Aspects of Epidemic spreading: Euclidean networks and comparison with data

Parongama Sen

Department of Physics
Calcutta University, Kolkata, India
www.caluniv.ac.in

In collaboration with Abdul Khaleque; J Phys A (2013) and ongoing work

Introduction

Many processes can be regarded as spreading phenomena through contacts.

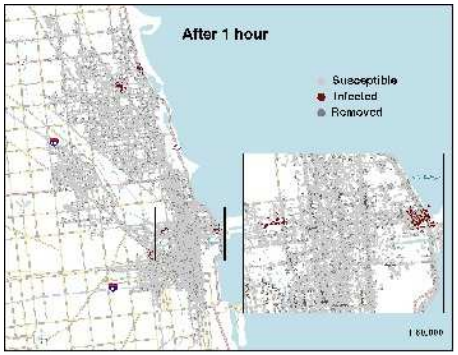
Examples range from epidemics, information, defects etc.

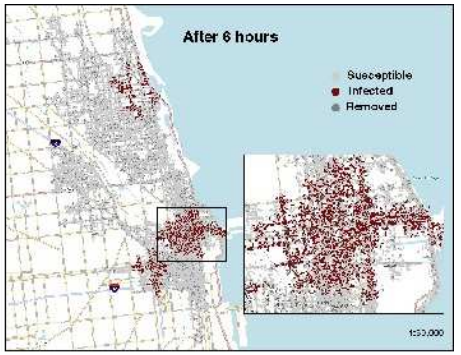
In a general process of contacts, the state of an 'unaffected' individual changes if one of the neighbour is 'infected'.

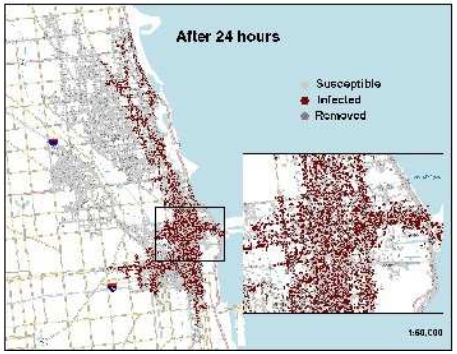
The study of spreading processes in the context of disease and epidemics dates back to almost a century

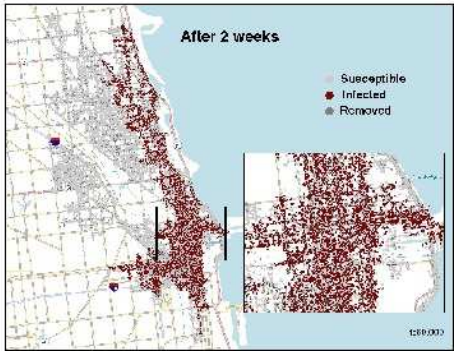
A more connected world has brought major consequences such as facilitate the spread of diseases all over the world to quickly become epidemics.

Examples are SAARS, HIV, Ebola etc.









Issues to be addressed in this talk:

- Threshold behaviour in epidemic models
- Effect of geographical factors
- Distribution of epidemic sizes and dynamics
- A study of the data for Ebola epidemic in West Africa
- Immunisation and tracing **
- Comparison with other cascading processes

Threshold behaviour in epidemic models

Mathematical modelling: Infected people infect their neighbours with a probability.

The occurrence of epidemic depends on the value of the infection probability.

When an epidemic takes place only above a minimum value of this probability, it is called a threshold behaviour.

That a threshold behaviour can exist in a model of disease propagation was already shown in some of the earliest works.

Models

The Susceptible-Infected (SI) model is the simplest of all the models, where the infected agents remain infected forever.

Ultimately all agents become infected for any value of the infection probability.

Hence it does not have a threshold behaviour.

Of the models showing a threshold behaviour, the Susceptible-Infected-Removed (SIR) and the Susceptible-Infected-Susceptible (SIS) models are the most studied and fundamental ones

Mean field theory for SIR and SIS models

SIR model

The SIR model is applicable for diseases which can be contracted only once.

S(t), I(t) and R(t) represent susceptible, infected and removed (recovered) fraction of population at time t with

$$S(t) + I(t) + R(t) = 1$$

They obey the following set of deterministic differential equations:

$$\frac{dS}{dt} = -q(k-1)IS, \quad (1)$$

$$\frac{dI}{dt} = -\mu I + q(k-1)IS, \quad (2)$$

$$\frac{dR}{dt} = \mu I. \quad (3)$$

q infection probability.

Note that these equations are deterministic.
Actual process will contain a degree of stochasticity.

Initial conditions: $R(0) = 0$; $S(0) \simeq 1$ and $I(0) \simeq 0$ when the initial infected population is very small.

Solution assuming $\mu = 1$:

$$S(t) = \exp[-q(k-1)R(t)].$$

Since at $t \to \infty$, $I = 0$,

$$R_\infty = 1 - \exp[-q(k-1)R_\infty].$$

One of the solutions is $R_\infty = 0$

A nonzero solution exists when $1 - \exp[-q(k-1)R_\infty]$ has a slope greater than unity at $R_\infty = 0$.
i.e. when $q > 1/(k-1)$.

SIS model

Two variables only, S and I, related by $S + I = 1$ and satisfying

$$\frac{dS}{dt} = \mu I - qkIS, \tag{4}$$

$$\frac{dI}{dt} = -\mu I + qkIS. \tag{5}$$

The above equations lead to a single equation (with $\mu = 1$),

$$\frac{dI}{dt} = -I + qkI(1 - I). \tag{6}$$

$q_c = 1/k$ above which the infected population remains non-zero in SIS.

Solution:

$$I = \frac{e^{(qk-1)t}}{qk(e^{(qk-1)t} - 1)/(qk - 1) + 1/I_0},$$

I_0 is the initially infected fraction

Leads to the expected result: for $q \geq q_c = 1/k$, a finite fraction of infected agents exists as $t \to \infty$.

Heterogeneous cases: e.g., the degree distribution is not uniform.

In this case, one considers separately the infected, susceptible and removed population density of individuals with degree k.

Since the total population with degree k remains constant, one has, for all k,

$$S_k + R_k + I_k = 1.$$

Heterogeneous mean field (HMF) theory for SIR leads to the result

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

For SIS, HMF gives

$$q_c = \frac{\langle k \rangle}{\langle k^2 \rangle}$$

Some controversies have been recently raised over the validity of HMF.

Epidemics, percolation and networks

Theoretical studies before 2000 were mainly on lattices.

Since social structures are more complex than lattices, one needs to consider the epidemic spreading on networks.

It is known that SIR and SIS models correspond to isotropic and directed percolation : infection probability corresponds to site/bond occupation probability.

An indirect method to estimate the threshold is to consider the percolation phenomena.

Epidemics, percolation and networks

Theoretical studies before 2000 were mainly on lattices.

Since social structures are more complex than lattices, one needs to consider the epidemic spreading on networks.

It is known that SIR and SIS models correspond to isotropic and directed percolation : infection probability corresponds to site/bond occupation probability.

An indirect method to estimate the threshold is to consider the percolation phenomena.

Watts-Strogatz small world network: in addition to k nearest neighbours, shortcuts are added randomly between chosen pairs of sites.

Percolation threshold can be calculated: for example, with $k = 1$,

$$q_c = \frac{\sqrt{4p^2 + 12p + 1} - 2p - 1}{4p}.$$

where p is the average number of shortcuts per bond.

Moore and Newman (2000)

SIR model on WS networks and scale-free networks.

For SIR on WS network, it was shown that

$$R_\infty \sim (q - q_c),$$

i.e., R_∞ vanishes at the threshold with the exponent equal to unity which corresponds to the exponent β in percolation, having the value equal to unity in mean field case.

Moreno *et al (2002)*

SIR model on WS networks and scale-free networks.

For SIR on WS network, it was shown that

$$R_\infty \sim (q - q_c),$$

i.e., R_∞ vanishes at the threshold with the exponent equal to unity which corresponds to the exponent β in percolation, having the value equal to unity in mean field case.

Moreno *et al (2002)*

On the Barabási-Albert network, where an incoming node connects to m existing nodes, the threshold value $q_c = \frac{\langle k \rangle}{\langle k^2 \rangle}$ (remember HMF) becomes zero as the second moment diverges and the first moment vanishes.

Consistently it was found that

$$R_\infty = \exp(-1/qm).$$

Hence R_∞ remains finite for any nonzero q.

Effect of geographical factors

While considering the long ranged epidemic spreading on lattices, one can regard this as a process where infection takes place at long distances in a Lévy flight manner, i.e., agents at distance l are infected with a probability $P(l) \propto l^{-d-\eta}$ in d dimensions.

It is found that the the long-range critical behavior for small η is mean field like and it changes continuously to its short-range counterpart for η greater than a critical value

Janssen *et al* (1999)

This field theoretical result was confirmed by simulations in two dimensions.

Linder *et al.* (2008)

One can regard this process also as infection spreading on an Euclidean network where the probability $P(\ell)$ of connections at distance ℓ decreases with ℓ in a prescribed manner.

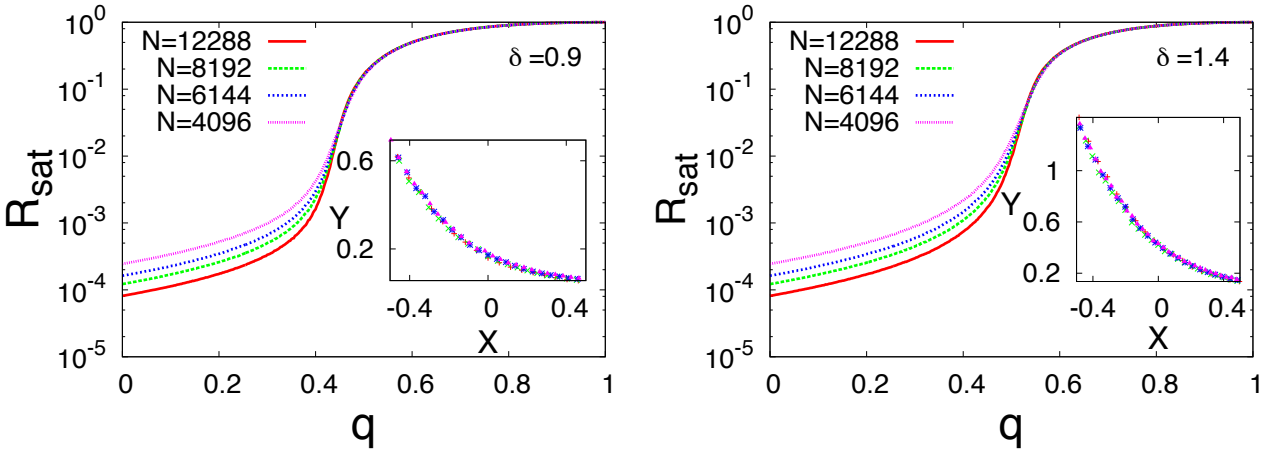
However, only a finite number of such bonds are allowed.

Extensive simulations in one and two dimensions have been made for the SIR model with $P(\ell) \propto \ell^{-\delta}$

Khaleque and Sen (2013); Grassberger (2012;2013)

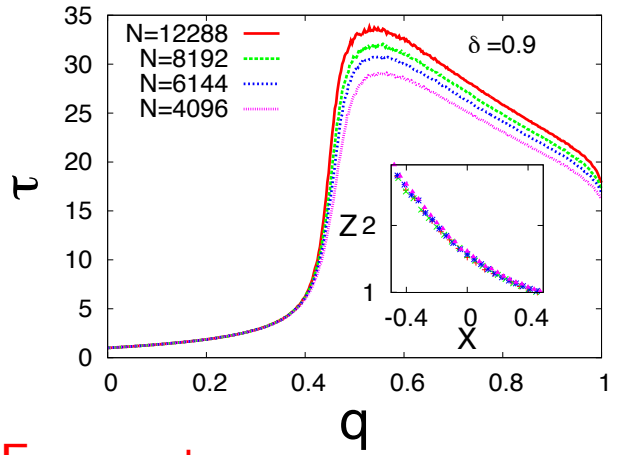
In the simulation, q_c and the exponents were estimated using finite size scaling.
The total duration of the disease also calculated.
Finite size scaling form for R_{sat}:

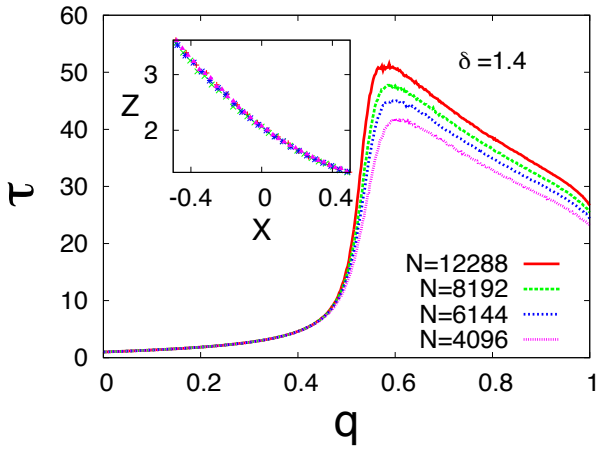
$$R_{sat} \propto N^{-\beta/\tilde{\nu}} g_1((q - q_c)N^{1/\tilde{\nu}})$$



The total duration τ is also analyzed by finite size scaling

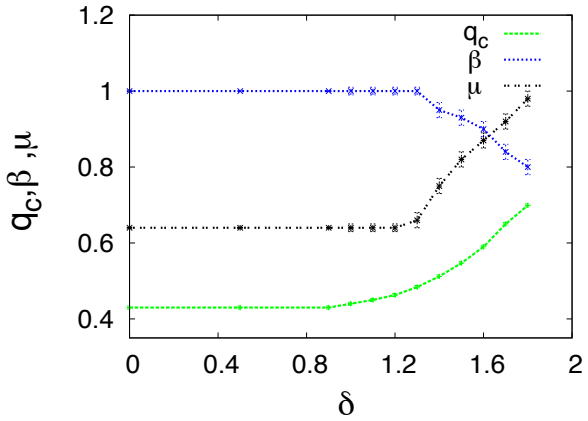
$$\tau \propto N^{\mu/\tilde{\nu}} g_2((q - q_c)N^{1/\tilde{\nu}})$$





Exponents

$\tilde{\nu}$ remains ≈ 3.0 in the entire parameter space



Other geographical networks

Other geographical networks

Contact processes, where each infected individual infects one neighbour at a time, on similar lattices were also considered leading to interesting phases

Muñoz *et al* (2010).

Other geographical networks

Contact processes, where each infected individual infects one neighbour at a time, on similar lattices were also considered leading to interesting phases

Muñoz *et al* (2010).

A general case when the spreading probability decreases with the network distance (i.e., with the number of steps separating two nodes) was considered: a finite threshold obtained in the scale free network.

Wu *et al* (2004)

Other geographical networks

Contact processes, where each infected individual infects one neighbour at a time, on similar lattices were also considered leading to interesting phases

Muñoz *et al* (2010).

A general case when the spreading probability decreases with the network distance (i.e., with the number of steps separating two nodes) was considered: a finite threshold obtained in the scale free network.

Wu *et al* (2004)

Geographical effect on the spreading on lattice embedded scale-free networks for the SI and the SIS models was also considered showing that epidemic spreads more smoothly in geographically constrained networks.

Xu et al (2006; 2007)

Other geographical networks

Contact processes, where each infected individual infects one neighbour at a time, on similar lattices were also considered leading to interesting phases

Muñoz *et al* (2010).

A general case when the spreading probability decreases with the network distance (i.e., with the number of steps separating two nodes) was considered: a finite threshold obtained in the scale free network.

Wu *et al* (2004)

Geographical effect on the spreading on lattice embedded scale-free networks for the SI and the SIS models was also considered showing that epidemic spreads more smoothly in geographically constrained networks.

Xu et al (2006; 2007)

Since human beings are mobile, this mobility factor has also been incorporated in a geographical hierarchical network

Zhao et al (2012)

Other geographical networks

Contact processes, where each infected individual infects one neighbour at a time, on similar lattices were also considered leading to interesting phases

Muñoz *et al* (2010).

A general case when the spreading probability decreases with the network distance (i.e., with the number of steps separating two nodes) was considered: a finite threshold obtained in the scale free network.

Wu *et al* (2004)

Geographical effect on the spreading on lattice embedded scale-free networks for the SI and the SIS models was also considered showing that epidemic spreads more smoothly in geographically constrained networks.

Xu et al (2006; 2007)

Since human beings are mobile, this mobility factor has also been incorporated in a geographical hierarchical network

Zhao et al (2012)

Traffic networks are strongly responsible for disease propagation - correlation between the two an important topic by itself.

Outbreak sizes and distribution; dynamics of spreading

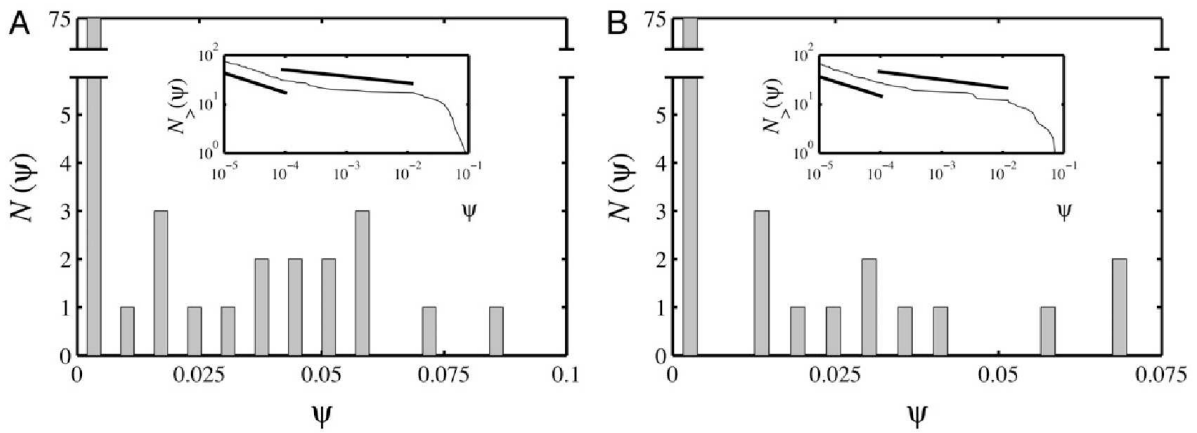
The total fraction of the population ever infected is the size of the outbreak.

May and Lloyd (2001) obtained theoretically the size of the outbreaks for both the SIS and SIR models on scale-free networks.

Distribution

Empirical data of childhood diseases showed that the distribution of the size of the outbreak has a peak for small outbreaks followed by a rather uniform distribution for larger sizes.

Measles and pertussis data for Iceland 1888-1990



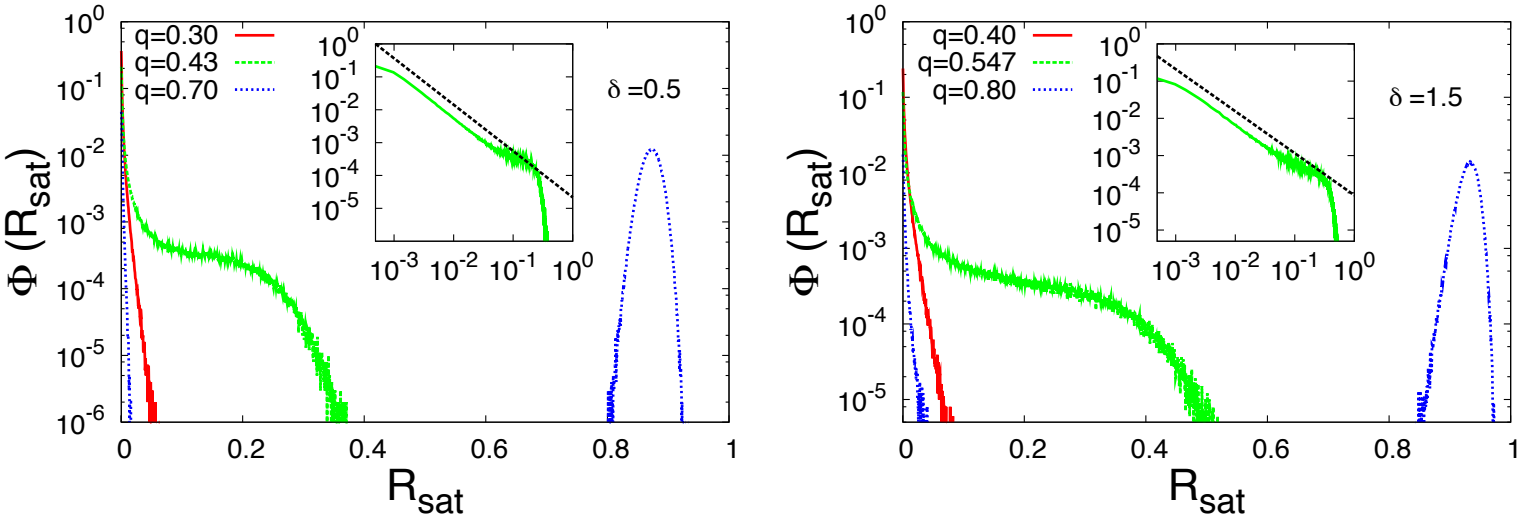
Distribution of epidemic size

Watts *et al.* (2005) analysed this data and showed that there is no power law behaviour, opposing the suggestion that the size distribution shows scale-free features (Rhodes and Anderson 1996).

For the SIR model, the distribution of outbreak size is unimodal (also called J shaped) below the threshold and bimodal (U shaped) above it on lattices.

This is also true for scale-free networks (Gallos and Argyrakis 2003) and Euclidean networks (Khaleque and Sen 2013).

Result for Euclidean network



$\delta = 0.5$ and 1.5

When there is only one infected node, a peak is always found at the value $1/N$ implying that the initially infected node has not been able to infect any one else.

A power law variation exactly at the threshold was obtained.

Spreading as a function of time

The fraction of newly infected agents can be studied as a function of time.

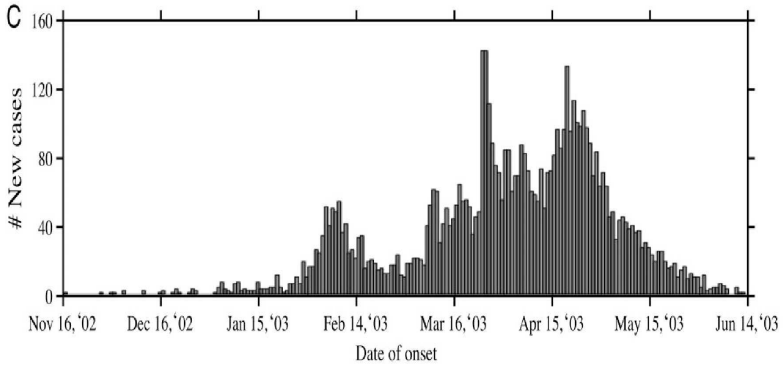
In SIR, it shows an initial increase, a peak value and then a decay.

Spreading as a function of time

The fraction of newly infected agents can be studied as a function of time.

In SIR, it shows an initial increase, a peak value and then a decay.

SAARS data

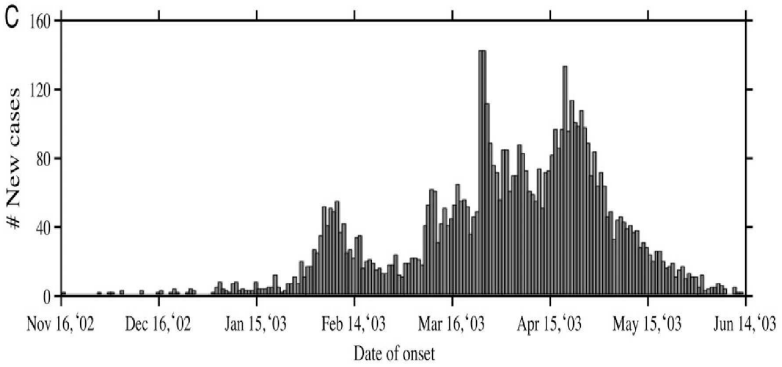


Spreading as a function of time

The fraction of newly infected agents can be studied as a function of time.

In SIR, it shows an initial increase, a peak value and then a decay.

SAARS data

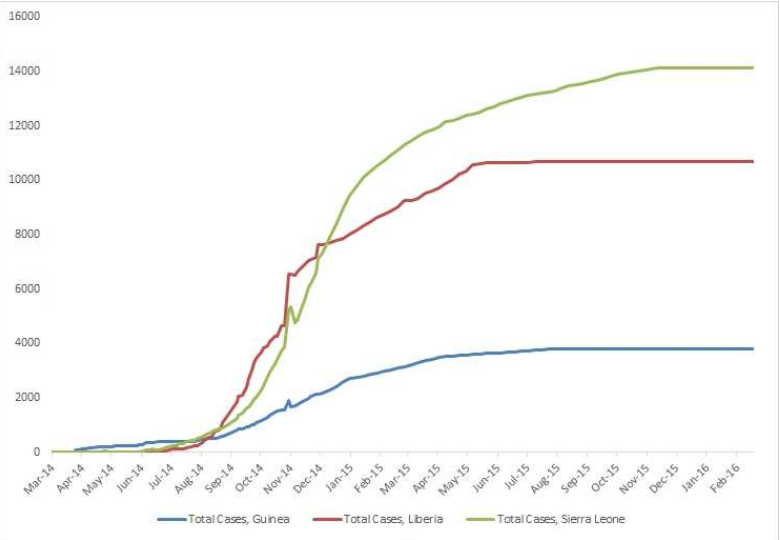


In real situations, even for diseases which can be described by the SIR model, There can be non-monotonic behaviour in the fraction of newly infected populations, with local peaks

Resurgent behaviour: reason may be demographic stochasticity.

For models with stochasticity, the number of infections die in time (for SIR) in a damped oscillatory manner (Hethcote 1974).

The cumulative infected population on the other hand shows an initial increase followed by a saturation.

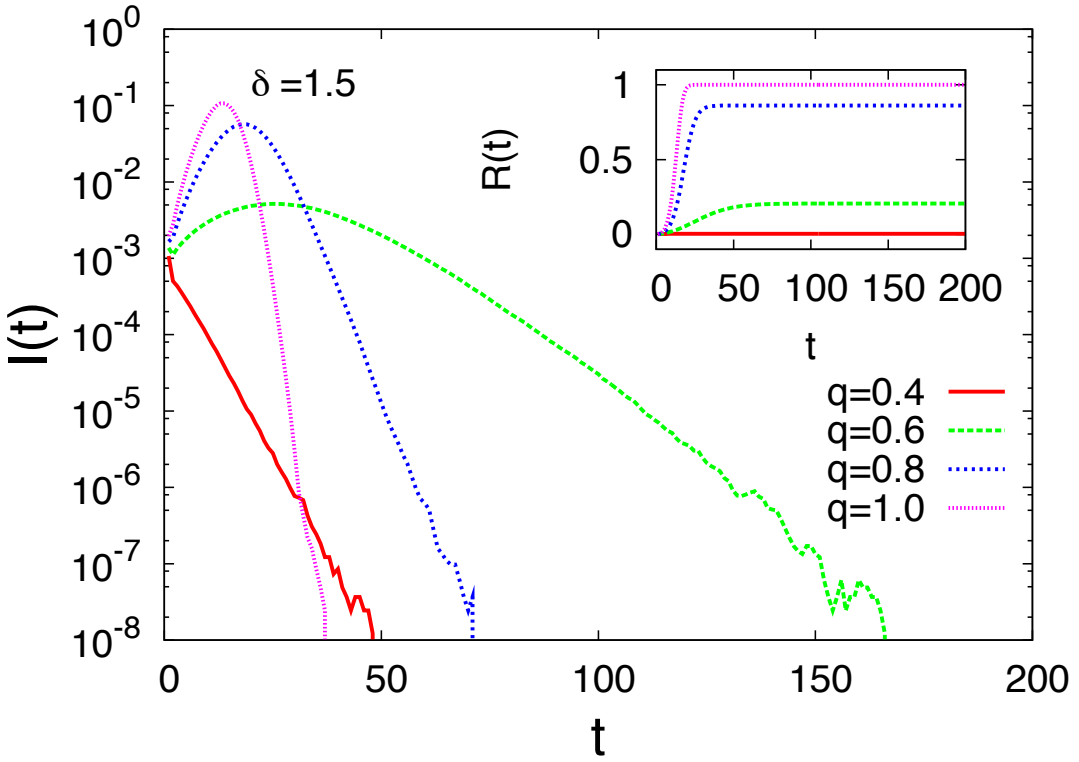


https://www.cdc.gov/vhf/ebola/outbreaks/2014-west-africa/cumulative-cases-graphs.html

In the SIS or SI model on the other hand, the fraction of newly infected population grows monotonically until it attains a saturation value (no peak).

The cumulative population is a sigmoid function in general (Moore and Newman 2000).

Dynamics: Result from simulation of SIR on Euclidean lattice:



Main plot: fraction of nodes infected at time t, $I(t) = R(t+1) - R(t)$.
Note that a peak is absent for $q < q_c$.
An initial growth and a peak value occurs at time $t = t_p$ only for $q > q_c$. For $q < q_c$, a monotonic decay.
Inset: cumulative data consistent with real data

Decay can occur if secondary infections are less than primary infections. Approximately,

$$(k-1)^2 q^2 < (k-1)q \tag{7}$$

which gives

$$q < 1/(k-1) = q_c. \tag{8}$$

This argument can explain the absence of the peak for $q < q_c$.

The fact that the recovered population is no longer susceptible has been ignored in this argument, but for initial times, this will not matter when the recovered population is very small.

Numerical data for $R(t)$ can be fitted to the form:

$$R(t) = \frac{a \exp(t/T)}{1 + c \exp(t/T)} - \frac{a}{1+c}, \tag{9}$$

where a, c and T depends on q and δ. The boundary condition assumed in the fitting is $R(0) = 0$.

From this one can expect that the peak will occur at time $t_p = T \log(1/c)$

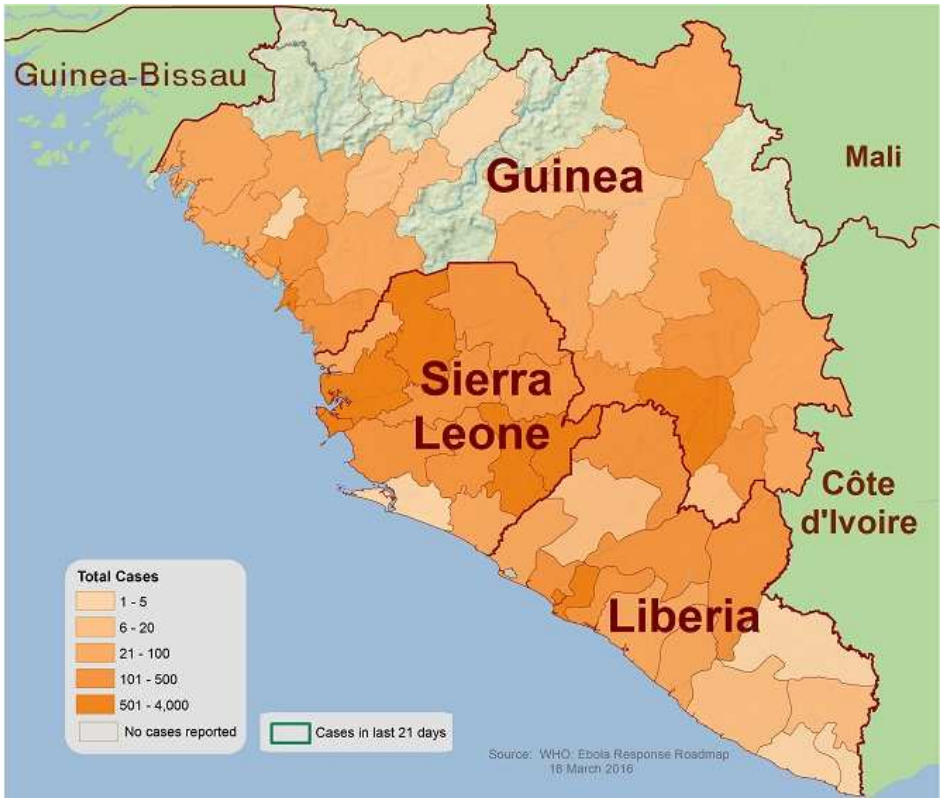
Table : a, c and T for three values of δ and comparison of t_p and $F(t_p)$ obtained from fitting and data($q = 0.58$)

δ	a $\times 10^{-3}$	c $\times 10^{-3}$	T	t_p (fit)	t_p (data)	$F \times 10^{-3}$ (fit)	$F \times 10^{-3}$ (data)
0.0	3.69	7.95	3.47	16.78	17	33.3	33.0
1.0	5.34	12.8	4.47	19.48	19	23.3	23.1
1.5	19.7	133	11.87	23.95	23	3.12	3.15

Analysis of the Ebola data

Ongoing work, A Khaleque and PS

Data from three West African countries:



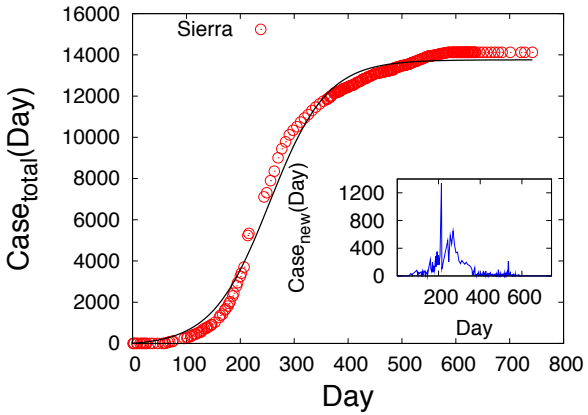
https://www.cdc.gov/vhf/ebola/outbreaks/2014-west-africa/distribution-map.html

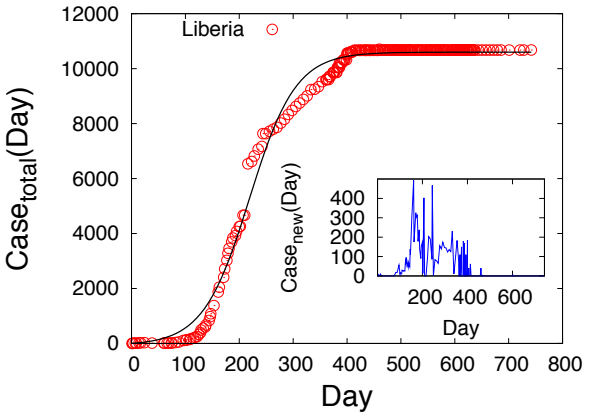
The outbreak began in Guinea in December 2013 and then spread to Liberia and Sierra Leone

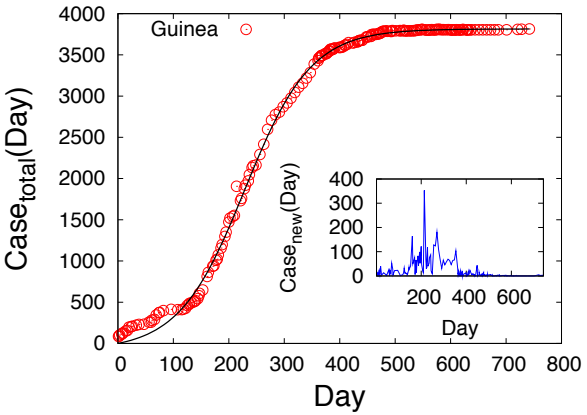
Ended 9 June 2016 in Liberia; 17 March 2016 in Sierra Leone and 1 June 2016 in Guinea

Data with fittings

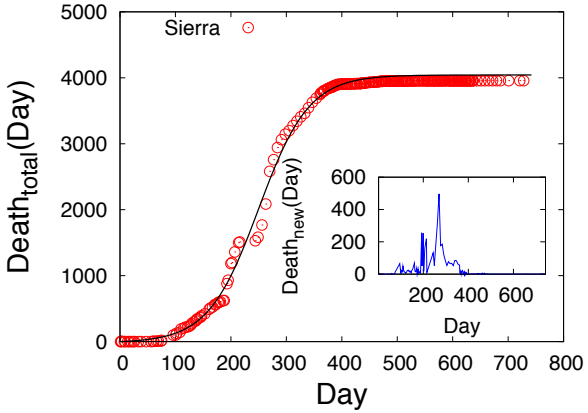
Cases

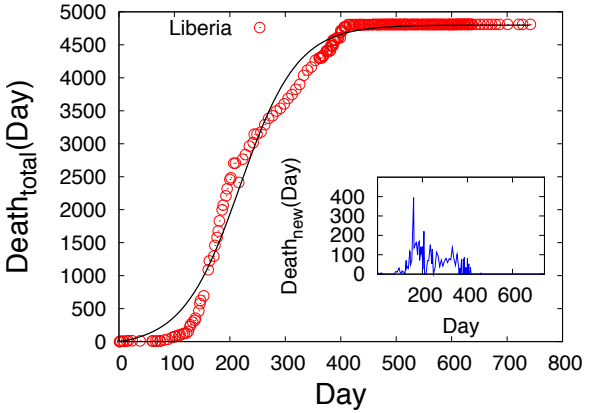


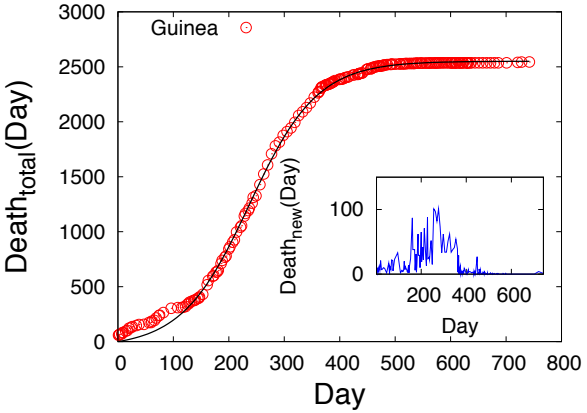




Deaths







Fitted form:

$$R(t) = \frac{a\exp(t/T)}{1+c\exp(t/T)} - \frac{a}{1+c}.$$

Table : Exponents a, c and T for three different countries

Country	a(Case)	c(Case)	T(Case)	a(Death)	c(Death)	T(Death)
Guinea	82.8	2.12×10^{-2}	59.43	53.26	2.05×10^{-2}	61.01
Liberia	64.9	6.08×10^{-3}	42.89	71.11	1.46×10^{-2}	51.72
Sierra Leone	117.2	8.44×10^{-3}	53.24	16.53	4.07×10^{-3}	45.43

The peak times can be calculated using the formula $t_p = T \log(1/c)$

Country	Peak Time (case)	Peak Time (Death)
Guinea	228.90	237.29
Liberia	218.85	218.57
Sierra Leone	254.14	250.00

Comments:

- It is hard to believe that the infection probability is different in different countries and thus a second factor may be present which is accounted by δ
- t_p usually increases with δ.
  δ is a measure of the length scale of disease spread.
  Sierra Leone being the smallest of the three, naively it can be assumed it has the maximum δ and therefore t_p is largest for it.
  However, saturated value of infected density largest for SL.
  Number of nearest neighbours in SL may be more as her population density is larger comparatively
  Probably lowers the value of q indirectly.
  Role of degree and δ to be simultaneously studied....
- Data is for two dimensional space while fitting form obtained from simulations made in one dimension. However, the network properties enable it to behave like a finite dimensional system below $\delta = 2$
- Effect of immunisation not incorporated in model

Controlling epidemic

The spread of epidemic may be controlled by making recovery rates faster and decreasing infection probability.

Some diseases may be prevented by vaccine.

Practically, it may not be possible to immunise the entire population.

Is it possible to stop the disease by vaccinating only a fraction?

This problem has direct analogy with network robustness:

A network becomes disconnected under an attack - usually targeted attacks are more effective; one needs to deactivate a fraction of nodes/edges only.

Vaccination effectively increases the threshold - linearly for WS networks and exponentially for Scale free (Peng et al 2013).

Method of immunisation other than vaccination : keeping the patient confined or quarantined (analogous to removing edges/nodes).

More modern methods: dynamically control the network (Selly et al 2015)

However, the network itself may be dynamic. Need to carefully choose the strategy (Stamini eta al 2013).

Useful reference:

Analysis and Control of Epidemics: A survey of spreading processes on complex networks

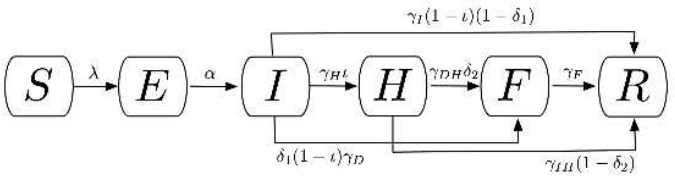
Cameron Nowzari, Victor M. Preciado, George J. Pappas arXiv:1505.00768

Practical issues like limited budget and feasibility of cutting links etc are relevant questions also.

Tracing the network: Contact tracing

Field based epidemiologists attempt to determine the source of infection in each case and each infected patient is linked to one or more other persons from whom they caught the disease and also a set of other individuals to whom they might transmit the disease to.

One can then identify the persons responsible for the infection and isolate them (note that the infection is possible over a considerable duration in reality)



Intermediate states: E = exposed, H = hospitalised, F = funeral (contracted the disease while handling diseased patient's body).
Statistical analysis using contact tracing have been made for diseases like Ebola in West Africa and HIV/AIDS in Cuba in the recent past.
Some models using contact tracing have also been proposed.

Comparison with other cascading processes

Many other phenomena where epidemic like feature is seen - especially cascading.

Examples - social phenomena like information spreading, popularity dynamics etc.

However, a major difference is, in such phenomena, the social choice is important. Not true for disease spreading where infection is involuntary- even a single diseased agent can lead to an epidemic.

How does social choice affect the distribution of "infected" population ?
In most cases, a log normal distribution was obtained.

Simple explanation of the log normal: take any quantity $\mathcal{R}$ which may represent, eg, the total revenue collected from the sales of books, movie tickets etc.,

$\mathcal{R}(t)$ grows at the rate

$$\frac{1}{\mathcal{R}(t)}\frac{d\mathcal{R}(t)}{dt} = r(t), \tag{10}$$

where $r(t)$ are randomly distributed.

This gives

$$\log(\mathcal{R}(t)) - \log(\mathcal{R}(0)) = \int_0^t r(t')dt'. \tag{11}$$

De Vany and Walls (1996)

For large enough times, the integral on the right hand side will be normally distributed such that $\mathcal{R}(t)$ has a log normal behaviour.

Similarly for popularity of items uploaded on the internet: Let N_t represent, eg, number of likes. It is a growing variable in time and it is expected that spread of popularity will happen through friends such that $N_t = (1 + X_t)N_{t-1}$ where where $X_1, X_2, \cdots$ are random positive variables, independently and identically distributed with mean value equal to 1.
As the growth in time is eventually curtailed by a decay in novelty, Wu and Huberman (2007) incorporated another factor γ_t which vanishes as $t \to \infty$ such that $N_t = (1 + \gamma_t X_t)N_{t-1}$. Approximately, for small t, one can write

$$N_t = \prod_t (1 + \gamma_t X_t) N_0 \approx \exp\left[\sum_t (\gamma_t X_t)\right] N_0$$

such that

$$\ln(N_t/N_0) = \sum_s \gamma_t X_t. \tag{12}$$

The right hand side being a sum of random variables follows a normal distribution and therefore N_t follows a log normal distribution.

A summary to end

- The complex process of disease spreading may be modelled in different ways.
- Basic models with threshold behaviour discussed - modifications numerous in the literature
- A model with geographical factor playing a role studied
- The Ebola data analysis shows consistency but also raises some questions

A summary to end

- The complex process of disease spreading may be modelled in different ways.
- Basic models with threshold behaviour discussed - modifications numerous in the literature
- A model with geographical factor playing a role studied
- The Ebola data analysis shows consistency but also raises some questions

Finally some advertisement: These and many other topics discussed in the book

Sociophysics: An Introduction: PS and B. K. Chakrabarti (OUP; 2013)

and

Thank you!