Propagazione di virus informatici in reti complesse

Michele Garetto

12/04/2006

E-mail virus propagation

- The virus propagates as an attachment to e-mail messages
- Requires human assistance
  - → random time elapses before the recipient reads the message
  - → the "click" probability
- The virus makes use of the recipient's address book to send copies of itself

E-mail virus propagation

- The virus propagates on the network graph induced by email address books:
  - Each node stands for an email address
  - Edges represents social or business relationships

- The resulting graph is expected to have "small world" and "scale-free" properties:
  → small *characteristic path length*
  → high *clustering coefficient*
  → *power-law degree distribution*

Our work on e-mail viruses

- Static analysis: what is the final size of an infection outbreak ? What is the probability of a large-scale epidemic ?

  → Percolation problem

- Dynamic analysis: how fast is the propagation of a virus in a network ? What is the (average) number of infected hosts as a function of time ?

  → Interactive Markov Chains

M. Garetto, D. Towsley, W. Gong, *Modeling Malware Spreading Dynamics, IEEE INFOCOM 2003*, San Francisco, CA, April 2003

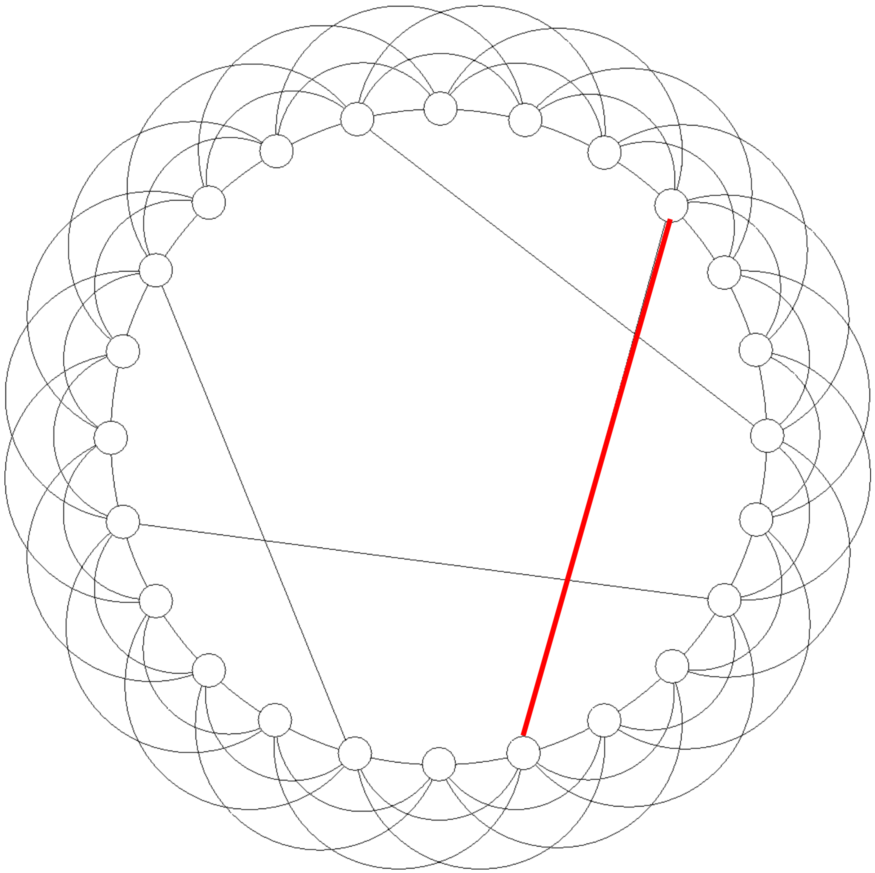
Introduction to graph percolation

- Consider a completely connected graph $G = \{V,E\}$
- Site percolation: mark each vertex with probability p
- Bond percolation: mark each edge with probability p

- For which values of p the resulting graph is still connected ?

→ percolation threshold
→ phase transition

The "small-world" model of Watts and Strogatz



A ring lattice with additional random shortcuts

N = 24 S = 4 k = 3

Parameters:

N = number of nodes

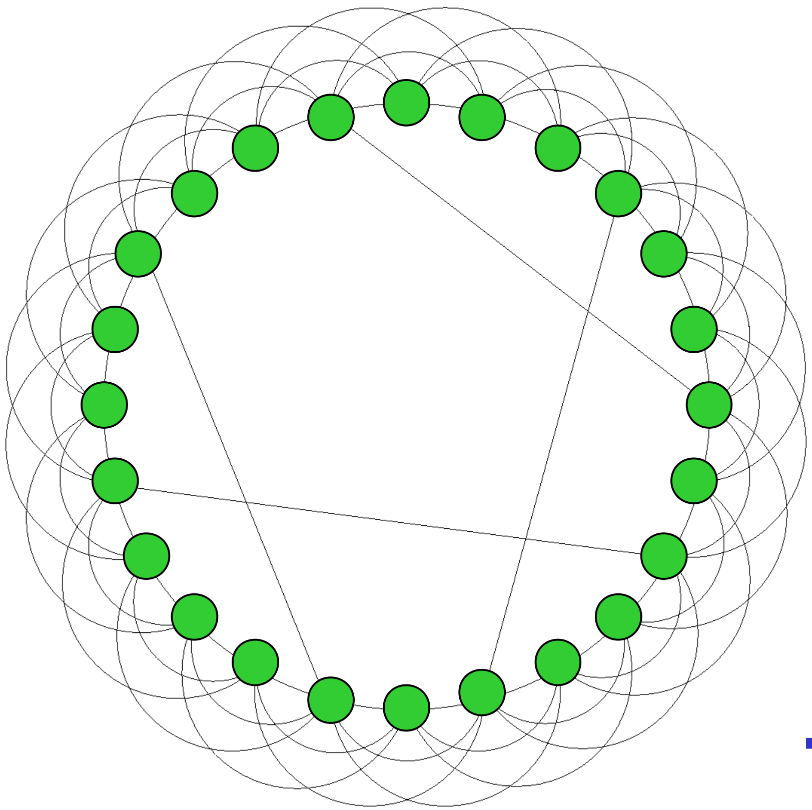
S = number of shortcuts

(ϕ = shortcuts density)

k = lattice connectivity

(number of neighbors on each side of a node)

Virus propagation on the small-world graph



● = susceptible
● = infected
● = immune

- Not all of the susceptible nodes receive a copy of the virus !

→ site percolation problem

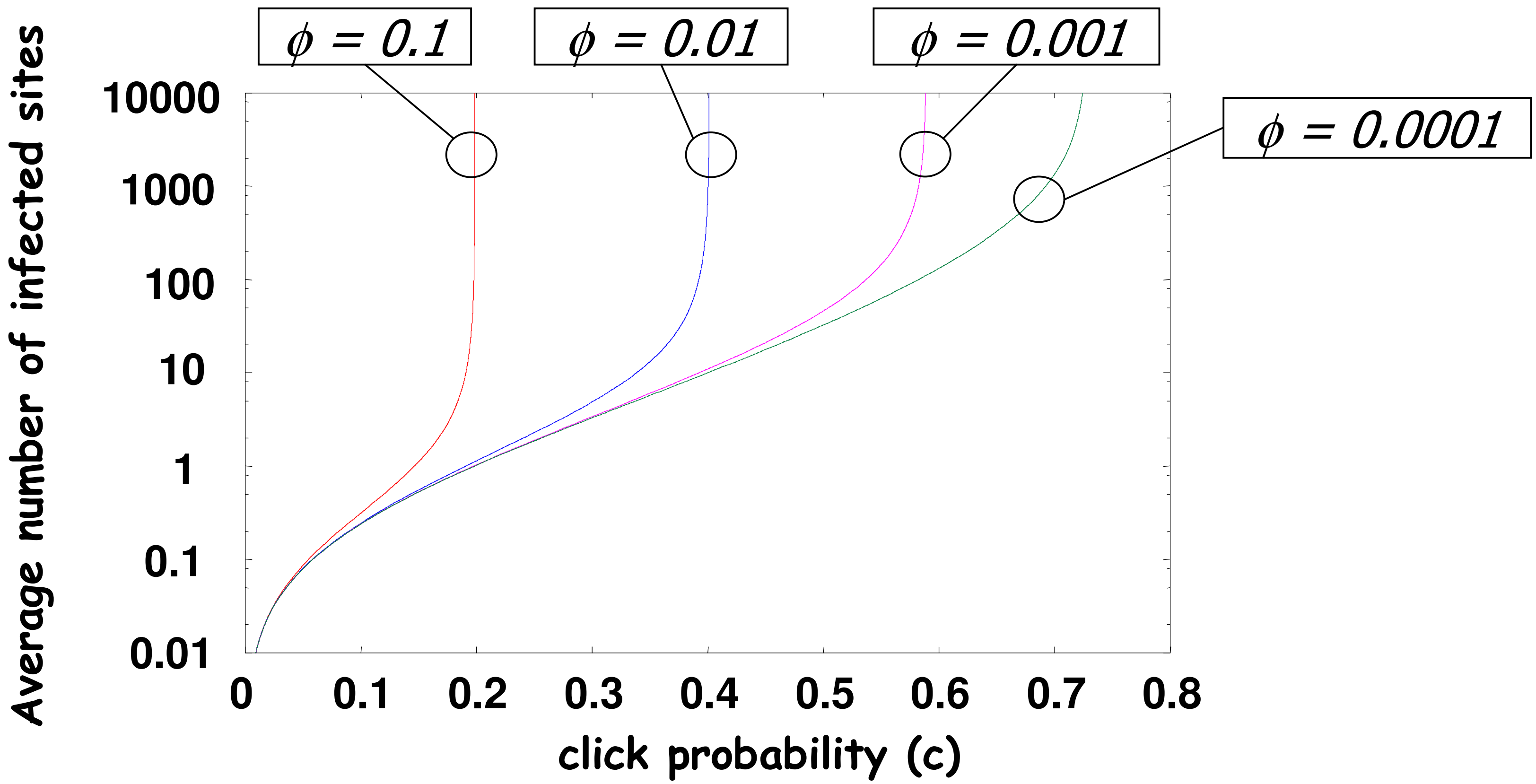
(node occupation probability = click probability)

Site percolation problem on the small world graph: exact asymptotic result

$$E[I_\infty] = \frac{c\,(1+q)}{1 - q - 2k\phi c\,(1+q)}$$

[Moore, Newman 2000]

$$q = 1 - (1-c)^k$$



Scale-free networks

- Power-law degree distribution

$$P(k) \sim k^{-\lambda}$$

- In case of random wiring, no percolation threshold exists when

$$2 < \lambda < 3$$

- Result holds also with connectivity correlations (assortative/disassortative mixing)
- In highly clustered networks ?

Our work on e-mail viruses

- Static analysis: what is the final size of an infection outbreak ? What is the probability of a large-scale epidemic ?

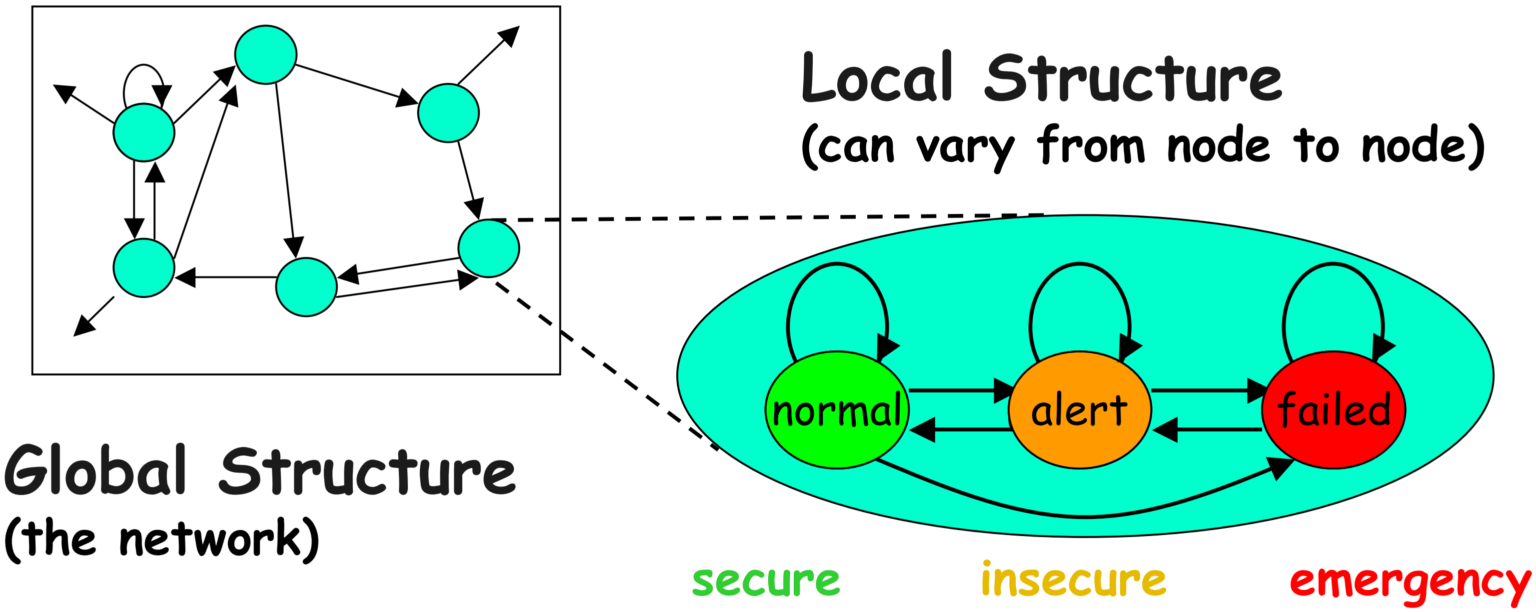
  → Percolation problem

- Dynamic analysis: how fast is the propagation of a virus in a network ? What is the (average) number of infected hosts as a function of time ?

  → Interactive Markov Chains

The "Interactive Markov Chains" (IMC) Modeling Framework

- Global network structure ... but locally a Markov chain



- Each node is represented by a Markov chain, whose state transitions are influenced by the states of its neighbors

Computational complexity issue

- The whole system evolves according to a global Markov chain G, whose state space dimension ($\#G$) is equal to the product of the local chain dimensions ($\#L$)

$$\#G = \#L^{N}$$

- The exact solution of the global Markov chain is feasible only for small systems

example:
- 20 nodes
- binary status (0 = not infected, 1 = infected)

→ 2^{20} states (~ 1 million) !

The Influence Model

- First proposed by

  C. Asavathiratham, "*The influence model: a tractable representation for the dynamics of networked Markov chains,*" Ph.D. dissertation, EECS Dept., MIT, Oct 2000

- Is a discrete-time Markov process, with state space dimension ($\#G$) equal to the sum of local chain dimensions ($\#L$)

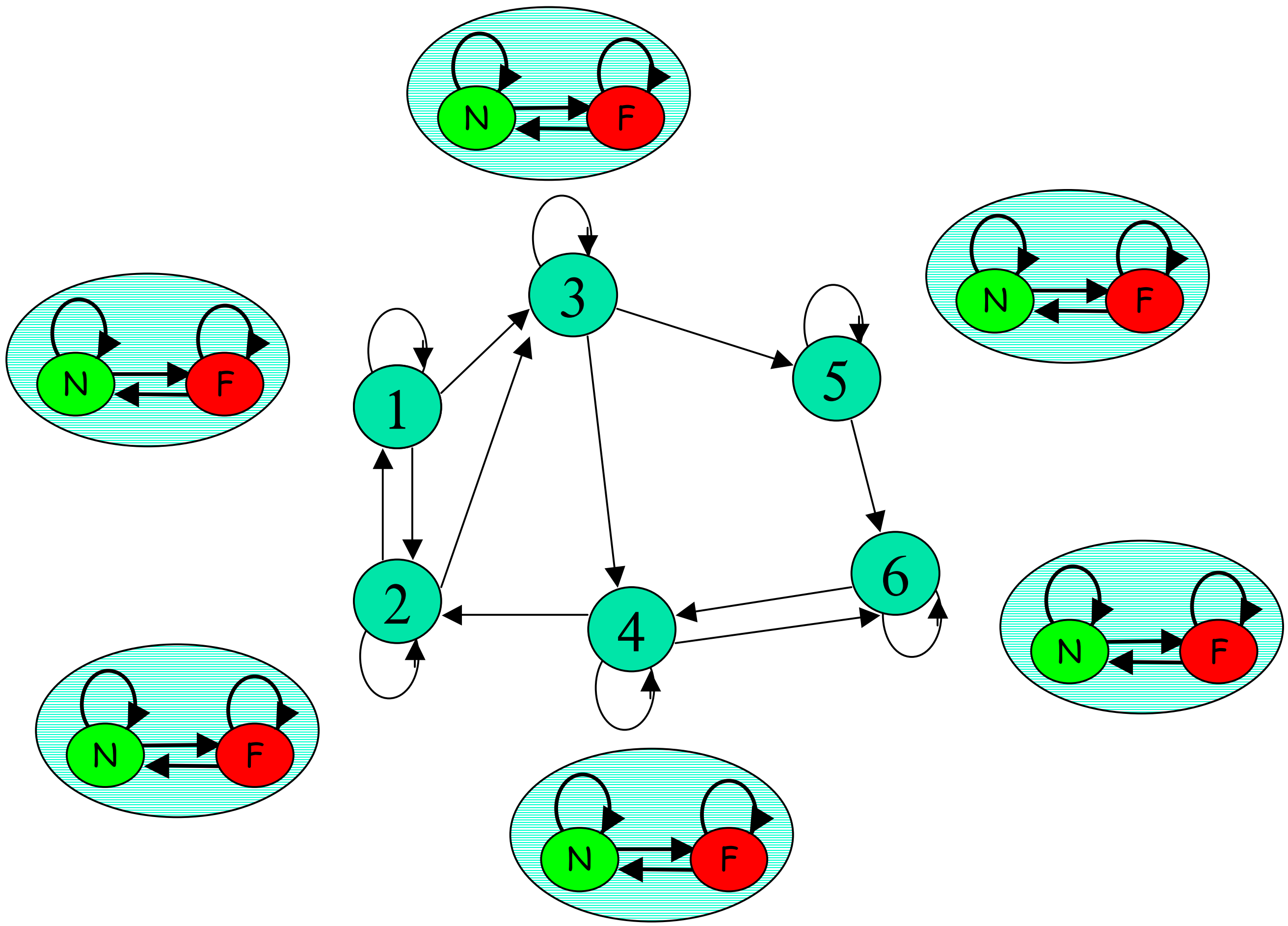
$$\#G = N \cdot \#L$$

- It allows to consider very large systems. Example:

  - 100000 nodes
  - binary status (0 = not infected, 1 = infected)

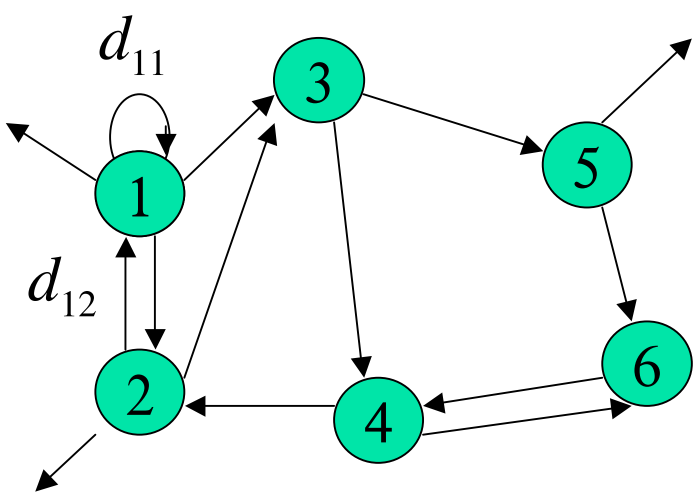
    → 200000 states

The Influence Model



The Influence Model

- At each time step, a node i is "influenced" by one of its neighbors j (or by itself) according to some probability $d_{i,j}$



➢ The probabilities $d_{i,j}$ are the weights associated with incoming edges

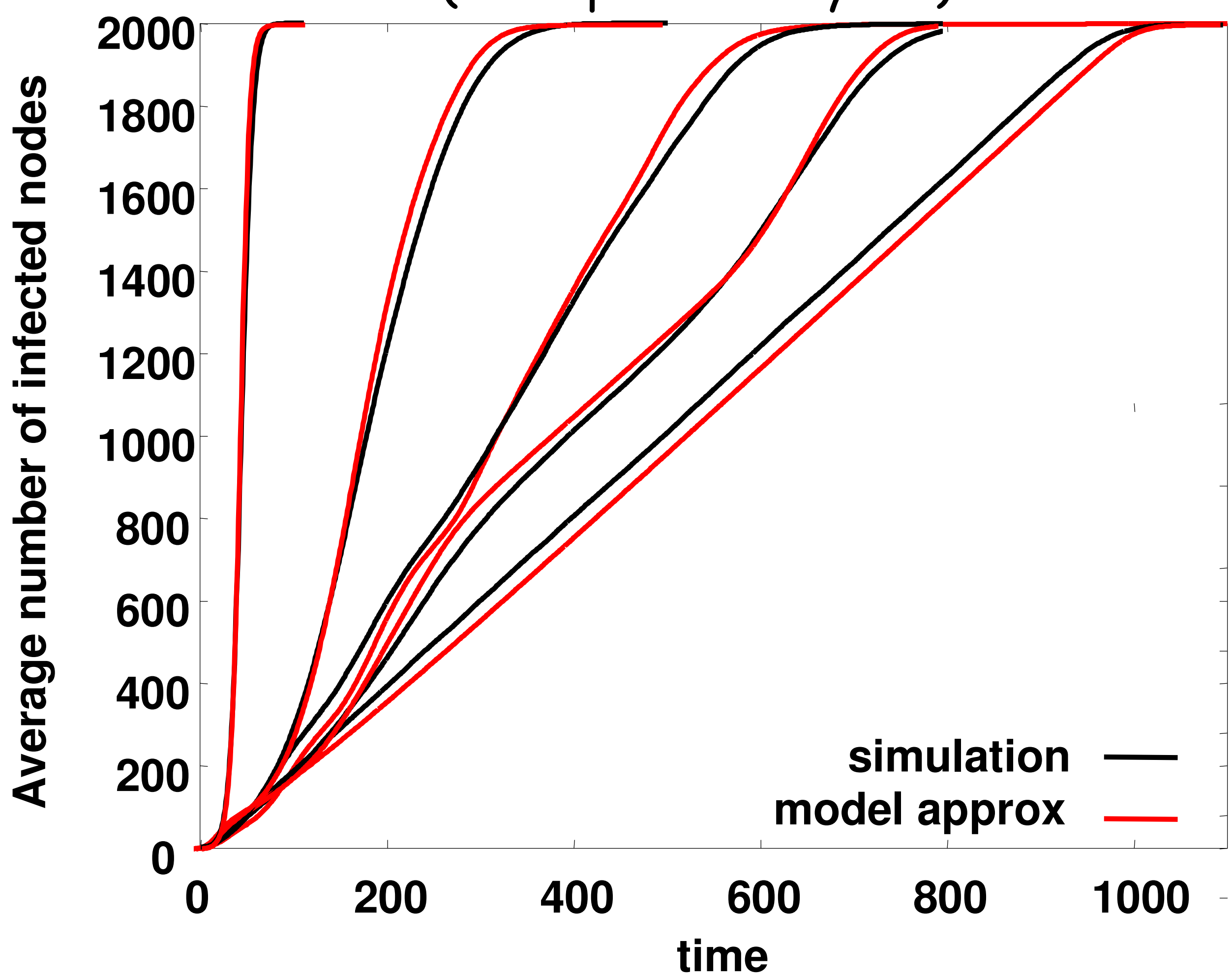
➢ $$\sum_j d_{i,j} = 1$$

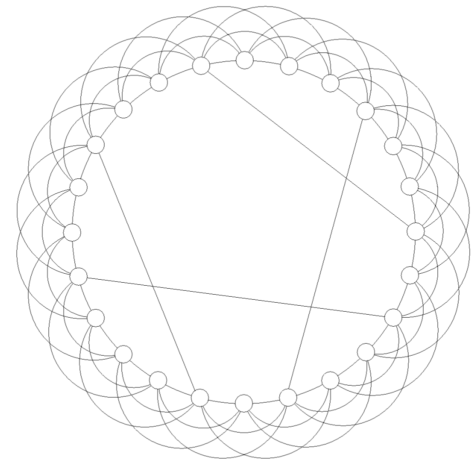
➢ self-influence corresponds to self-loop

- The new state of i is determined by the state of the influencing node according to a transition matrix $A_{i,j}$

Propagation on the small world graph: the impact of topology

(click probability = 1)



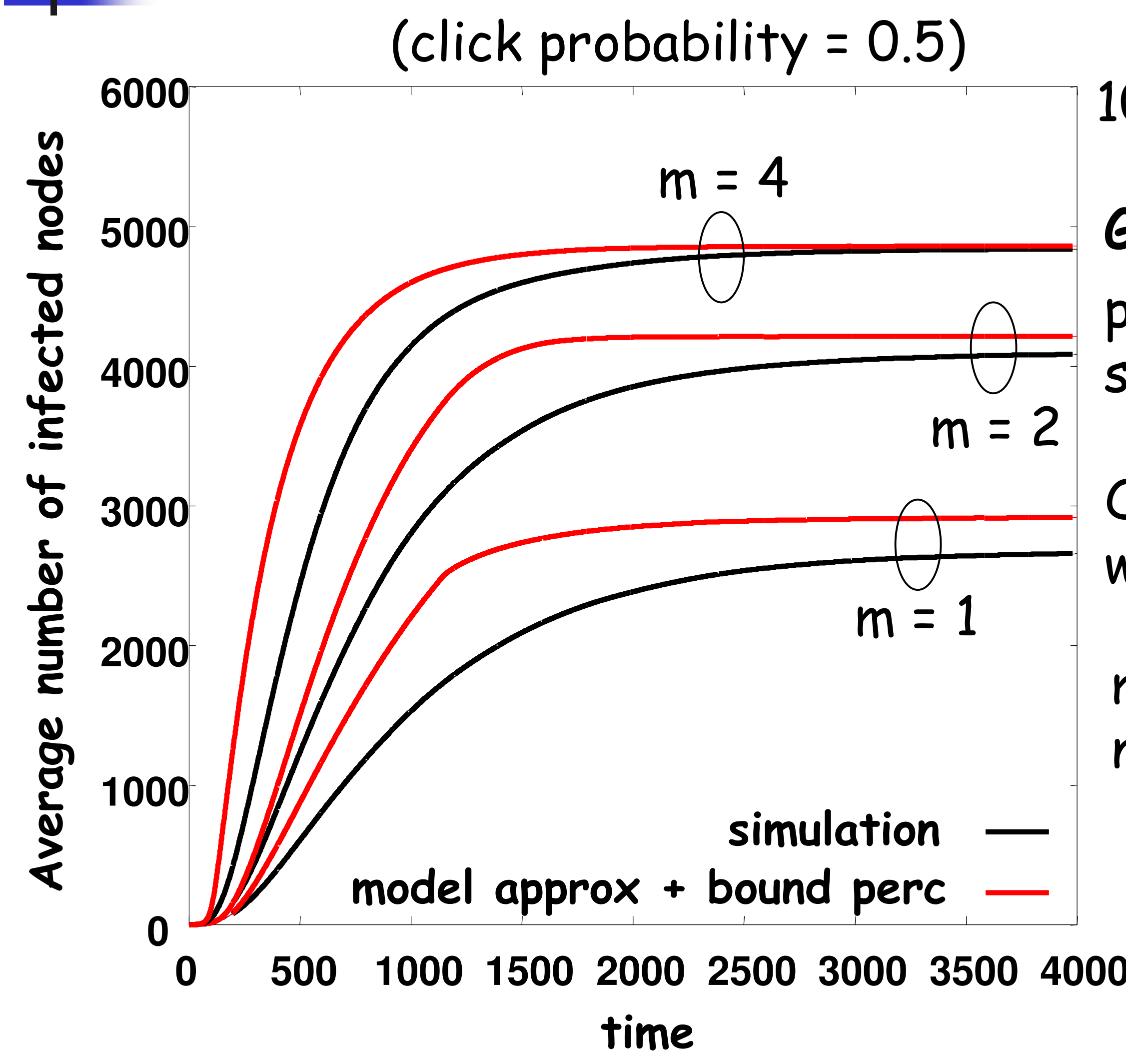


2000 nodes

[illegible]

no shortcuts

Transient analysis and percolation on power-law random graphs



10000 nodes

GLP algorithm (Bu 2002)

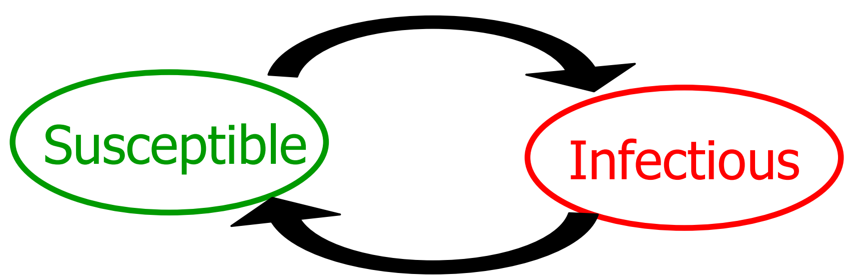
power-law node degree, small-world properties

One initially infected node with degree 10

m = initial connectivity of nodes added to the graph

Spectral methods

- The impact of network topology on the spread of epidemics can be predicted by looking at the eigenvalues and eigenvectors of the network adjacency matrix
- Example: SIS model



A. Ganesh, L. Massoulié, D. Towsley, "The Effect of Network Topology on the Spread of Epidemics," *IEEE Infocom 2005*

Combining transient analysis and percolation on general topologies

- A simple upper bound of the probability to receive a copy of the virus (valid on general topologies) :

$$P_{R_i}[\infty] = 1 - \prod_{j \in n_i} (1 - c_j)$$

- The probability not to receive a copy of the virus can be regarded as an initial immunization

$$P_{M_i}[0] = 1 - P_{R_i}[\infty]$$

→ Global upper bound for the infection process