University Ss. Cyril and Methodius
FACULTY OF ELECTIRCAL ENGINEERING
AND INFORMATION TECHNOLOGIES

Influence spreading in complex networks

Sasho Gramatikov, M.Sc.

Motivation

- **Power Blackout** – a short circuit caused by a tree near of a transmission line result in power outages for nearly 7.5 million customers along the entire western United States
- **Traffic Congestion** - An accident at a major street intersection during rush hour blocks the traffic flow through that intersection and soon causes an area of grid lock because the congestion spreads to nearby intersections.
- **Cold Spreading** – a person caught a cold and spreads it in a couple of days to a lot of other people
- **Product Popularity** – completely unknown product becomes enormously popular because a popular person uses it.

Motivation

- What is common in all the examples?
  - There are nodes
  - Nodes are connected in a network
  - Nodes have state
  - There is an interaction between nodes
  - Initial change of the state of a node causes spreading of a dynamic process in the network
- Questions
  - Is it better for a station to join the grid?
  - What is the probability that two intersections are blocked
  - What is the scale of spreading the cold
  - Is a product going to become popular?
- There is a necessity to find a model for representing dynamic processes in complex networks.

Complex Networks

- A **complex network** is a network with non-trivial topological features—features that do not occur in simple networks.
- Network topologies:
  - Random
  - Geographic random
  - Small-world
  - Scale-free

Weight calculation techniques

- In most of the dynamic models, the network structure is not enough for complete description of the network.
- It is required that each outgoing link is assigned a weight which will quantitatively define the node capability of spreading the dynamic process.
- The weight calculation techniques are based on network properties defined in graph theory.
- Main condition for calculating the weights: sum of incoming weights equals 1.
- Suggested techniques:
  - Equal weight
  - Degree
  - Node betweenness centrality
  - Edge betweenness centrality

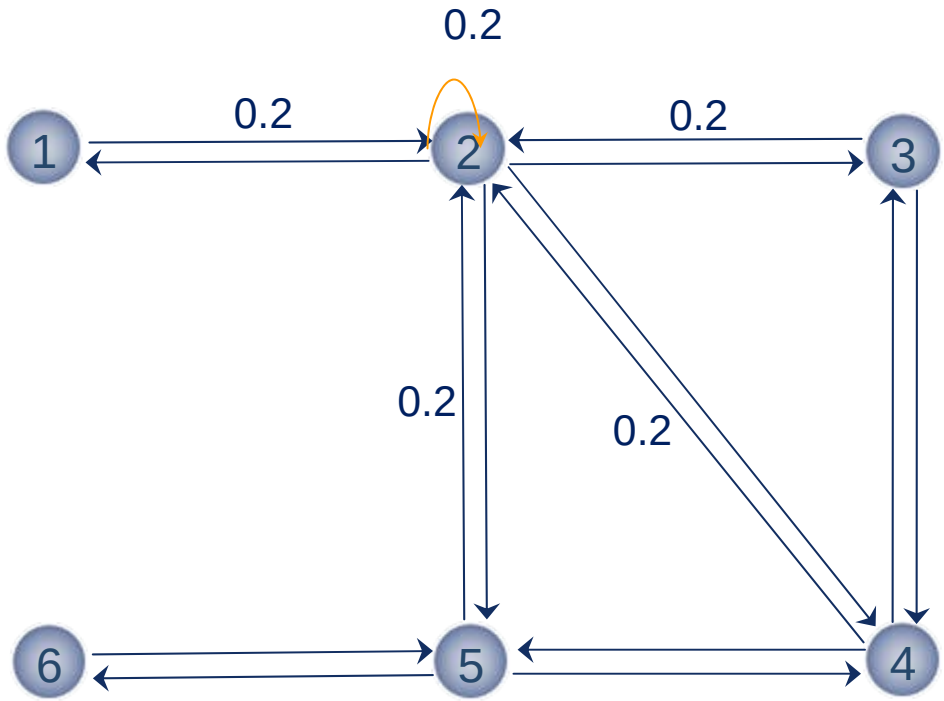
Equal weight distribution

- Assigns equal weights to every incoming edge.
- Every node has self-influence (self-loops)
- The weight of the links with destination to node *i*, originating from its neighbors (m) is:

$$d_{ij} = \frac{1}{m+1}$$

- Very simple technique
- Local character
- Does not give a complete picture of the network.

Equal weight distribution



$$d_{2j} = \frac{1}{4+1} = 0.2$$

Equal weight distribution

$$D = \begin{bmatrix} 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 1 & 0 & 0 \\ 1 & 1 & 0 & 1 & 1 & 0 \\ 0 & 1 & 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 & 0 & 1 \\ 0 & 0 & 0 & 0 & 1 & 0 \end{bmatrix} \Rightarrow \begin{bmatrix} 0.5 & 0.2 & 0 & 0 & 0 & 0 \\ 0 & 0.2 & 0.33 & 0.25 & 0 & 0 \\ 0.5 & 0.2 & 0.33 & 0.25 & 0.25 & 0 \\ 0 & 0.2 & 0.33 & 0.25 & 0.25 & 0 \\ 0 & 0.2 & 0 & 0.25 & 0.25 & 0.5 \\ 0 & 0 & 0 & 0 & 0.25 & 0.5 \end{bmatrix}$$

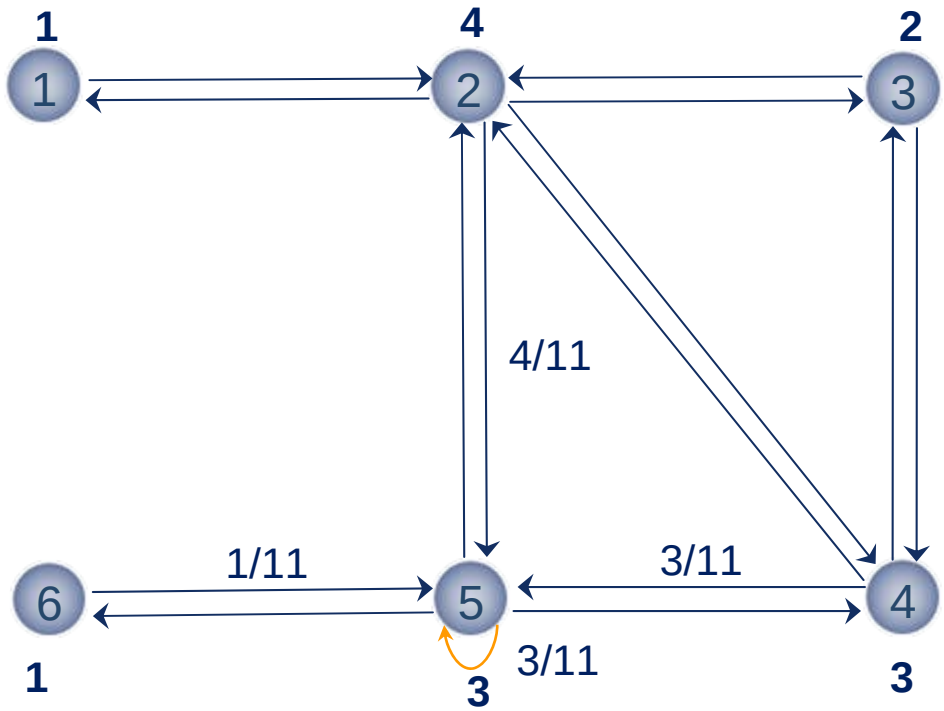
Node degree distribution

- Improved technique
- Assigns weights to incoming links proportional to the degree of the originating nodes.
- Weight of links directed to node *i* are proportional to the degrees of its neighbors:

$$d_{ji} = \frac{D(j)}{D(i) + \sum_{k \in P_i} D(k)}$$

- Local character

Node degree distribution



$$d_{25} = \frac{D(2)}{D(5) + \sum_{k \in P_5} D(k)} = \frac{4}{3+8} = \frac{4}{11}$$

Node betweenness centrality distribution

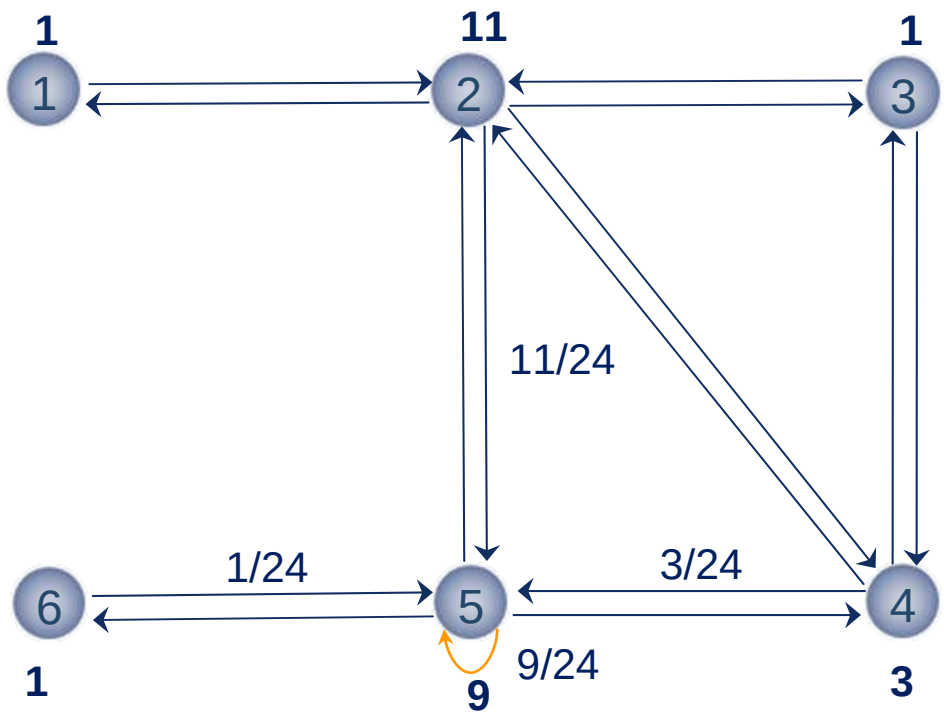
- Node betweenness centrality is a fraction of the shortest paths between any pair of nodes that pass through a certain node:

$$C(v) = \sum_{s \neq v \neq t \in V} \frac{\sigma_{st}(v)}{\sigma_{st}}$$

- It is measure of the importance of a node in a network.
- Node betweenness centrality has a global character – every part of the network is included in calculation of its value.
- Weights of the incoming links of a node are proportional to the importance of the originating nodes:

$$d_{ji} = \frac{C(j)}{C(i) + \sum_{k \in P_i} C(k)}$$

Node betweenness centrality distribution



$$d_{25} = \frac{C(2)}{C(5) + \sum_{k \in P_5} C(k)} = \frac{11}{9+15} = \frac{11}{24}$$

Edge betweenness centrality distribution

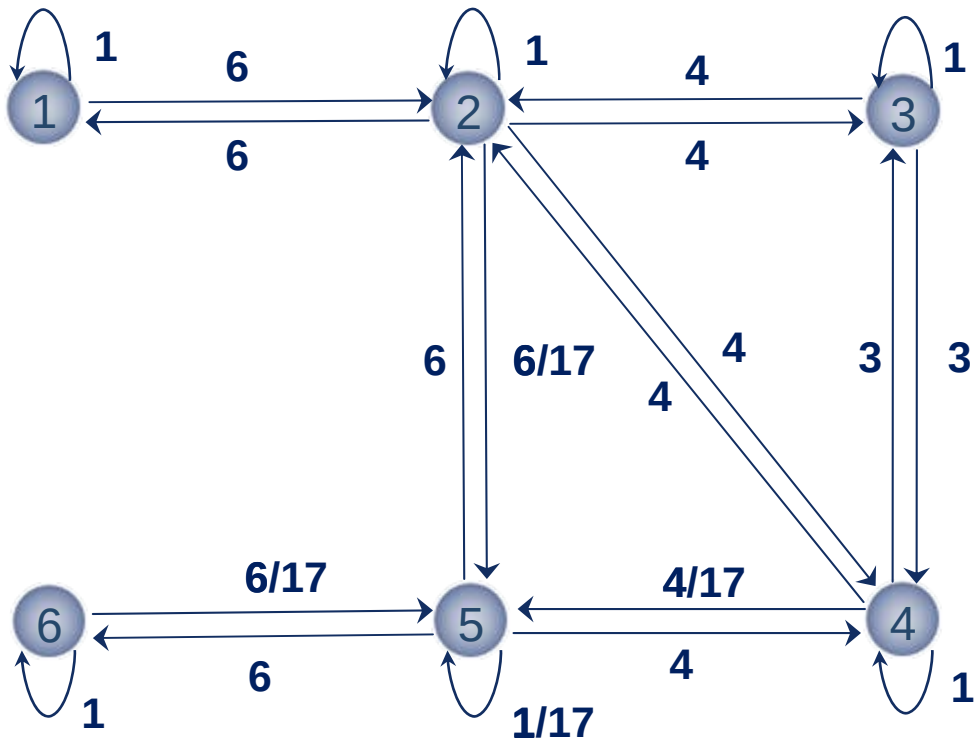
- Edge betweenness centrality is fraction of the shortest paths between all pairs of nodes that lays on a certain edge.

$$C(e) = \sum_{s \neq t \in V} \frac{\delta_{st}(e)}{\delta_{st}}$$

- Measure of importance of edges in a network.
- Global character.
- Weights are easily calculated by normalization of incoming links, so that their sum equals 1.

$$d_e = \frac{C(e)}{\sum_{k \in Q_j} C(k)}$$

Edge betweenness centrality distribution

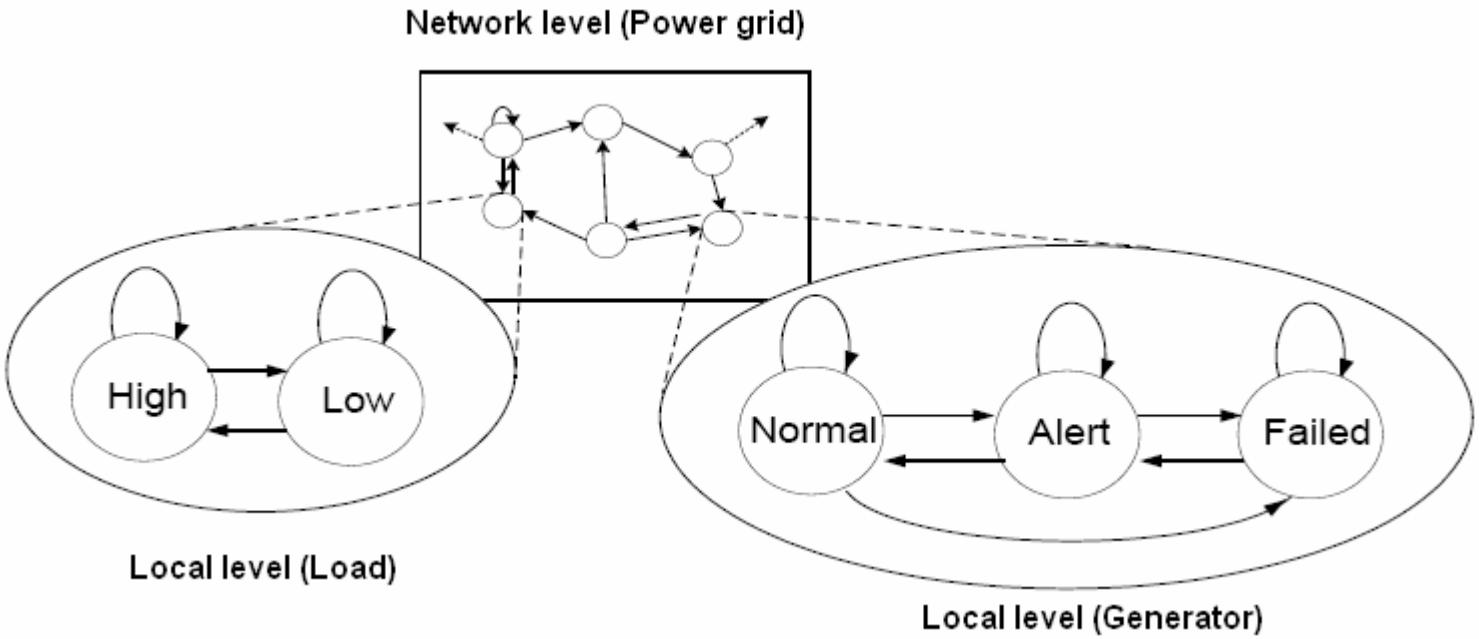


$$d_{25} = \frac{C(e_{25})}{\sum_{k \in Q_5} C(k)} = \frac{6}{17}$$

General Influence Model

- Stochastic dynamic model which treats the network on two levels:
  - network level – each node is treated as an active entity called site.
  - local level – Markov chain representing the possible states of each site.
- With the network connections, the transition probabilities of each local chain are likely to depend not only on the current status of its site, but also on those of its neighboring sites.
- The general influence model is a network of interacting Markov chains.
- Depending on the structure of local level of each site, there are two types of general influence model:
  - homogeneous – all nodes have equal local structure
  - heterogeneous – nodes have different structure of local Markov chain

General Influence Model



- A generator which is currently on alert can have a high probability of moving to failed status if it is surrounded by high loads and failed generators.
- Generator influenced by low loads and normal generators would have a high probability of reverting to normal status.

Homogeneous General Influence Model

- D_{nxn} - stochastic network matrix ($d_{ij} \neq 0$ if link from i to j exists)
- A_{mxm} - transition matrix, defining the local Markov chain
- $s_i^{'}[k] = [0 \cdots 1 0 \cdots 0]$ - status vector of site *i* at time *k*.
- $p_i[k]$ - probability of *i* being in a certain status at *k*
- State and probability vectors of the network

$$s[k] = \begin{bmatrix} s_1[k] \\ \vdots \\ s_n[k] \end{bmatrix} \qquad p[k] = \begin{bmatrix} p_1[k] \\ \vdots \\ p_n[k] \end{bmatrix}$$

- Definition of Network influence matrix (Kronecker product)

$$H = D^{'} \otimes A = \begin{pmatrix} d_{11}A & \dots & d_{n1}A \\ \vdots & \ddots & \vdots \\ d_{1n}A & \cdots & d_{nn}A \end{pmatrix}$$

Homogeneous General Influence Model

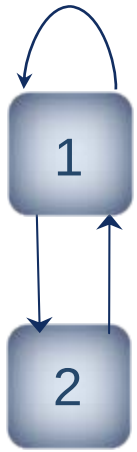
- Model definition:

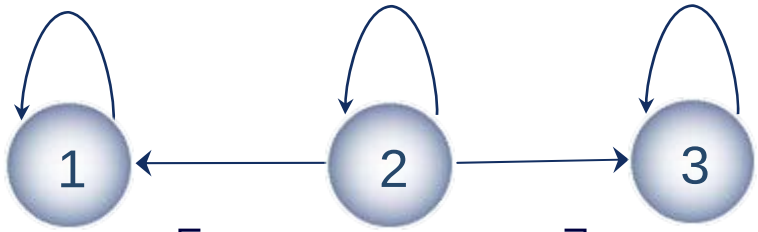
$$p'[k+1] = s'[k]H$$

$$s'[k+1] = \text{MultiRealize}(p'[k+1])$$

- MultiRealize treats each block of PMF's within p'[k + 1] separately, independently realizing the new status vectors block by block.
- MultiRealize is equivalent to *n* independent throwing of a m-sided dice. Each outcome determines the status of a node.
- Influence matrix – similar properties to a stochastic matrix
  - Nonnegative
  - Eigen values are positive
  - The dominant (maximum) eigen value is 1

Homogeneous General Influence Model



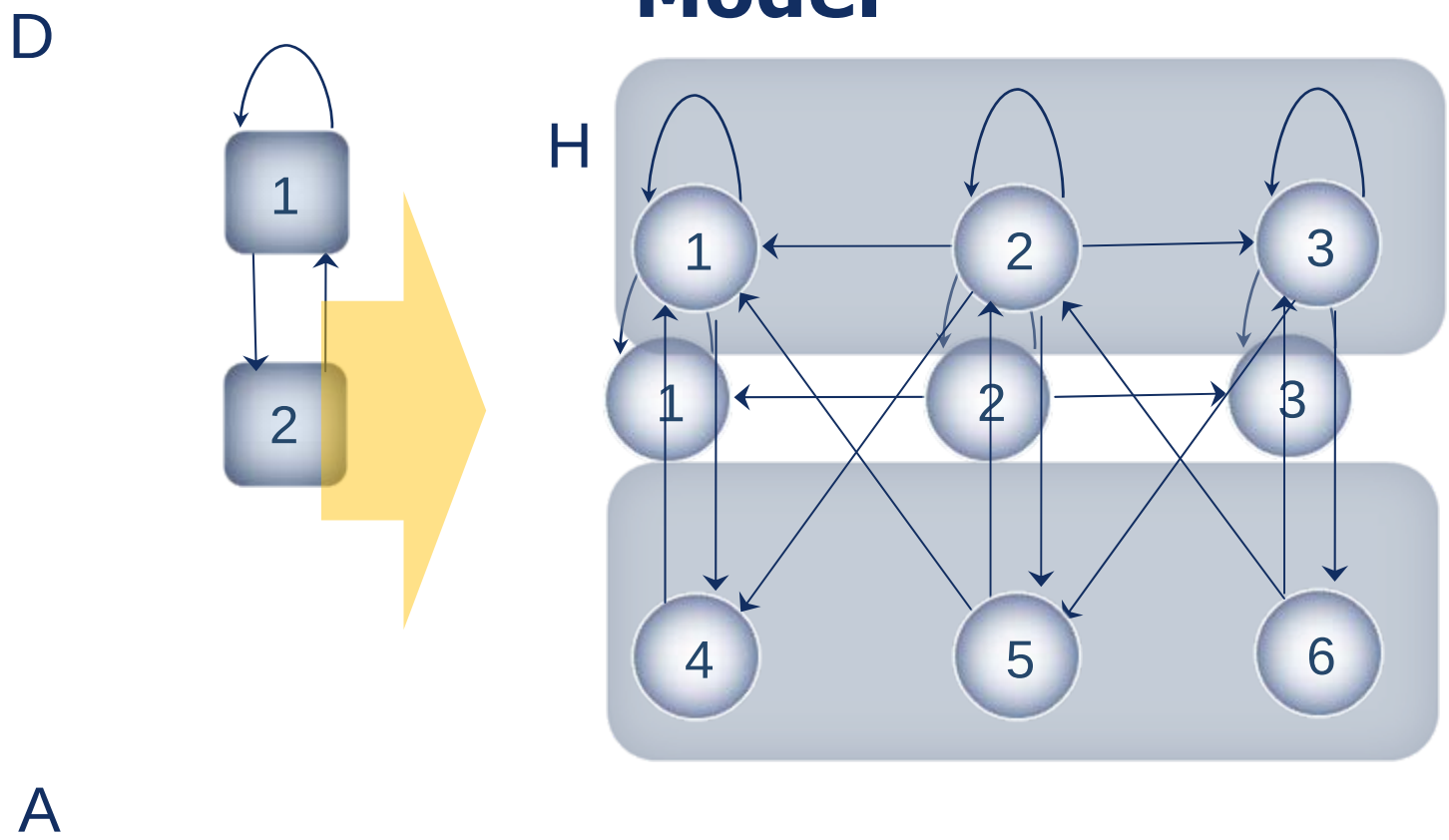


$$D' = \begin{bmatrix} 0.5 & 1 \\ 0.5 & 0 \end{bmatrix}$$

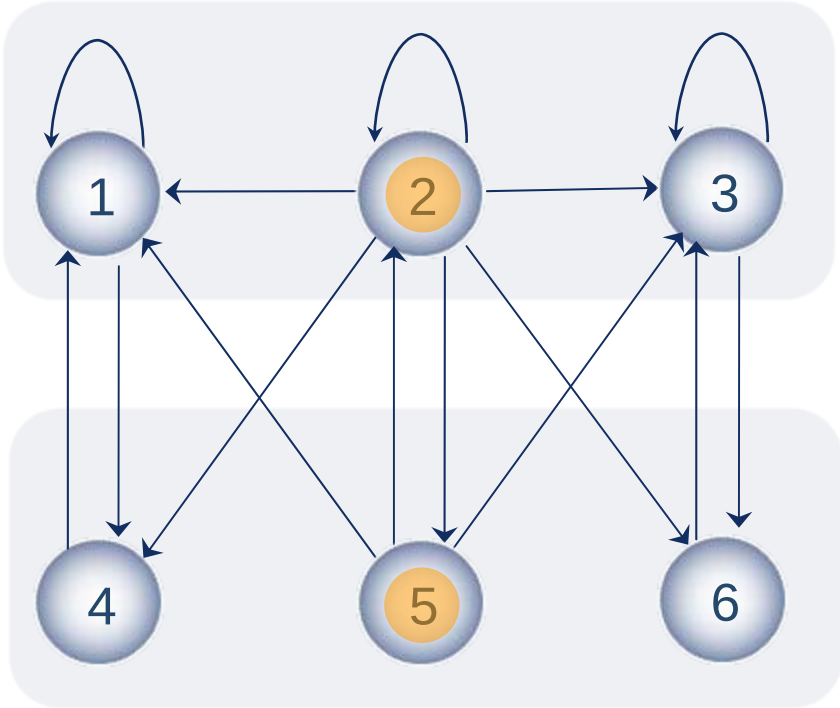
$$A = \begin{bmatrix} 1 & 0 & 0 \\ 0.4 & 0.2 & 0.4 \\ 0 & 0 & 1 \end{bmatrix}$$

$$H = D' \otimes A = \begin{bmatrix} 0.5 & 1 & 0 & 1 & 0 & 0 \\ 0.2 & 0.1 & 0.2 & 0.4 & 0.2 & 0.4 \\ 0 & 0 & 0.5 & 0 & 0 & 1 \\ 0.5 & 0 & 0 & 0 & 0 & 0 \\ 0.2 & 0.1 & 0.2 & 0 & 0 & 0 \\ 0 & 0 & 0.5 & 0 & 0 & 0 \end{bmatrix}$$

Homogeneous General Influence Model



Homogeneous General Influence Model



Homogeneous General Influence Model

- What is the expected values of the state in a certain time period?
- What is the stationary state of the system?

$$E\left(s'[k]\right) = E\left(s'[0]\right)H^k$$

$$E(s'[k]) = E(s'[0])v_0 w_0^{'} + \sum_{i=1}^{v_n - 1} \lambda_i^k E(s'[0])v_i w_i^{'}$$

$$\lim_{k \to \infty} E(s'[k]) = E(s'[0])v_0 w_0^{'} = w_0^{'}$$

$$E(s'[k]) \to E(s'[0])v_0 w_0^{'} = w_0^{'} = (1_m \otimes \alpha)^{'} = [\alpha^{'} \cdots \alpha^{'}]$$

- Stationary state does not depend on network structure, but only on local Markov chain structure.

Homogeneous General Influence Model

- Analysis of statistical parameters:
  - Mean value of number of nodes in certain state
  - Standard deviation of the number of nodes in certain state.
- Homogenwous model with two-state Markov chain:

$$A = \begin{bmatrix} 1-p & p \\ q & 1-q \end{bmatrix}$$

- Stationary state of Markov chain:

$$\alpha = [\alpha_1 \quad \alpha_2] \qquad \alpha_1 = \frac{p}{p+q} \qquad \alpha_2 = \frac{q}{p+q}$$

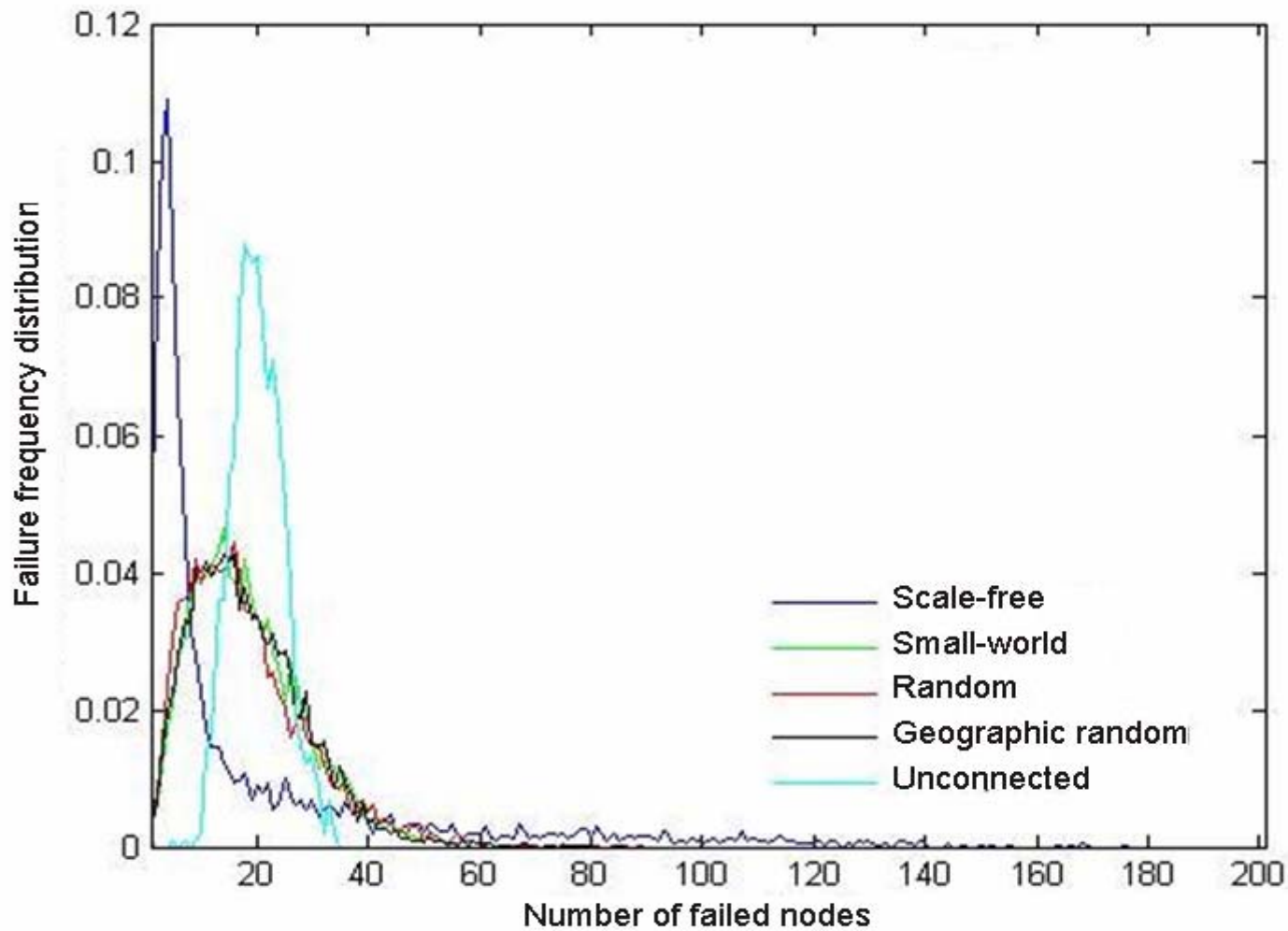
- Mean value of failed nodes:

$$\bar{N} = N \frac{p}{p+q}$$

Homogeneous General Influence Model

N=200
t=1000
i=10

$$A = \begin{bmatrix} 0.99 & 0.01 \\ 0.1 & 0.9 \end{bmatrix}$$

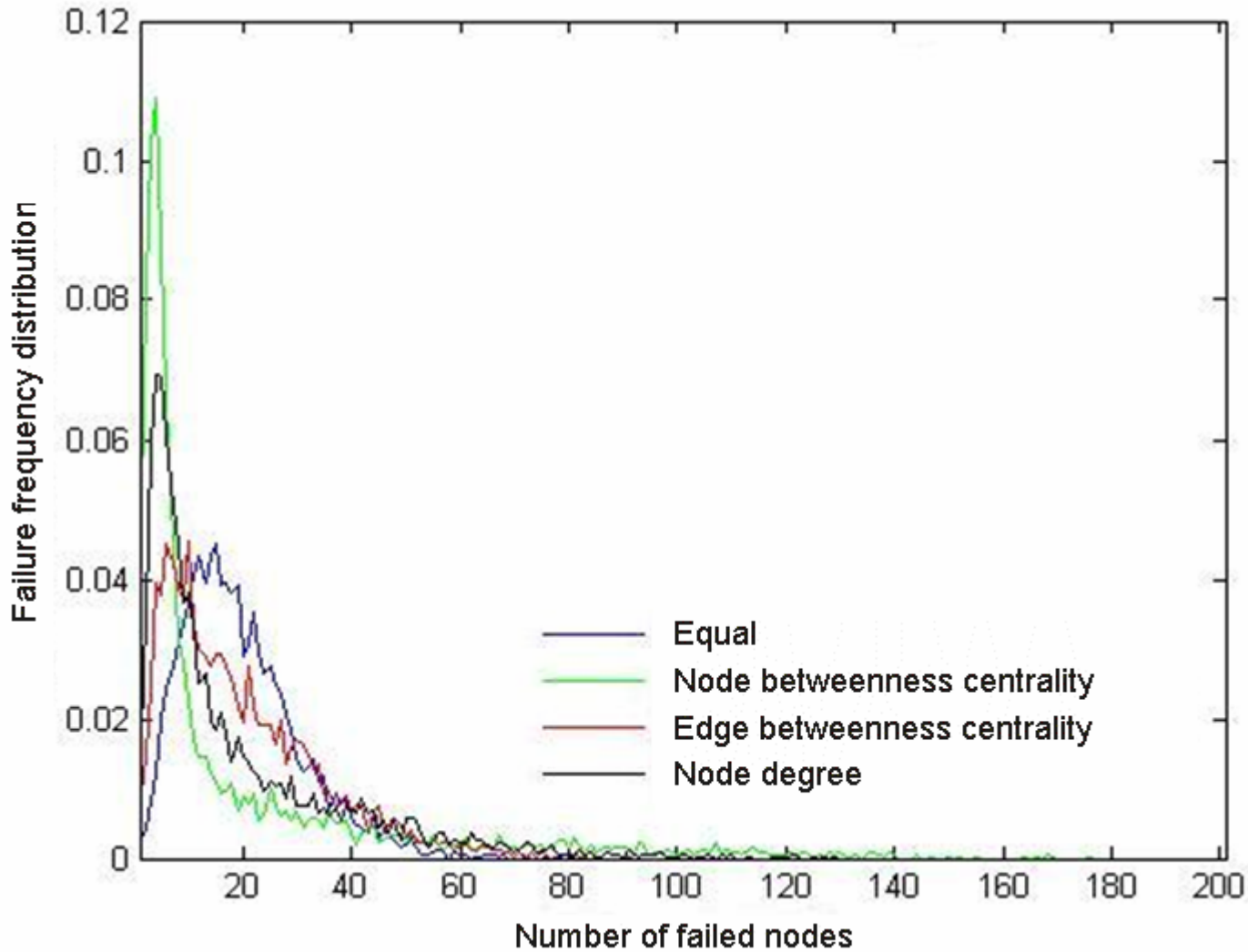


Failure frequency distribution for different complex networks with weights assigned using the **node betweenness centrality** technique

Homogeneous General Influence Model

N=200
t=1000
i=10

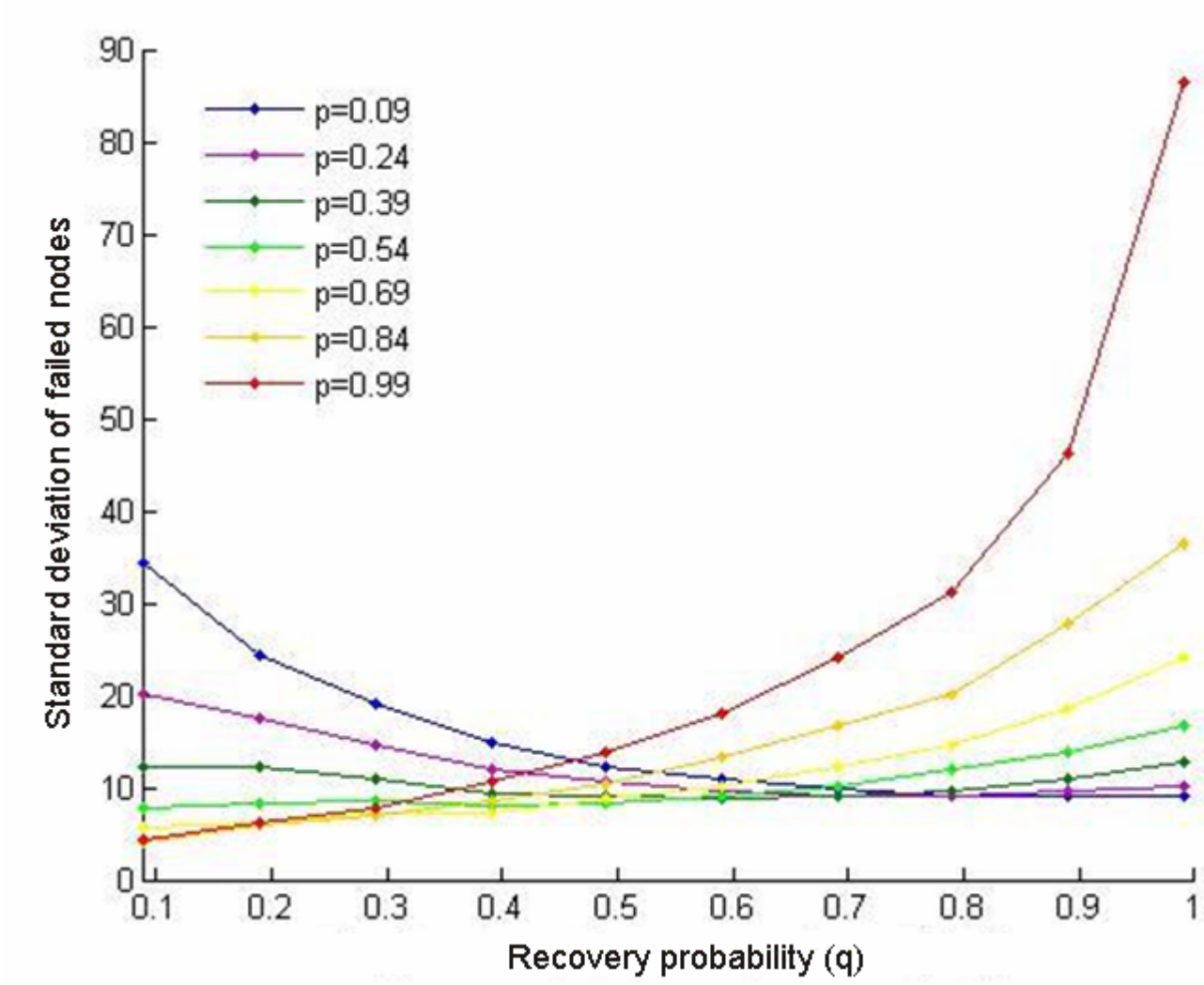
$$A = \begin{bmatrix} 0.99 & 0.01 \\ 0.1 & 0.9 \end{bmatrix}$$



Failure frequency distribution for weight assignment techniques for **scale-free** networks

Homogeneous General Influence Model

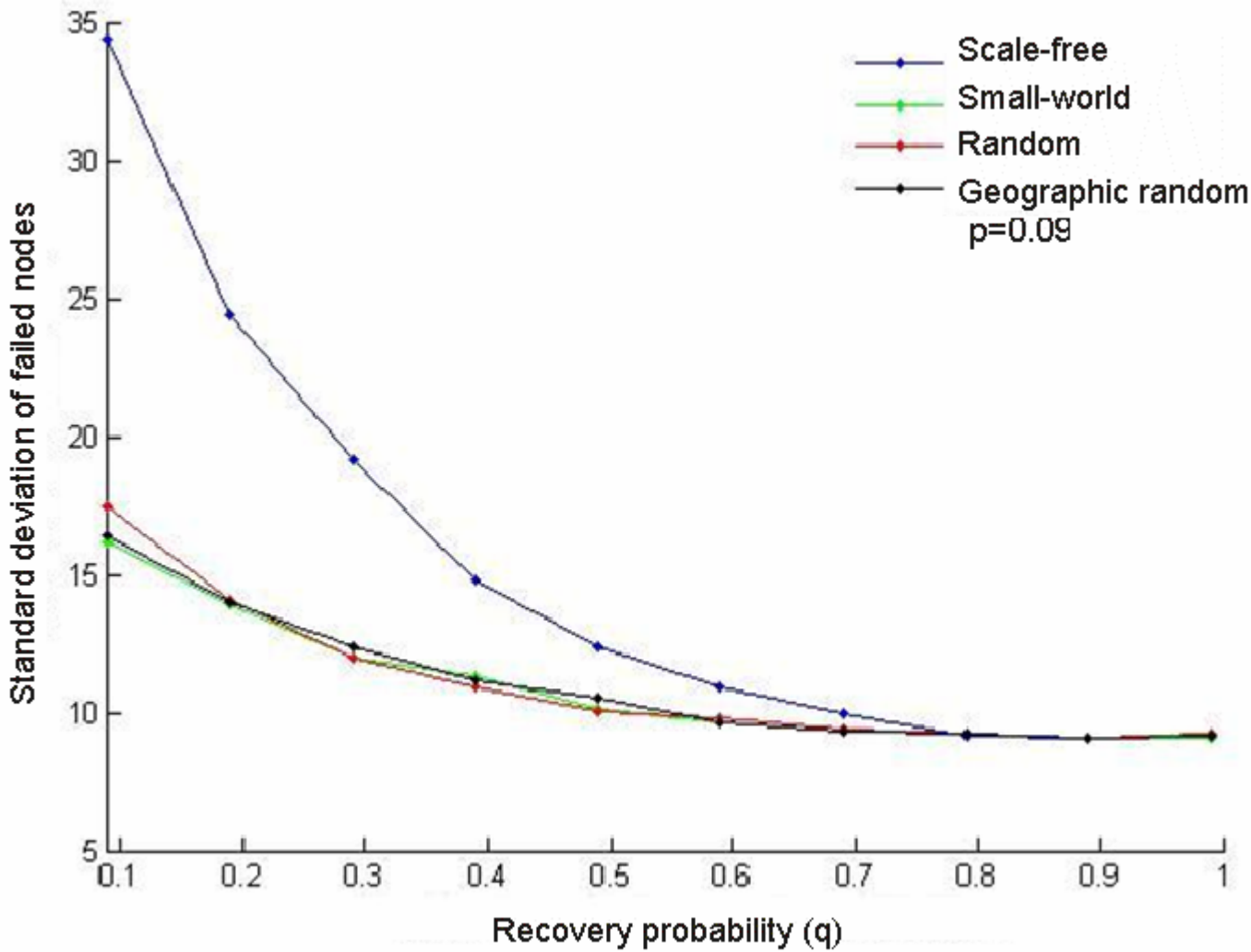
N=200
t=1000
i=20



Dependence of standard deviation on structure of local Markov chain for **scale-free** networks

Хомоген генерален модел на влијание

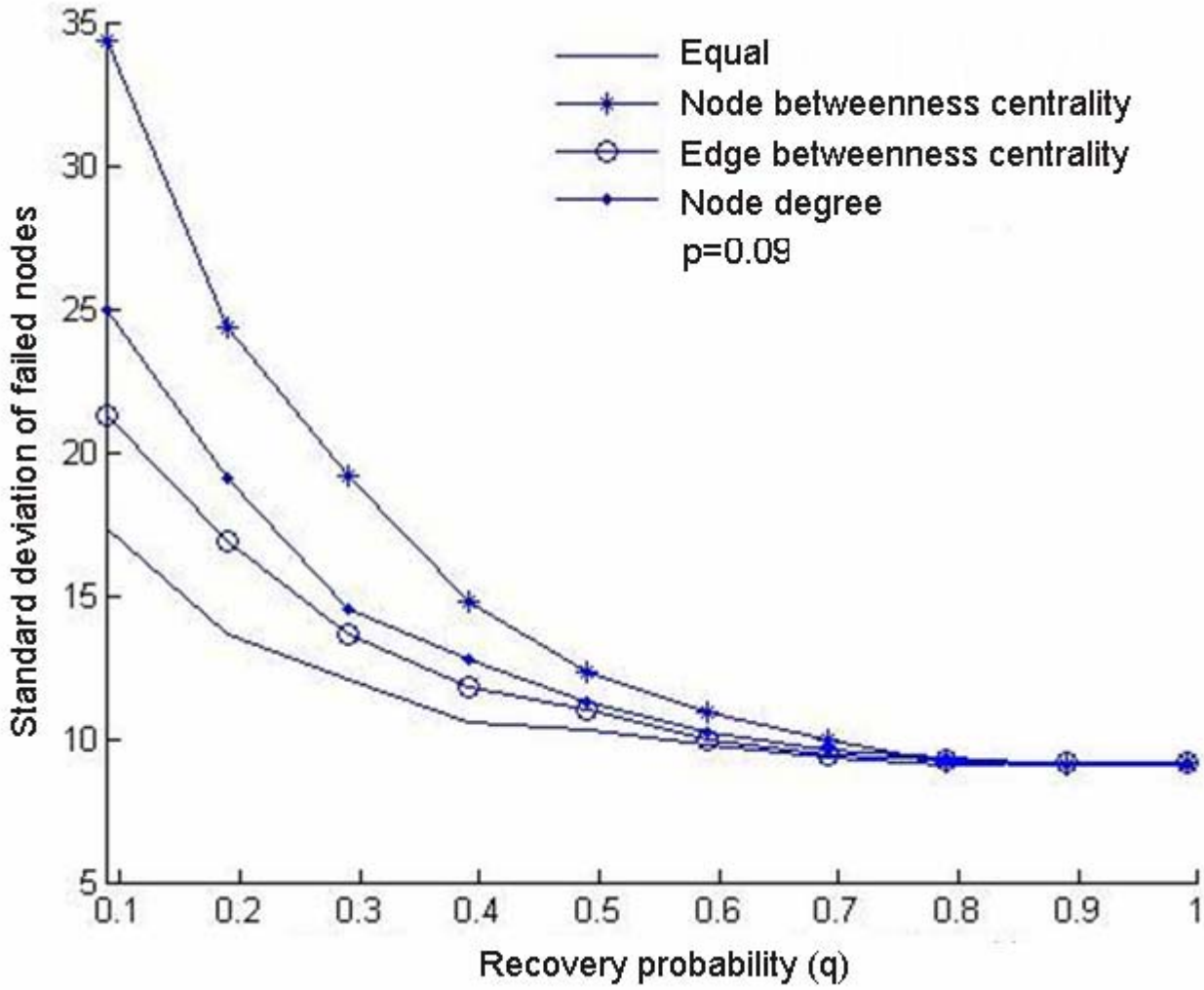
N=200
t=1000
i=20
p=0.09



Dependence of standard deviation on recovery probability q for different network topologies for **equal weight** distribution technique

Хомоген генерален модел на влијание

N=200
t=1000
i=20
p=0.09



Dependence of standard deviation on recovery probability q for different weight distribution techniques for **scale-free** networks

Heterogeneous General Influence Model

- Gives more real picture of processes in real networks
- The Markov chain of each node has a different structure
- Influence matrix:

$$H = D' \otimes \{A_{ij}\} = \begin{pmatrix} d_{11}A_{11} & \cdots & d_{n1}A_{1n} \\ \vdots & \ddots & \vdots \\ d_{1n}A_{n1} & \cdots & d_{nn}A_{nn} \end{pmatrix}$$

- Introducing heterogeneity
  - Nodes have same structure of Markov Chain, but different transition probabilities
  - example with the internet network – more important nodes (routers) are better protected against failure

Heterogeneous General Influence Model

- Two state Markov Chain

$$A_i = \begin{bmatrix} 1-p_i & p_i \\ q_i & 1-q_i \end{bmatrix}$$

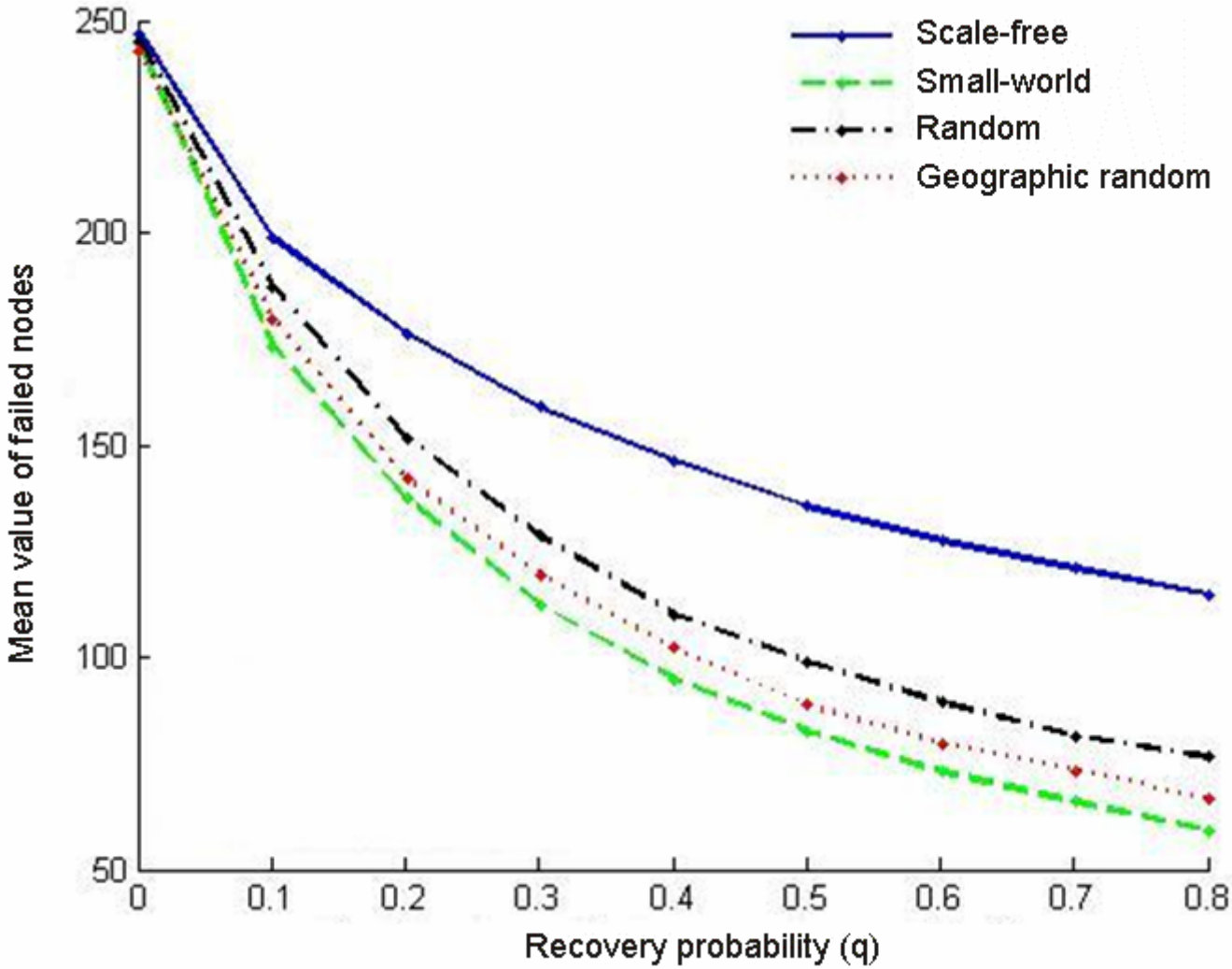
- Importance defined by node degree
- Failure probability is proportional to node importance
- Minimum and maximum values are pre set

$$p_i = p_{\max} - \frac{p_{\max} - p_{\min}}{D_{\max} - D_{\min}} (D(i) - D_{\min})$$

- Recovery probability is the same for each node

Хетероген генерален модел на влијание

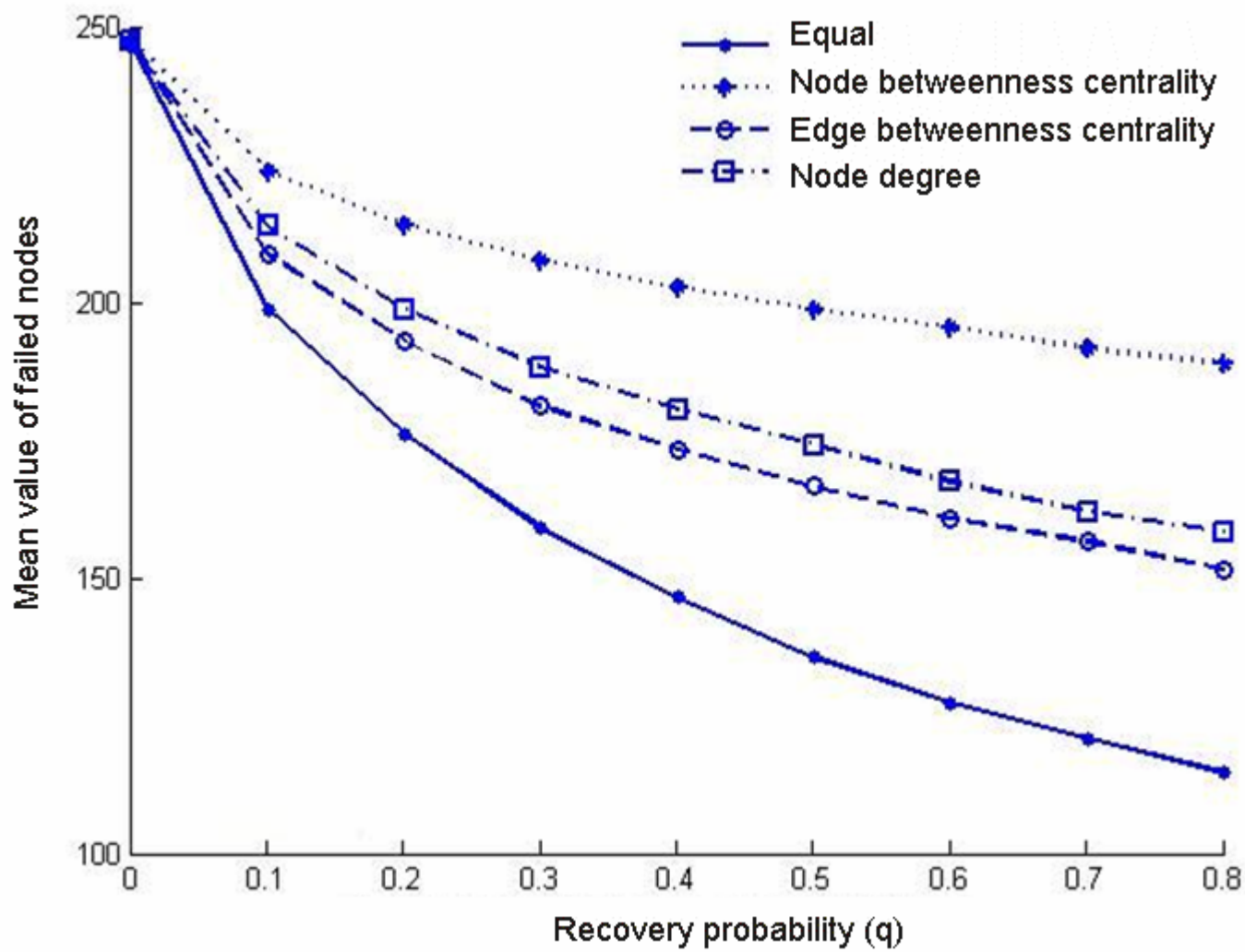
N=250
t=1000
i=10
p_{min}=0.01
p_{max}=0.41



Dependence of mean value of failed nodes on recovery probability q for different network topologies and **equal weight** distribution technique

Хетероген генерален модел на влијание

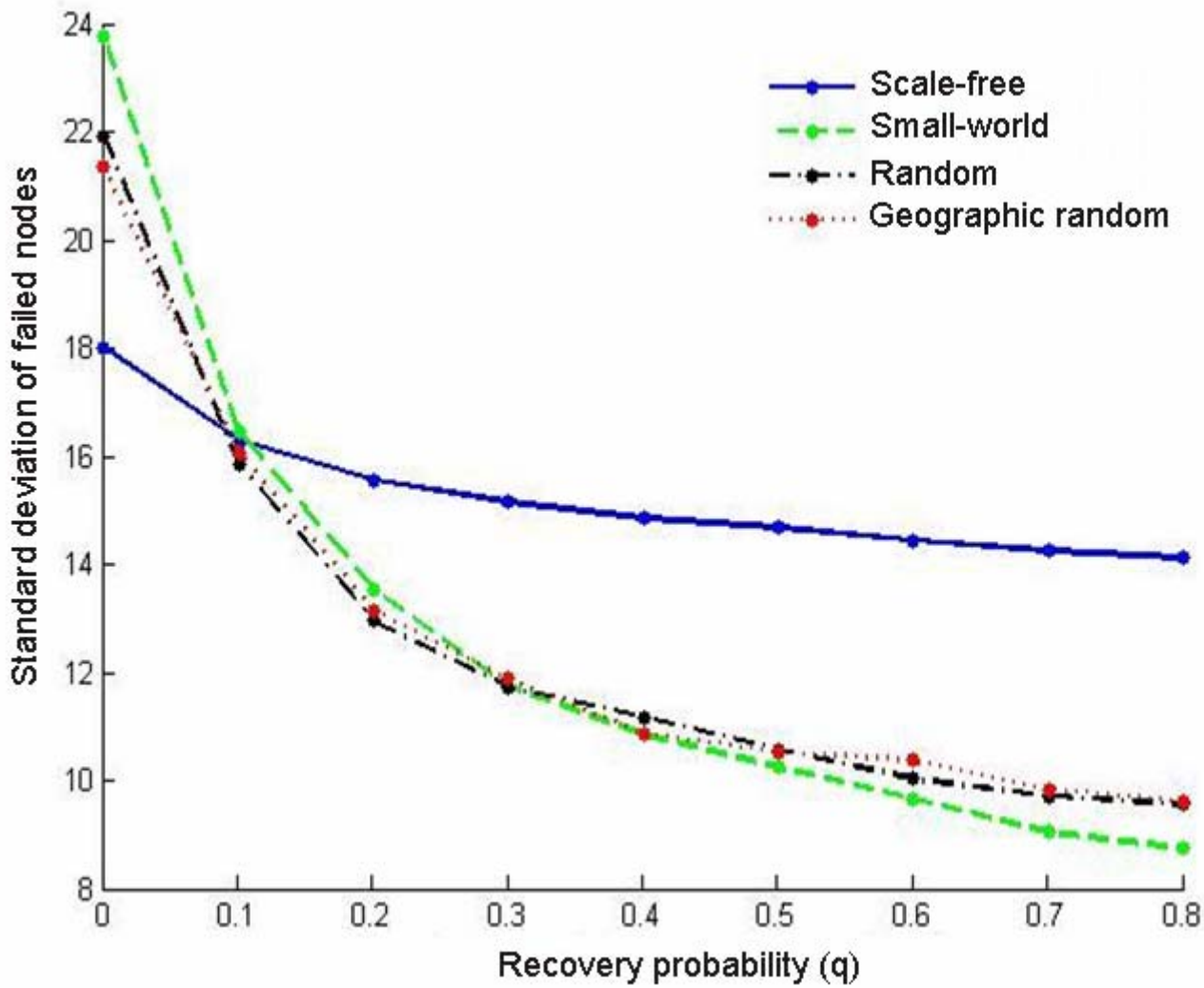
N=250
t=1000
i=10
p_{min}=0.01
p_{max}=0.41



Dependence of mean value of failed nodes on recovery probability q for different weight distribution techniques for **scale-free** networks

Хетероген генерален модел на влијание

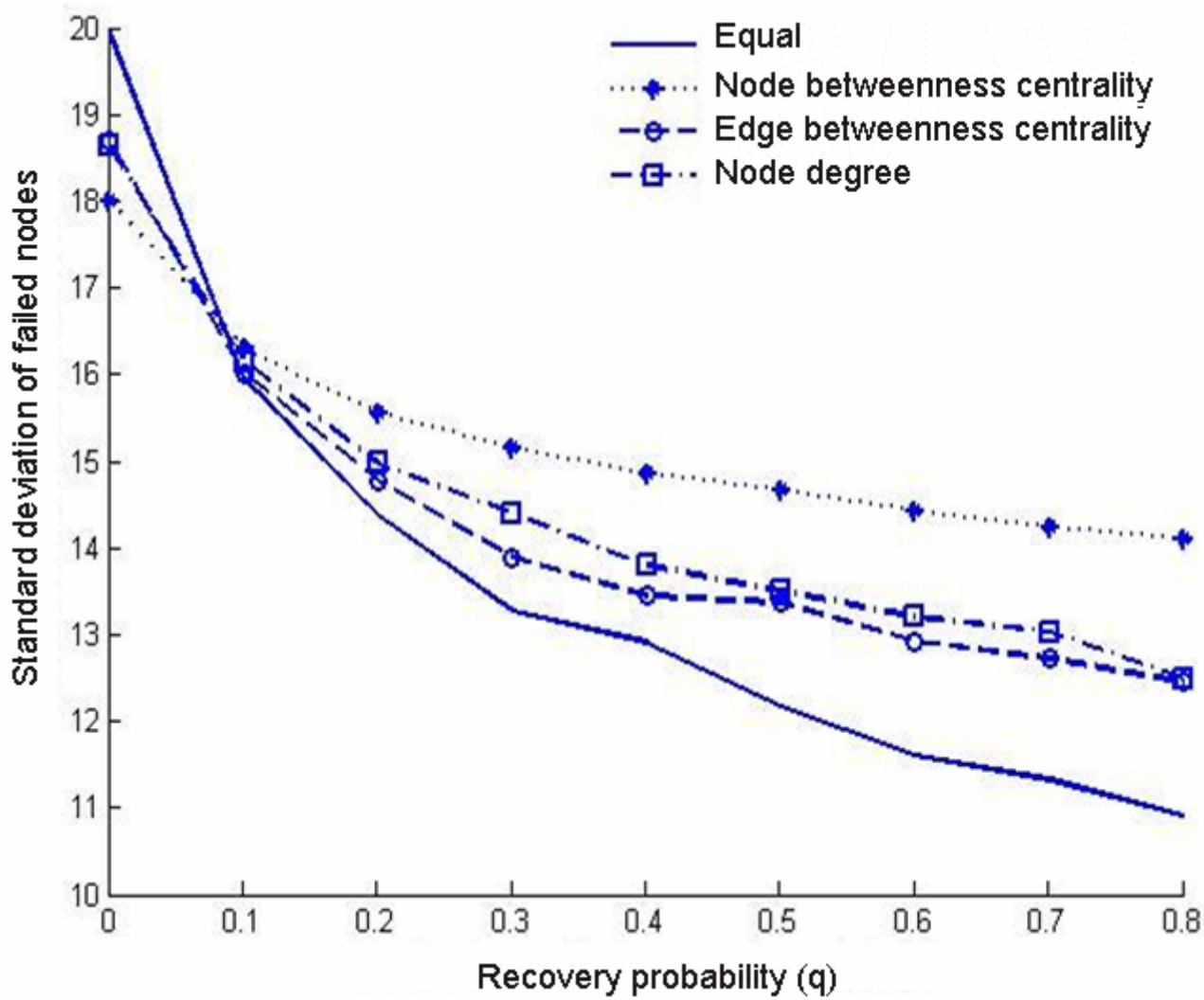
N=250
t=1000
i=10
p_{min}=0.01
p_{max}=0.41



Dependence of standard deviation on recovery probability q for different network topologies for **node betweenness centrality** distribution technique

Хетероген генерален модел на влијание

N=250
t=1000
i=10
p_{min}=0.01
p_{max}=0.41



Dependence of standard deviation on recovery probability q for different weight distribution techniques for **scale-free** networks

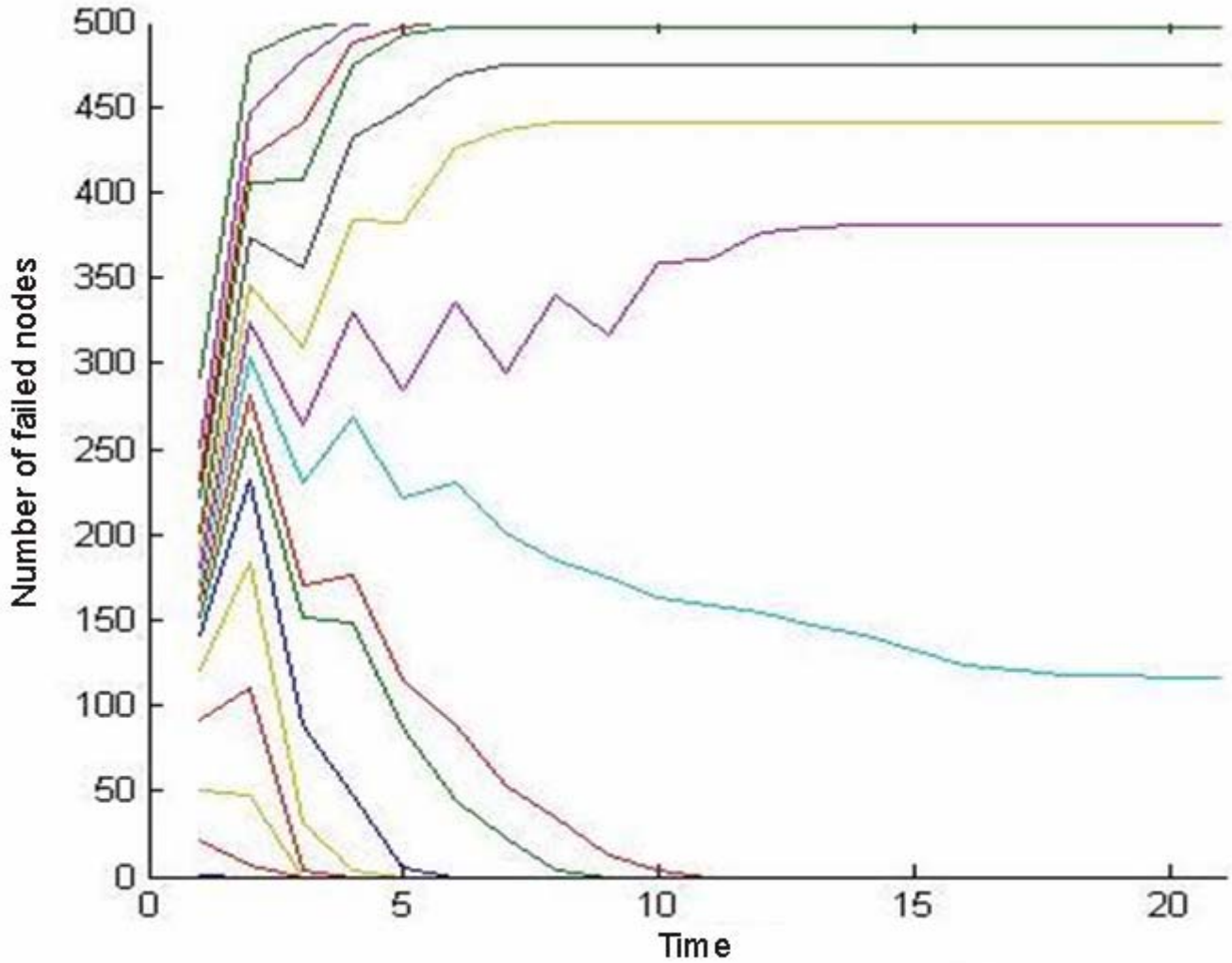
Threshold model

- Deterministic model of influence spreading where the new state is determined according to the quantity of received influence compared to given threshold.

$$s_i[k+1] = \begin{cases} 0 & \sum_{j=1}^{n} d_{ij} s_j[k] \leq \Theta(i) \\ 1 & \sum_{j=1}^{n} d_{ij} s_j[k] > \Theta(i) \end{cases}$$

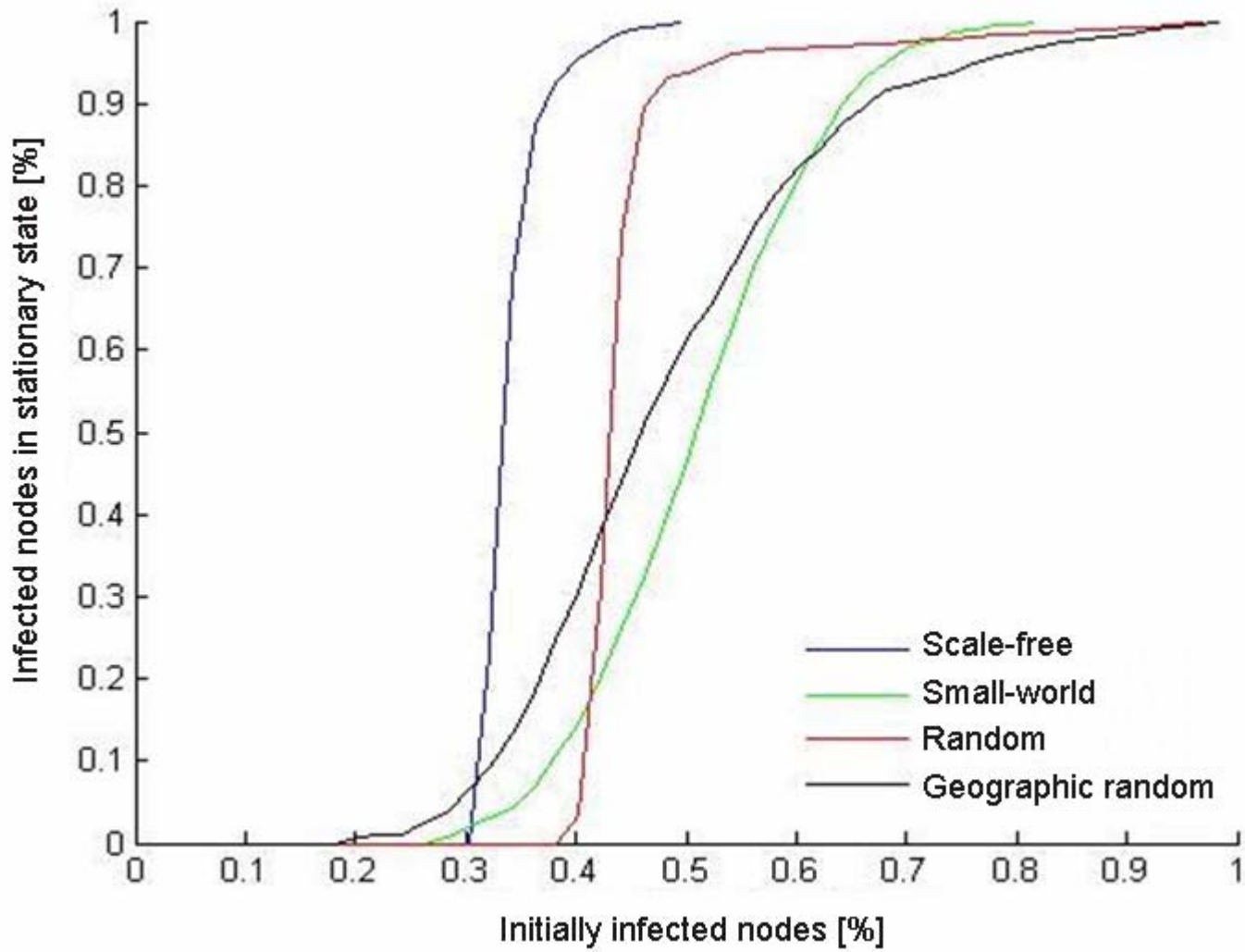
- $\Theta(i)$ is threshold funcition
  - constant value
  - value determined as a function of some property from graph theory.
  - random value

Threshold model



Time dependence of failed nodes on different initial conditions

Threshold model



Dependence of stationary state on initially infected nodes (threshold=0.5)

Epidemiologic Influence Modles

- Initially used for modeling the process of spreading natural viruses.
- Suitable representation for modeling spreading of
  - computer viruses,
  - viral marketing (product promotion),
  - social processes
- Types of epidemiologic models:
  - SIR (Susceptible - Infected - Recovered),
  - SI (Susceptible - Infected)
  - SIS (Susceptible - Infected - Susceptible)

SIR

- Model definition:

$$p_i^S(k+1) = 1 - p_i^I(k+1) - p_i^R(k+1)$$

$$p_i^I(k+1) = s_i^S(k)\left[1 - \prod_{j=1}^{n}(1 - \beta d_{ij} s_i^I(k))\right] + (1-\alpha)s_i^I(k)$$

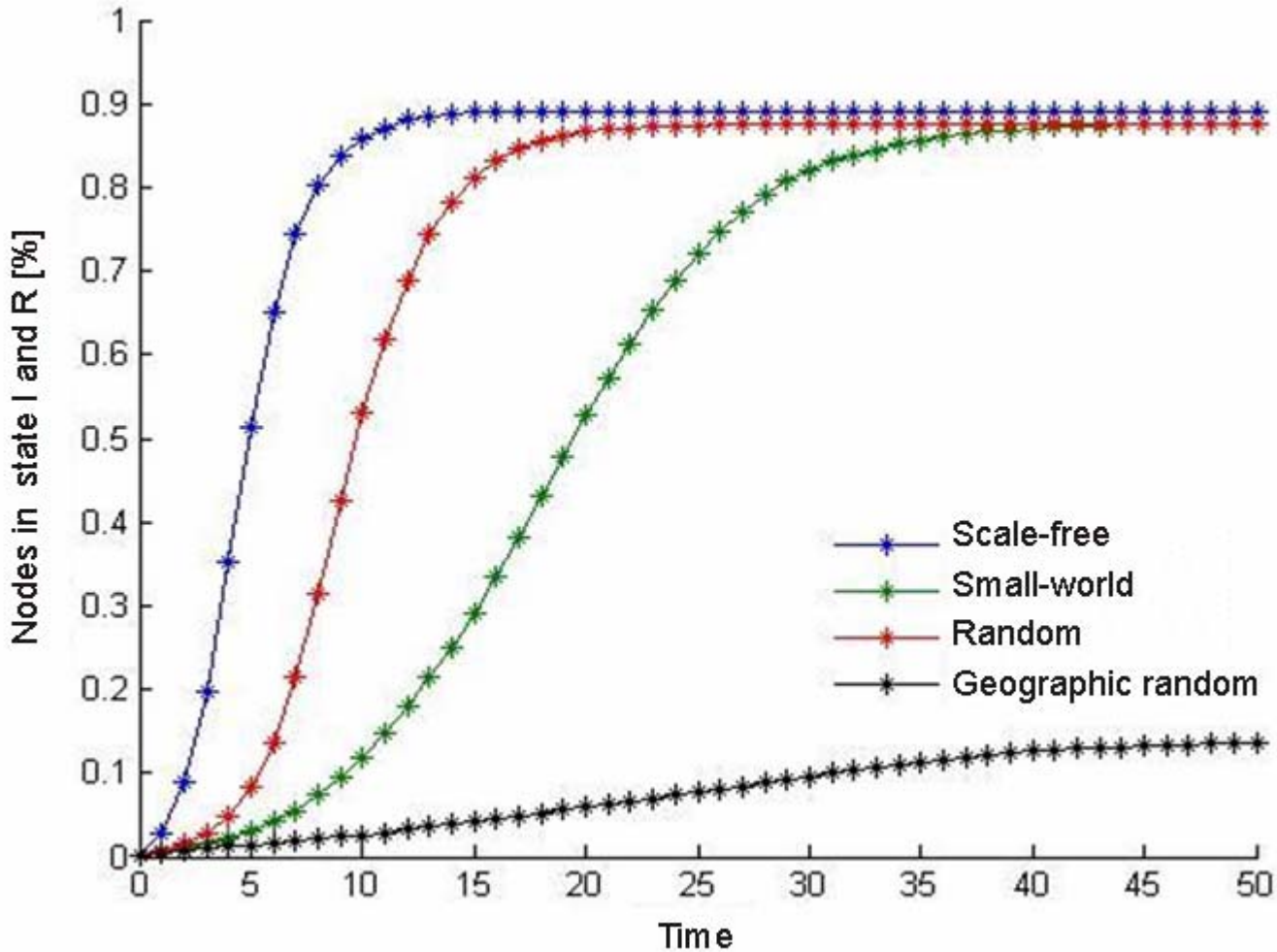
$$p_i^R(k+1) = \alpha s_i^I(k) + s_i^R(k)$$

- $S_i(k) = [s_i^S(k)\ s_i^I(k)\ s_i^R(k)]$ - state vector of node i
- $P_i(k) = [p_i^S(k)\ p_i^I(k)\ p_i^R(k)]$ - probability vector of node i

- β infection probability
- α recovery probability
- d_{ij} connectivity between nodes *i* and *j* (0 or 1)

SIR

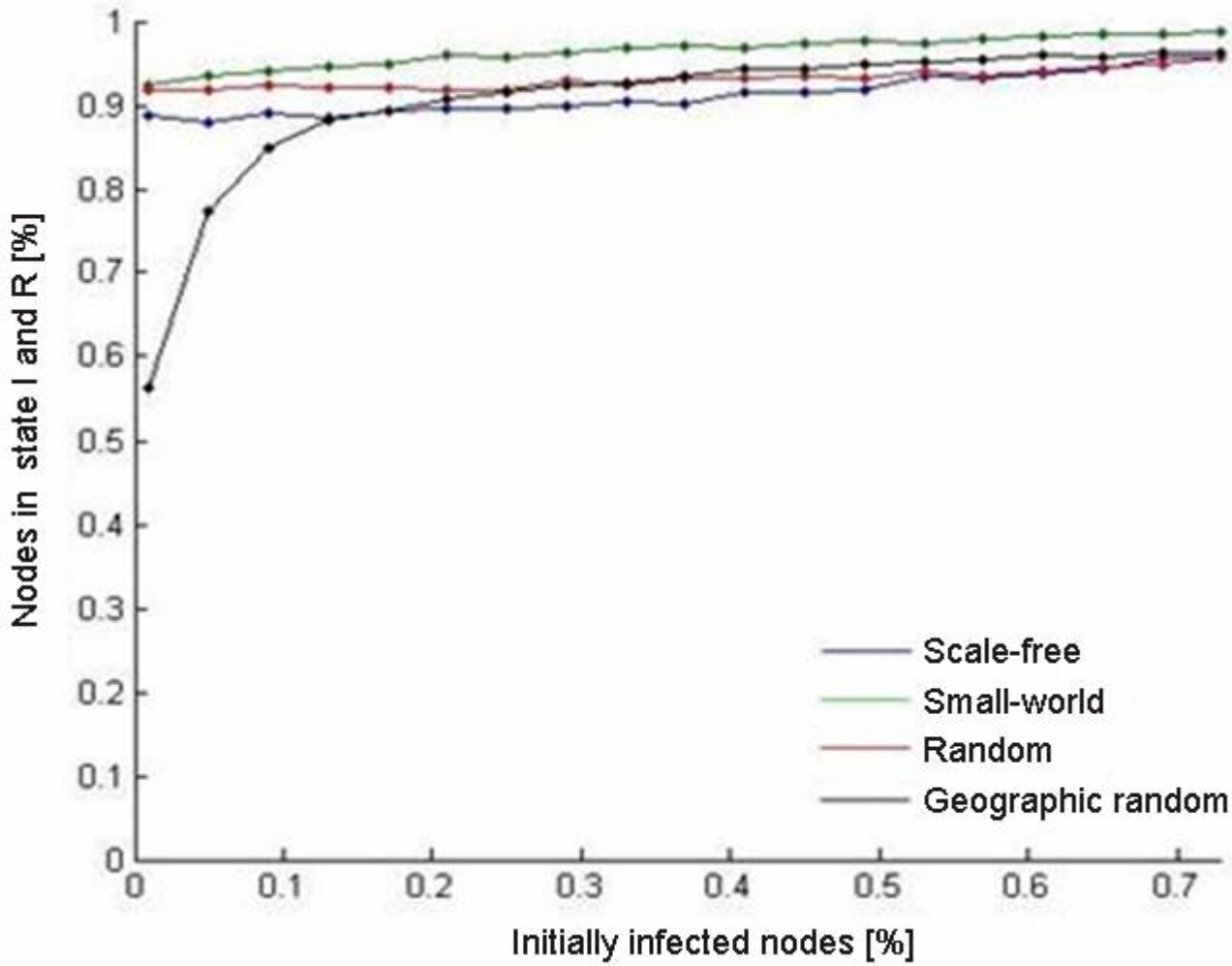
N=500
i=20
α=0.2
β=0.3



Time dependence of number of infected nodes

SIR

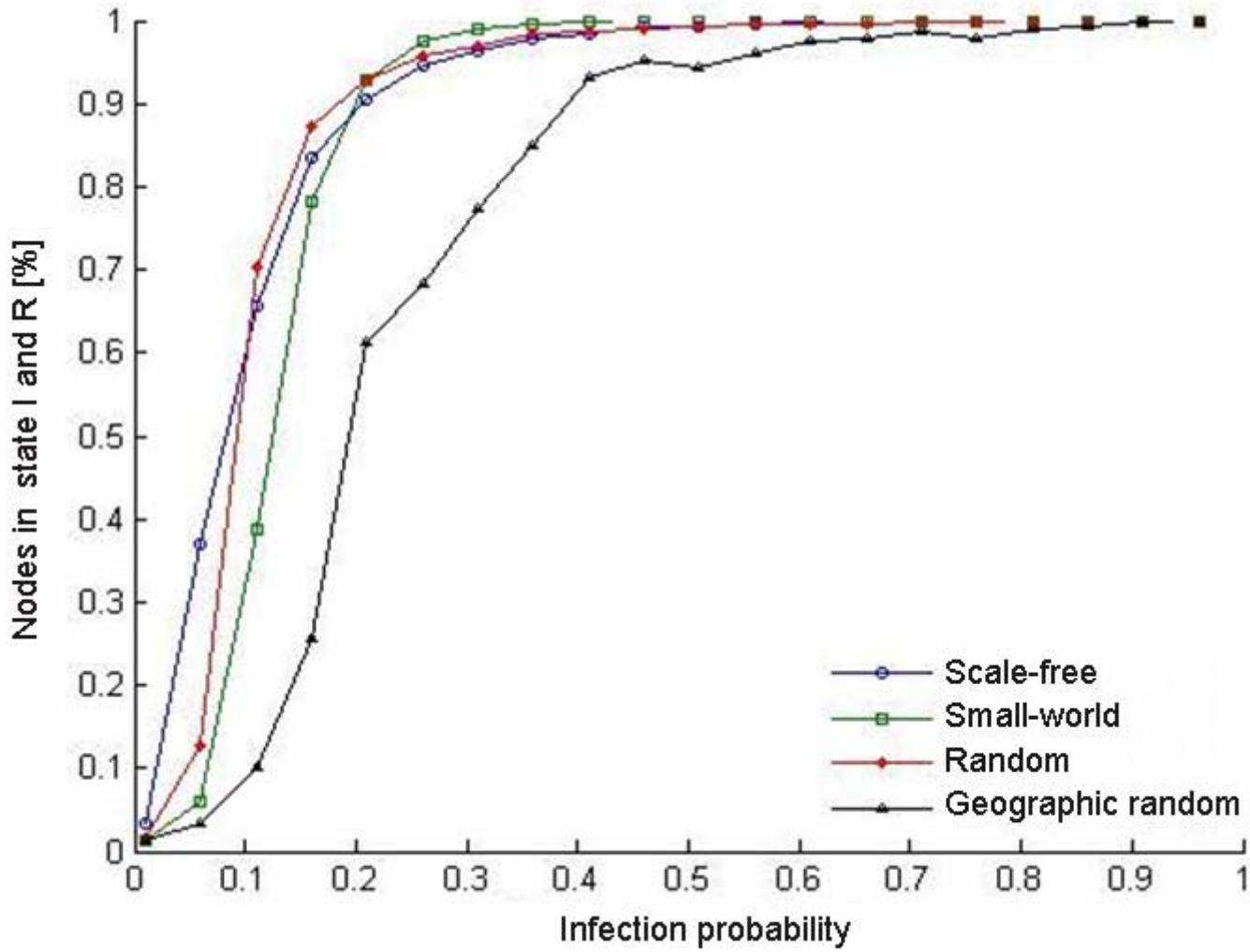
N=500
i=20
α=0.2
β=0.3



Dependence of number of infected nodes in stationary state on number of initially infected nodes

SIR

N=500
i=20
α=0.2



Dependence of infected nodes in stationary state on infection probability

SIS

- Model definition:

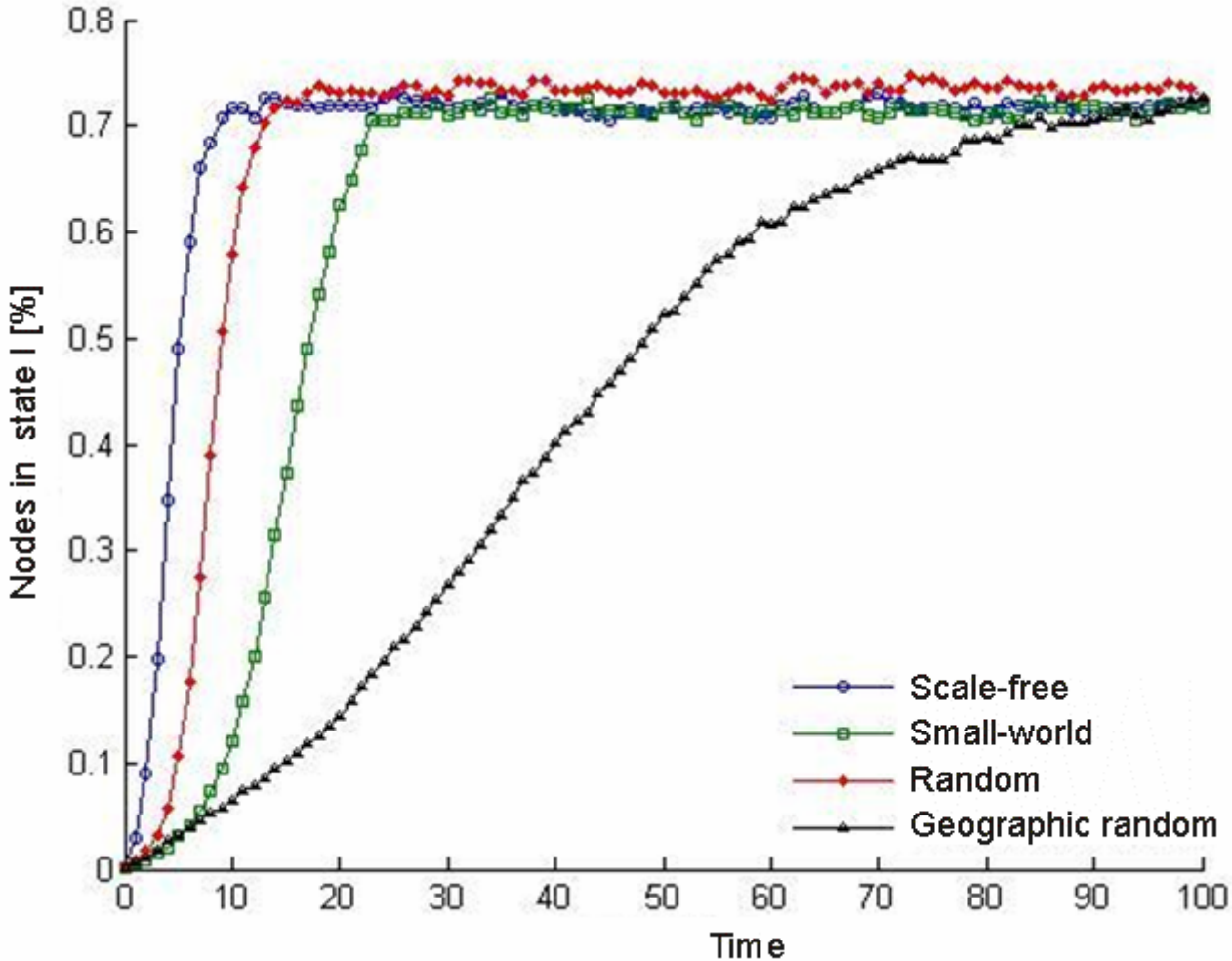
$$p_i^S(k+1) = 1 - p_i^I(k+1)$$

$$p_i^I(k+1) = s_i^S(k)\left[1 - \prod_{j=1}^{n}(1 - \beta d_{ij} s_i^I(k))\right] + (1-\delta)s_i^I(k)$$

- $S_i(k) = [s_i^S(k)\ s_i^I(k)]$ - state vector of node i
- $P_i(k) = [p_i^S(k)\ p_i^I(k)]$ - probability vector of node i
- β infection probability
- δ recovery probability
- d_{ij} connectivity between nodes *i* and *j* (0 or 1)

SIS

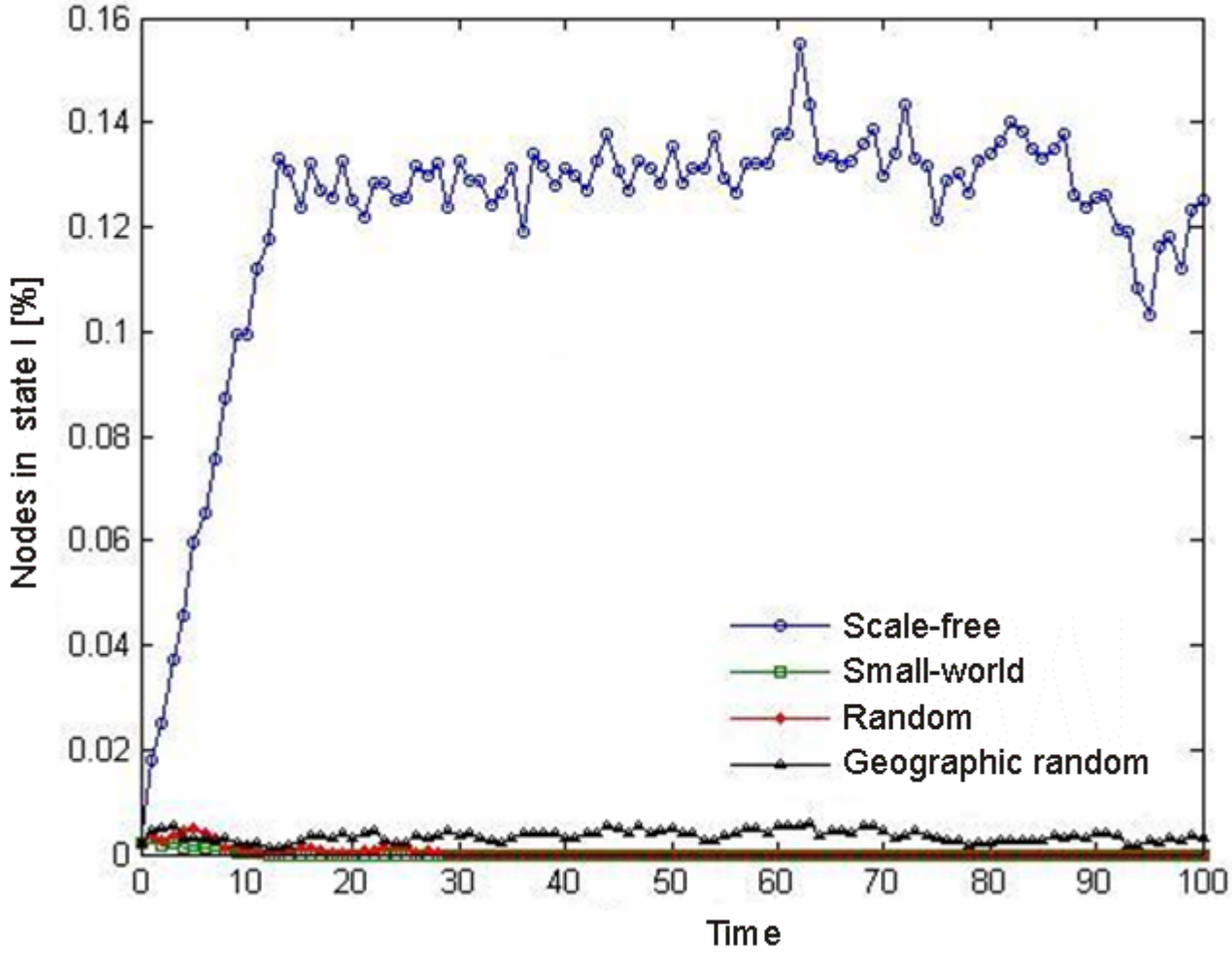
N=500
i=20
β=0.3
δ=0.2



Time dependence of number of infected nodes

SIS

N=500
i=20
β=0.12
δ=0.8



Time dependence of number of infected nodes

Conclusion

- The General influence model is the base for most of the models and with certain modification and specific initial condition can be derived from it.
- Combination of the dynamic processes can be used for modeling real behaviors of networks.
- Discovering network vulnerabilities.
- Finding optimal ways for spreading influence.
- Preventing influence from spreading (diseases).
- Steps forward
  - more complex methods for heterogeneity
  - new weight calculation techniques
  - mobility