Network theory

Part V: errors, attacks and walks

Complexity in Social Systems
AA 2023/2024
Maxime Lucas
Lorenzo Dall'Amico

Recap last lecture

Models

Barabasi-Albert model
Bianconi-Barabasi model
Link/Copying model

Concepts

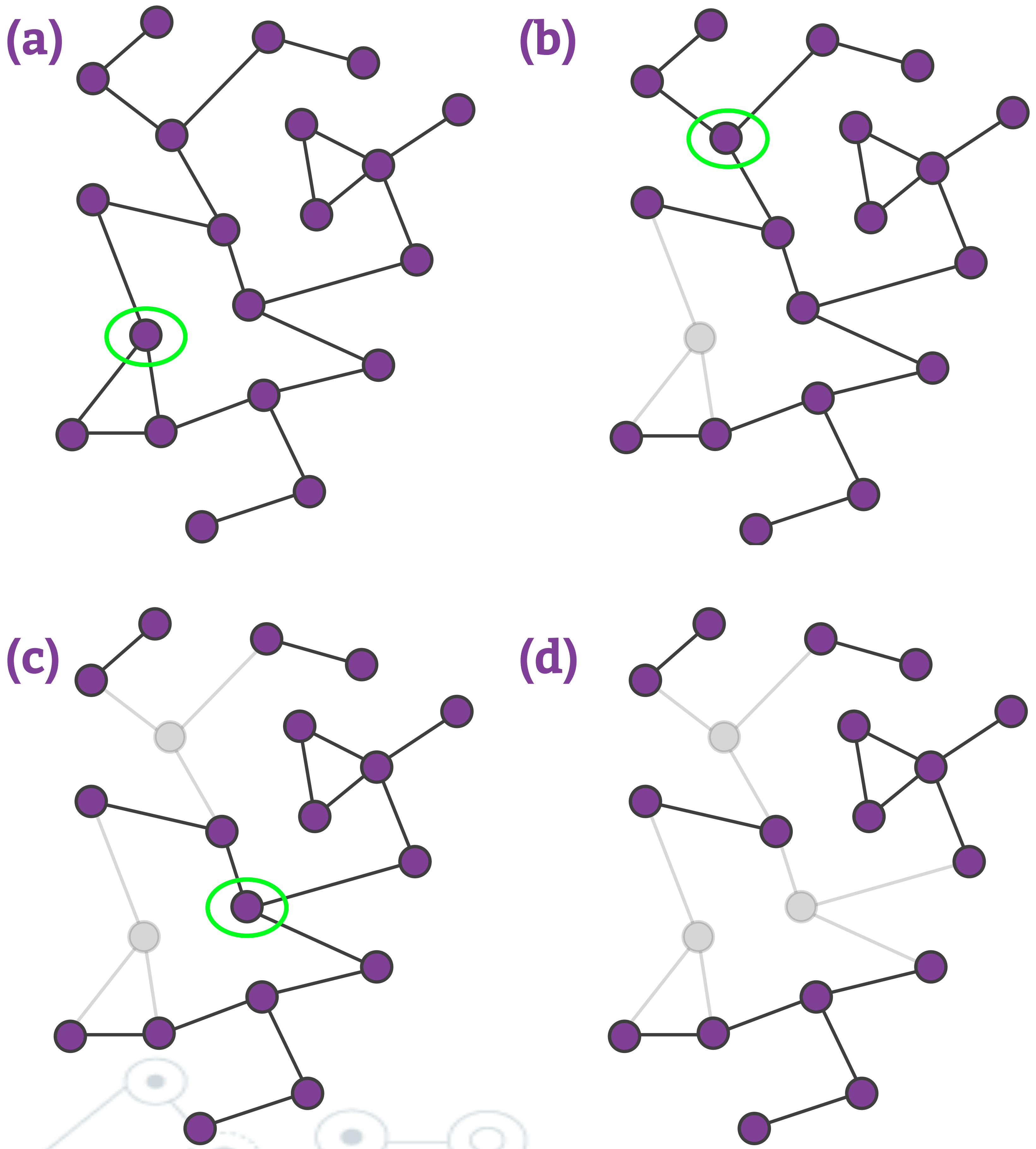
Origins of scale-free distributions
Growing networks
Assortativity and correlations
“Robustness”

Today!

Percolation!
Random Walks
Spectral properties

Robustness

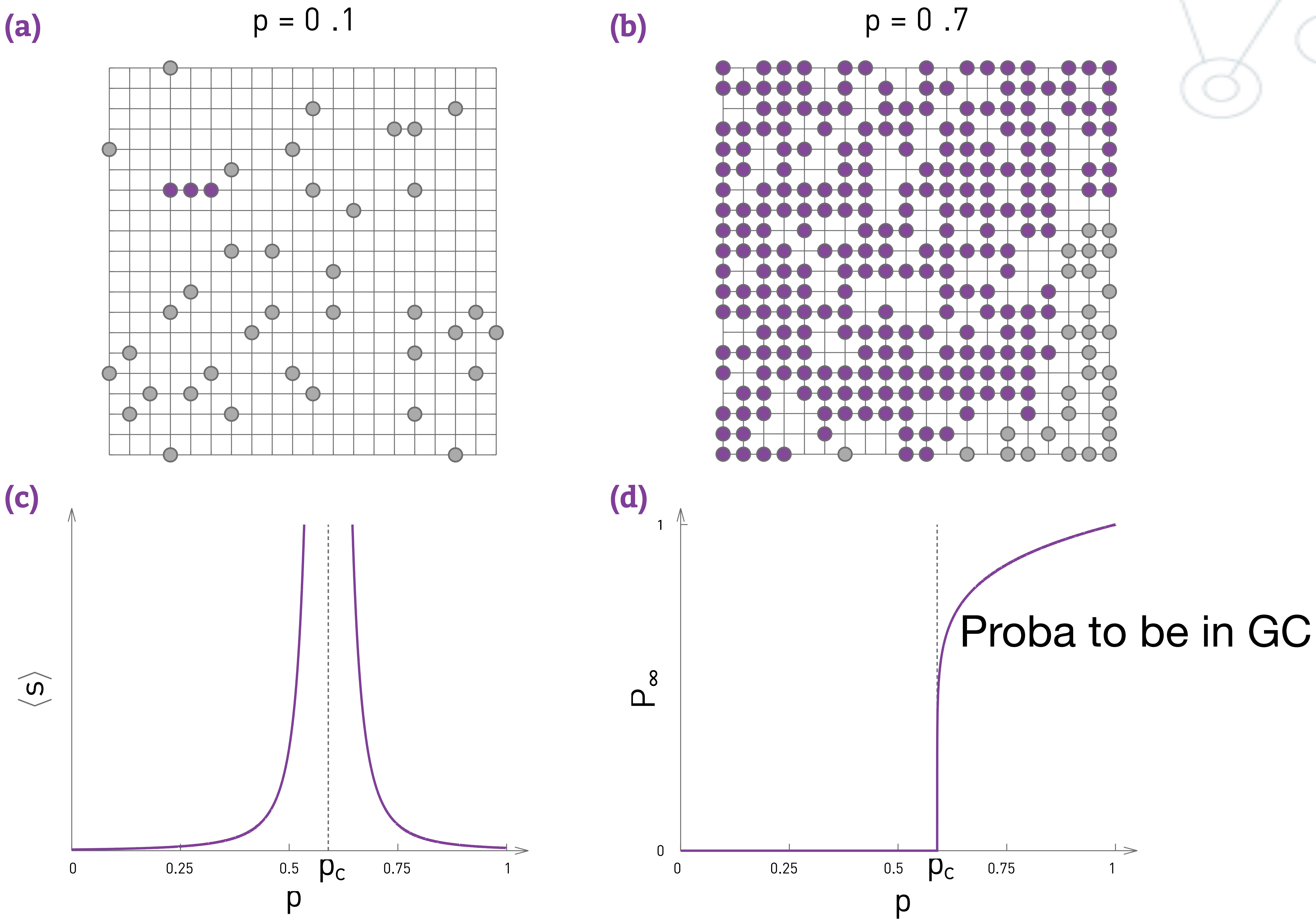
Introduction



Percolation

- What is the expected size of the largest cluster?
- What is the average cluster size?

Add dot with proba p



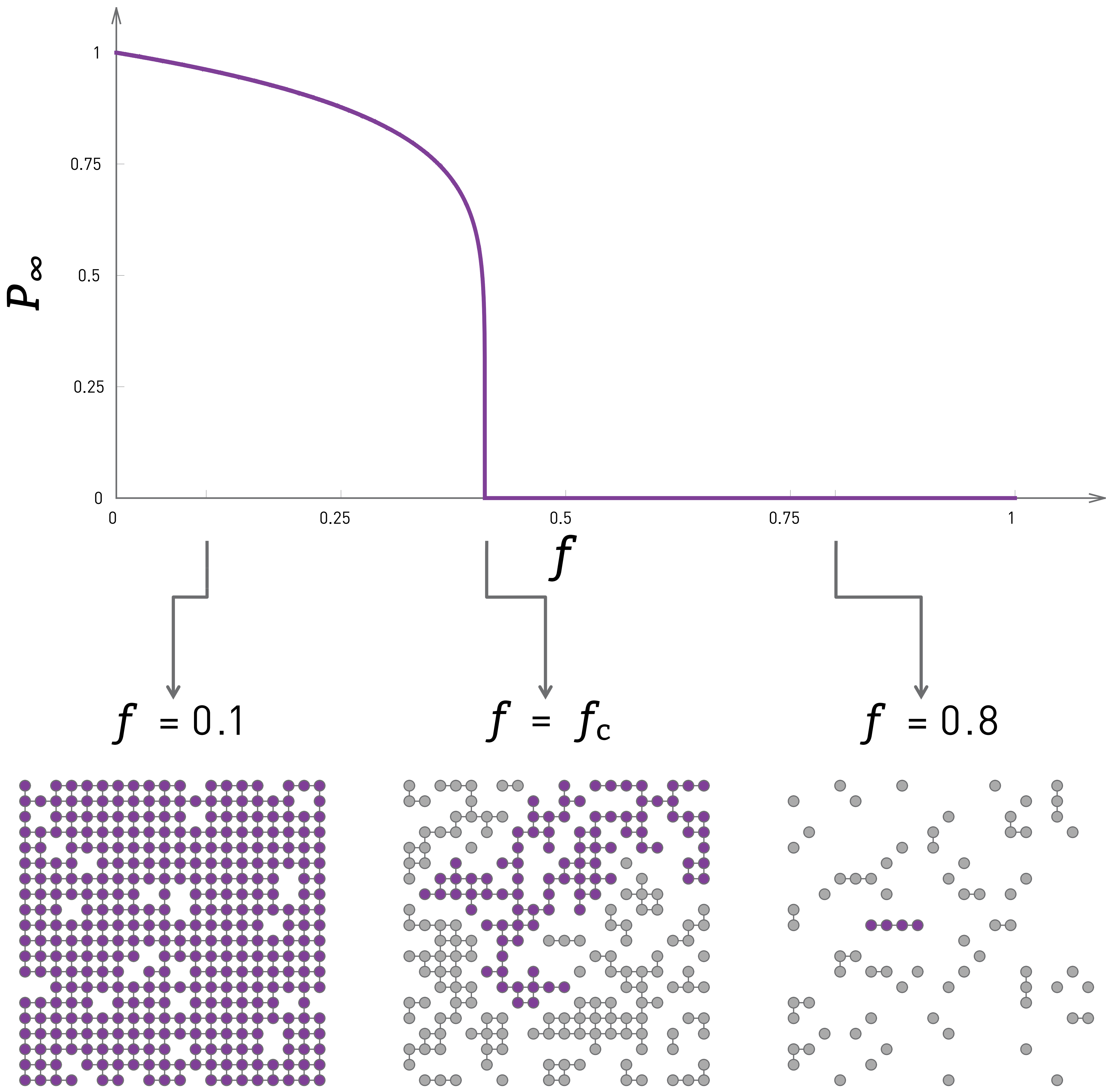
$$\langle s \rangle \sim |p - p_c|^{-\gamma_p}$$

$$P_\infty \sim (p - p_c)^{\beta_p}.$$

Robustness

Inverse percolation

Remove a fraction f of nodes



$0 < f < f_c$:

There is a giant component.

$P_\infty \sim |f - f_c|^\beta$

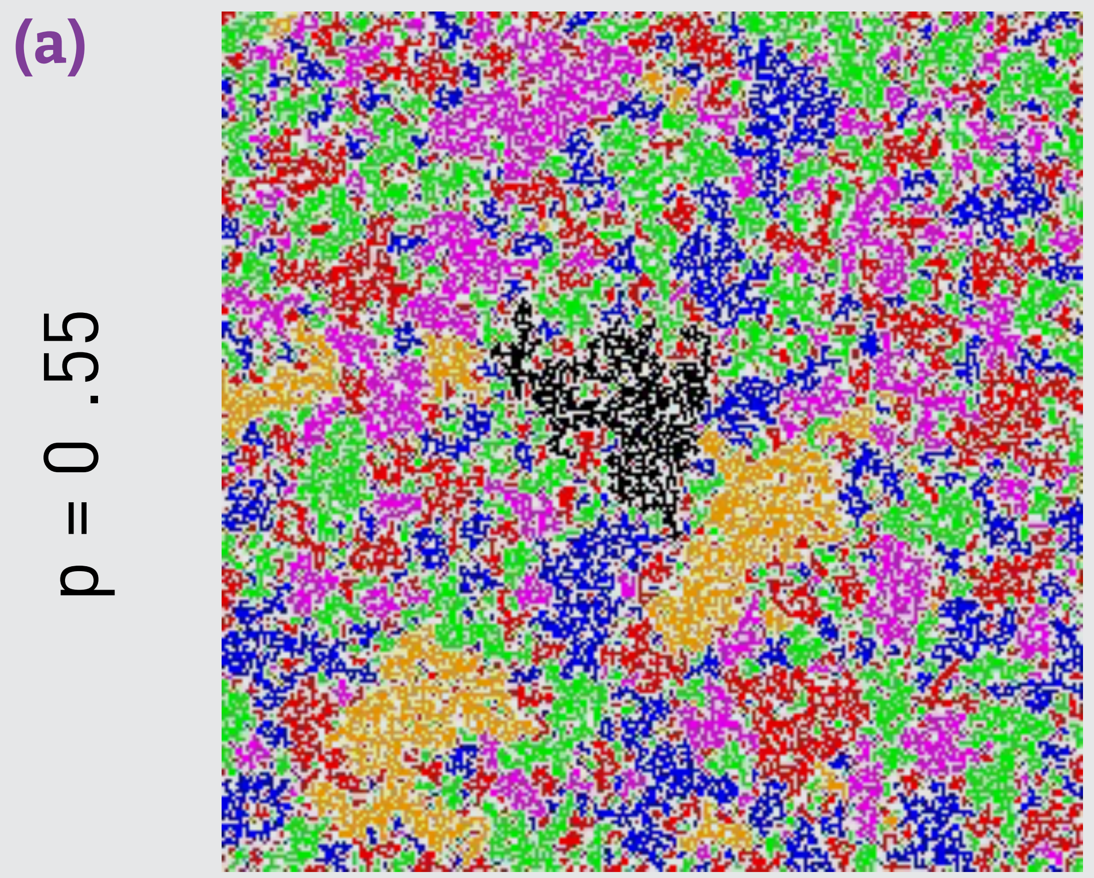
$f = f_c$:

The giant component vanishes.

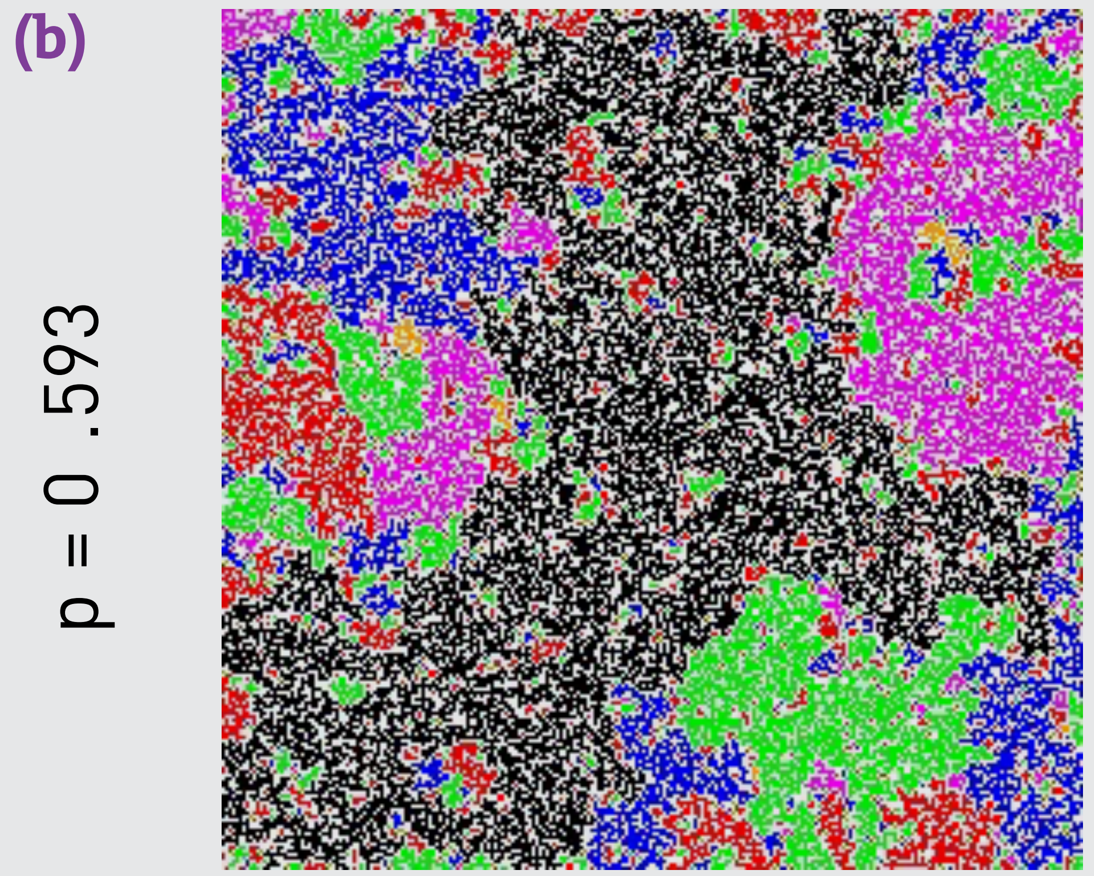
$f > f_c$:

The lattice breaks into many tiny components.

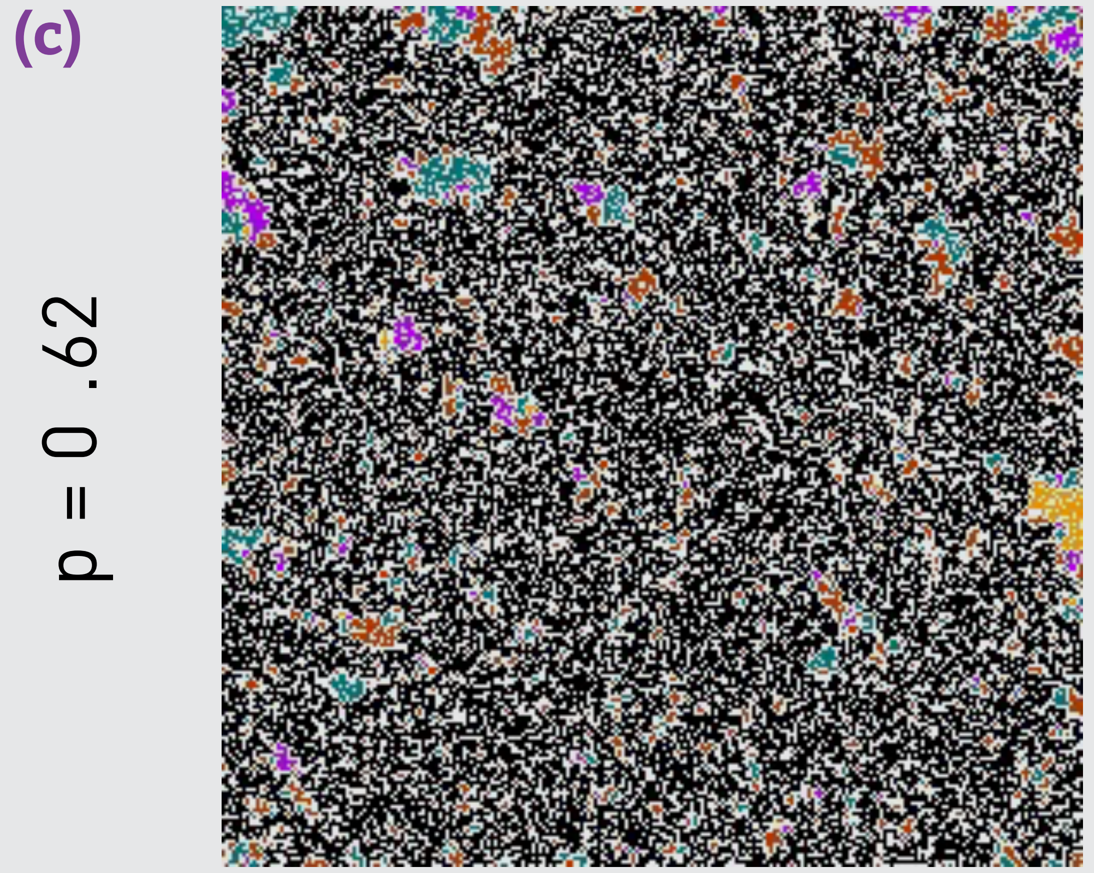
Small islands



Spanning cluster



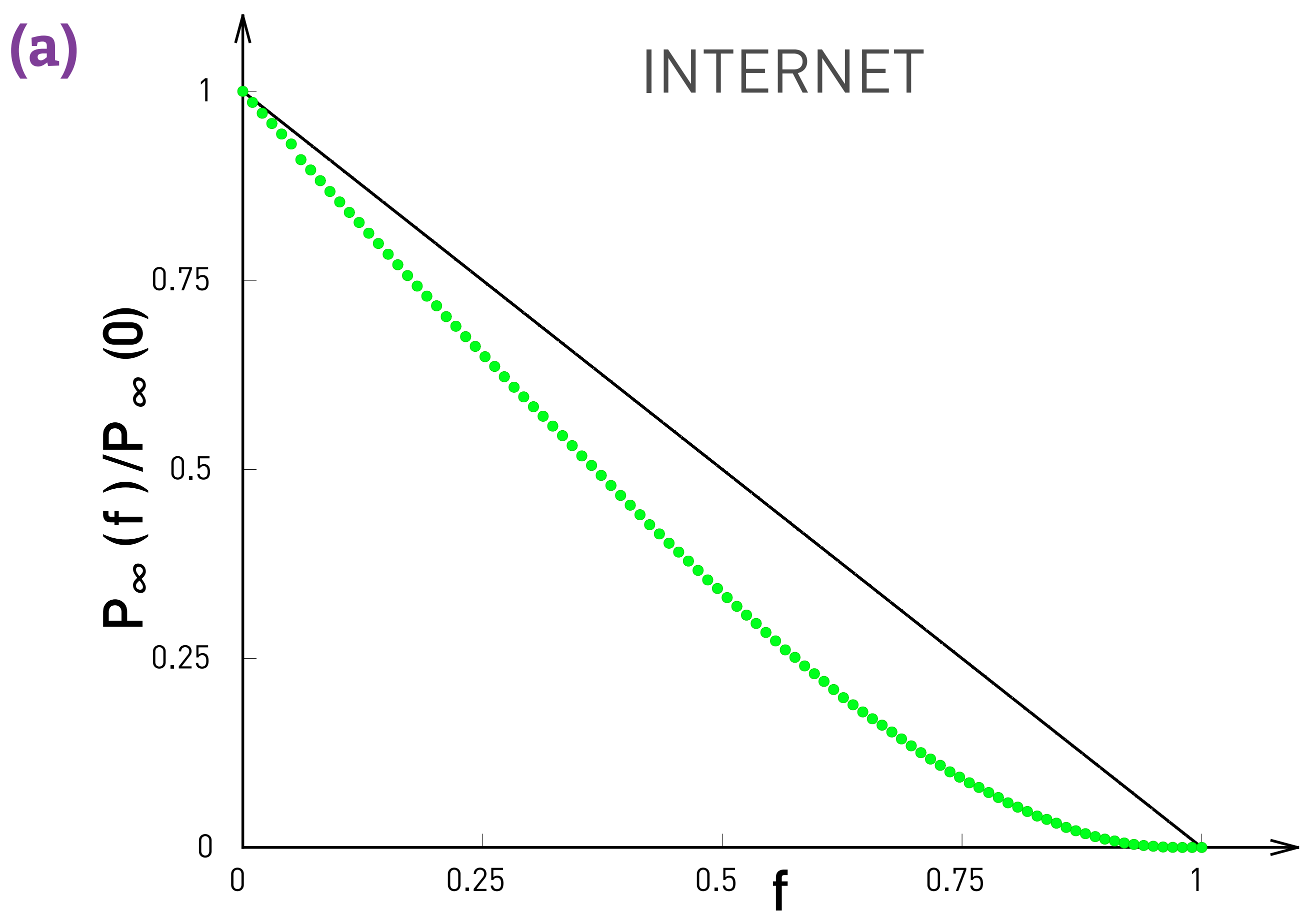
Fully burned down

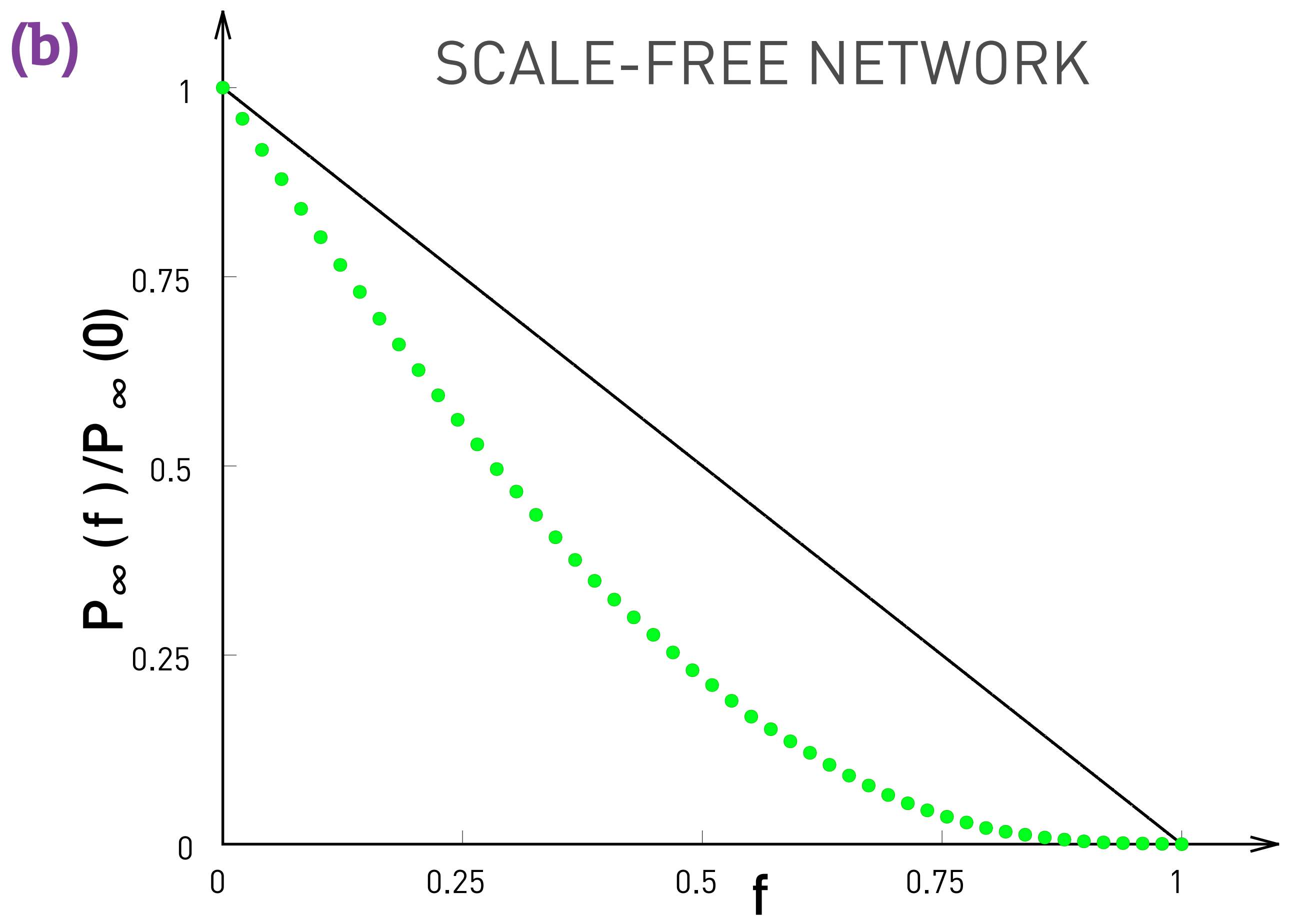


Robustness

What it looks like on (random) networks?

Internet more robust than lattice





When is there a GC?

Iff giant component GC , each node in GC must be connected to at least two other nodes on average. Hence, average degree k_i of node i attached to j in GC:

$$\langle k_i \mid i \leftrightarrow j \rangle = \sum_{k_i} k_i P(k_i \mid i \leftrightarrow j) = 2 \,.$$

$$P(k_i \mid i \leftrightarrow j) = \frac{P(k_i, i \leftrightarrow j)}{P(i \leftrightarrow j)} = \frac{P(i \leftrightarrow j \mid k_i)\, p(k_i)}{P(i \leftrightarrow j)}$$ Bayes

$$P(i \leftrightarrow j) = \frac{2L}{N(N-1)} = \frac{\langle k \rangle}{N-1}\,, \qquad P(i \leftrightarrow j \mid k_i) = \frac{k_i}{N-1}\,,$$ No deg corr

$$\sum_{k_i} k_i P(k_i \mid i \leftrightarrow j) = \sum_{k_i} k_i \frac{P(i \leftrightarrow j \mid k_i)\, p(k_i)}{P(i \leftrightarrow j)} = \sum_{k_i} k_i \frac{k_i p(k_i)}{\langle k \rangle} = \frac{\sum_{k_i} k_i^2 p(k_i)}{\langle k \rangle}$$

Molloy-Reeds criterion (once again!)

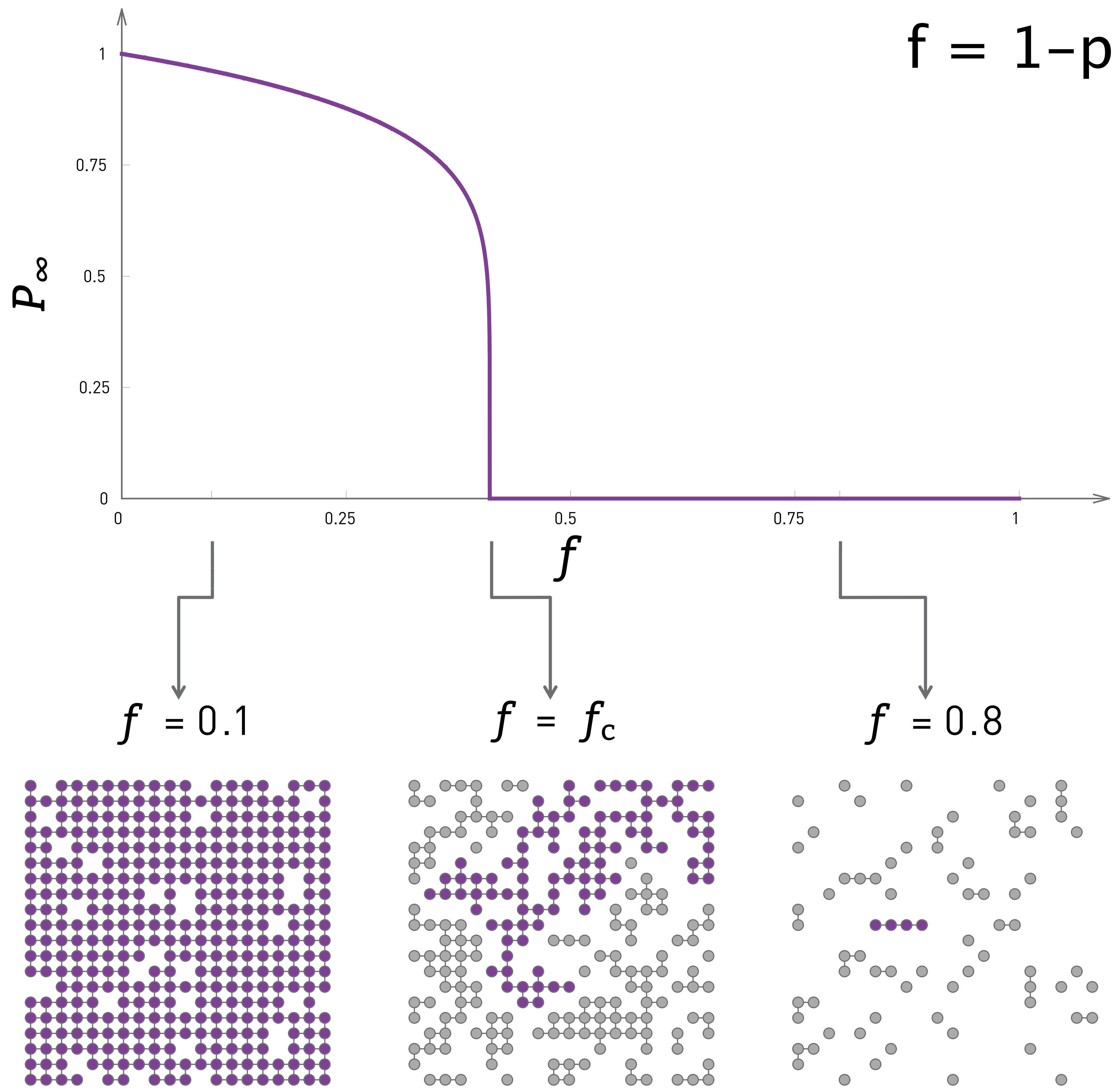
$$\kappa = \frac{\langle k^2 \rangle}{\langle k \rangle} > 2 \,.$$

$$\text{for Poisson } p_k \quad \rightarrow \quad \langle k^2 \rangle = \langle k \rangle^2 + \langle k \rangle \quad \rightarrow \quad \langle k \rangle > 1$$

Robustness

Random removal: find critical f

The problem becomes: does the damaged network still fulfil the Molloy-Reeds?



$0 < f < f_c$:
There is a giant component.
$P_\infty \sim |f - f_c|^\beta$

$f = f_c$:
The giant component vanishes.

$f > f_c$:
The lattice breaks into many tiny components.

Initial network: p(k), <k>, <k^2>

Final network?

Proba that node with degree k goes to k' with prob

$$\binom{k}{k'} f^{k-k'} (1-f)^{k'}$$

k' < k

Resulting degree distribution

$$p'(k') = \sum_k p(k) \binom{k}{k'} f^{k-k'} (1-f)^{k'}$$

Compute <k> and <k^2> for Molloy Reeds

Robustness

Random removal: average degree

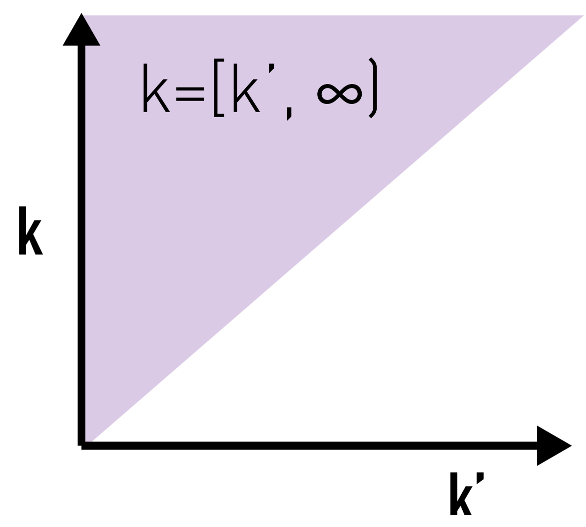
$$p'_{k'} = \sum_{k=k'}^{\infty} p_k \binom{k}{k'} f^{k-k'} (1-f)^{k'}.$$

$$\langle k' \rangle_f = \sum_{k'=0}^{\infty} k' p'_{k'}$$

$$= \sum_{k'=0}^{\infty} k' \sum_{k=k'}^{\infty} p_k \left(\frac{k!}{k'!(k-k')!} \right) f^{k-k'} (1-f)^{k'}$$

$$= \sum_{k'=0}^{\infty} \sum_{k'=k'}^{\infty} p_k \frac{k(k-1)!}{(k'-1)!(k-k')!} f^{k-k'} (1-f)^{k'-1} (1-f).$$

$$\sum_{k'=0}^{\infty} \sum_{k=k'}^{\infty} = \sum_{k=0}^{\infty} \sum_{k'=0}^{k}.$$



$$\langle k' \rangle_f = \sum_{k=0}^{\infty} k \sum_{k'=0}^{k} p_k \frac{k(k-1)!}{(k'-1)!(k-k')!} f^{k-k'} (1-f)^{k'-1} (1-f)$$

$$= \sum_{k=0}^{\infty} (1-f) k p_k \sum_{k'=0}^{k} \frac{(k-1)!}{(k'-1)!(k-k')!} f^{k-k'} (1-f)^{k'-1}$$

$$= \sum_{k=0}^{\infty} (1-f) k p_k \sum_{k'=0}^{k} \binom{k-1}{k'-1} f^{k-k'} (1-f)^{k'-1}$$

$$= \sum_{k=0}^{\infty} (1-f) k p_k$$

Sum of binomial over all possibilities = 1

$$= (1-f) \langle k \rangle.$$

Robustness

$$p'_{k'} = \sum_{k=k'}^{\infty} p_k \binom{k}{k'} f^{k-k'} (1-f)^{k'}.$$

Random removal: average degree

$$\langle k'^2 \rangle_f = \langle k'(k'-1) + k' \rangle_f$$

$$= \langle k'(k'-1) \rangle_f + \langle k' \rangle_f$$

$$= \sum_{k'=0}^{\infty} k'(k'-1) p'_{k'} + \langle k' \rangle_f .$$

$$\langle k'(k'-1) \rangle_f = \sum_{k'=0}^{\infty} k'(k'-1) p'_{k'}$$

$$= \sum_{k'=0}^{\infty} k'(k'-1) \sum_{k=k'}^{\infty} p_k \binom{k}{k'} f^{k-k'} (1-f)^{k'}$$

Robustness

Random removal: average degree

$$\langle k'^2 \rangle_f = \langle k'(k'-1) + k' \rangle_f$$

$$= \langle k'(k'-1) \rangle_f + \langle k' \rangle_f$$

$$= (1-f)^2 \langle k(k-1) \rangle + (1-f)\langle k \rangle$$

$$= (1-f)^2 \left(\langle k^2 \rangle - \langle k \rangle \right) + (1-f)\langle k \rangle$$

$$= (1-f)^2 \langle k^2 \rangle - (1-f)^2 \langle k \rangle + (1-f)\langle k \rangle$$

$$= (1-f)^2 \langle k^2 \rangle - \left(-f^2 + 2f - 1 + 1 - f \right)\langle k \rangle$$

$$= (1-f)^2 \langle k^2 \rangle + f(1-f)\langle k \rangle .$$

$$\langle k' \rangle_f = (1-f)\langle k \rangle$$

$$\langle k'^2 \rangle_f = (1-f)^2 \langle k^2 \rangle + f(1-f)\langle k \rangle$$

$$\kappa = \frac{\langle k'^2 \rangle_f}{\langle k' \rangle_f} = 2 .$$

So, for any Pk, with random removal

$$f_c = 1 - \frac{1}{\frac{\langle k^2 \rangle}{\langle k \rangle} - 1}$$

Only depends on 1st and 2nd moment of p(k)

$$f_c^{ER} = 1 - \frac{1}{\langle k \rangle} .$$

Denser
-> more robust

Robustness

Scale-free networks?

$$\langle k^m \rangle = (\gamma-1)k_{\min}^{\gamma-1}\int_{k_{min}}^{k_{\max}} k^{m-\gamma}dk = \frac{(\gamma-1)}{(m-\gamma+1)}k_{\min}^{\gamma-1}[k^{m-\gamma+1}]_{k_{\min}}^{k_{\max}}.$$

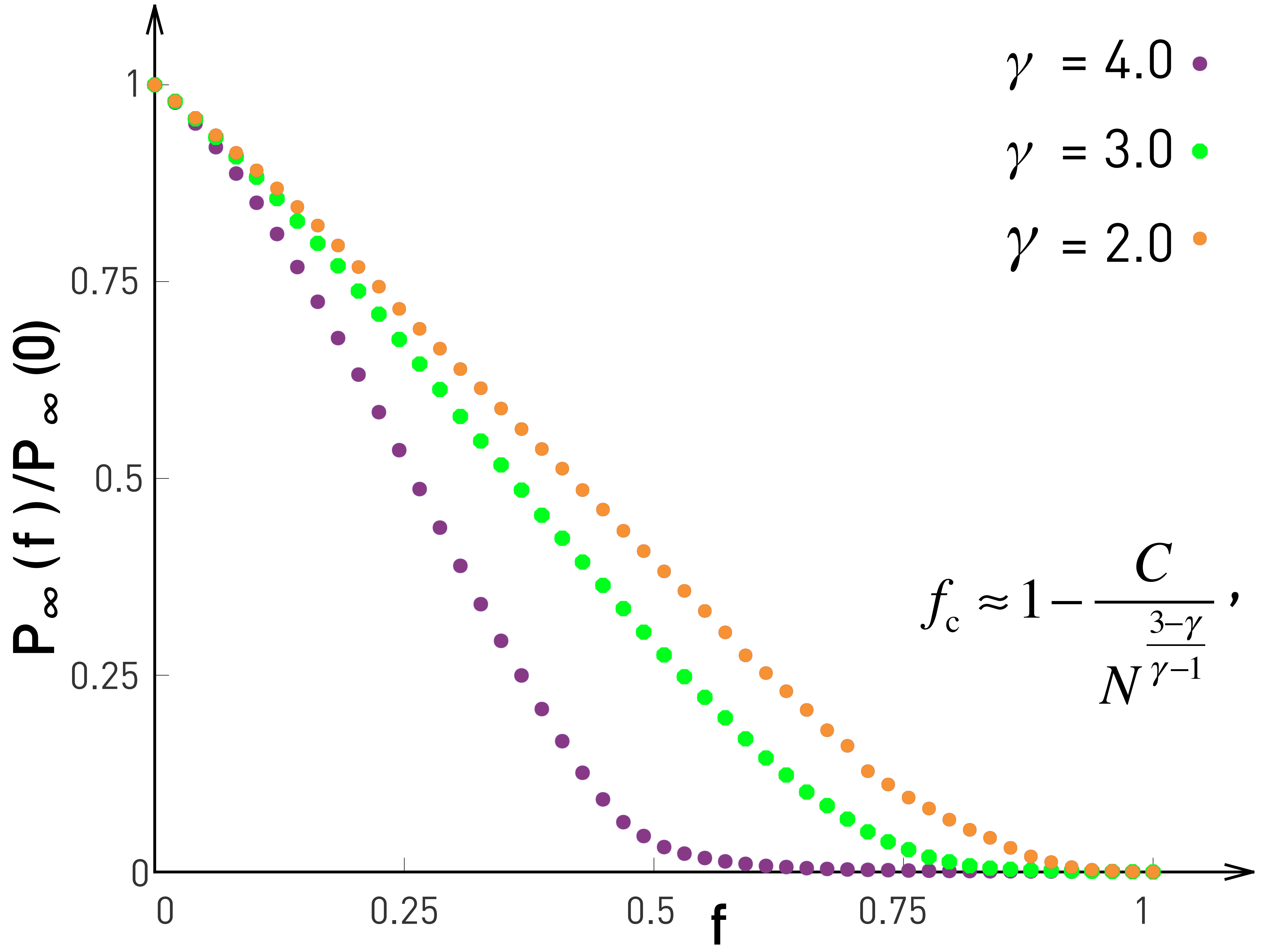
$$k_{\max} = k_{\min}N^{\frac{1}{\gamma-1}}$$

$$\langle k^m \rangle = \frac{(\gamma-1)}{(m-\gamma+1)}k_{\min}^{\gamma-1}[k_{\max}^{m-\gamma+1} - k_{\min}^{m-\gamma+1}].$$

$$\kappa = \frac{\langle k^2 \rangle}{\langle k \rangle} = \frac{(2-\gamma)}{(3-\gamma)}\frac{k_{\max}^{3-\gamma} - k_{\min}^{3-\gamma}}{k_{\max}^{2-\gamma} - k_{\min}^{2-\gamma}},$$

$$f_c = \begin{cases} 1 - \dfrac{1}{\frac{\gamma-2}{3-\gamma}k_{\min}^{\gamma-2}k_{\max}^{3-\gamma} - 1} & 2 < \gamma < 3 \\ 1 - \dfrac{1}{\frac{\gamma-2}{\gamma-3}k_{\min} - 1} & \gamma > 3 \end{cases}$$

-> 1 if N-> inf



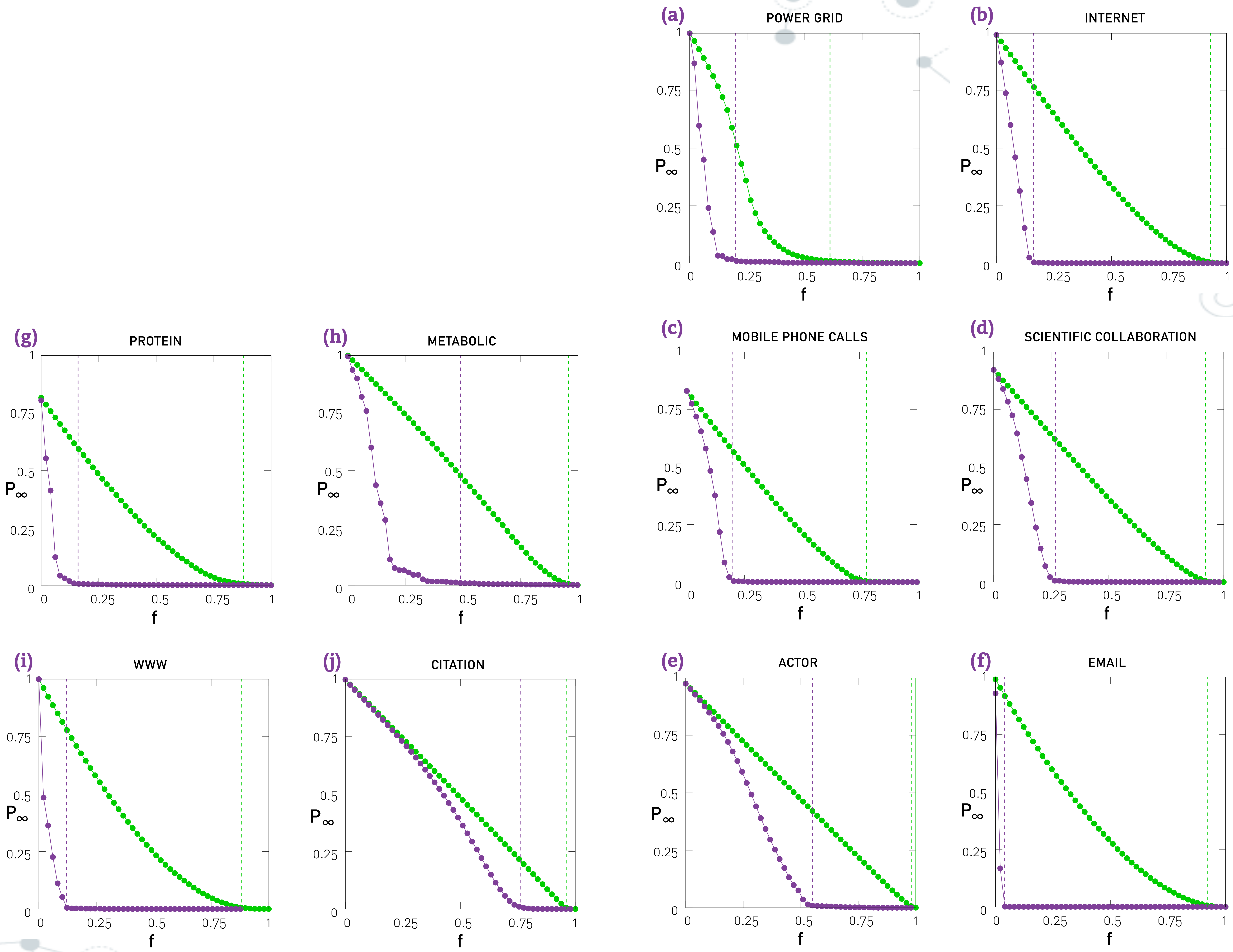
$$f_c \approx 1 - \frac{C}{N^{\frac{3-\gamma}{\gamma-1}}},$$

Removing lots of small nodes in SF

Attacks

Many types of attacks…

- **Degree**
- Betweenness
- Link-based analogues
- **Cascades…**



Attacks

Critical threshold

Hub removal

$$f = \int_{k'_{max}}^{k_{max}} p_k dk = \frac{\gamma-1}{\gamma-1} \frac{k'^{\,-\gamma+1}_{max} - k^{-\gamma+1}_{max}}{k^{-\gamma+1}_{min} - k^{-\gamma+1}_{max}}.$$

$$k_{max} \gg k'_{max} \gg k_{min}$$

$$f = \left(\frac{k'_{max}}{k_{min}}\right)^{-\gamma+1}, \quad k'_{max} = k_{min} f^{\frac{1}{1-\gamma}}.$$

$$\tilde{f} = f^{\frac{2-\gamma}{1-\gamma}}.$$

-> 1 if gamma = 2

Link removal

$$\tilde{f} = \frac{\int_{k'_{max}}^{k_{max}} k p_k dk}{\langle k \rangle} = \frac{1}{\langle k \rangle} c \int_{k'_{max}}^{k_{max}} k^{-\gamma+1} dk$$

$$= \frac{1}{\langle k \rangle} \frac{1-\gamma}{2-\gamma} \frac{k'^{\,-\gamma+2}_{max} - k^{-\gamma+2}_{max}}{k_{min}^{\,-\gamma+1} - k^{-\gamma+2}_{max}}.$$

$$\langle k \rangle \approx \frac{\gamma-1}{\gamma-2} k_{min}$$

$$\tilde{f} = \left(\frac{k'_{max}}{k_{min}}\right)^{-\gamma+2}.$$

Attacks

Critical threshold

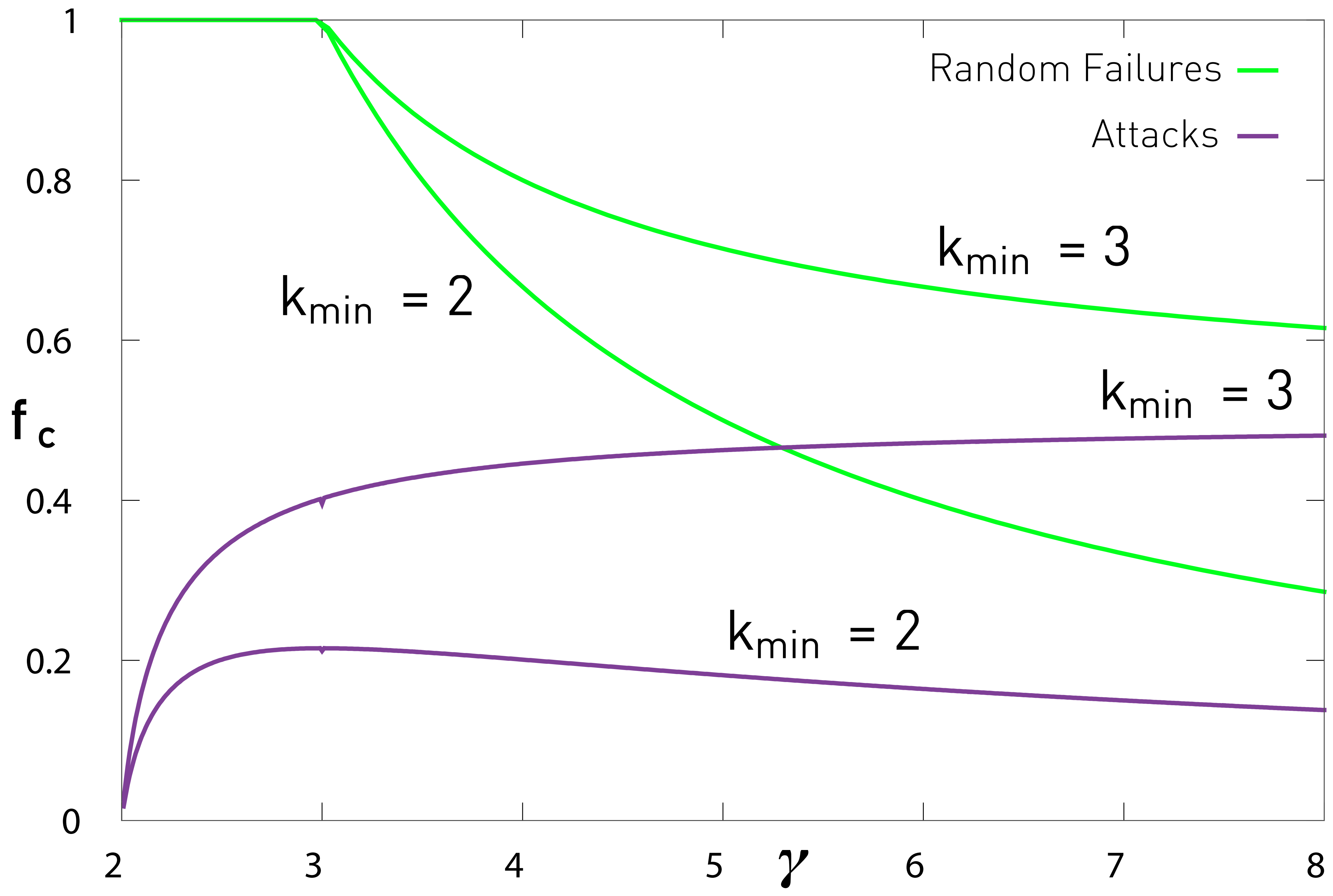
$$p'_{k'} = \sum_{k=k_{\min}}^{k'_{\max}} \binom{k}{k'} \tilde{f}^{\,k-k'} (1-\tilde{f})^{k'} p_k .$$

$$\langle k^m \rangle = \frac{(\gamma-1)}{(m-\gamma+1)} k_{\min}^{\gamma-1} [k_{\max}^{m-\gamma+1} - k_{\min}^{m-\gamma+1}] \cdot$$

$$\kappa = \frac{2-\gamma}{3-\gamma} \frac{k'^{\,3-\gamma}_{\max} - k^{3-\gamma}_{\min}}{k'^{\,2-\gamma}_{\max} - k^{2-\gamma}_{\min}} .$$

$$\kappa = \frac{2-\gamma}{3-\gamma} \frac{k^{3-\gamma}_{\min} f^{(3-\gamma)/(1-\gamma)} - k^{3-\gamma}_{\min}}{k^{2-\gamma}_{\min} f^{(2-\gamma)/(1-\gamma)} - k^{2-\gamma}_{\min}} = \frac{2-\gamma}{3-\gamma} k_{\min} \frac{f^{(3-\gamma)/(1-\gamma)} - 1}{f^{(2-\gamma)/(1-\gamma)} - 1} .$$

$$f_c^{\frac{2-\gamma}{1-\gamma}} = 2 + \frac{2-\gamma}{3-\gamma} k_{\min} \left(f_c^{\frac{3-\gamma}{1-\gamma}} - 1 \right)$$



Errors and attacks

Real networks

NETWORK	RANDOM FAILURES (REAL NETWORK)	RANDOM FAILURES (RANDOMIZED NETWORK)	ATTACK (REAL NETWORK)
Internet	0.92	0.84	0.16
WWW	0.88	0.85	0.12
Power Grid	0.61	0.63	0.20
Mobile-Phone Call	0.78	0.68	0.20
Email	0.92	0.69	0.04
Science Collaboration	0.92	0.88	0.27
Actor Network	0.98	0.99	0.55
Citation Network	0.96	0.95	0.76
E. Coli Metabolism	0.96	0.90	0.49
Yeast Protein Interactions	0.88	0.66	0.06

Published: 27 July 2000

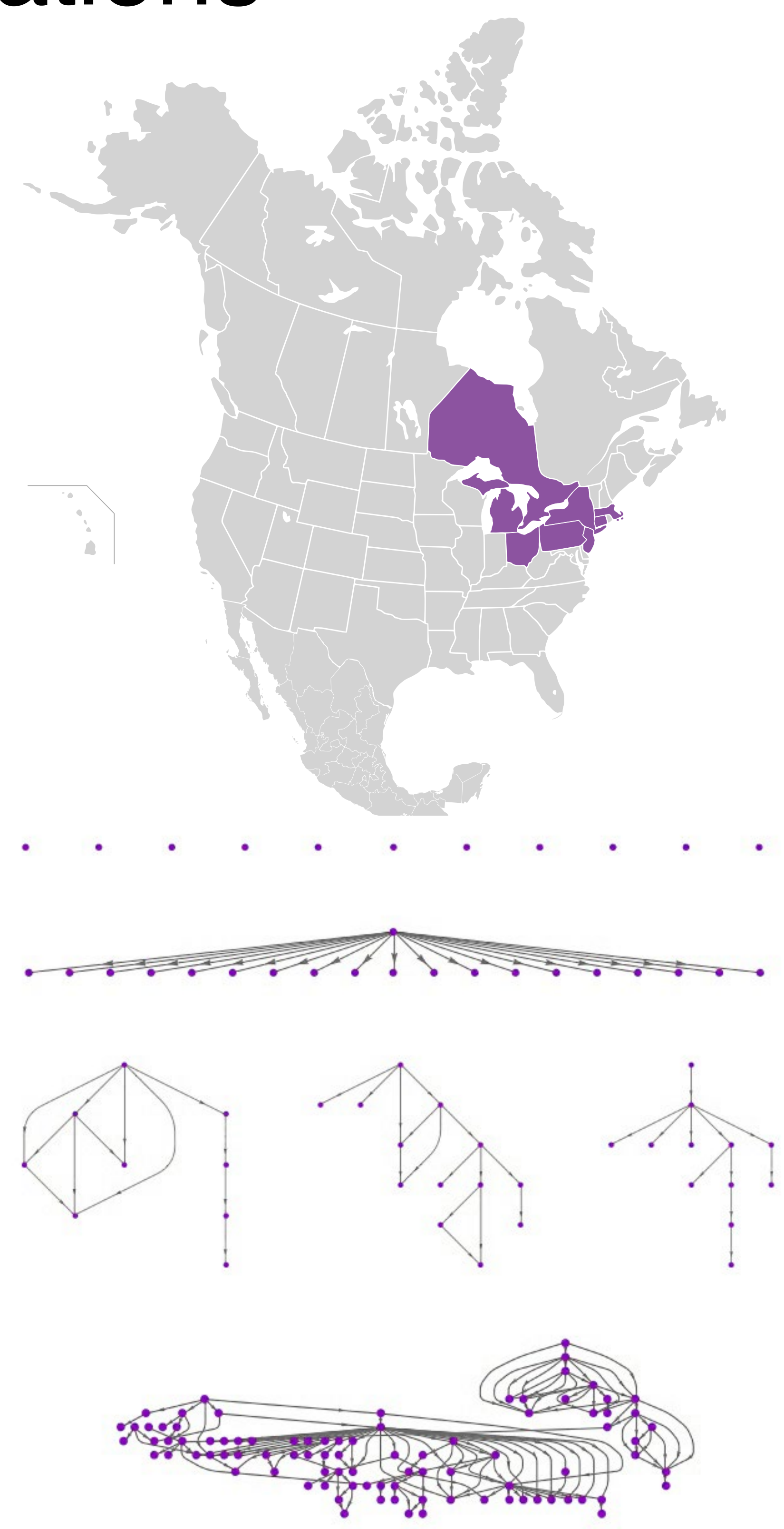
Error and attack tolerance of complex networks

Réka Albert, Hawoong Jeong & Albert-László Barabási

Nature **406**, 378–382 (2000) | Cite this article

Cascades

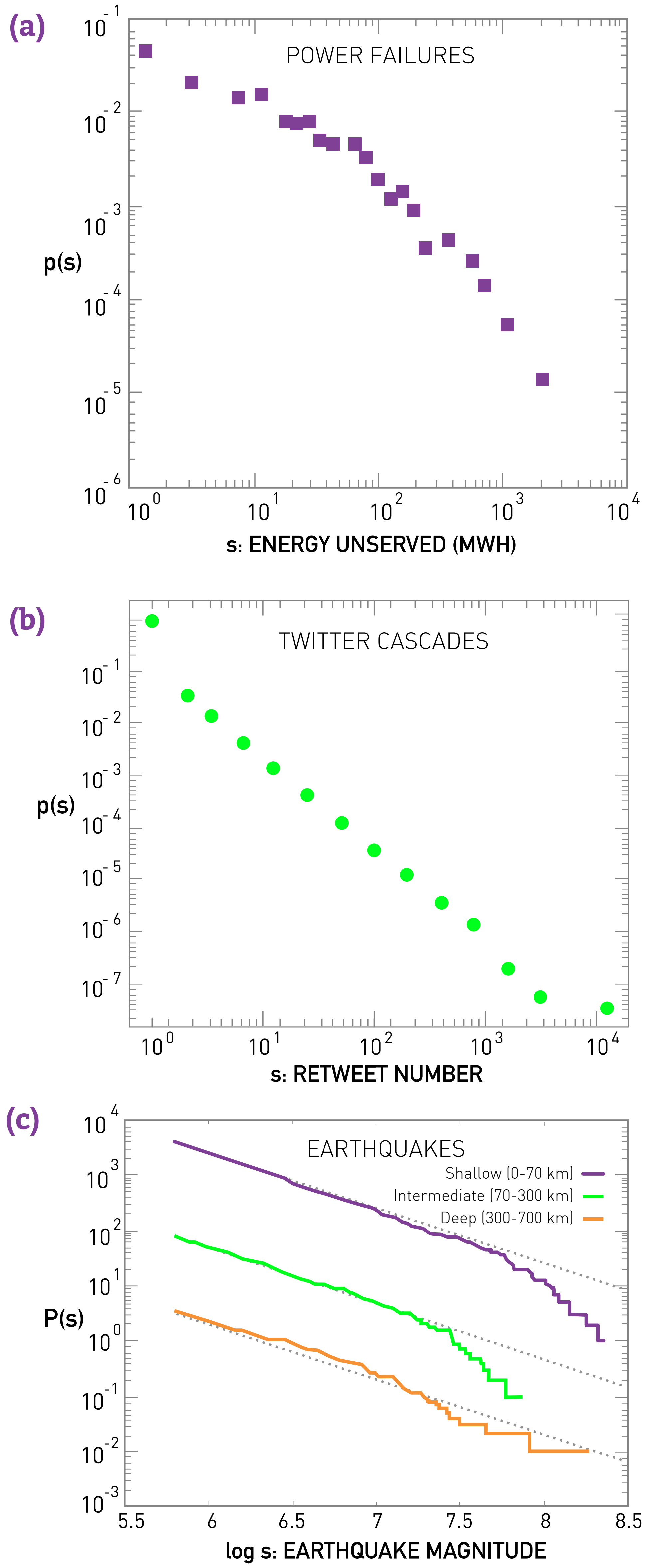
Observations



SOURCE	EXPONENT	CASCADE
Power grid (North America)	2.0	Power
Power grid (Sweden)	1.6	Energy
Power grid (Norway)	1.7	Power
Power grid (New Zealand)	1.6	Energy
Power grid (China)	1.8	Energy
Twitter Cascades	1.75	Retweets
Earthquakes	1.67	Seismic Wave

Avalance exponent

$$p(s) \sim s^{-\alpha},$$



"Hard" Model

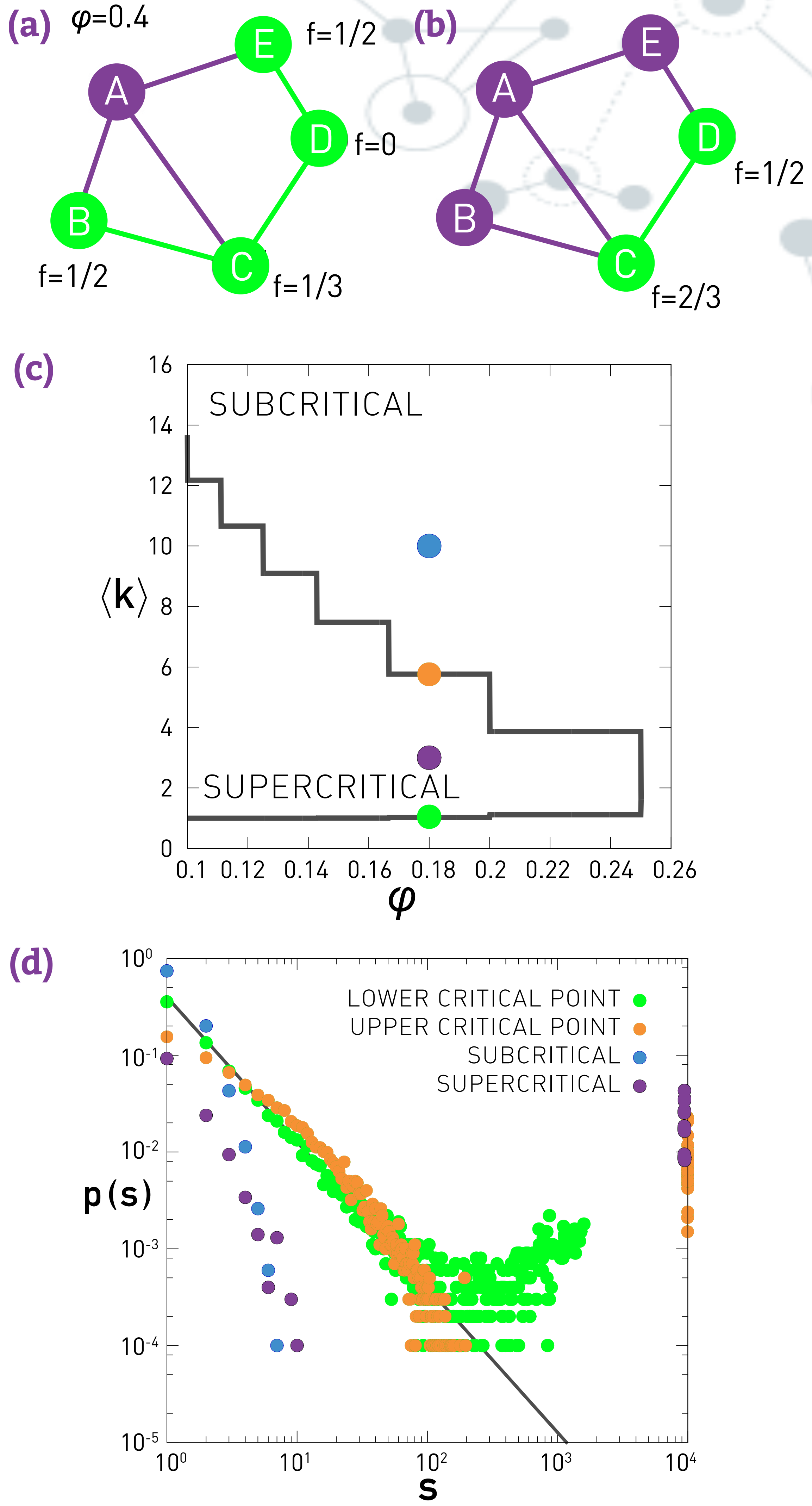
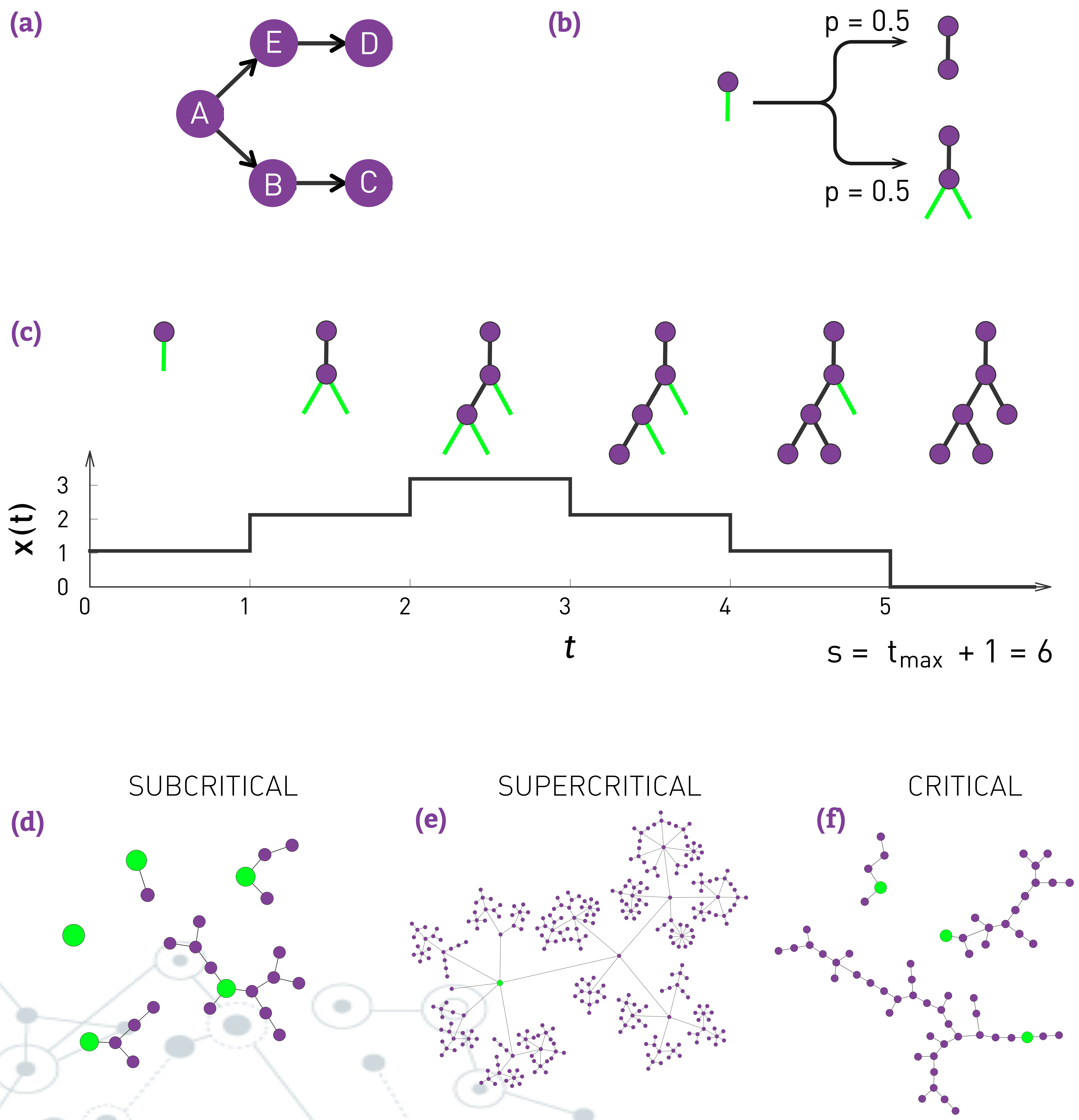


Figure 8.20
Failure Propagation Model

Cascades

Easier model



The branching model can be solved analytically, allowing us to determine the avalanche size distribution for an arbitrary p_k. If p_k is exponentially bounded, *e.g.* it has an exponential tail, the calculations predict $\alpha = 3/2$. If, however, p_k is scale-free, then the avalanche exponent depends on the power-law exponent γ, following (Figure 8.22) [32, 33]

$$\alpha = \begin{cases} 3/2, & \gamma \geq 3 \\ \gamma/(\gamma-1), & 2 < \gamma < 3 \,. \end{cases} \quad (8.15)$$

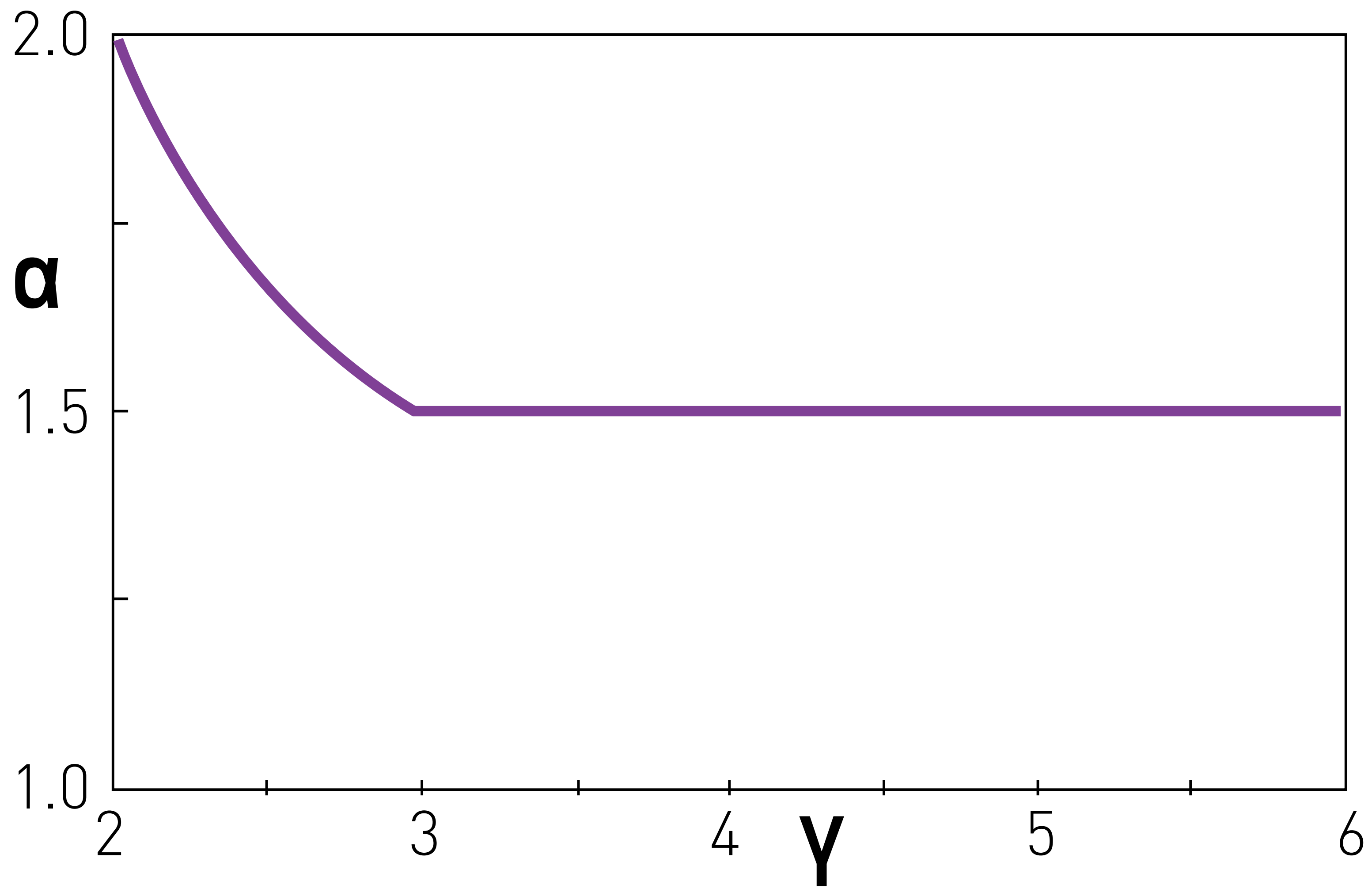


Figure 8.22

The Avalanche Exponent

Pause

Walks in networks

- Simple process, well studied in finite dimensional lattices
- Simplest way to explore or search in a network
- Basic element of diffusion processes
- Basis of PageRank

Num walkers $$W = \sum_i W_i$$ Number walkers on node i

Markovian random walker $$d_{ij} = \frac{r}{k_i}$$ Rate out of i to j

Total diffusion rate Out of I $$r = \sum_{j \sim i} d_{ij}$$

Hypothesis = statistical equivalence of nodes with the same degree

Degree-block Variables $$W_k = \frac{1}{N_k} \sum_{i | k_i = k} W_i$$

Diffusion equation $$\partial_t W_k(t) = -rW_k(t) + k \sum_{k'} p(k' | k) \frac{r}{k'} W_{k'}(t)$$

P: probability of having neighbour with degree k'

Walks in networks

$$\partial_t W_k(t) = -rW_k(t) + k\sum_{k'} p(k'|k)\frac{r}{k'}W_{k'}(t)$$

Uncorr. networks $p(k'|k) = \frac{k'p(k')}{\langle k \rangle}$

$$\partial_t W_k(t) = -rW_k(t) + \frac{k}{\langle k \rangle} r \sum_{k'} p(k')W_{k'}(t)$$

Stationary solution: $\partial_t W_k(t) = 0$

using $\sum_k p(k)W_k = W/N$

$$W_k = \frac{k}{\langle k \rangle}\frac{W}{N}$$

Finally, probability to find one walker in degree k: $p_k = \frac{W_k}{W} = \frac{k}{\langle k \rangle}\frac{1}{N} \propto k$

Walks in (directed) networks

PageRank

Previous ranking: crawl around a starting page and return the ranking based on the #matches to word query, index, etc
The PageRank algorithm: major breakthrough based on idea that ranking depends on network topology
“Google” defines the importance of each document by a combination of the probability that a random walker surfing the web will visit that document, and some heuristics based in the text disposition, [cit. Barrat/Barth/Vesp]

Probability that a random walker will visit page i:

$$P_R(i) = \frac{q}{N} + (1-q)\sum_j x_{ij}\frac{P_R(j)}{k_{out,j}}$$

q = **damping / teleportation**

x_ij: adjacency

Brin and Page, 1998

Degree-block variables $k = (k_{in}, k_{out})$

$$P_R(k) = \frac{1}{N_k}\sum_{i\in k} P_R(i)$$

$$P_R(k) = \frac{q}{N} + \frac{(1-q)}{N_k}\sum_{i\in k}\sum_{k'}\frac{1}{k'_{out}}\sum_{j\in k'} x_{ij}P_R(j)$$

Mean-field approx: $P_R(j) = P(k)$

$$P_R(k) = \frac{q}{N} + \frac{(1-q)}{N_k}\sum_{k'}\frac{P_R(k')}{k'_{out}}\sum_{i\in k}\sum_{j\in k'} x_{ij} = \frac{q}{N} + \frac{(1-q)}{N_k}\sum_{k'}\frac{P_R(k')}{k'_{out}}E_{k'\to k}$$

Walks in networks

PageRank

$$P_R(k) = \frac{q}{N} + \frac{(1-q)}{N_k} \sum_{k'} \frac{P_R(k')}{k'_{out}} E_{k' \to k}$$

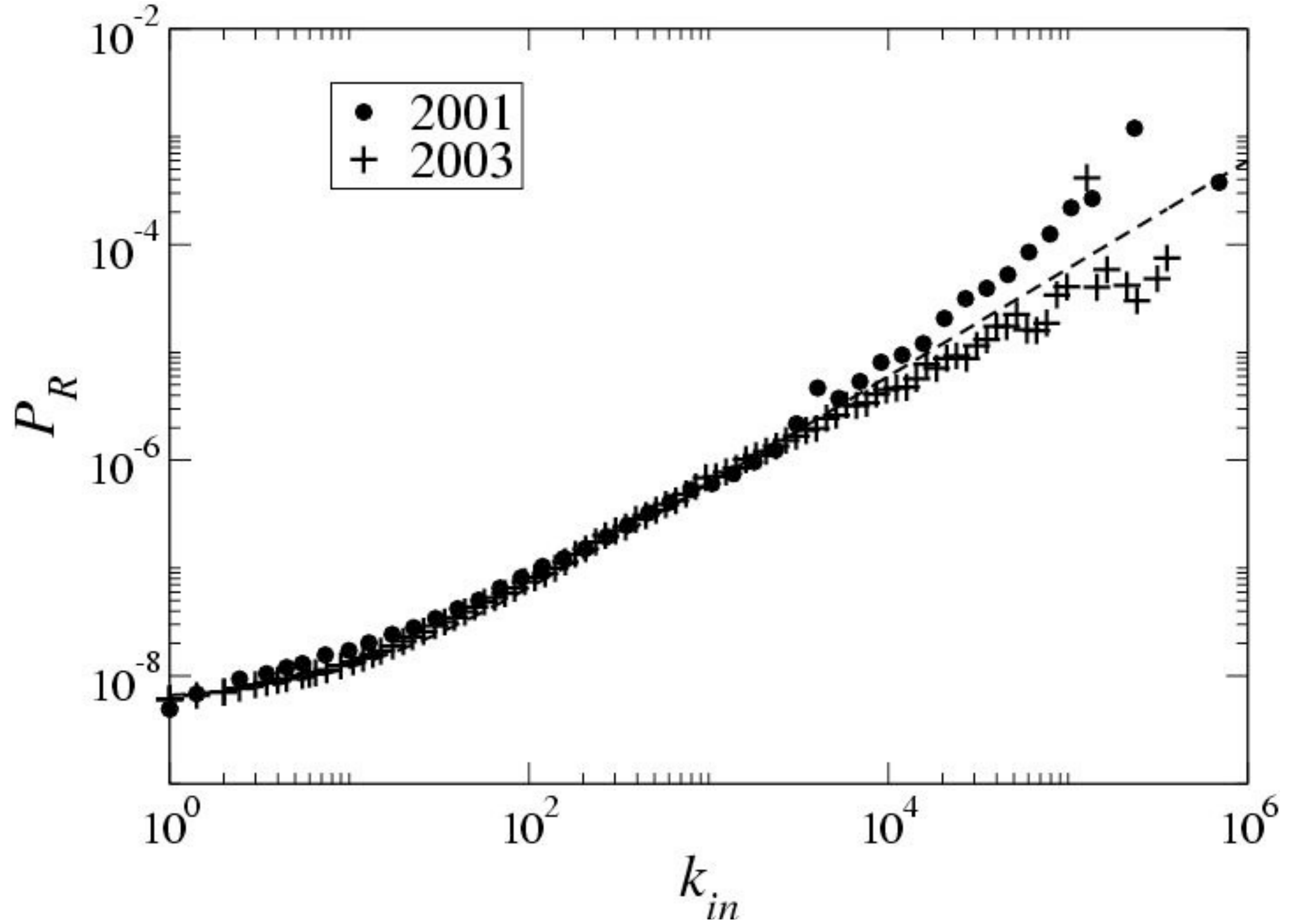
$$E_{\mathbf{k}' \to \mathbf{k}} = k_{in} P(\mathbf{k}) N \frac{E_{\mathbf{k}' \to \mathbf{k}}}{k_{in} P(\mathbf{k}) N} = k_{in} P(\mathbf{k}) N P_{in}(\mathbf{k}'|\mathbf{k}),$$

Uncorr. networks

$$P_{in}(\mathbf{k}'|\mathbf{k}) = \frac{k'_{out} P(\mathbf{k}')}{\langle k_{in} \rangle}$$

$$P_R(k) = \frac{q}{N} + (1-q)\frac{k_{in}}{\langle k_{in} \rangle} \sum_{k'} P_R(k')P(k') = \frac{q}{N} + \frac{(1-q)}{N} \frac{k_{in}}{\langle k_{in} \rangle}$$

Data from the Web



PageRank: also to quantify importance of scientific papers

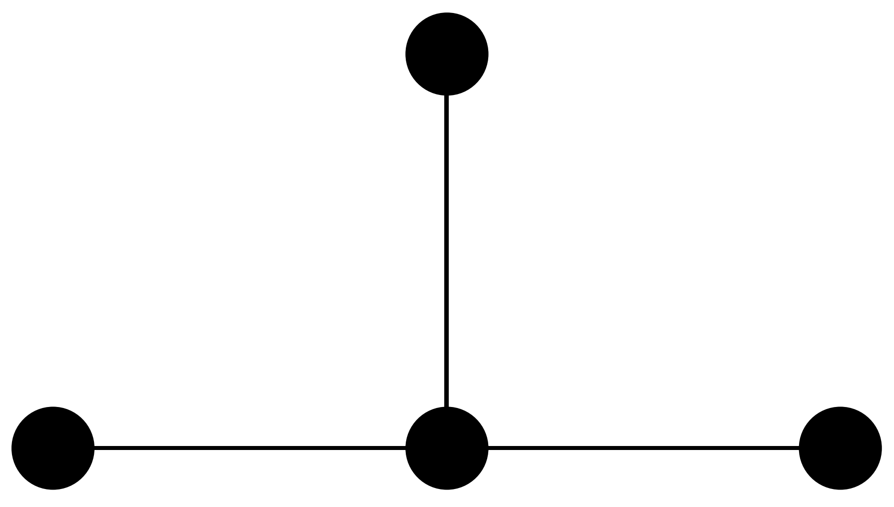
Walks in networks

Laplacian

$$\Delta\phi(v) = \sum_{w \in \nu_{(v)}} (\phi(w) - \phi(v)) .$$

$$\mathbf{L} = \mathbf{D} - \mathbf{X} \qquad D_{ij} = \delta_{ij} k_i \qquad X_{ij} = \textbf{adj}$$

$$L_{ii} = k_i$$
$$L_{ij} = -x_{ij}$$



S_4

$$L(S_4) = \begin{pmatrix} 3 & -1 & -1 & -1 \\ -1 & 1 & 0 & 0 \\ -1 & 0 & 1 & 0 \\ -1 & 0 & 0 & 1 \end{pmatrix}$$

Undirected graphs:

- on infinite square lattice == continuous Laplacian
- L symmetric ==> spectrum is positive semidefinite $0 \leq \lambda_1 \leq \lambda_2 \leq \ldots \leq \lambda_N$
- The multiplicity of *0* as an eigenvalue of *L* is equal to the number of connected components of the graph.
- The second smallest eigenvalue λ_1 is called the algebraic connectivity.
  It is non-zero only if the graph is formed of a single connected component.

Walks in networks

Laplacian and return times

$$\partial_t p(i,t|i_0,0) = -\sum_j L_{ij}\, p(j,t|i_0,0) \qquad p(i,0|i_0,0) \;=\; \delta_{ii_0} \qquad \sum_i L_{ij} = 0 \;\;\rightarrow\;\; \sum_i p(i,t\,|\,i_0,0) = 1$$

Spectral density

$$\rho(\lambda) = \left\langle \frac{1}{N}\sum_{i=1}^{N}\delta(\lambda-\lambda_i)\right\rangle$$

Laplace transform

$$\tilde{p}_{ii_0}(s) = \int_0^\infty \mathrm{d}t e^{-st}\, p(i,t|i_0,0).$$

Integration by parts

$$\int_0^\infty \mathrm{d}t e^{-st}\,\partial_t p(i,t|i_0,0) = -p(i,0|i_0,0) + s\,\tilde{p}_{ii_0}(s) = s\,\tilde{p}_{ii_0}(s) - \delta_{ii_0}$$

Finally, rewrite as

$$s\tilde{p}_{ii_0}(s) - \delta_{ii_0} = -\sum_j L_{ij}\,\tilde{p}_{ji_0}(s) \qquad \sum_j \left(s\delta_{ij} + L_{ij}\right)\tilde{p}_{ji_0}(s) = \delta_{ii_0}.$$

Return time, definition

$$p_0(t) = \left\langle \frac{1}{N}\sum_{i_0} p(i_0,t|i_0,0)\right\rangle.$$

Laplace transform

$$\tilde{p}_0(s) = \left\langle \frac{1}{N}\sum_{i_0}\tilde{p}_{i_0 i_0}(s)\right\rangle = \left\langle \frac{1}{N}\mathrm{Tr}\,\tilde{\mathbf{p}}(s)\right\rangle, \qquad \tilde{p}_0(s) = \left\langle \frac{1}{N}\sum_i \frac{1}{s+\lambda_i}\right\rangle.$$

Laplace inverse transform

$$p_0(t) \;=\; \int_{c-\mathrm{i}\infty}^{c+\mathrm{i}\infty} \mathrm{d}s e^{ts}\left\langle \frac{1}{N}\sum_j \frac{1}{s+\lambda_j}\right\rangle \;=\; \left\langle \frac{1}{N}\sum_j \mathrm{e}^{-\lambda_j t}\right\rangle,$$

Or equivalently

$$p_0(t) = \int_0^\infty \mathrm{d}\lambda \mathrm{e}^{-\lambda t}\rho(\lambda).$$

Searching in networks

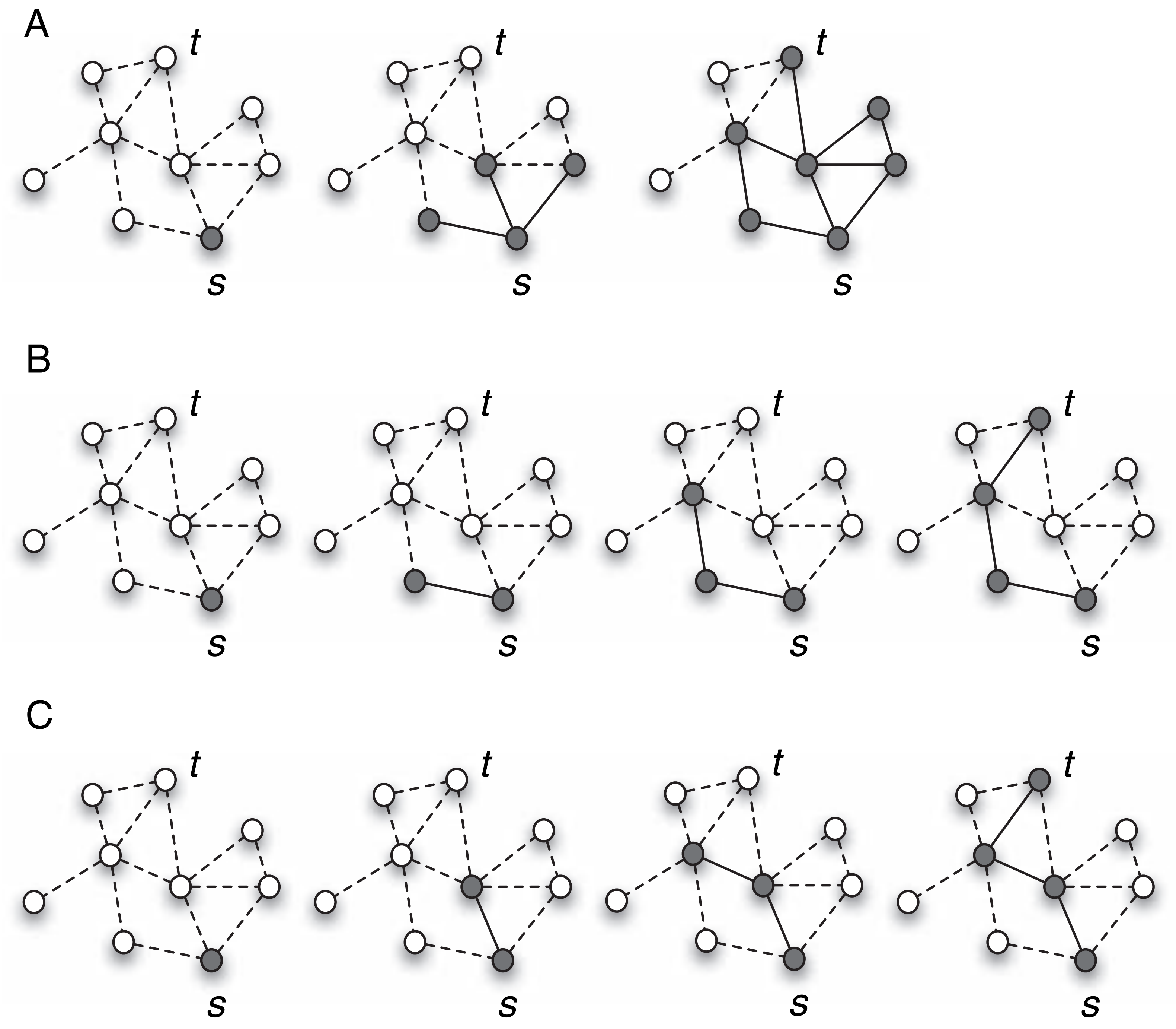


Fig. 8.4. Schematic comparison of various searching strategies to find the target vertex t, starting from the source s. A, Broadcast search; B, Random walk; C, Degree-biased strategy. The broadcast search finds the shortest path, at the expense of high traffic.

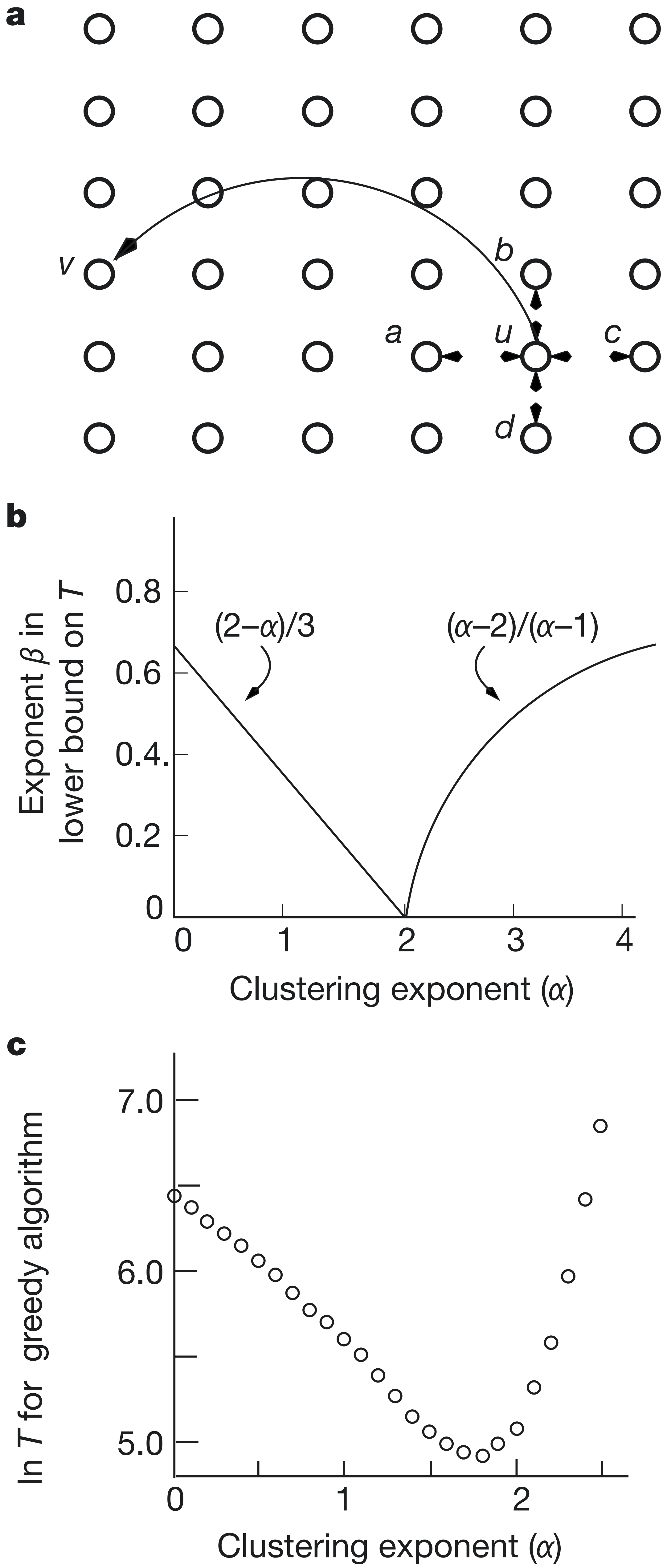
Finds shortest path, generated traffic: traffic $\propto N$

Not shortest path, less traffic: $T \propto N^{0.79} \quad \gamma = 2.1$

Peer-to-Peer networks, Adamic et al. (2001)

What about this?

Searching in small world networks



- Start from a D-dimensional hypercubic lattice:
- add a link node i, and connect to node j with at geographical distance r_{ij} with prob $\sim r^{-\alpha}$,
- Each node knows its own position and that of its neighbours.
- Greedy search process:
    - a message has to be sent to a certain target node t whose geographical position is known.
    - A node i receiving the message forwards it to the neighbor node j geographically closest to the target $(\min r_{jt})$

Kleinberg (2000a)
if $\alpha = D$, the delivery time scales as log^2(N) with the size N of the network.

What is the dimension then of real networks? Are they navigable?
Barrat/Barth/Vesp Chapter 8

What did we talk about today?

- Formalism of robustness
- Errors vs attacks
- Cascades (qualitatively)

- Walks and random walkers
- Pagerank
- Introduction to Laplacian

What didn’t we talk about today?

- Math-y Cascades …
- How to engineer robustness?
- Deeper Laplacian spectral theory

- Calculations of the spectral densities
- Specific search results
- Dynamical systems in general