Identifying Influential Spreaders in Complex Networks

Maksim Kitsak, Lazaros K. Gallos, Shlomo Havlin,
Fredrik Liljeros, Lev Muchnik, H. Eugene Stanley, and Hernan A. Makse

Nature Physics *6* 888 (2010).

Q1: **Definition of spreading efficiency?**

Q2: **What determines spreading efficiency?**

Q3: **Who/what are the most efficient spreaders?**

BRIDGING PSYCHOPHYSICS AND NEUROPHYSIOLOGY,

Denton, TX, March 19, 2011

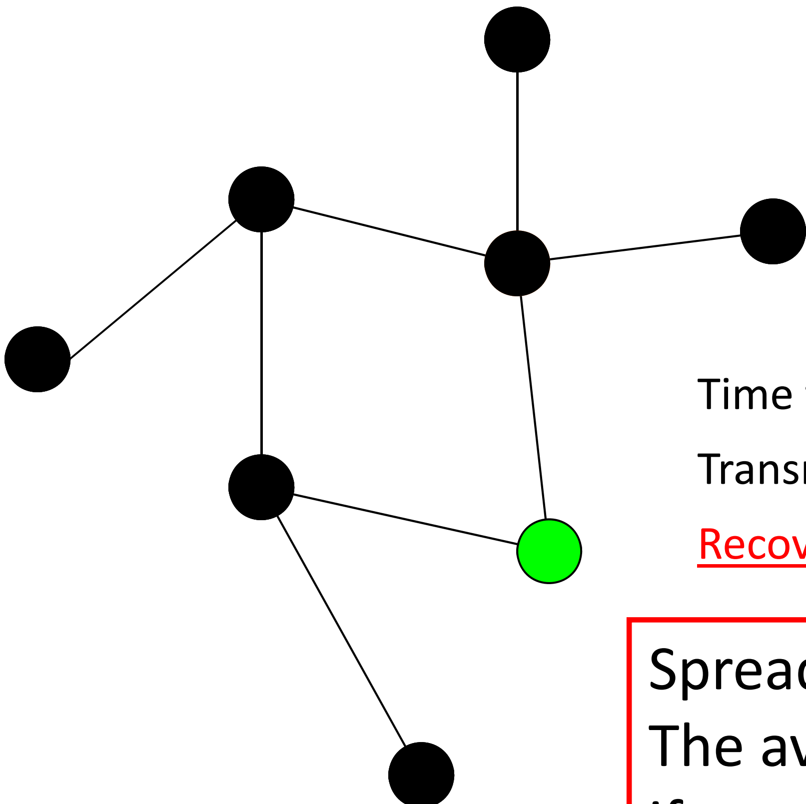
caida

UC San Diego
Local Impact, National Influence, Global Reach

BOSTON UNIVERSITY

Spreading Processes: Examples and Models

- **Examples:**
- **Infectious Diseases *(smallpox, influenza…)***
- **Rumor, Ideas**
- ***Email, bluetooth viruses***



The SIR Model

- "Susceptible" (unaffected) individual.
- "Infected" (affected) individual.
- "Recovered" individual.

Time to "recover" $T_R = 2$

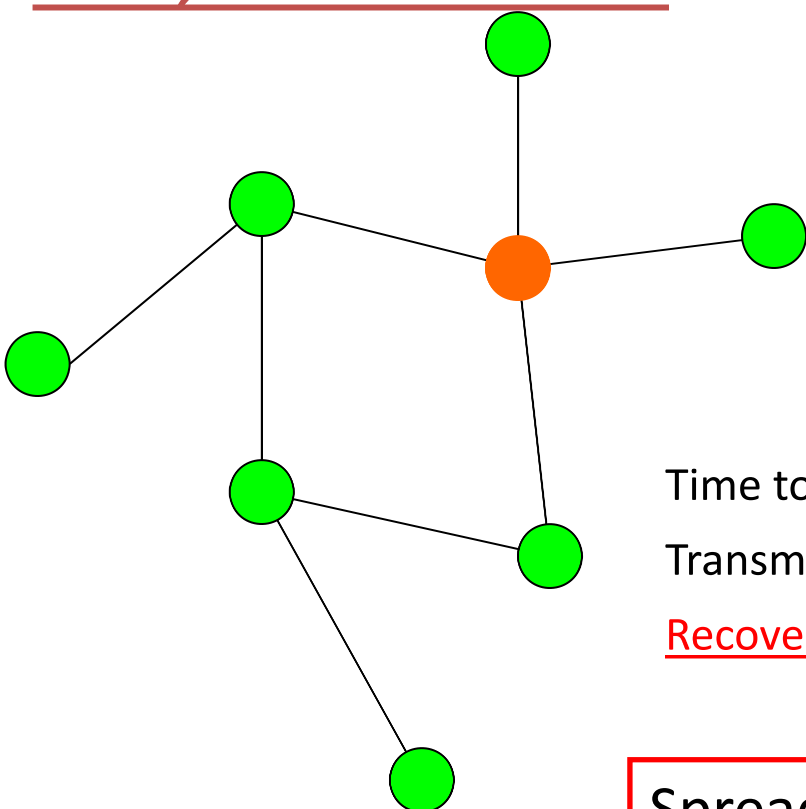
Transmission probability $\beta = 0.5$

Recovered individuals can not be infected!!!

Spreading efficiency: ***$\langle M_i \rangle$***
The average number of infected nodes
if spreading starts at node ***i***

Spreading Processes: Examples and Models

- **Examples:**
- **Infectious Diseases *(smallpox, influenza…)***
- **Rumor, Ideas**
- ***Email, bluetooth viruses***



The SIS Model

● "Susceptible" (unaffected) individual.

● "Infected" (affected) individual.

● ~~"Recovered" individual.~~

Time to "recover" $T_R = 2$

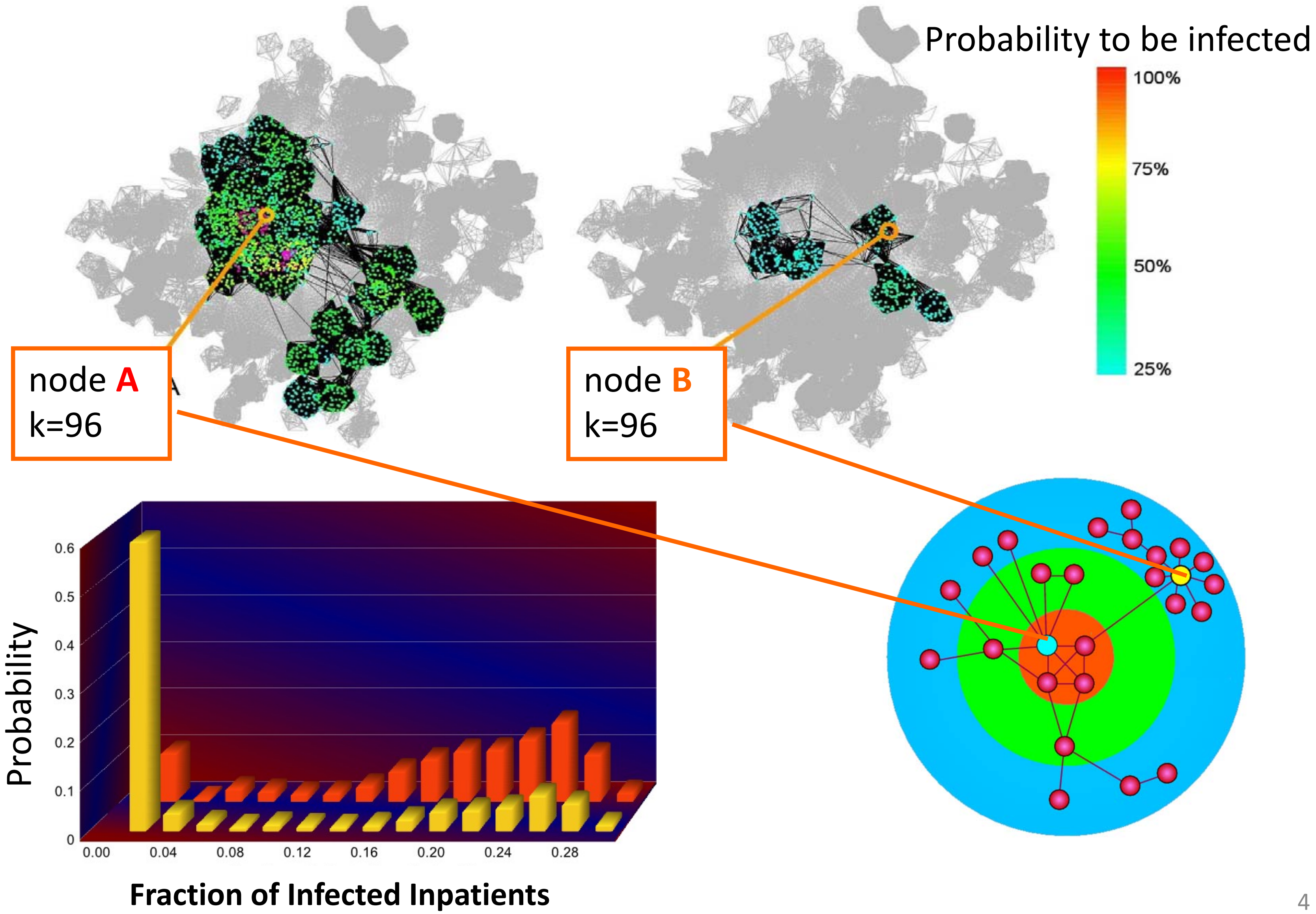
Transmission probability $\beta = 0.5$

Recovered individuals can be infected again!!!

Spreading efficiency (persistency): ***$\rho_i(t)$***
Probability node ***i*** is infected at time ***t***

Spreading efficiently determined by node placement!

Hospital Network: Inpatients in the same quarters connected with links



k-cores and *k*-shells determine node placement

***K*-core: sub-graph with nodes of degree at least k inside the sub-graph.**

Pruning Rule:

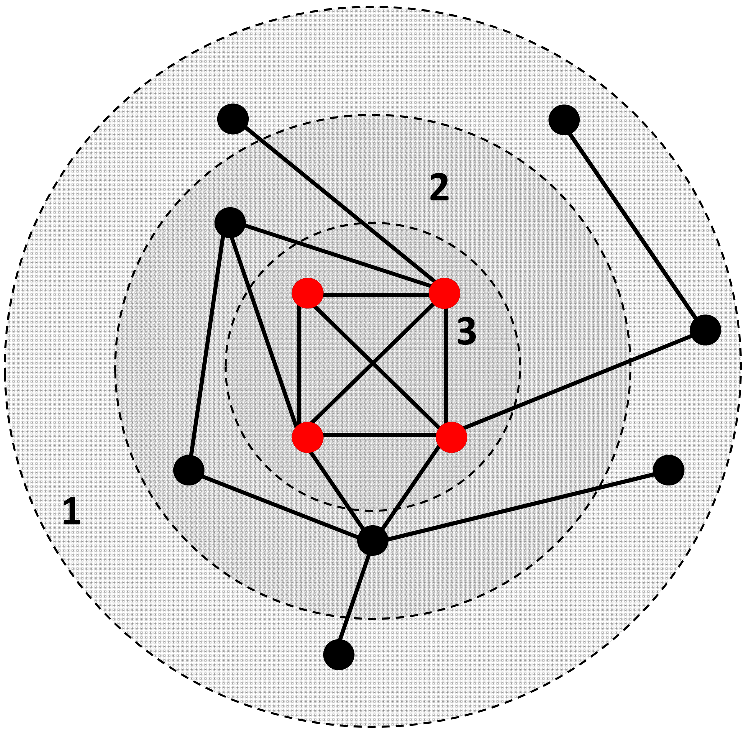
1) Remove all nodes with k=1.

Some remaining nodes may now have k = 1.

2) Repeat until there is no nodes with k = 1.

3) The remaining network forms the 2-core.

4) Repeat the process for higher k to extract other cores



S. B. Seidman, Social Networks, **5**, 269 (1983).

K-shell is a set of nodes that belongs to the *K*-core

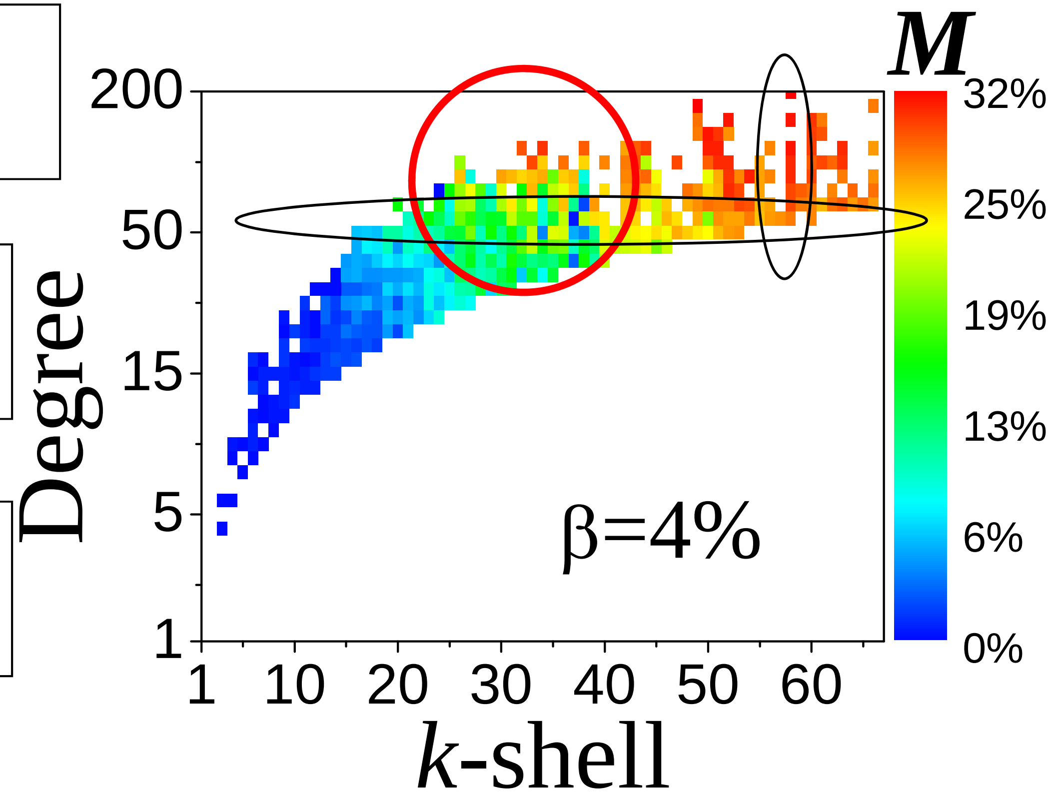
but NOT to the *K+1*-core

(1) For every individual ***i*** measure the average fraction of individuals M_i he or she would infect (spreading efficiency).
(2) Group individuals based on the number of connections and the k-shell value.

A. Most efficient spreaders occupy high *k*-shells.

B. For fixed *k*-shell **<*M*>** is independent of ***k*.**

C. A lot of hubs are inefficient spreaders.



Three candidates:
Degree, *k*-shell, betweenness centrality

Imprecision functions test the merits of degree, k-shell and centrality

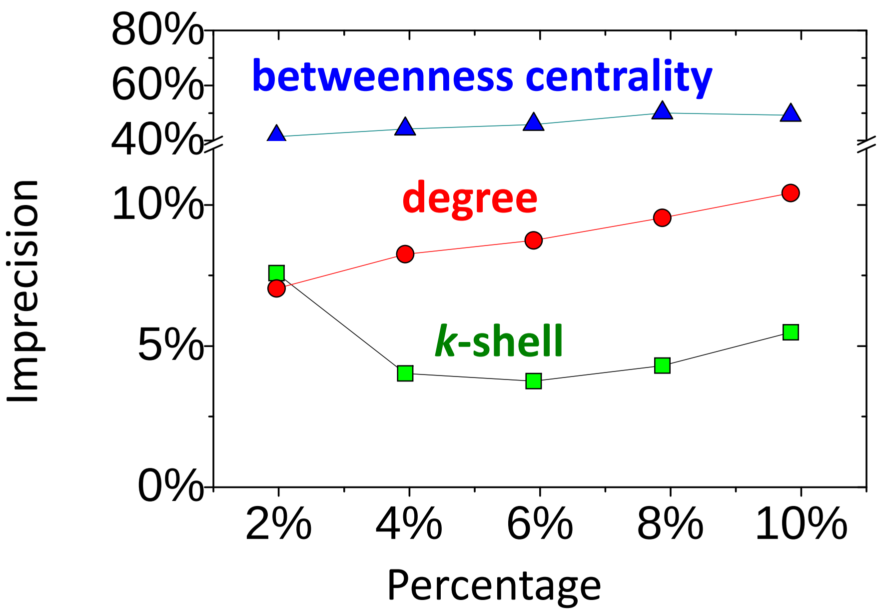
For given percentage ***p***

- Find ***p%*** most efficient spreaders (as measured by ***M***)
- Calculate the average infected mass M_{EFF}.
- Find ***p%*** nodes with highest ***k*-shell** indices.
- Calculate the average infected mass M_{kshell}.

Imprecision function:

$$\varepsilon(p) = 1 - \frac{M_{kshell}(p)}{M_{EFF}(p)}$$

Measure the imprecision for
K-shell, degree and centrality.

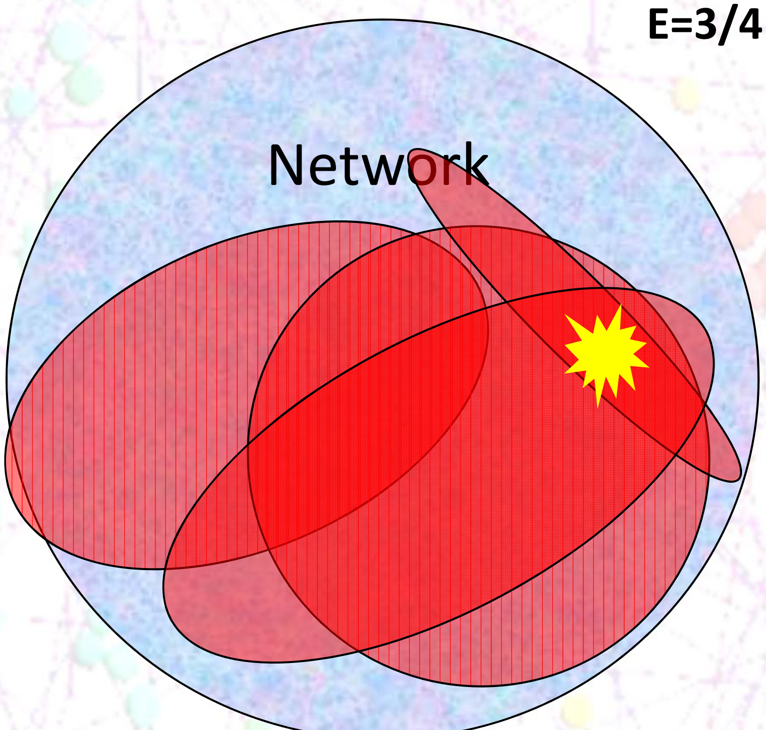


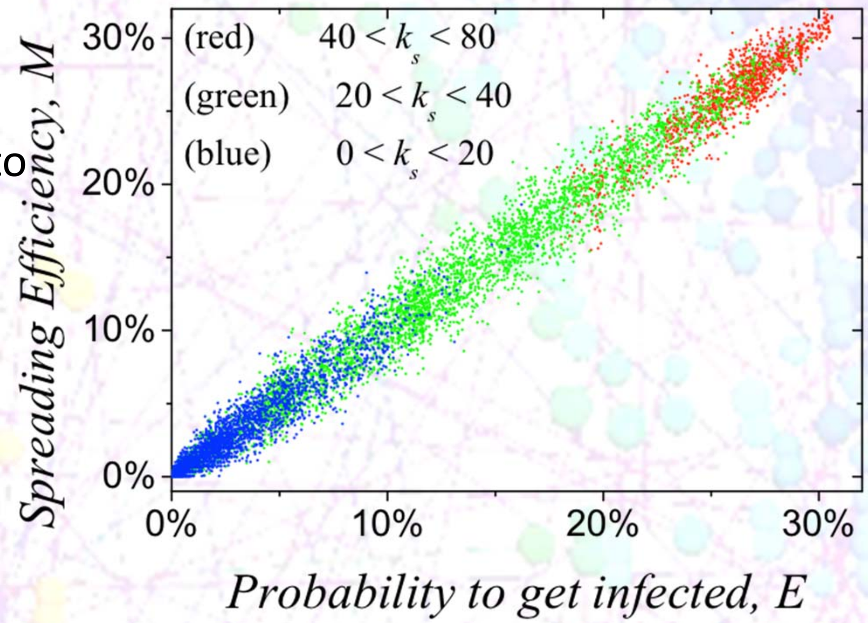
k-shell is the most robust spreading efficiency indicatior.
(followed by degree and betweenness centrality)

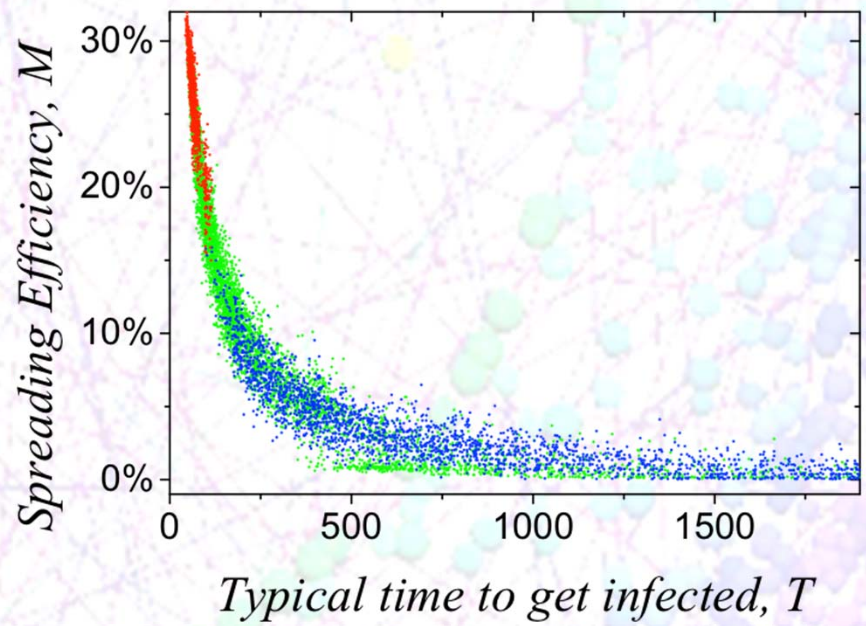
How often do we get sick? How quickly do we get sick?

- An epidemic starts at a randomly chosen node.
- Probability that given node ***i*** gets infected?, ***E***?
- How many time steps does it take for the virus to reach node ***i*** ?, ***T***

E=3/4







Most efficient spreaders:

a) are more likely to be infected.

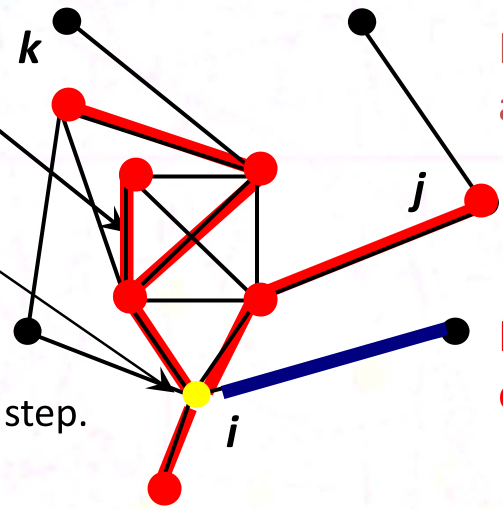
b) get infected at early stages.

Why are M, E and $\langle T^{-1} \rangle$ related?

Epidemic path

Epidemic origin

Neighbors are infected with probability β per time step.



Probability that virus **will not** propagate along given edge in **1** day: $1-\beta$

…in **2** days: $(1-\beta)(1-\beta)$

…in $\boldsymbol{T_R}$ days: $(1-\beta)^{T_R}$

Probability virus **will** propagate along given edge in ***T*** days:

$$p = 1-(1-\beta)^{T_R}$$

SIR spreading can be mapped onto the edge percolation problem!

M.E.J. Newman Phys. Rev. E **66** 016128 (2002).

Pre-select (mark) edges at random with probability: $p = 1-(1-\beta)^{T_R}$

The average infected mass $\mathbf{M_i}$ ⟷ $\mathbf{S_i}$ The average cluster size reached from node ***i*** along marked edges.

Epidemics centrality $\mathbf{E_i}$ ⟷ $\mathbf{S_i/N}$ The fraction of the average cluster size reached from node I along marked edges

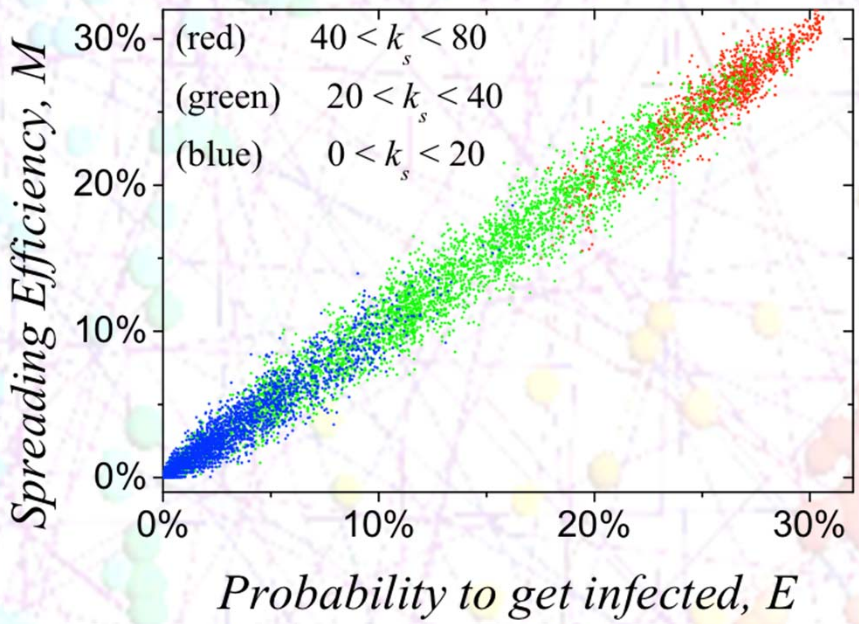
Virus originating at node ***i*** infects node ***j*** in l_{ij} time steps.

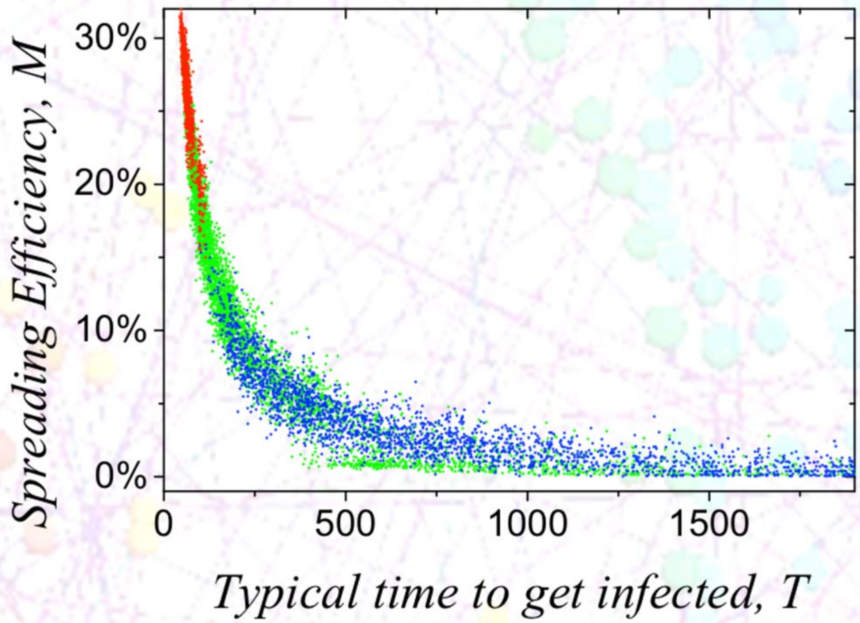
l_{ij} is the shortest path between nodes ***i*** and ***j*** along marked edges.

Average Inverse time $\langle \mathbf{T^{-1}} \rangle_i$ ⟷ $\dfrac{1}{N}\sum_j 1/l_{ij} \sim M/\ln M$

Why are M, E and <T^{-1}> correlated?

- It is relatively easy to measure how often each individual gets sick.





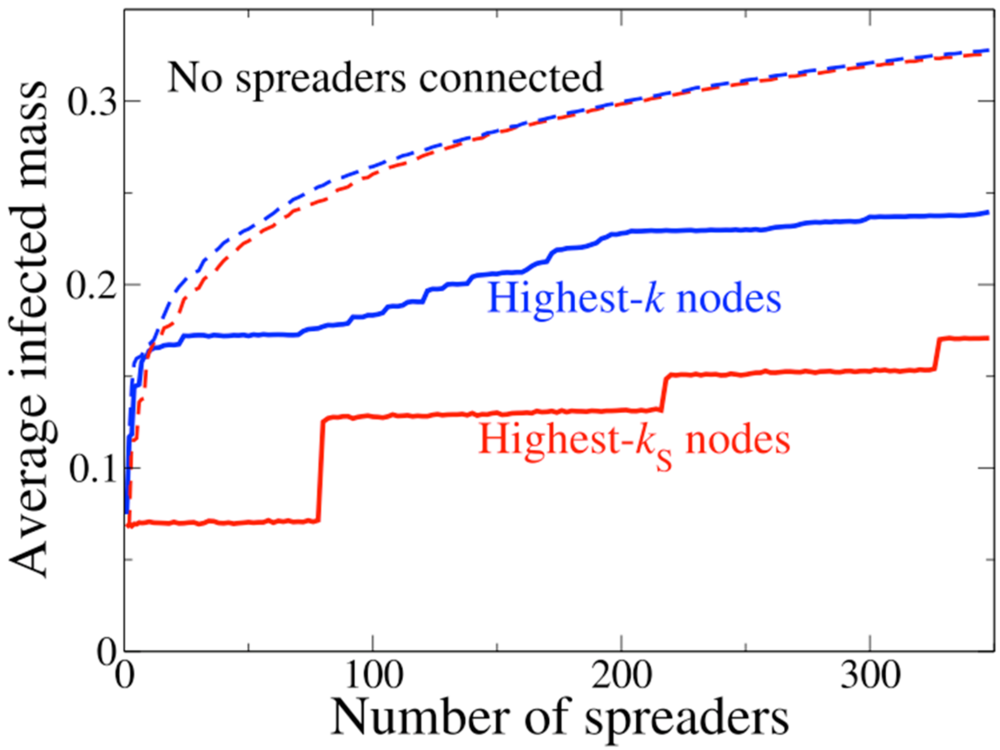
Are the most efficient spreaders those who get sick often?

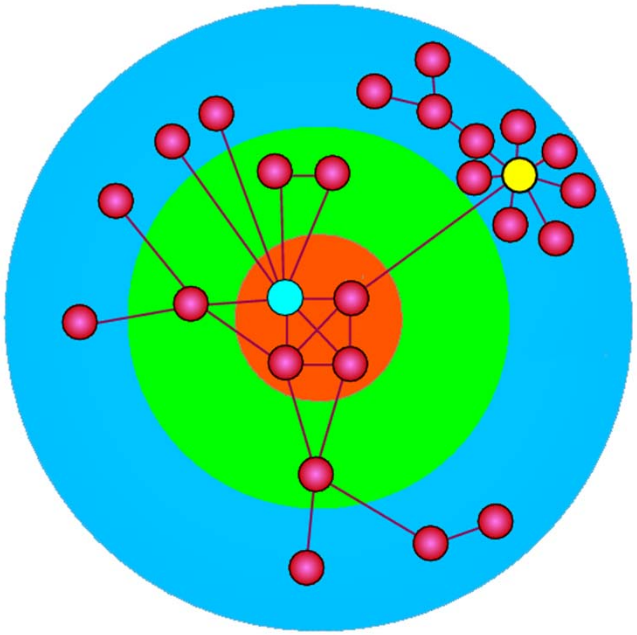
(Further analysis is necessary in the case of inhomogeneous society)

(disorder in transmissibility etc.)

The most efficient spreaders are located in the highest k-shell layers. The most efficient spreaders are also likely to be infected early during an epidemic outbreak.

What happens if virus starts from several origins simultaneously?



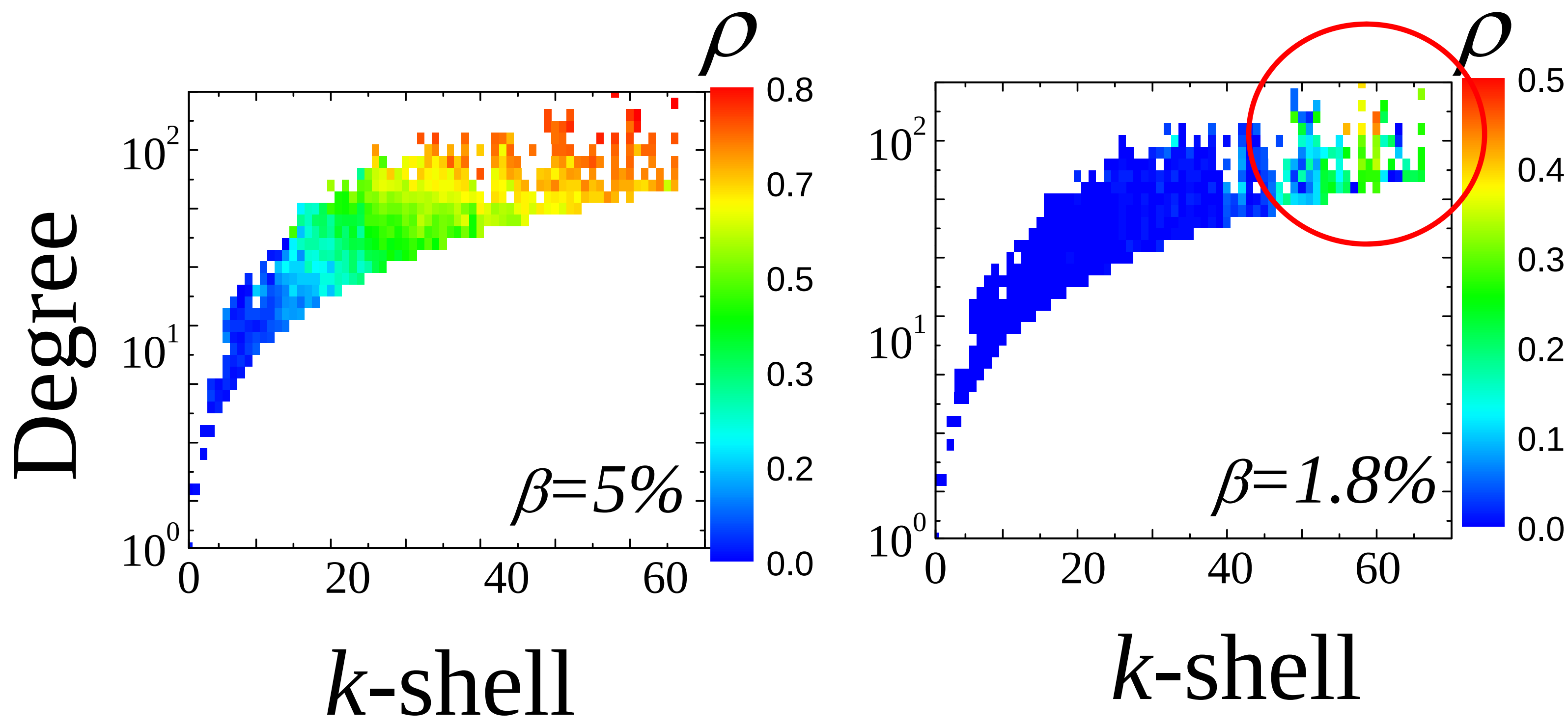


Multiple source spreading is enhanced when one “repels” sources.

Identifying efficient spreaders in the hospital network (SIS)

SIS: Number of infected nodes reaches endemic state (equilibrium)

Persistence $\rho_i(t)$ (probability node ***i*** is infected at time ***t***)



High k-shells form a reservoir where virus can exist locally.

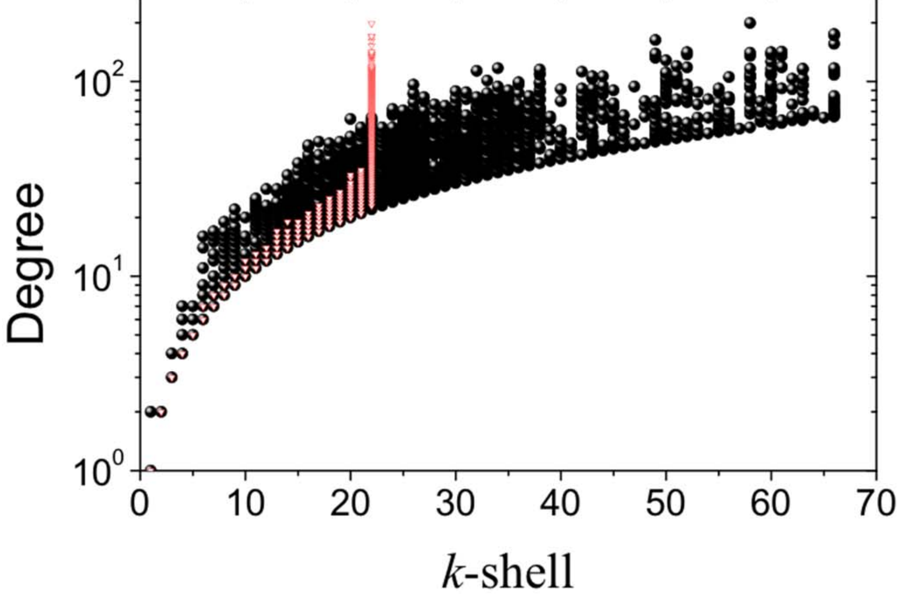
Consistent with core groups (H. Hethcote et al 1984)

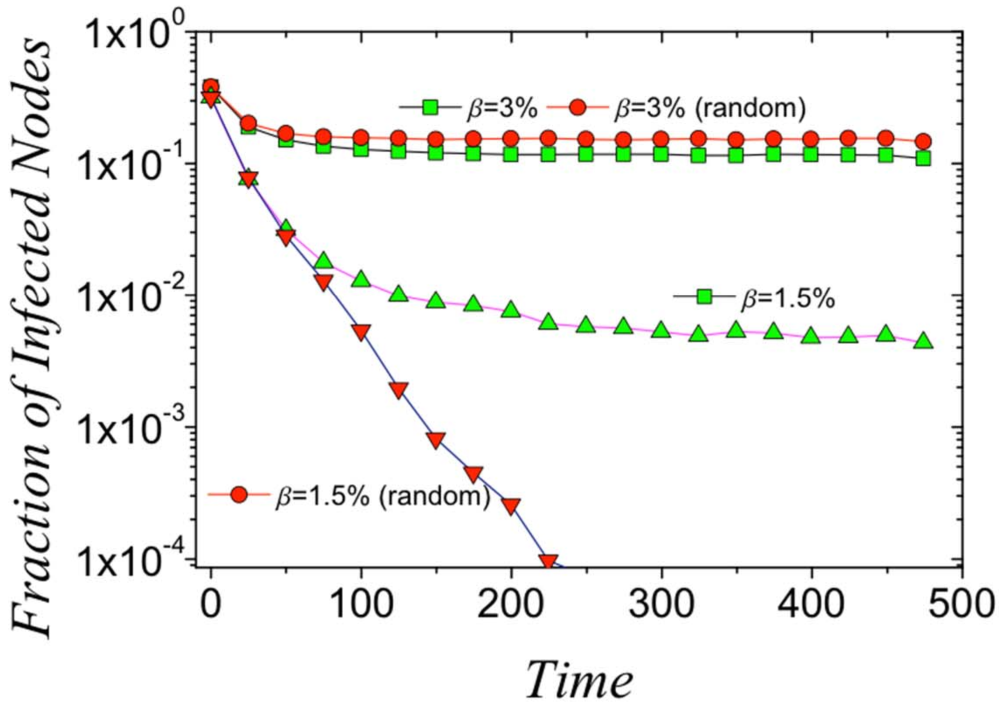
Destroying highest *k*-shells suppresses viruses.

1) Randomly rewire network preserving degree distribution.
2) Compare virus persistence in original and rewired networks.`

Original Contact Network of Inpatients

Randomized Contact Network of Inpatients





Virus of low contagiousness persists only in high *k*-shells.
"Destruction" of high *k*-shells suppresses virus persistence.

Summary

SIR

1) *k*-shell value is a reliable indicator of spreading efficiency.
The most efficient spreaders occupy the innermost k-shells.

2) Multiple source spreading is enhanced when one "repels" sources.

3) Most efficient spreaders are more likely to be infected. It takes less time for the virus to reach most efficient spreaders.

4) Immunization strategies are not reciprocal to spreading strategies.

SIS

5) High k-shells form a reservoir where virus can survive locally and infect neighbor nodes.

6) High k-shells may decrease epidemic threshold.

7) Immunization/Removal of high k-shells helps to suppress virus persistence.

Nature Physics *6* 888 (2010).