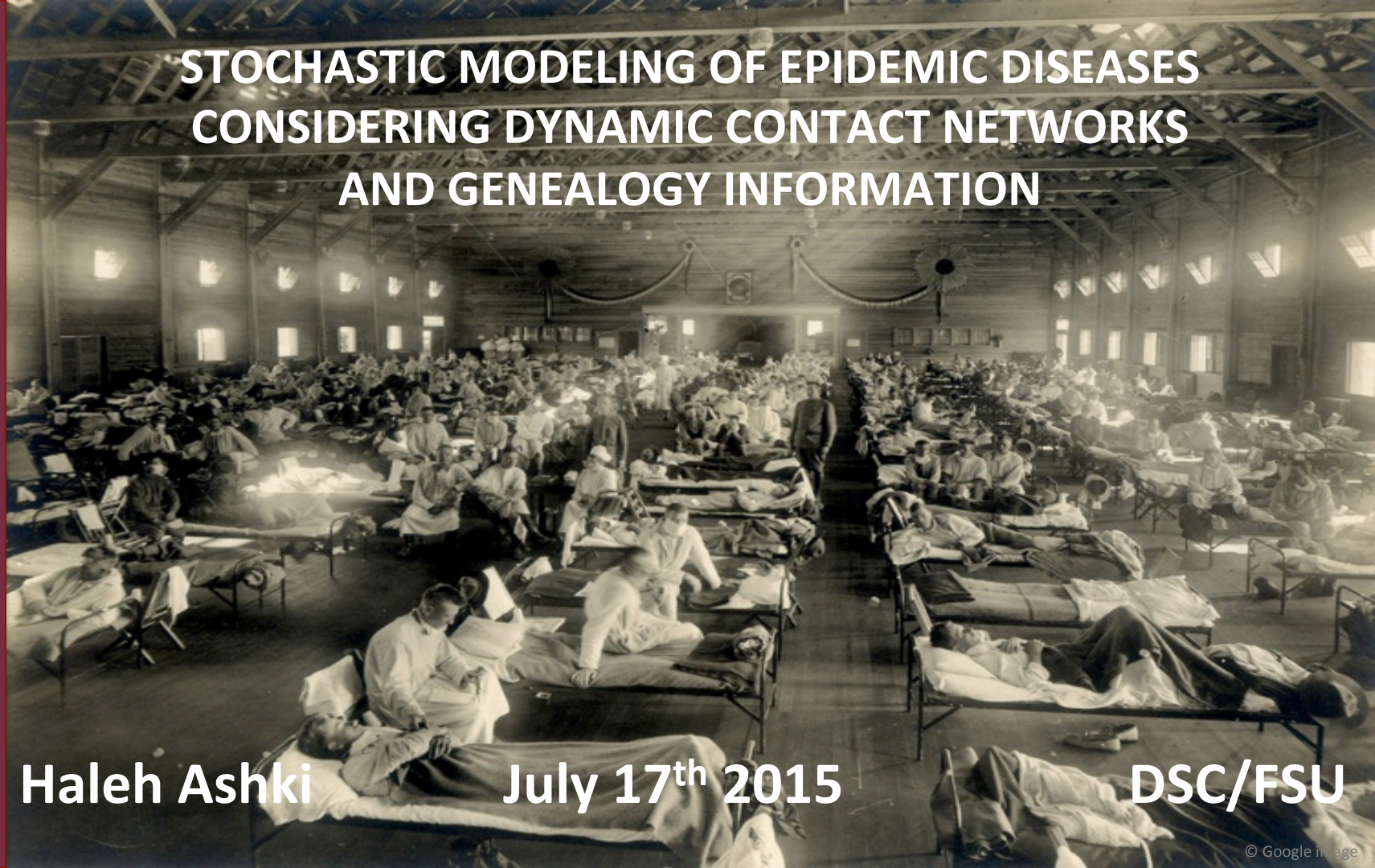
STOCHASTIC MODELING OF EPIDEMIC DISEASES CONSIDERING DYNAMIC CONTACT NETWORKS AND GENEALOGY INFORMATION

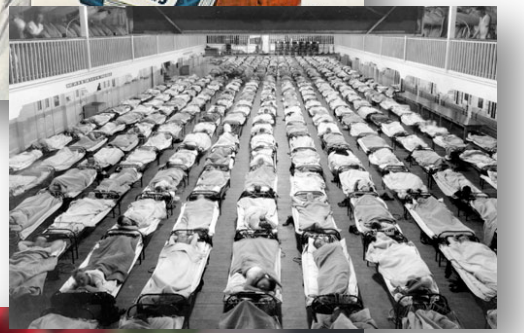


Haleh Ashki

July 17th 2015

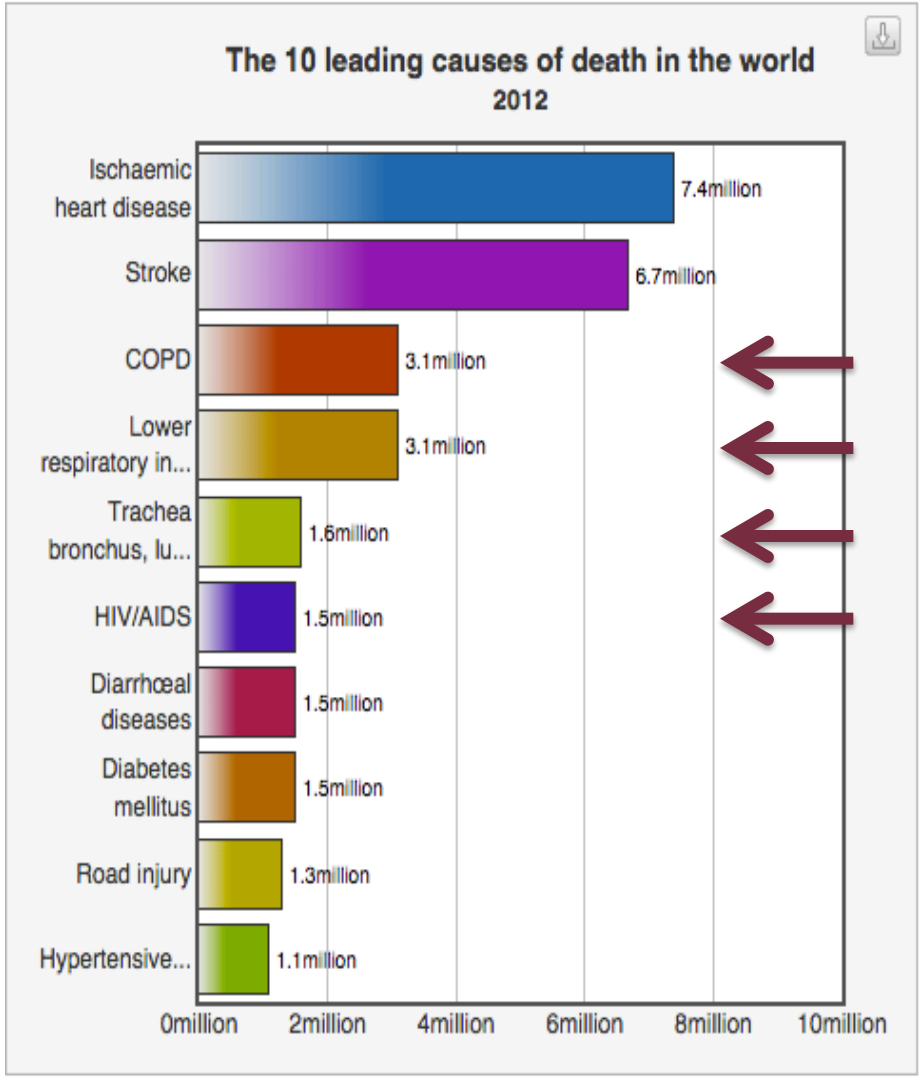
DSC/FSU

- Introduction
- Model Parameters
  - Epidemiological parameters
  - Contact Network
  - Genealogical Parameters
- Hidden Markov Model
- Parameter Estimation
  - Dynamics of The System
  - R0
  - Intervention methods



Infectious diseases

- Viruses are one of the sources for human diseases (Infectious diseases).
- Viruses can transmit from one host to other one.
- About 10,000 years ago, with the agricultural revolution and larger population groupings resulted in the first epidemiological transition.
- Infectious diseases are among top 10 of causes of human death.

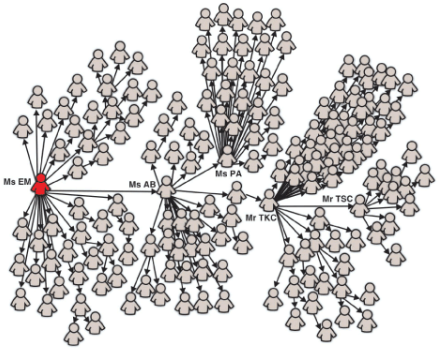


Infection

- An infection results when a pathogen **invades** and begins **growing** within a host.
- The first step is the **transmission** into the host.

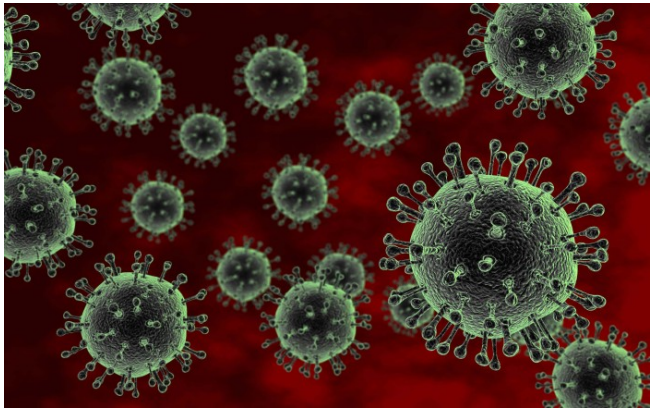
- Virus will be transmitted by **contact** of an infected individual to a susceptible individual for a **long enough** period of time.

Epidemic



In an epidemic, the infection **escapes** the initial group of cases into the community and results in **Epidemic** incidence of the disease.

Spanish flu
H1N1
HIV/AIDS
Ebola



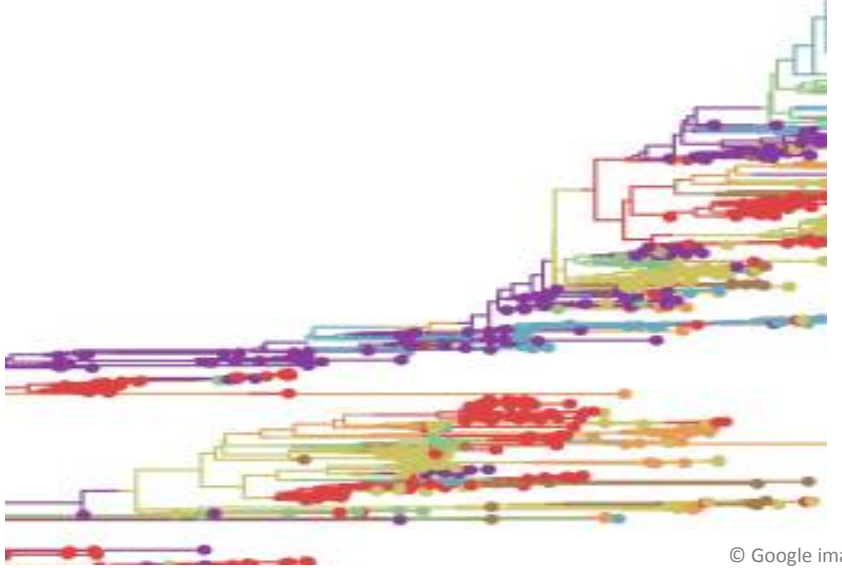
© Google image

Parameters in epidemic modeling

Epidemiology parameters

Contact Network

Genealogy information



Epidemiology parameters

- Transition rate: **β**

  is an average probability that an infectious individual will transmit the disease to a susceptible individual.

- Recovery rate: γ

  is a rate that infected individual gets better or dies.

- Basic reproductive number: *R0*

- *R0* it is the number of infected individuals one already infected generates on average during the time of being infected.

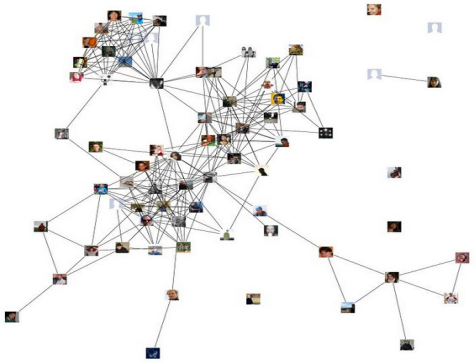
- *R0* < 1 the epidemic is not self-sustainable, *R0* > 1 an epidemic is possible

Diseases	R0
Measles	12-18
HIV/AIDS	2-5
SARS	2-5
Influenza	2-3
Ebola	1.5-2.5

Contact Network

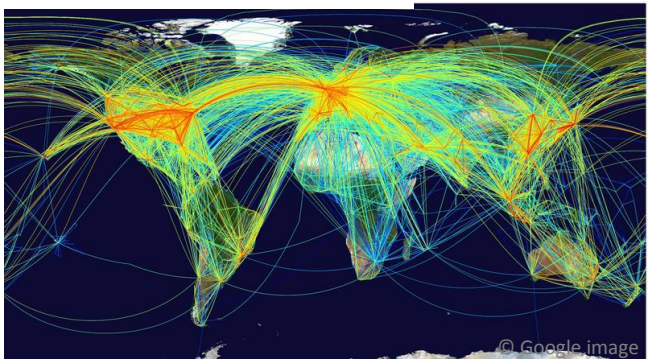
- The Internet, the World Wide Web, Facebook, and eBay.

- Each of us is also part of a bigger network, the **social contact network**.



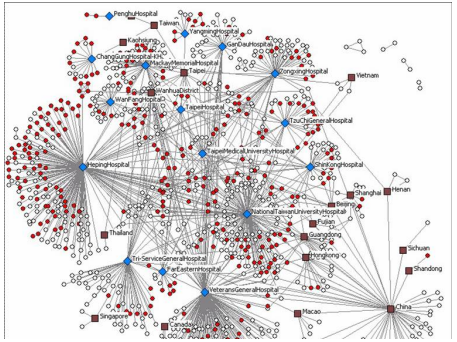
- **Nodes** in social networks represent: individuals, farms, cities, courtiers, etc.

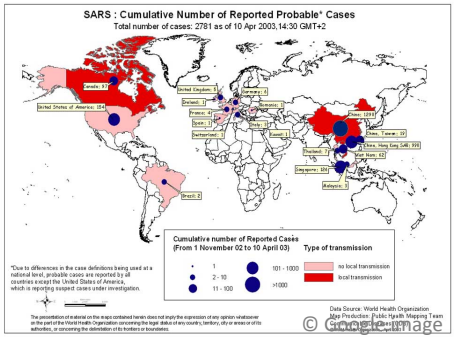
- **links** in social networks represent: friendship, a sexual relationship, a past communication, a co-authorship, or a citation.



why is the contact network important in epidemic modeling?

- Severe Acute Respiratory Syndrome (SARS) in 2002 and 2003
- **R**0 estimated by mathematical epidemiologist was in the range of **2.2 to 3.6**, much higher that 1.0 (the epidemic threshold).
- Despite of this estimate, SARS has **not emerged as a global pandemic**.
- The **discrepancy** between the estimated **R**0 and the observed epidemiology comes from the **model assumptions**:
- The models assumed that the network is **fully mixed** but this assumption did not hold. The rate of connection is not the same for every individual.

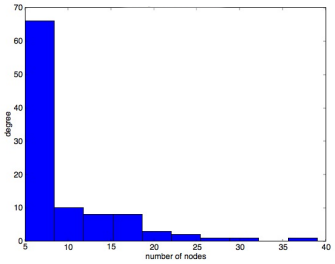


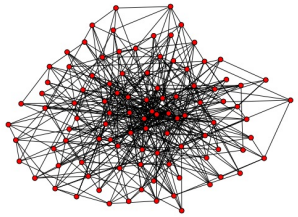


Static vs Dynamic network

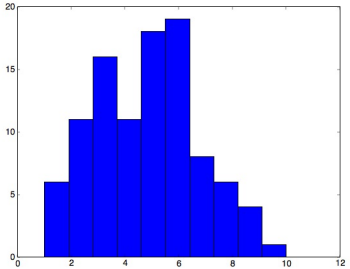
Static network: Number of nodes and connections are static.
Dynamic network: Number of nodes, number of edges, connections and durations of each connection can vary over time.

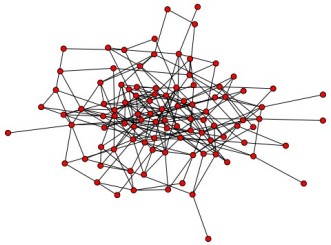
BA: Barabási–Albert
Power-law distribution



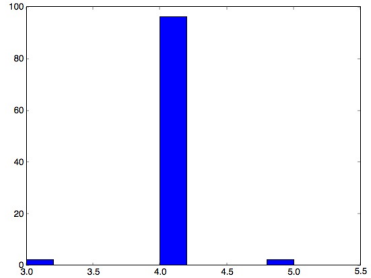


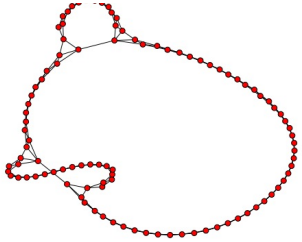
ER: Erdős–Rényi
Binomial distribution





WS: Watts-Strogatz
Dirac distribution



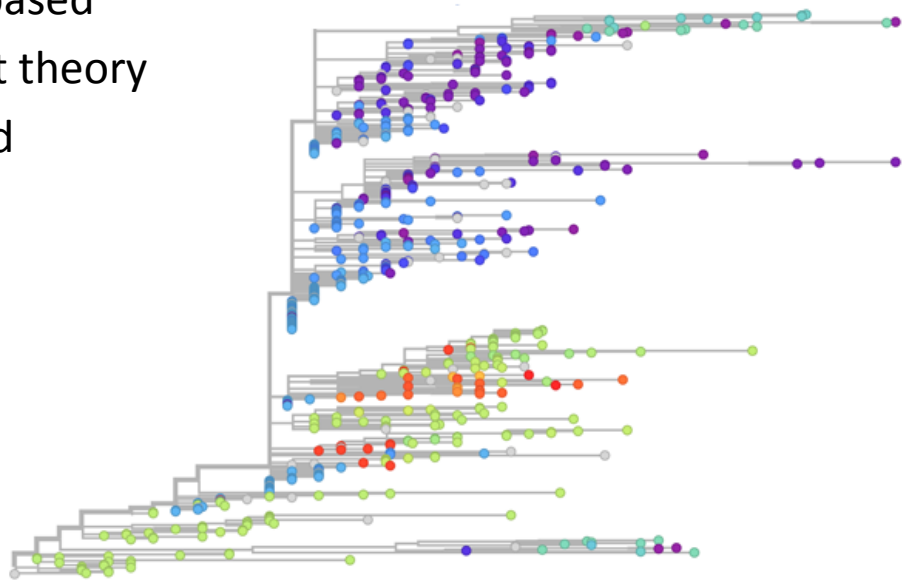


Genealogy information

- **genealogy information** has been used to estimate epidemic parameters very recently.

- Sequencing the genome of pathogens has become more **affordable and faster.**

- Reconstruct the **evolutionary tree** based on the data and then use Coalescent theory to estimate the **origin of disease** and **epidemic parameters**.



How to get the dynamic networks from data sources?

1.Samples pathogen:
DNA sequences

2.Location, time of sampled data:
Network of sampled data

3.Epidemiological Information about disease:
Transition rate, recovery rate, incubation time, etc.

Set of networks showing the connectivity and mobility between hosts.

Hidden Markov Model (HMM)

Is composed of (S, K, Π, A, B)

$S = 1, ..., N$ Set of hidden states

$K = k_1, ..., k_M$ output alphabet of observation states

$\Pi = \pi_i$ Initial state distribution

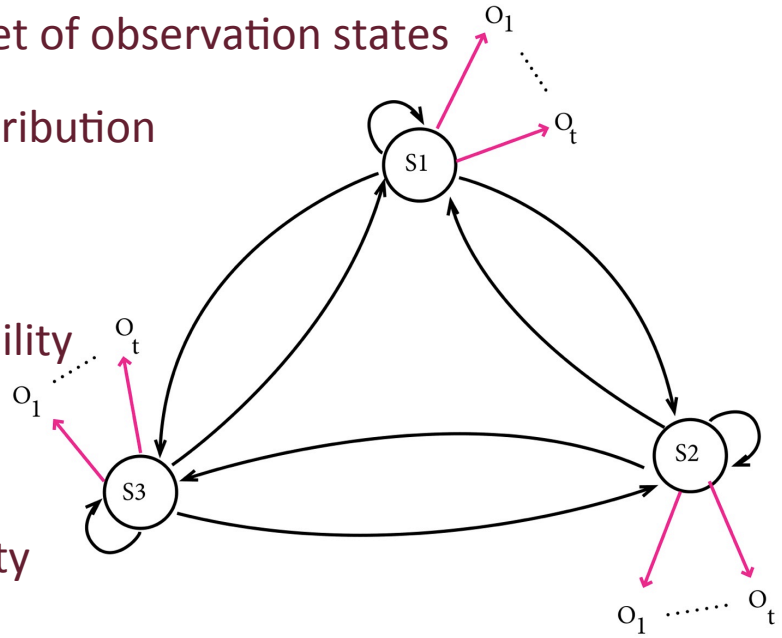
$\pi_i = P(s_1 = i)$

$A = a_{ij}$ State transition probability

$a_{ij} = P(s_{t+1}|s_t) 1 \leq i, j \leq N$

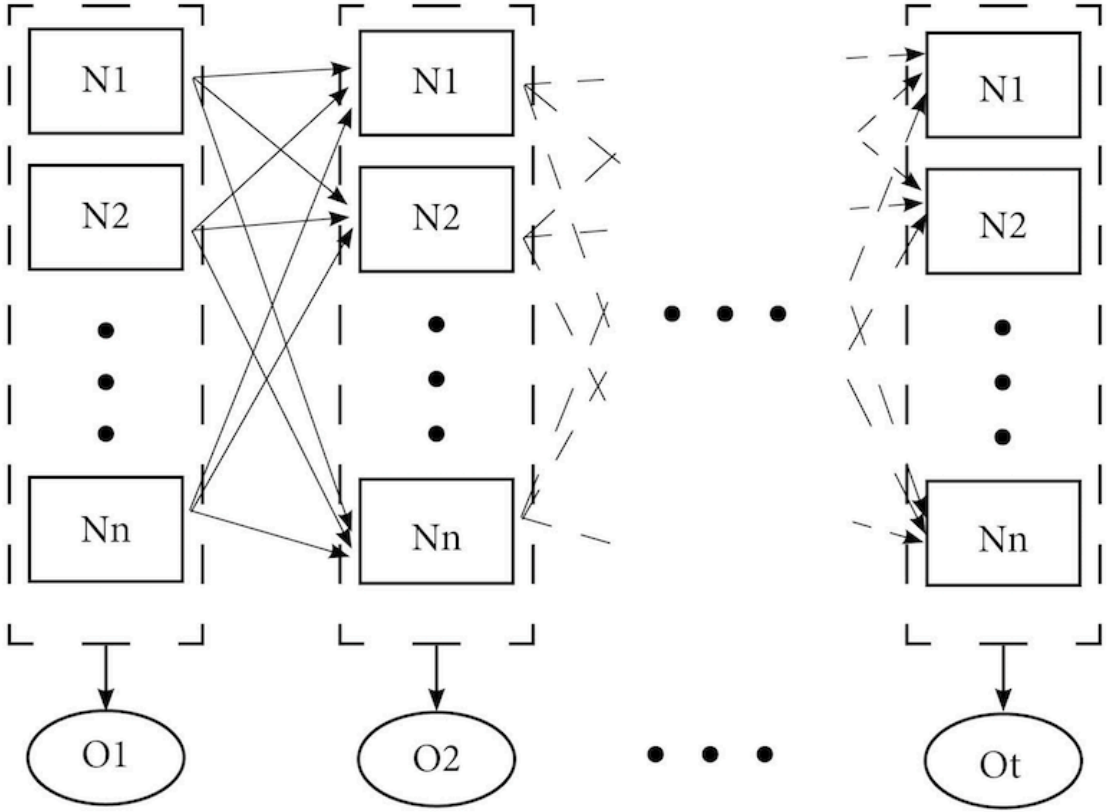
$B = bj(o_t)$ Emission probability

$bj(o_t) = P(o_t|s_t = j)$



HMM for my model

- The transition probabilities, and from which state an observation is generated are all unknown.
- N: hidden states: set of static networks
- O: Observed states: Genome data of samples pathogens



HMM formulation

Initial Probability

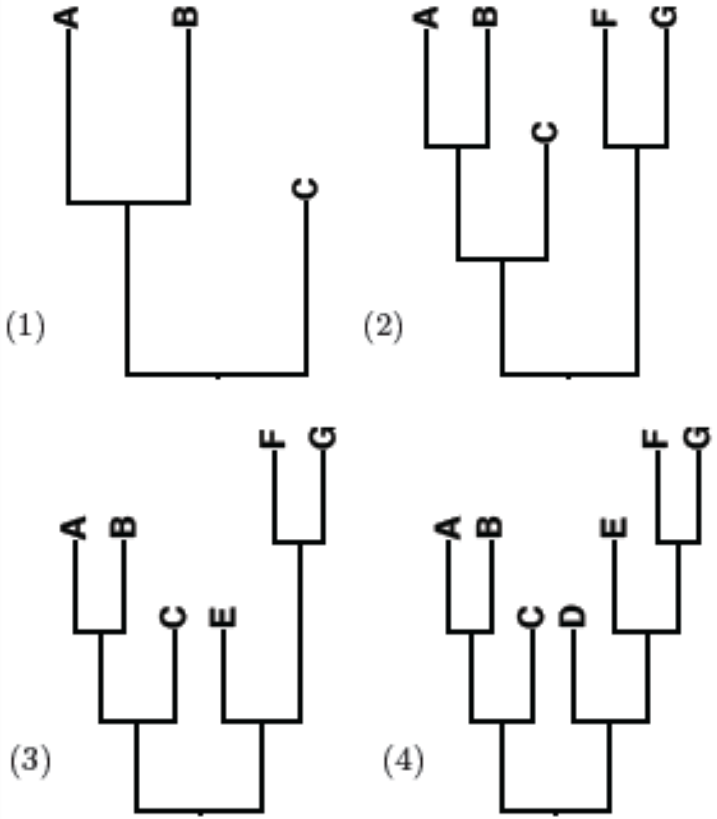
Transition Probability

$$P(S_{1:T}, O_{1:T}) = P(S_1)P(O_1|S_1)\prod_{t=2}^{T} P(S_t|S_{t-1})P(O_t|S_t)$$

Emission Probability

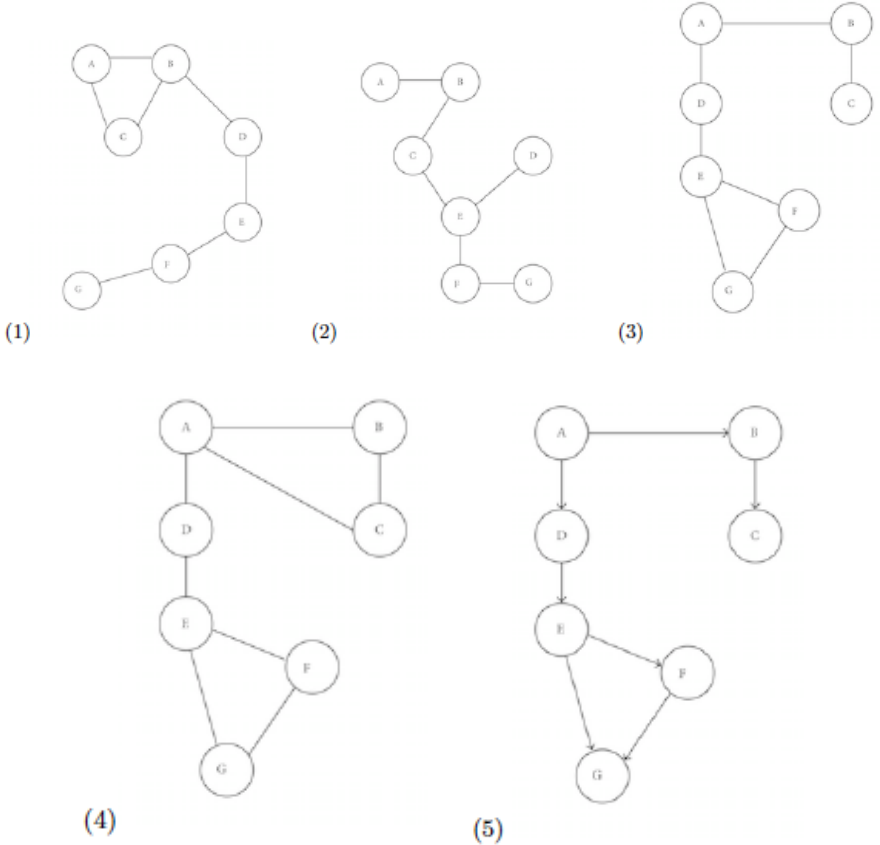
Observed data:

DNA sequence of sample pathogen →
Evolutionary tree for each time step



Hidden data:

Location and time of sampled pathogen →
Set of arbitrary contact networks

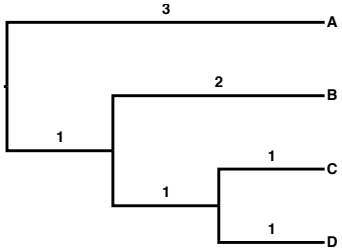


Hidden Markov Model:

My Formulation

Transforming inputs in a same format: Matrix

- **Observed data**: evolutionary tree T
- Patristic Distance matrix: the number of mutational differences between two tips on a tree.



	A	B	C	D
A	0	6	6	6
B	6	0	4	4
C	6	4	0	2
D	6	4	2	0

- **Hidden data**

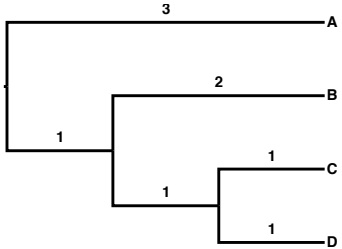
Constructing Matrix *G*

$\mathbf{G_{ij}}$ = Normalized (amount of divergence between i and j in tree T)

	A	B	C	D
A	0	1	1	1
B	1	0	0.66	0.66
C	1	0.66	0	0.33
D	1	0.66	0.33	0

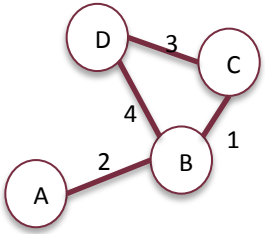
Transforming input in a same format: Matrix

- **Observed data**: evolutionary tree T
- Patristic Distance matrix: the number of mutational differences between two tips on a tree.



	A	B	C	D
A	0	6	6	6
B	6	0	4	4
C	6	4	0	2
D	6	4	2	0

- **Hidden data**: Network M
- Distance matrix: shortest path length among each two nodes in network.



	A	B	C	D
A	0	2	3	6
B	2	0	1	4
C	3	1	0	3
D	6	4	3	0

Constructing Matrix *G*

$\mathbf{G_{ij}}$ = Normalized (amount of divergence between i and j in tree T)

	A	B	C	D
A	0	1	1	1
B	1	0	0.66	0.66
C	1	0.66	0	0.33
D	1	0.66	0.33	0

Constructing Matrix *N*

$\mathbf{N_{ij}}$ = Normalized (shortest path between i and j in network M)

	A	B	C	D
A	0	0.33	0.50	1.00
B	0.33	0	0.16	0.66
C	0.50	0.16	0	0.50
D	1.00	0.66	0.50	0

Hidden Markov Model:

My Formulation

Emission probability

$$P(O_t|S_t) = P(\text{genealogy tree}|\text{network structure})$$

Euclidean distance between two matrices:

$$A = (a_{ij});$$

$$a_{ij} = \|N_i - G_j\|_2^2$$

Emission probability matrix

$$P_{ij} = \frac{A_{ij}}{\|\sum_j A_{ij}\|_1}$$

The distance value is : 1.12224

To make the emission probability matrix we need the distance value between all networks and trees.

Three Fundamental Problems HMMs can solve

1. Given a model μ = (A,B,Π), and an observation sequence $O = (o_1,...,o_T)$, it efficiently computes the probability of the observation sequence given the model : $P(O|\mu)$

2. Given a model μ and the observation sequence *O*, calculates the state sequence (1, ..., N) that best ”explains” the observations.

Viterbi algorithm

3. Given an observation sequence *O*, and a space of possible models, it adjusts the parameters so as to find the model μ that maximizes $P(O|\mu)$.

Calculates the model parameter that best describes the observation sequence:

Baum-Welch algorithm

Viterbi algorithm $\underset{S'}{\operatorname{argmax}}\ P(S', O|\mu)$

$$\delta_j(t) = \max_{s_1 \ldots s_{t-1}} P(s_1, \ldots, s_{t-1}, o_1, \ldots, o_{t-1}, s_t = j|\mu)$$

Initialization:

$$\delta_i(1) = \pi_i b_i(o_1)$$

$$\psi_i(1) = 0$$

Induction:

$$\delta_i(t) = b_{ijo_t} \max \delta_i(t-1) a_{ij}$$

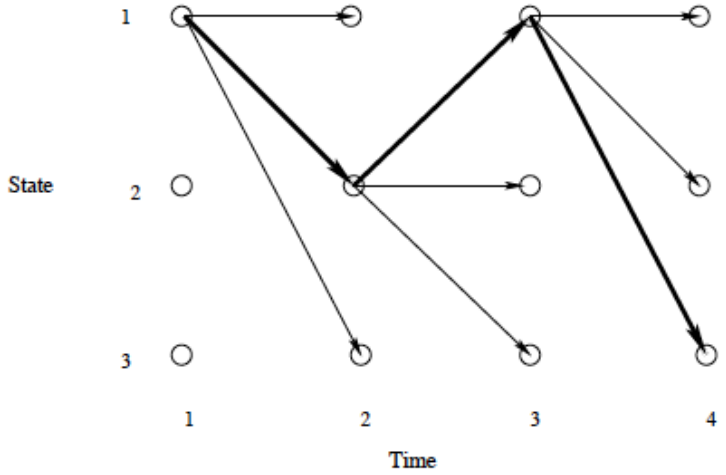
$$\psi_i(1) = \operatorname{argmax}[\delta_i(t-1) a_{ij}]$$

Best path:

$$s_t^* = \psi_{t+1}(s_{t+1}^*)$$

Where

$$s_T^* = \operatorname{argmax}[\delta_i(T)]$$



Baum-Welch Algorithm: $\underset{\mu}{\operatorname{argmax}}\ P(O|\mu)$

Forward procedure

$$\alpha_i(t) = P(o_1, ..., o_{t-1}, s_t = i|\mu)$$

$$\alpha_i(1) = \pi_i$$

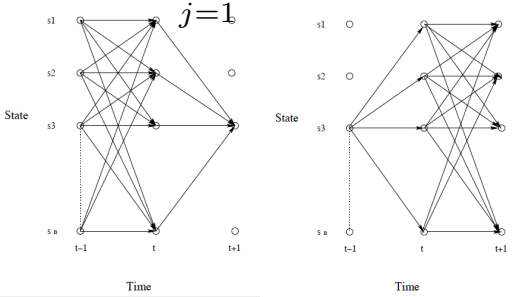
$$\alpha_j(t+1) = \sum_{i=1}^{N} \alpha_i(t) a_{ij} b_j(o_t)$$

Backward procedure

$$\beta_i(t) = P(o_{t+1}, ..., o_T|s_t = i, \mu)$$

$$\beta_i(T) = 1$$

$$\beta_i(t) = \sum_{j=1}^{N} a_{ij} b_j(o_t) \beta_j(t+1)$$



Parameter estimation

$$p_t(ij) = \frac{\alpha_i(t) a_{ij} b_j(o_t) \beta_j(t+1)}{\sum_{m=1}^{N} \sum_{n=1}^{N} \alpha_m(t) a_{mn} b_n(o_t) \beta_n(t+1)}$$

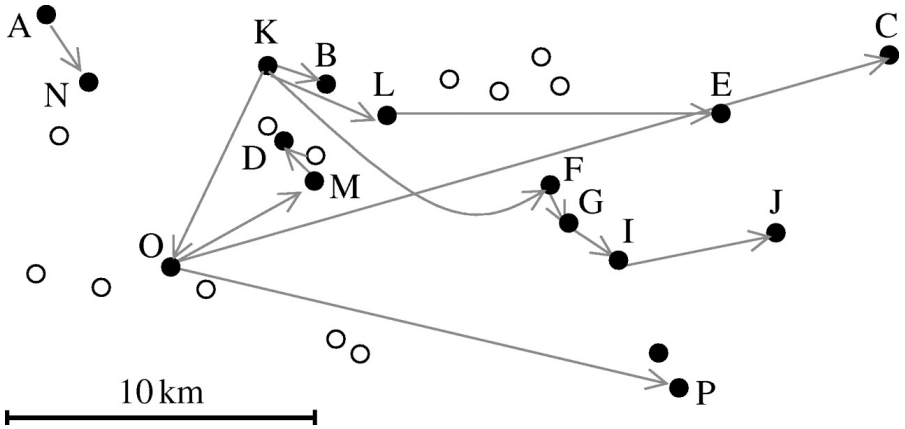
These variables are used to re-estimate the model parameters:

$$a'_{ij} = \frac{\sum_{t=1}^{T} p_t(ij)}{\sum_{t=1}^{T} \gamma_t}$$

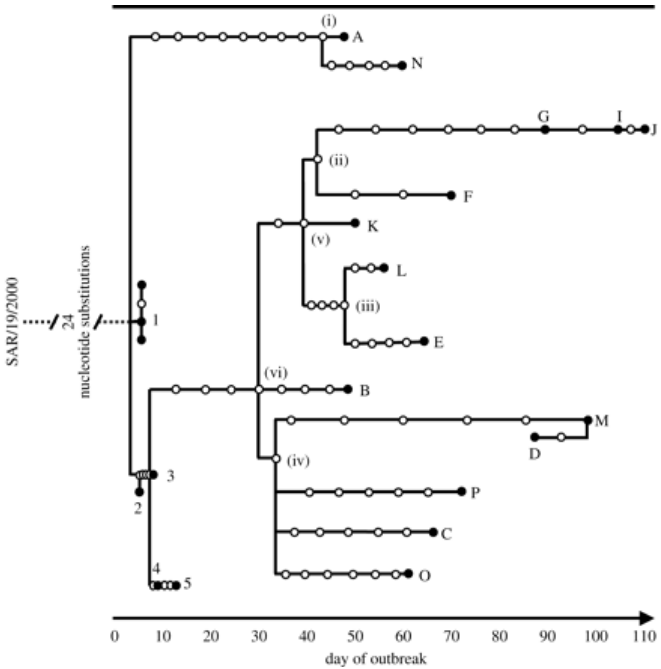
$$b'_{ink} = \frac{\sum_{t}^{k} p_t(ij)}{\sum_{t=1}^{T} \gamma_t}$$

Foot-and-mouth disease UK 2001

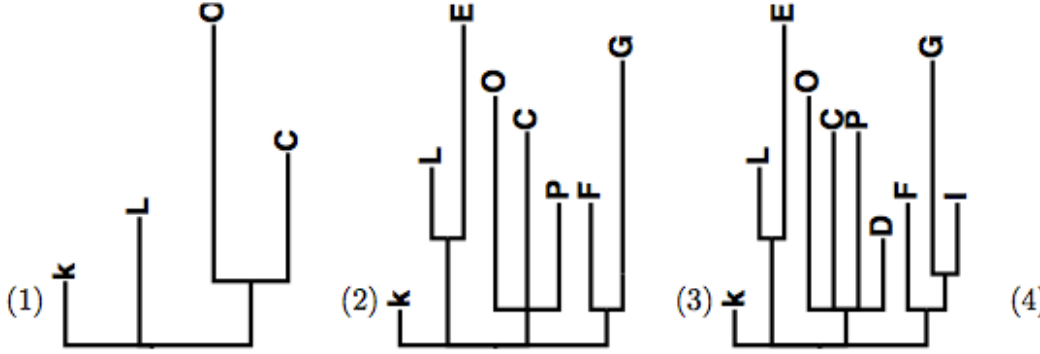
Map showing the spatial relationship of 15 infected farms confirmed by laboratory testing.

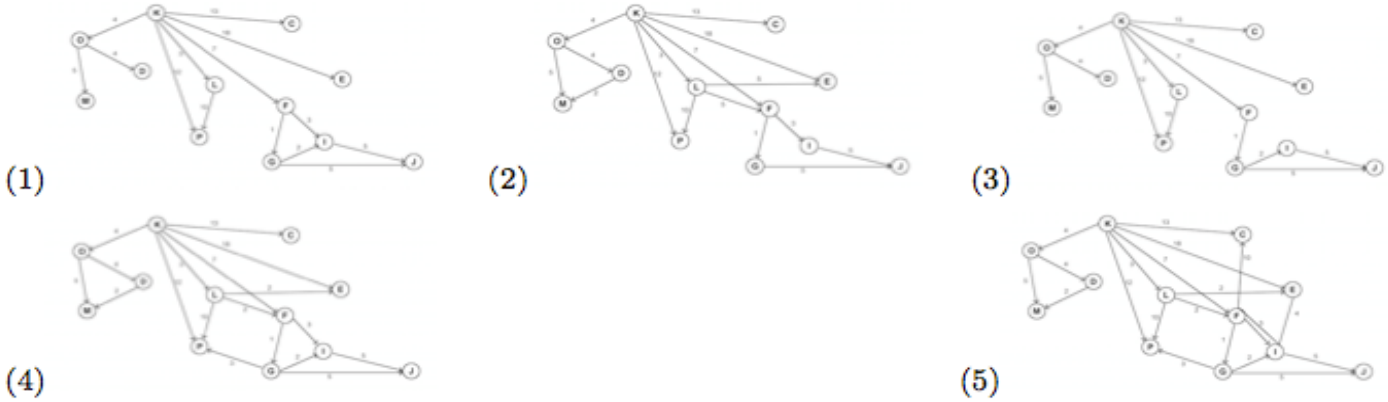


Statistical parsimony analysis of 22 FMDV complete genome sequences



Tree and networks constructed from MFDV data





Emission probabilities

Hidden States / Observed States	1	2	3	4
1	0.6557	0.7444	**0.7839**	0.8161
2	**0.6680**	**0.7476**	0.7595	0.8249
3	0.6542	0.7415	0.7818	0.8226
4	0.6531	0.7473	0.7763	0.8233
5	0.6628	0.7409	0.7602	**0.8361**

The best network that describe the genealogy tree 4, the tree containing all taxa, is network 5

Experiment

Experiment	Observed Sequence	Best Hidden State Sequence
First	(4,4,4,4,4,4,4,4,4,4)	(5,5,5,5,5,5,5,5,5,5)
Second	(1,1,1,2,2,2,2,3,3,3,3,4,4,4,4,4)	(2,2,2,2,2,2,2,1,1,1,1,5,5,5,5,5)

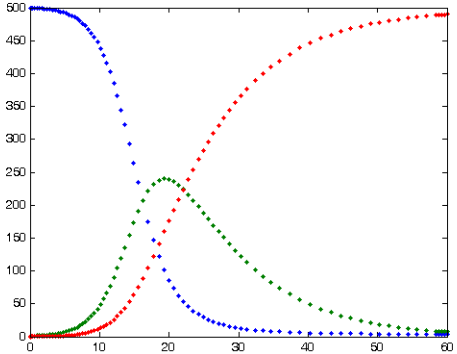
First: there is only one tree as observed state.

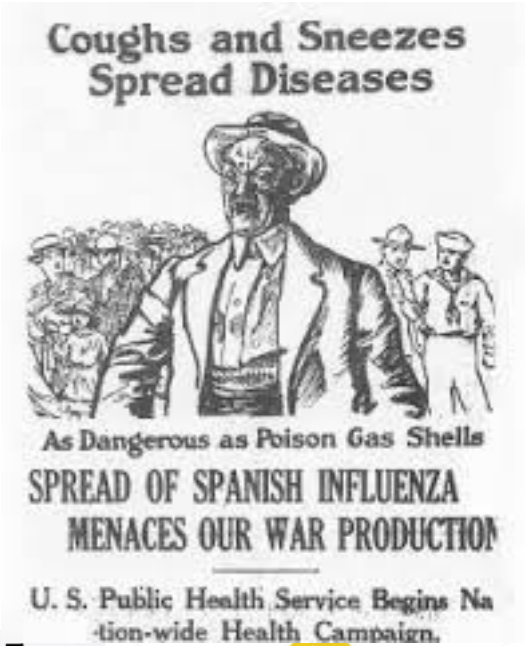
Second: there are more observed sequences for a longer time period.

© Google image

What to estimate

- **Dynamics of system**
- **R0**
- **Intervention methods**





Dynamics of system

Epidemic model: SIR

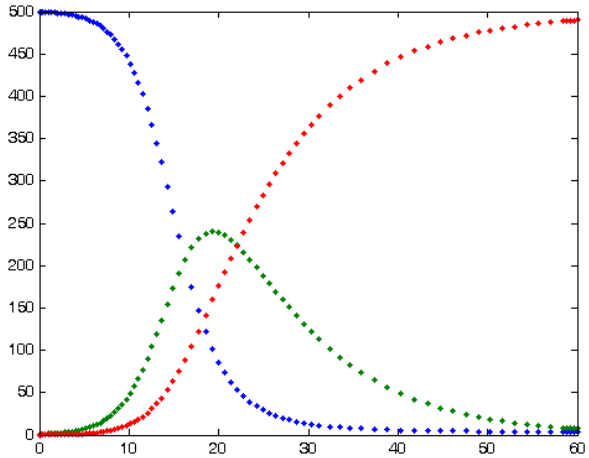
Susceptible are individuals who are not yet infected and are susceptible to disease.

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

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

SIR model formulation

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



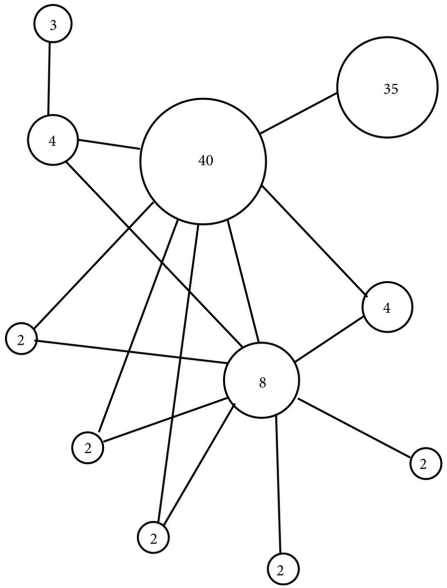
β: Transition rate

γ: Recovery rate

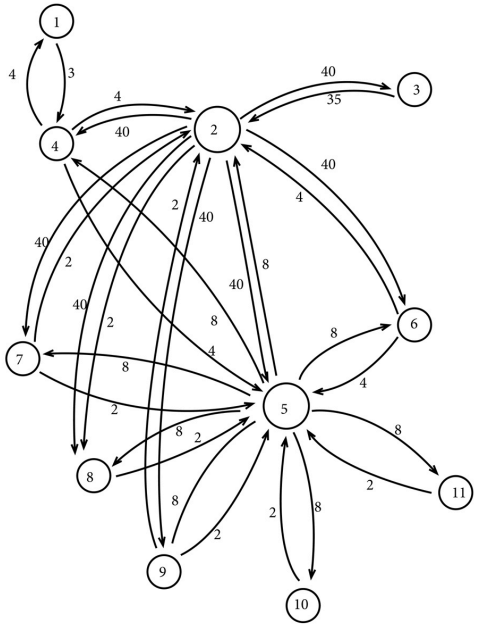
Blue=Susceptible
Green=Infected
Red=Recovered

My model: SIR model + dynamic contact network

- **Connections** among individuals: Adjacency matrix
- **Transition rate:** Weight on nodes



Transforming node weight to **edge weight**



Weighted adjacency matrix

0	0	3	0	0	0	0	0	0	0
0	40	40	40	40	40	40	40	0	0
35	0	0	0	0	0	0	0	0	0
4	0	0	4	0	0	0	0	0	0
8	0	8	0	8	8	8	8	8	8
4	0	0	4	0	0	0	0	0	0
2	0	0	2	0	0	0	0	0	0
2	0	0	2	0	0	0	0	0	0
2	0	0	2	0	0	0	0	0	0
0	0	0	2	0	0	0	0	0	0
0	0	0	2	0	0	0	0	0	0

Model and Parameters

Three **state vectors** contains the probability of being in each state of all individuals in the population.

$$P_I = [P_{I,1}, P_{I,2}, ..., P_{I,N}]$$
$$P_S = [P_{S,1}, P_{S,2}, ..., P_{S,N}]$$
$$P_R = [P_{R,1}, P_{R,2}, ..., P_{R,N}]$$
$$P_{R,i} = 1 - (P_{I,i} + P_{S,i})$$

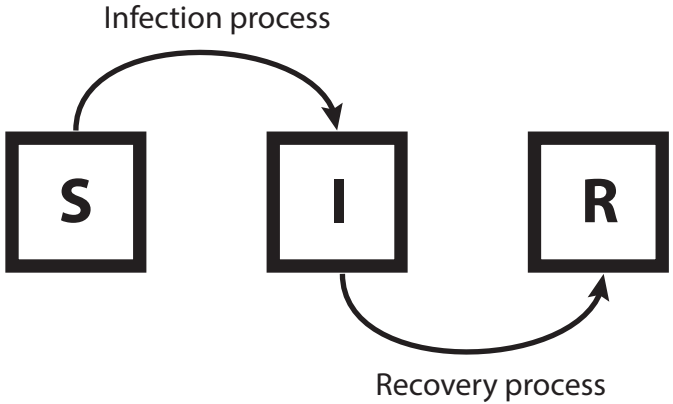
The initial states:

PI(0) = (0,0,1,0,0,1,0,...) has few ones and the rest are zeros.

PS(0) = (1,1,0,1,1,0,1,...) has the value 0 for those who are initially infected and the rest are ones.

PR(0) = (0, 0, 0, ...) is a vector of size n of zeros.

Two processes:



$$P_S = 1 - P_I$$
$$P_I = 1 - P_R$$

Formulation

$$\frac{dP_k}{dt}(t) = \sum_j C_{jk} P_j(t)$$

Kolmogorov forward equation

$$\frac{d(P_I)}{dt} = CP_I$$

$$\frac{d(P_I)}{dt} = (\sum_i \sum_j \beta_i C_{ij} (P_I)_j)$$

$$\frac{d(P_R)}{dt} = -\gamma_i P_R$$

$$\frac{d(P_I)}{dt} = \delta(\sum_i \sum_j \beta_i C_{ij} (P_I)_j) + (\delta - 1)(\gamma_i P_I)$$

$$\delta = \begin{cases} 1 & \text{if } \frac{PI}{\tau_{SI}} \leqslant 1, \\ 0 & \text{otherwise,} \end{cases}$$

β_i: transition rate for individual i

C_{ij}: connection between individual i an j

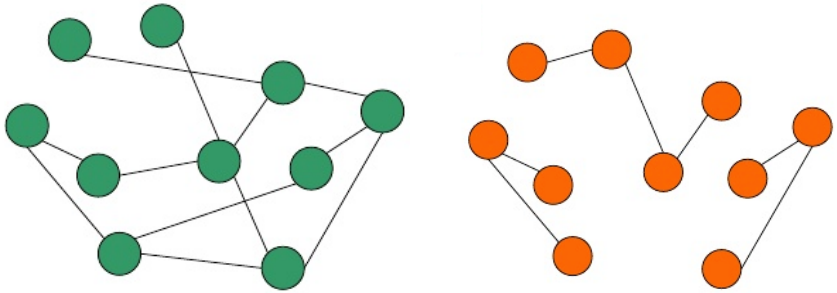
γ_i: recovery rate for individual i

τ_{SI} is an infection threshold or the duration of infectiousness. The range is between 0 and 1. its based on life history of the virus or disease incubation time.

Experiment on Static and Dynamic network

Dynamic network: Day and Night

two static networks are combined with different connectivity and this pattern is repeated for some time. This would represent a standard social network, where we go to work during the day and interact with others, and during the night we only interact with our friend and families.



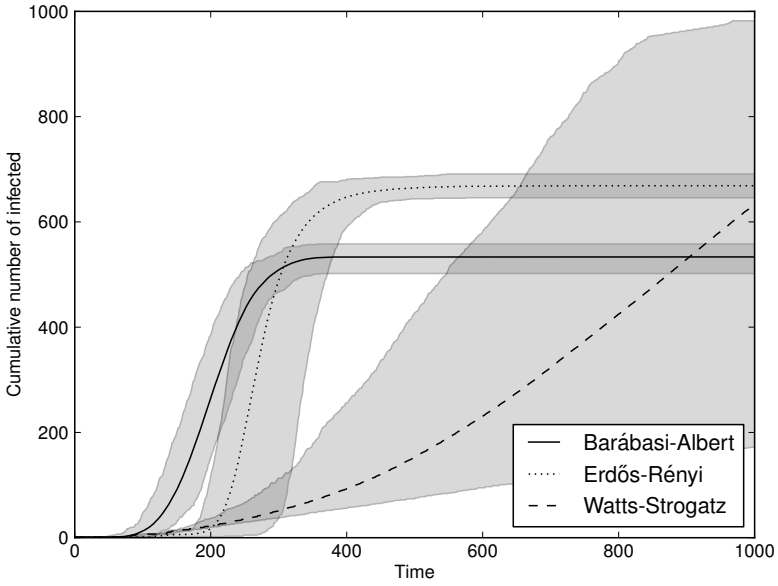
Mean degree $\langle k \rangle = \sum_k kp(k)$

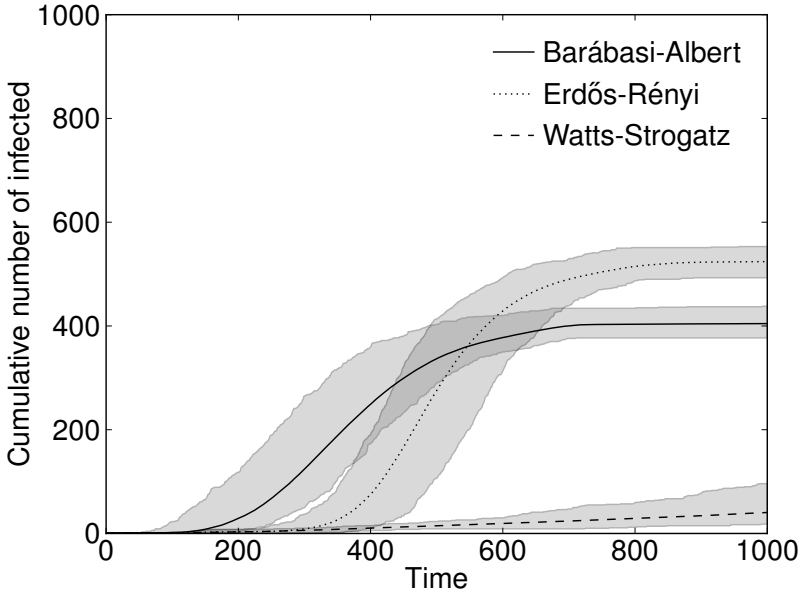
is the mean of the probability distribution of nodes degree.

Static network :
Mean degree= 6

Dynamic network :
Mean degree for Day network = 9.59-10
Mean degree for Night network = 2.5-3
Average mean degree =6.5

Static vs Dynamic result





- **similar** mean degree.
- The course of an epidemic is different in these two simulations.
- Fewer infected individuals in the dynamic network.
- The epidemic **peaks later** and less abruptly in the dynamic network.

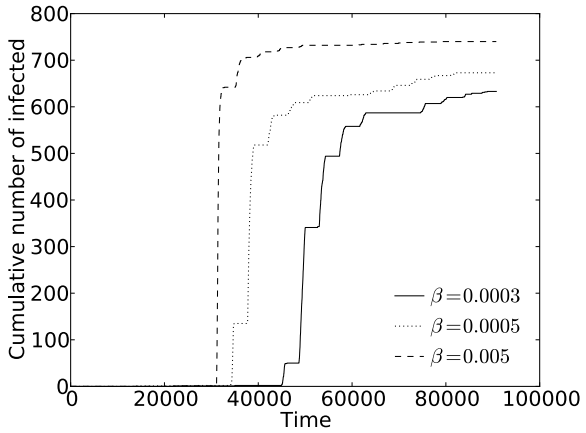
Experiment on BA, ER, WS
Number of nodes=1000
β=0.005
γ= 0.05
Average of 500 runs

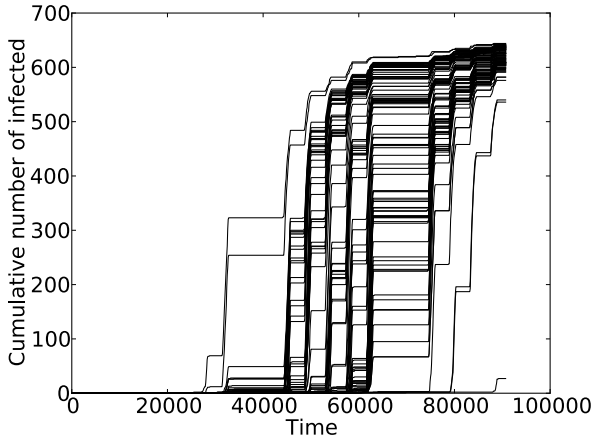
Experiment on real data

- Potential challenge is the scarcity of **empirical data** for dynamic contact networks.
- Recording each person's contacts for a **long period** of time in a **large society** is a time consuming process.
- Wearable active Radio-Frequency Identification Devices (RFID) are able to detect face-to-face contact among individuals with a spatial resolution of about 1.5 meters, and a time resolution of 20 seconds.

Dynamic network was used from real live data collected in high school of 789 individuals.

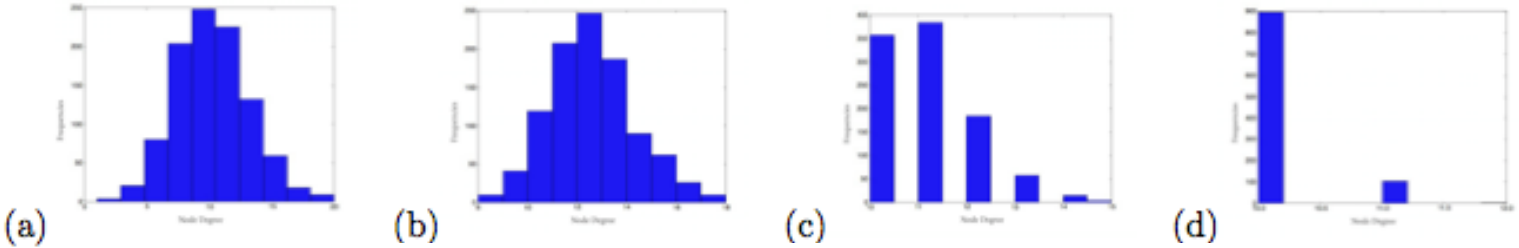
data were collected from 6:00 am until the end of the school day.



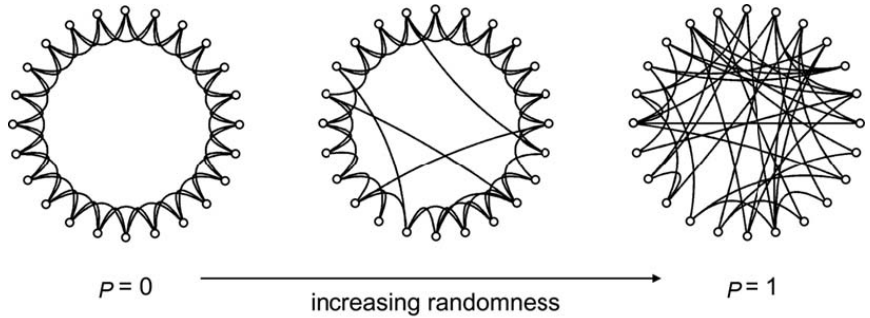


R0 estimation on Benchmark networks

ER and WS networks with different rewiring probability. By increasing the rewiring probability the randomness property of network is increased.



(a) ER (b) WS $p = 0.5$ (c) WS $p = 0.1$ (d) WS $p = 0.01$

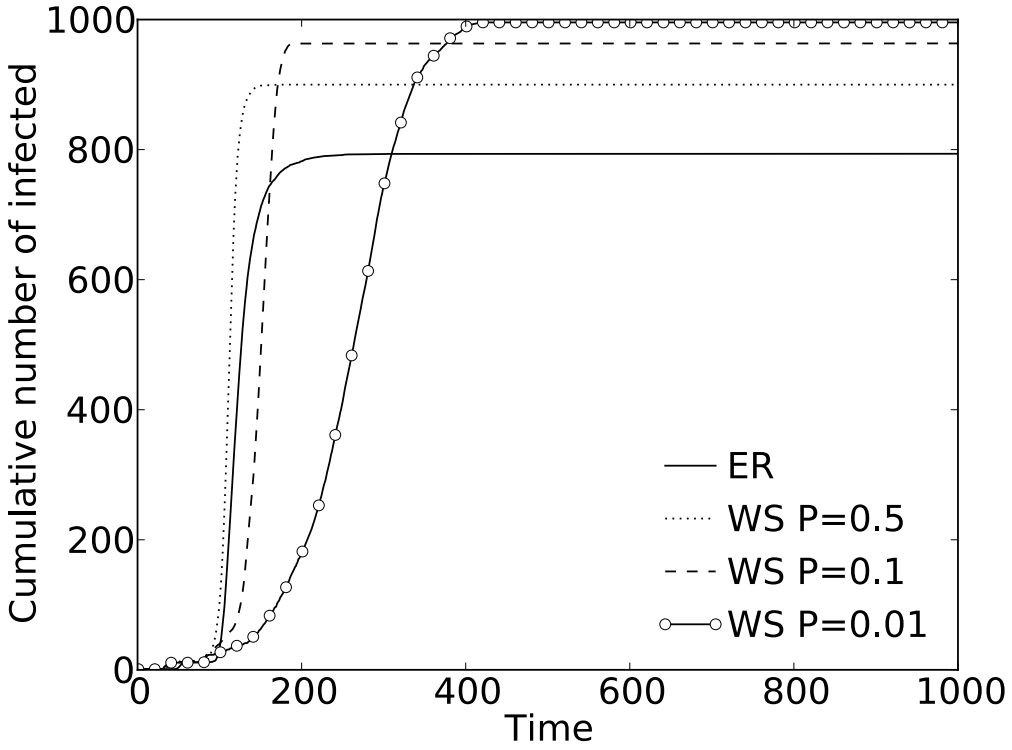


Parameter Estimation:

R0

Epidemic behavior in benchmark network

- Mean degree and mean squa degree are the same for all networks.
- Most infected individuals in WS, P=0.01
- Fewer infected individuals in ER



Parameter Estimation:

R0

My Formulation

- Using Eigenvalue propertied of adjacency matrix.
- Spectral Gap (SG): the difference between largest and second largest eigenvalue.

$$R0^* = (1 + \frac{1}{SG}) \frac{\beta}{\beta + \gamma} \frac{\langle k^2 \rangle - \langle k \rangle}{\langle k \rangle}$$

Parameter	ER	WS p=0.5	WS p=0.1	WS p=0.01
$\langle k \rangle$	10.18	12.02	10.95	10.10
$\langle k^2 \rangle$	103.18	132.62	109.93	92.49
Spectral gap	4.5	3	0.867	0.09
Max number of infected	793.6	900	963.3	995.60
Time of early stage	109.4	102.30	125.3	173.20
$R0^*$	1.0049	1.2040	1.75	8.9805

Intervention methods

- Mitigating or preventing the spread of infectious diseases is the ultimate goal of infectious disease epidemiology.
- Intervention method control the spread of disease and reduce the epidemic threshold.
- Intervention methods: Pharmaceutical (PI) vs Non-Pharmaceutical (NPI).
- PI: using antivirals, antibiotics, and vaccinations.
- NPI: change the social network structure in a population without administering any drugs such as closure of public spaces, quarantine.

My Formulation

- Finding important nodes with high connectivity and transition rate
- Eigenvalue centrality: Eigenvector of dominant eigenvalue gives the order of important nodes
- works for unweighted adjacency matrix
- Transition rates are the edge weighs on directed contact network
- PageRank method : Eigenvalue problem for weighted matrix

$$R = (\frac{1-d}{n}E + dM)R$$

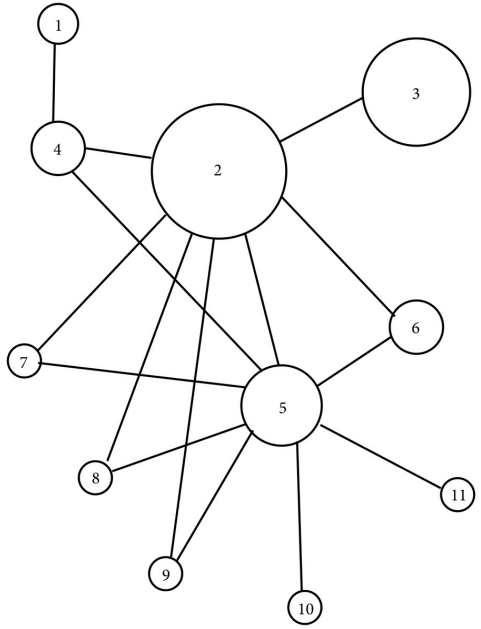
$$R = \hat{M}R$$

$$\hat{M}R = 1R$$

adjacency matrix M is calculated by dividing column i by the number of out-links li.
E is a n × n matrix full of 1s

To get the largest eigenvalue the **Power method** with complexity $O(n^2)$ is used.

Experiment



Method	Node orders
PageRank	2,3,5,4,6,7,8,9,1,10,11
Centrality	2,5,4,6,7,8,9,10,11,1,3

Intervention methods

Edge removal

- Changing the behavior of individuals.
- Decreasing or eliminating the connectivity of infected or highly at risk infectious persons in the population.
- Closing public places, schools, quarantine.

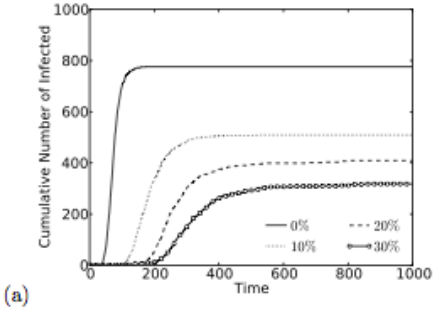
Immunity increasing

- Increasing the immunity of individuals and reducing their risk of being infected.
- A vaccinated person still visits the same places, and meets the same people.
- Vaccinated person will become less likely to get infected or will be less likely to infect other people.
- Using mask, vaccination and antiviral drugs.

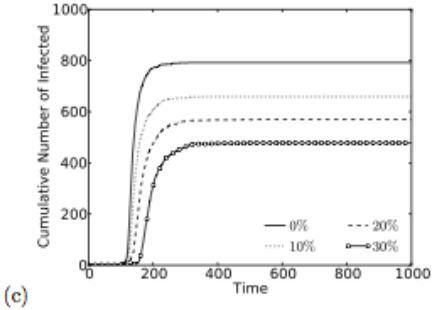
Result on edge removal

Method:

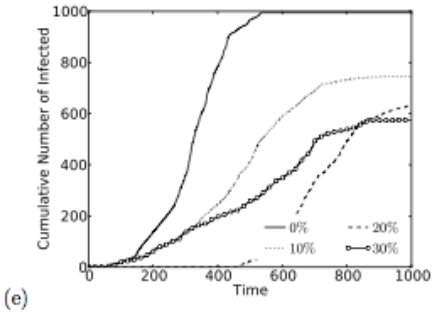
removes different percentages of targeted nodes at a particular time in the early stages. For example, here, I have disconnected 10%, 20% and 30% of targeted nodes.



(a)



(c)

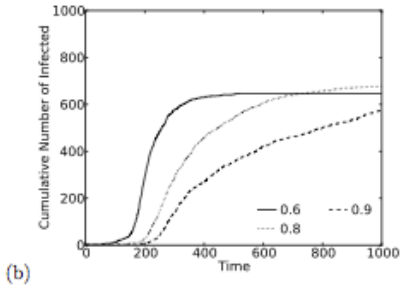


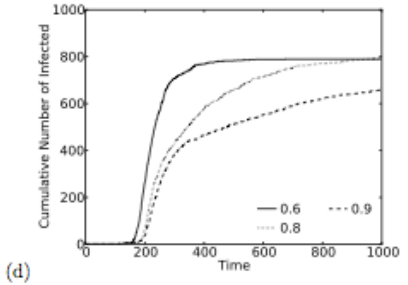
(e)

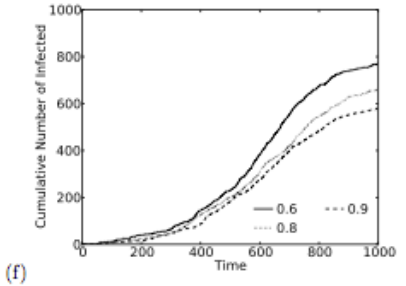
Result on immunity increasing

Method:

intervention methods by increasing the immunity rate on the same number of individuals in population. For example, here, I have used the same number of people to be immunized, 30 percent, but have increased the immunity rate to 80% and 90%.







Summary

- I developed a method that uses a Hidden Markov model to detect a time series of possible contact networks taking a time series of genealogies of the viruses and location data into account.

- I developed a method that simulate epidemics on dynamic contact network and showed how this approach can change our conclusions of the course of the epidemic. I show the effect of the dynamic network in intervention approaches that target important nodes of the network using an Eigenvalue/PageRank method.

- I developed an extension to current estimators of the reproduction number R0 that can be applied to networks where their randomness increase using the Eigenvalue property (Spectral Gap) of the weighted adjacency matrix.

Acknowledgements

- Dr.Beerli
- Committee members:
  - Dr.Wilgenbusch
  - Dr. Sachin Shanbhag
  - Dr. Dennis Slice
  - Dr. Alan Lemmon
  - Dr. Christopher Coutts
- Scientific Computing faculties and staff at Florida State University.
- Beerli and Wilgenbusch families.
- Last but not least: My family:
  - My lovely parents and sister (Abdollah, Fahimeh and Yassi)
  - My beloved husband and daughter (Mehdi and Hannan)