

V6 Time-dependent properties of Epidemic Networks

Review (V5): An SI outbreak starting with a single randomly chosen vertex eventually spreads to all members of the component containing that vertex.

Let us assume that vertex i belongs to the giant component.

With probability s_i , vertex i is **susceptible**.

To become infected an individual must catch the disease from a neighboring individual j that is already infected.

The probability for j being **infected** is

$$x_j = 1 - s_j$$

The **transmission** of the disease during the time interval t and $t + dt$ occurs with probability βdt .

Review: Time-dependent properties of Epidemic Networks

Multiplying these probabilities and then summing over all neighbors of i , yields the total probability of i becoming infected :

$$\beta s_i \sum_j A_{ij} x_j$$

where A_{ij} is an element of the adjacency matrix.

Thus, the s_i obey a set of n non-linear differential equations

$$\frac{ds_i}{dt} = -\beta s_i \sum_j A_{ij} x_j = -\beta s_i \sum_j A_{ij} (1 - s_j)$$

From $s_i + x_i = 1$ we get the complementary equation for x_i .

$$\frac{dx_i}{dt} = \beta s_i \sum_j A_{ij} x_j = \beta (1 - x_i) \sum_j A_{ij} x_j$$

We will assume again that the disease starts either with a single vertex or a small randomly selected number c of vertices.

Thus $x_i = c/n \rightarrow 0$, $s_i = 1 - c/n \rightarrow 1$ in the limit for large n .

Review: Time-dependent properties of Epidemic Networks

The equation $\frac{ds_i}{dt} = -\beta s_i \sum_j A_{ij} x_j = -\beta s_i \sum_j A_{ij} (1 - s_j)$ is not solvable in closed form for general A_{ij} .

By considering suitable limits, we can calculate some features of its behavior.

Let us e.g. consider the behavior of the system at early times.

For large n and the given initial conditions, x_i will be small in this regime.

By ignoring terms of quadratic order, we can approximate

$$\frac{dx_i}{dt} = \beta s_i \sum_j A_{ij} x_j \approx \beta \sum_j A_{ij} x_j$$

or in matrix form

$$\frac{d\mathbf{x}}{dt} = \beta \mathbf{A} \mathbf{x} \quad \text{where } \mathbf{x} \text{ is the vector with elements } x_i.$$

Review: Time-dependent properties of Epidemic Networks

Write \mathbf{x} as a linear combination of the eigenvectors of the adjacency matrix

$$\mathbf{x}(t) = \sum_{r=1}^n a_r(t) \mathbf{v}_r$$

where \mathbf{v}_r is the eigenvector with eigenvalue σ_r . Then

$$\frac{d\mathbf{x}}{dt} = \sum_{r=1}^n \frac{da_r}{dt} \mathbf{v}_r = \beta \mathbf{A} \sum_{r=1}^n a_r(t) \mathbf{v}_r = \beta \sum_{r=1}^n \sigma_r a_r(t) \mathbf{v}_r$$

By comparing terms in \mathbf{v}_r we get $\frac{da_r}{dt} = \beta \sigma_r a_r$

This has the solution $a_r(t) = a_r(0) e^{\beta \sigma_r t}$

Substituting this expression back gives $\mathbf{x}(t) = \sum_{r=1}^n a_r(0) e^{\beta \sigma_r t} \mathbf{v}_r$

The fastest growing term in this expression is the term corresponding to the largest eigenvalue σ_1 .

Review: Time-dependent properties of Epidemic Networks

Assuming this term dominates over the others we will get

$$\mathbf{x}(t) \sim e^{\beta \sigma_1 t} \mathbf{v}_1$$

So we expect the number of infected individuals to grow exponentially, just as in the fully mixed version of the SI model, but now with an exponential constant that depends not only on β but also on the leading eigenvalue of the adjacency matrix.

The probability of infection in this early period varies from vertex to vertex roughly as the corresponding element of the leading eigenvector \mathbf{v}_1 .

Propagation of SI model on network

For long times, the network version of the SI model predicts that the probability of infection of a vertex in the giant component tends to 1.

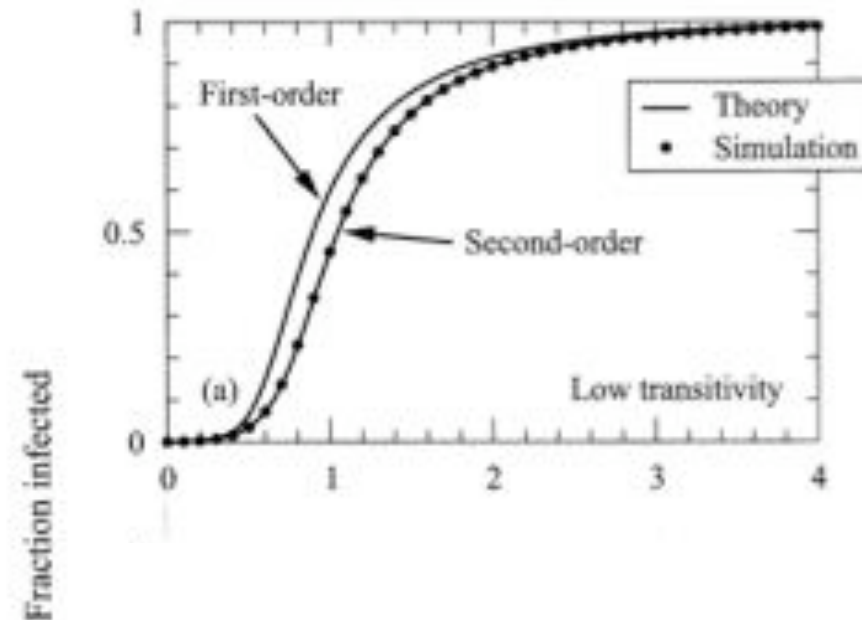
This is qualitatively similar to the fully mixed version of the SI model.

What happens if we integrate $\frac{dx_i}{dt} = \beta(1 - x_i) \sum_j A_{ij}x_j$ numerically?

The curve „first-order“ shows the results of a numerical integration.

The points labelled „simulation“ are average values from many simulations of simulated epidemics with the same spreading β .

→ the agreement is ok but not perfect.



Propagation of SI model on network

The reason for this discrepancy is that the probabilities of the two nodes i and j in $\frac{ds_i}{dt} = -\beta s_i \sum_j A_{ij} x_j$ are not independent.

In general, the quantities s_i and x_j between neighboring vertices are **correlated**.

We can incorporate this into our calculations, at least approximately, by using a so-called **pair approximation**.

Pair approximations

Let us denote by $\langle s_i \rangle$ the average probability that vertex i is susceptible.

Similarly, $\langle x_i \rangle$ is the average probability that i is infected and $\langle s_i x_j \rangle$ is the average probability that i is susceptible AND j is infected at the same time.

This yields now a truly exact version of our previous eq. $\frac{ds_i}{dt} = -\beta \sum_j A_{ij} \langle s_i x_j \rangle$. (1)

The previously used equation $\frac{ds_i}{dt} = -\beta s_i \sum_j A_{ij} x_j$

is an approximation to this true equation where we assumed that $\langle s_i x_j \rangle \cong \langle s_i \rangle \langle x_j \rangle$

The trouble with the new eq. is that we cannot solve it directly because it contains the unknown quantity $\langle s_i x_j \rangle$.

Pair approximations

-> we need an expression for $\langle s_i x_j \rangle$.

To reach the state where i is susceptible and j is infected, j must have been infected by another neighboring vertex k .

The probability for such a previous configuration where i and j are susceptible and k is infected is $\langle s_i s_j x_k \rangle$.

When starting with this configuration, j will become infected via k with rate β .

Summing over all neighbors k except for i , the total rate at which j becomes infected is

$$\beta \sum_{k \neq i} A_{jk} \langle s_i s_j x_k \rangle$$

Pair approximations

However, $\langle s_i x_j \rangle$ can also decrease when i becomes infected.

This can happen in 2 ways.

- Either i can be infected by its infected neighbor j what happens with rate $\beta \langle s_i x_j \rangle$, or
- i can be infected by another neighbor $l \neq j$ so that i is infected with rate $\beta \langle x_l s_i x_j \rangle$.

Summing the latter expression over all neighbors l other than j gives a total rate of

$$\beta \sum_{l \neq j} A_{il} \langle x_l s_i x_j \rangle.$$

Pair approximations

Putting all these terms together, we get

$$\frac{d\langle s_i x_j \rangle}{dt} = \beta \sum_{k \neq i} A_{jk} \langle s_i s_j x_k \rangle - \beta \sum_{l \neq j} A_{il} \langle x_l s_i x_j \rangle - \beta \langle s_i x_j \rangle$$

In theory this equation now allows us to calculate $\langle s_i x_j \rangle$.

In practice, it involves yet more terms that we don't know on the right-hand side, the three-variable averages $\langle s_i s_j x_l \rangle$ and $\langle x_l s_i x_j \rangle$.

Such a succession of equations will never end – in the jargon of mathematics, **it doesn't close.**

Moment closure

However, we can make progress by approximating the 3-variable averages by appropriate combinations of 1-variable and 2-variable averages.

This will allow us to close the equations and get a set that we can solve.

This process is called **moment closure**.

The moment closure method at the level of 2-variable averages that we discuss here is also called **pair approximation method**.

Starting with $\frac{d\langle s_i x_j \rangle}{dt} = \beta \sum_{k \neq i} A_{jk} \langle s_i s_j x_k \rangle - \beta \sum_{l \neq j} A_{il} \langle x_l s_i x_j \rangle - \beta \langle s_i x_j \rangle$
our goal is to approximate the 3-variable averages with lower-order ones.

Moment closure

We will make use of **Bayes theorem** for probabilities:

$$\langle s_i s_j x_k \rangle = P(i, j \in S, k \in I) = P(i, j \in S)P(k \in I | i, j \in S)$$

where $P(i \in S)$ means the probability that vertex i is in the set S of susceptible vertices.

We know that i and j are neighbors in the network and that j and k are neighbors.

We will assume that the disease state of k does not depend on the disease state of i .

If the only path in the network from i to k is through j , this is a good approximation.

If there is another path, this remains an approximation.

Moment closure

Assuming the state of k to be independent of the state of i , we have

$$P(k \in I | i, j \in S) = P(k \in I | j \in S) = \frac{P(j \in S, k \in I)}{P(j \in S)} = \frac{\langle s_j x_k \rangle}{\langle s_j \rangle}$$

After putting this into our previous equation we get

$$\begin{aligned} \langle s_i s_j x_k \rangle &= P(i, j \in S, k \in I) = P(i, j \in S) P(k \in I | i, j \in S) \\ &= \langle s_i s_j \rangle \frac{\langle s_j x_k \rangle}{\langle s_j \rangle} \end{aligned}$$

In a similar way, we can derive an expression for the other 3-variable average

$$\langle x_l s_i x_j \rangle = \langle x_l s_i \rangle \frac{\langle s_i x_j \rangle}{\langle s_i \rangle}$$

Moment closure

We can now substitute both expressions into our previous equation and get the pair approximation equation

$$\frac{d\langle s_i x_j \rangle}{dt} = \beta \frac{\langle s_i s_j \rangle}{\langle s_j \rangle} \sum_{k \neq i} A_{jk} \langle s_j x_k \rangle - \beta \frac{\langle s_i x_j \rangle}{\langle s_i \rangle} \sum_{l \neq j} A_{il} \langle s_i x_l \rangle - \beta \langle s_i x_j \rangle$$

This equation now contains only averages over 2 variables at a time.

It contains also a new average $\langle s_i s_j \rangle$ that we have not encountered before.

This can be rewritten as $\langle s_i s_j \rangle = \langle s_i (1 - x_j) \rangle = \langle s_i \rangle - \langle s_i x_j \rangle$. With this, our eq. becomes

$$\frac{d\langle s_i x_j \rangle}{dt} = \beta \frac{\langle s_i \rangle - \langle s_i x_j \rangle}{\langle s_j \rangle} \sum_{k \neq i} A_{jk} \langle s_j x_k \rangle - \beta \frac{\langle s_i x_j \rangle}{\langle s_i \rangle} \sum_{l \neq j} A_{il} \langle s_i x_l \rangle - \beta \langle s_i x_j \rangle \quad (2)$$

We will simplify this eq. further.

Moment closure

Let us define p_{ij} to be the conditional probability that j is infected given that i is not:

$$p_{ij} = P(j \in I | i \in S) = \frac{P(i \in S, j \in I)}{P(i \in S)} = \frac{\langle s_i x_j \rangle}{\langle s_i \rangle}$$

Then the time evolution of p_{ij} is given by

$$\frac{dp_{ij}}{dt} = \frac{d}{dt} \left(\frac{\langle s_i x_j \rangle}{\langle s_i \rangle} \right) = \frac{1}{\langle s_i \rangle} \frac{d\langle s_i x_j \rangle}{dt} - \frac{\langle s_i x_j \rangle}{\langle s_i \rangle^2} \frac{d\langle s_i \rangle}{dt} \quad \text{Use (1) and (2)}$$

$$= \beta \left(1 - \frac{\langle s_i x_j \rangle}{\langle s_i \rangle} \right) \sum_{k \neq i} A_{jk} \frac{\langle s_j x_k \rangle}{\langle s_j \rangle} - \beta \frac{\langle s_i x_j \rangle}{\langle s_i \rangle} \sum_{l \neq j} A_{il} \frac{\langle s_i x_l \rangle}{\langle s_i \rangle} - \beta \frac{\langle s_i x_j \rangle}{\langle s_i \rangle} + \beta \frac{\langle s_i x_j \rangle}{\langle s_i \rangle} \sum_l A_{il} \frac{\langle s_i x_l \rangle}{\langle s_i \rangle}$$

$$= \beta (1 - p_{ij}) \sum_{k \neq i} A_{jk} p_{jk} - \beta p_{ij} \sum_{l \neq j} A_{il} p_{il} - \beta p_{ij} + \beta p_{ij} \sum_l A_{il} p_{il}$$

The terms in the 2 sums over l now cancel out except for $l = j$, leaving

$$\frac{dp_{ij}}{dt} = \beta (1 - p_{ij}) \left[-p_{ij} + \sum_{k \neq i} A_{jk} p_{jk} \right]$$

where we have also used $A_{ij} = 1$

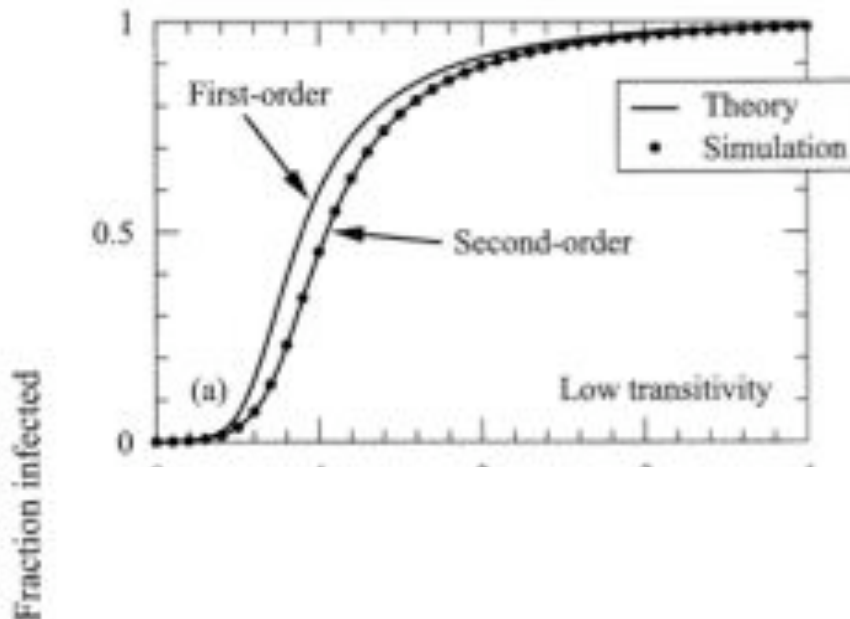
Moment closure

Rewriting $\frac{ds_i}{dt} = -\beta \sum_j A_{ij} \langle s_i x_j \rangle$ in terms of p_{ij} gives

$$\frac{d\langle s_i \rangle}{dt} = -\beta \langle s_i \rangle \sum_j A_{ij} p_{ij}$$

which has the solution

$$\langle s_i(t) \rangle = \langle s_i(0) \rangle \exp \left(-\beta \sum_j A_{ij} \int_0^t p_{ij}(t') dt' \right)$$



Numerical integration of these 2 equations for $\frac{dp_{ij}}{dt}$ and $\langle s_i(t) \rangle$ gives the „second order“ curve, which agrees very nicely with the simulated epidemics.

Recent extensions in modelling disease spreading

Modelling disease outbreaks in realistic urban social networks

Stephen Eubank¹, Hasan Guclu², V. S. Anil Kumar¹, Madhav V. Marathe¹, Aravind Srinivasan³, Zoltán Toroczkai⁴ & Nan Wang⁵

¹*Basic and Applied Simulation Science Group, Los Alamos National Laboratory, MS M997, Los Alamos, New Mexico 87545, USA*

²*Department of Physics, Applied Physics and Astronomy, Rensselaer Polytechnic Institute 110 8th Street, Troy, New York 12180-3590, USA*

³*Department of Computer Science and Institute for Advanced Computer Studies, University of Maryland, College Park, Maryland 20742, USA*

⁴*Centre for Nonlinear Studies and Complex Systems Group, Los Alamos National Laboratory, MS B258, Los Alamos, New Mexico 87545, USA*

⁵*Department of Computer Science, University of Maryland, College Park, Maryland 20742, USA*

Nature 429, 180 (2004)

Authors used the Transportation Analysis and Simulation System (TRANSIMS) developed at Los Alamos Natl. Laboratories.

It is estimated that by 2030 more than 60% of the world's population live in cities.

Urban areas form a perfect fabric for fast, uncontrolled disease propagation.

-> Need to model inter-person contacts in more detail than by differential equations.

TRANSIMS

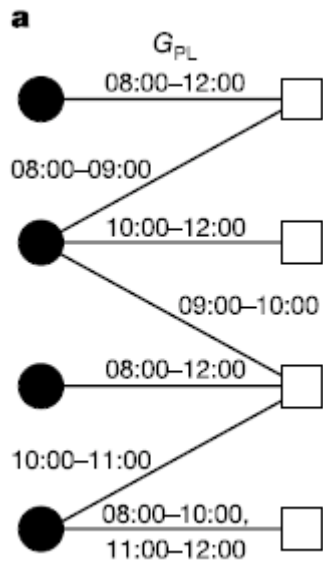
TRANSIMS creates a synthetic population with a demographic distribution (age, income) consistent with joint distributions in census data.

The program uses an agent-based paradigm to simulate the movement of all individuals during e.g. one day.

By this it estimates positions and activities of all travelers using a transportation infrastructure on a second-by-second basis.

Nature 429, 180 (2004)

Model social contact network as bipartite graph



a, A bipartite graph G_{PL} with two types of vertices representing four people (P) and four locations (L).

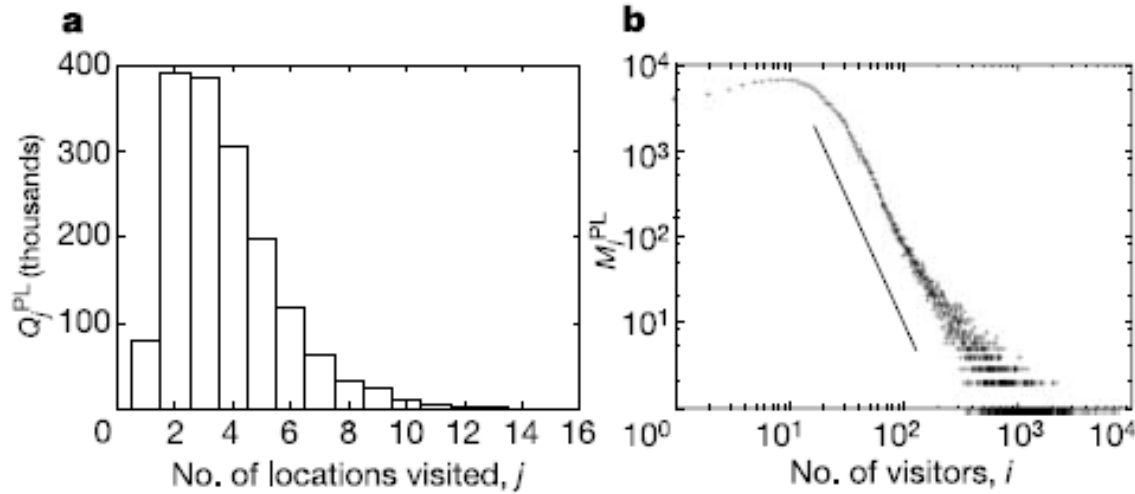
If person p visited location l , there is an edge in this graph between p and l .

Vertices are labelled with appropriate demographic or geographic information, edges with arrival and departure times.

For the city Portland / US, G_{PL} has about 1.6 million vertices with a giant component of 1.5 million people and 180.000 locations.

Nature 429, 180 (2004)

Degree distributions for estimated Portland social network.



a, The number of people Q_j^{PL} who visited j different locations in the bipartite people–locations graph G_{PL} .

b, The number of locations M_i^{PL} in G_{PL} that are visited by exactly i different people.

On average people visited 4 locations / day.

This distribution decays exponentially (nobody visited more than 15 locations in 1 day).

This plot follows a power-law tail. The slope of the straight-line graph is -2.8.

Nature 429, 180 (2004)

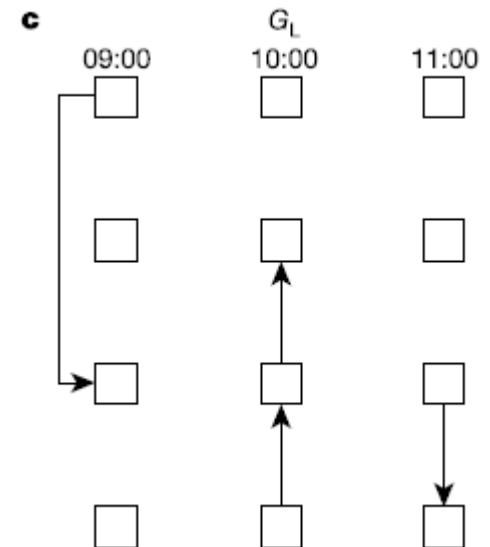
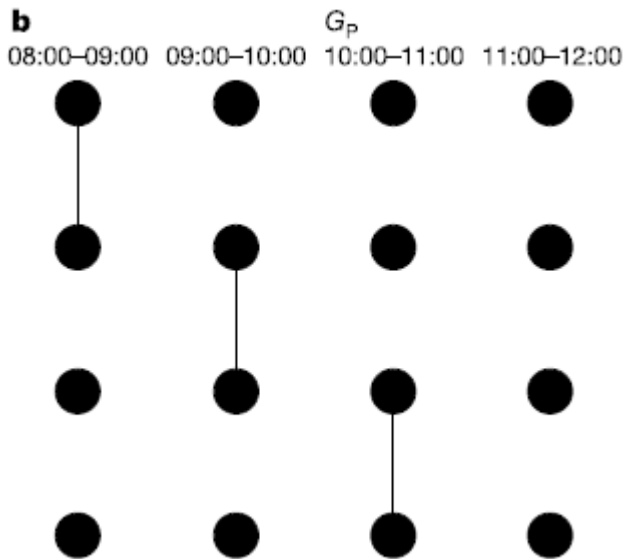
Disease transmission

Many infectious diseases get transmitted between people in the same location.

Diseases spread mainly due to people's movement.

-> investigate two projections of G_{P_L} obtained by drawing an edge between all pairs of vertices that have a distance of 2 from each other on the bipartite graph.

This yields two disconnected graphs G_P and G_L .



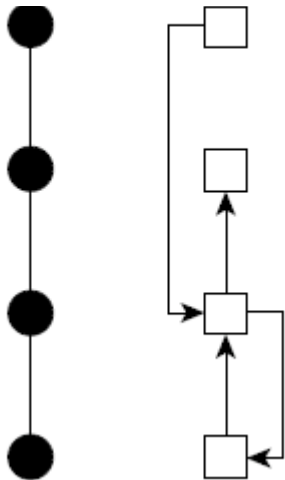
Nature 429, 180 (2004)

Static projections of graph G_{PL}

In G_P the edges are labelled with the sets of time intervals during which the people were co-located.

For simplicity, we will discard these time labels and look at the static projections shown below

d



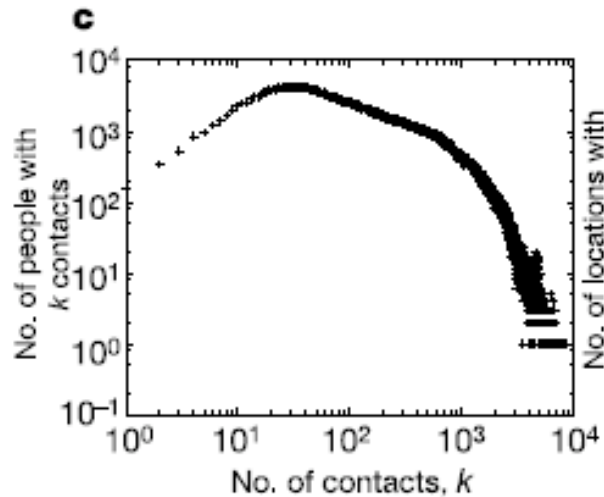
d, People (such as 24-year-old male) are represented by filled circles, and locations (such as 34 Elm Street) by open squares.

This projection yields a worst-case scenario for the spreading of a disease because the static graph is much more connected than G_P .

Consider also the idea of an “effective contact” by removing all edges for which the duration of contact was shorter than 1 hour.

Nature 429, 180 (2004)

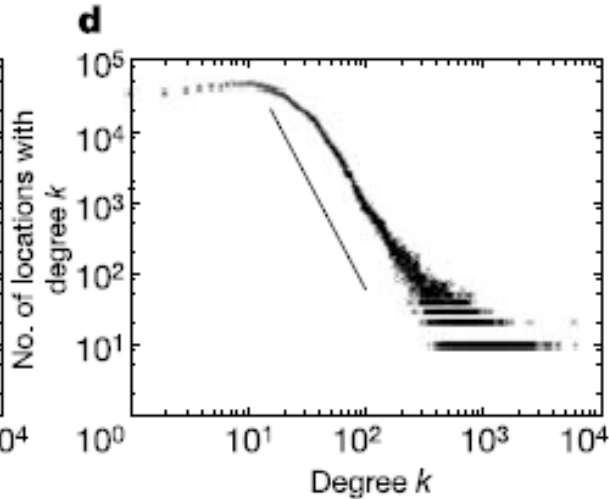
Degree distributions of static graph



c, The number of people who have k neighbors in the static people–contact graph P on log–log scale.

This network resembles a small-world network.

Nature 429, 180 (2004)



d, The in and out degree distributions of the locations network G_L . This network resembles a scale-free network with a power-law distribution. The slope of the straight-line graph is -2.8 .

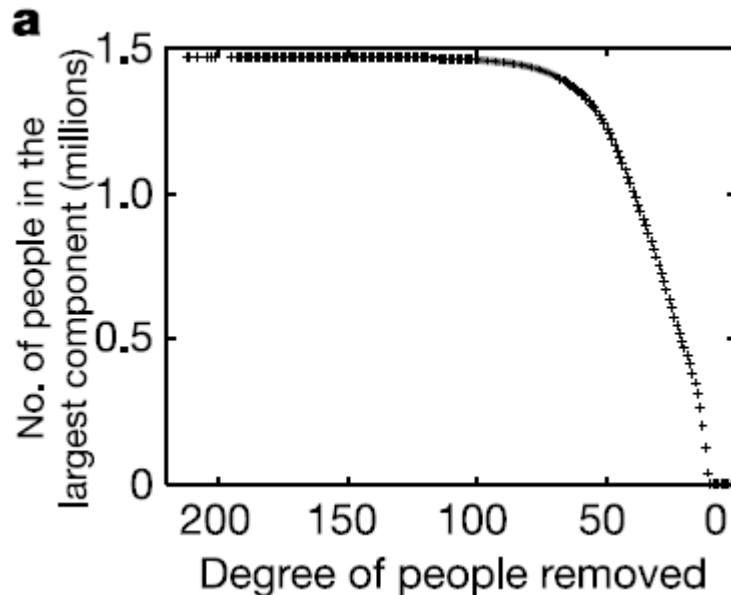
The scale-free nature of the localization network reflects the capacity distribution of locations.

Shopping malls take up large numbers of people who normally live in **small homes**.

Effects of removing high-degree vertices

Can one e.g. prevent disease propagation in this social network by vaccinating (or putting into quarantine) a small number of high-degree vertices?

-> investigate the size of the giant component after removing from the bipartite graph G_{PL} all vertices with degree higher than k .

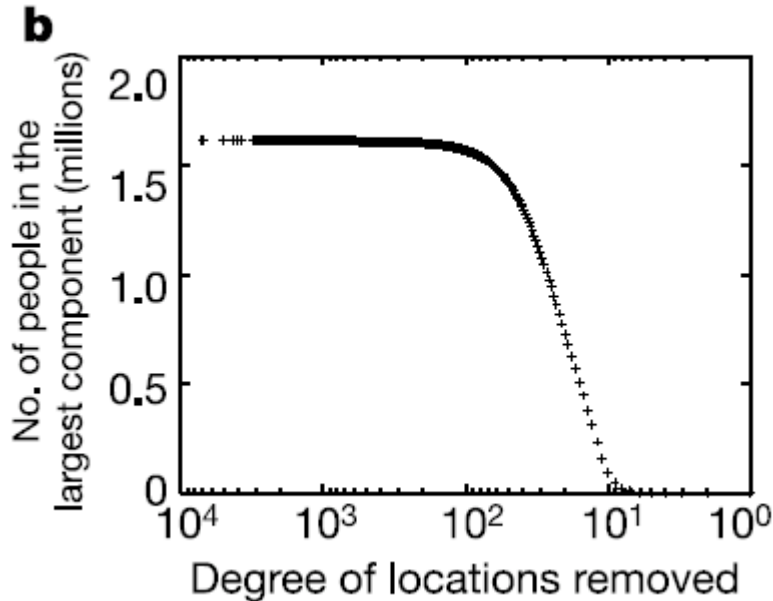


-> even when all vertices of degree 11 and higher are removed, a unique giant component persists.

Thus, any effective vaccination strategy must be mass vaccination.

Nature 429, 180 (2004)

Effect of removing most visited locations



In **b** we remove all locations with degree k or higher from G_{PL} and monitor the size of the largest connected component in the static people–contact graph induced by the remaining bipartite graph

-> Even closing the most-visited locations – or vaccinating everyone who visits them – does not shatter the induced people-contact graph until large fractions of the population have been effected.

Q: Can epidemics be stopped at all without resorting to mass vaccination?

Nature 429, 180 (2004)

Early detection

An alternative strategy could involve early detection of infected individuals and efficient targeting.

Consider an idealized situation in which sensors at a location can detect whether any person is infected.

The feasibility of early detection depends on the number of sensors required.

This problem is equivalent to finding the minimum dominating set.

Nature 429, 180 (2004)

Overlap ratio

We wish to find a subset L' of locations so that all (or most) people visit some location in L' .

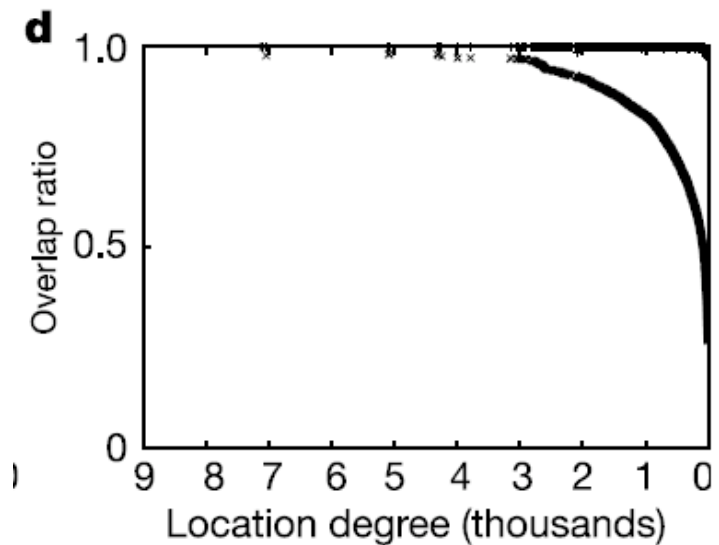
The overlap ratio $\omega(J)$ of a set of locations $J \subset L$ is $\frac{n(J)}{\sum_{l \in J} \text{deg}(l)}$

where $n(J)$ is the total number of people visiting any location in J and $\text{deg}(l)$ is the number of different people visiting location $l \in J$.

The smaller the overlap ratio, the larger is the number of different locations in J visited by a single person.

d. Overlap ratio by degree. The lower curve shows the cumulative overlap ratio by degree, which is the overlap ratio for locations having degree k or less. The upper curve shows the overlap ratio for locations having degree = k .

-> Not many people visit more than one high-degree location per day. **Early detection could be accomplished by placing sensors there.**



Nature 429, 180 (2004)

Recent extensions in modelling disease spreading

The role of the airline transportation network in the prediction and predictability of global epidemics

Vittoria Colizza*, Alain Barrat†, Marc Barthélemy**‡, and Alessandro Vespignani*§

*School of Informatics and Center for Biocomplexity, Indiana University, Bloomington, IN 47401; and †Centre National de la Recherche Scientifique, Unité Mixte de Recherche 8627, Université Paris-Sud, Bâtiment 210, F-91405 Orsay, France

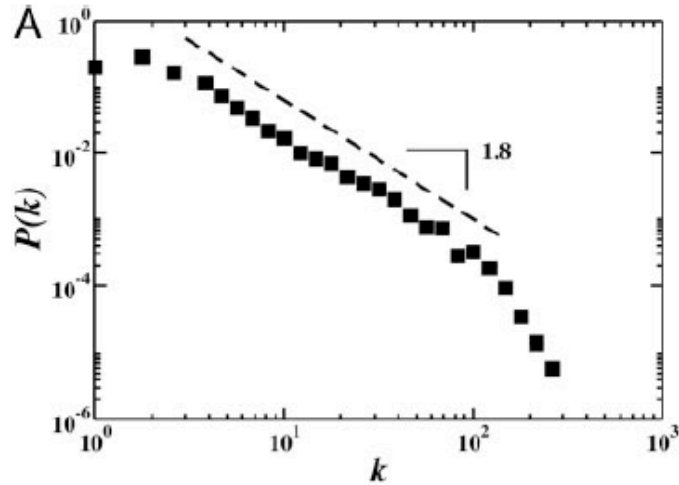
Use IATA database : 3880 airports , 18.810 weighted edges, where weight w_{jl} denotes passenger flow between airports j and l .

Consider also populations N_j of the metropolitan areas served by each airport.

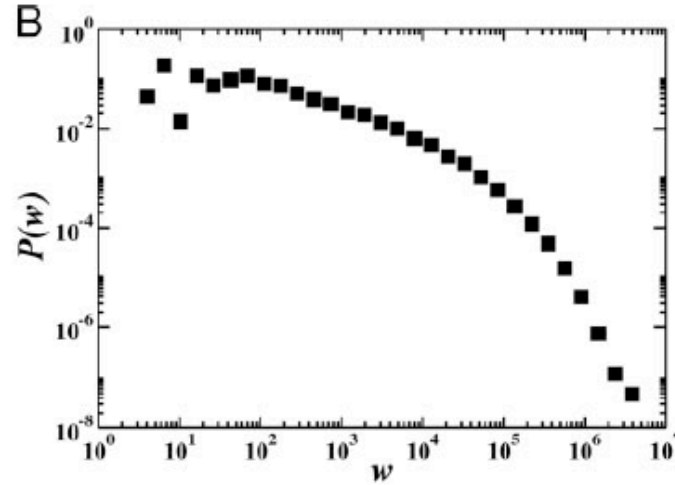
Focus on 3100 largest airports and 17182 edges accounting for 99% of the world-wide air traffic.

PNAS 103, 2015 (2006)

Properties of the world-wide airport network



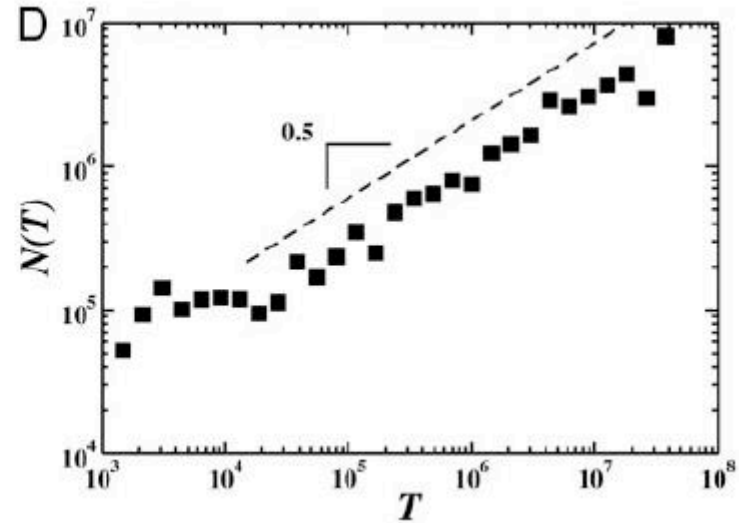
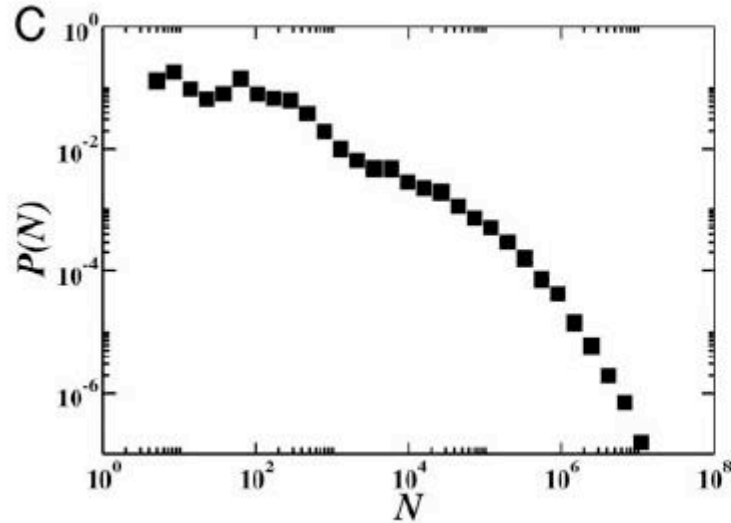
(A) The degree distribution $P(k)$ that an airport j has k_j connections to other airports follows a power-law behavior on almost two decades with exponent 1.8 ± 0.2 .



(B) The distribution of the weights (fluxes) is skewed and heavy-tailed.

PNAS 103, 2015 (2006)

Properties of the world-wide airport network



(C) The distribution of populations N_j is heavy-tailed distributed.

(D) The city population varies with the traffic of the corresponding airport as $N \approx T^\alpha$ with $\alpha \cong 0.5$.

PNAS 103, 2015 (2006)

Model disease propagation by SIR model for each airport

Use compartment model with 3100 compartments.

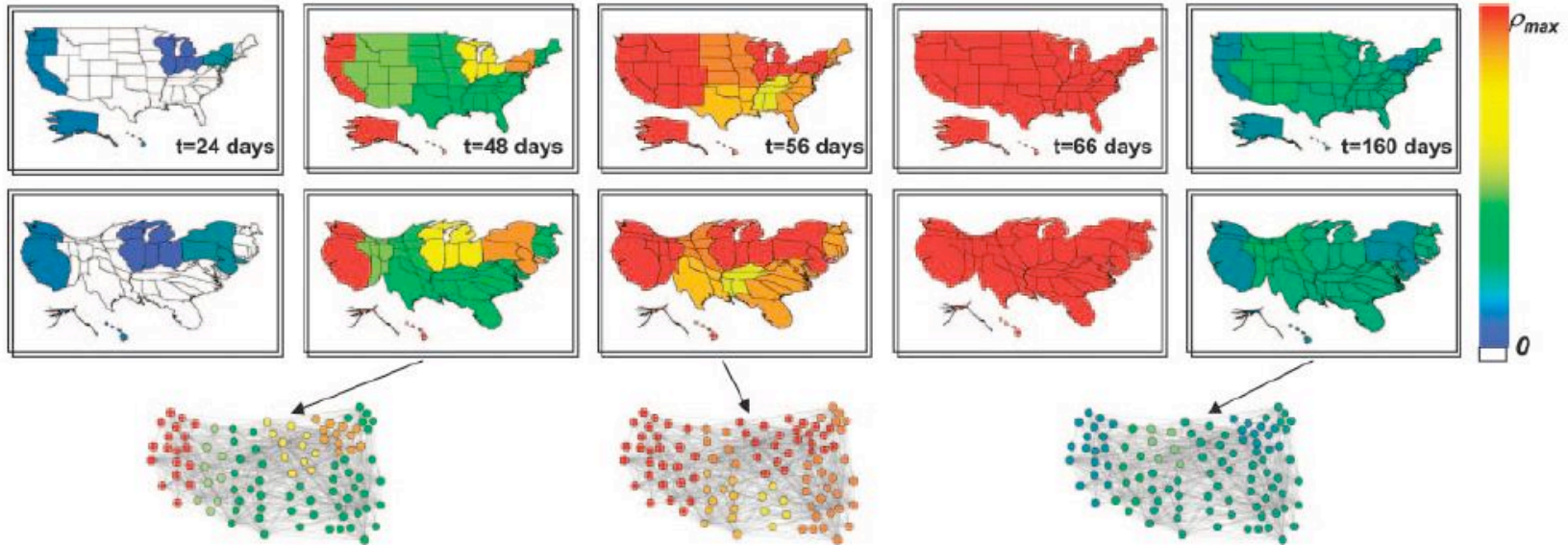
In each city (airport), individuals can exist in one of 3 discrete states:

susceptible (S), infected (I), permanently recovered (R)

Model dynamics by numerical integration of a set of 3×3100 stochastic differential equations.

PNAS 103, 2015 (2006)

Recent extensions in modelling disease spreading



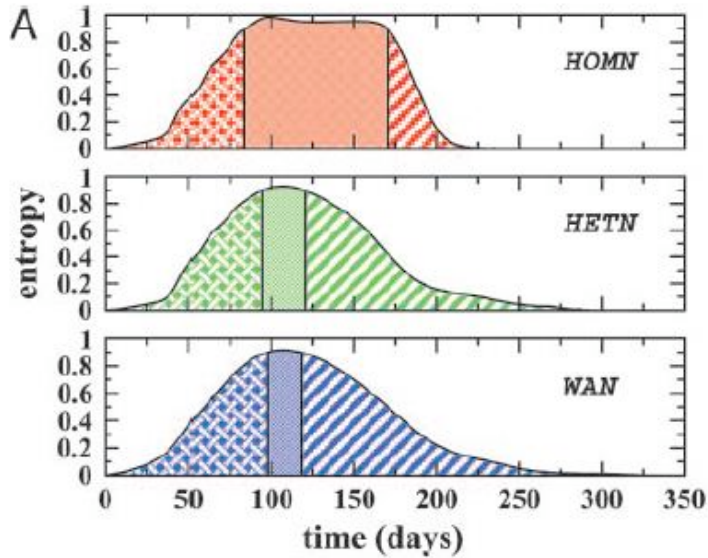
Geographical representation of the disease evolution in the United States for an epidemics starting in Hong Kong based on a SIR dynamics within each city. The color code corresponds to the prevalence in each region, from 0 to the maximum value reached (ρ_{max}).

Top row: the original United States maps,

Bottom row: obtained by rescaling each region according to its population.

Shown below are 3 representations of the airport network restricted to the 100 airports in the United States with highest traffic; the color is assigned in accordance to the color code adopted for the maps.

Heterogeneity of epidemic pattern



heterogeneity of epidemic pattern in actual network (*WAN*) is compared with 2 network models (*HOMN* and *HETN*).

HOMN: homogenous Erdos-Renyi random graph

HETN: topology of real network

For *HOMN* and *HETN*, all fluxes and populations are set to average value.

Shown is entropy $H(t) = \sum_j \rho_j(t) \log \rho_j(t)$.

$H(t) = 1$ means that epidemics is homogeneously distributed over all nodes.

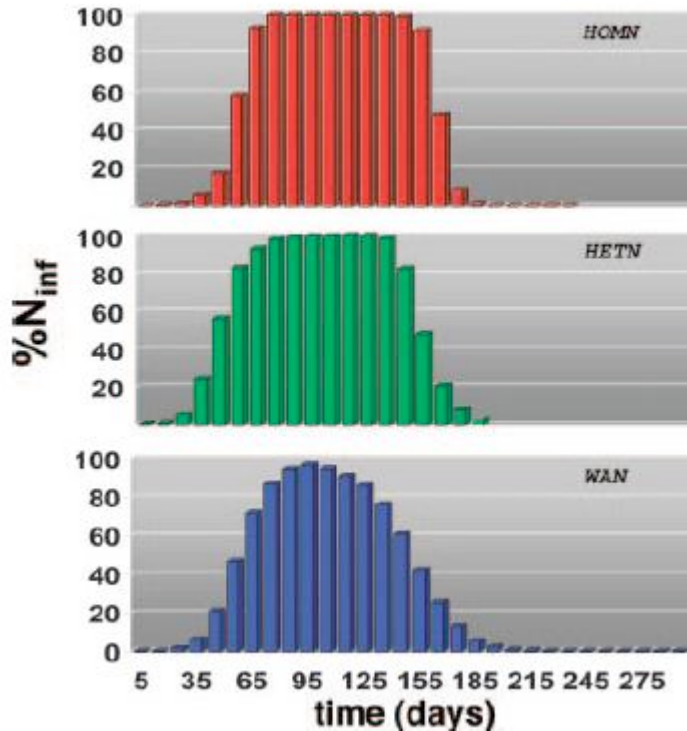
$H(t) = 0$ means that only 1 city is affected

Each profile is split into 3 different phases. The central phase corresponds to $H > 0.9$; i.e., to a homogeneous geographical spread of the disease.

This phase is much longer for the *HOMN* than for the real airport network *WAN*.

The behavior observed in *HETN* is close to the real case meaning that **the connectivity pattern plays a leading role in the epidemic behavior.**

What cities are affected?



Percentage of infected cities as a function of time for an epidemics starting in Hong Kong based on a SIR dynamics within each city.

The *HOMN* case displays a large interval in which all cities are infected.

The *HETN* and the real case show a smoother profile with long tails, signature of a long lasting geographical heterogeneity of the epidemic diffusion.

PNAS 103, 2015 (2006)

Quality of disease forecasts?

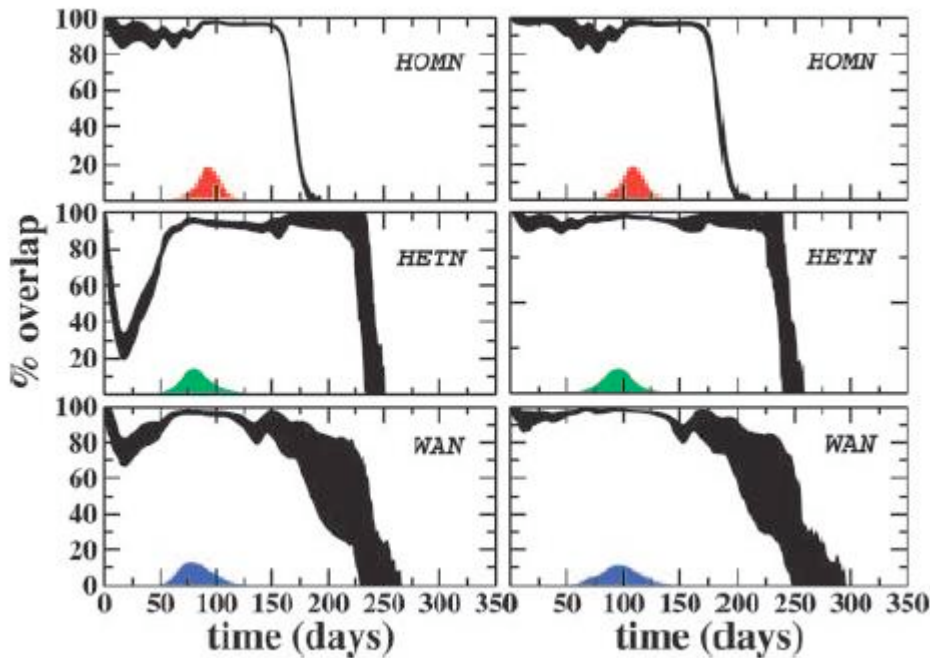
Measure similarity of 2 outbreak scenarios.

Overlap = 1 means that the very same cities have the very same number of infectious individuals in both realizations.

Overlap = 0 means that the two realizations do not have any common infected cities at time t .

PNAS 103, 2015 (2006)

Overlap as a function of time



(Left): initially infected city is an airport hub

(Right): initially infected city is a poorly connected city.

(Right): Outbreaks from poorly connected cities lead in all models to well reproducible behavior because only few connections are available.

(Left): Outbreaks from hubs lead to well reproducible behavior in the HOMN model.

The HETN and WAN models lead to considerable variation of the results here because the epidemics can take one of several possible channels.

PNAS 103, 2015 (2006)

Conclusions

- air-transportation-network properties seems responsible for the global pattern of emerging diseases.
- Large-scale mathematical models that take fully into account the complexity of the transportation matrix can be used to obtain detailed forecast of emergent disease outbreaks.
- To make the forecast more realistic, it is necessary to introduce more details in the disease dynamics. In particular, seasonal effects and geographical heterogeneity in the basic transmission rate (due to different hygienic conditions and health care systems in different countries) should be addressed.
- Also need to connect air transportation network with other transportation systems such as railways and highways for forecasting on longer time scales.

PNAS 103, 2015 (2006)