

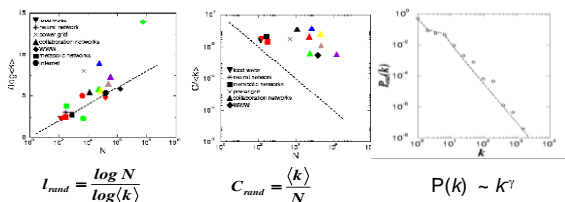
1

Today's Lecture

1. **Barabasi – Albert Model** of preferential attachment – how do we get scale free networks?
2. **Robustness** of networks – why is it important?
3. **Centrality Measures** – How important are certain nodes?

2

Empirical findings for real networks



Small World:
distances scale logarithmically with the network size

Clustered:
clustering coefficient does not depend on network size.

Scale-free:
The degrees follow a power-laws distribution.

Network Science: Evolving Network Models January 2012

3

BENCHMARK 2: Random Network Model

Erdős-Rényi Model- Publ. Math. Debrecen 6, 290 (1959)



Degree distribution:

$$P_{rand}(k) \cong C_{N-1}^k p^k (1-p)^{N-1-k}$$

Path length:

$$l_{rand} \approx \frac{\log N}{\log \langle k \rangle}$$

Clustering coefficient:

$$C_{rand} = p = \frac{\langle k \rangle}{N}$$

Network Science: Evolving Network Models January 2012

4

BENCHMARK 3: Small World Model

Watts-Strogatz algorithm – Nature 2008



- fixed node number N
- connecting pairs of nodes with probability p

Degree distribution:

Exponential

Path length:

$$l_{rand} \approx \frac{\log N}{\log \langle k \rangle}$$

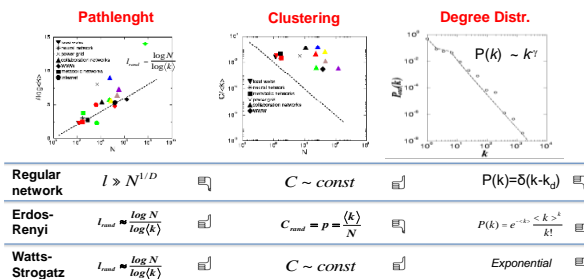
Clustering coefficient:

$$C(l) = C(0)(1 - b)$$

Network Science: Evolving Network Models January 2012

5

EMPIRICAL DATA FOR REAL NETWORKS



Network Science: Evolving Network Models January 2012

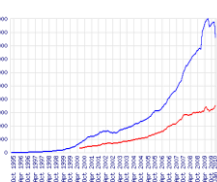
6

SCALE-FREE MODEL (BA model)

7

BA MODEL: Growth (www/Pubs)

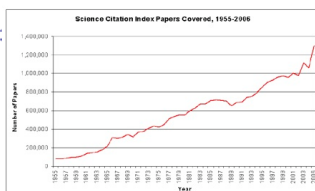
www



<http://website101.com/define-ecommerce-web-terms-definitions/>

Barahási & Albert. *Science* **286**, 509 (1999)

Scientific Publications



http://www.kk.org/thetechnium/archives/2008/10/the_expansion_o.php

9

BA MODEL

To ~~find~~ find shortest path from one to another

Dijkstra finds shortest path from source to target in weighted graph

works if all edges are non-negative

Algorithm

1. Initialize distance to infinity for all vertices except source which is 0

2. Select the vertex with minimum distance from source among unvisited vertices and mark it as visited

3. Update the distance of all unvisited vertices adjacent to the selected vertex if the new distance is less than the current distance

4. Repeat steps 2 and 3 until all vertices are visited

5. The distance of the target vertex is the shortest path from source to target

6. Return the shortest path from source to target

7. End

[illegible]

11

BA MODEL: Growth

ER, WS models: the number of nodes, N , is fixed (static models)

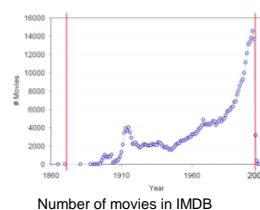
Real networks continuously expand by the addition of new nodes

Barabási & Albert, *Science* **286**, 509 (1999)

8

BA MODEL: Growth (Actors/Internet)

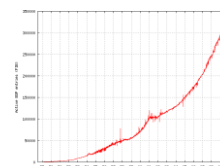
Actor network



Herr II, Bruce W., Ke, Weimao, Hardy, Elisha, and Börner, Katy. (2007) Movies and Actors: Mapping the Internet Movie Database. In Conference Proceedings of 11th Annual Information Visualization International Conference (IV 2007), Zurich, Switzerland, July 4-6, pp. 465-469.

10

Internet



Growth of the Internet routing table

<http://www.trainsignaltraining.com/ccna-ipv6>

BA MODEL: Preferential Attachment

(A) Growth: At each timestep we add a new node with m ($\leq m_0$) links that connects the new node to m nodes already in the network.

(B) Preferential attachment: The probability $\Pi(k_i)$ that a link of the new node connects to node i depends on the degree k_i of node i as:

$$\Pi(k_i) = \frac{k_i}{\sum_i k_i}$$

we need both!

After t timesteps the Barabási-Albert model generates a network with $N = t + m_0$ nodes and $m_0 + mt$ links

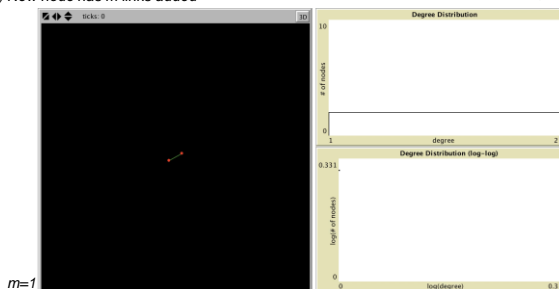
12

BA MODEL

(1) Networks continuously expand by the addition of new nodes

(2) New node has m links added

Barabási & Albert, Science **286**, 509 (1999)

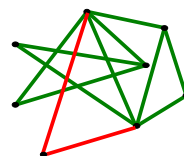


13

BA MODEL: Growth

(1) Networks continuously expand by the addition of new nodes

Add a new node with m links



Barabási & Albert, Science **286**, 509 (1999)

Network Science: Evolving Network Models January 2012

14

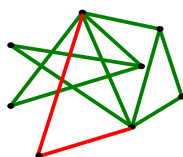
BA MODEL: Preferential Attachment

Where will the new node link to?
ER, WS models: choose randomly.

New nodes prefer to link to highly connected nodes (www, citations, IMDB).

PREFERENTIAL ATTACHMENT:

the probability that a node connects to a node with k links is proportional to k .



Barabási & Albert, Science **286**, 509 (1999)

$$\Pi(k_i) = \frac{k_i}{\sum_j k_j}$$

Network Science: Evolving Network Models January 2012

15

Origin of SF networks: Growth and preferential attachment

(1) Networks continuously expand by the addition of new nodes

WWW: addition of new documents

(2) New nodes prefer to link to highly connected nodes.

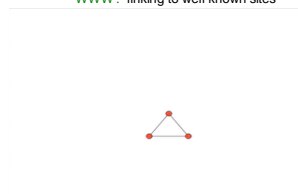
WWW: linking to well known sites

GROWTH:

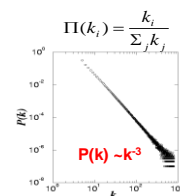
add a new node with m links

PREFERENTIAL ATTACHMENT:

the probability that a node connects to a node with k links is proportional to k .



Barabási & Albert, Science **286**, 509 (1999)



Network Science: Evolving Network Models January 2012

16

BA model I

GROWTH:

add a new node with m links

PREFERENTIAL ATTACHMENT:

the probability that a node connects to a node with k links is proportional to k .

$$\Pi(k_i) = \frac{k_i}{\sum_j k_j}$$

After t time steps in BA model

$$N = t + m_0 \text{ nodes and } m_0 + mt \text{ links}$$

Network Science: Evolving Network Models January 2012

17

BA model II

Assume k_i is a time dependent continuous variable, the rate at which each node i acquires links follows:

$$\frac{\partial k_i}{\partial t} = m \Pi(k_i) = m \frac{k_i}{\sum_j k_j}$$

m describes that every node arrives with m links, hence node j has m chances to be Chosen. The sum above goes over all nodes in the network except the newly added Nodes

$$\sum_j k_j = 2mt - m$$

The top equation becomes

$$\frac{\partial k_i}{\partial t} = \frac{k_i}{2t - 1}$$

Network Science: Evolving Network Models January 2012

18

BA model III

$$\frac{\partial k_i}{\partial t} = \frac{k_i}{2t-1}$$

We neglect -1 for $t \gg 1$

$$\frac{\partial k_i}{\partial t} \equiv \frac{k_i}{2t} \quad \text{or} \quad \frac{\partial k_i}{k_i} = \frac{1}{2} \frac{\partial t}{t}$$

We can integrate

$$k_i(t) = m \left(\frac{t}{t_i} \right)^\beta, \quad \beta = \frac{1}{2}$$

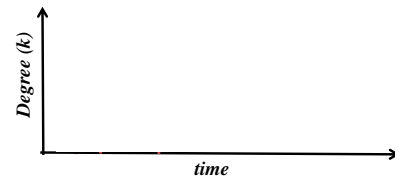
Meaning node i joins the network at time t_i with m links

Network Science: Evolving Network Models January 2012

19

Fitness Model: Can Latecomers Make It?

SF model: $k(t) \sim t^{-1/2}$ (first mover advantage)



Network Science: Evolving Network Models January 2012

20

Degree distribution I

The probability that degree $k_i(t)$ of node i is smaller than a value of k :

$$P(k_i(t) < k) = P(t_i > \frac{m^\beta t}{k^\beta})$$

In the model we add nodes at equal time intervals, we write the probability that a node arrives at time t_i as a random variable with a constant probability density

$$P(t_i) = \frac{1}{m_0 + t}$$

Substitute this in top equation we obtain a cumulative distribution function:

$$P(k) = P(t_i \leq \frac{m^\beta t}{k^\beta}) = 1 - \frac{m^\beta t}{(t + m_0)k^\beta}$$

Network Science: Evolving Network Models January 2012

21

Degree distribution I

We obtain the degree distribution function $p(k)$ by taking the derivative of the cumulative Function $P(k)$:

$$p(k) = \frac{\partial P(k_i(t) < k)}{\partial k} = \frac{2m^\beta t}{m_0 t} \frac{1}{k^{\beta+1}}$$

Which for $t \gg m_0$ reduces to

$$p(k) \approx 2m^\beta k^{-\gamma}$$

$$\text{With } \gamma = \frac{1}{\beta} + 1 = 3$$

Network Science: Evolving Network Models January 2012

22

Degree distribution

$$k_i(t) = m \frac{t^{\frac{1}{2}}}{t_i^{\frac{1}{2}}} \quad b = \frac{1}{2}$$

$$P(k) = \frac{2m^2 t}{m_0 + t} \frac{1}{k^3} \sim k^{-\gamma} \quad \boxed{\gamma = 3}$$

(i) The degree exponent is independent of m .

(ii) As the power-law describes systems of rather different ages and sizes, it is expected that a correct model should provide a time-independent degree distribution. Indeed, asymptotically the degree distribution of the BA model is independent of time (and of the system size N)
→ the network reaches a stationary scale-free state.

(iii) The coefficient of the power-law distribution is proportional to m^2 .

A.-L. Barabási, R. Albert and H. Jeong, *Physica A* 272, 173 (1999)

Network Science: Evolving Network Models January 2012

23

NUMERICAL SIMULATION OF THE BA MODEL

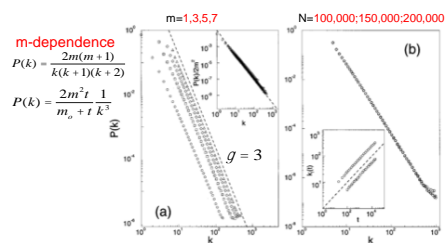
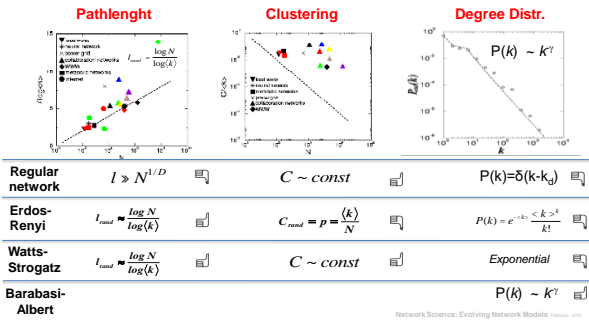


FIG. 21. Numerical simulation of network evolution: (a) Degree distribution of the Barabási-Albert model, with $N = m_0 + t \approx 300,000$ and $\Delta, m_0 = m = 1$; $\square, m_0 = m = 3$; $\diamond, m_0 = m = 5$; and $\triangle, m_0 = m = 7$. The slope of the dashed line is $\gamma = 3$, providing the best fit to the data. The inset shows the rescaled distribution (see text) $P(k)/2m^2$ for the same values of m , the slope of the dashed line being $\gamma = 3$; (b) $P(k)$ for $m_0 = m = 5$ and various system sizes: $\square, N = 100,000$; $\triangle, N = 150,000$; $\diamond, N = 200,000$. The inset shows the time evolution for the degree of two vertices, added to the system at $t_1 = 5$ and $t_2 = 95$. Here $m_0 = m = 5$, and the dashed line has slope 0.5, as predicted by Eq. (61). After Barabási, Albert, and Jeong (1999).

Network Science: Evolving Network Models January 2012

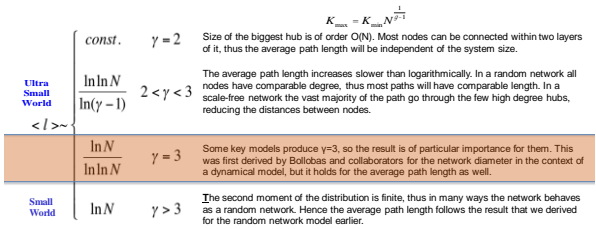
24

EMPIRICAL DATA FOR REAL NETWORKS



25

DISTANCES IN SCALE-FREE NETWORKS

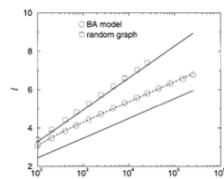


26

PATH LENGTHS IN THE BA MODEL

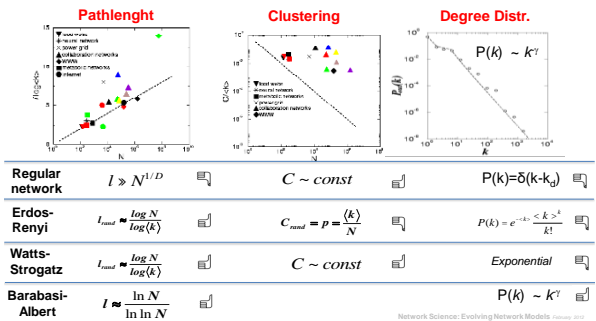
$$l \approx \frac{\ln N}{\ln \ln N}$$

Bollobas, Riordan, 2002



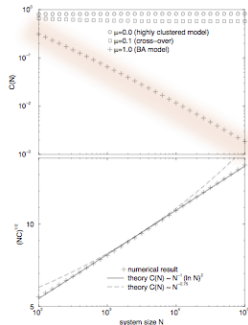
27

EMPIRICAL DATA FOR REAL NETWORKS



28

CLUSTERING COEFFICIENT OF THE BA MODEL



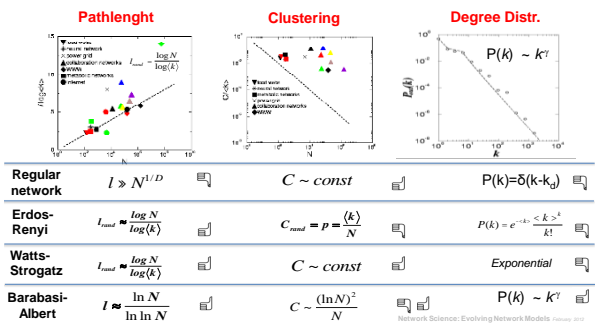
$$C = \frac{m (\ln N)^2}{8 N}$$

Konstantin Klemm, Victor M. Eguiluz, Phys. Rev. E 65, 057102 (2002)

Network Science: Evolving Network Models January 2012

29

EMPIRICAL DATA FOR REAL NETWORKS



30

The origins of preferential attachment.

Network Science: Evolving Network Models January 2012

31

LOCAL MECHANISMS

Local mechanisms. The *link selection model* offers perhaps the simplest example of a local or random mechanism capable of generating preferential attachment.

Growth: at each time step we add a new node to the network.

Link selection: we select a link at random and connect the new node to one of nodes at the two ends of the selected link

- The higher is the degree of a node, the higher is the chance that it will be located at the end of the chosen link.
- The more degree- k nodes are in the network (i.e., the higher is p_k), the more likely that a degree k node will be at the end of the link.

Network Science: Evolving Network Models January 2012

33

LOCAL MECHANISMS

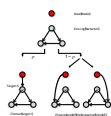
Copying model – Examples:

Social networks:

the more acquaintances an individual has, the higher is the chance that she will be introduced to new individuals by her existing acquaintances. Without friends, it is difficult to make new friends.

Citation Networks:

no scientist can be familiar with all papers published on a certain topic. If we assume that authors decide what to cite by randomly selecting references from the papers they have already read, then papers with more citations are more likely to be cited again.



Network Science: Evolving Network Models January 2012

35

MECHANISMS RESPONSIBLE FOR PREFERENTIAL ATTACHMENT

There are two philosophically different takes on this question.

LOCAL

The first approach views preferential attachment as an interplay between random events and some structural network property. These mechanisms do not require global knowledge of the network. Hence we will call them **local** or **random mechanisms**.

GLOBAL

The second approach assumes that the addition of each new node or link is preceded by a cost-benefit analysis, balancing various needs with the available resources. This assumes familiarity with the whole network and relies on optimization principles, prompting us to call them **global** or **optimized mechanisms**.

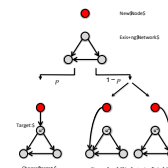
Network Science: Evolving Network Models January 2012

32

LOCAL MECHANISMS

Copying model.

When building a new webpage, authors tend to borrow links from webpages covering similar topics, a process captured by the copying model [17, 18]. In the model in each time step a new node with a single link is added to the network. To choose the target node we randomly select a node u and follow a two-step procedure [17]:
(a) **Random Connection:** with probability p the new node links to u .
(b) **Copying:** with probability $1 - p$ we randomly choose an outgoing link of node u and connect the new node to the selected link's target. Hence the new node "copies" one of the links of an earlier node.



Network Science: Evolving Network Models January 2012

34

GLOBAL MECHANISMS

Optimization. A longstanding assumption of economics is that humans make rational decisions, balancing cost against benefits. In other words, each individual aims to maximize its personal advantage.

Consider the Internet, whose nodes are routers or autonomous systems, connected via cables. Establishing a new Internet connection between two routers requires laying down a cable between them. As this can be costly, each new link is preceded by careful cost-benefit analysis.

Requires **global** information about network and node degrees.

Network Science: Evolving Network Models January 2012

36

SUMMARY: PROPERTIES OF THE BA MODEL

• Nr. of nodes:	$N = t$
• Nr. of links:	$L = m t$
• Average degree:	$\langle k \rangle = \frac{2L}{N} \rightarrow 2m$
• Degree dynamics	$k_i(t) = m e^{\frac{2t}{t_0}} \quad b = \frac{1}{2} \quad \beta: \text{dynamical exponent}$
• Degree distribution:	$P(k) \sim k^{-\beta} \quad g = 3 \quad \gamma: \text{degree exponent}$
• Average Path Length:	$l \approx \frac{\ln N}{\ln \ln N}$
• Clustering Coefficient:	$C \sim \frac{(\ln N)^2}{N}$

The network grows, but the degree distribution is stationary.

Network Science: Evolving Network Models January 2012

37

EVOLVING NETWORK MODELS

The BA model is only a minimal model.

Makes the simplest assumptions:

- linear growth
- linear preferential attachment

$$\langle k \rangle = 2m$$

$$\Pi(k_i) \propto k_i$$

Does not capture

variations in the shape of the degree distribution
variations in the degree exponent
the size-independent clustering coefficient

The BA model can be adapted to describe most features of real networks.

We need to incorporate mechanisms that are known to take place in real networks: addition of links without new nodes, link rewiring, link removal; node removal, constraints or optimization (see Barabasi book)

Network Science: Evolving Network Models January 2012

38

LESSONS LEARNED: evolving network models

1. There is no universal exponent characterizing all networks.
2. Growth and preferential attachment are responsible for the emergence of the scale-free property.
3. The origins of the preferential attachment is system-dependent.

Network Science: Evolving Network Models January 2012

39

LESSONS LEARNED: evolving network models

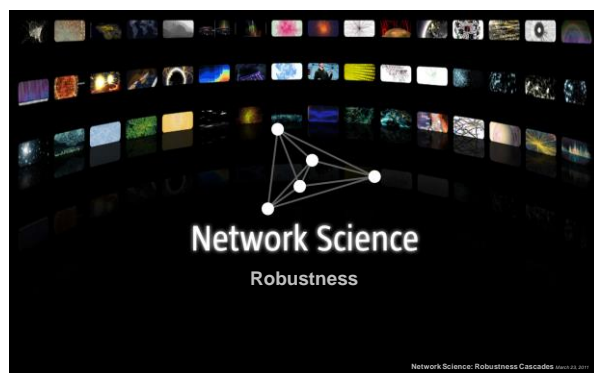
Philosophical change in network modeling:

ER, WS models are static models – the role of the network modeler it to cleverly place the links between a fixed number of nodes so that the network topology mimic the networks seen in real systems.

BA and evolving network models are dynamical models: they aim to reproduce how the network was built and evolved.
Thus their goal is to capture the network dynamics, not the structure.
→ as a byproduct, you get the topology correctly

Network Science: Evolving Network Models January 2012

40



Network Science: Robustness Cascades March 2011

41

ROBUSTNESS IN COMPLEX SYSTEMS

Complex systems maintain their basic functions even under errors and failures

cell → mutations

There are uncountable number of mutations and other errors in our cells, yet, we do not notice their consequences.

Internet → router breakdowns

At any moment hundreds of routers on the internet are broken, yet, the internet as a whole does not loose its functionality.

Where does robustness come from?

There are feedback loops in most complex systems that keep tab on the component's and the system's "health".

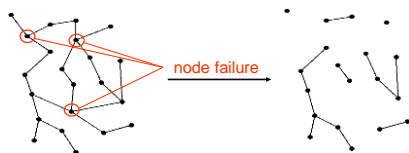
Could the network structure affect a system's robustness?

Network Science: Robustness Cascades March 2011

42

ROBUSTNESS

Could the network structure contribute to robustness?

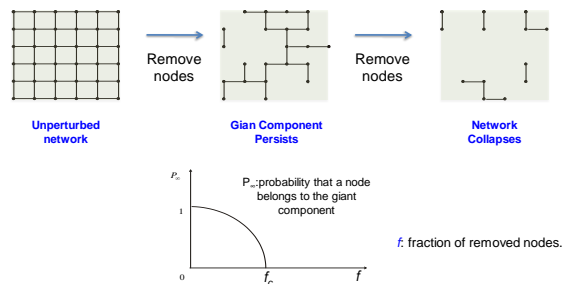


How do we describe in quantitative terms the breakdown of a network under node removal?
~percolation theory~

Network Science: Robustness Cascades (Nov 23, 2017)

43

ROBUSTNESS: INVERSE PERCOLATION TRANSITION

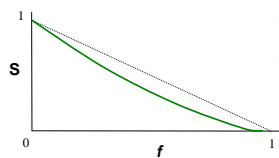
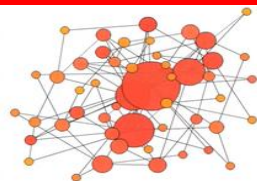


Network Science: Robustness Cascades (Nov 23, 2017)

44

ROBUSTNESS OF SCALE-FREE NETWORKS

Scale-free networks do not appear to break apart under random failures.
Reason: the hubs.
The likelihood of removing a hub is small.



Albert, Jeong, Barabási, *Nature* **406** 378 (2000)

Network Science: Robustness Cascades (Nov 23, 2017)

45

ROBUSTNESS OF SCALE-FREE NETWORKS

$$f_c = 1 - \frac{1}{K-1} \quad K = \frac{\langle k^2 \rangle}{\langle k \rangle} = \left| \frac{2-g}{3-g} \right| \frac{K_{\min}^{3-g} K_{\max}^{g-2}}{K_{\max}} \quad \begin{matrix} g > 3 \\ 3 > g > 2 \\ 2 > g > 1 \end{matrix}$$

$$K_{\max} = K_{\min}^{\frac{1}{g-1}}$$

$\gamma > 3$: K is finite, so the network will break apart at a finite f_c that depends on K_{\min}

$\gamma < 3$: K diverges in the $N \rightarrow \infty$ limit, so $f_c \rightarrow 1$!!!
for an infinite system one needs to remove all the nodes to break the system.

For a finite system, there is a finite but large f_c that scales with the system size

Network Science: Robustness Cascades (Nov 23, 2017)

46

NUMERICAL EVIDENCE

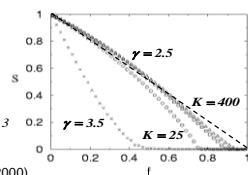
Scale-free random graph with

$$P(k) = Ak^{-\gamma}, \text{ with } k = m, \dots, K$$

$$f_c = 1 - \frac{1}{\frac{\gamma-2}{\gamma-3}m-1} \quad \text{if } \gamma > 3$$

$$f_c = 1 - \frac{1}{\frac{\gamma-2}{3-\gamma}m^{\gamma-2}K^{3-\gamma}-1} \quad \text{if } 2 < \gamma < 3$$

Cohen et al., Phys. Rev. Lett. 85, 4626 (2000).

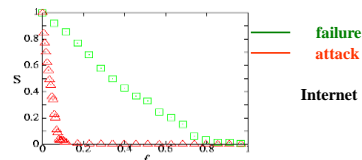


Infinite scale-free networks with $\gamma < 3$ do not break down under random node failures.

Network Science: Robustness Cascades (Nov 23, 2017)

47

INTERNET'S ROBUSTNESS TO RANDOM FAILURES



R. Albert, H. Jeong, A.L. Barabasi, *Nature* **406** 378 (2000)

$$f_c = 1 - \frac{1}{K-1}$$

Internet: Router level map, $N=228,263$; $\gamma=2.1 \pm 0.1$; $\kappa=28 \rightarrow f_c=0.962$

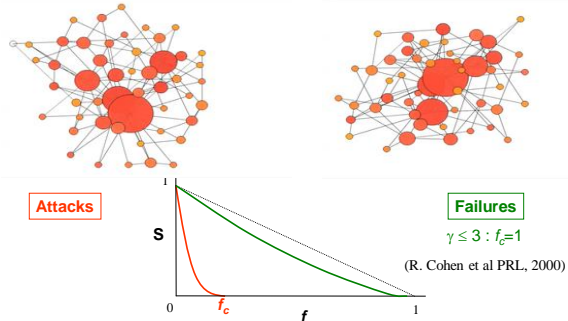
AS level map, $N=11,164$; $\gamma=2.1 \pm 0.1$; $\kappa=264 \rightarrow f_c=0.996$

Internet parameters: Pastor-Satorras & Vespignani, *Evolution and Structure of the Internet* Table 4.1 & 4.4

Network Science: Robustness Cascades (Nov 23, 2017)

48

Achilles' Heel of scale-free networks



Albert, Jeong, Barabási, *Nature* 406 378 (2000)

Network Science: Robustness Cascades (Nov 23, 2017)

49

Numerical simulations of network resilience

Two networks with equal number of nodes and edges

- ER random graph
- scale-free network (BA model)

Study the properties of the network as an increasing fraction f of the nodes are removed.

Node selection: random (errors)

the node with the largest number of edges (attack)

Measures:

S : the fraction of nodes in the largest connected cluster

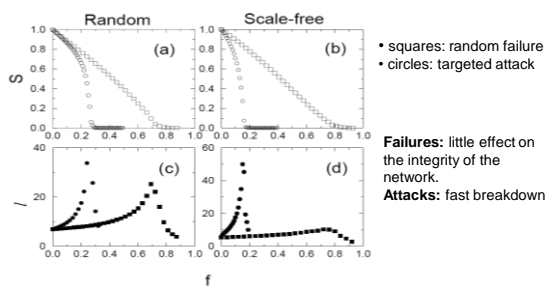
l : the average distance between nodes in the largest cluster

R. Albert, H. Jeong, A.-L. Barabási, *Nature* 406, 378 (2000)

Network Science: Robustness Cascades (Nov 23, 2017)

50

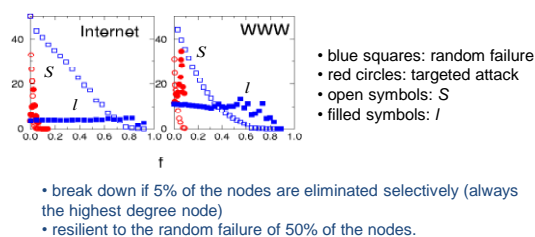
Scale-free networks are more error tolerant, but also more vulnerable to attacks



Network Science: Robustness Cascades (Nov 23, 2017)

51

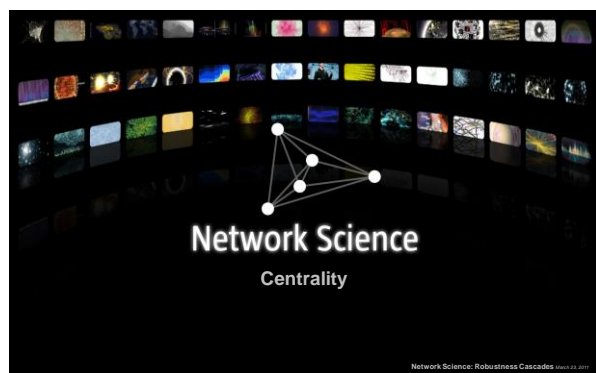
Real scale-free networks show the same dual behavior



Similar results have been obtained for metabolic networks and food webs.

Network Science: Robustness Cascades (Nov 23, 2017)

52

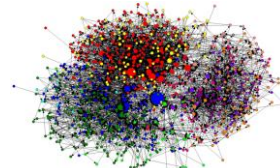


Network Science: Robustness Cascades (Nov 23, 2017)

53

Centrality Measures

- Describing how central a *node* or *edge* is.
- Network-level statistics may also be derived:
 - Minimum
 - Maximum
 - Average,
 - Etc.
- Indegree / Outdegree / All



54

54

Degree Centrality (Freeman, '79)

- Number of links: d_i
- Normalized: $\frac{d_i}{(N-1)}$
- "Popularity", "Gregariousness".
 - Indicates the *probability* of getting the information / infection spreading in the network.

55

Closeness Centrality (Freeman, '79)

- The sum of the shortest paths leading to the other nodes:
 - Distance ~ as inverse measure of centrality. $\sum_{i \neq j} l_{ij}$
- Normalized: $\frac{N-1}{\sum_{i \neq j} l_{ij}}$
 - A value in $[0,1]$ + "inverted"
- Indicates the *speed* of getting the information / the infection.

56

Betweenness Centrality (Freeman, '79)

- The number of crossing shortest paths: $\sum_{j,k} \frac{\tilde{p}_{jik}}{\tilde{p}_{jk}}$
- Normalized (divided by max value): $\sum_{j,k} \frac{\tilde{p}_{jik}}{\tilde{p}_{jk}} \bigg/ \max_i \sum_{j,k} \frac{\tilde{p}_{jik}}{\tilde{p}_{jk}}$
- Ability to *control* information / infection, or "Brokerage" – *linking* far regions, or Ability to *maintain connectivity*.

57

Eigenvector Centrality (Bonacich '72)

- The main Eigen vector of the (possibly weighted) adjacency matrix.
- Recursive definition:
 - Start with centrality 1 at each node.
 - Re-calculate centralities as the weighted sum of the neighbors' centrality
 - Normalize (divide by $\max(c_i)$).
 - Repeat until it reaches a fix point.
- A measure of "being linked to central nodes". (Notice the recursion.)
 - Corresponds to the *probability* of getting the information / infection.

58

The end