INTRODUCTION TO NETWORK SCIENCE

János Kertész
janos.kertesz@gmail.com

10. SPREADING ON NETWORKS

Network Dynamic Phenomena

- Dynamic processes on networks
  - Diffusion, random walk
  - Transport
  - Packet transfer according to protocol
  - Synchronization
  - Spreading

- Dynamics of networks
  - Network growth and evolution
  - Network restructuring
  - Network adaptation
  - Temporal networks

Random walk: Page Rank

Google ranking is a combination of heuristic elements and the probability that a random walker will find the page.

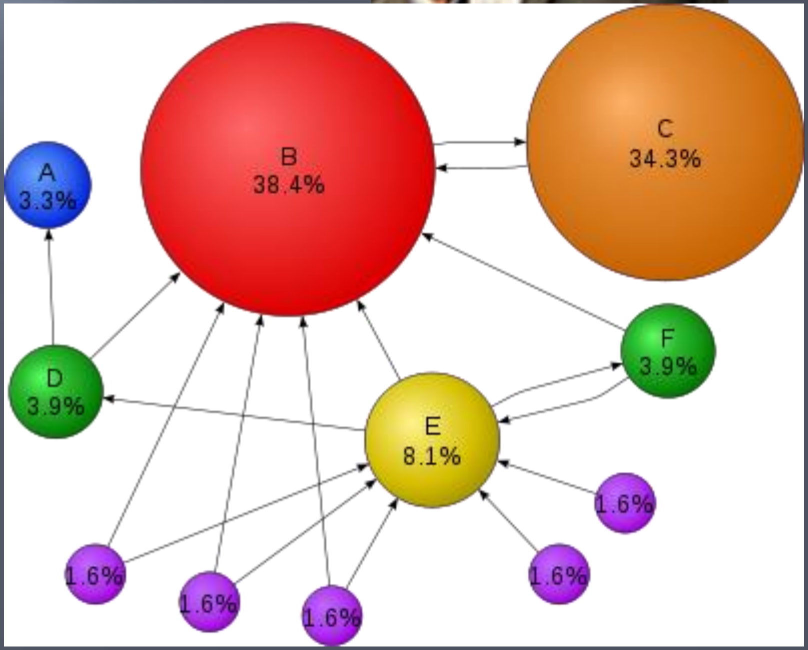
$$P_R(i) = \frac{q}{N} + (1-q)\sum_j A_{ij} \frac{P_R(j)}{k_{\text{out},j}}$$

self-consistent eq. iterative sol'n

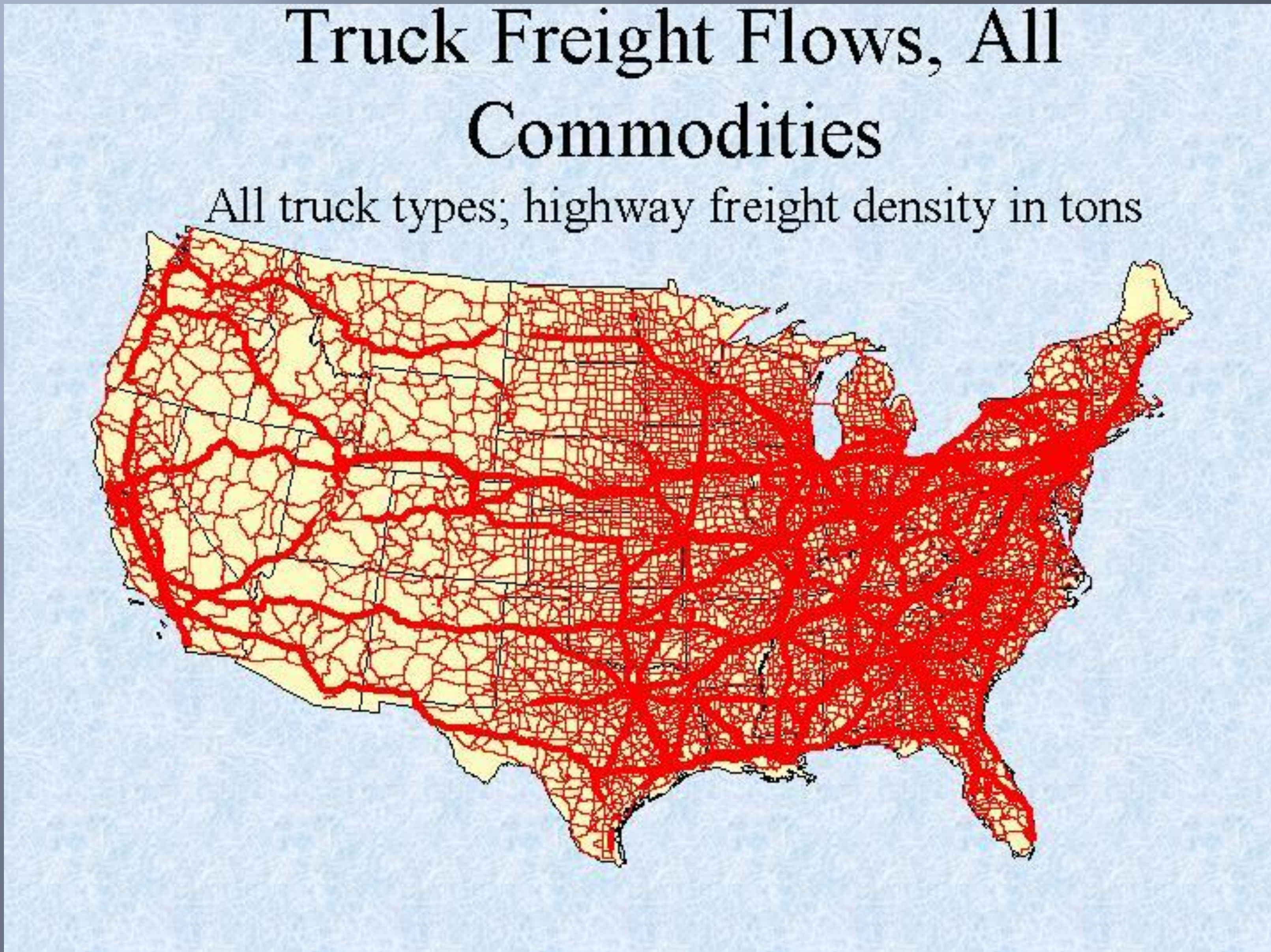
q is a damping factor: it mimics that after having not found, what we were looking for, we get bored and make random trials. It also avoids getting trapped (directed NW!). *N* is the total number of pages. *q* ~*0*.15 is used.

There are refined, unpublished algorithms but the core is PR.

Larry Page

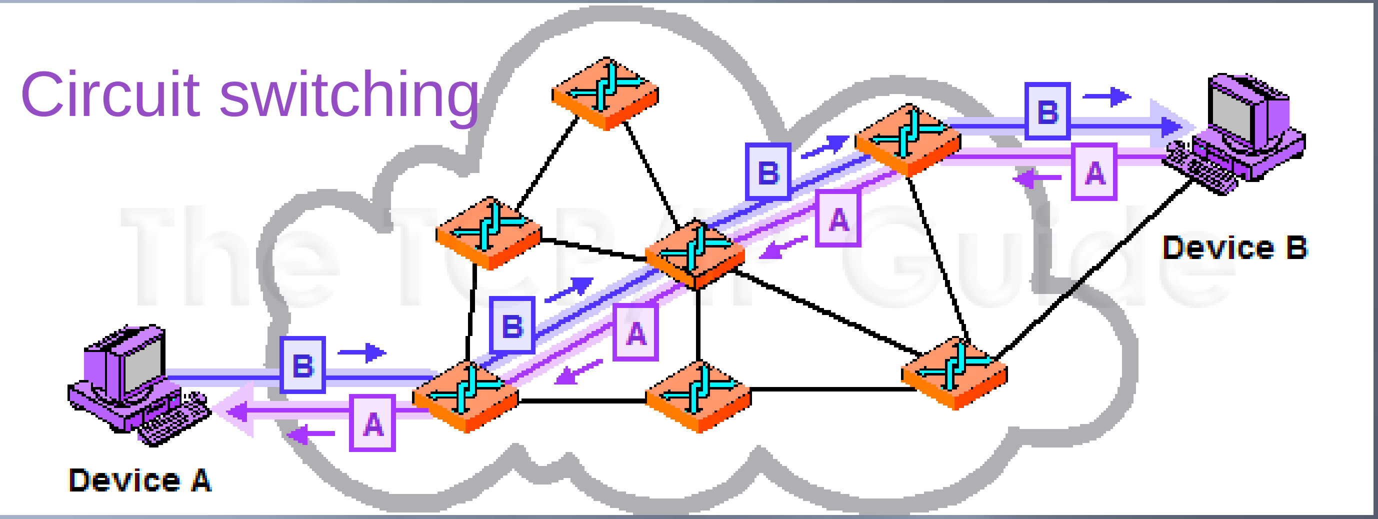


Transport

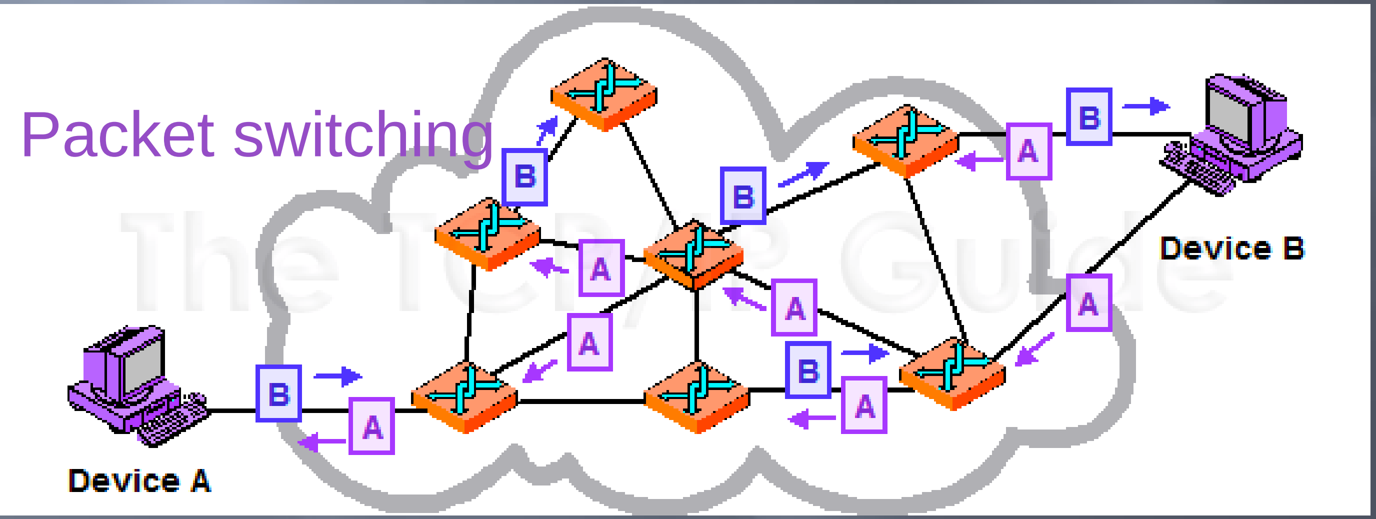


http://www.ops.fhwa.dot.gov/freight/Memphis/

Packet transfer according to protocol



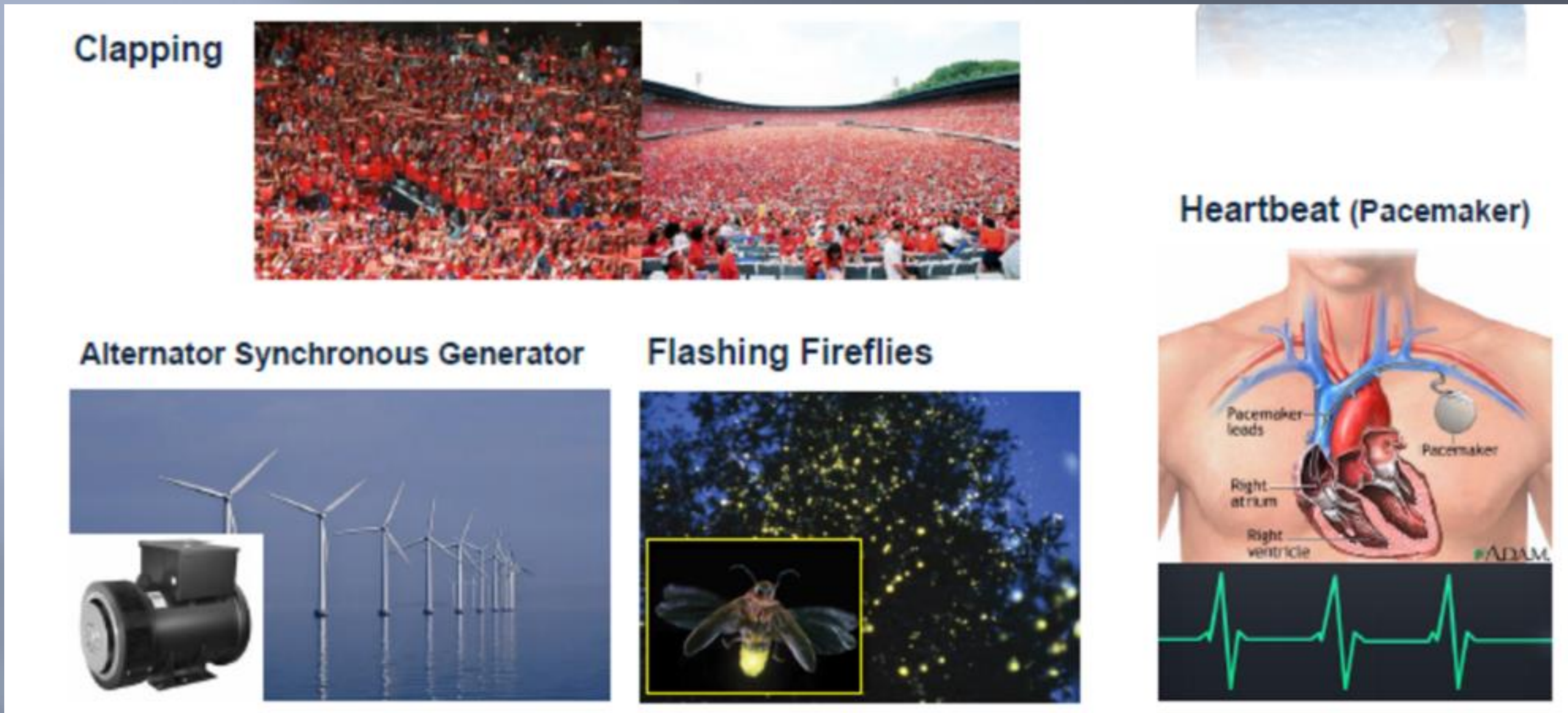
For communi-cation a route has to be established and kept open throughout the exchange of information



Information is chopped into pieces (packets), which travel on different routes and get reassembled finally

www.tcpipguide.com

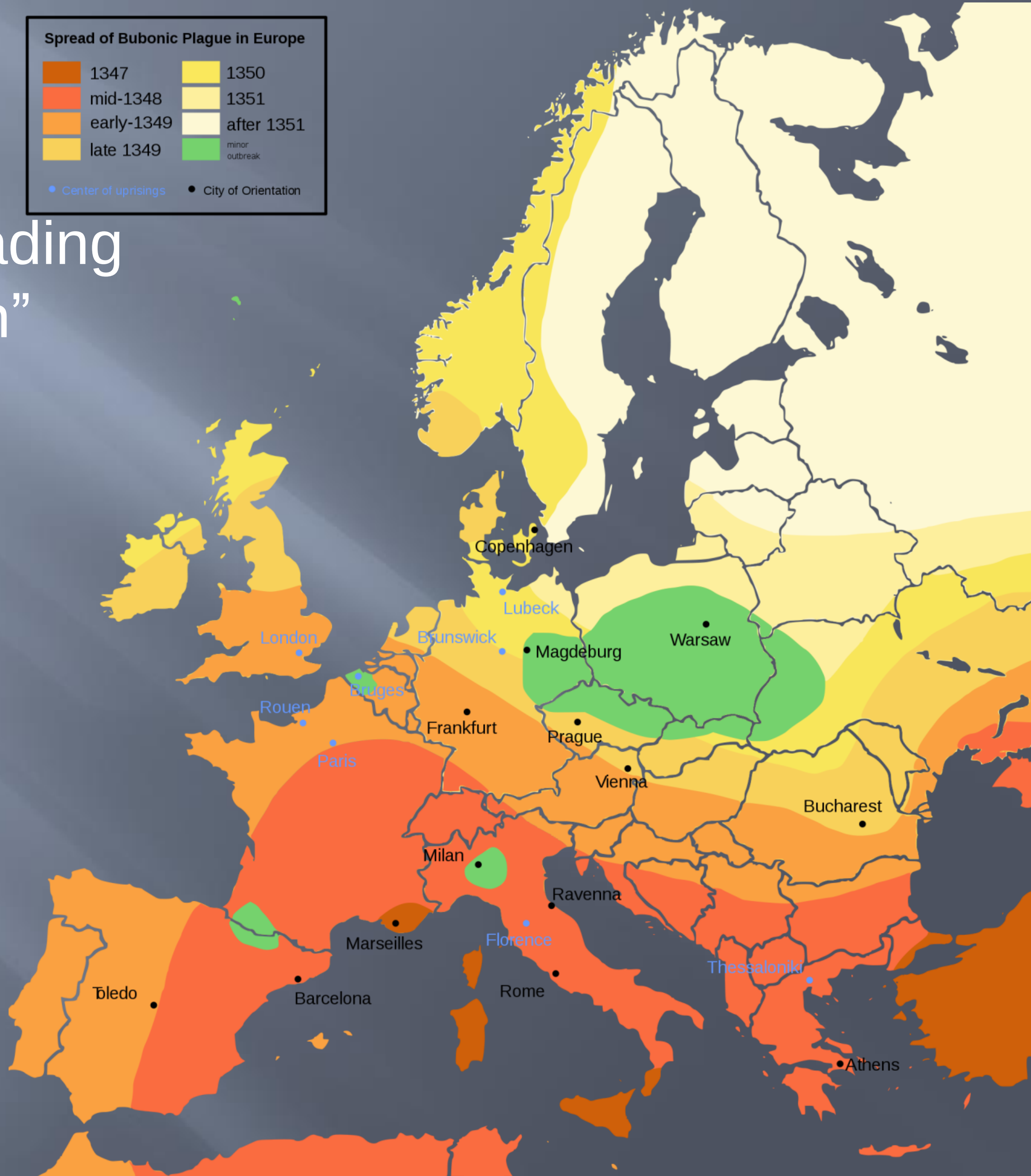
Synchronization



Couples oscillators show interesting phenomena on networks including phase transitions and structure sensitive behavior.

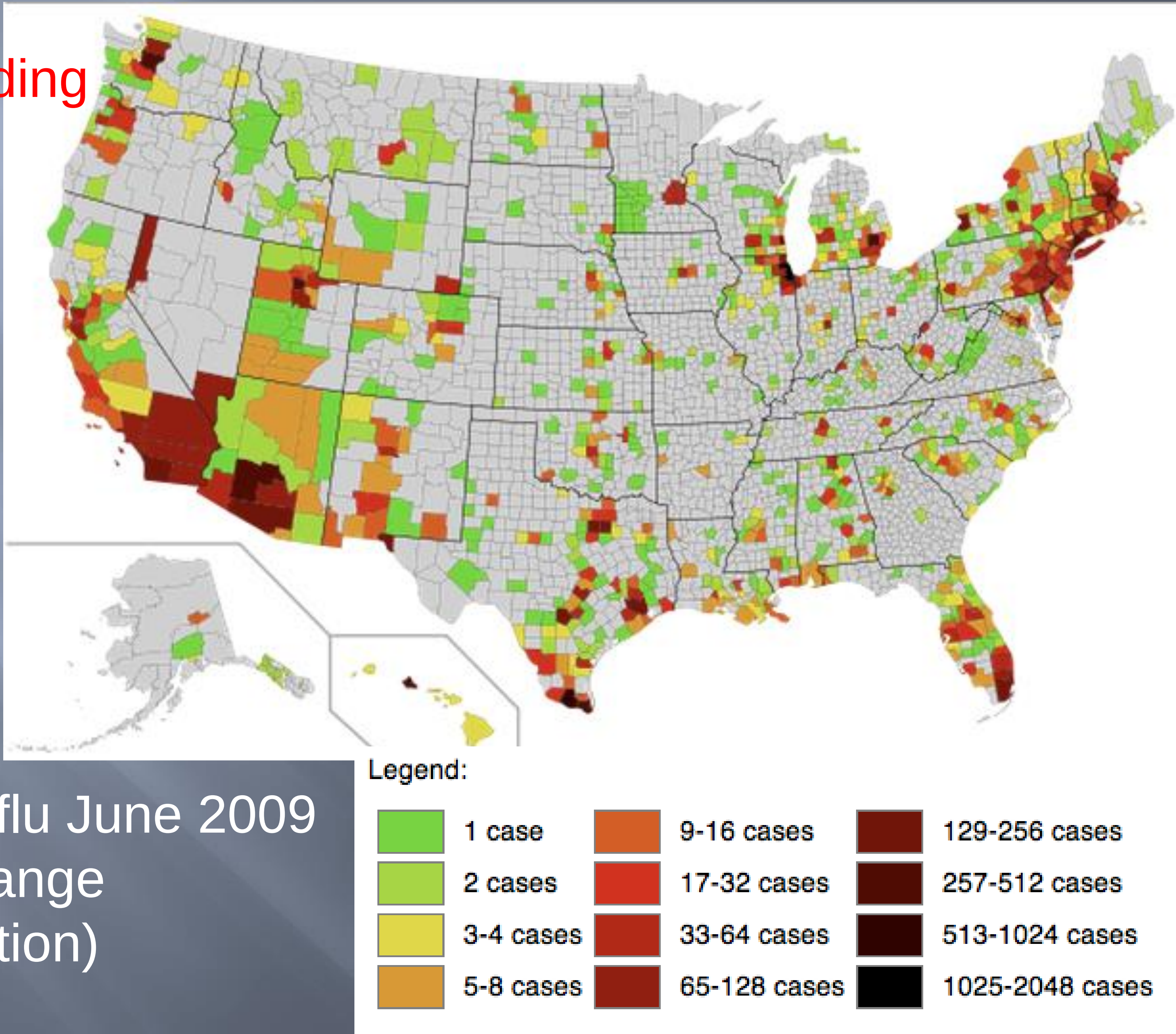
Spreading

Medieval spreading of „Black Death” (short range interaction)



http://www.historyofinformation.com/

Spreading



Swine flu June 2009 (long range interaction)

Spreading

Spreading of

- Disease
- Computer viruses
- Innovations
- Ideas
- Fashion
- Behavioral traits and cultural patterns
- …

In complex social contagion more than binary interactions are needed: Peer pressure

Important for

- Epidemiology
- Computer science
- Sociology
- Economics
- …

Spreading

Many approaches:

- Compartmental (mean field)
- Heterogeneous network
- Multi agent models
- Immunization strategies
- …

Relation to diffusion:
Agents move following a rule (e.g., diffusion) but *carry* an infectious property, which can be transmitted.

Another picture:
Temporal networks
Links are activated temporarily on a network enabling connection between infected and susceptible

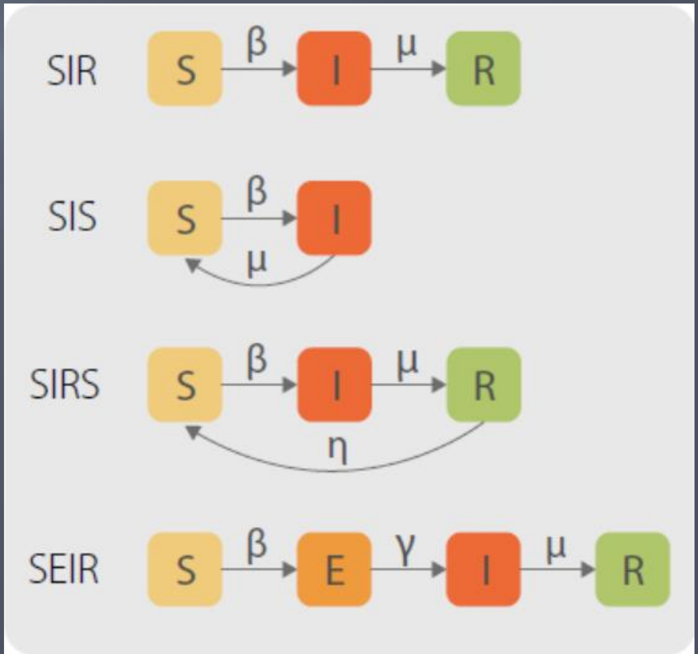
Basic notions

Epidemic spreading among individuals
Different states – compartments:

- **S**usceptible
- **I**nfected
- **R**ecovered (immune)
- **E**xposed (infected but not yet infecting)

Resulting in different models in the spirit of reaction-diffusion processes, e.g., $S + I \rightarrow 2I$.

β, μ, η, γ are rates by which the reactions happen. In the simplest case "homogeneous or perfect mixing" is assumed: Everybody can meet everybody with the probability proportional to the concentrations (mean field approximation).



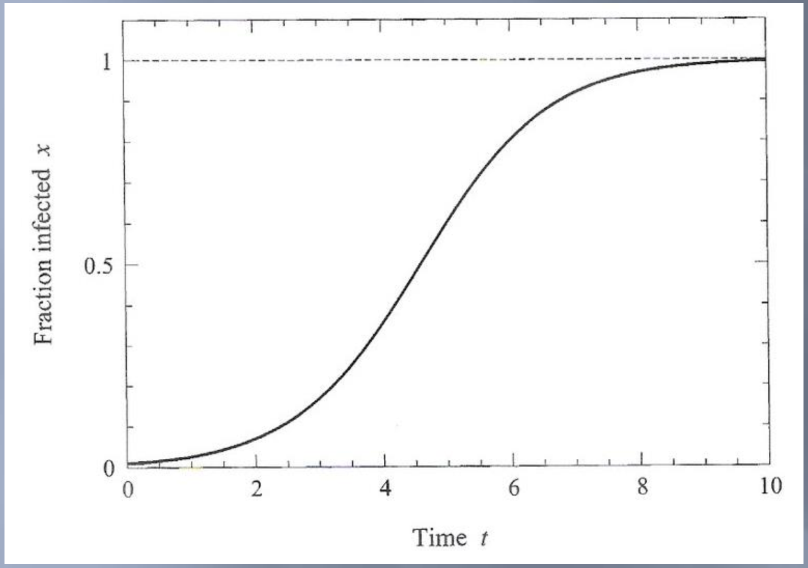
In the simple reactions not involving meeting between individuals from different compartment the description based on rates have a simple interpretation. E.g., for $I \to R$ reaction with rate μ a Poisson process of recovery is assumed, indicating that the probability density of recovery time is $\mu e^{-\mu t}$ with the average recovery time $1/\mu$. (In many cases the memory-less Poissonian assumption is not valid.)

If individuals from two compartments are involved, as for $S + I \to 2I$ we have to take into account the probability of meeting.

Perfect mixing means $\rho^\alpha = \frac{N^\alpha}{N}$, the densities of the individuals in compartment α characterize the situation. The equations for SIS and SIR:

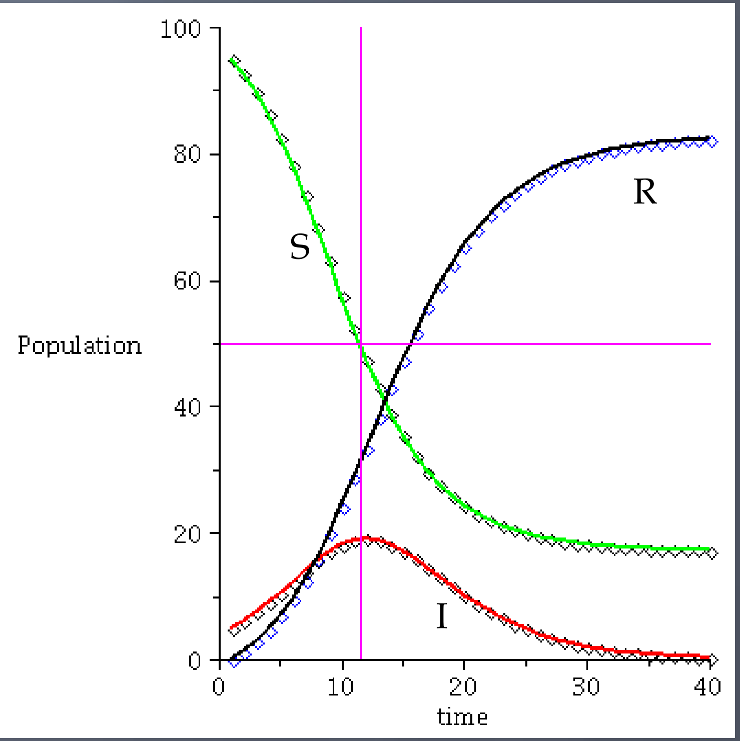
$$\frac{d\rho^I}{dt} = \beta\rho^I\rho^S - \mu\rho^I$$
$$\frac{d\rho^S}{dt} = -\beta\rho^I\rho^S + \chi\rho^I$$

These are deterministic equations (no fluctuations); $\chi = \mu$ for SI and 0 for SIR. With the normalization $\varrho^I = 1 - \varrho^S$ for SIS; $\varrho^R = 1 - \varrho^S - \varrho^I$ for SIR the equations are complete.



SI

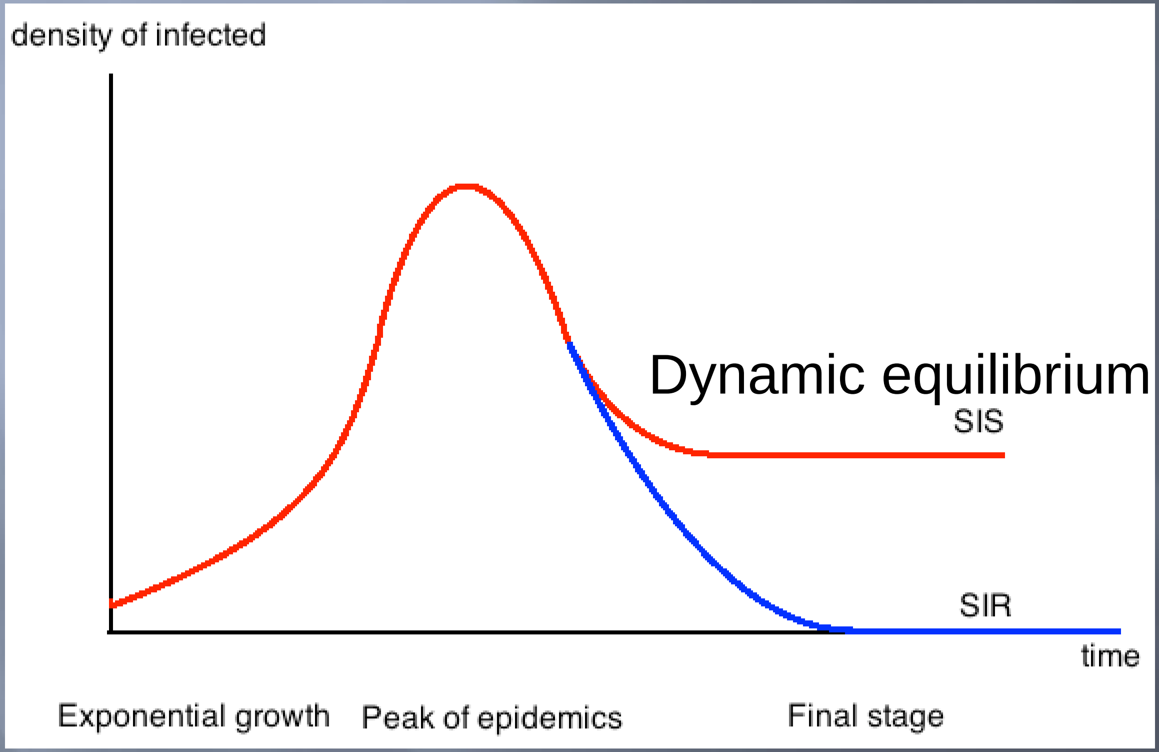
$$\rho^{I}(t) = \frac{\rho_0^I e^{\beta t}}{1 - \rho_0^I + \rho_0^I e^{\beta t}}$$



SIR

$$\frac{d\rho^{R}(t)}{dt} = \gamma\left[1 - \rho^{R}(t) - \rho_0^{S} \exp\left(-\beta\rho^{R}(t)/\gamma\right)\right]$$

Comparison of SIS and SIR: Above the epidemic threshold



Epidemic threshold

At the beginning $0 < \varrho^I \ll 1$ → linearization:

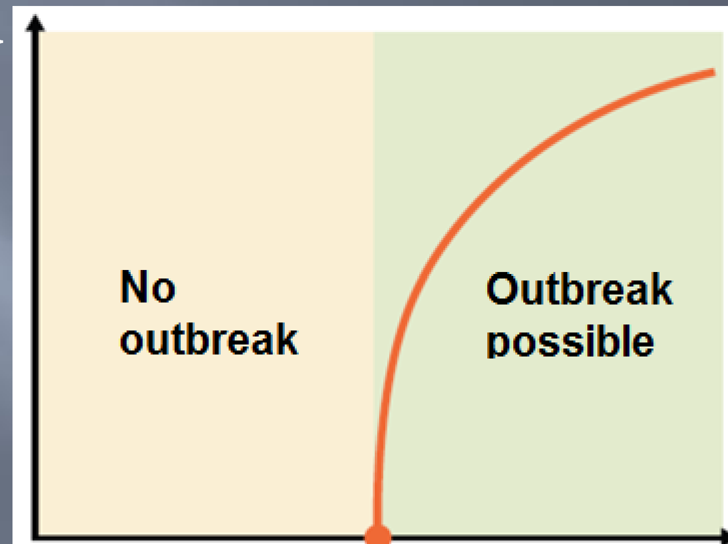
$$\frac{d\rho^I}{dt} \simeq (\beta - \mu)\rho^I \rightarrow \rho^I(t) \simeq \rho^I(0)e^{(\beta-\mu)t}$$

Exponential growth for $\beta - \mu > 0 \quad \Rightarrow \quad R_0 = \frac{\beta}{\mu} > 1$ with R_0 *basic reproduction number*. $R_0 = 1$ is the *epidemic threshold* above which there is a macroscopic outbreak in the SIR and a nonzero asymptotic density of *I* in the SIS model.

Relation to phase transititons

For SIR: Mapping to percolation

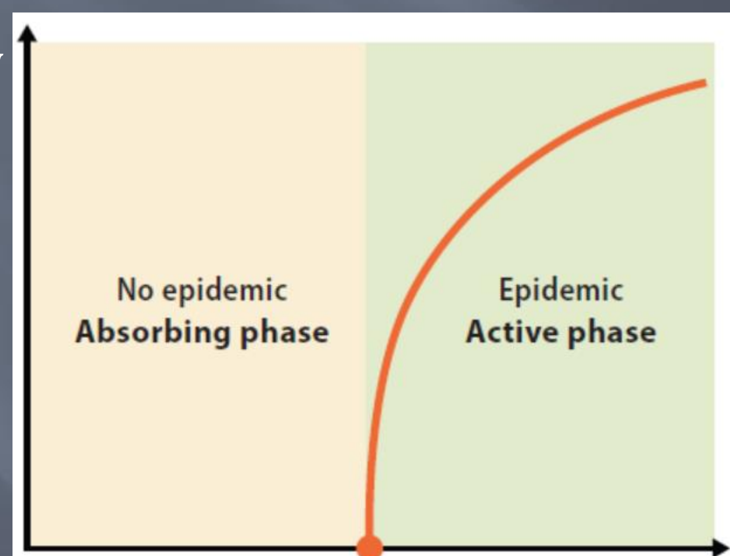
$P\{\rho^R \to \infty\}$



R

For SIS: dynamic phase transition with an absorbing phase

ρ^I



R

Homogeneous models

Limitations of the homogeneous models:
Fluctuations are ignored!

Fluctuations in:
i) Number of individuals in the different compartments
ii) Contacts
iii) Transmission etc. rates (ignored here; multi agent)

i) The epidemic threshold has a probabilistic meaning. Above the threshold the probability of an outbreak is non-zero but less than 1. Fluctuations may lead to extinction above the threshold!

ii) Spreading takes place on the contact network

Spreading on networks

We denote the density of susceptible nodes with degree k as s_k and that of the infected as i_k. The corresponding equations will be, e.g., (we normalize by $N_k=NP(k)$):

$$s_k + i_k + r_k = 1$$

$$\frac{ds_k}{dt} = -s_k \sum_j \beta_{kj} i_j$$

$$\frac{di_k}{dt} = s_k \sum_j \beta_{kj} i_j - \gamma i_k$$

This is also mean field but much better.
It is sensitive to the special role of the hubs.

Spreading on networks

The SI model:

$$s_k + i_k = 1$$

$$\frac{di_k}{dt} = (1 - i_k) \sum_j \beta_{kj} i_j$$

$$\beta_{kj} = k \frac{j-1}{j} \beta P(j|k)$$

k possibilities; $\frac{j-1}{j}$ because neighbor node got infected from somewhere

In the last step we ignored degree-degree correlations ($P(j|k) = jP(j)/\langle k \rangle$).

$$\frac{di_k(t)}{dt} = \beta(1 - i_k) k \Theta(t)$$

$$\Theta(t) = \frac{\sum_j (j-1) P(j) i_j(t)}{\langle k \rangle}$$

independent of k

The linearized (early stage) equations will then be:

$$\frac{di_k(t)}{dt} = \beta k \Theta(t)$$

$$\frac{d\Theta(t)}{dt} = \beta \left(\frac{\langle k^2 \rangle}{\langle k \rangle} - 1 \right) \Theta(t)$$

Spreading on networks

$$\frac{di_k(t)}{dt} = \beta k \Theta(t)$$

$$\frac{d\Theta(t)}{dt} = \beta\left(\frac{\langle k^2\rangle}{\langle k\rangle} - 1\right)\Theta(t)$$

Can be solved for uniform initial condition: $i_k\ (t = 0) = i_0$

$$i_k(t) = i_0\left(1 + \frac{k(\langle k\rangle - 1)}{\langle k^2\rangle - \langle k\rangle}\left(e^{t/\tau} - 1\right)\right)$$

Exponential growth.
Larger degree nodes display faster prevalence.

with $$t = \frac{\langle k\rangle}{b\left(\langle k^2\rangle - \langle k\rangle\right)}$$

Total rate of infected:

$$i(t) = \sum_k i_k(t)P(k) = i_0\left(1 + \frac{\langle k\rangle^2 - \langle k\rangle}{\langle k^2\rangle - \langle k\rangle}\left(e^{t/\tau} - 1\right)\right) = i_0\left(1 + \frac{\langle k\rangle - 1}{\kappa - 1}\left(e^{t/\tau} - 1\right)\right)$$

With $k = \frac{\langle k^2\rangle}{\langle k\rangle}$ inhomogeneity ratio

Spreading on networks

SIR model

Following a similar line of thought and taking into account the correction due to recovery:

$$\frac{di_k(t)}{dt} = bks_k(t)Q(t) - mi_k(t)$$

leading to

$$t = \frac{\langle k \rangle}{b\langle k^2 \rangle - (m+b)\langle k \rangle} = \frac{1}{bk - (m+b)}$$

Spreading if

$$k > \frac{m}{b} + 1$$

Epidemic threshold. For an infinite scale free network with a degree exponent ≤ 3 we get $\kappa = \infty$, → null epidemic threshold, i.e., for any nonzero rates there is spreading! The inhomogeneity parameter governs the epidemic threshold, similarly to the percolation and resilience thresholds

Immunization

The outspread of dangerous diseases should be preventively hindered by vaccination, a process which intentionally transforms S to R.

Of course, if every newborn baby is vaccinated, the population is safe. This is the way, how smallpox (Variola) was defeated.
Estimated death in 20th century: 300 Million
Estimated infected in 1967: 15 Million
1979: WHO declared smallpox eradicated

Compulsory vaccination of all babies.

Vaccination is expensive

Immunization

What is the good strategy if only a part of the population can be vaccinated?

Simplest: Uniform immunization density.

Mean field: $\tau^{-1} = a_{S\to I} - a_I$

$$R_0 = \frac{a_{S\to I}}{a_I} \begin{cases} >1 \text{ outbreak} \\ =1 \text{ threshold} \\ <1 \text{ localized} \end{cases}$$

If the density of immune vertices is g

$$a_{S\to I} \to a_{S\to I}(1-g)$$

We have to choose g such

$$\frac{a_{S\to I}(1-g)}{a_I} < 1 \Rightarrow g_c = 1 - \frac{a_I}{a_{S\to I}}$$

Immunization

Uncorrelated inhomogeneous network:

$$\frac{b(1-g_c)}{m} = \kappa^{-1} = \frac{\langle k \rangle}{\langle k^2 \rangle}$$

For an infinite scale free network with a degree exponent ≤ 3 we get $\kappa = \infty$, thus g_c=1, i.e., everybody has to be vaccinated. This is in full accord with previous results that the epidemic threshold is 0 and the percolation threshold is 1.

Uniform vaccination is not a good strategy in a scale free network!

Immunization

We have seen that a scale free network is robust against random failures ←→ the epidemics is "robust" against uniform vaccination. Reason: Hubs

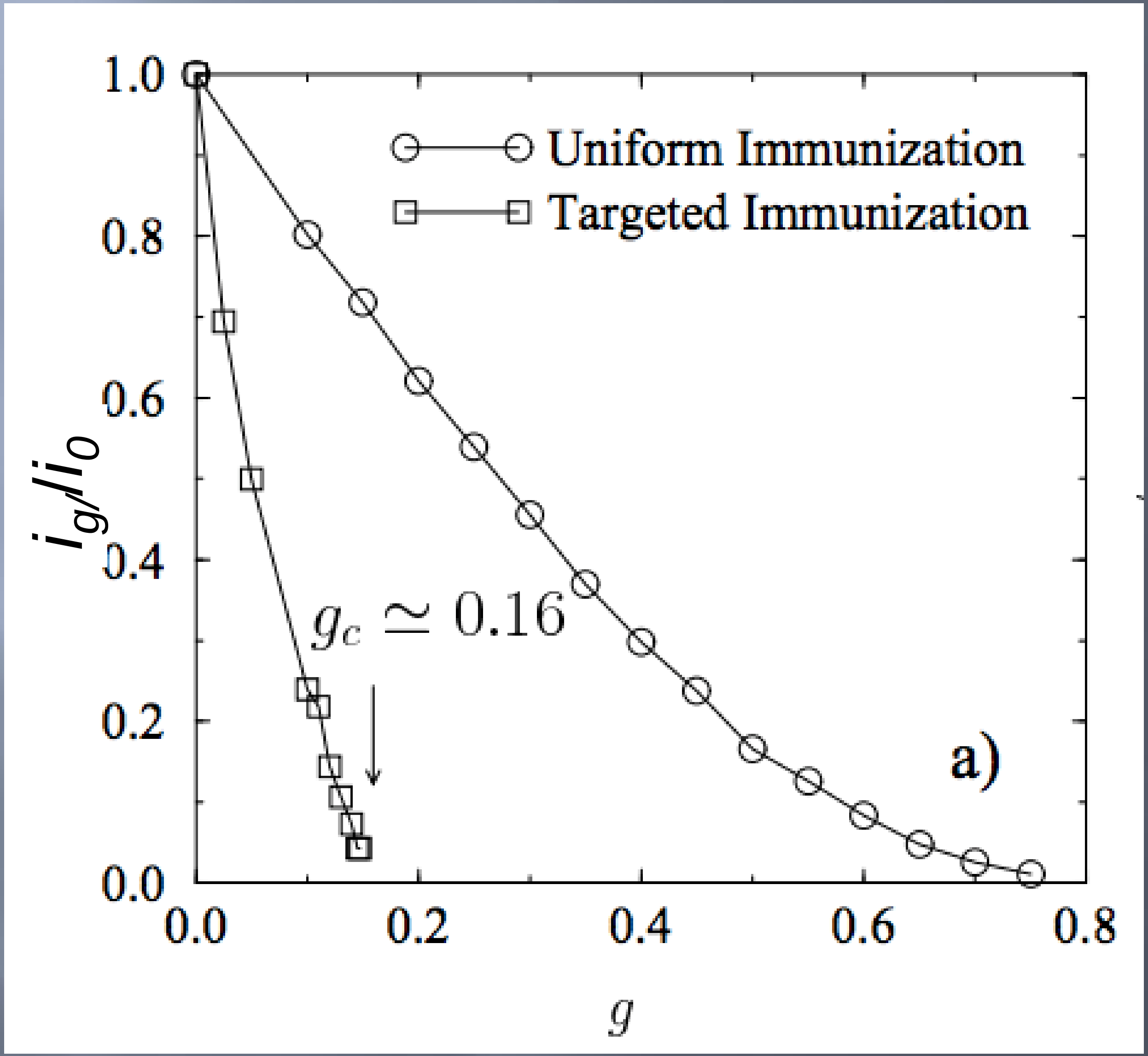
A scale free network is vulnerable against intentional attacks ←→ targeted vaccinations. Reason: Hubs

What happens if we remove fraction *g* of the nodes with highest degree? This introduces an upper cutoff in the degrees: $k_c(g)$; all vertices with $k > k_c$ will be immune. The protection of the network will be achieved

$$\frac{\langle k \rangle_g}{\langle k^2 \rangle_g} > \frac{b}{m}$$

which defines the critical value of *g*

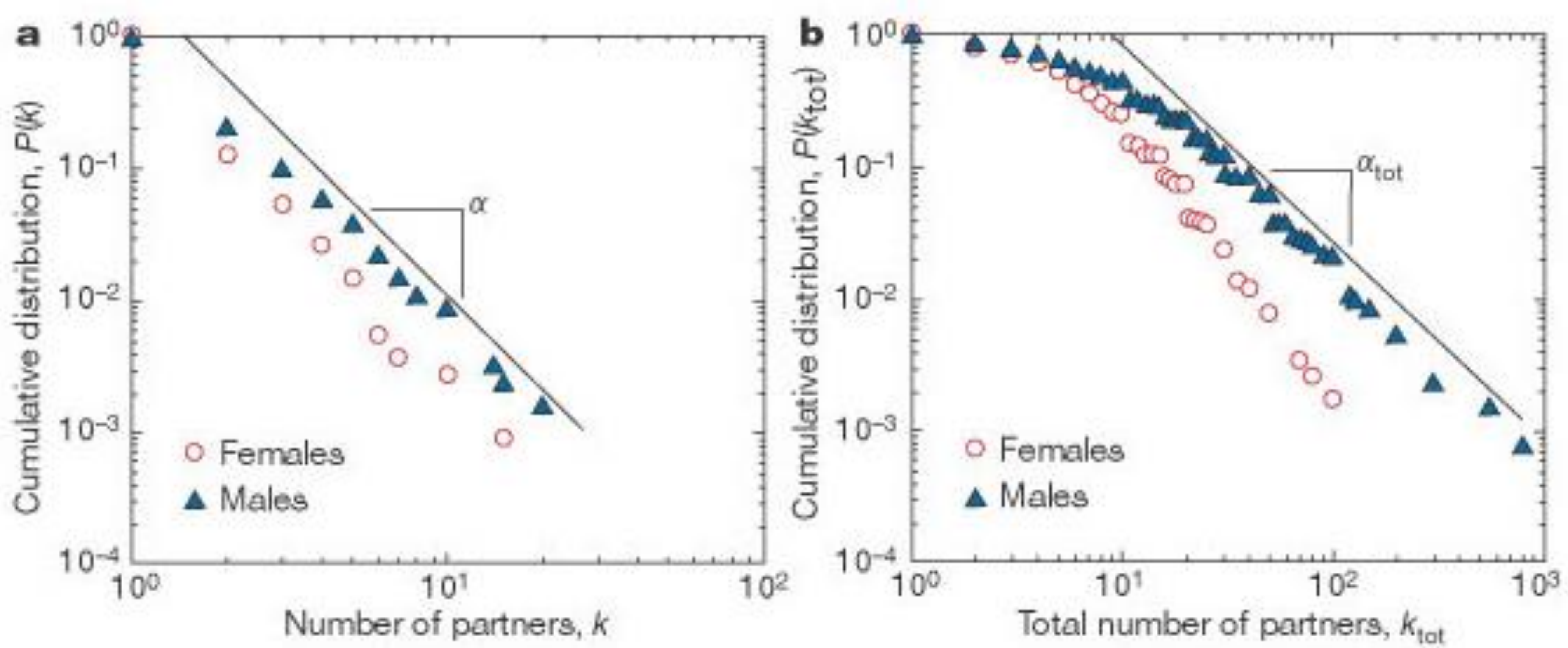
Immunization



Pastor-Satorras, Vespignani, 2001

Immunization

This strategy assumes knowledge about the degrees of individuals – which is not known!



Liljeros et al. 2000

Efficient immunization without global knowledge

Immunization

Select a fraction *g* at random.
Go to their neighbors and immunize them!

This strategy has a high chance to find the hubs.

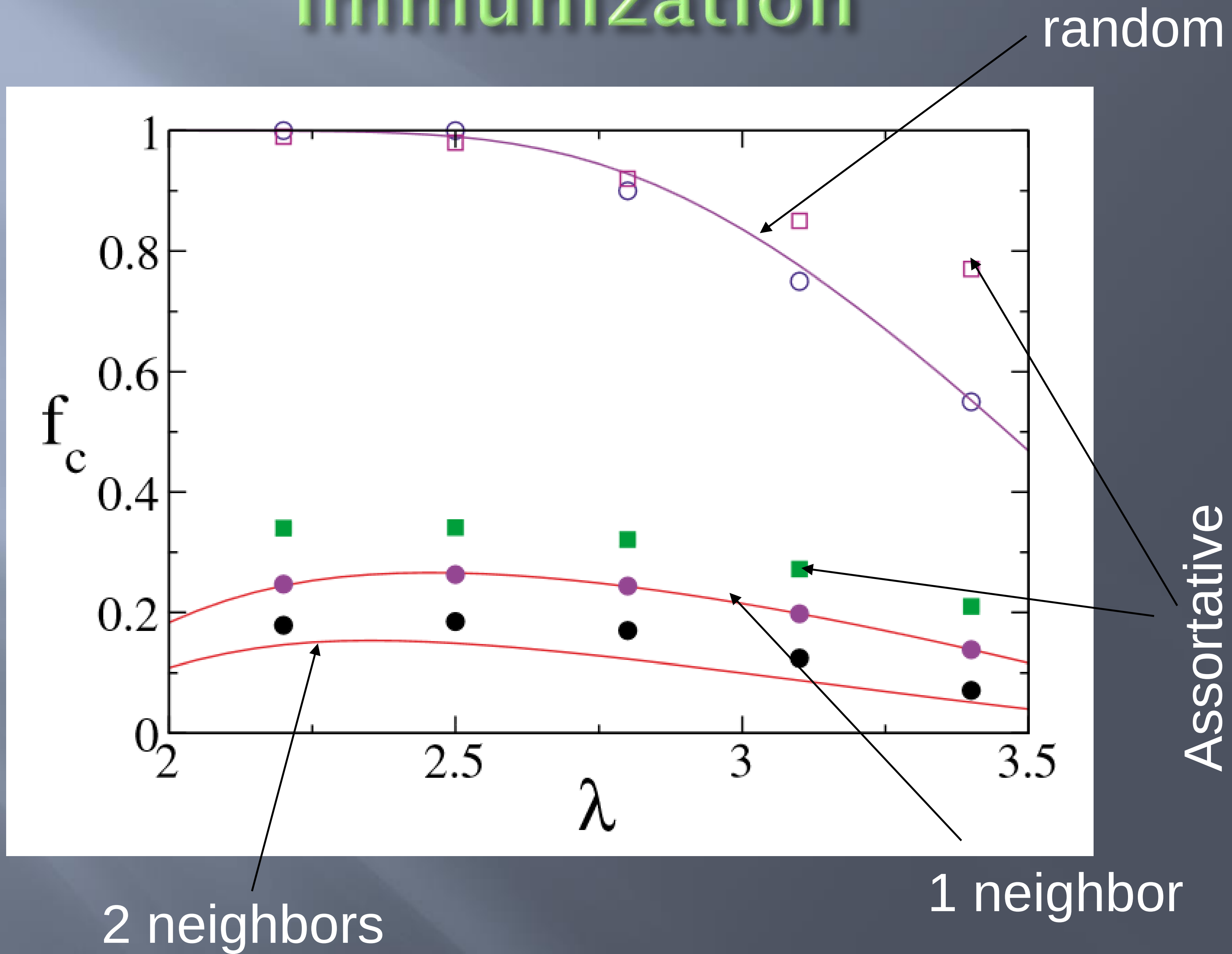
Even better: Chose the neighbor with highest degree.

Make a 2 (or n) step walk towards highest degree neighbors.

Since networks are small worlds, we find a hub.

Cohen et al . 2003

Immunization



Social contagion: info, rumors, innovation

First: How is the society structured?

The strength of weak ties (M.Granovetter, 1973)

Hypothesis about the small scale (micro-) structure of the society:

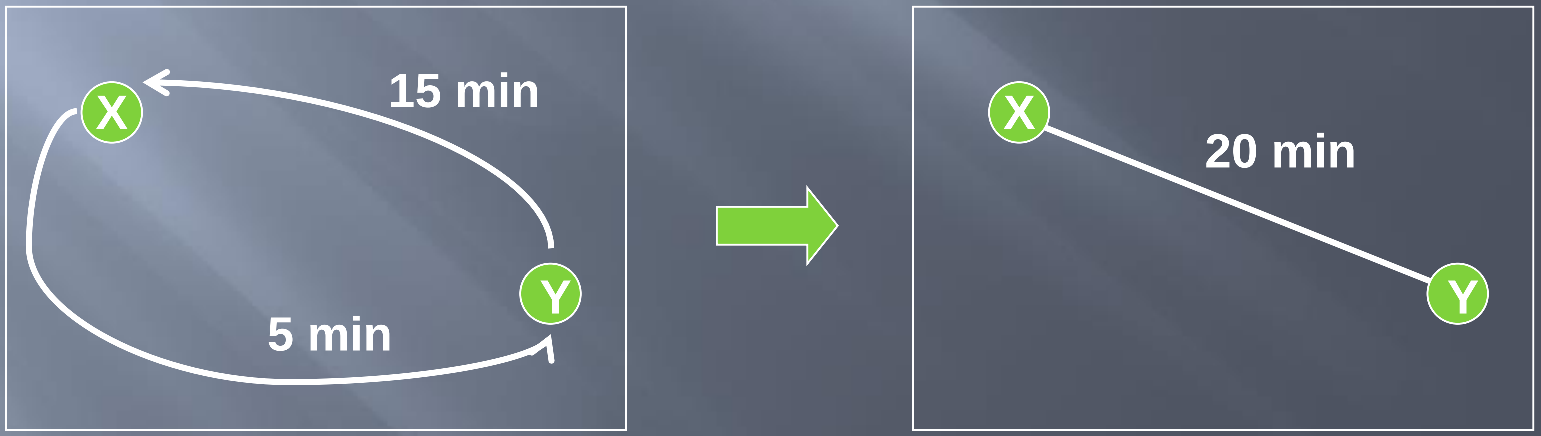
1. "The strength of a tie is a (probably linear) combination of the amount of time, the emotional intensity, the intimacy (mutual confiding), and the reciprocal services which characterize the tie."
2. "The stronger the tie between A and B, the larger the proportion of individuals S to whom both are tied."

Consequences on large (macro-) scale:
Society consists of strongly wired communities linked by weak ties.
The latter hold the society together.

Constructing social network from mobile phone data

- Over 7 million private mobile phone subscriptions
- Focus: voice calls within the home operator

- Data aggregated from a period of 18 weeks
- Require reciprocity (X→Y AND Y→X) for a link



- Customers are anonymous (hash codes)
- Data from an European mobile operator

J.-P. Onnela, et al. (2007)

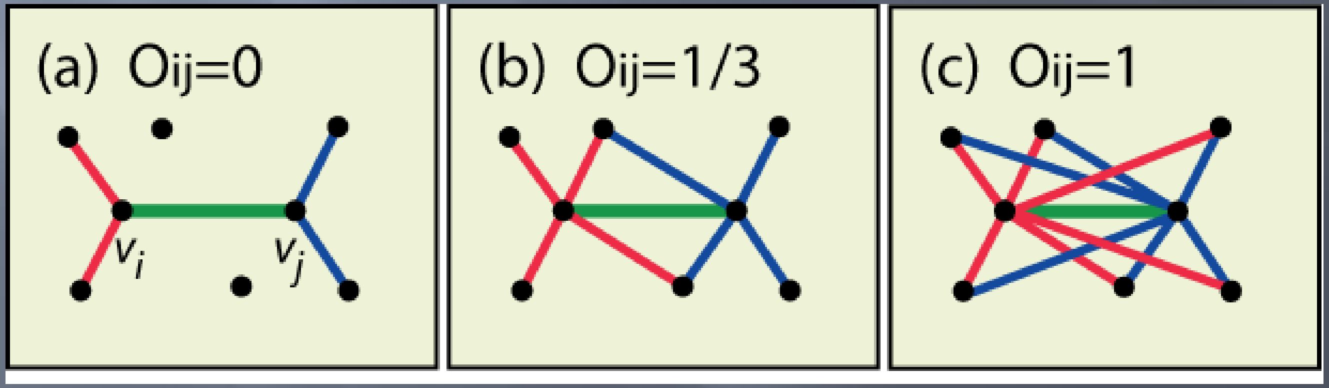
Overlap

- **Definition: relative neighborhood overlap (topological)**

$$O_{ij} = \frac{n_{ij}}{(k_i - 1) + (k_j - 1) - n_{ij}}$$

where the number of triangles around edge (*vi*, *vj*) is *nij*

- Illustration of the concept:

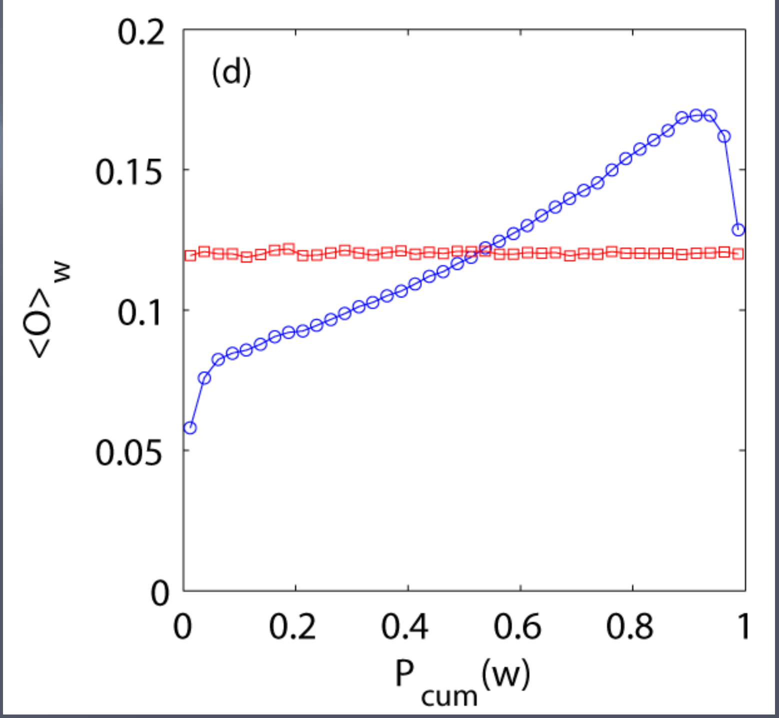


Empirical Verification

- Let $\langle O \rangle_w$ denote O_{ij} averaged over a bin of *w*-values

$$P_{\text{cum}}(w') = \sum_{w \le w'} P(w)$$

- Use cumulative link weight distribution:
  (the fraction of links with weights less than w')
  - Relative neighbourhood overlap increases as a function of link weight
    ⇒Verifies Granovetter's hypothesis (~95%)
    (Exception: Top 5% of weights)



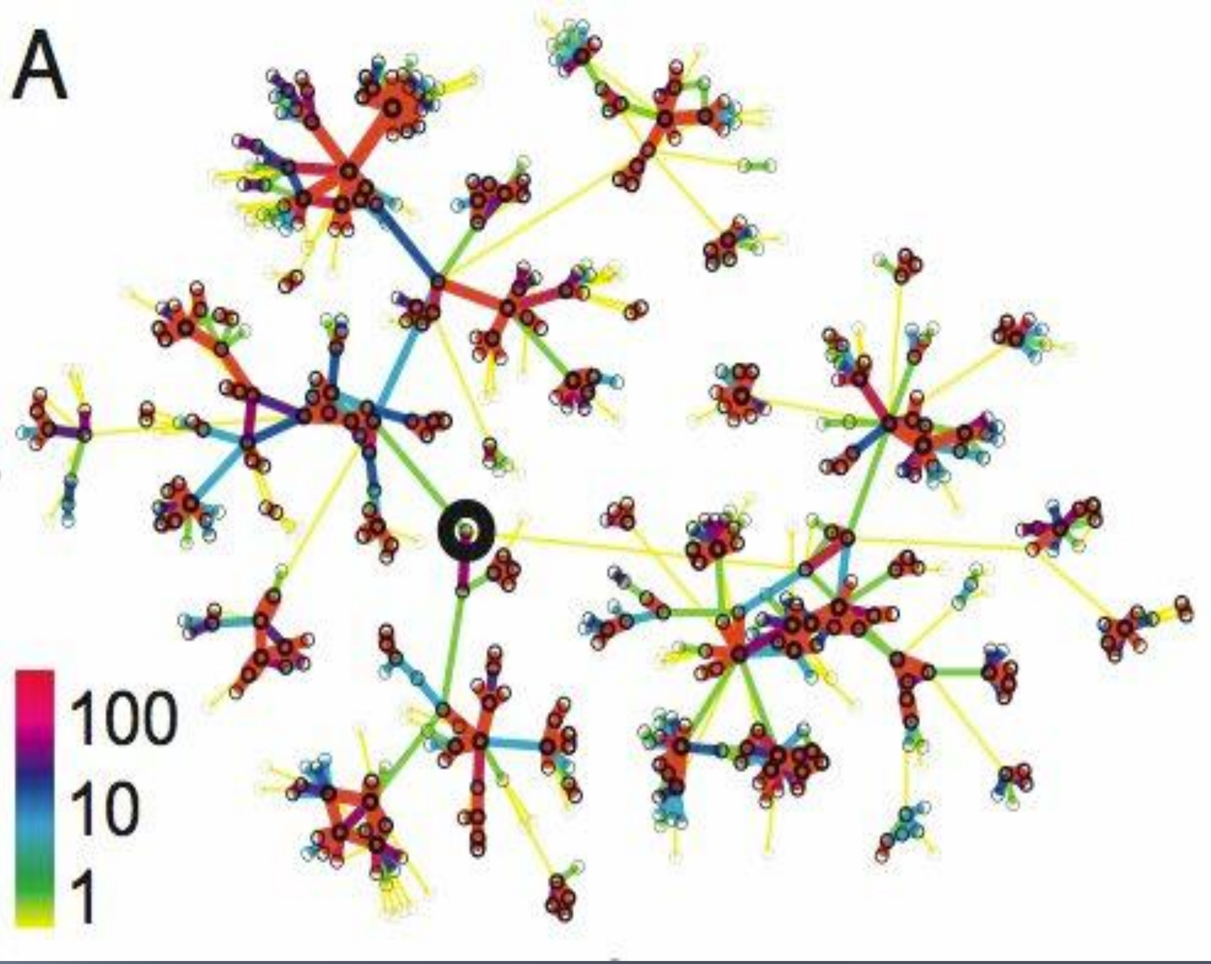
Blue curve: empirical network

Red curve: weight randomised network

Aggregate networks

The picture depends on type of question we ask.

Assuming that mobile phone calls represent social contacts, the aggregate network of call events is a proxy for the weighted human interaction network at sociatal level.



Granovetterian structure strongly wired communities linked by weak ties.

Onnela et al. PNAS 2007

Spreading of information

Knowledge of information diffusion based on unweighted networks

Use the present network to study diffusion on a weighted network:

Does the topology and tie strength relationship affect spreading?

Spreading simulation: infect one node (with information); play SI

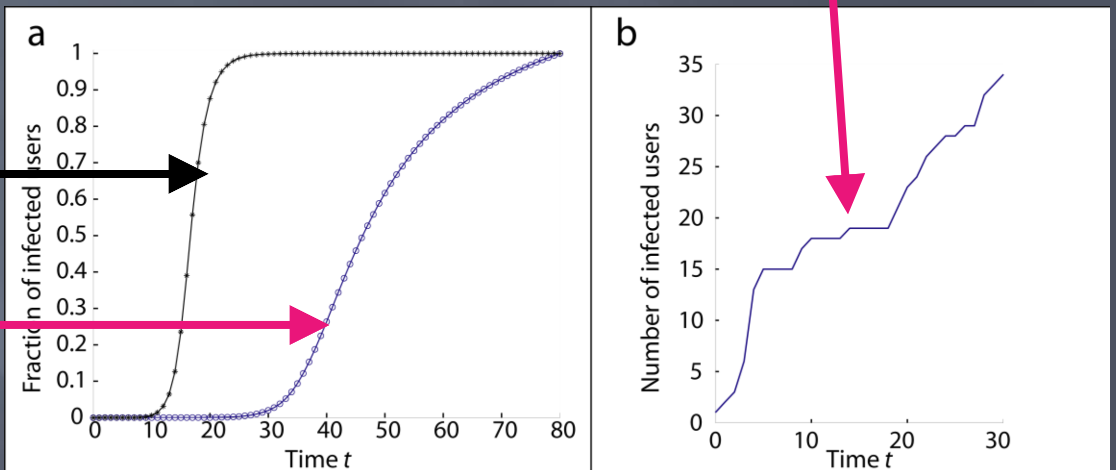
(1) Empirical: $p_{ij} \propto w_{ij}$

(2) Reference: $p_{ij} \propto \langle w \rangle$

Spreading significantly faster on the reference (average weight) network because information gets trapped in communities in the real network

Reference

Empirical



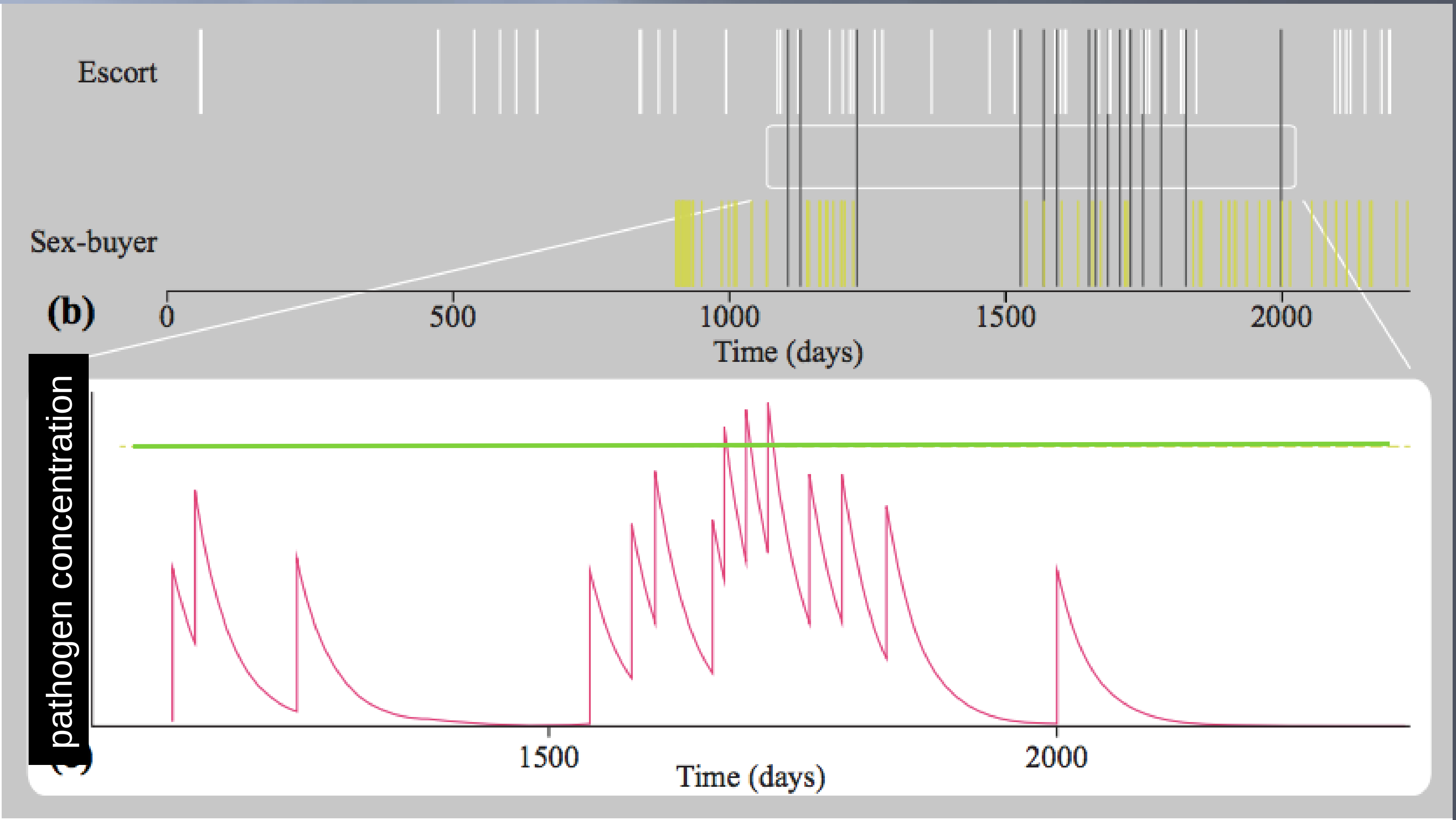
Spreading on a temporal network: Information on a call network

Compared to the simple SI model correlations:

- Topology (communities)
- Granovetterian structure: topology – weight corr.
- Burstiness
- Periodicities
- Triggered events

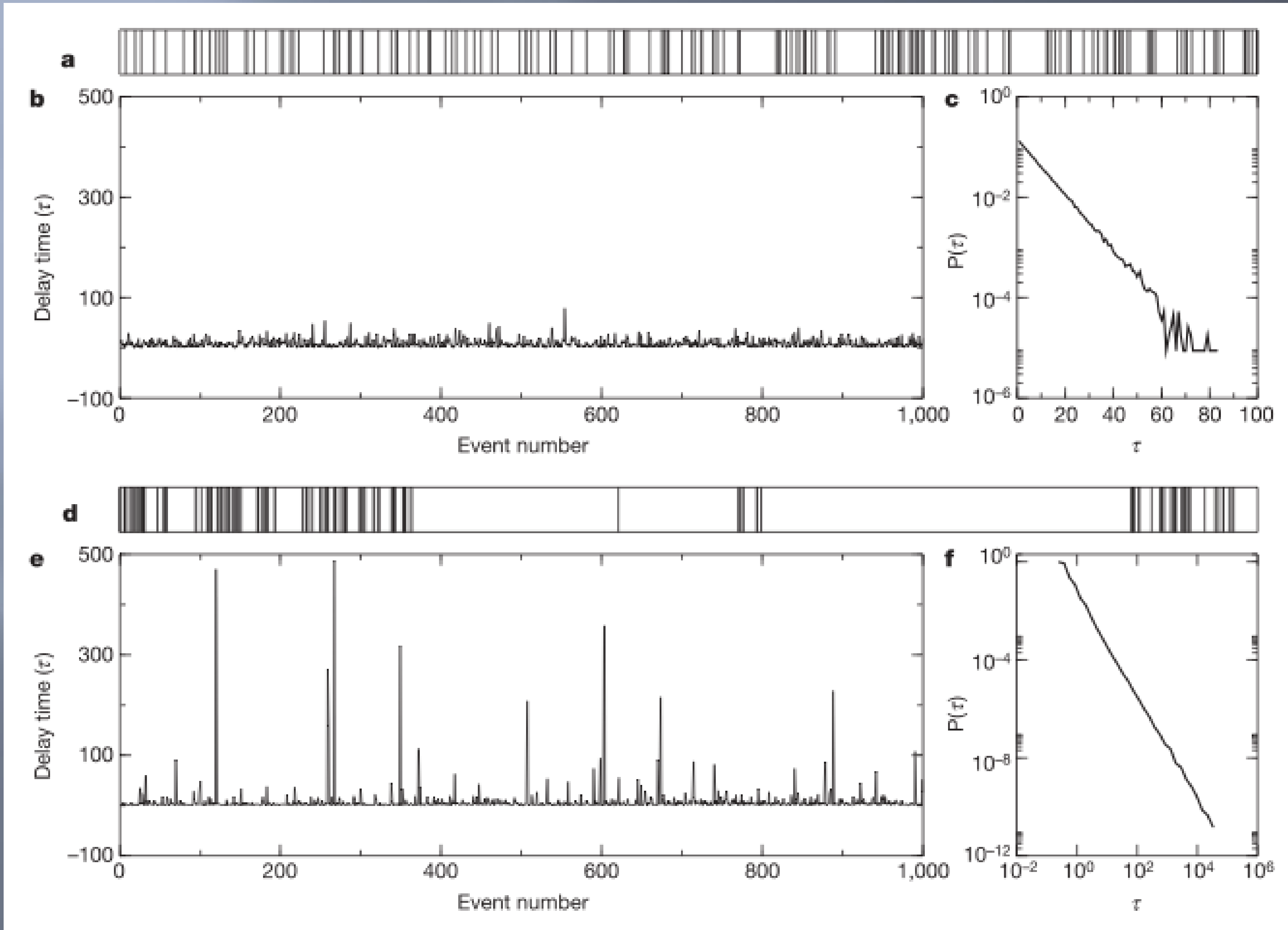
Strong temporal inhomogeneities

Temporal behavior is often non-Poissonian, bursty.
This can be due to seasonalities, to external stimuli and to intrinsic burstiness.



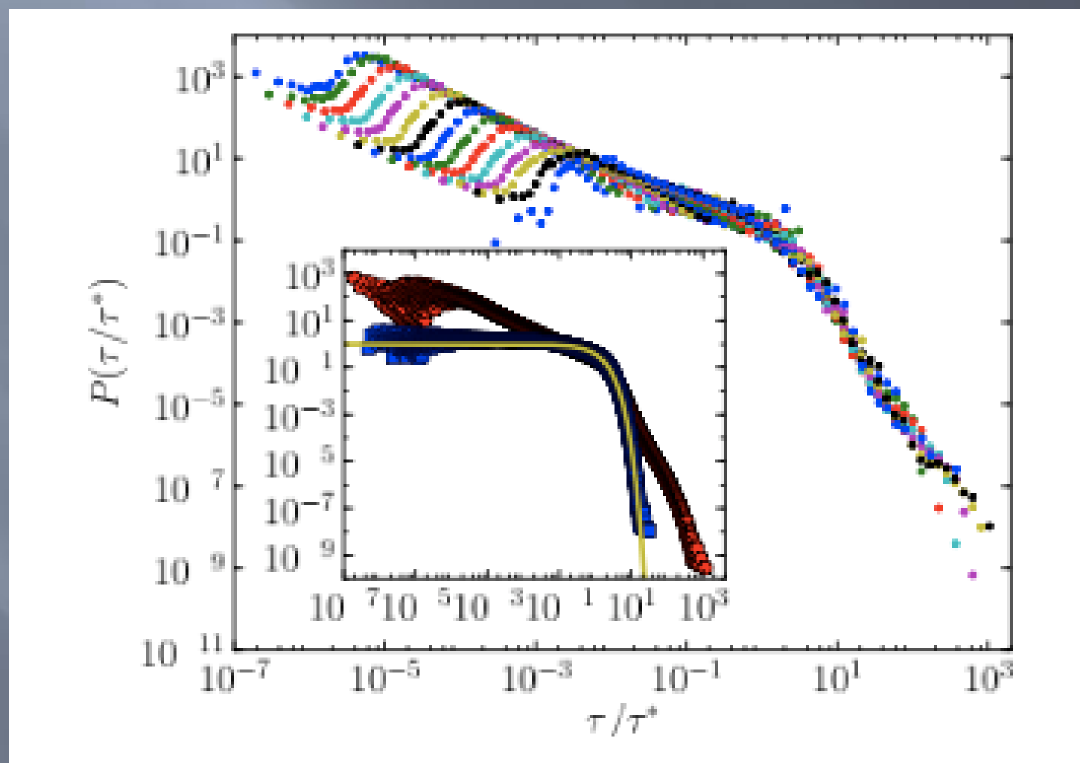
Rocha et al. PNAS (2011)

Burstiness

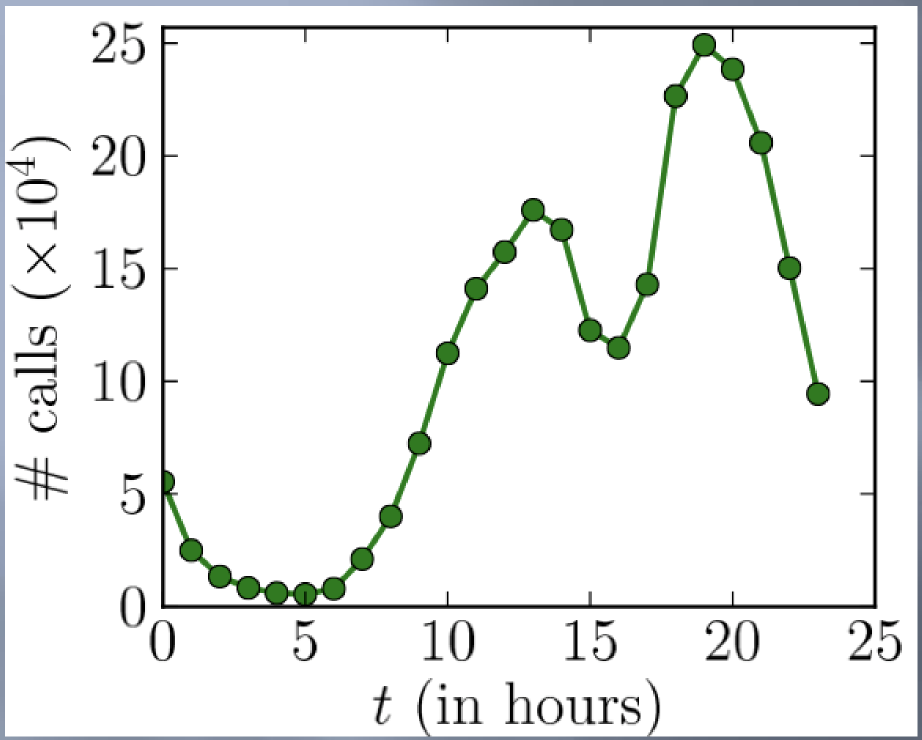


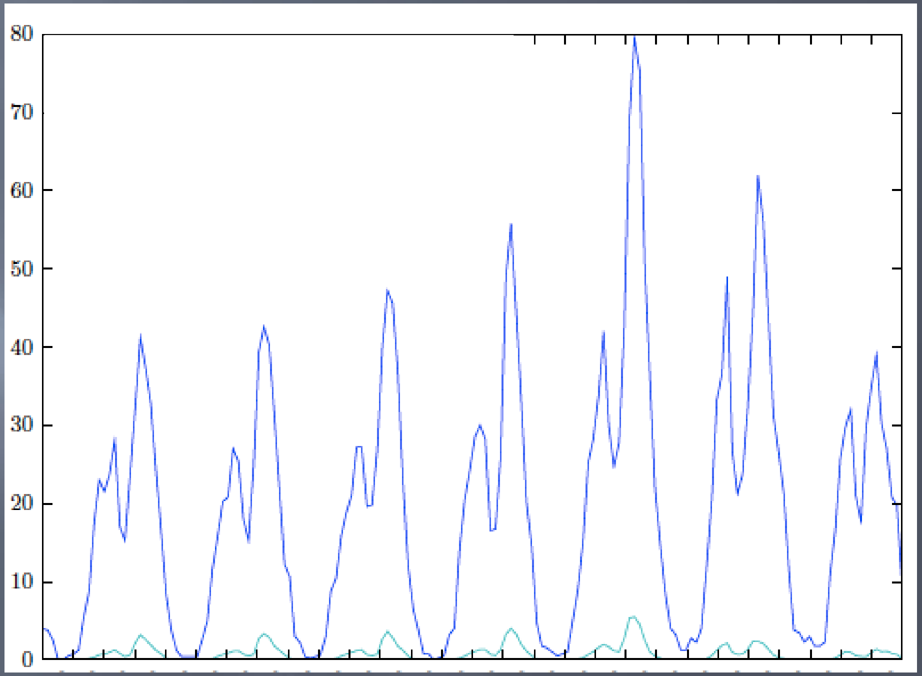
Barabasi Nature 2005

Why is bursty dynamics interesting? Affects spreading, gives insight into the nature of human behavior.



Periodic patterns





days

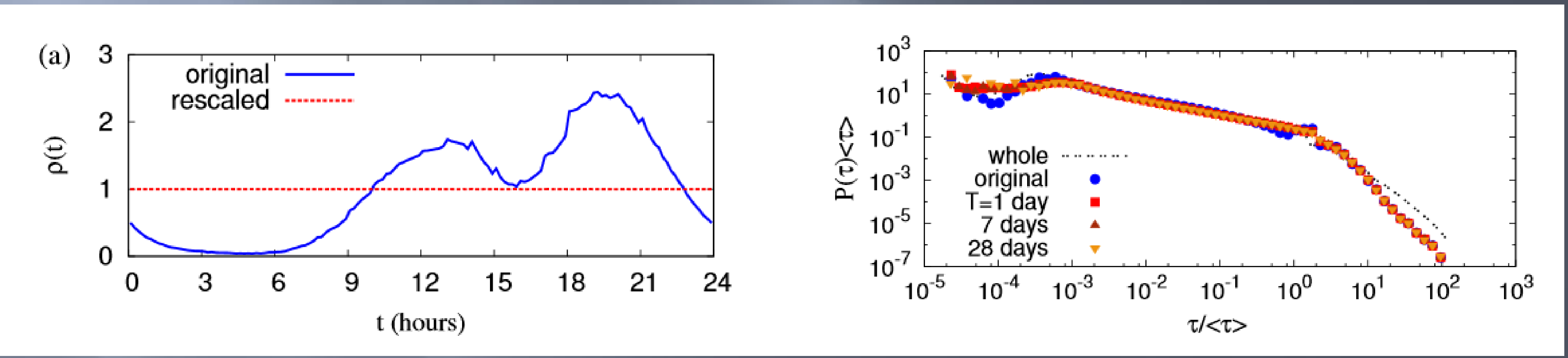
Deseasoning

Deseasoning the data by time rescaling

$$t^*(t) = \sum_{0 \le t' < t} \rho_{\Lambda,T}(t')$$

with

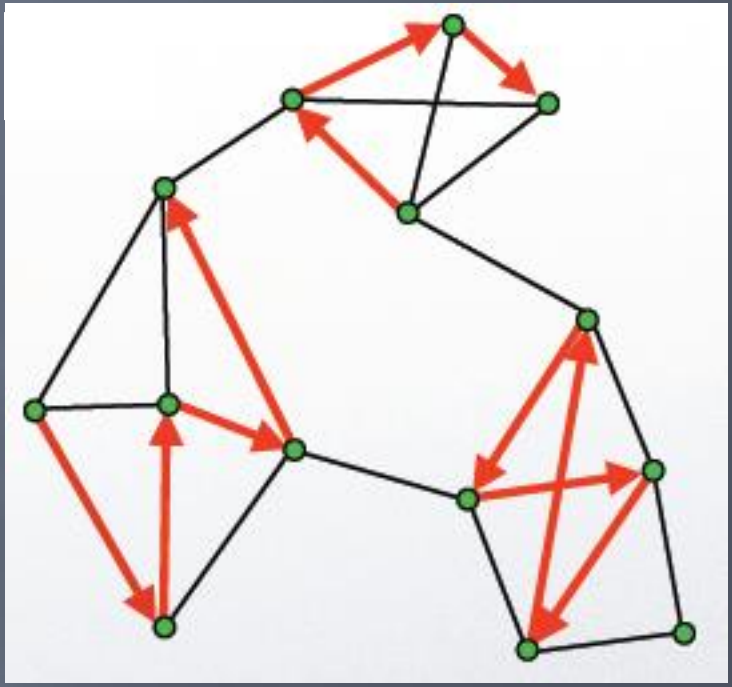
$$\rho_{\Lambda,T}(t) = \frac{T}{s_\Lambda} \sum_{k=0}^{\lfloor T_f/T \rfloor} n_\Lambda(t+kT), \;\; s_\Lambda = \sum_{t=0}^{T_f} n_\Lambda(t)$$



Thus the non-Poissonian character is not due to the circadian pattern.

Triggered events

- Link-link dynamic correlations



Experiment: ”Infect” a random node, the empirical call data and assume that ”infection” is transmitted by each call.

How to identify the effect of the different correlations on spreading?

Introduce different null models by appropriate shuffling of the data.

Problem of null models: E.g. Time shuffling

Link1	Link2	Link3...	LinkN
t_{11}	t_{21}	t_{31}...	t_{N1}
t_{12}	t_{22}	t_{32}...	t_{N2}
.	.	.	.
.	.	t_{3n_3}...	.
t_{1n_1}	.		.
	t_{2n_2}		.
			t_{Nn_N}

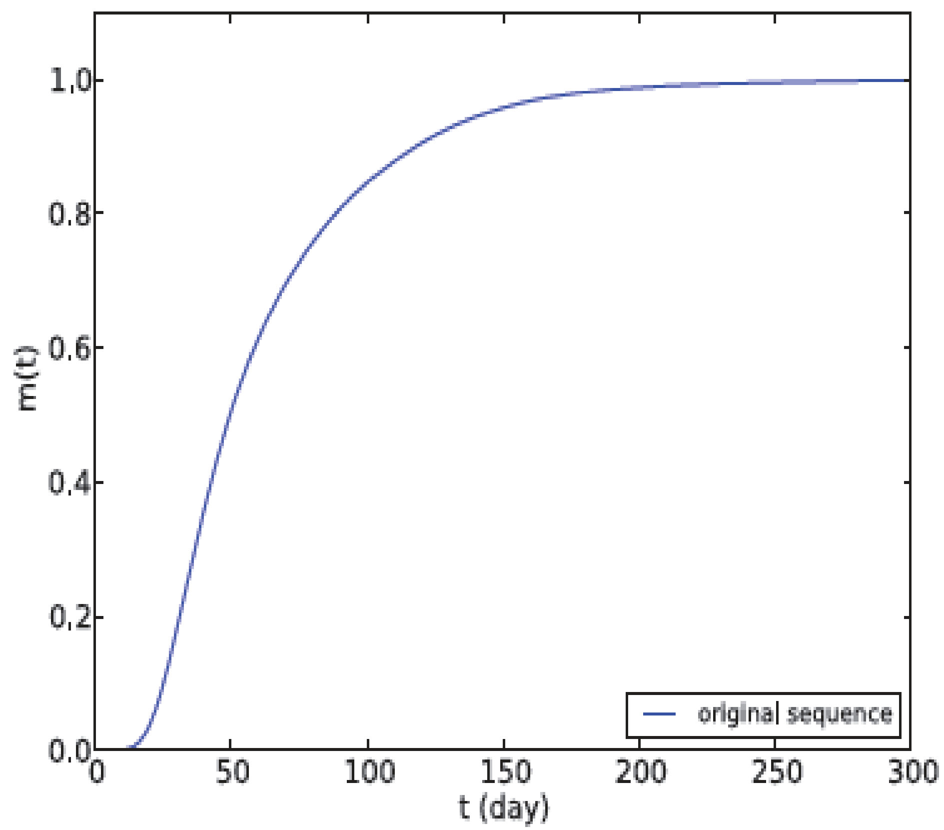
Destroyes burstiness (and link-link correlations) but keeps weight and daily pattern

Original event sequence

- Time ordered sequence of original call events
- It contains all possible correlations which take place in the system

WT: Weight-topology
BD: Bursty dynamics
LL: Link-link (triggered events)
CS: Community structure

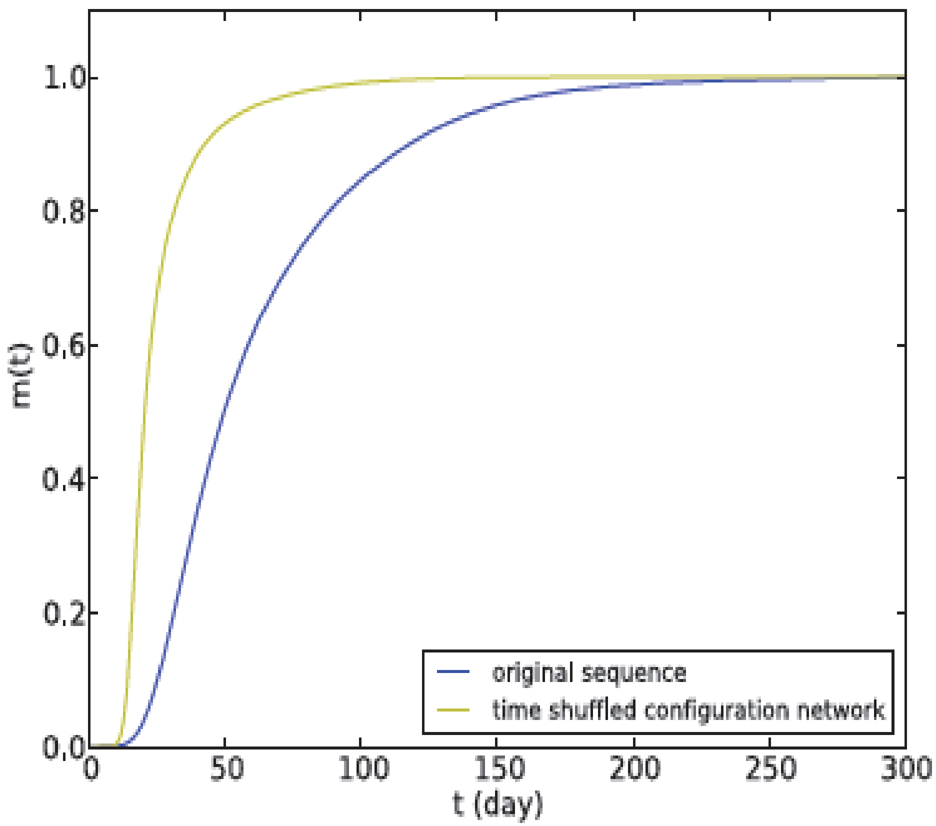
	WT	BD	LL	CS	25%m
Original	✓	✓	✓	✓	33.7



Time shuffled configuration network

- Using configuration model to **destroy community structure**, but keep N, |E| and the network connected
- Shuffle the event times to **destroy bursty dynamics**
- No correlation takes place in the system

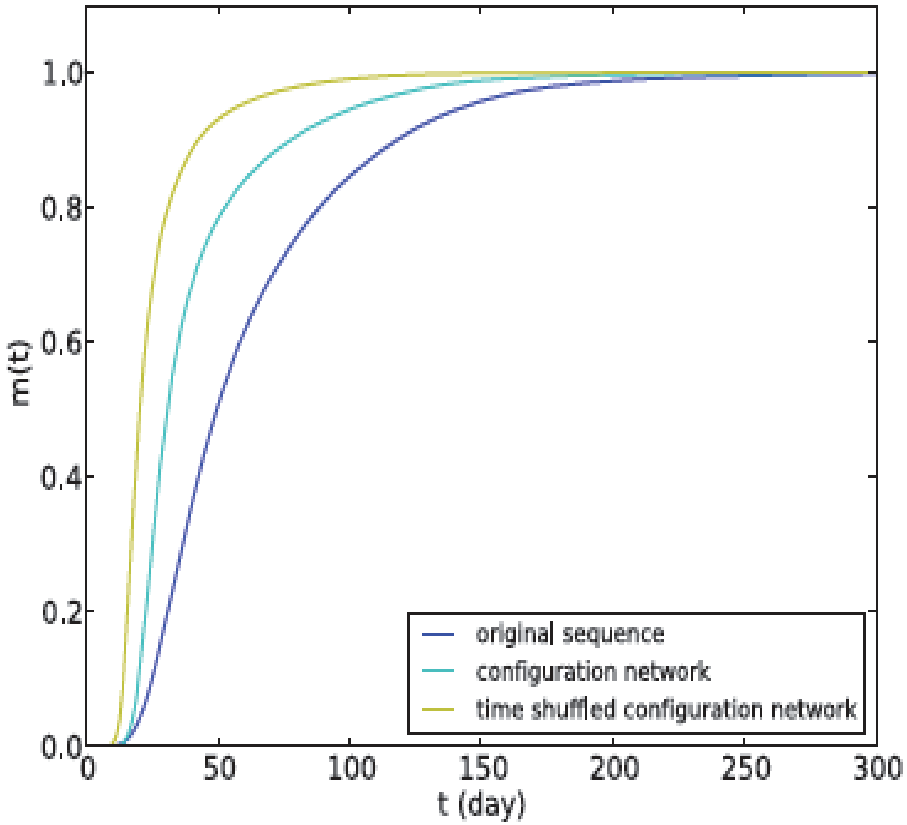
	WT	BD	LL	CS	25%m
Original	✓	✓	✓	✓	33.7
TimeConf	X	X	X	X	16.4



Configuration network

- Using the same configuration method to **destroy community structure**
- Only **bursty dynamical behavior** is kept
- The infection speed is slowed down by bursty dynamics

	WT	BD	LL	CS	25%m
Original	√	√	√	√	33.7
TimeConf	X	X	X	X	16.4
Config.	X	√	X	X	23.8



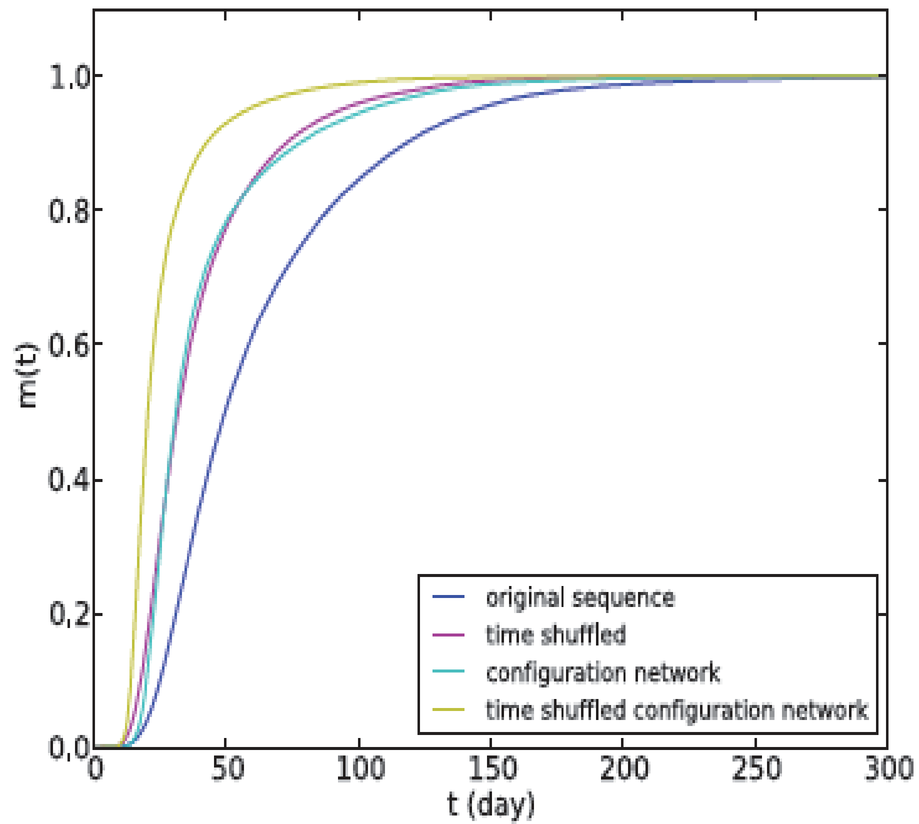
Time shuffled event sequence

- Shuffle the event times but keep **community structure** and **weight-topology** correlations unchanged

- Bursty dynamics and link-link correlations are switched off

- Bursty event clustering is slowing down the dynamics

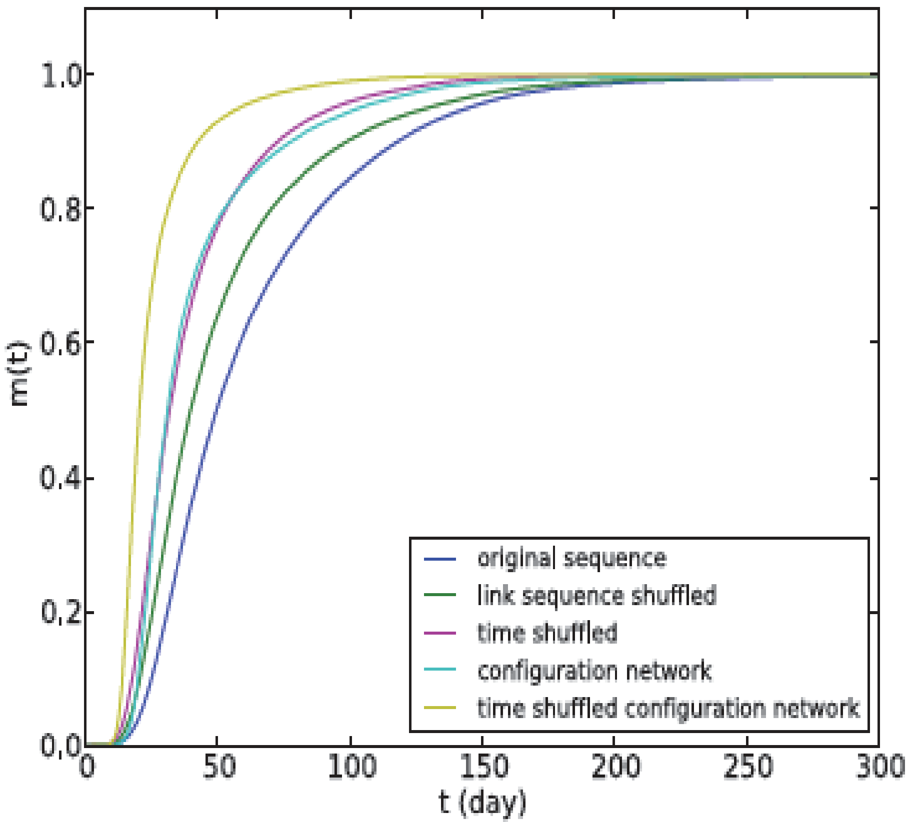
	WT	BD	LL	CS	25%m
Original	✓	✓	✓	✓	33.7
TimeConf	✗	✗	✗	✗	16.4
Config.	✗	✓	✗	✗	23.8
Time	✓	✗	✗	✓	22.9



Link sequence shuffled event sequence

- Shuffle link call sequences between randomly chosen links
- **Link-link** and **weight-topology** correlations are switched off
- Weight-topology correlations also slow down the dynamics

	WT	BD	LL	CS	25%m
Original	✓	✓	✓	✓	33.7
TimeConf	X	X	X	X	16.4
Config.	X	✓	X	X	23.8
Time	✓	X	X	✓	22.9
Link	X	✓	X	✓	27.5



Link sequence shuffling

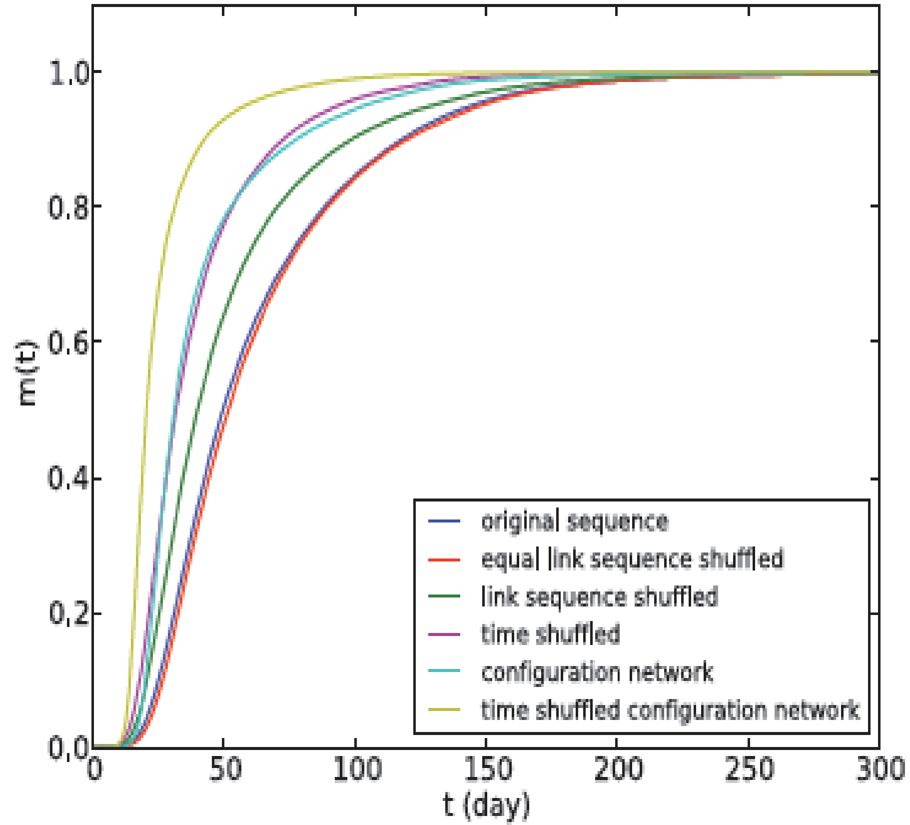
Link1	Link2	Link3...	LinkN
t_{11}	t_{21}	t_{31}...	t_{N1}
t_{12}	t_{22}	t_{32}...	t_{N2}
.	.	.	.
.	.	t_{3n_3}...	.
t_{1n_1}	.		.
	t_{2n_2}		.
			t_{Nn_N}

Select random pairs of links sequences and exchange
Destroys topology-weight and link-link correlation, keeps burstiness

Equal link sequence shuffled event sequence

- Shuffle call sequences between links having the same weight
- Only **link-link correlations** are destroyed
- Multilink correlations accelerate the spreading process

	WT	BD	LL	CS	25%m
Original	✓	✓	✓	✓	33.7
TimeConf	X	X	X	X	16.4
Config.	X	✓	X	X	23.8
Time	✓	X	X	✓	22.9
Link	X	✓	X	✓	27.5
Equal link	✓	✓	X	✓	35.3

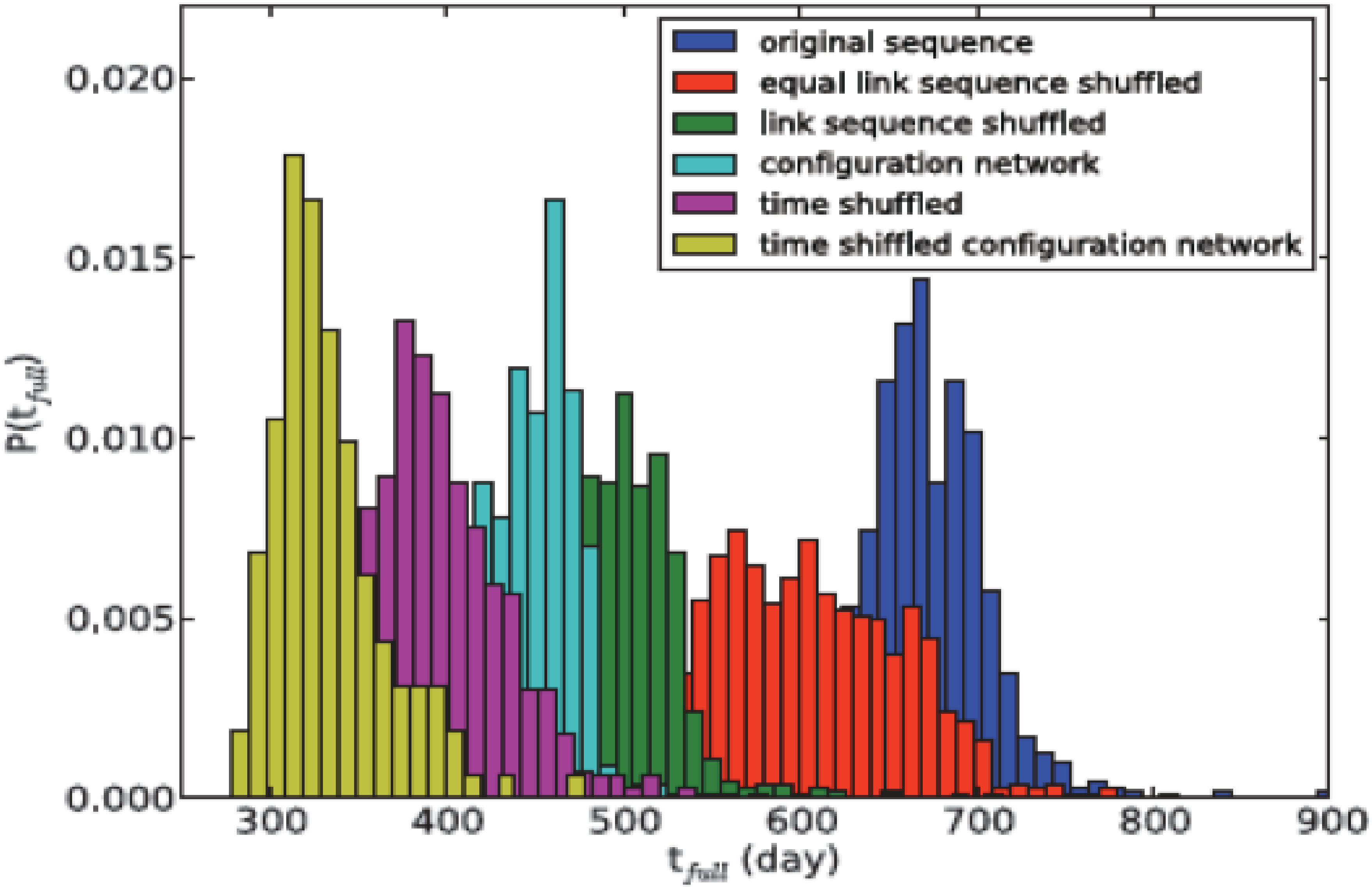


Equal weight link sequence shuffling

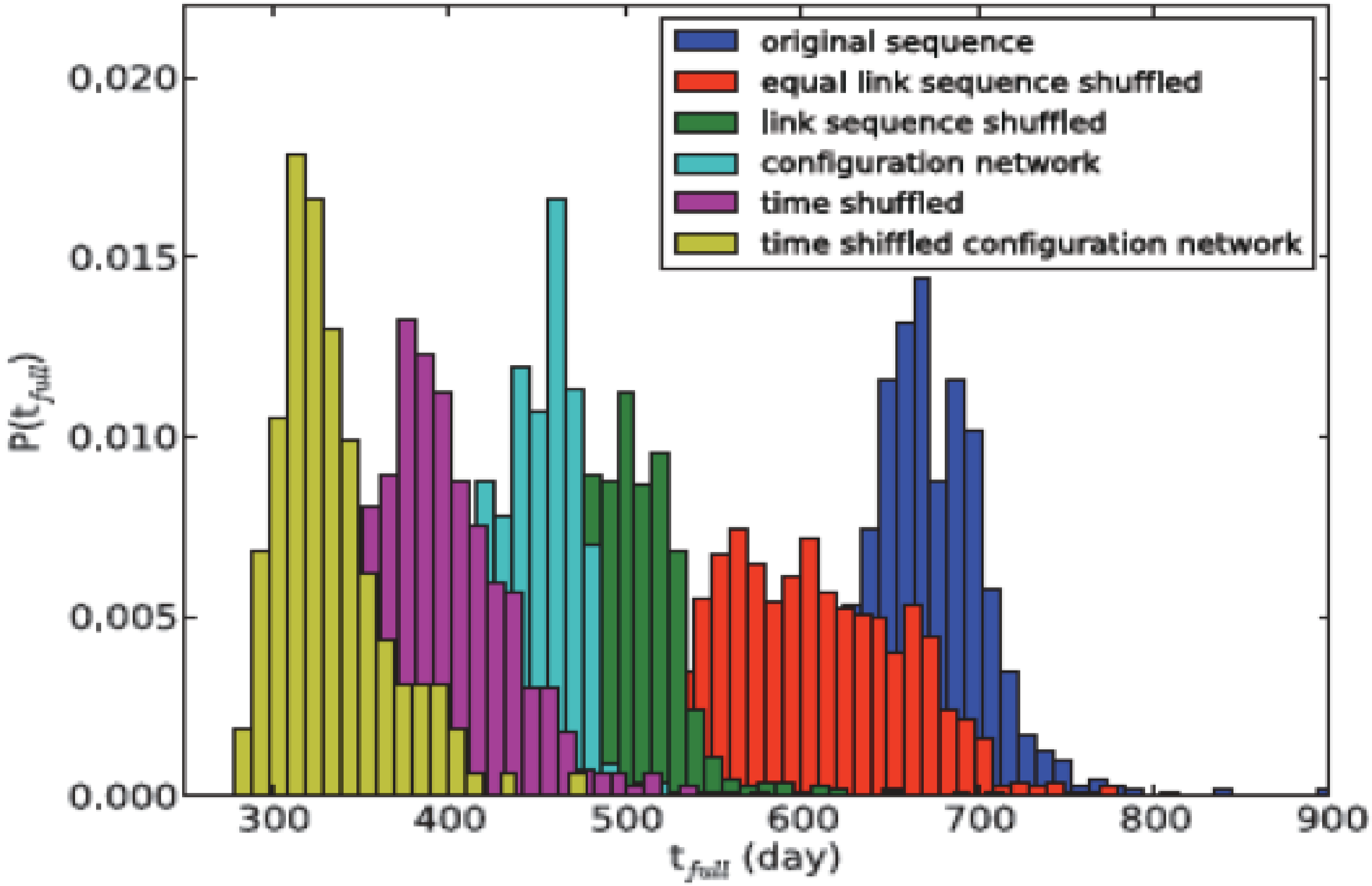
Link1$_{w1}$	Link2$_{w2}$	Link3$_{w2}$	LinkN$_{w1}$
$t_{1'1}$	$t_{2'1}$	$t_{3'1}$…	$t_{N'1}$
$t_{1'2}$	$t_{2'2}$	$t_{3'2}$…	$t_{N'2}$
.	.	.	.
.	.	.	.
$t_{1'n\_1'}$	$t_{2'n\_2'}$	$t_{3'n\_3'}$…	$t_{N'n\_N'}$

Destroys link-link correlations but keeps weight-topology correlations and bursty dynamics

Long time behavior



Long time behavior (total infection)



Complex contagion

SI,SIR, etc models are not good for social contagion of information, rumors, innovations etc.

There the transmission is not a 2-body interaction.

Sometime spreading within the society can be extremely fast (e.g., rumor about the accident in a nuclear power plant in Hungary in 2004

Threshold model

M. Granovetter (Am. J. Sociology 1978) Threshold models

D. Watts (PNAS 2002) Mathematical form

Random network with degree distribution p_k and average degree $\langle k \rangle = z$. Every node has a threshold ϕ indicating the **critical ratio** of adopting neighbors needed to make the node adopt. Initiate the process by infecting a node.

There are **vulnerable** nodes, which get infected if they have one adopting neighbor: $\phi \leq 1/k$.

The others are **stable**.

The phase diagram can be calculated.

Generating function method

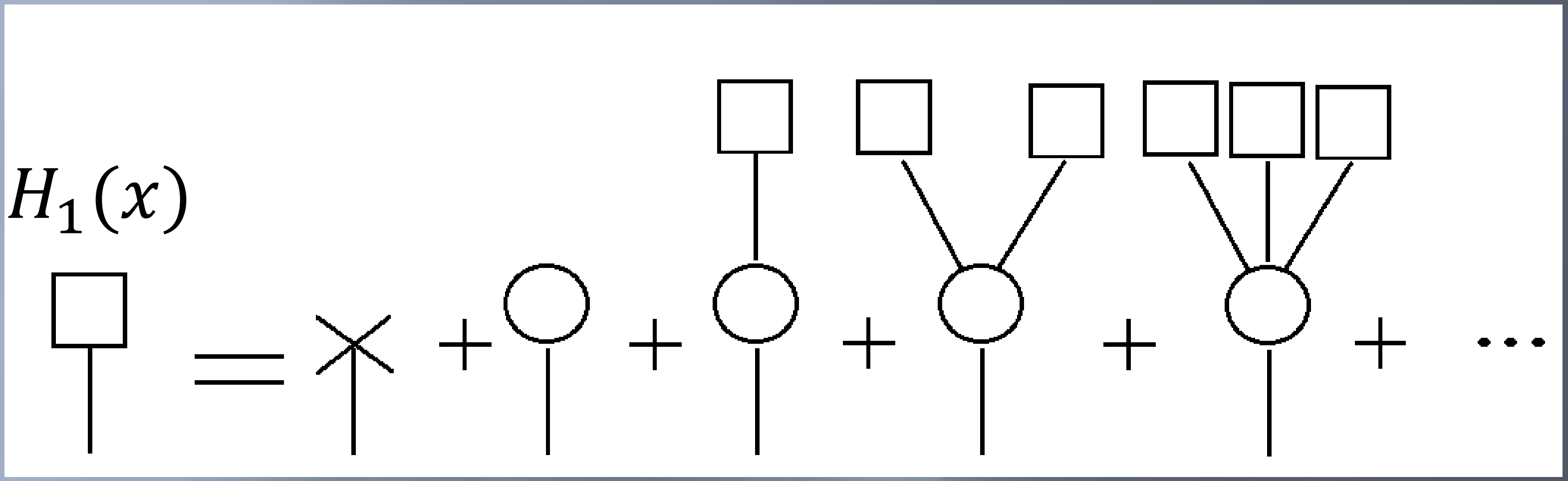
p_k Prob that a node has degree *k*
ρ_k Prob that a node of degree *k* is vulnerable $(1/k > \phi)$
q_n Prob that a node belongs to vulnerable cluster of size *n*
w_n Prob that a node's neighbor –" – –" – –" – of size *n*

$G_0(x) = \sum_k p_k \rho_k \, x^k$ gen. fn.: a k-node →vuln.
$G_1(x) = \sum_k r_k \, x^{k-1} = \sum_k \frac{k p_k \rho_k}{z} \, x^{k-1}$ gen. fn.: a node's neighbor → vuln with $k - 1$ outgoing degrees: $G_1(x) = G_0'(x)/z$

$H_0 = \sum_n q_n \, x^n$ gen. fn.: node belongs to vuln. cluster
$H_1 = \sum_n w_n \, x^n$ gen. fn.: node's neighbor –" – –" – –" –

Sparse, random, uncorrelated networks are tree like

Using tree-like property:



$H_1(x) = \left(1 - G_1(1)\right) + xr_0 + xr_1H_1(x) + xr_2H_1^2(x) + \ldots$

$H_1(x) = 1 - G_1(1) + xG_1(H_1(x))$ and similarly

$H_0(x) = 1 - G_0(1) + xG_0(H_1(x))$

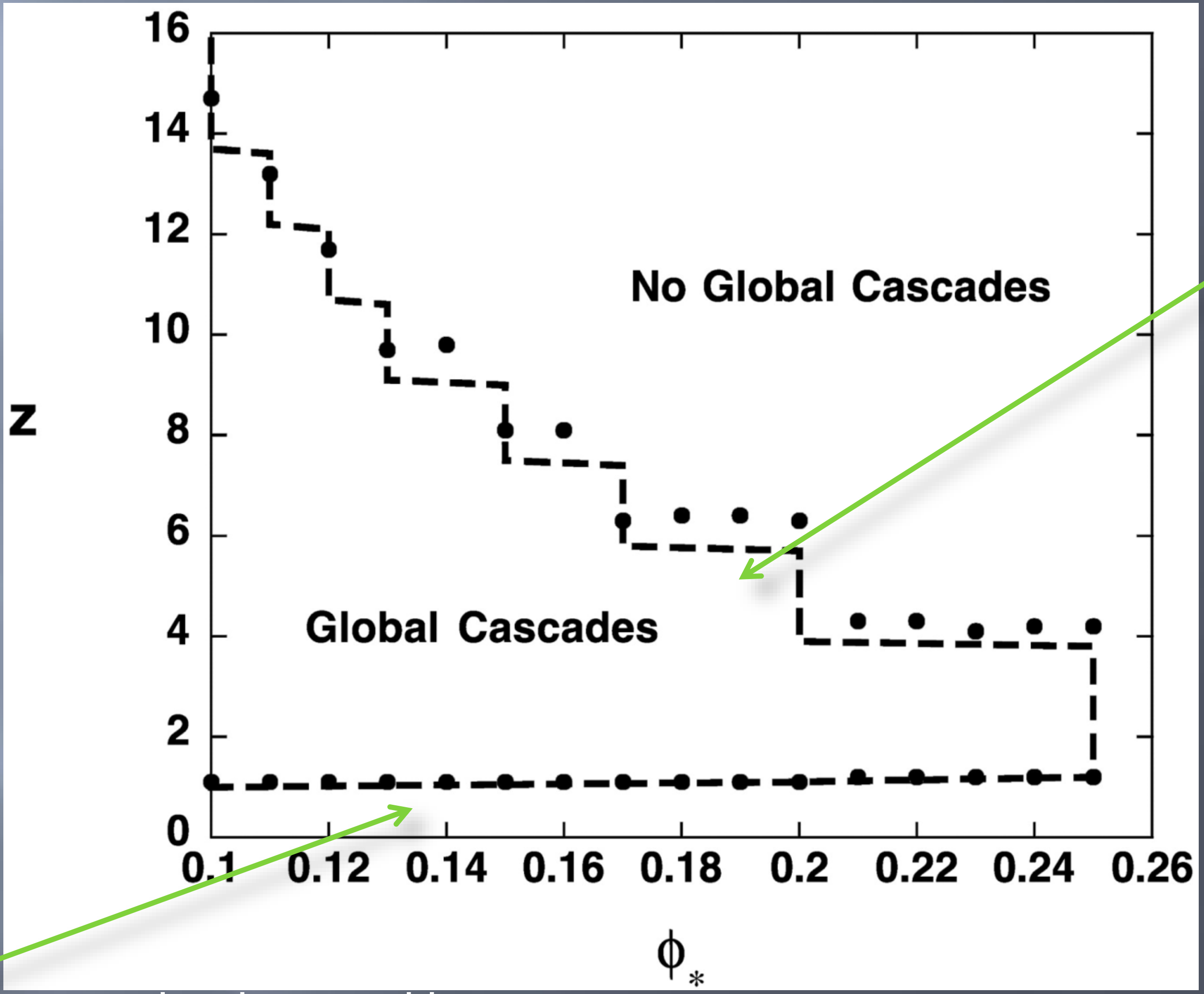
Using $H_1(1) = 1$

$\langle n \rangle = H_0'(1) = G_0(1) + \frac{(G_0(1))^2}{z - G_0''(1)}$ from which the criterion

$G_0''(1) = \sum_k k(k-1)p_k\rho_k = z$ for the transition

Cascade windows for the threshold model.

(ER graph)



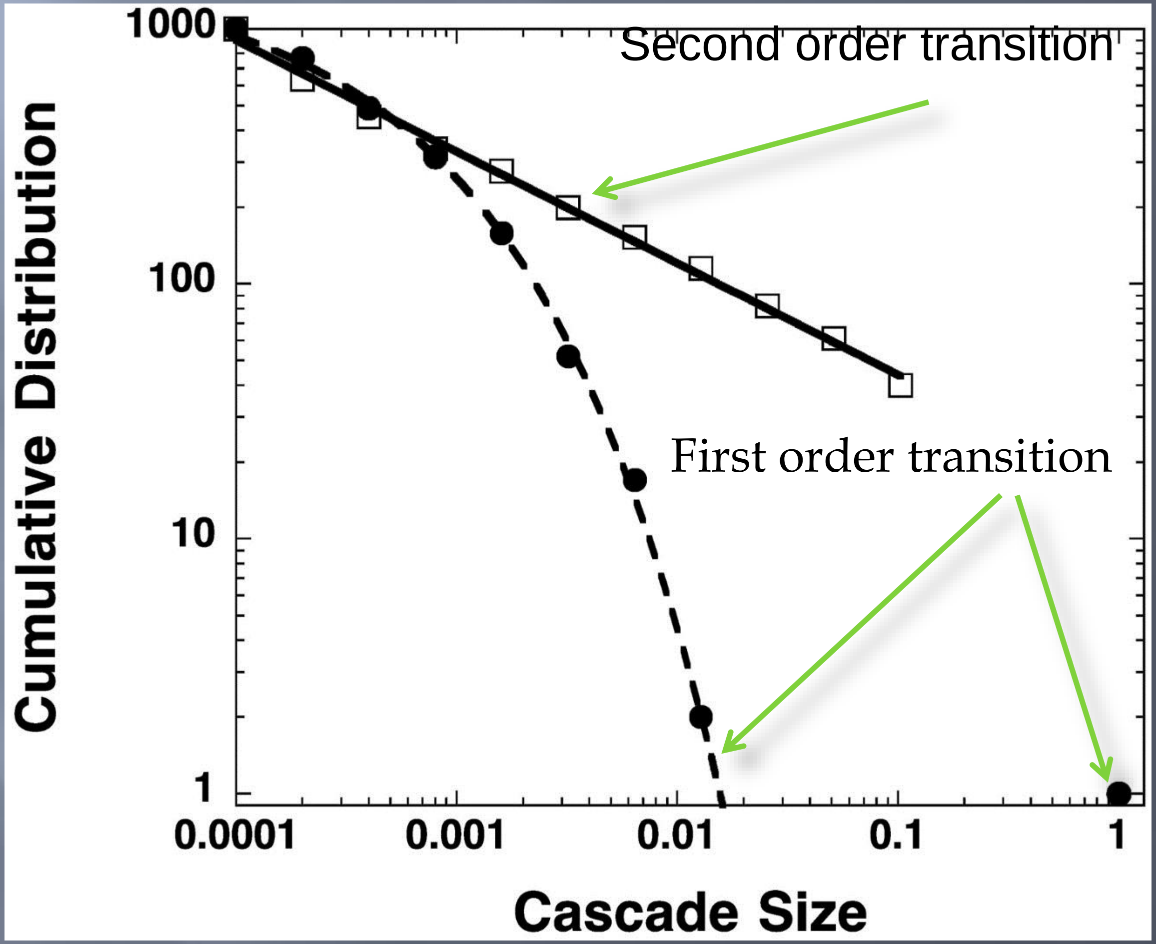
First order tr.

Fragmentation, second order transition

Watts D J PNAS 2002;99:5766-5771

PNAS

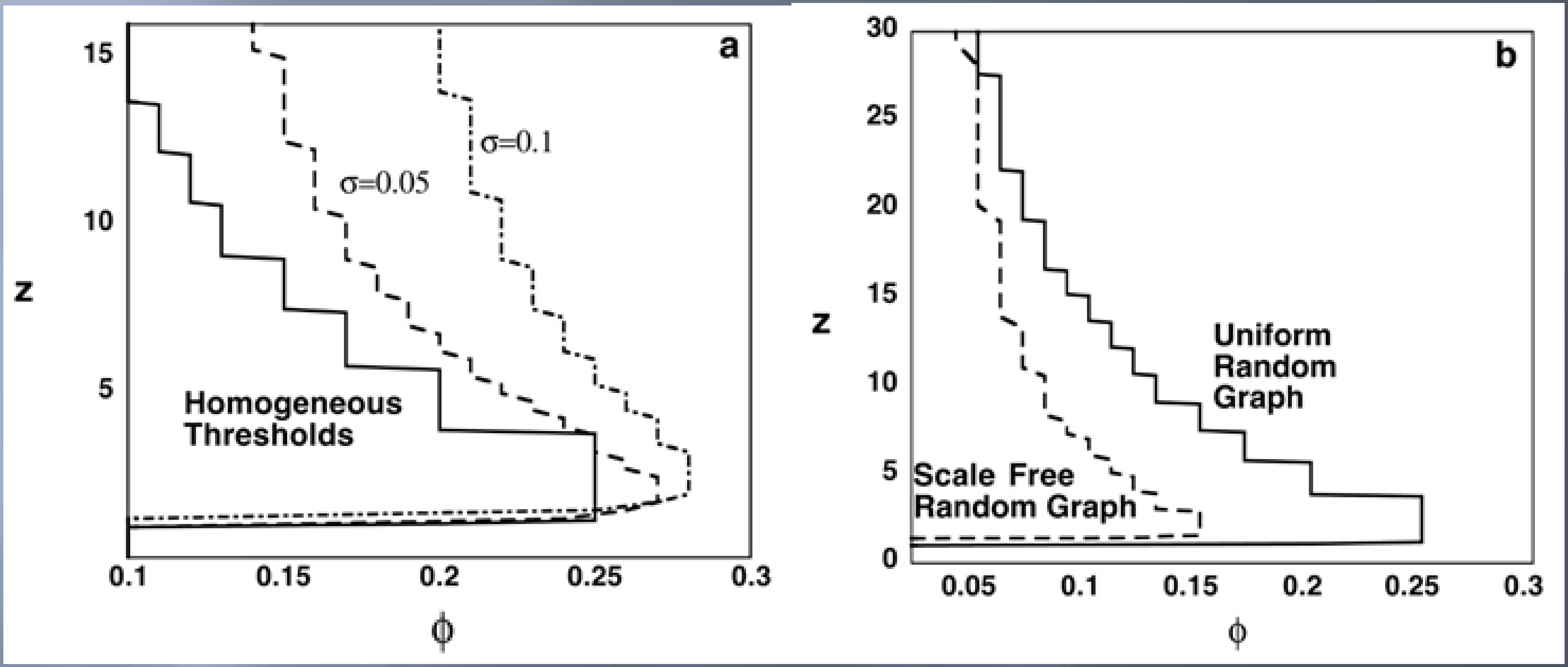
Cumulative distributions of cascade sizes at the lower and upper critical points, for n = 1,000 and z = 1.05 (open squares) and z = 6.14 (solid circles), respectively.



Watts D J PNAS 2002;99:5766-5771

PNAS

Cascade windows for heterogeneous networks.



The two transitions differ qualitatively: the lower one is due to a connectivity transition (similarly to the usual percolation fragmentation) while the upper one is the consequence of too high degree – the threshold criterion cannot be fulfilled.

Watts D J PNAS 2002;99:5766-5771

Take home messages

- Functioning complex systems can be modelled by dynamic processes on networks, spreading being one of the most important one.
- Spreading can be described at different levels: Perfect mixing, degree based mean field, etc.
- Hubs boost spreading → 0 epidemic threshold for SF
- Vaccination should focus on hubs and they can be found by local algorithms
- Studying empirical data about spreading on temporal networks by the method of null models reveals that the main decelerating factors are Granovetterian structure and bursty communication patterns.
- Threshold model describes transition between global and local spreading. Global spreading can be very fast.

Last homework:

Create a Barabasi Albert network with $N = 10^5$.
Simulate the SIS model.
We have discrete times. An infected node infects its neighbors with probability β and gets immune in the next step with probability 1.
The order parameter of the problem is the asymptotic (average) value of infected nodes.
Calculate the order parameter, when the initial infection is
- the at the largest hub
- the last attached node

Technically: You generate the network and repeat the experiment with different random number sequences starting from the same node.
Compare the two cases!