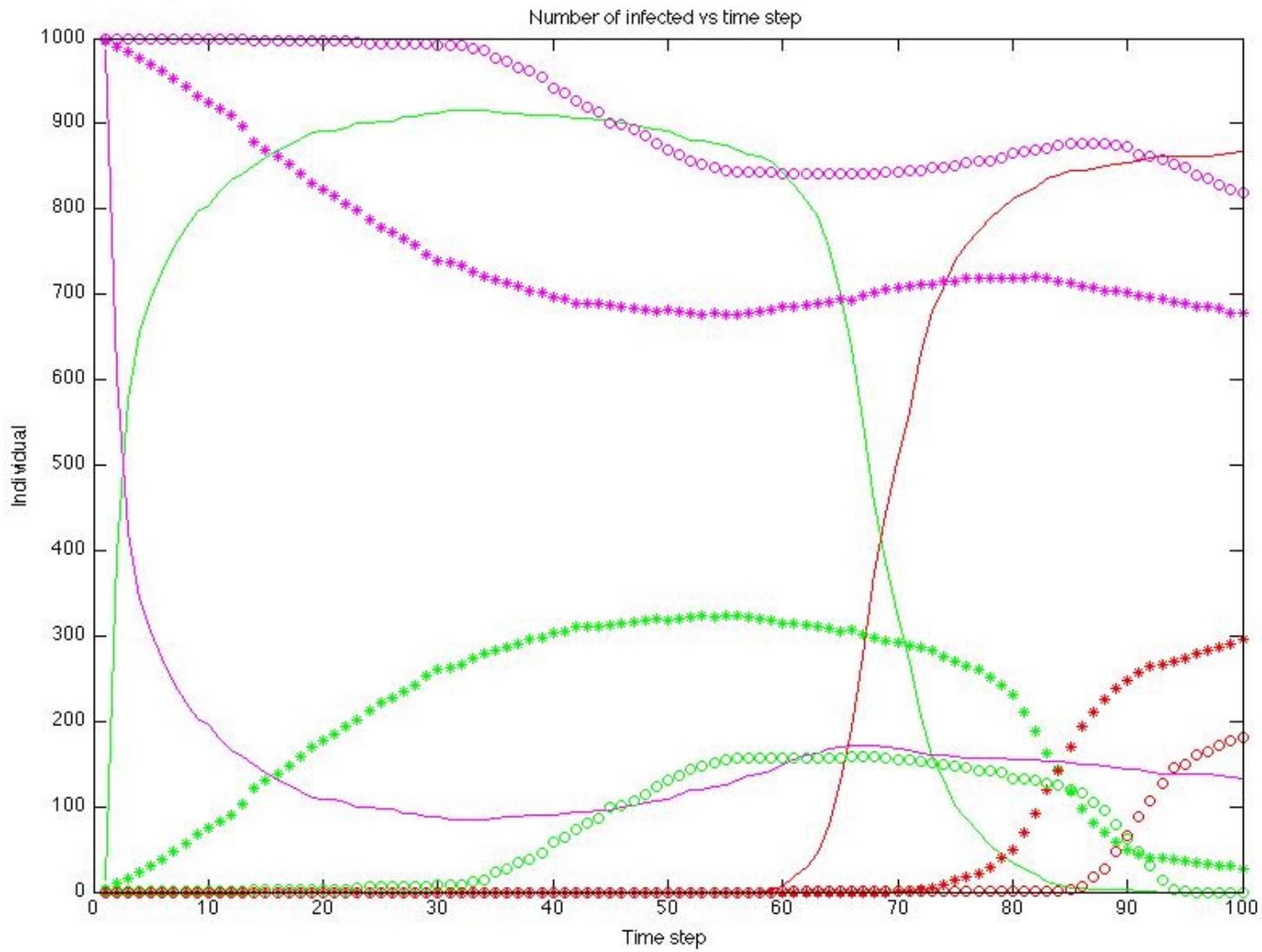
Number of individual in each compartment for a population size of 400 individuals over 500 time steps.

Change of connectivity ER, WS, and BA networks.



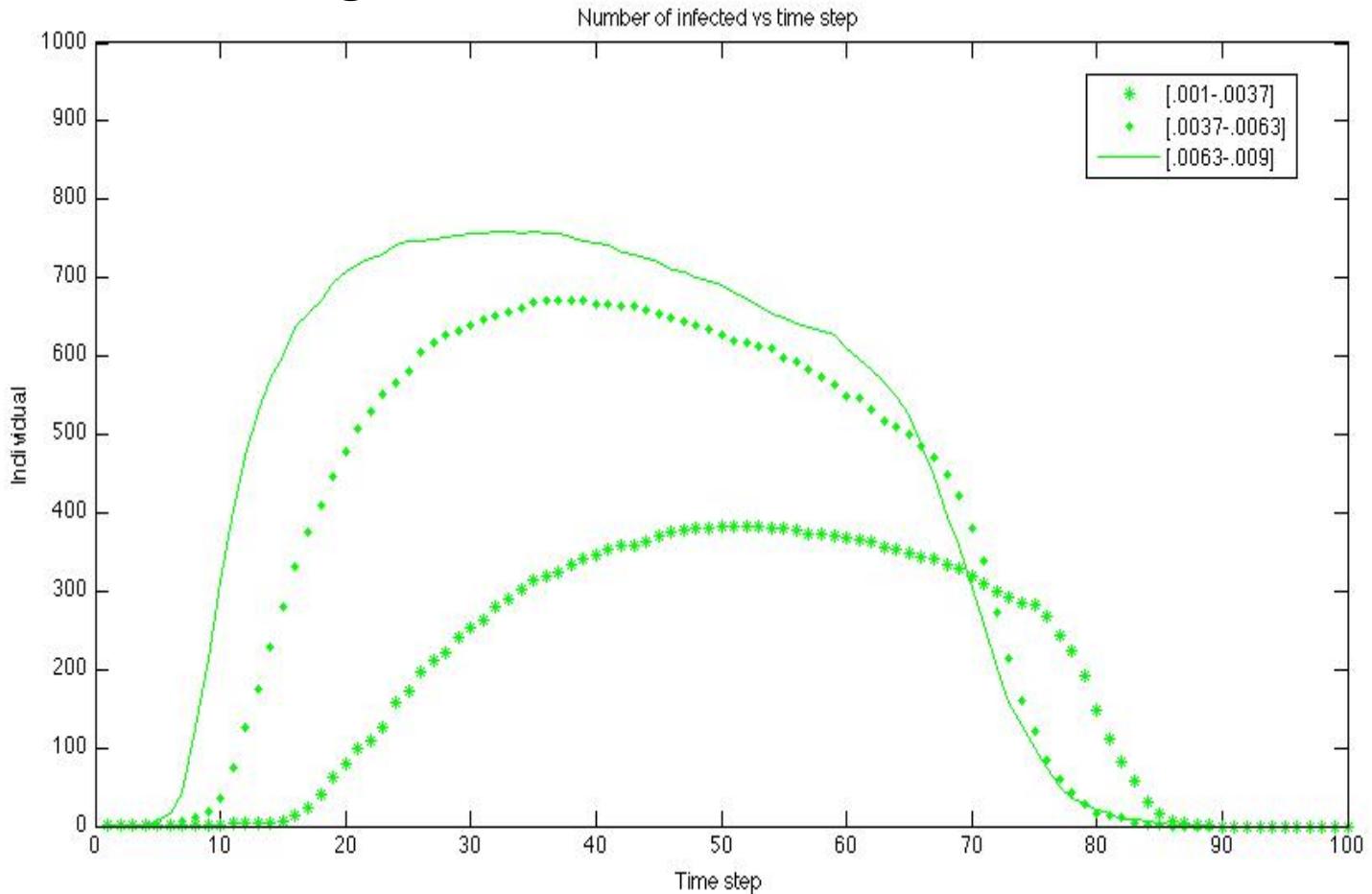


BA graph



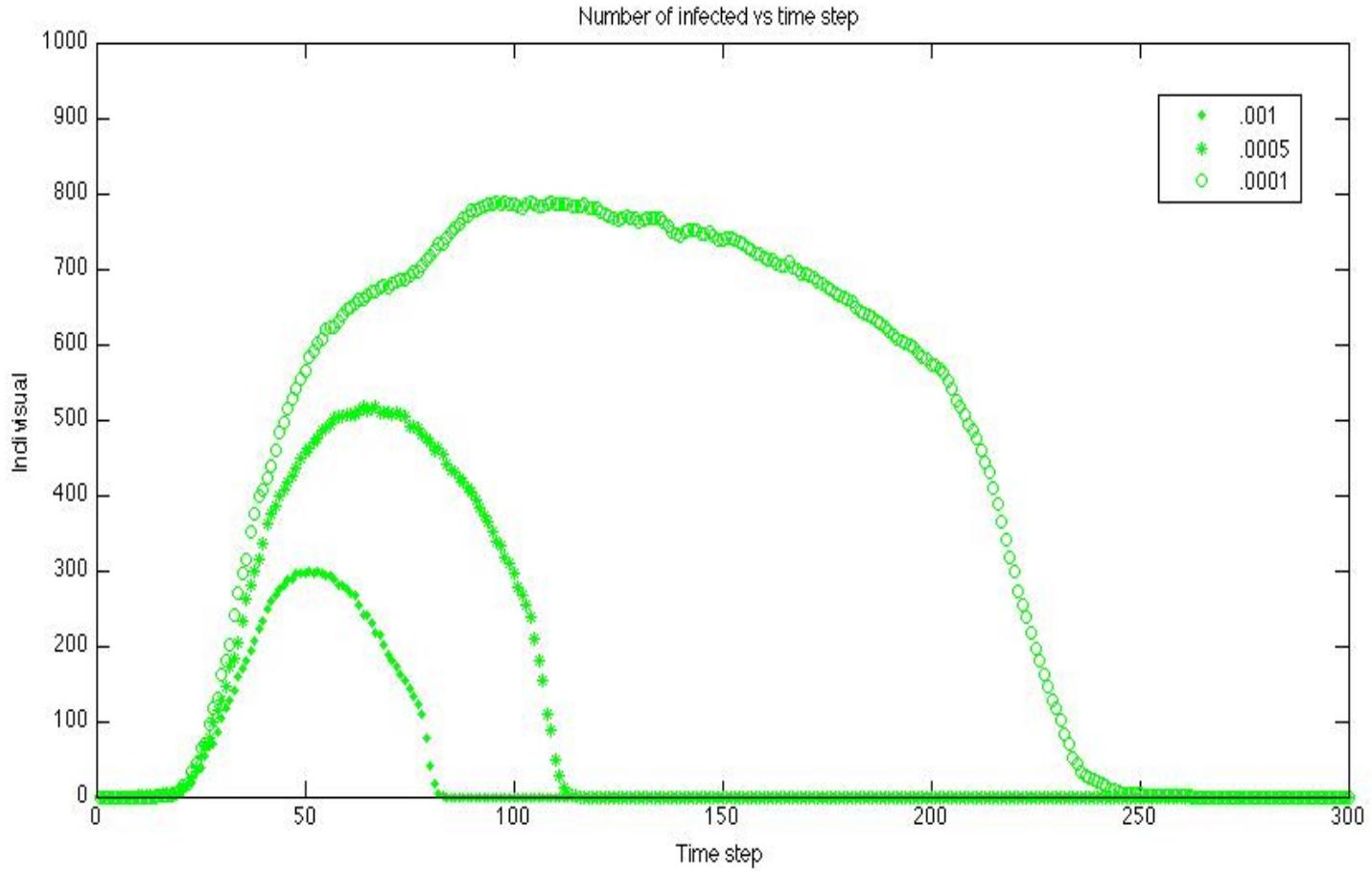
WS graph

Change of disease transition rate



Different transition rate domain is tested on the same network for 1000 individual. [.001-.0037], [.0037-.0063], [.0063-.009]

Change of recovery rate



Three recovery rates were simulated: 0.001, 0.0001, 0.0005

Impact of intervention methods

The method is applied once at the time that was $2/3$ before the time it reached the maximum number of infected individuals.

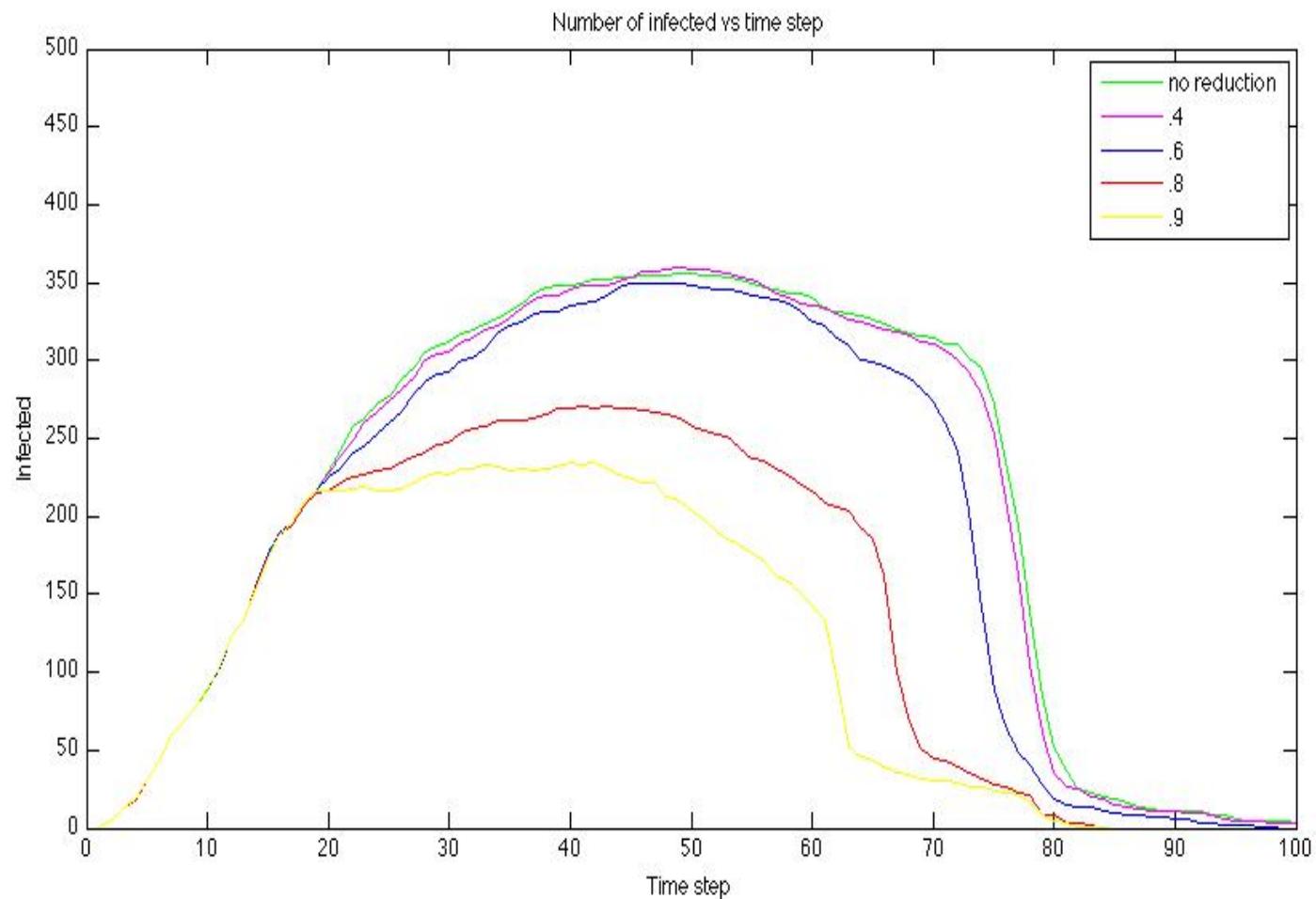
- Connection reduction:**

The algorithm picks the nodes with a higher degree than average and then reduces their degree by a fraction (0.4, 0.6, 0.8, and 0.9)

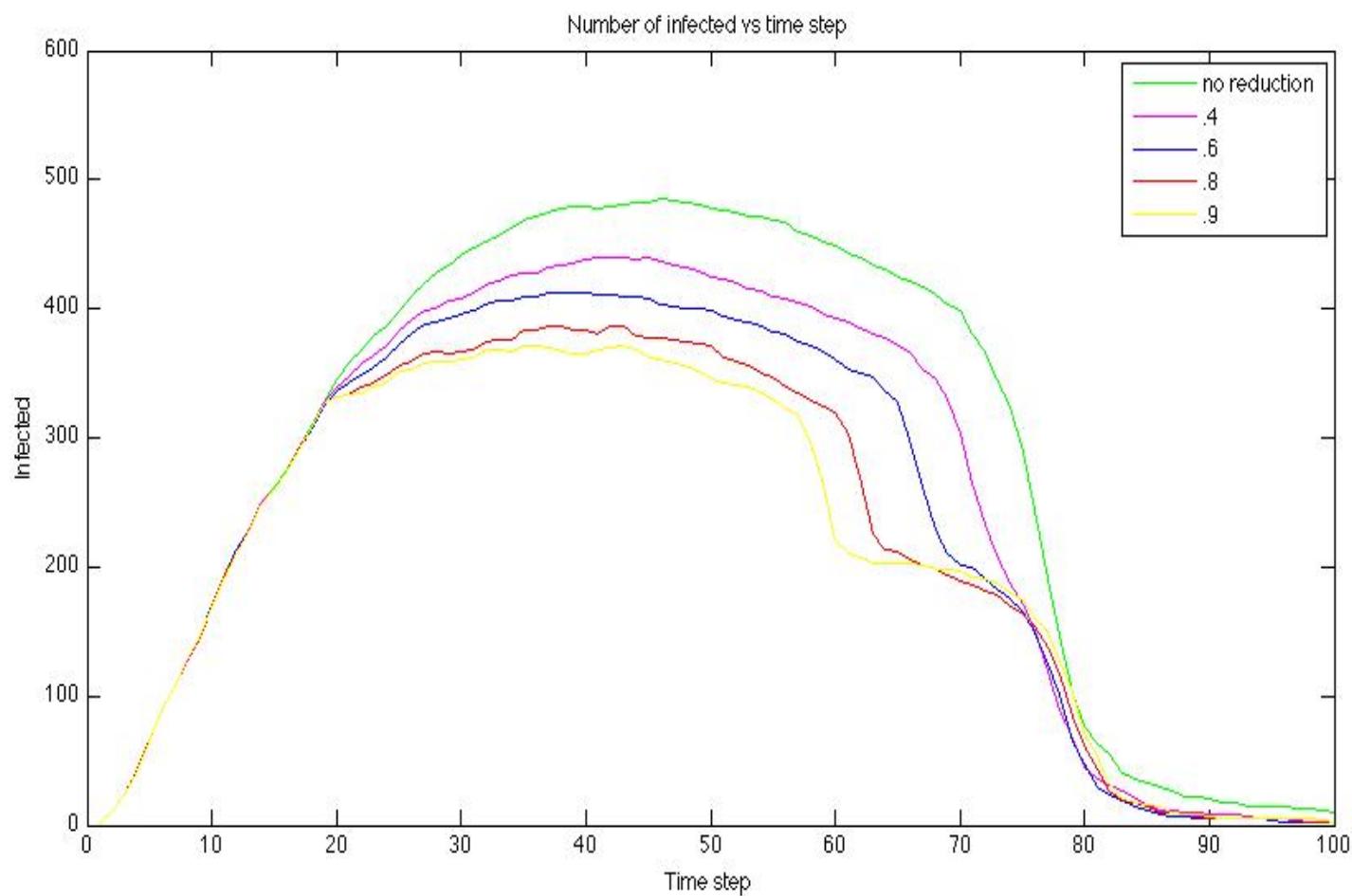
- Transition rate reduction.**

Nodes with edge weights higher than average are picked and the rate is reduced by subtracting some percentage of edge weight (0.4, 0.6, 0.8, and 0.9).

Node degree reduction on 500 individual network

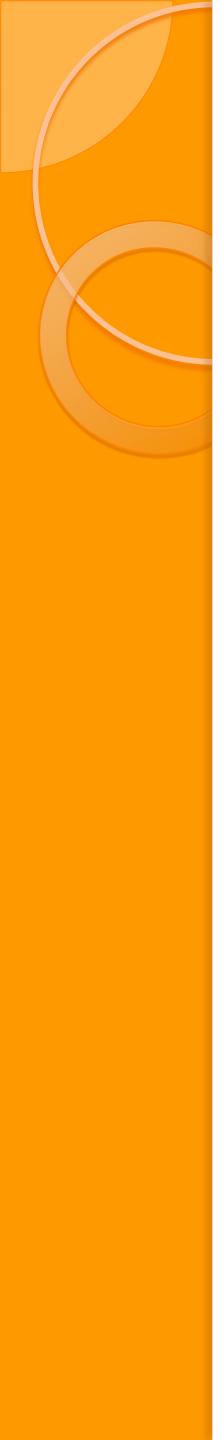


Transition rate reduction on 500 individual network



Model comparison

	Dynamics of epidemic	Dynamic Transition and Recovery rate	Dynamic Network	Dynamic Intervention methods	R0 and outbreak size estimation
Newman et al.	No	No	No	No	Yes
Volz et al.	Yes	No	Yes	No	Yes
Mine	Yes	Yes	Yes	Yes	No(not yet)



Future Works

Complexity and memory usage

- Calculating derivative of three state vectors of population size n repeated m times needs $O(n^2m)$ operation.
- In matlab it needs $8 \times n^2$ bytes to store a adjacency matrix of population of size n . But since this matrix is sparse, it needs less space.

Improving on memory usage and operation.

Reducing the size of transition matrix

lump or group the individuals that have similar properties to a node with higher weight.

Properties we may consider:
same connectivity rate, same immunity,
same biological background, being in same
public places,...

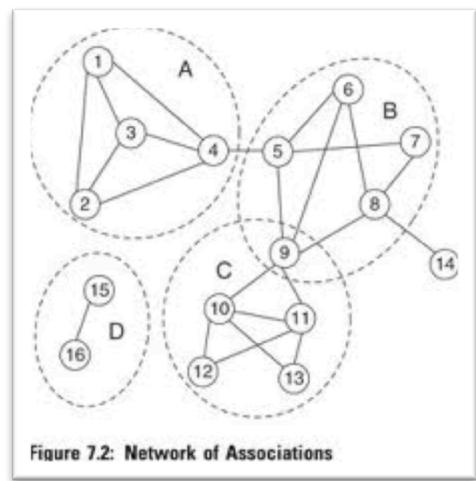
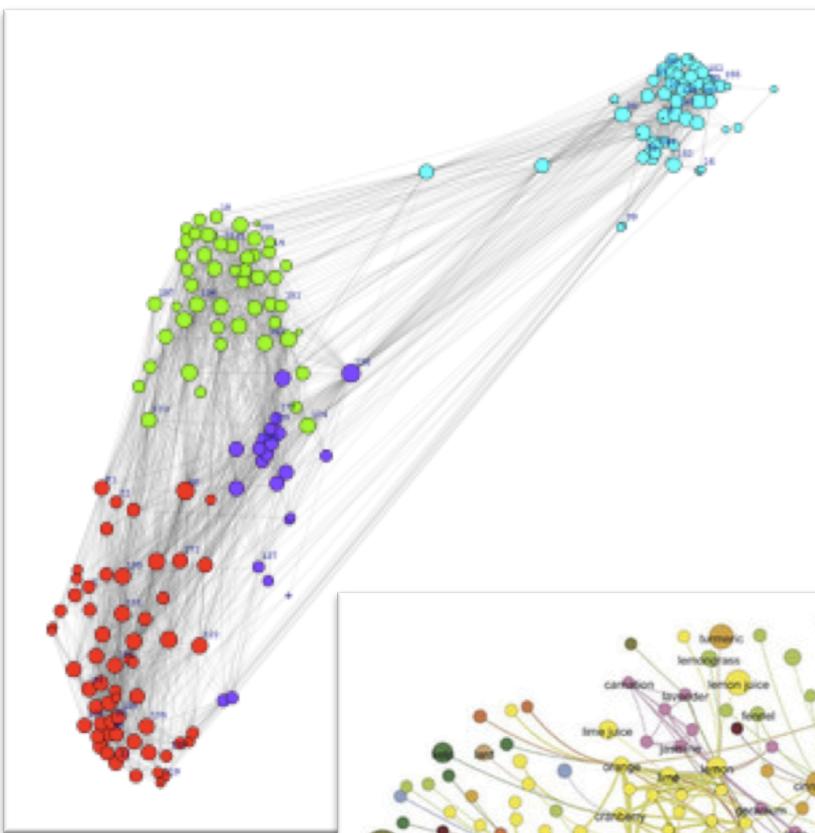
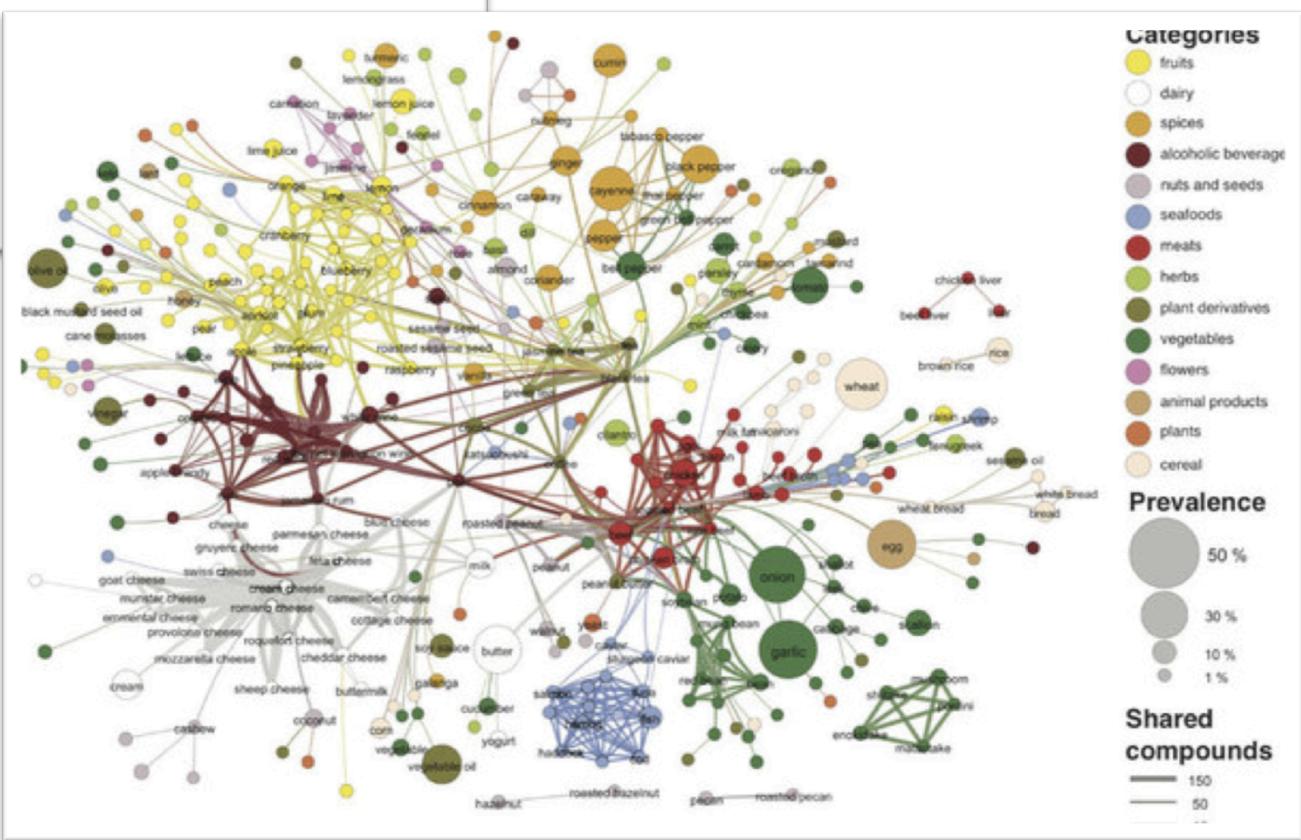


Figure 7.2: Network of Associations



A method for reducing matrix size

generating the network and corresponding matrix based on some data, information from population. Such as sub-groups in population(schools, health centers,...)

- Immunity of different group of people based on age, ...
- Race and biological background,..

It leads to having network with bigger nodes having higher weight.

Lumping methods that I have reviewed

- K-mean clustering
- Optimal Markov Chain Lumping
(Derisavi et al., 2003, 2005; Derisavi, 2007)
- Normalized Cut and Image Segmentation
Shi and Malik (2000)
- Lumping Complex Networks
Filliger and Hongler (2008)
- SVD

K-mean clustering

partition the n observations into k sets ($k \leq n$)

$$S = \{S1, S2, \dots, Sk\}$$

$$\arg \min_{\mathbf{S}} \sum_{i=1}^k \sum_{\mathbf{x}_j \in S_i} \|\mathbf{x}_j - \boldsymbol{\mu}_i\|^2$$

where $\boldsymbol{\mu}_i$ is the mean of points in S_i

Optimal Markov Chain Lumping

(Derisavi et al., 2003, 2005; Derisavi, 2007)

This method lump matrix diagram representation of Markov model.

Calculates the sum of each row in matrix and group the rows having same value. This is done by kind of binary search algorithm.

I modified it by using hash table instead.

Normalized Cut and Image Segmentation

Shi and Malik (2000)

- Make the weight matrix representation of pixels in an image(similar intensity, close to each ,etc)
- Solve $(D - W)x = \lambda Dx$ for eigenvectors with the smallest eigenvalues.
- Use eigenvector with the second smallest eigenvalue to bipartition the graph.
- Decide if the current partition should be sub-divided, and recursively repartition.

Lumping Complex Networks

- This method reduce the size of large matrices representing the network preserving at the same time the spectral properties of the underlying graph.
- They calculate the largest and second eigenvalue of the matrix and their corresponding eigenvector
- Divide the interval between largest and smallest value in each eigenvector based on the arbitrary number of clusters.
- Identify the component lye in each partition.
- Combine these two matrix to get the lumped matrix.

SVD :Singular Value Decomposition

- Calculate the SVD of the matrix.
- Choose the some of the first eigenvalues (based on the error ...)
- Make the approximated matrix by using these eigenvalues and SVD of matrix and.

Complexity Issues

- Calculating eigenvalue and eigenvector of a matrix it's a complex operation.
- I have to optimize the increasing the complexity of calculating of eigenvalues and the reducing of memory usage.

More future work

- Calculating R₀ and average outbreak size.
- Applying more intervention methods.
 - New methods
 - Testing existence method by applying at different time.
- Implementation
 - Coding in Python and adding interactive graphs that show the propagation of disease population.

Phyldynamics

- Phylodynamics, a term introduced by Grenfell et al. (2004) as “The melding of immunodynamics, epidemiology and evolutionary biology”.
- Volz et al. (2009) has used the coalescent theory to fit an epidemic SIR model to sequence data.
- Leventhal et al. (2012) also has used the SIR epidemic model and has construct the phylogenetic tree based on who has infected whom and have compare that with phylogenetic tree of Swiss HIV data.

Future work

As a future work I will use the network topology on calculating model parameters by integrating over many phylogenetic trees(coalescent theory).

I will use the genetic data, phylogenetic theory as well as population dynamics of infectious agent for modeling the dynamics of infectious disease.

I thank:

Dr. Beerli

Dr. Gallivan

Dr. Lemmon

Dr. Slice

Dr. Wilgenbusch

Department of Scientific Computing

All my friends in department in particular

Michal

Last but not least, my husband