Complex networks

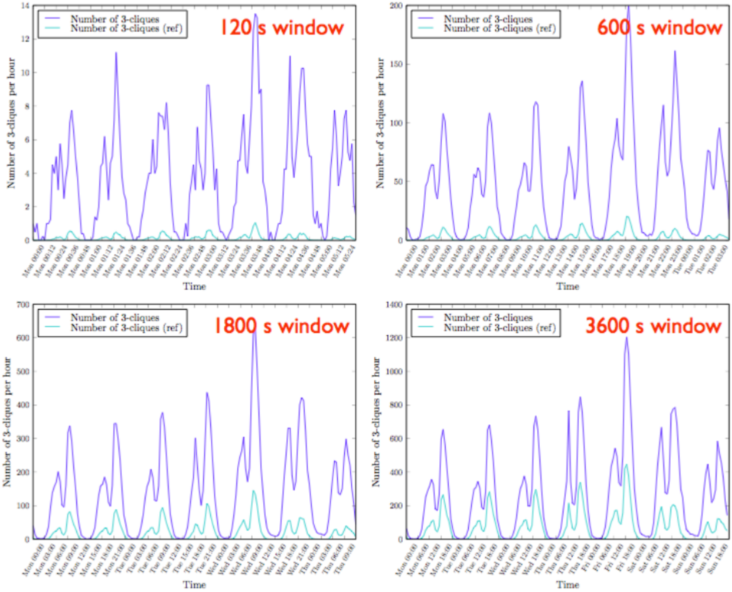
Diffusion and spreading on networks

János Török

Department of Theoretical Physics

April 26, 2023

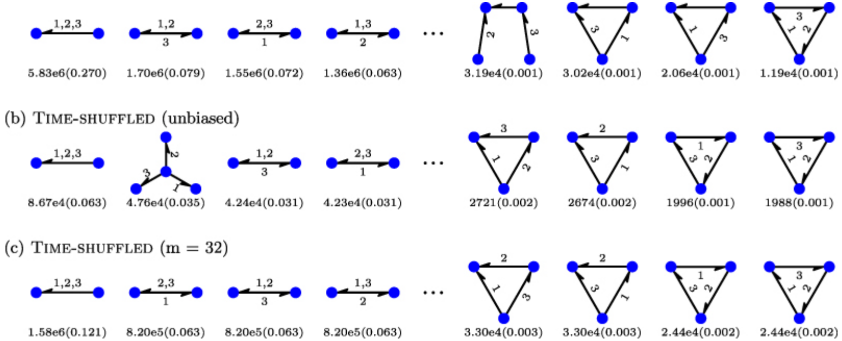
Temporal motifs: occurrence of triangles



Temporal motifs: occurrence of ordered sequences

Most frequent ones

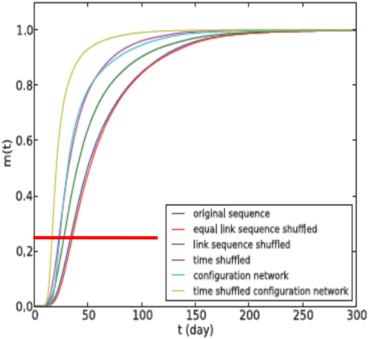
Least frequent ones



Importance of different effects in temporal spreading

- Equal-weight link-sequence shuffled: Whole single-link event sequences are randomly exchanged between links having the same number of events
- Only link-link correlation is destroyed

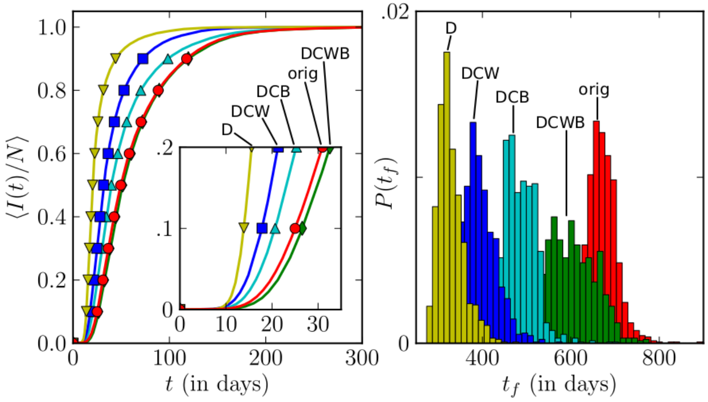
Event sequence	D	C	W	B	E	25%
Original	✓	✓	✓	✓	✓	33.7
Config. shuffle	✓	×	×	×	×	16.4
Config. keep	✓	×	×	✓	×	23.8
Orig. shuffle	✓	✓	✓	×	×	22.9
Shuffle. keep	✓	✓	×	✓	×	27.5
W keep sh.,keep	✓	✓	✓	✓	×	35.3



Importance of different effects in temporal spreading

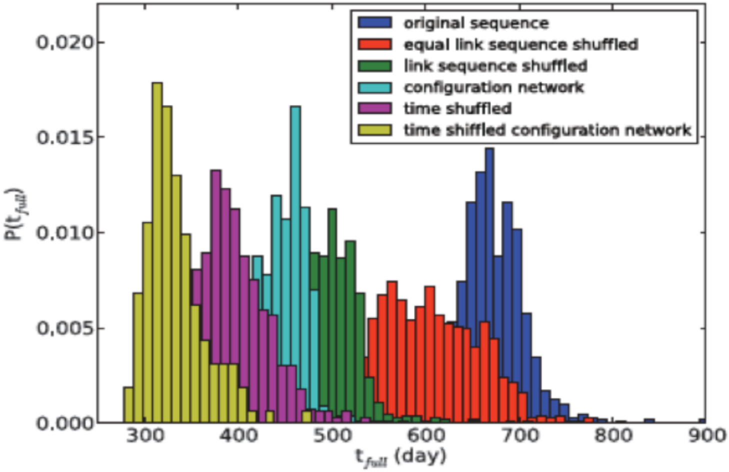
- Long time behaviour:

Event sequence	D	C	W	B	E	25%
Original	✓	✓	✓	✓	✓	33.7
Config. shuffle	✓	×	×	×	×	16.4
Config. keep	✓	×	×	✓	×	23.8
Orig. shuffle	✓	✓	✓	×	×	22.9
Shuffle. keep	✓	✓	×	✓	×	27.5
W keep sh.,keep	✓	✓	✓	✓	×	35.3



Importance of different effects in temporal spreading

- Everything slows down the spreading
- Burstiness has higher impact than topological structures

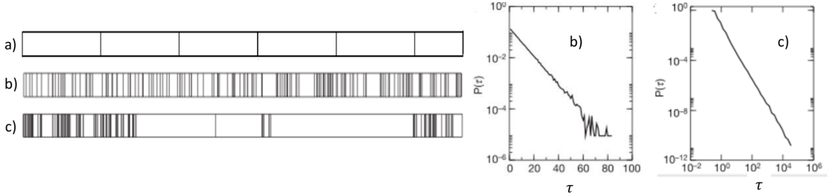


Interevent time

- Time interval between successive events τ
- Distribution of τ is $P(\tau)$
- Distribution is characterized by the average $\langle\tau\rangle$ and the variance σ
- Burstiness:

$$B = \frac{\sigma - \langle\tau\rangle}{\sigma + \langle\tau\rangle}$$

- (a) $B = -1$: deterministic, (b) $B = 0$: Poisson, (c) $B = 1$: bursty

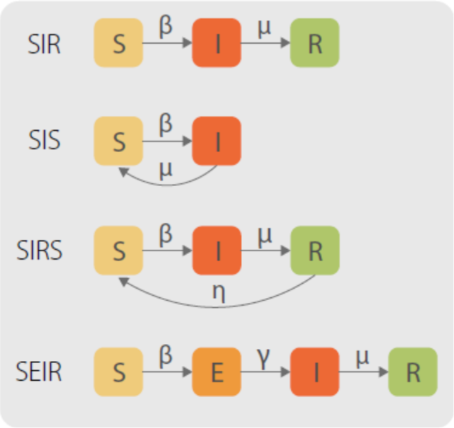


Spreading on networks

- One of the most important problems on networks
- Also one of the real success
- This lecture:
  - Advanced mean-field calculations
  - Cascade models
  - Spreading in temporal networks

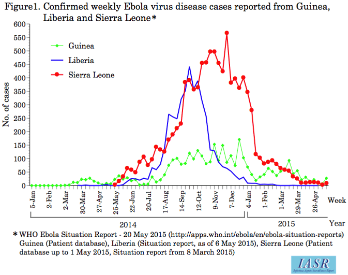
Epidemic models: notations

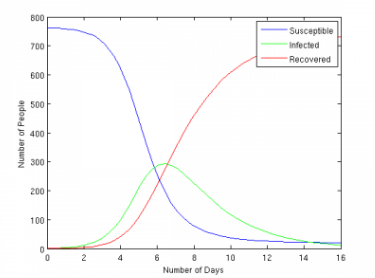
- States:
  - S: susceptible
  - I: Infected
  - R: Recovered (immune)
  - E: Exposed (infected but not yet infecting)
- Rates: β, μ, η, γ



SIR reality vs. model

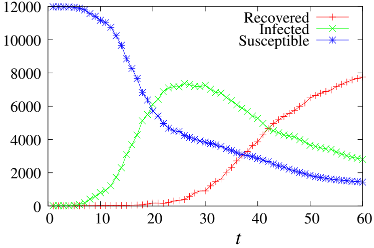
- Perfect mixing
- Everybody can meet everybody
- Ebola

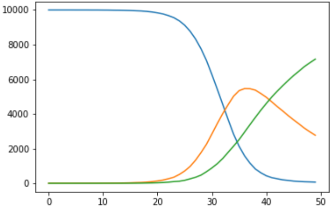




SIR reality vs. model

- Perfect mixing
- Everybody can meet everybody
- Covid-19, South Korea
- Susceptible approximated





SIS model: mean field

- Perfect mixing
- Everybody can meet everybody
- The different type meet with probability proportional to their density
- Density of types:

$$\rho^\alpha = N^\alpha / N$$

- The mean field SIR equations:

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

where $\chi = \mu$ for SIS and $\chi = 0$ for SIR.

Epidemic threshold

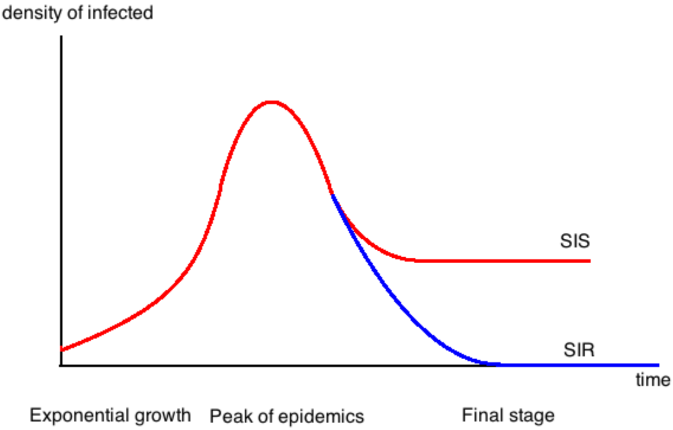
- Linearization: $\rho^I \ll 1$, $\rho^S \simeq 1$

$$\frac{d\rho^I}{dt} = \beta\rho^I\rho^S - \mu\rho^I$$
$$\frac{d\rho^I}{dt} \simeq (\beta - \mu)\rho^I$$
$$\rho^I(t) \simeq \rho^I(0)\exp[(\beta - \mu)t]$$

- Two regimes:
  - $\beta < \mu$: Disease dies out
  - $\beta > \mu$: Disease spreads
- Reproduction number: $R_0 = \beta/\mu$
- The epidemic threshold for perfectly mixing population is $R_0 = 1$ above which the epidemic spreads.

SIR/SIS

- Above the epidemic threshold
- In SIS dynamic equilibrium



SIS: Individual based mean field (IBMF)

- Markov chain approach
- Two state $X_i = 1$ for I and $X_i = 0$ for S.
- $E[X_i(t)]$ expected value of X_i
- a_{ij} element of the adjacency matrix
- The Master equation:

$$\frac{dE[X_i(t)]}{dt} = E\left[-\mu X_i(t) + (1 - X_i(t))\beta \sum_j a_{ij} X_j(t)\right]$$

- Introducing $\lambda = \beta/\mu$ and rescaling the time by $1/\mu$
- For static network:

$$\frac{d\rho_i^I(t)}{dt} = -\rho_i^I(t) + \lambda \sum_j a_{ij}\rho_j^I(t) - \lambda \sum_j a_{ij} E[X_i(t)X_j(t)]$$

Pastor-Satorras et el., Epidemic processes in complex networks (2015)

SIS: Individual based mean field (IBMF)

- Markov chain approach
- The Master equation for static network:

$$\frac{d\rho_i^I(t)}{dt} = -\rho_i^I(t) + \lambda \sum_j a_{ij}\rho_j^I(t) - \lambda \sum_j a_{ij} E[X_i(t)X_j(t)]$$

- No explicit solution due to the two term correlations $E[X_i(t)X_j(t)]$.
- Joint probability distribution cannot be calculated
- Assumption: neighboring nodes are statistically independent:

$$E[X_i(t)X_j(t)] \equiv E[X_i(t)]E[X_j(t)] = \rho_i^I(t)\rho_j^I(t)$$

- The Master equation for the SIS model thus reads:

$$\frac{d\rho_i^I(t)}{dt} = -\rho_i^I(t) + \lambda[1 - \rho_i^I(t)] \sum_j a_{ij}\rho_j^I(t)$$

SIS: Individual based mean field (IBMF)

- The Master equation for static network SIS model in the independent neighbors limit:

$$\frac{d\rho_i^I(t)}{dt} = -\rho_i^I(t) + \lambda[1 - \rho_i^I(t)] \sum_j a_{ij}\rho_j^I(t)$$

- Loss term: probability that node i is infected times the rate of recovery (hidden in the rescaled time)
- Gain term: probability that node i is susceptible, times the total probability that any of its nearest neighbors is infected, times the effective transmission rate $\lambda = \beta/\mu$

SIS: Individual based mean field (IBMF)

- The Master equation for static network SIS model in the independent neighbors limit:

$$\frac{d\rho_i^I(t)}{dt} = -\rho_i^I(t) + \lambda[1 - \rho_i^I(t)] \sum_j a_{ij}\rho_j^I(t)$$

- Linear stability analysis

$$\frac{d\rho_i^I(t)}{dt} \simeq -\rho_i^I(t) + \lambda \sum_j a_{ij}\rho_j^I(t) = \sum_j J_{ij}\rho_j^I(t)$$

with $J_{ij} = -\delta_{ij} + \lambda a_{ij}$

- An endemic state occurs when Λ_1 the largest eigenvalue of J is positive.
- The epidemic threshold is thus:

$$\lambda > \lambda_c^{\mathrm{IBMF}} \equiv \frac{1}{\Lambda_1}$$

SIS: Individual based mean field (IBMF)

- For networks with power law degree distribution $P(k) \sim k^{-\gamma}$
- Largest eigenvalue:

$$\Lambda_1 = \min(\sqrt{k_{\max}}, \langle k^2 \rangle / \langle k \rangle)$$

- The epidemic threshold:

$$\lambda_c^{\text{IBMF}} = \begin{cases} \frac{1}{\sqrt{k_{\max}}} & \text{if } \gamma \geq 5/2 \\ \frac{\langle k \rangle}{\langle k^2 \rangle} & \text{if } 2 < \gamma < 5/2 \end{cases}$$

- In both cases $\lim_{N \to \infty} \lambda_c^{\text{IBMF}} = 0$ for scale free networks
- Of course in finite system there is a small deviation but in infinite systems there is no epidemic threshold since the largest degree is infinite.

SIS: Degree based mean field (DBMF)

- All nodes with the same degree are statistically equivalent.
- Degree has a maximum k_{max}
- Number of equations k_{max}
- Conditional probabilities: $P(k'|k)$ probability that a node with degree k is connected to a node of degree k'
- $P(k'|k)$ is the same for all k degree nodes.
- In the case of uncorrelated networks:

$$P(k'|k) = \frac{k' P(k')}{\langle k \rangle}$$

SIS: Degree based mean field (DBMF)

- $\rho^I_k(t)$ is the probability that a node of degree k is infected at time t
- Master equation:

$$\frac{d\rho^I_k(t)}{dt} = -\rho^I_k(t) + \lambda k[1 - \rho^I_k(t)] \sum_{k'} P(k'|k)\rho^I_{k'}(t)$$

- Note that the factor k in the gain term is for the number of links the node of degree k has with that chance to get infected
- Linearized version

$$\frac{d\rho^I_k(t)}{dt} \simeq -\rho^I_k(t) + \lambda k \sum_{k'} P(k'|k)\rho^I_{k'}(t) = \sum_{k'} J_{kk'}\rho^I_{k'}(t)$$

- With

$$J_{kk'} = -\delta_{kk'} + \lambda k P(k'|k)$$

SIS: Degree based mean field (DBMF)

- Linearized version

$$\frac{d\rho_k^I(t)}{dt} \simeq -\rho_k^I(t) + \lambda k \sum_{k'} P(k'|k)\rho_{k'}^I(t) = \sum_{k'} J_{kk'}\rho_{k'}^I(t)$$

- With

$$J_{kk'} = -\delta_{kk'} + \lambda k P(k'|k)$$

- There is an epidemic state if

$$\lambda > \lambda_c^{DBMF} = \frac{1}{\Lambda_1}$$

where again Λ_1 is the largest eigenvalue of J

SIS: Degree based mean field (DBMF)

- The epidemic threshold, for uncorrelated networks with

$$P(k'|k) = \frac{k'P(k')}{\langle k \rangle}$$

- Probability to find an infected node following a randomly chosen edge

$$\Theta = \sum_{k'} P(k'|k) = \frac{k'P(k')}{\langle k \rangle} \rho^I_{k'}(t)$$

The Master equation of the Degree based mean field is

$$\frac{d\rho^I_k(t)}{dt} = -\rho^I_k(t) + \lambda k[1 - \rho^I_k(t)]\Theta$$

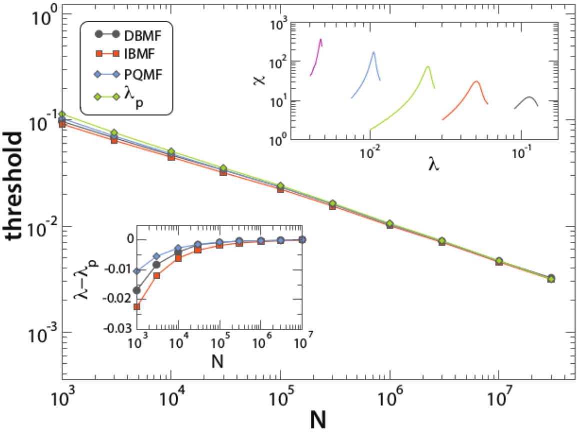
- The two latter equations can be solved in self-consistently.

SIS: Degree based mean field (DBMF)

- The self-consistent solution allows for an epidemic state only if

$$\lambda > \lambda_c^{DBMF} = \frac{\langle k \rangle}{\langle k^2 \rangle}$$

- For power law degree distribution with exponent $2 < \gamma \leq 3$
  The threshold is 0 in the infinite limit.



SIS: Comparison

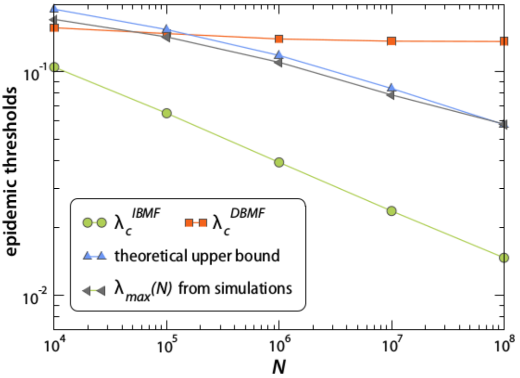
- IBMF

$$\lambda_c^{\text{IBMF}} = \begin{cases} \frac{1}{\sqrt{k_{max}}} & \text{if } \gamma \geq 5/2 \\ \frac{\langle k \rangle}{\langle k^2 \rangle} & \text{if } 2 < \gamma < 5/2 \end{cases}$$

- DBMF

$$\lambda > \lambda_c^{DBMF} = \frac{\langle k \rangle}{\langle k^2 \rangle}$$

- Network: scale free with $\gamma = 2.25$



SIS: Epidemic threshold

- It seems that in the scale free networks in the infinite system there is no epidemic threshold
- Numerical simulations show also this picture
- Note that in the SIS model there is a dynamic steady state with a fraction of infected nodes
- In scale-free networks only part of the system will be infected, the hubs and the immediate neighborhood.
- Concepts of:
  - **Epidemic state**: Homogeneously infected
  - **Active state**: Small finite active part
- Thankfully real systems are never infinite and never scale-free

Problems with SIR model

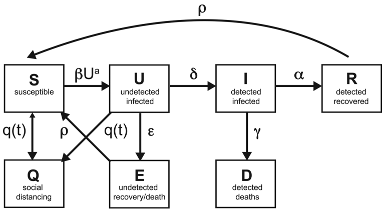
- Fully mixed society $\rightarrow$ network
- Disease either dies out fast or infects the whole society even on networks $\rightarrow$ geographical location with travel links effected by the infection
- Inhomogeneous society $\rightarrow$ age groups and connections as in SBM

Age	0-9	10-19	20-29	30-39	40-49	50-59	60-69	70-79	80+
0-9	19.2	4.8	3.0	7.1	3.7	3.1	2.3	1.4	1.4
10-19	4.8	42.4	6.4	5.4	7.5	5.0	1.8	1.7	1.7
20-29	3.0	6.4	20.7	9.2	7.1	6.3	2.0	0.9	0.9
30-39	7.1	5.4	9.2	16.9	10.1	6.8	3.4	1.5	1.5
40-49	3.7	7.5	7.1	10.1	13.1	7.4	2.6	2.1	2.1
50-59	3.1	5.0	6.3	6.8	7.4	10.4	3.5	1.8	1.8
60-69	2.3	1.8	2.0	3.4	2.6	3.5	7.5	3.2	3.2
70-79	1.4	1.7	0.9	1.5	2.1	1.8	3.2	7.2	7.2
80+	1.4	1.7	0.9	1.5	2.1	1.8	3.2	7.2	7.2

Ram and Schaposnik, Sci. Rep. 2021

Problems with SIR model

- Fully mixed society $\rightarrow$ network
- Disease either dies out fast or infects the whole society even on networks $\rightarrow$ geographical location with travel links effected by the infection
- Inhomogeneous society $\rightarrow$ age groups and connections as in SBM
- Computational limitations $\rightarrow$ two level systems
- Government measures $\rightarrow$ new states in the SIR model



Khan, Van Bussel, Hussain, Epidemiology and Infection 2020

Immunization

- Often the task is to stop the spreading
- Sometimes one can immunize part of the society
- Can we stop the spreading?
- Example:
    - Of course, if every newborn baby is vaccinated, the population is safe. This is the way, how smallpox (Variola) was defeated.
    - Estimated death in 20 th century: 300 Million
    - Estimated infected in 1967: 15 Million
    - 1979: WHO declared smallpox eradicated

Immunization

- Epidemic threshold (complete graph/fully mixed state):

$$R_0 = \frac{\beta}{\mu} \begin{cases} > 1 & \text{outbreak} \\ = 1 & \text{threshold} \\ < 1 & \text{localized} \end{cases}$$

- The density of the immune vertices is g, then:

$$\beta' = \beta(1-g)$$

- The threshold for networks

$$\frac{\beta(1-g)}{\mu} = \frac{\langle k \rangle}{\langle k^2 \rangle}$$

- For infinitely large scale free network with $\gamma \leq 3$ we get $g_c = 1$
- For random immunization everybody must be vaccinated

Immunization

- Epidemic threshold for networks

$$\frac{\beta(1-g)}{\mu} = \frac{\langle k \rangle}{\langle k^2 \rangle}$$

- Targeted immunization: immunize high degree nodes
- This decreases the variance faster than the average

$$\frac{\langle k \rangle_g}{\langle k^2 \rangle_g} > \frac{\beta(1-g)}{\mu}$$

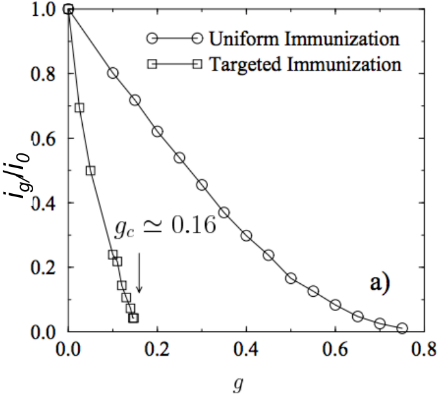
which defines the critical value of g

Immunization

- Targeted immunization: immunize high degree nodes

$$\frac{\langle k \rangle_g}{\langle k^2 \rangle_g} > \frac{\beta(1-g)}{\mu}$$

which defines the critical value of g

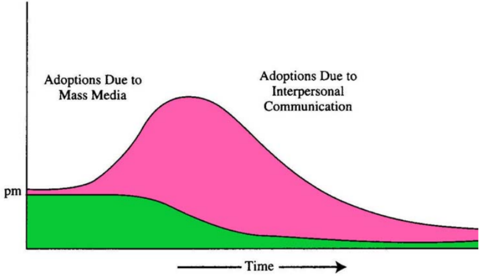


Pastor-Satorras, Vespignani, 2001

Innovation spreading

- Rogers (1962)

- Mahajan, Muller and Bass (1990)



Threshold model

- Sometimes the spreading is due to load from the neighbors
- E.g. if too many of my neighbors are infected I will also get infected
- Innovation spreading: many of my friends have iPhone I will also get one.

Threshold model

- Networks with average degree $\langle k \rangle = z$
- Nodes have threshold ϕ_i
- If the number of active nodes in the neighborhood reach ϕ_i then the node becomes active (too many friends have some product I will also buy it)
- Start from a small seed
- If thresholds are sufficiently low cascades may propagate through the whole system (size $\sim \mathcal{O}(N)$)

Watts, A simple model of global cascades on random networks (2002)

Threshold model

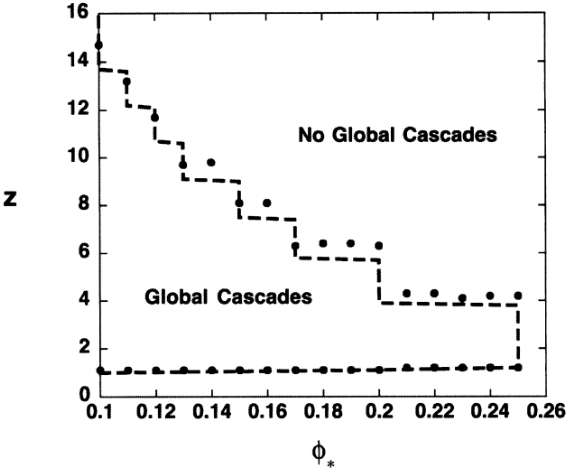
- In large uncorrelated random networks there are hardly any triangles
- *Vulnerable* nodes are the ones where the threshold is less than $\phi_i < 1/k_i$, one neighbor is enough to get infected
- Global cascade is possible if these nodes percolate
- This is the *cascade condition*

$$z > \sum_k k(k-1)P(k)P(\phi \leq 1/k)$$

- $k(k-1)$ increases with k
- $P(\phi \leq 1/k)$ decreases with k
- Two or 0 solutions

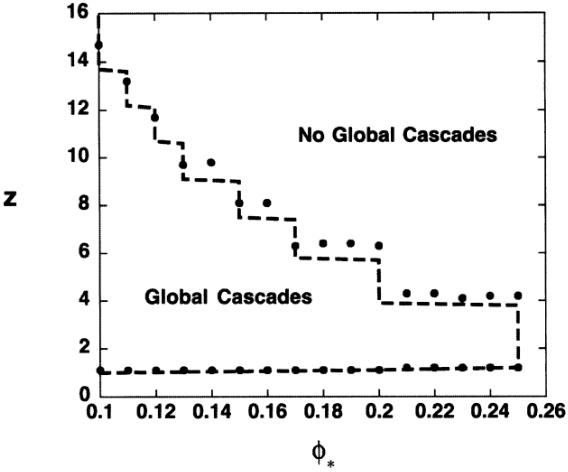
Threshold model: Phase diagram

- Points simulation
- Dashed line calculated threshold



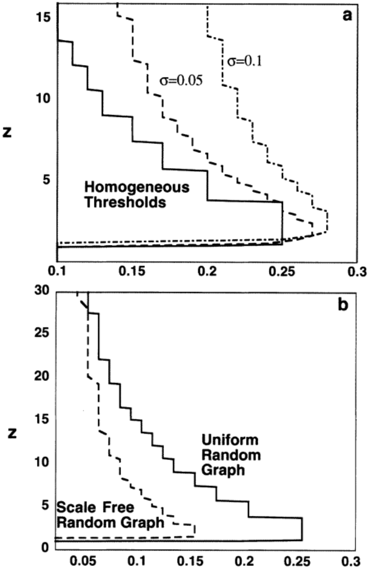
Threshold model: Phase diagram

- Top line: first order phase transition of cascades
- Bottom line: second order phase transition of network percolation limit



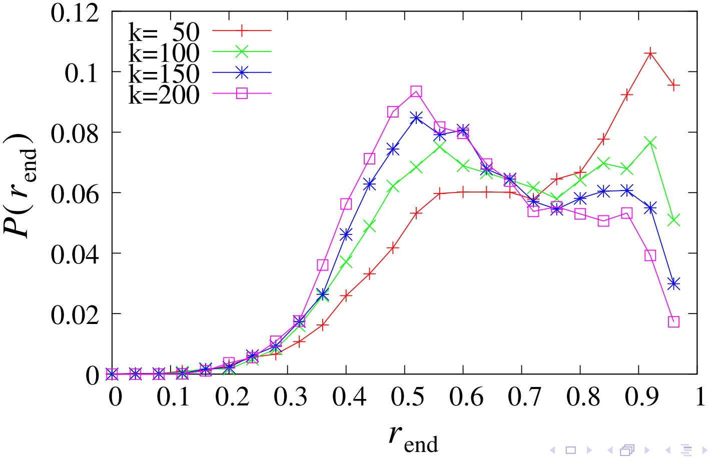
Threshold model: Phase diagram

- ϕ With normal distribution and σ variance
- Scale free graph



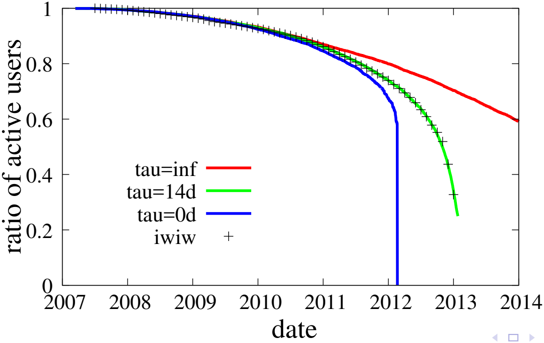
Fall of a social network site

- Users leave due to exogenous effects (advertisements, news, etc.)
- Users leave if some part of their friends leave.
- This depends on the embeddedness of the user



Fall of a social network site: Model

- Users leave due to exogenous effects (advertisements, news, etc.):
  - Here rate of leave increases with time as was the popularity of the alternative site
  - Users with low degree are more susceptible to global effects
- Users leave if their friends leave.
  - Threshold model with threshold above 50%
  - Leave is not immediate one needs time (τ) to recognize friend is inactive

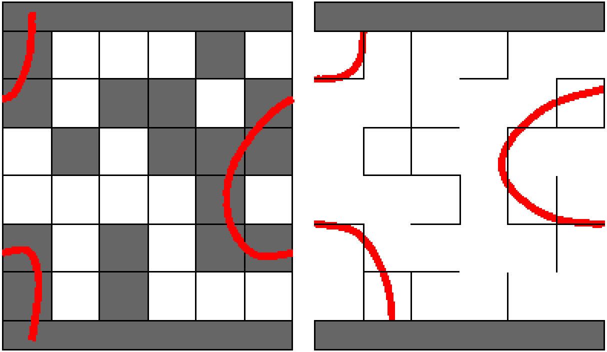


Percolation

Percolation

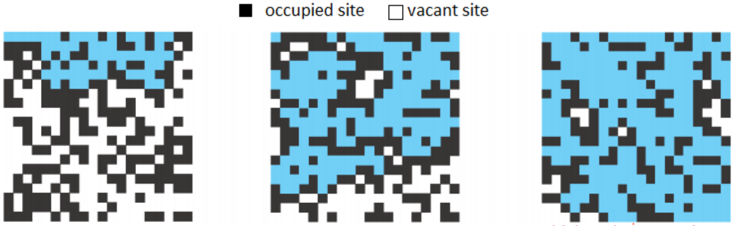
Behavior of connected cluster

- Site percolation
- Bond percolation



Percolation model

- Random environment
- With probability p site vacant (conducts)
- Two states: percolates or not!
- Percolation: presence of infinite cluster, in infinitely large system the cluster holds finite fraction of nodes.



low p: does not percolate

high p: does percolate

Percolation theory

Questions (in infinite systems):

1. Is there an infinite cluster in infinite systems?
2. How many infinite clusters are there?
3. Mean cluster size (without the infinite one)?
4. Cluster size distribution

Answers:

1. Above a critical density with probability 1 below it with probability 0
2. Only 1!
3. Decreases as a power low away from the critical density
4. Power law

Percolation theory

Questions (in infinite systems):

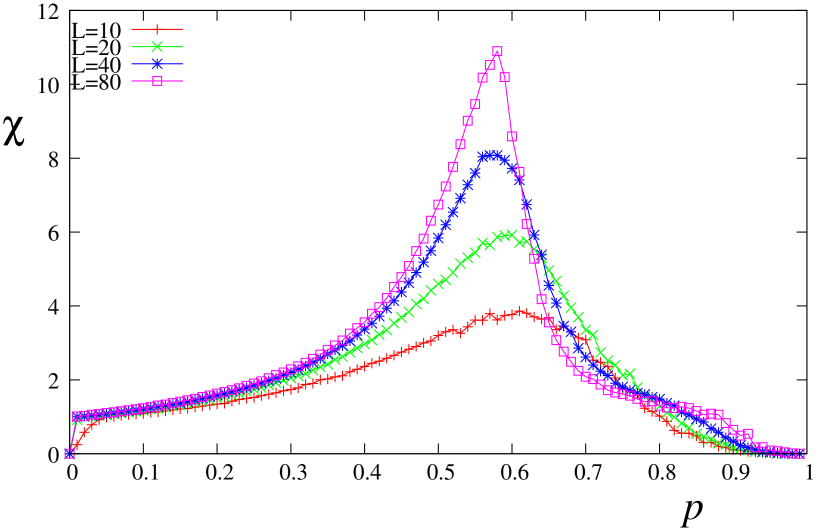
1. Is there an infinite cluster in infinite systems?
2. How many infinite clusters are there?
3. Cluster size distribution (n_s)
4. Mean cluster size (without the infinite one)? ($S = \sum_s s^2 n_s$)

Answers:

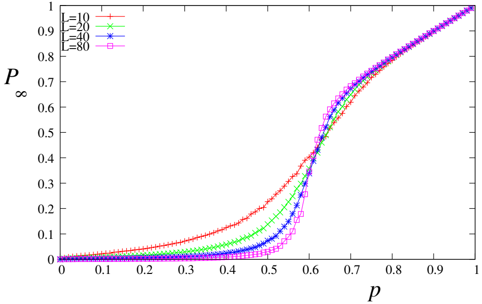
1. if $p > p_c$ then yes, otherwise no
2. Only 1!
3. $n_s \sim s^{-\tau}$
4. $S \sim |p - p_c|^{-\gamma}$

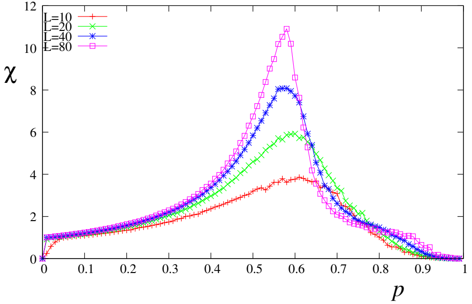
Like a second order phase transition in a geometric system!

Percolation model

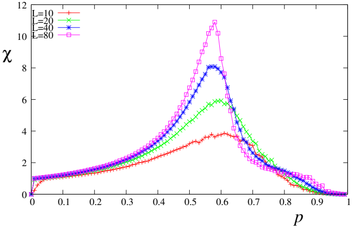


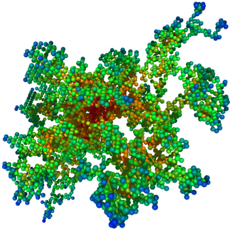
Percolation model





Percolating cluster





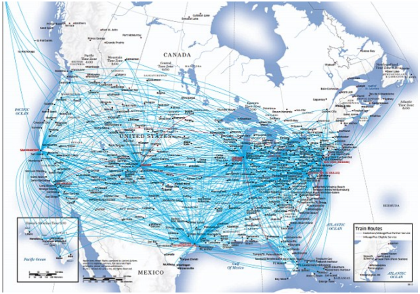
- Largest cluster
  - fractal with fractal dimension of d_f
  - $S_\infty \sim \begin{cases} \xi_f^d \log(N/\xi^d) & p < p_c \\ N^{d_f/d} & p = p_c \\ NP_\infty(p) & p > p_c \end{cases}$
- Largest not infinite cluster: size $\sim |p - p_c|^{-\nu}$

Percolation on networks (graphs)

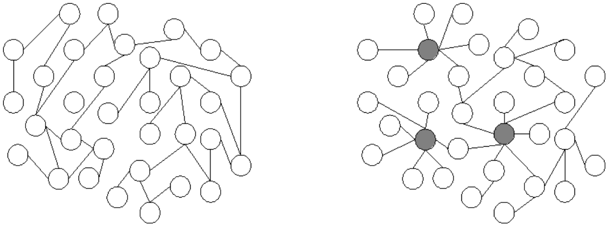
- Network is defined by nodes and links
- Percolation gives us connected components
- Link removal percolation gives information about robustness, and structure

Percolation and attack on random networks

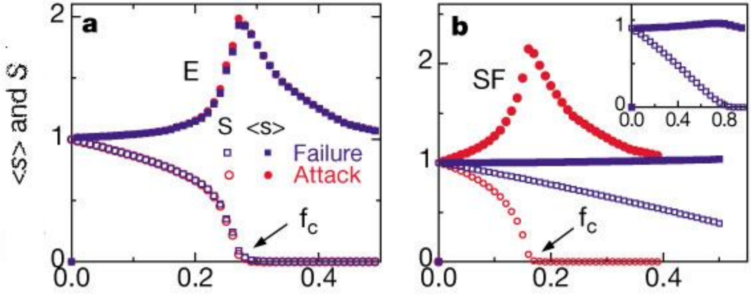
- Failure: equivalent to percolation: remove nodes at random
- Attack: remove most connected nodes



Error vs. attacks

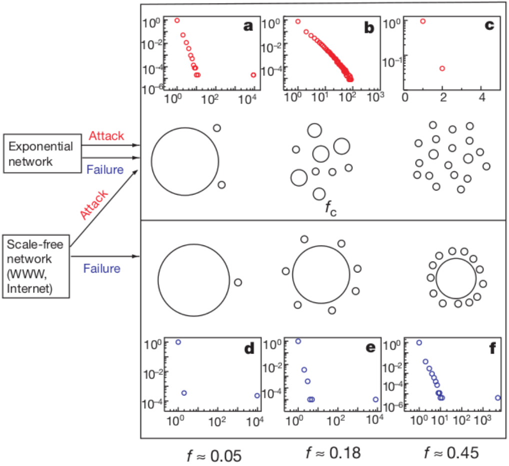


(a) Random network (b) Scale-free network



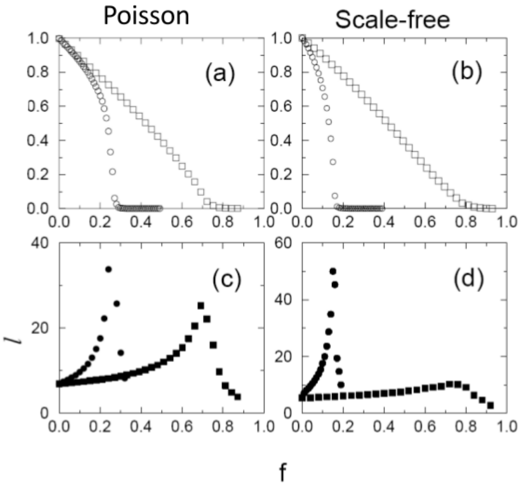
Percolation and attack on random networks

- Failure: equivalent to percolation: remove nodes at random
- Attack: remove most connected nodes



Robustness

- Link/node removal percolation
- Here: random, and largest first
- There is also weakest first



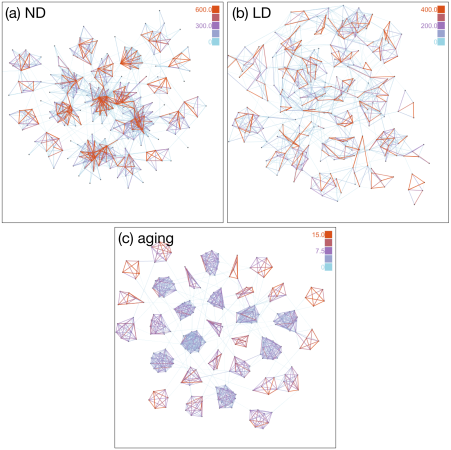
- squares: random failure
- circles: targeted attack

Failures: little effect on the integrity of the network if scale free.
Attacks: fast breakdown

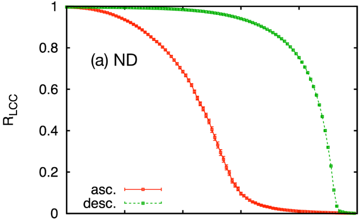
l: average component size

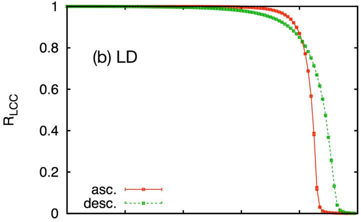
Link removal percolation on networks

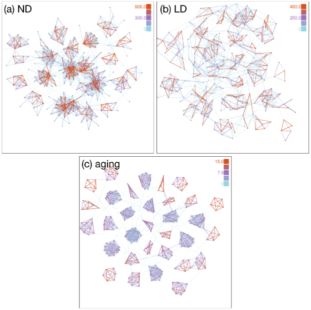
- Granovetter hypothesis: The strength of the weak ties
- Human communities have strong connections
- These communities are connected with weak ties
- Test the structures with Link removal percolation

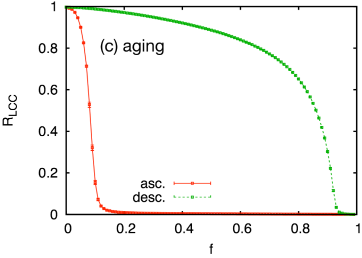


Link removal percolation on networks



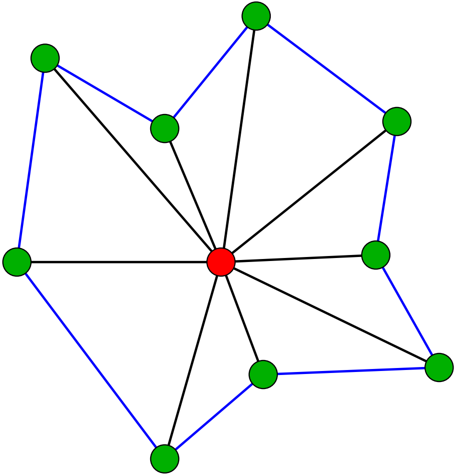






Robustness

- Resistant both against random and targeted attacks.
- Must have hubs to resist random attacks
- Small degree nodes should be interconnected so they remain viable after removal of the hubs



Robustness against attacks

- Malicious attacks target central nodes, hubs
- Solution: central nodes should be connected
- Assortative mixing is preferred (high degree nodes are connected between each other)
- (Barabasi-Albert is thus a bad example)
- Robustness measure:

$$R = \frac{1}{N} \sum_{Q=1}^{N} s(Q)$$

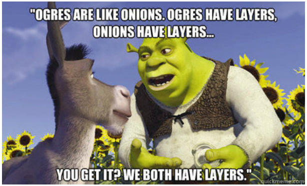
- $s(Q)$ fraction of nodes in the largest connected cluster after removing $Q = qN$ nodes
- Optimize for R

Onion structures

- Robustness measure:

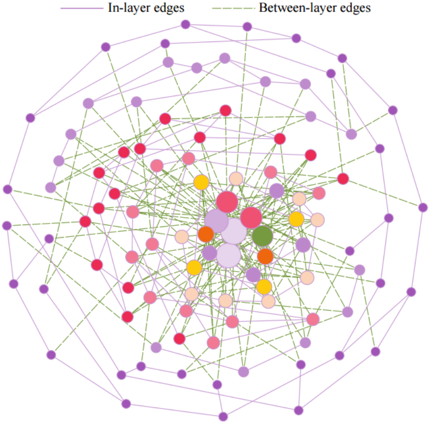
$$R = \frac{1}{N}\sum_{Q=1}^{N} s(Q)$$

- $s(Q)$ fraction of nodes in the largest connected cluster after removing $Q = qN$ nodes
- Optimize R by only rewiring and keeping degree distribution constant
- Onion structures are the most robust
  - Assortative
  - Layers with similar degree nodes
  - Inter-layer connections



Onion structures

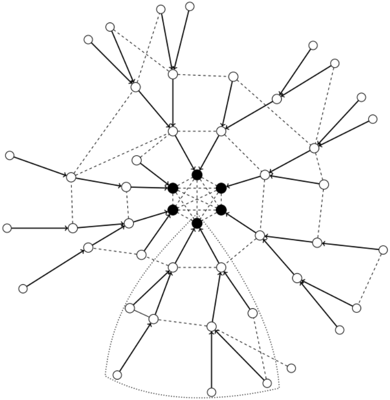
- Assortative
- Layers with similar degree nodes
- Inter-layer connections



Château de Vincennes

Optimization: costs

- Internet Autonomous system topology
- Providers can connect to the top tier or be a customer
- They are responsible for directing the Internet traffic
- Simple protocols define the routing (mainly greedy)
- Many optimizes the structure



Flight route optimization

- Suppose weight of a link is defined as

$$w_{ij} = d_{ij} / t_{ij}$$

where d_{ij} is the distance, and t_{ij} is the traffic between two cities

- When more paths are possible the most economical is used:

$$C_{ij} = \min_{p \in \mathcal{P}} \sum_{l \in p} w_l$$

- Keep total traffic constant
- Function to be optimized is the average cost to pay to travel from any node to any other

$$\mathcal{L} = \frac{2}{N(N-1)} \sum_{i<j} C_{ij}$$

Flight route optimization

- Check a small circle:
- Let us assume $d_1 = d(A, B) = d(B, C) > d(A.C) = d'$
- Cost function (T is the average traffic between two cities):

$$\mathcal{L}_1 = \frac{2d + d'}{T}$$

- Cut connection (B, C). The new cost function

$$\mathcal{L}_2 = \frac{d + d'}{2T} < \mathcal{L}_1$$

- The optimal path is a tree!

Tree model

- If it is known that the network is a tree task is easier:

$$\mathcal{L}_{\sqcup} = \sum_{e \in T} b_e \frac{d_e}{t_e}$$

  where b_e is the link betweenness centrality
- The optimal traffic

$$t_e = \frac{T\sqrt{b_e d_e}}{\sum_e \sqrt{b_e d_e}}$$

- The optimal traffic tree can then be obtained by minimizing

$$\mathcal{L} = \sum_{e \in T} \sqrt{b_e d_e}$$

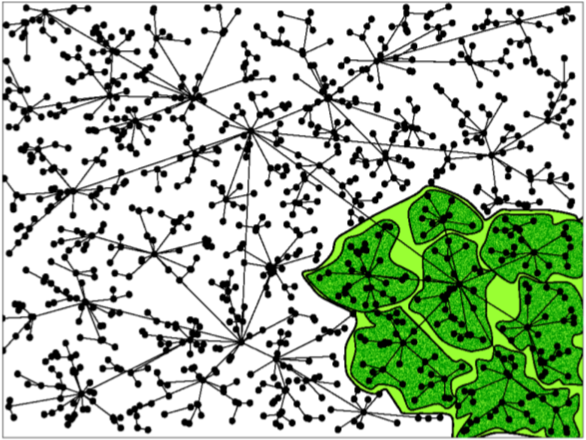
- More generally

$$\mathcal{L} = \sum_{e \in T} b_e^{\mu} d_e^{\nu}$$

  where μ and ν control the relative importance of distance against topology as measured by centrality

Optimal traffic on networks

- Exponential degree distribution
- Power law betweenness distribution
- Hierarchical organizations
- $\mu = \nu = 0.5$



Optimal traffic on networks