Epidemics on Networks

Anderson McKendrick
1876-1943
Pioneered mathematical methods in epidemiology

Analysis of infection spread:

- In theory we should consider all biological processes
- In practice we can only model some of them because of the complexity of the problem

S(usceptible)-I(nfected) Model

Someone who doesn't have the disease but potentially could catch one if comes into contact with infected individual

Someone who has the disease and can potentially pass it to susceptible individual if they come into contact

Susceptible → Infected

Model 1: mass-action approximation (any individual has equal chances to come into contact with every other)

Model 2: network-based (limited possible number of contacts)

n	number of individuals in the system
$S(t)$	number of susceptible individuals at time t
$X(t)$	number of infected individuals at time t
$s = S/n,$	$x = X/n$
S, X	expectations of $S(t)$ and $X(t)$ if we run the same process many times
β	contact rate per-individual

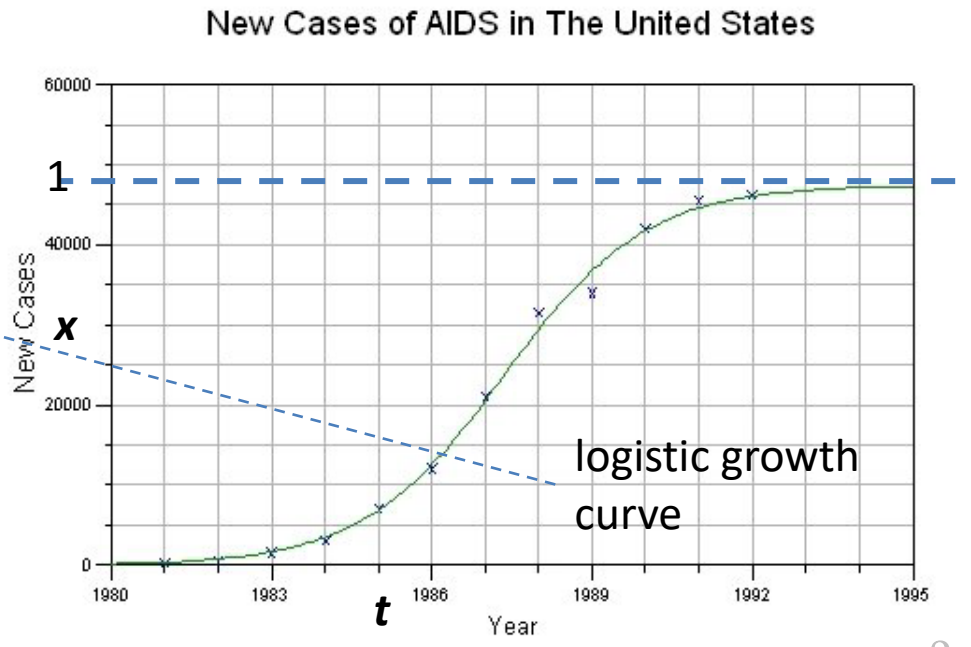
number of susceptible people got infected from one person

s+x = 1

average rate of new infections is $X\beta\frac{S}{n} \Rightarrow \frac{dX}{dt} = X\beta\frac{S}{n}$ or $\frac{dx}{dt} = \beta(1-x)x$

with solution $x(t) = \frac{x_0 e^{\beta t}}{1 - x_0 + x_0 e^{\beta t}}$

value of *x* at *t=0*



http://www.nlreg.com/aids.htm

SIR model

Susceptible

Infected

Recovered = Removed or Dead

SIR Model

n	number of individuals in the system
$S(t)$	number of susceptible individuals at time t
$X(t)$	number of infected individuals at time t
$R(t)$	number of recovered individuals at time t
S, X, R	expectations of $S(t)$, $X(t)$, $R(t)$
$s = S/n,$	$x = X/n$, $r = R/n$
β	contact rate per-individual
γ	recovery rate per-individual

number of susceptible decreases

$$\frac{ds}{dt} = -\beta s x$$

$$\frac{dx}{dt} = \beta s x - \gamma x$$

$$\frac{dr}{dt} = \gamma x$$

Note: $s + x + r = 1$

eliminate *x* and integrate both parts

$$\frac{1}{s}\frac{ds}{dt} = -\frac{\beta}{\gamma}\frac{dr}{dt} \Rightarrow s = s_0 e^{-\frac{\beta r}{\gamma}}$$

$$x = 1 - s - r \Rightarrow \frac{dr}{dt} = \gamma\left(1 - r - s_0 e^{-\frac{\beta r}{\gamma}}\right)$$

$$r = 1 - s_0 e^{-\frac{\beta r}{\gamma}}$$

in the limit of *n*

Remember the size of giant component in Poisson graph?
$S = 1 - e^{-cS}$

SIR Model

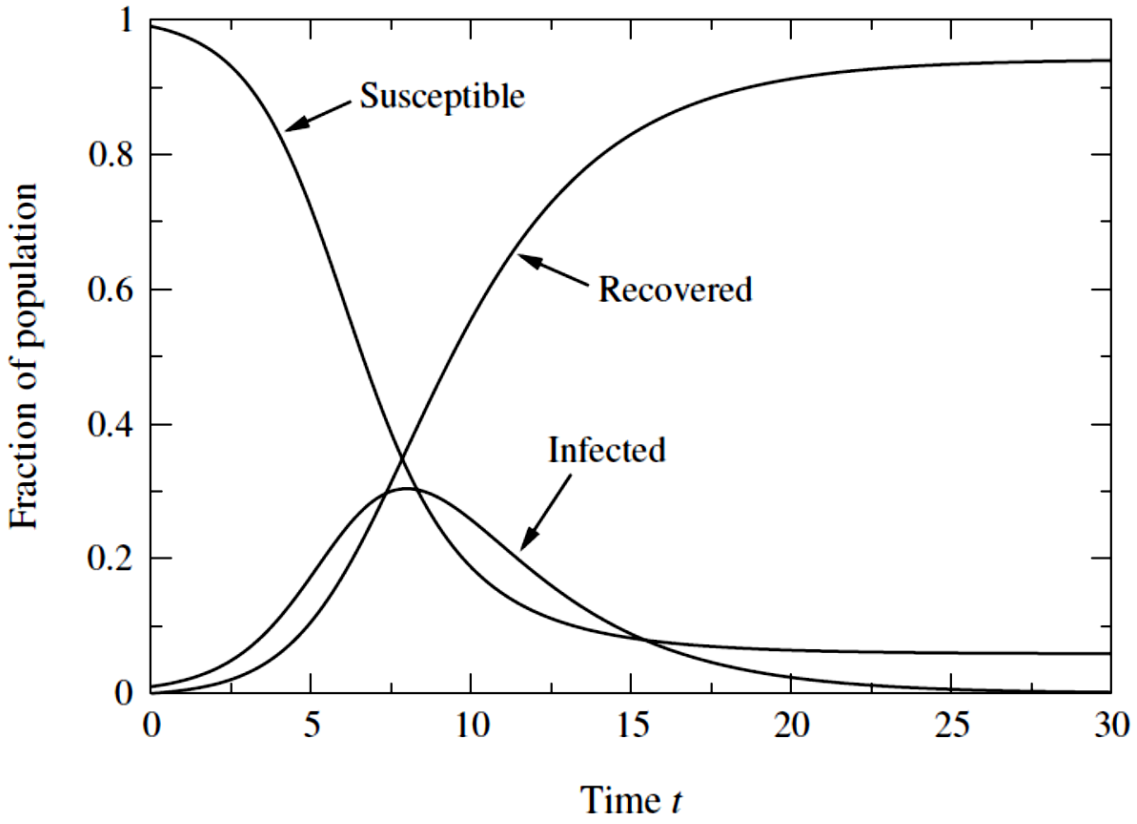


Figure 16.2: Time evolution of the SIR model. A numerical solution of the SIR equations (16.9). The three curves show the fractions of the population in the susceptible, infected, and recovered states as a function of time. The parameters are $\beta = 1$, $\gamma = 0.4$, $s_0 = 0.99$, $x_0 = 0.01$, and $r_0 = 0$.

SIR Model

β contact rate per-individual
γ recovery rate per-individual

Epidemic threshold: the transition between epidemic and non-epidemic regimes happens when

$$\beta = \gamma$$

Basic reproduction number is the number of cases one case generates on average over the course of its infectious period

$$R_0 = \beta/\gamma$$

SIS Model

Susceptible → Infected → Susceptible

n	number of individuals in the system
$S(t)$	number of susceptible individuals at time t
$X(t)$	number of infected individuals at time t
$R(t)$	number of recovered individuals at time t
S, X, R	expectations of $S(t)$, $X(t)$, $R(t)$
$s = S/n,$	$x = X/n$, $r = R/n$
β	contact rate per-individual
γ	recovery rate per-individual

$$\frac{ds}{dt} = \gamma x - \beta s x$$

$$\frac{dx}{dt} = \beta s x - \gamma x$$

transmission rate

recovery rate

$$s + x = 1$$

$$x(t) = x_0 \frac{(\beta - \gamma)\, e^{(\beta-\gamma)t}}{\beta - \gamma + \beta x_0 e^{(\beta-\gamma)t}}$$

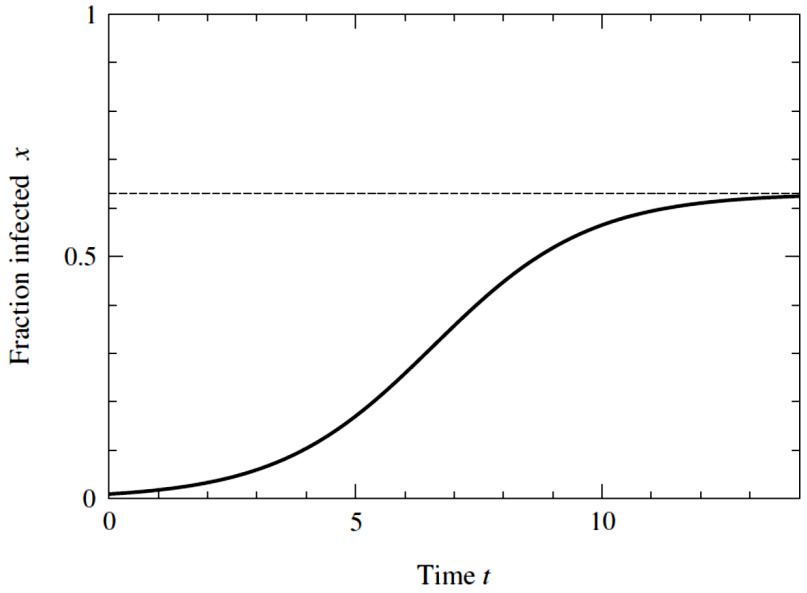


Figure 16.3: Fraction of infected individuals in the SIS model. The fraction of infected individuals in the SIS model grows with time following a logistic curve, as in the SI model. Unlike the SI model, however, the fraction infected never reaches unity, tending instead to an intermediate value (dashed line) at which the rates of infection and recovery are balanced. (Compare this figure with Fig. 16.1 for the SI model.)

Networks (not full mixing), SIR

Transmission/infection rates have the same meaning but they will work via edges only

β transmission or infection rate (similar to the contact rate per-individual in full mixing model)

γ recovery rate per-individual

Probability (in full mixing) that the individual is still infected after time τ is

$$\lim_{\delta\tau \to 0} (1 - \gamma\delta\tau)^{\tau/\delta\tau} = e^{-\gamma\tau}$$

In the same way the probability that the disease is transmitted

$$\phi = 1 - e^{-\beta\tau}$$

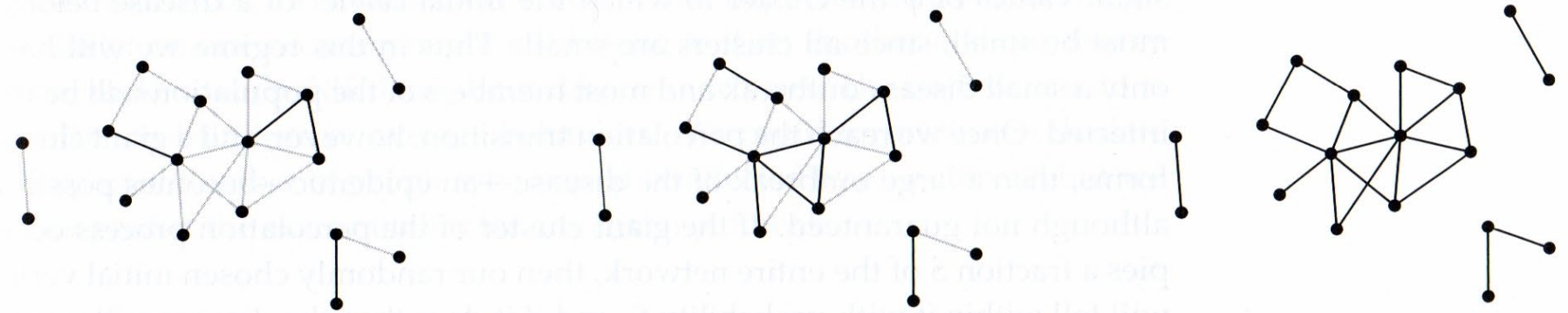
Assumption for networks: every infected individual remains infected for the same length of time, i.e., ϕ is constant.

Consider edge percolation with occupation probability ϕ, i.e., leave edges for disease transmission.

Edges do not necessarily transmit the disease. They can only potentially transmit it with some probability.

- Edge (or bond) percolation starts at randomly chosen node
- Let us take our network "occupy" each edge with probability ϕ, or not with probability $1 - \phi$. The occupied edges represent those along which disease *may be* transmitted if it reaches either of the nodes at the ends of the edge.

For small values of ϕ the cluster to which the initial carrier of a disease belongs must be small, since all clusters are small. Thus, in this regime we will have only a small disease outbreak. Once we reach the percolation transition, however, and a giant cluster forms, then a large outbreak becomes possible, although not guaranteed.



(a) $\phi = 0.2$ (b) $\phi = 0.5$ (c) $\phi = 1$

Figure 17.4: Bond percolation. In bond percolation, a fraction ϕ of the edges in a network are filled in or "occupied" at random to create connected clusters of vertices. (a) For small occupation probability ϕ the clusters are small. (b) Above the percolation threshold a large cluster forms, though there are usually still some small clusters as well. (c) When $\phi = 1$ all edges are occupied but the large cluster may still not fill the whole network: at $\phi = 1$ the largest cluster corresponds to the largest component of the network, which is often just a subset of the whole network.

Consider SIR on network with degree distribution p_k

- u - average prob that $i \in V$ is not connected to gc via one particular edge ij, i.e., either ij is unoccupied $(1-\phi)$ or $j \notin$ gc (u^k is j has k neighbors)

$$u = 1 - \phi + \phi \sum_{k=0}^{\infty} q_k u^k = 1 - \phi + \phi g_1(u)$$

gen function of excess deg distribution

- Thus, gc fraction is $S = 1 - \sum_{k=0}^{\infty} p_k u^k = 1 - g_0(u) \Rightarrow \phi_c = \frac{\langle k \rangle}{\langle k^2 \rangle - \langle k \rangle}$

in node percolation it was $S = \phi(1 - g_0(u))$

$$\phi = 1 - e^{-\beta\tau} \;\Rightarrow\; \beta\tau = -\ln(1-\phi_c) = \ln \frac{\langle k^2 \rangle - \langle k \rangle}{\langle k^2 \rangle - 2\langle k \rangle}$$

Epidemic threshold if >= then possible epidemic

Epidemic Thresholds in Real Networks

see paper by Chakrabarti, Wang, Wang, Leskovec, Faloutsos (will follow their notation)

Q: How to model viral propagation on arbitrary network? Which node is best to immunize? When is the outbreak?

Homogeneous model – everybody has equal contact to others (rate of infection depends on the density of population). ***Nonhomogeneous model*** – otherwise.

Kephart-White model: directed graph; $i \to j \in E$ if i can infect j; β is rate of infection (virus birth rate); δ is virus death rate (node-curing).

η_t - size of infected population at time t

$\langle k \rangle$ - average degree

Solution when value change over time

$$\frac{d\eta_t}{dt} = \beta\langle k\rangle\eta_t\left(1 - \frac{\eta_t}{N}\right) - \delta\eta_t \qquad \left(\text{ its steady state } \eta = N(1 - \frac{\delta}{\beta\langle k\rangle})\right)$$

Intuition: epidemic threshold τ is a value such that a viral outbreak dies out quickly

$$\beta/\delta < \tau,\ \tau_{KW} = 1/\langle k\rangle$$

good approximation for homogeneous networks

Consider small time steps $\Delta t \to 0$; the system is a Markov chain with 2^N configurations of nodes S/I.

$p_{i,t}$ - probability i is infected at time t
$\zeta_{i,t}$ - probability that i **will not receive** infections from $N(i)$ in the next step

$$\zeta_{i,t} = \prod_{j \in N(i)} (p_{j,t-1}(1-\beta) + (1-p_{j,t-1})) = \prod_{j \in N(i)} (1 - \beta p_{j,t-1})$$

Non-linear dynamical system

$$1 - p_{i,t} = (1 - p_{i,t-1})\zeta_{i,t} + \delta p_{i,t-1}\zeta_{i,t}$$

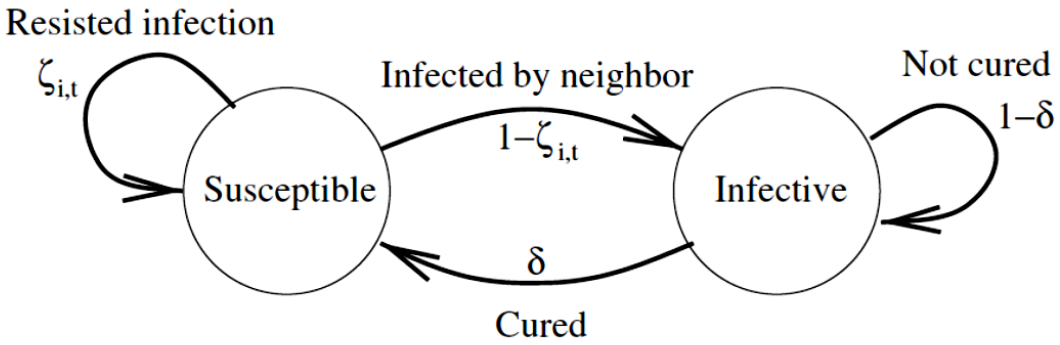


Fig. 1. The SIS model, as seen from a single node. Each node, at each time-step t, is either susceptible (S) or infective (I). A susceptible node i is currently healthy, but can be infected (with probability $1 - \zeta_{i,t}$) by receiving the virus from a neighbor. An infective node can be cured with probability δ; it then goes back to being susceptible. Note that $\zeta_{i,t}$ depends on the both the virus birth rate β and the network topology around node i.

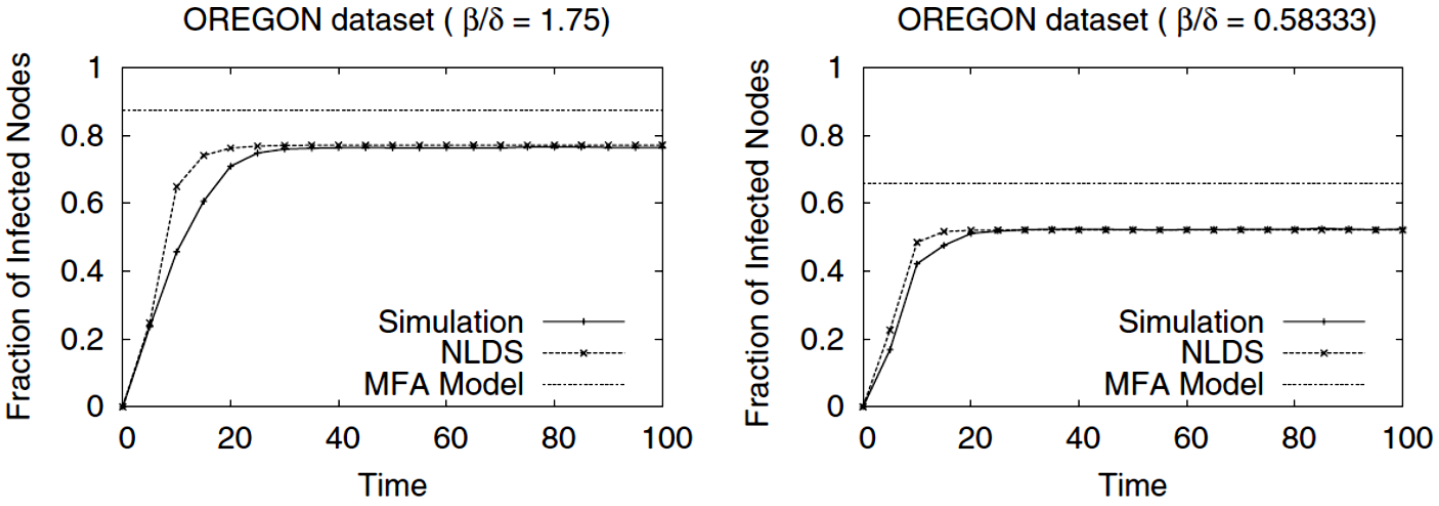


Fig. 2. Experiments on the real-world Oregon graph. The plots show the time evolution of infection in the *Oregon* network. Both simulations were performed with fixed β, but varying δ. In both cases, our model conforms more precisely to the simulation results than the MFA model.

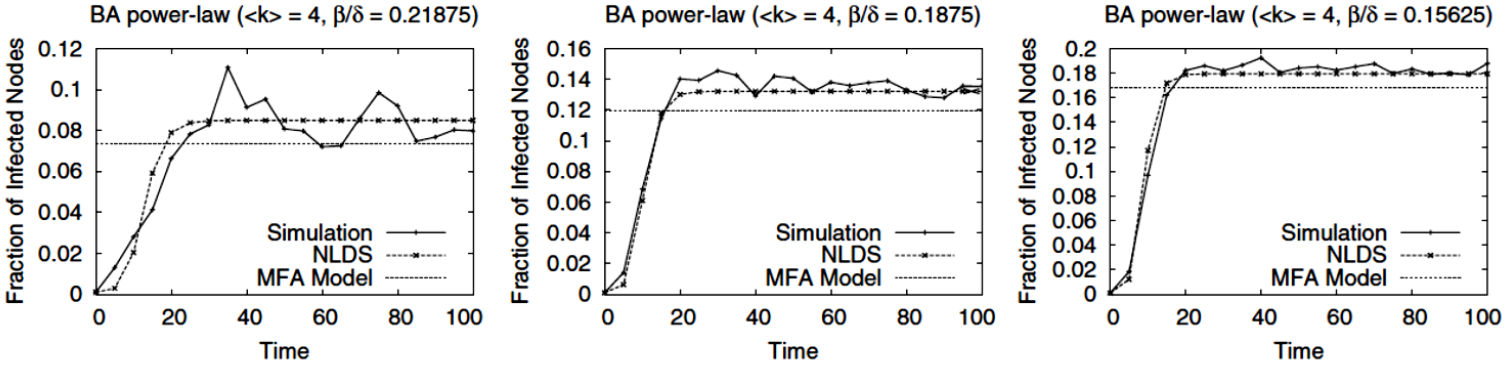


Fig. 3. Experiments on BA power-law topology. We compare our model and the MFA model to the simulation results for several choices of β, keeping δ fixed. The plots show time evolution of infected population in a 1000-node BA power-law network. Our model outperforms the other model in steady state predictions by a slight margin.

Definition (NLDS Epidemic Threshold). The epidemic threshold τ for NLDS is a value such that

$$\begin{aligned} \beta/\delta < \tau &\Rightarrow \text{infection dies out over time} \\ \beta/\delta > \tau &\Rightarrow \text{infection survives and becomes an epidemic} \end{aligned}$$

Theorem 1. In NLDS, the epidemic threshold τ for an undirected graph is

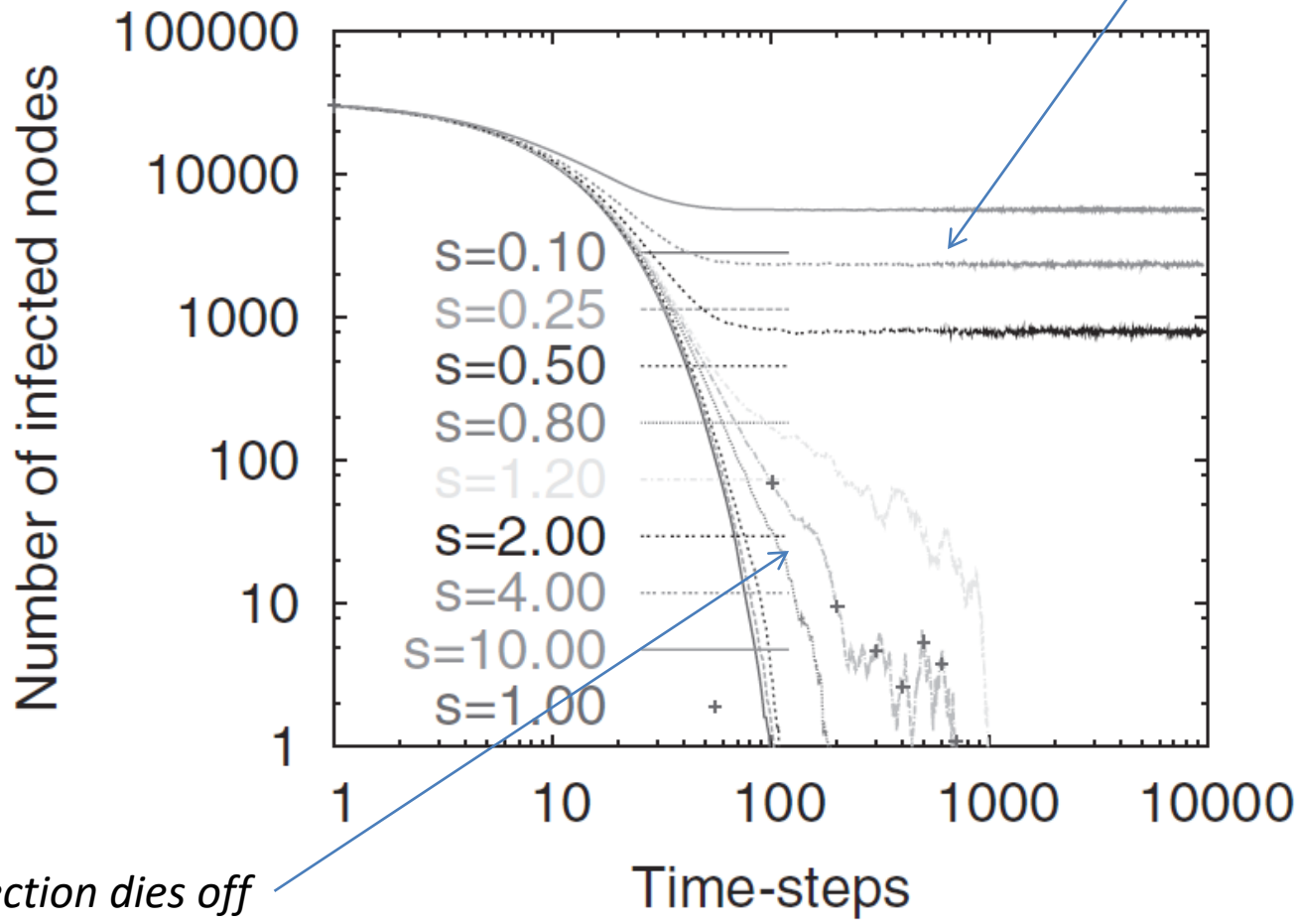
$$\tau = 1/\lambda_{1,A},$$

where $\lambda_{1,A}$ is the largest eigenvalue of the adjacency matrix A.
$s = \beta/\delta \cdot \lambda_{1,A}$ is a score of infection.

Theorem 2. When an epidemic is diminishing, i.e., $\beta/\delta < 1/\lambda_{1,A}$, the probability of infection decays at least exponentially over time.

s>1 - infection survives

Enron database graph (email exchanges)
33K nodes, 361K edges



s< 1 – infection dies off

Immunization

Q: We have a budget of *k* nodes that can be immunized. How to choose these nodes?

Theorem 1 → We need to decrease the maximal eigenvalue of the adjacency matrix.

Observation: Popular "targeted" immunization strategy (immunization of high-degree nodes) does not agree with maximum eigenvalue decrease! Example:

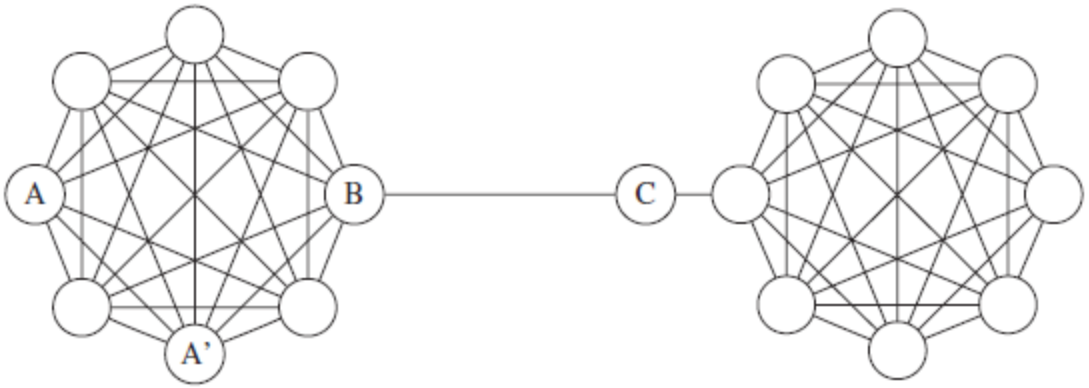
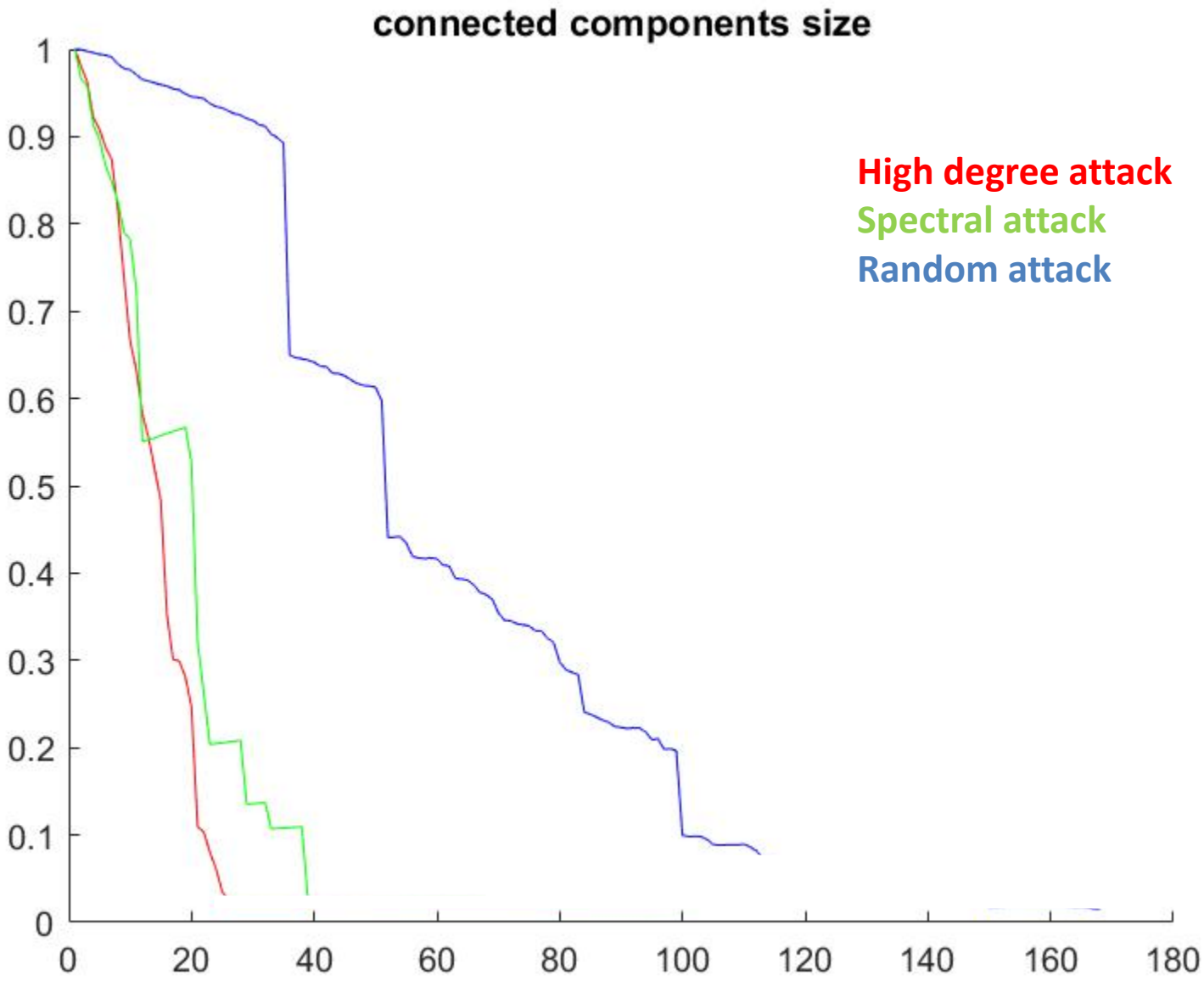


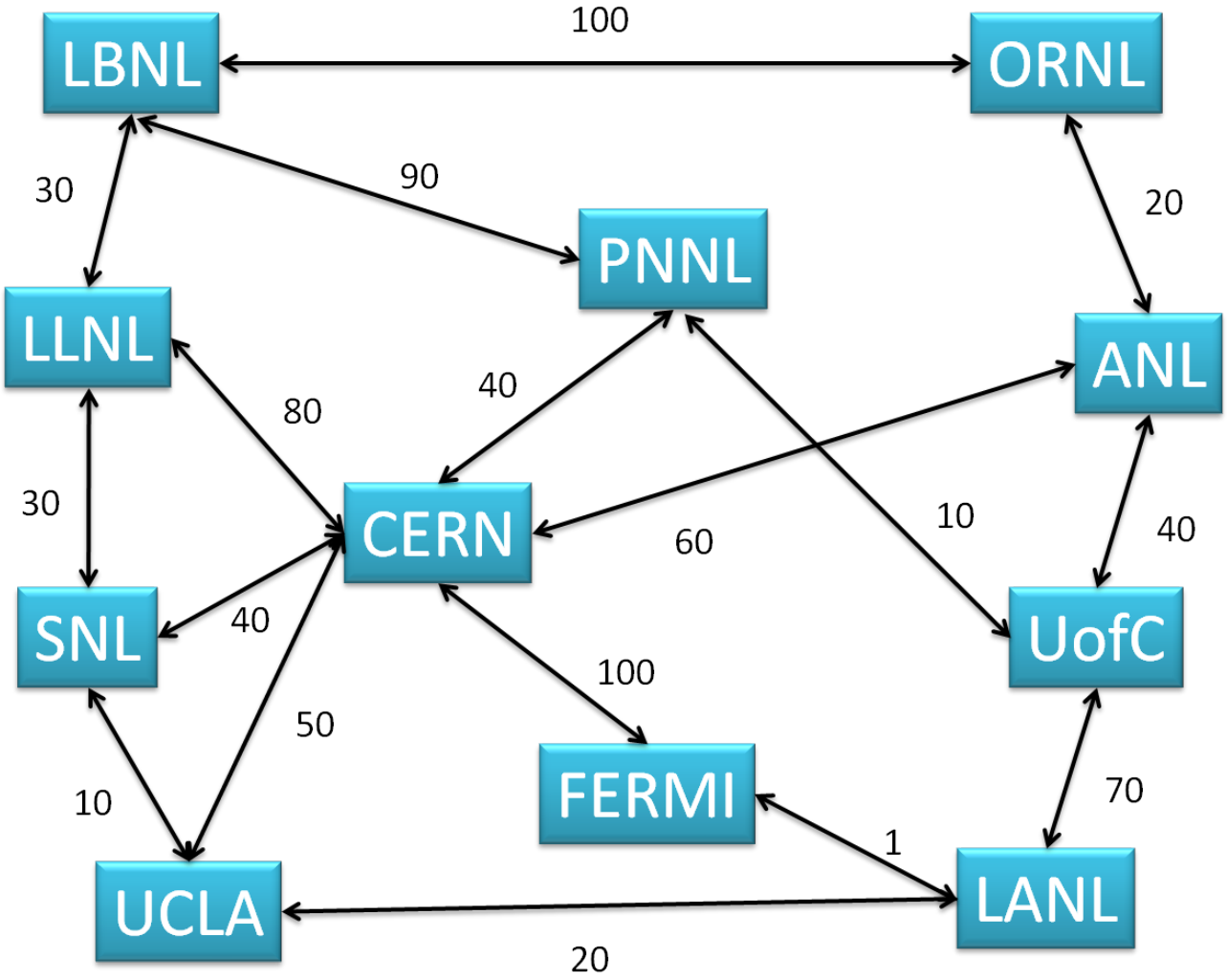
Fig. 8. The "bar-bell" graph. Two cliques of the same size connected with a bridge.

Matlab immunization demo

Three attacks on power grid

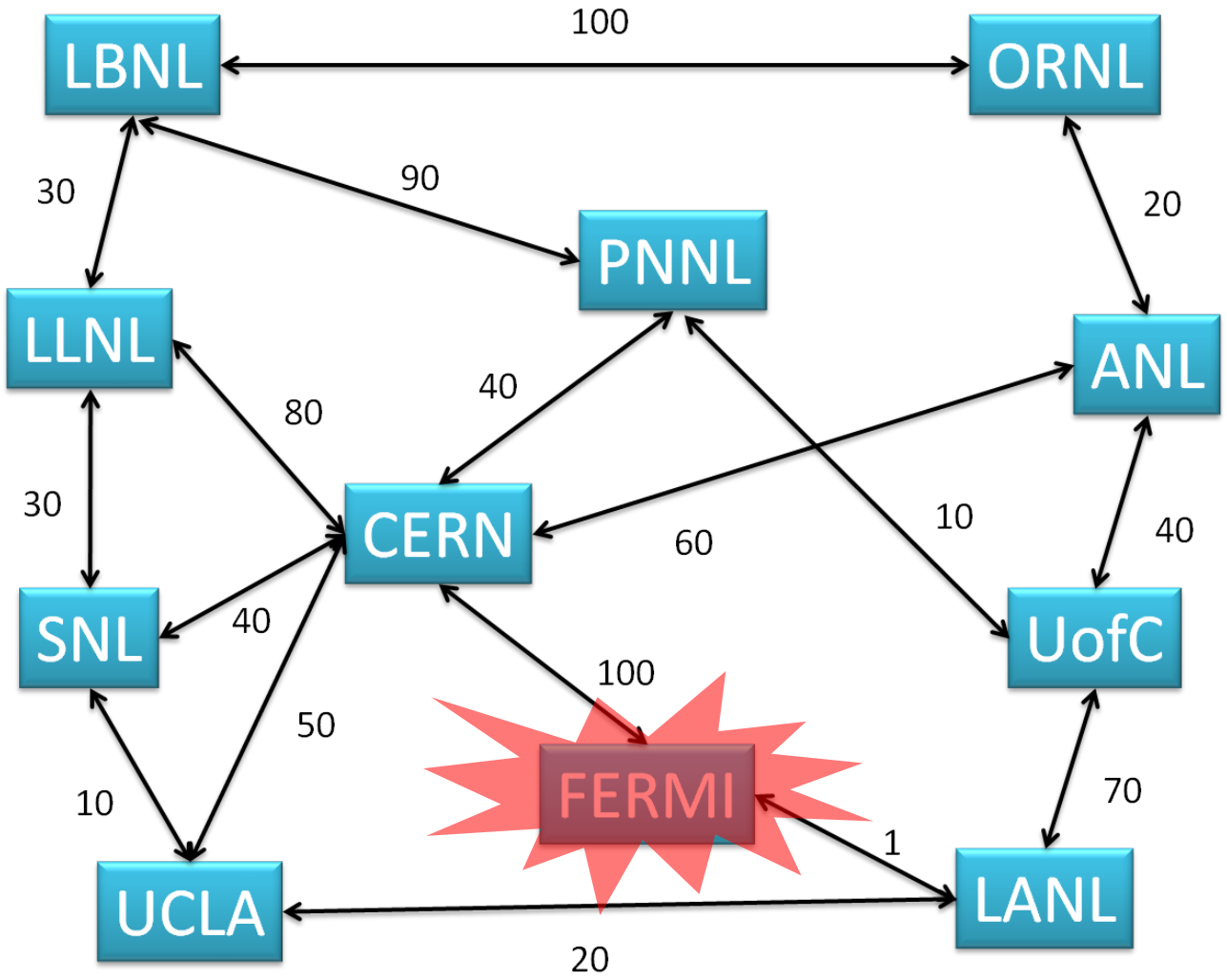


Example of a Model: Response to Epidemics and Cyber Attacks



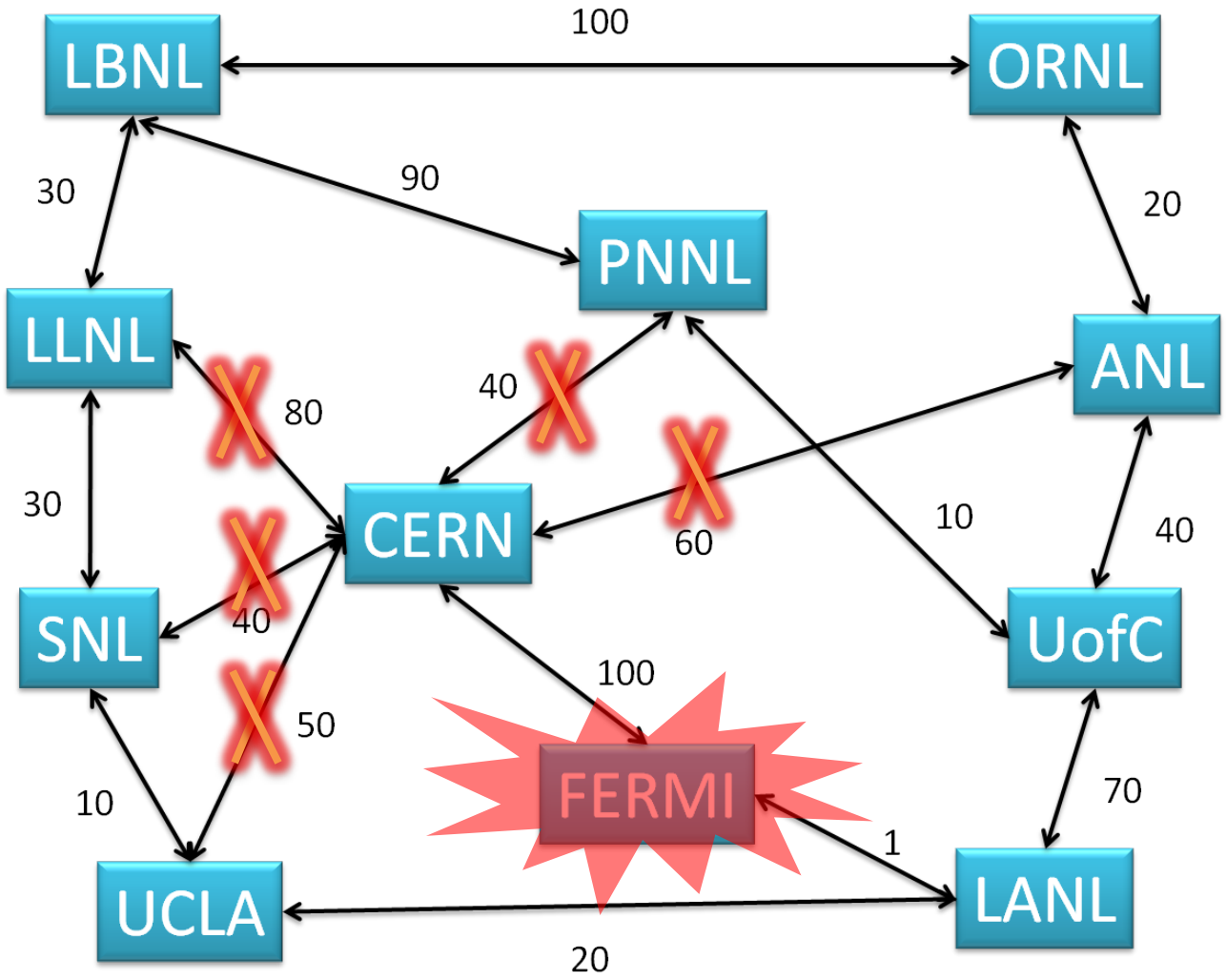
Open Science Grid: collaboration network example

Example of a Model: Response to Epidemics and Cyber Attacks



Open Science Grid: collaboration network example

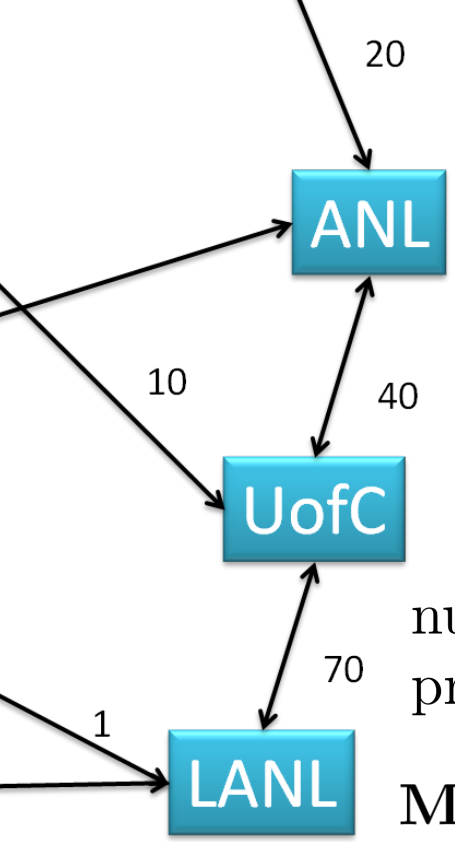
Example of a Model: Response to Epidemics and Cyber Attacks



Open Science Grid: collaboration network example

Example of a Model: Response to Epidemics and Cyber Attacks

Goldberg, Leyffer, S "Optimal Response to Epidemics and Cyber Attacks in Networks", 2011



site i closed/open $\quad x_i \in \{0,1\}$

infection probability at i $\quad \phi_i$

number of shared users $\quad w_{ij}$

probability of $j \to i$ spread $\quad p_{ij}$

Model

connections between open sites, i.e., the utility of network

$$\underset{x}{\text{maximize}} \quad \sum_{ij \in E} w_{ij} x_i x_j$$

$$\text{subject to} \quad x_i - \prod_{j \in N(i)} (1 - p_{ij}\phi_j x_j) \leq t_i \quad \forall i \in V$$

infection at node *i* is less than some constant

$$x \in \{0,1\}^n$$

Large-scale networks

Ratios Between MA and ILS