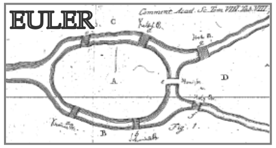


For IN2045@TUM: Internet Models, Graph Models and Metrics
Shortened by Heiko Niedermayer for the purpose of the lecture. Thanks to Dimitri for providing the slides.

Internet topology and routing structure, analysis and models

Dimitri Papadimitriou
Alcatel-Lucent Bell N.V.

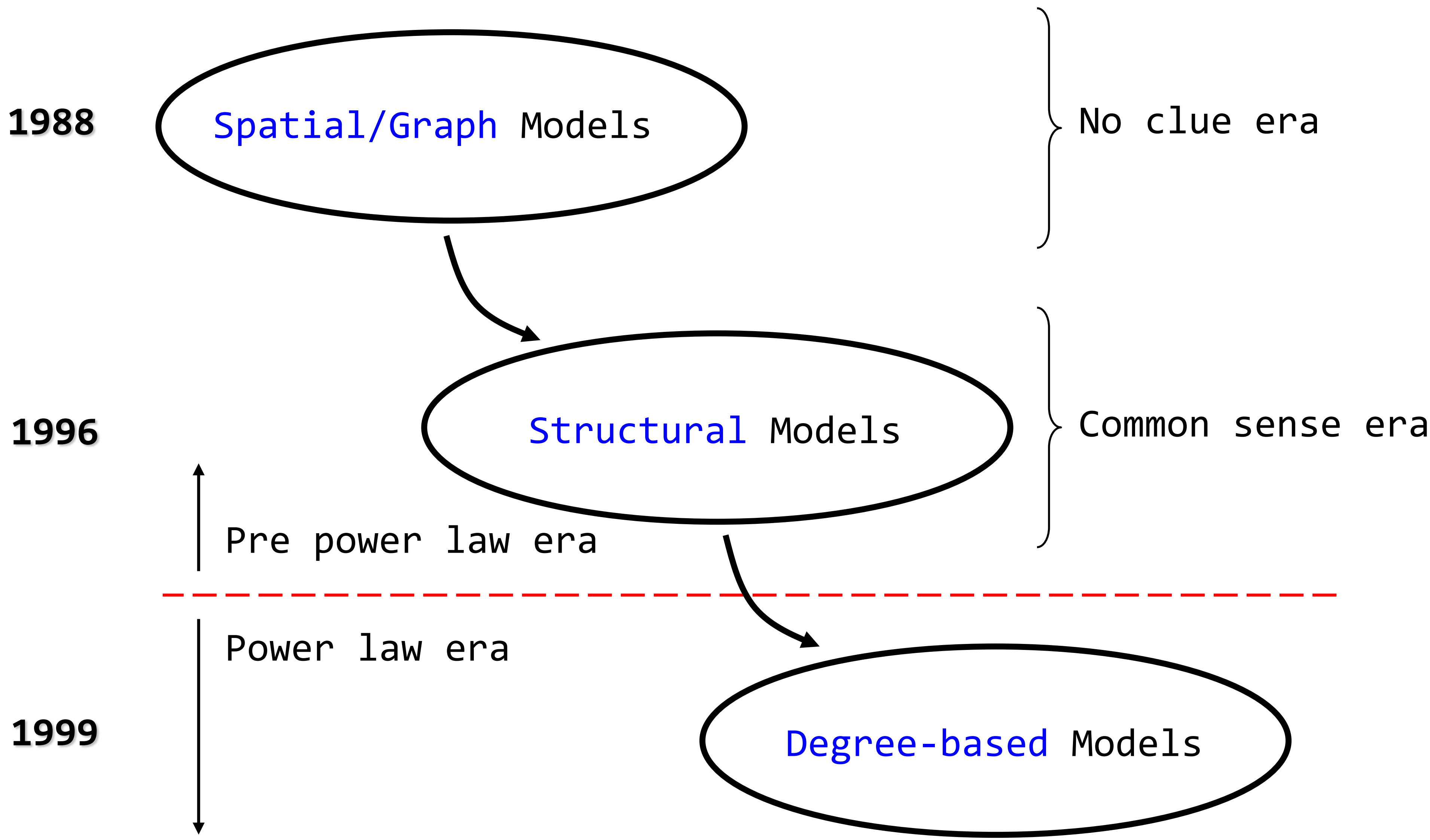
October 6, 2010
Alcatel-Lucent Bell N.V.
Antwerpen, Belgium

Outline (shortened slide set)

Internet Topology modelling

- Random Graphs models and generators
- Power Law relationships
- Degree-based models and generators
- Internet topology metrics

Models for Network Topology



Random Graphs models

Model	Probability	Year
Pure random model (ER model)	$P(u,v) = p$	1960
Waxman model	$P(u,v) = \alpha\ e^{-d/(\beta D)}$	1988
Exponential model	$P(u,v) = \alpha\ e^{-d/(D-d)}$	
Locality model	α if $d < r$ β if $d \geq r$	

where

- $d \equiv d(u,v)$ is the distance from u to v
- D is the maximum distance between any two nodes
- $0 < \alpha,\ 0 < \beta \leq 1$
  - Increasing α increases the number of edges in the graph
  - increasing β increases the ratio of long edges to short edges
- r is the boundary

Cf. "Statistical mechanisms of Networks" Lecture for Exponential Random Graph

Random Graph Model
Erdös-Renyi (ER) Model

Basic random graph model: given n vertices, an edge between any two vertices exists with a probability p, independently of any other edge in the network

- Initially: number of vertices $|V| = n$ and no edges
- To obtain a random element $G_{n,p}$, select the edges (u,v) independently with probability p ($0 \leq p \leq 1$)

  → element $G_{n,p}$ of the set G(n,P(edge)=p) appears with probability $p^m(1-p)^{M-m}$

  where $M = \frac{n(n-1)}{2}$ is max. possible number of edges

  Limit for large value of n:
  - expected number of edges E[m] = $M\,p$
  - expected average node degree E[λ] = $(n-1)\,p$

Note: another variant of the random graph model G(n,m) assigns uniform probability p to all graphs with n nodes and m edges

Random Graph Model
Erdös-Renyi (ER) Model

Probability p(*k*) that a node has degree *k* is Binomial

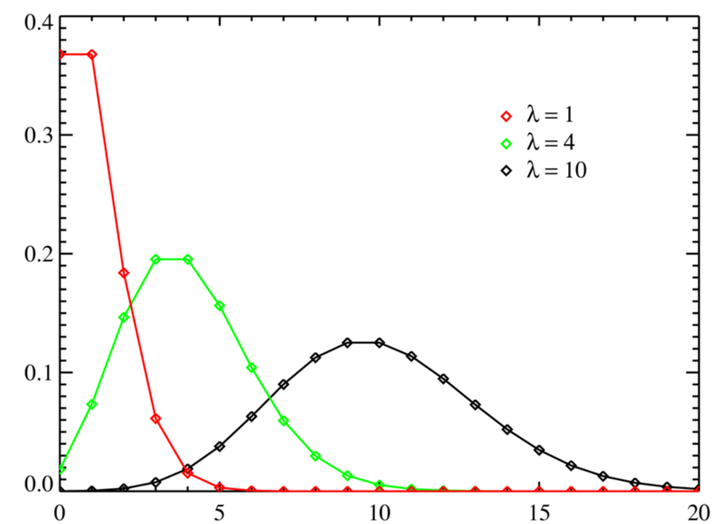
$$p(k) = \binom{n-1}{k} p^k (1-p)^{(n-1)-k}$$

number of ways to attach the endpoint of k edges from a particular node probability that there are exactly k edges

For large *n* (*n* >> *k* λ) and degree *k* fixed, where λ is the average node degree (λ = 2m/n = (n-1)p ≈ np), this is the Poisson distribution

$$p(k) = \frac{\lambda^k e^{-\lambda}}{k!}$$

The expected value of a Poisson-distributed random variable P(k) is equal to λ and its variance P(k,λ) is equal to λ

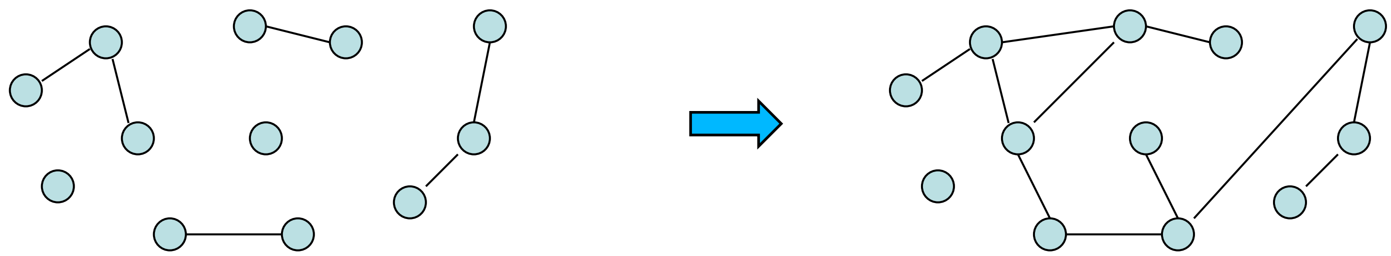


Random Graph Properties (1)

Expected structure of the random graph varies with the value of p

- The edges join vertices together to form components, i.e., (maximal) subsets of vertices that are connected by paths through the network

Phase transition property (most important property)

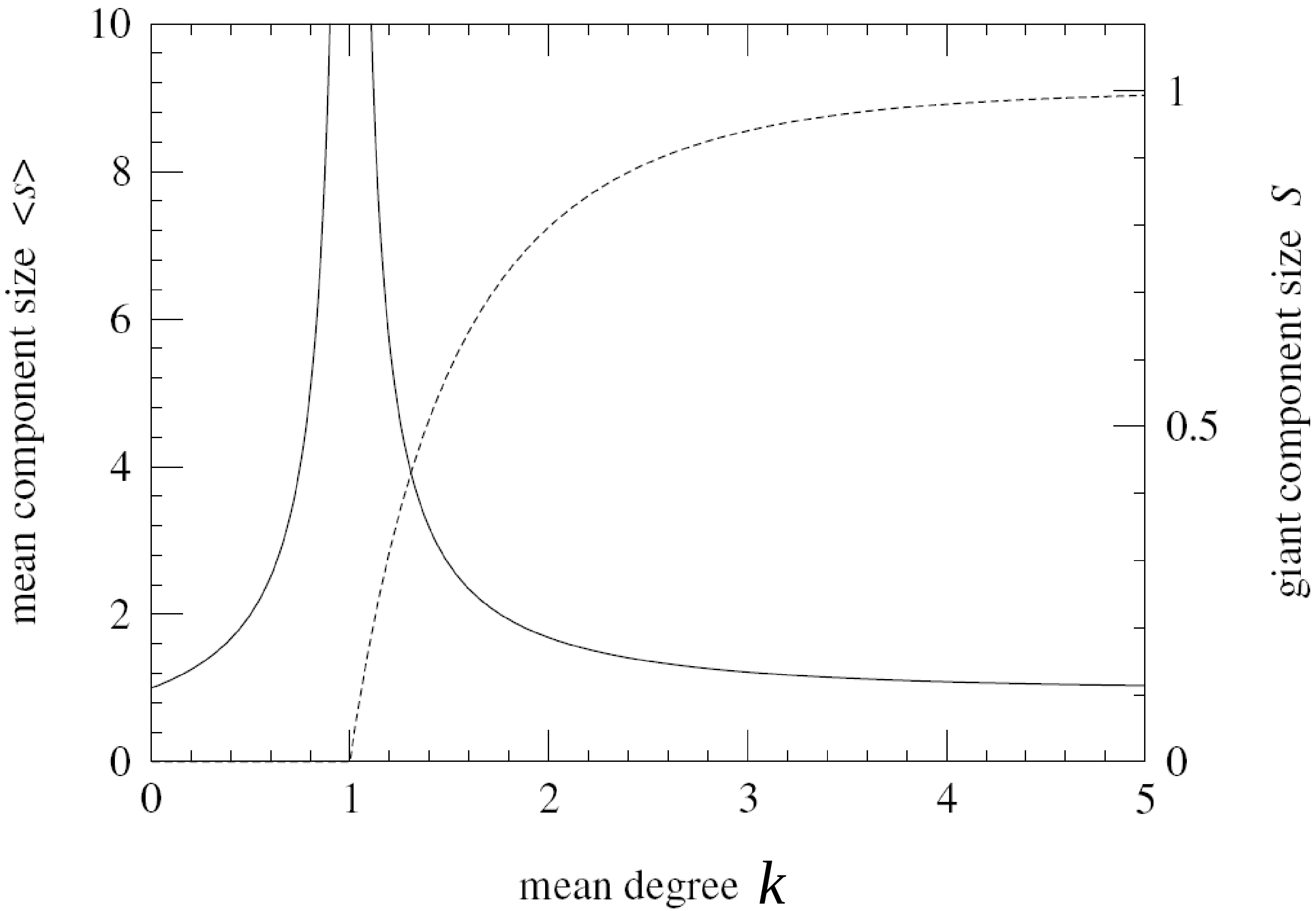


From a low-density, low-p state in which there are few edges and all components are small (i.e., O(log(n))

High-density, high-p state in which an extensive (i.e., O(n)) fraction of all vertices are joined together in a single giant component (remainder of the vertices occupying smaller components)

Random Graph Properties (2)

The mean component size (solid line), excluding the giant component if there is one, and the giant component size (dotted line) for the Poisson random graph



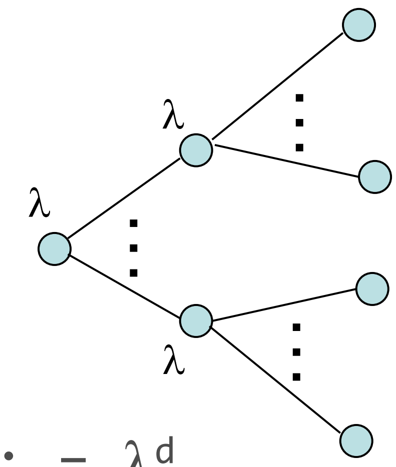
source: [Newman03]

Random Graph Properties (3)

Diameter

The number of nodes at distance d from a given node is given by λ^d

- If a graph has λ as average degree then
- Then
  - The number of first neighbours: λ
  - The number of second neighbours: λ^2
  - ...
  - The number of neighbours at distance d: = λ^d



- When extended to include all nodes in the graph

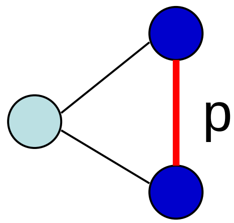
  $N = \lambda^D \rightarrow \log(N) = \log(\lambda^D) = D \log(\lambda)$

  => Diameter $D = \dfrac{\log(N)}{\log(\lambda)}$

Clustering coefficient

- If the graph is sparse enough, the probability that two neighbors of a node are connected is the independent probability, p:

$$E(C) \cong p \cong \frac{\lambda}{N}$$



- For a complete graph clustering coefficient = 1

Random Graph Model
Erdös-Renyi (ER) Model

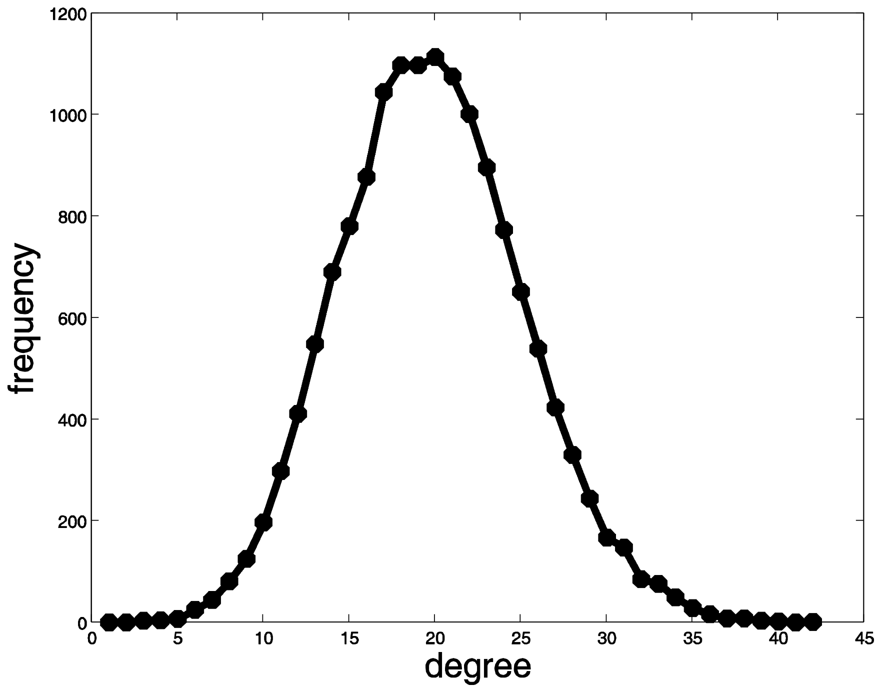
The measurements on real networks are usually compared against those obtained from “random networks”

Problem: find the probability distribution that best fits the observed data

f_k = fraction of nodes with degree $k \equiv$ probability that a randomly selected node has degree k

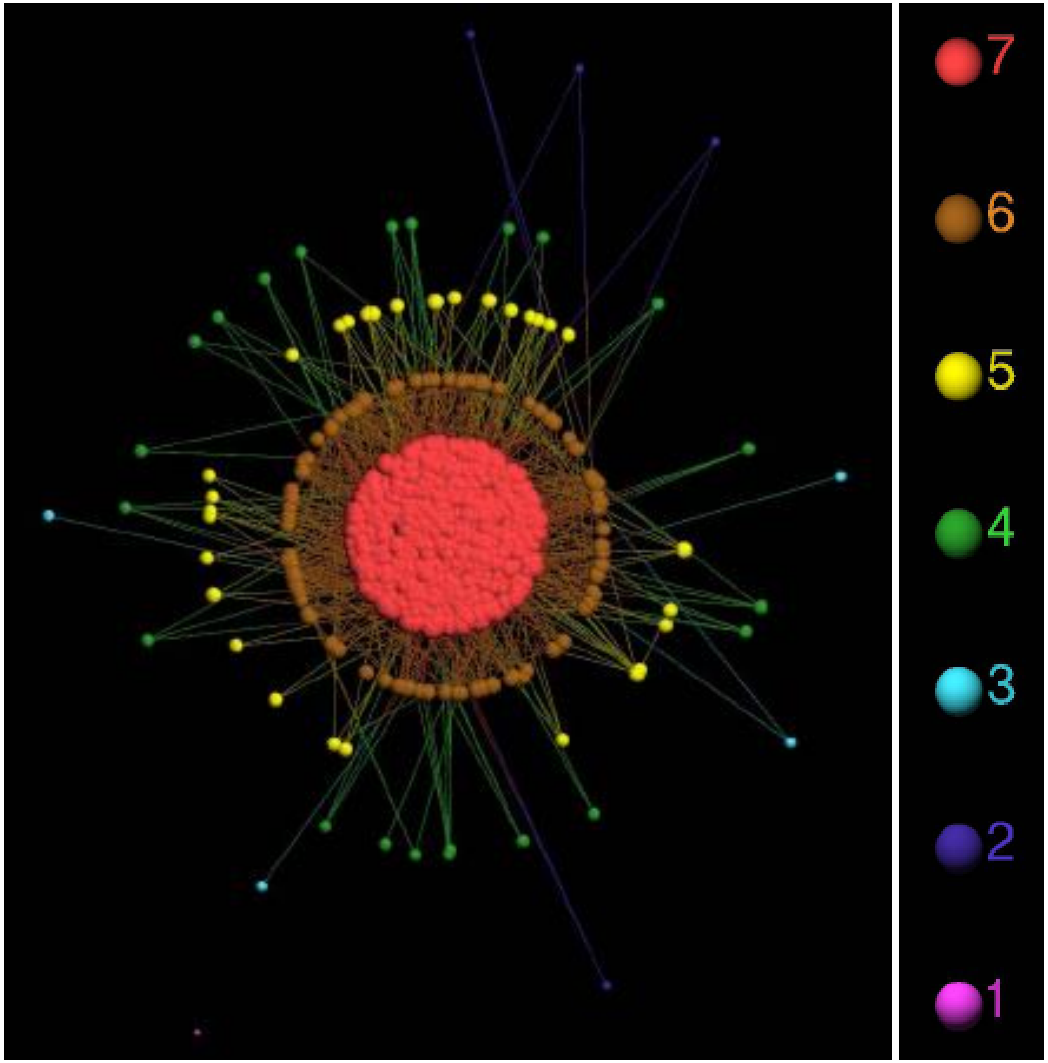
$$p(k) = P(k;\lambda) = \frac{\lambda^k}{k!} e^{-\lambda}$$

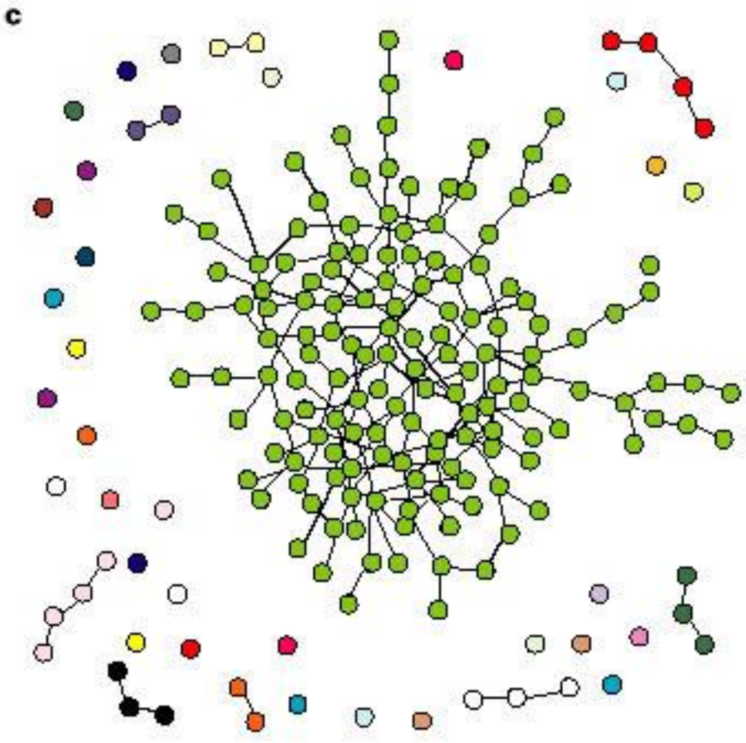
With the random graph model the node degree distribution is Poisson of mean $\lambda = n\,p$ (average node degree)



Highly concentrated around the mean λ (average degree) -> the probability of very high degree nodes is exponentially small

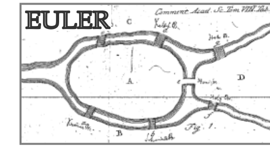
Random Graph Model
Erdös-Renyi (ER) Model





Poisson random graph

Source: http://www.caida.org (ISMA 2006)



Are E-R graphs realistic?

They have small world properties (diameter is logarithmic in the size of the graph) but **low clustering coefficient**

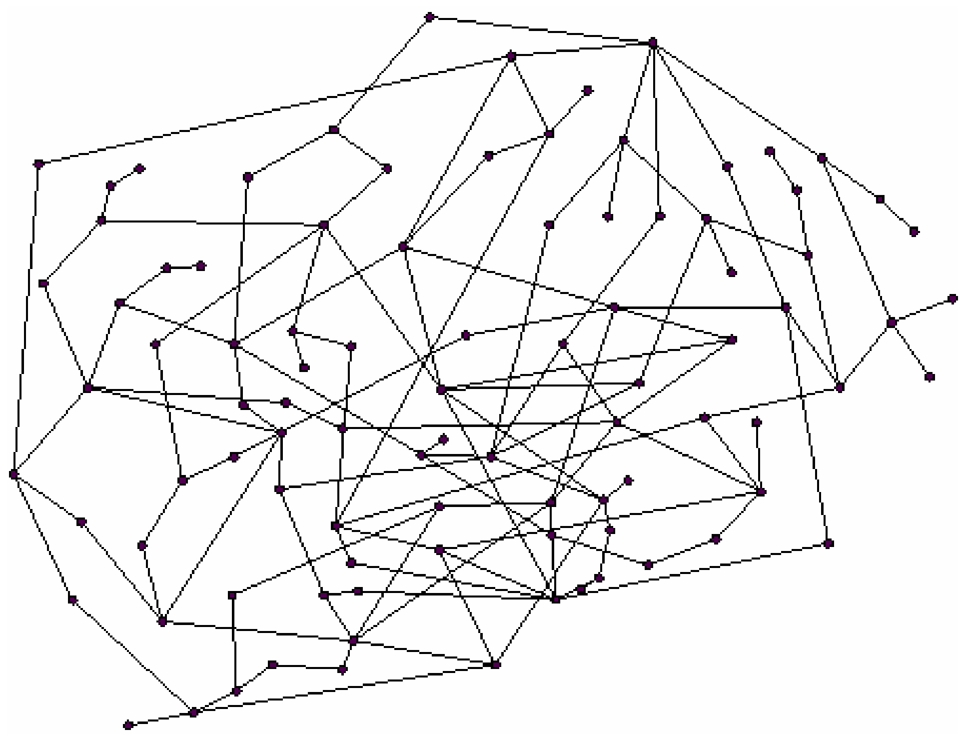
- Example for Internet AS topology: compare 0.30 with 0.0004 [Pastor-Satorras and Vespignani]

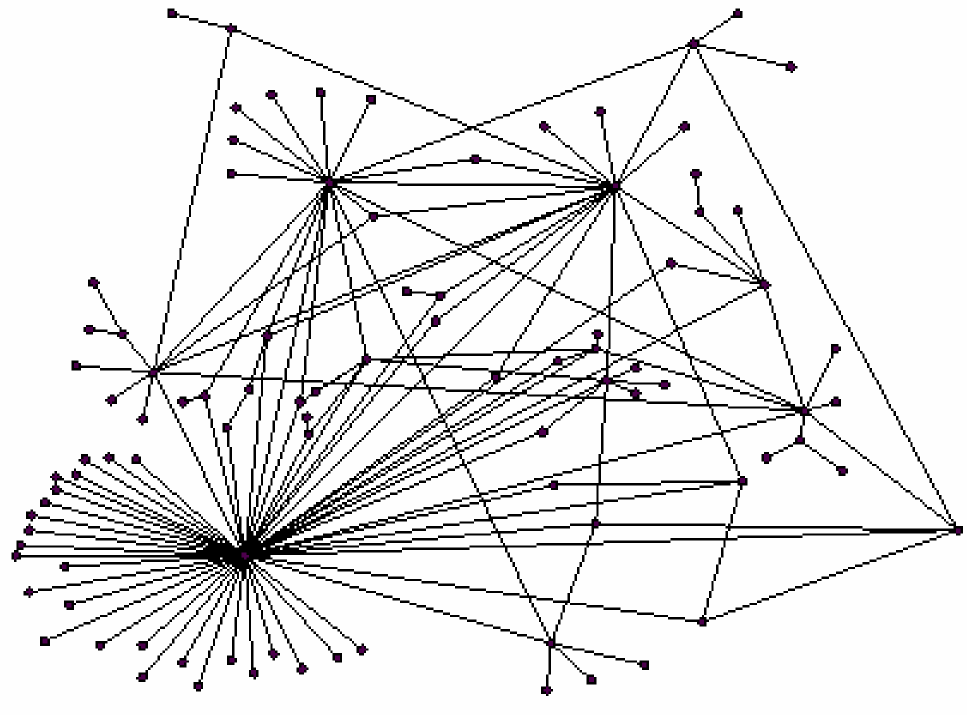
Unrealistic degree distributions

- Degrees not concentrated around mean (characteristic of Poisson distribution)
- Exponential tails (instead of heavy tailed degree distribution)

Result: departure from ER model and variants

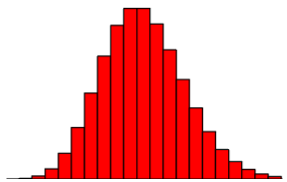
Poisson vs. Power-law network



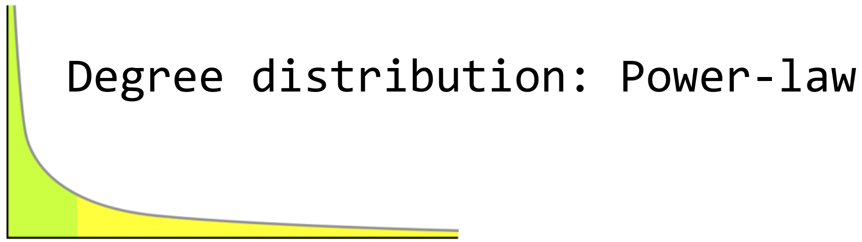


Poisson network
(Erdos-Renyi random graph)

Power-law distribution of
node degree: $P(k) \sim k^{-\gamma}$



Degree distribution: Poisson



Outline

Internet Topology modelling

- Random Graphs models and generators
- Power Law relationships
- Degree-based models and generators
- Internet topology metrics

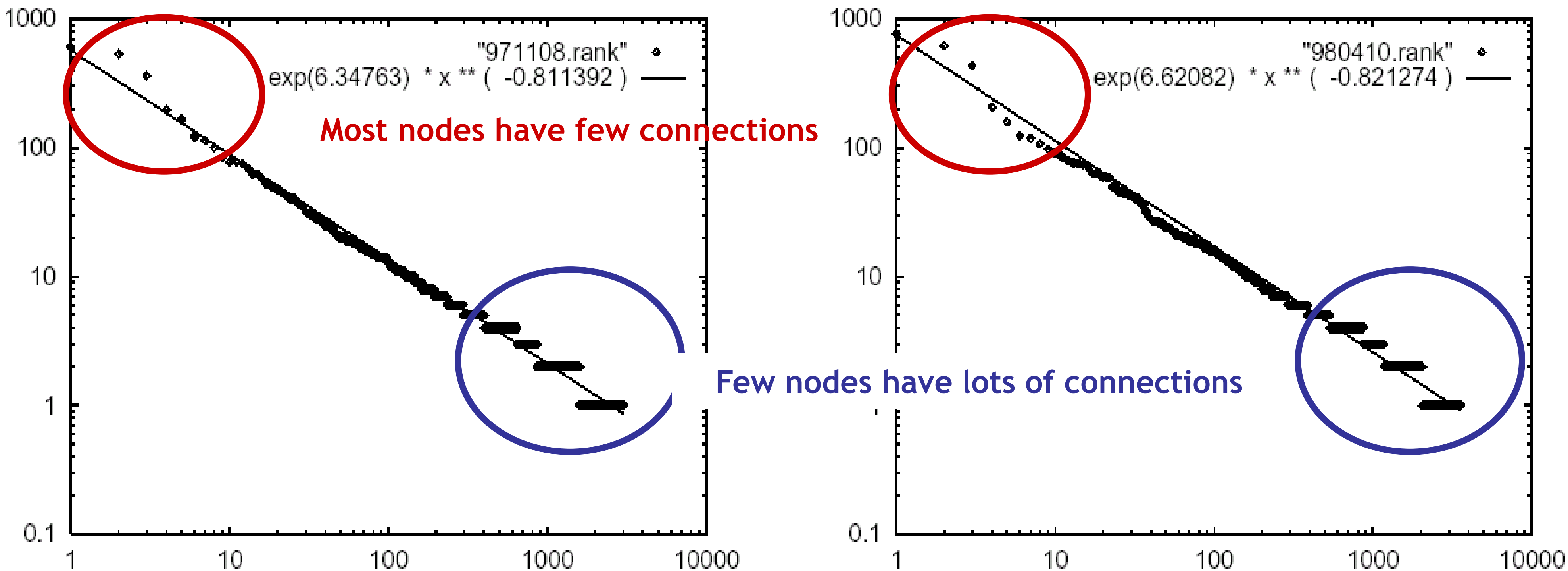
Initial Observation

Faloutsos et al. (1999) identify power law in node degree distribution at both router- and AS-level graph

A random variable X is said to follow a **power law** distribution with scaling index $\gamma > 0$ if

$$P[X > x] \approx c\, x^{-\gamma} \quad \text{as } x \to \infty$$

Rank plots: log-log plot of the out-degree of the nodes (# of edges incident) vs rank of the nodes (index in the order of decreasing out-degree)



(a) Int-11-97 (b) Int-04-98

Power Law and Scale Invariance

A random variable X (or its corresponding distribution function F) is said to follow a power law or is scaling with (scale or tail) index $\gamma > 0$ if

$$P[X > x] = 1 - F(x) \approx c\, x^{-\gamma} \quad \text{as } x \to \infty \qquad (*)$$

For $1 < \gamma < 2$: F has infinite variance but finite mean

For $0 < \gamma \leq 1$: F has infinite variance and infinite mean

Since (*) implies $\log(P[X > x]) \approx \log(c) - \gamma \log(x)$ -> doubly logarithmic plots of x versus 1 – F(x) yield straight lines of slope $-\gamma$ (at least for large x)

Power-law distributions are called scaling distributions because if the random variable X satisfies relationship (*) and x > w, then the conditional distribution of X given that X > w is given by

$$P[X > x \mid X > w] = \frac{P[X > x]}{P[X > w]} \approx c_1 x^{-\gamma}$$

where constant c_1 is independent of x and given by $c_1 = w^{\gamma}$

Thus, at least for large values of x, $P[X > x|X > w]$ is identical to the (unconditional) distribution $P[X > x]$, except for a change in scale

Power Laws

Power Law 1: Out-degree of nodes vs. rank

Power Law 2: Frequency of out-degree

Power Law 3: Pairs of nodes within *h* hops

Power Law 4: Eigenvalues of adjacency matrix

Power Laws	**Expression**	**Value**
Rank exponent (R)	$d_v \propto r_v^R$	R ~ -0,8
Outdegree exponent (O)	$f_d \propto d^O$	O ~ -2,2
Hop-plot exponent (H)	$P(h) \propto h^H$	H ~ 4,7
Eigenvalue exponent (ε)	$\lambda_i \propto i^\varepsilon$	ε ~ -0,48

Power Laws - Summary

Rank Exponent *R*: The out-degree d_v of a node v is proportional to the rank of the node r_v to the power a constant *R* (~ -0.8): $\mathbf{d_v \propto r_v^R}$

Out-degree Exponent *O*: The frequency f_d of an out-degree d is proportional to the out-degree to the power of a constant *O* (~ -2.2): $\mathbf{f_d \propto d^O}$

Hop-plot Exponent *H*: The total number of pairs of nodes P(h) within h hops is proportional to the number of hops to the power of a constant *H* (~ 4.7): $\mathbf{P(h) \propto h^H}$

Effective Diameter: given a graph with N nodes and E edges, define the effective diameter as:

$$\delta_{ef} = \left(\frac{N^2}{N+2E} \right)^{1/H}$$

Eigen Exponent *ε*: The eigenvalues λ_i of a graph are proportional to the order i to the power of a constant ε (~ -0.48): $\mathbf{\lambda_i \propto i^{\varepsilon}}$

Outline

Internet Topology modelling

- Random Graphs models and generators
- Power Law relationships
- Degree-based models and generators
- Internet topology metrics

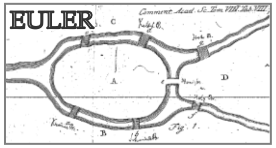
Degree-based Network Topology: Models and Generators

Faloutsos et al. (1999) find power law in node degree distribution at router-level graph and Autonomous System (AS) graph

Basic Idea: traditional random graphs [Erdös & Renyí, 1959] do not produce power laws, so develop new models that explicitly attempt to match the observed (power law) distribution in node degree

=> Led to active research in degree-based network models: focus on generators that match degree distribution of observed graph (descriptive methods)

Degree-based Network Topology: Models and Generators



Two methods for generating random networks having power law distributions in node degree

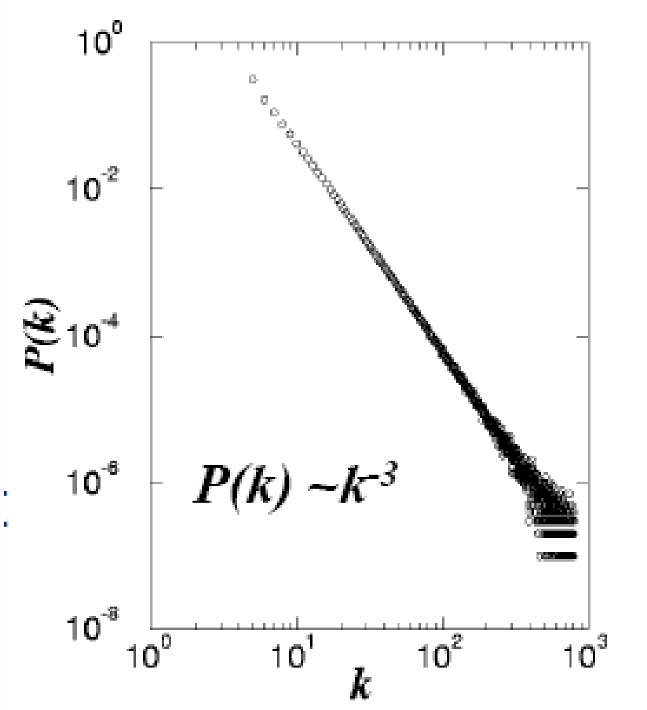
- **Growth modelling** (evolutionary)
  - **Barabasi-Albert (BA)** model: scale free networks characterized by Incremental growth and Preferential Attachment
  - Albert-Barabasi (AB) model: variant of BA model
  - Inet 3.0: enforced power law degree distribution and Preferential Attachment
  - BRITE: model based on Incremental growth and Preferential Attachment
  - **Generalized Linear Preference (GLP)** model
- **Distribution modelling** (non-evolutionary)
  - Exact degree sequence: **Power Law Random Graph (PLRG)**
  - Expected degree sequence: **Generalized Random Graph (GRG)**

Barabasi-Albert (BA) model [Barabasi 1999]

Power-law degree distribution can arise from two mechanisms

- Incremental growth: continuous addition of new nodes and edges to the system
- Preferential attachment: new nodes are preferentially attached to nodes that are already well connected

Probability of attachment to node i: $P_i(t) = \dfrac{k_i(t)}{\sum_{j=1}^{N(t)} k_j(t)}$



It is estimated that BA model generates networks with node degree distribution $P(k) \sim k^{-3}$

Barabasi-Albert (BA) model [Barabasi 1999]

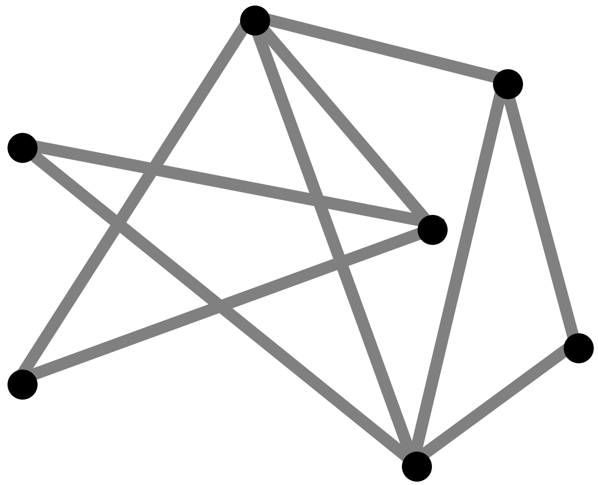
Method

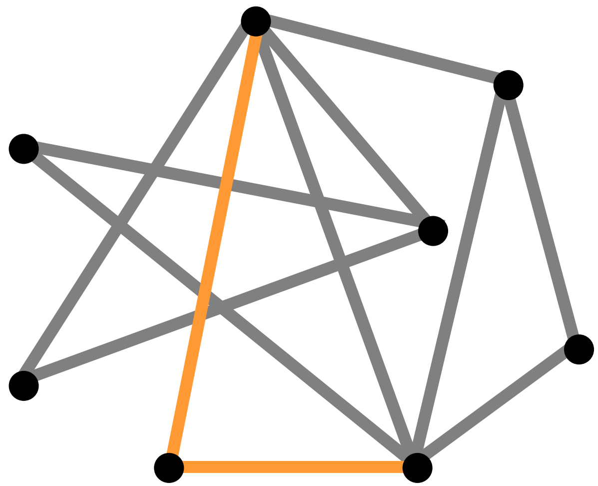
- Start with a small number (m_0) of nodes
- At every step, add a new node with $m \leq m_0$ edges that link the new node to m different nodes already present in the graph

  Probability $P_i(t)$ that a new node will be connected to an existing node i depends on the connectivity (degree) k_i of that node at time t

  -> at each step: $P_i(t) = k_i(t) / \sum_{j=1,N} k_j(t)$

- After t steps the model leads to a random network with N(t) = t + m0 nodes and E(t) = m . t links



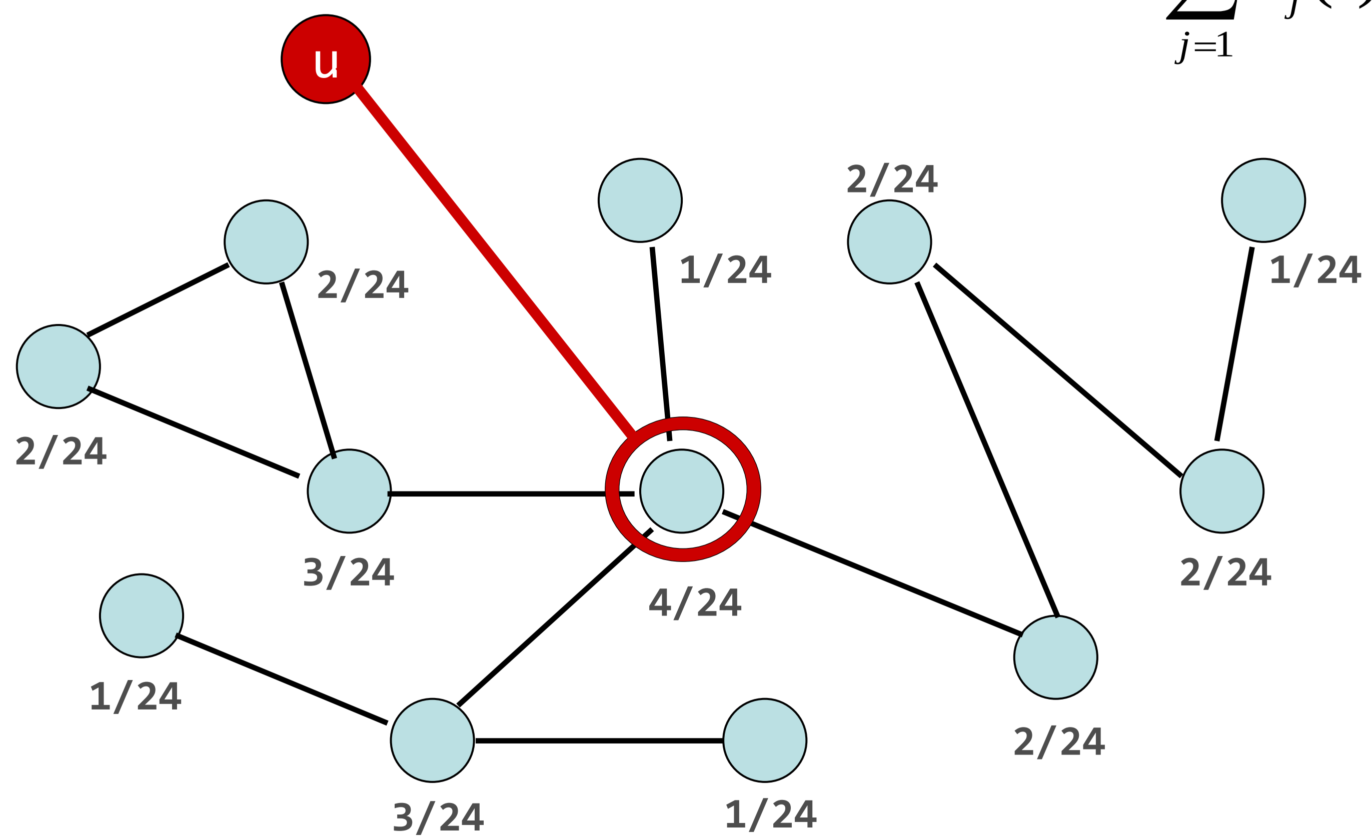


Barabasi-Albert (BA) model [Barabasi 1999]

Preferential attachment model: paradigm the "rich gets richer"

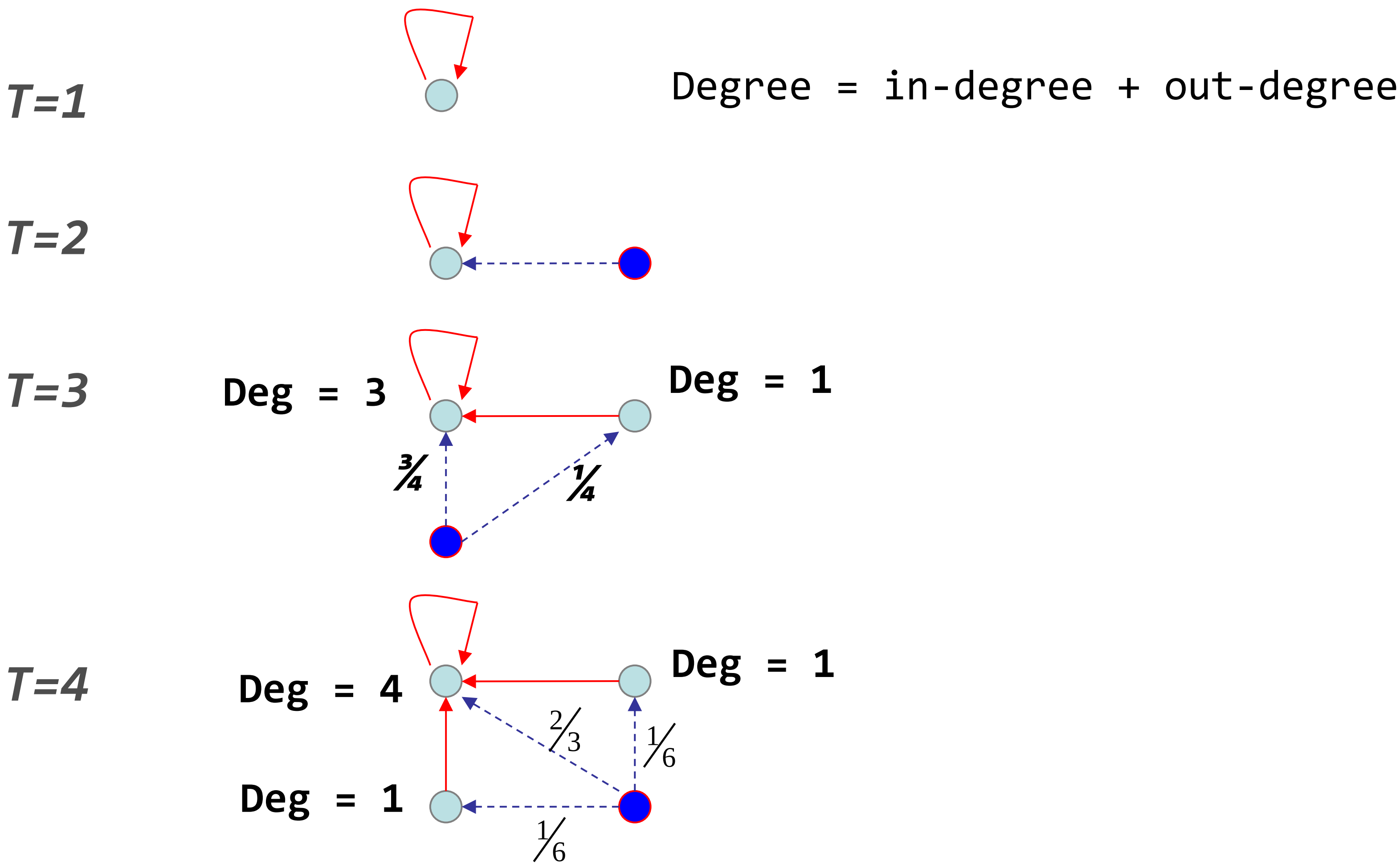
-> Attach new node u to existing node u in graph with probability

$$P_i(t) = \frac{k_i(t)}{\sum_{j=1}^{N} k_j(t)}$$



Preferential Attachment Model

Preferential Attachment model: each new node connects to the existing nodes with a probability **proportional to their degree**



BA Model -> Topology properties

After t steps the model leads to a topology with

- Number of nodes: $N(t) = m_0 + t$
- Number of edges: $E(t) = e_0 + m\,t$
- The degree of node i increases in time as a power-law with exponent 1/2: $k_i(t) = m\,(t/t_i)^{1/2}$

- Average degree $\langle k \rangle = 2E/N \to 2m$
- Degree distribution $P(k) \to 2m^2\,k^{-3}$ for $t \to \infty$

  The probability that a node has k links follow a power-law with exponent $\gamma=3$ (degree distribution becomes stationary)

PA Model Variations

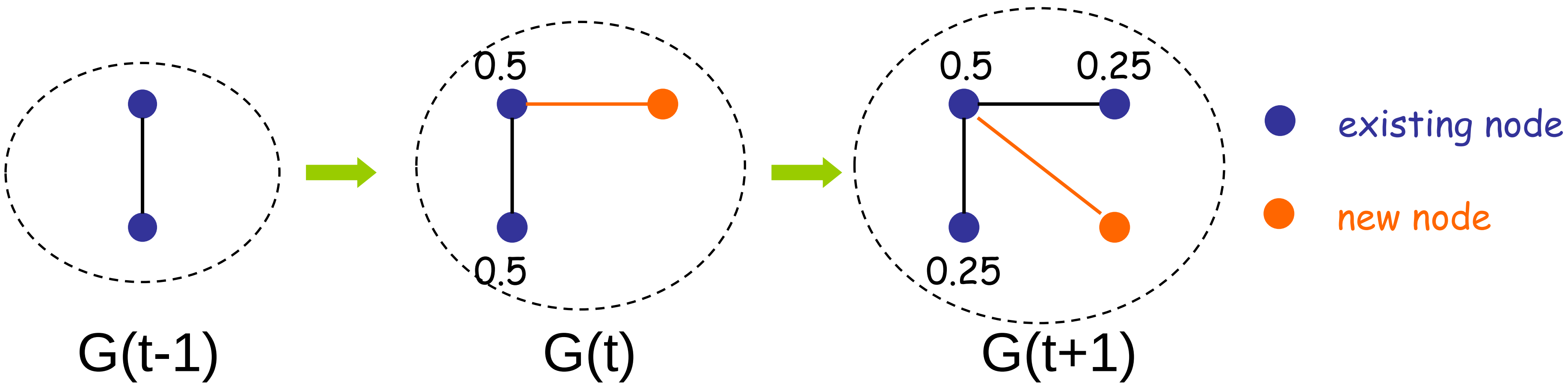
Variations of the Preferential Attachment model (scale index γ depends on model details)

Model	Constraint	Reference
$P(k) \sim k$	$\gamma = 3$	Barabasi and Albert (1999)
$P(k) \sim k + a$	$\gamma = 2 + a/m$ $-m < a < \infty$, $m > 1$	Initial attractiveness *a* (-> shift of γ): Dorogovtsev-Mendes-Samukhin model (2000)
$P(k) \sim k - \beta$	$-\infty < \beta \leq 1$	Generalized linear model: Bu (2002)
$P(k) \sim k^{\alpha}$	$\alpha < 1$ (exp.distribution) $\alpha > 1$ (fully connected)	Non-linear model: Krapivsky and Redner (2000)
$P(k) \sim \eta k$	$P(k) = k^{-1-C} / \ln(k)$	Intrisic Fitness: Bianconi and Barabasi (2000)

Barabasi-Albert (BA) model (Barabasi, 1999)
Albert-Barabasi (AB) model (Albert, 2000)

Incremental growth: starting from graph $G(t=t_0)$: G_0

- Add new nodes to graph G
- Add new links to graph G
- **Rewire links**: re-arrangement of already existing links



Linear preferential attachment: new nodes prefer existing nodes with large-degree

- $P_i(t)$ probability of selecting an existing node i of degree k_i at time t

  BA Model: $P_i(t) = k_i(t) / \Sigma_{j=1,N}\, k_j(t)$

  AB Model: $P_i(t) = [k_i(t) + 1] / \Sigma_{j=1,N}\, [k_j(t) + 1]$

Extended model

Start with m_0 isolated (unconnected) nodes

At each step, perform one of following three operations

- Add m new links with probability ***p***:

  For each of link, one end of the link is selected at random, while the other is preferentially selected with probability

  $$P_i(t) = [k_i(t) + 1] / \Sigma_{j=1,N} [k_j(t) + 1]$$

- Rewire m links with probability ***q***:

  For each link, randomly select a node i and a link (i,j) connected to it. This link is removed and replaced by a new link (i',j) connecting the node j to a new node i' selected with probability $P_i(t)$

  $$\left(\frac{\partial k_i}{\partial t}\right) = -q\frac{m}{N} + qm\frac{k_i+1}{\sum_j (k_j+1)}$$

- Add new node with m links with probability ***1-p-q***:

  Preferentially select the m links (that are connected to nodes already present) with prob. $P_i(t)$

  $$\left(\frac{\partial k_i}{\partial t}\right) = (1-p-q)m\frac{k_i+1}{\sum_j (k_j+1)}$$

Albert-Barabasi (AB) model [Albert 2000]

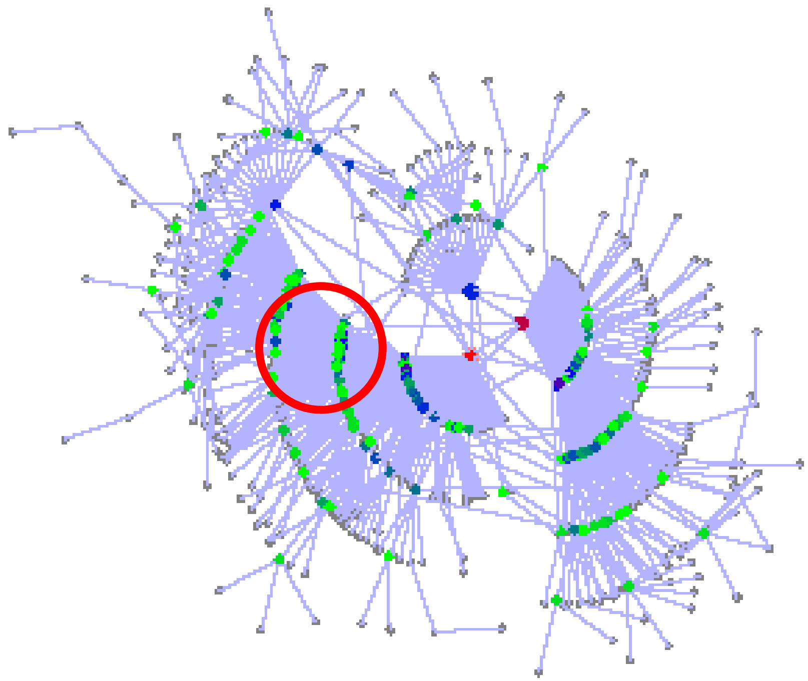
The preferential attachment probability

$$P_i(t) = \frac{k_i(t)+1}{\sum_{j=1}^{N}(k_j(t)+1)}$$

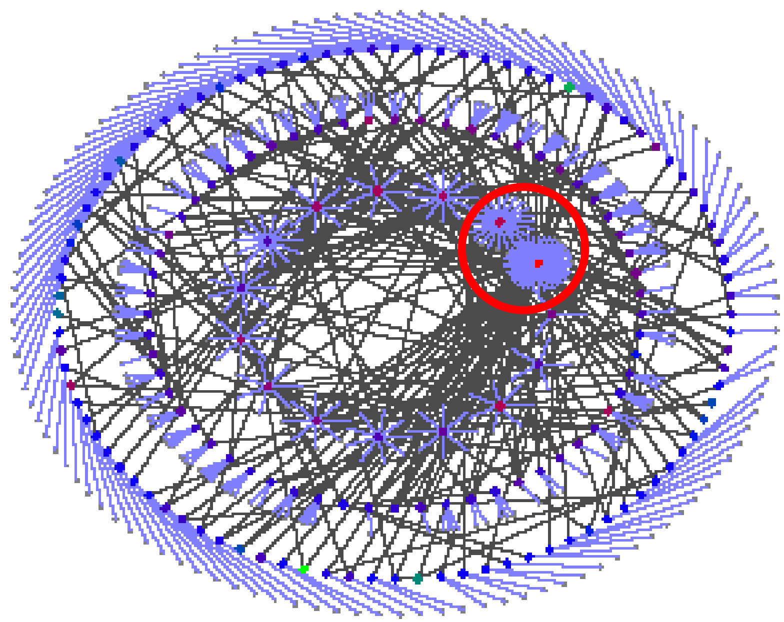
leads to a power-law distributed connectivity, whose exponent depends on the parameters q and p.

Properties of Degree-based Models

Preferential Attachment



Expected Degree Sequence (PLRG)



Degree sequence follows a power law (by construction)

High-degree nodes correspond to highly connected central hubs, which are crucial to the system

Achilles' heel: robust to random failure, fragile to specific attack (to hubs)

Power Laws Relationships of the Internet Topology: Revisited

Main Findings

- AS paths (BGP routing system) might not cover the complete AS topology
- Distribution of node degrees is not exactly a power law but definitely a heavy tailed distribution
- A vast majority of new ASes are born with vertex degree 1 or 2
- ASs can die also!! (*deaths not included in the BA Model*)
- ASs have much stronger preference to connect to high vertex degree ASs than predicted by the linear preferential model
- Rewiring not a significant factor in the evolution of the Internet

Scale Free networks

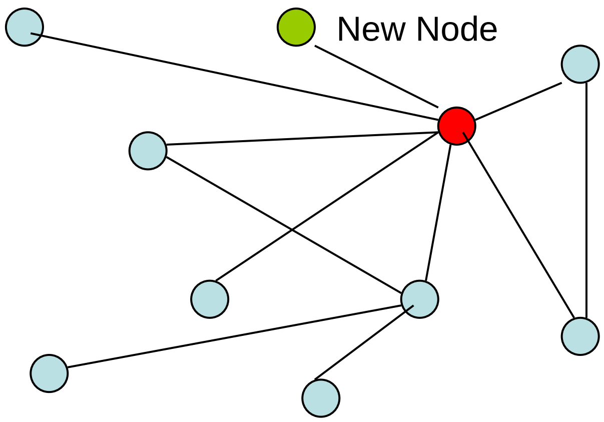
Scale free networks (term introduced by Barabasi)

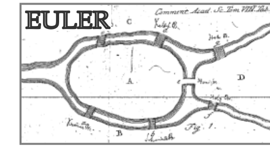
Idea: universal model of network topologies that exhibit power law distributions in the network node connectivity

Definition of scale free: any function f(x) that remains unchanged to within a multiplicative factor under a rescaling of the independent variable x

-> Power law function since only solutions to

f(a x) = g(a) f(x)





Scale Free networks

1. Continuous incremental growth
   - Existing models of networks did not include the addition of nodes over time (graphs remained static)
   - Scale free networks are in a state of continuous growth by incremental addition of new nodes and links to the system
2. Preferential attachment
   - New nodes tend to connect to nodes that are already well connected. New nodes have higher probability of connecting to the existing nodes with high connectivity, i.e., a “rich gets richer”
   - **“Rich club” phenomenon** - power laws in asymptotic limit: new nodes attach preferentially to high-degree nodes (well-connected nodes) in linear proportion to their degree

   <u>Note</u>: role of rewiring process (re-arrangement of the already existing links)

Rich Club Phenomenon

Rich nodes

- Power-law technologies have small number of nodes having large number of links

AS graph shows this phenomenon

- Rich nodes are well connected to each other
- Rich nodes are connected preferentially to the other rich nodes

Measured in the

- Original-maps of the AS graph (BGP Routing tables by University of Oregon Route Views Project)
- Extended-maps of the AS graph (BGP Routing tables + Looking Glass (LG) data + Internet Routing Registry data)

Scale Free networks Controversy

Scientists spot Achilles heel of the Internet

Fact: scale-free networks have approximately power law degree distributions

Claim

- If the Internet has power law degree distribution
- Then, the Internet must be scale-free

=> The Internet has the properties of a scale-free network

Scale Free networks Controversy

Scientists spot Achilles heel of the Internet

Implications of “scale free” network structure: Few centrally located and highly connected hubs (high-degree nodes correspond to highly connected central “hubs”, critical to the system)

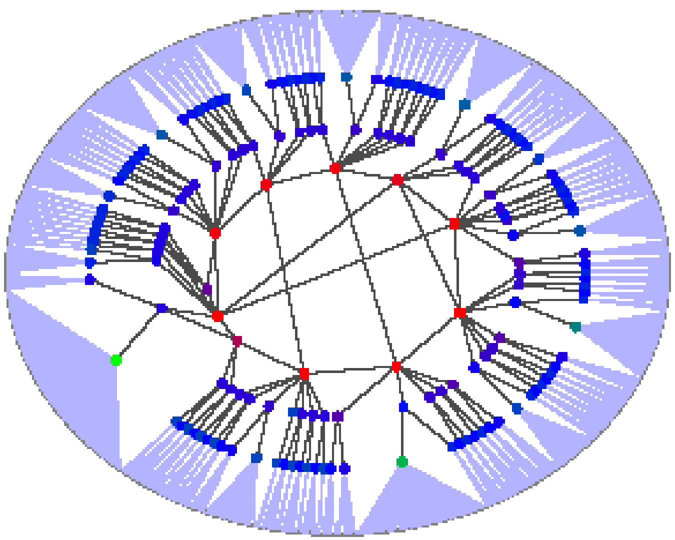
=> Achilles’ heel: robust to random attack/node failures (probability of targeting hub very low) but vulnerable to targeted attacks

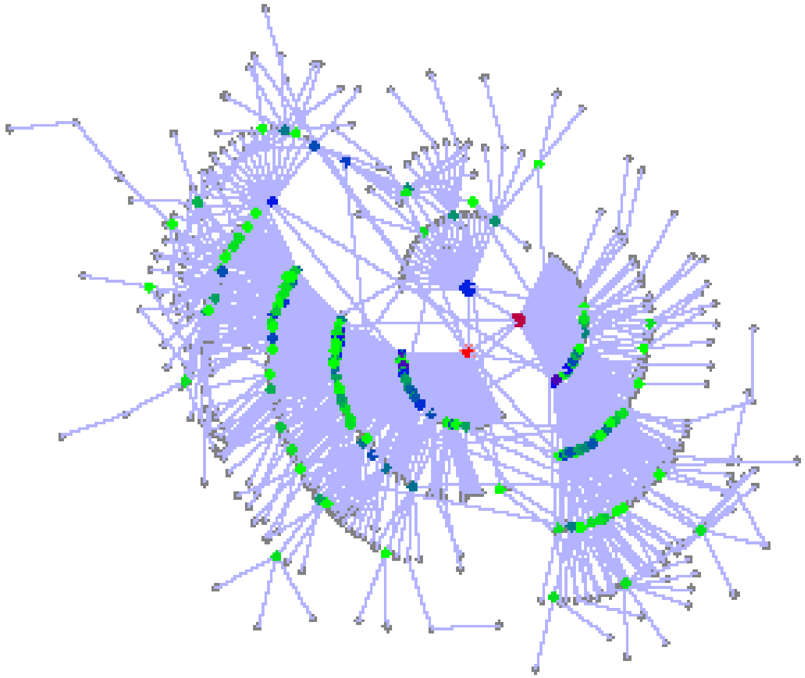
> "The reason this is so is because there are a couple of very big nodes and all messages are going through them. But if someone maliciously takes down the biggest nodes you can harm the system in incredible ways. You can very easily destroy the function of the Internet,..."

-- “Achilles heel of the Internet” Albert, Jeong, Barabasi, Nature 2000

Scale-Free Networks Controversy

Networks with the same statistical features can be OPPOSITES in terms of engineering





Approx. Real network
- Meshed, low-degree core
- Result of design
- High performance and robustness

Preferential Attachment
- High degree central “hubs”
- From random construction
- Poor performance and robustness

Scale-Free Networks Controversy

Scale-free claims: based critically on the implied relationship between power laws and a network structure that has highly connected “central hubs”

- Not all networks with power law degree distributions have properties of scale free networks (The Internet is just one example!)
- Building a model to replicate power law data is no more than curve fitting (descriptive, not explicative)

The scale-free models ignore all system-specific details in making their claims

- Ignore architecture e.g. hardware, protocol stack
- Ignore objectives e.g. performance
- Ignore constraints e.g. geography, economics

Scale-Free Networks Controversy

Conclusion (from "opponents")

- The scale-free claims of the Internet are not merely wrong, they suggest properties that are opposite to the real thing
- Fundamental difference: random vs. designed

Outline

Internet Topology modelling

- Random Graphs models and generators
- Power Law relationships
- Degree-based models and generators
- Internet topology metrics

Internet Topology Metrics

A network topology is characterized by topology metrics including (non-exhaustive list)

- Average degree
- Degree Distribution (DD)
- Joint Degree Distribution (JDD) or Degree correlation
- Characteristic path length
- Distance
- Clustering and clustering coefficient
- Betweenness
- Spectrum

Average degree

Definition: average node degree $$\boxed{\bar{k} = \frac{2m}{n}}$$

where m = number of links
n = number of nodes (a.k.a graph size)

Interpretation

- Coarsest connectivity characteristic of the topology
- Networks with higher $\bar{k}$ are "better-connected" on average and, consequently, are likely to be more robust
- Detailed topology characterization based only on the average degree is limited

  Reason: graphs with the same average node degree can have very different structure

Degree Distribution (1)

Definition: node degree distribution (DD) P(k) is the probability that a randomly selected node is k-degree

$$\boxed{P(k) = \frac{n(k)}{n}}$$

where n(k) = number of nodes of degree k (k-degree nodes)

- Degree distribution contains more information about connectivity than the average degree

  Reason: given a specific form of P(k) we can always restore the average degree by

$$\bar{k} = \sum_{k=1}^{k_{\max}} k\, P(k)$$

where k_{max} is the maximum node degree in the graph

Degree Distribution (2)

Interpretation

- Most frequently used topology characteristic
- [Faloutsos99] observation that Internet's degree distribution (both router and AS-level) follows a *power law* had significant impact on network topology research
  - *Structural Internet models* before failed to exhibit power laws ⇔ organized hierarchy existence among ASes
  - [Tangmunarunkit02]: topologies derived from structural generators that incorporated hierarchies of AS tiers did not have much in common with topologies obtained from real observed data
- Smooth power law degree distribution indicates
  - Indicates no organized tiers among ASes
  - The power law distribution also implies substantial variability associated with degrees of individual nodes

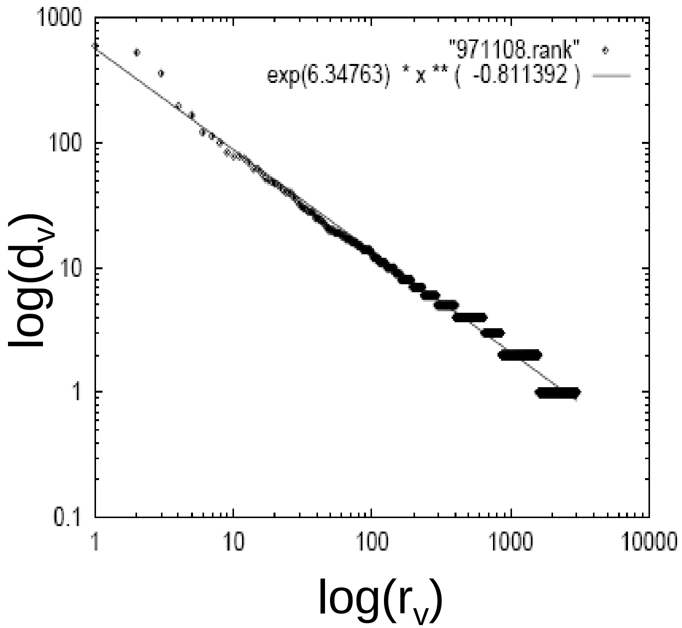
Degree Distribution (3)

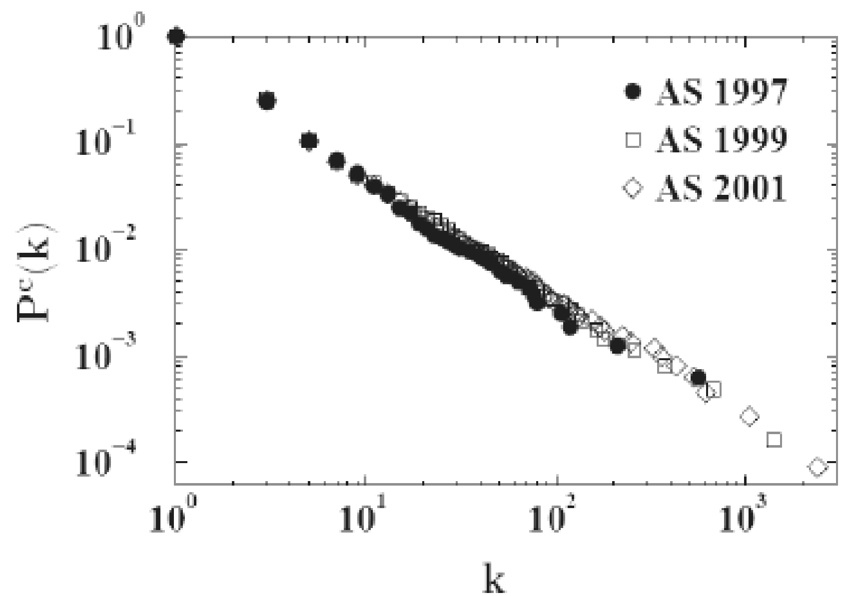
Note

- Node Degree Distribution (DD) tells how many nodes of a given degree are in the network but it does NOT provide information on the interconnection between these nodes
- Reason: given P(k), structure of the neighborhood of the average node of a given degree is still unknown

Degree Distribution (4)

Approximated by long tail power law distribution of node degree k: **$P(k) \sim k^{-\gamma}$**, where power-law exponent $\gamma = 2.254$





In practice, the distribution is not a strict power law

- The Internet contains more 2-degree nodes than 1-degree nodes
- The distribution has a longer tail, i.e. the maximum degree is much larger large than expected by the power-law

The Internet is characterized by a **fewer nodes with a large degree** a **large number of nodes with a low degree**

Source: Faloutsos et al (1999)

Joint Degree Distribution (1)

Definition: joint degree distribution (JDD) $P(k_1,k_2)$, or the *node degree correlation matrix* is the probability that a randomly selected edge connects k_1-degree and k_2-degree nodes

$$\boxed{P(k_1,k_2) = \mu(k_1,k_2)\frac{m(k_1,k_2)}{2m}}$$

where

$\mu(k_1,k_2) = 1$ if $k_1 = k_2$ and 2 otherwise

$m(k_1,k_2)$ is the total number of edges connecting nodes of degrees k_1 and k_2

- JDD contains more information about the graph connectivity than the degree distribution

  Reason: given a specific form of P(k1,k2) one can always restore both the degree distribution P(k) and $\bar{k}$

Joint Degree Distribution (2)

- Summary statistic of JDD: Average neighbor connectivity k_{nn}

$$k_{nn}(k) = \sum_{k'=1}^{k_{\max}} k' P(k' | k)$$

Average neighbor degree of the average k-degree node

- JDD shows whether AS of a given degree preferentially connect to high- or low-degree AS

- JDD provides more information than DD (information about 1-hop neighborhoods around a node) but JDD does not tell us how neighbors interconnect

- Note: in a full mesh graph, $k_{nn}(k)$ reaches its maximal possible value: $n - 1$. Therefore, for uniform graph comparison plot normalized values $k_{nn}(k)/(n - 1)$

JDD and Assortativity coefficient r

Summary statistic of JDD: assortativity coefficient r

$$r \sim \sum_{k_1,k_2=1}^{k_{max}} k_1 k_2 (P(k_1,k_2) - k_1 k_2 P(k_1) P(k_2)/\bar{k}^2)$$

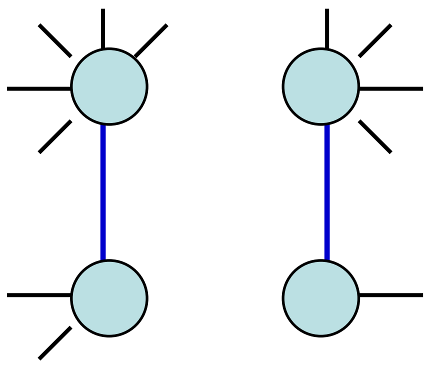
where $-1 \leq r \leq 1$

Interpretation of r

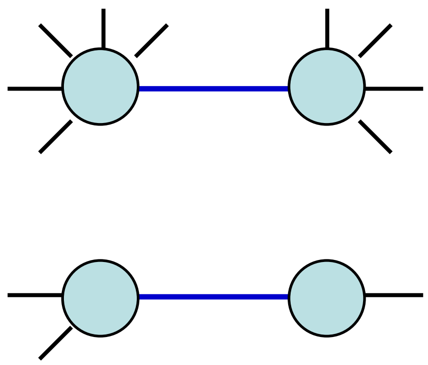
- Disassortative networks ($r < 0$) have an excess of radial links (links connecting high-degree nodes to low-degree nodes) i.e. links connecting nodes of dissimilar degrees
  - Cons: more vulnerable to both random failures and targeted attacks
  - Pros: vertex covers in disassortative graphs are smaller, which is important for applications such as traffic monitoring and prevention of DoS attack
- Assortative networks ($r > 0$) have an excess of tangential links i.e links connecting nodes of similar degrees

Assortative coefficient r

The Internet exhibits a negative correlation between a node’s degree k and its nearest-neighbors average degree



Diassortative mixing
$r < 0$



Assortative mixing
$r > 0$

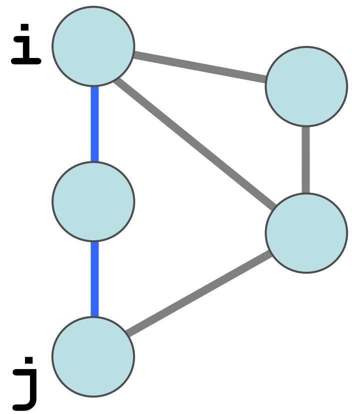
$\Rightarrow$ Disassortative mixing ($r = -0.236 < 0$): high-degree nodes tend to connect with low-degree nodes and visa versa

Characteristic Path Length

Definition: characteristic path length L (of a graph G=(V,E): average of path length d(i,j) over all pairs of vertices i, j ∈ V

$$L = \frac{1}{n(n-1)} \sum_{\substack{i,j \in V \\ i \neq j}} d(i,j)$$

Path length or distance d(i,j): number of edges of the shortest path between vertices i and j



d(i,j) = 2

Distance (1)

Definition: distance distribution d(x) is the probability that a random pair of nodes are at distance x of each other

- Divided by the total number of pairs n^2 (self-pairs included)

Associated statistics with distance distribution of a graph

- Average distance d_m
- Standard deviation σ (a.k.a distance distribution width since distance distributions in Internet graphs have a characteristic Gaussian-like shape)

Distance (2)

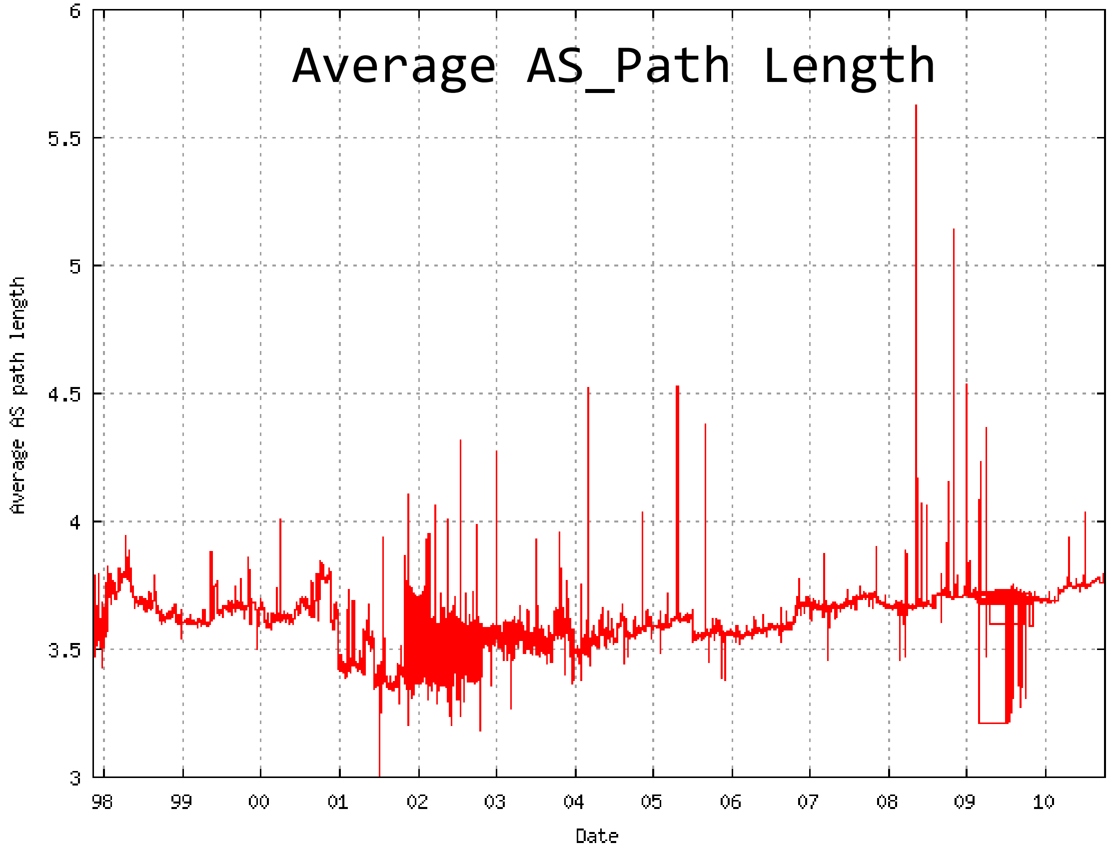
Interpretation

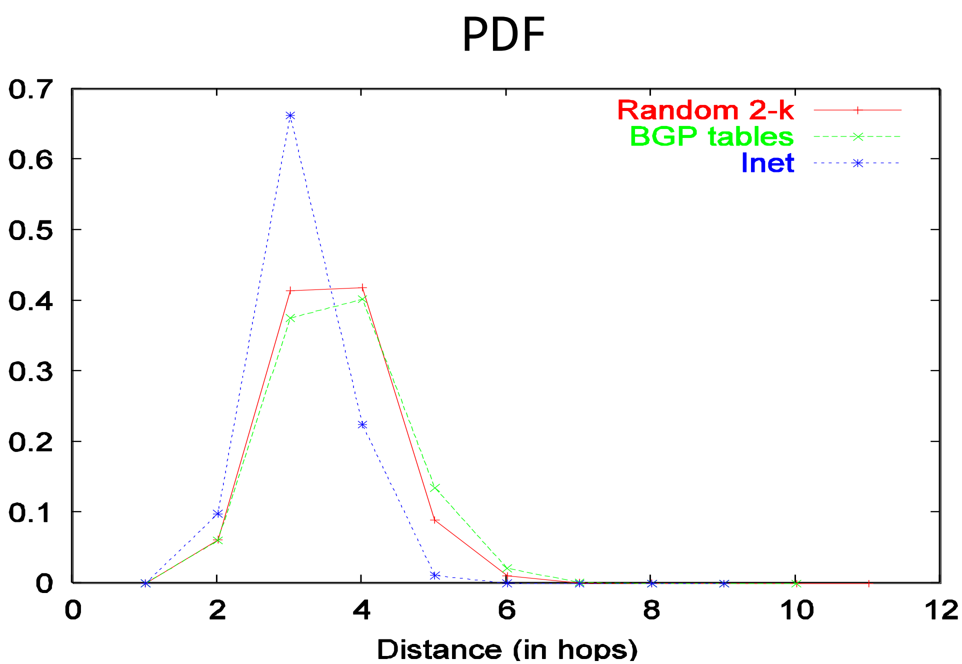
- Distance distribution is important for routing
  - Distance-based locality-sensitive approach as root of most modern routing algorithms: performance of routing algorithms depend mostly on the distance distribution
  - Short average distance and narrow distance distribution width break the efficiency of traditional hierarchical routing: root causes of inter-domain routing scalability issues in the Internet

Distance (3)

Distance distribution (shortest path length)

- Performance parameters of routing algorithms depend solely on distance distribution
- Internet: 86% of AS pairs are at distance 3 to 4 AS hops





Distance (4)

Consequence: efficient application of hierarchical, aggregation-based routing to Internet-like topologies is hopeless

- **Distance**: Hierarchical routing performs well for topologies where average distance d between nodes increases polynomially with network size n: `d(n) ~ n exp(m), m>0`

  > < Internet topology average distance d growths at most logarithmically with network size (n): `d(n) ~ log(n)`

- **Path length increase**: Hierarchical routing performs well when ratio

  $$\frac{\texttt{distance (a,b)}}{\texttt{routing path length (a,b)}} \rightarrow \texttt{cte}$$

  > < Internet characteristic routing path length is almost constant hence, ratio → ∞ for distance → ∞

  Note: applying hierarchical aggregation-based routing to the Internet AS-level topology would incur about 15-times AS-path length increase

Clustering (1)

Quantifies how close node's neighbors are to forming a clique (complete graph i.e. every pair of distinct vertices is connected by an edge)

Definition: Local clustering coefficient c(i) of vertex *i* of degree k_i (has k_i neighbors)

$$c(i) = |E(\Gamma(i))| \binom{k_i}{2}^{-1} = \frac{2|E(\Gamma(i))|}{k_i(k_i - 1)}$$

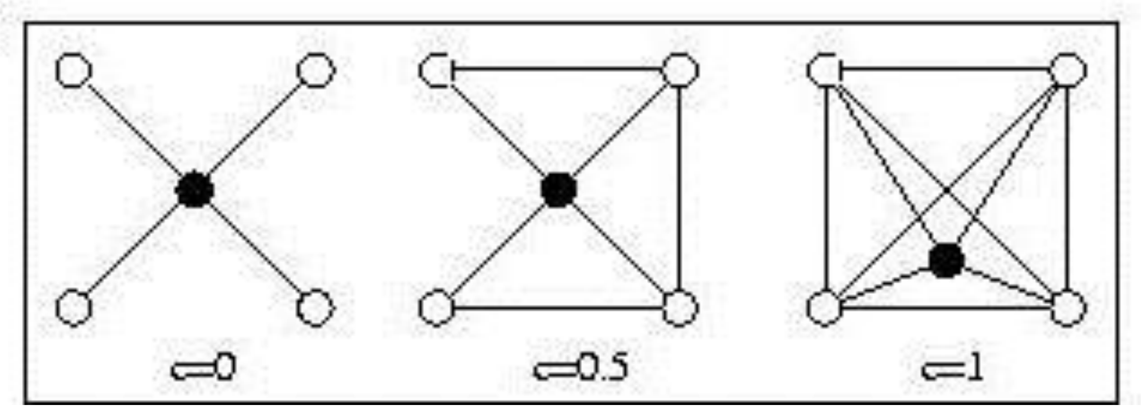
where

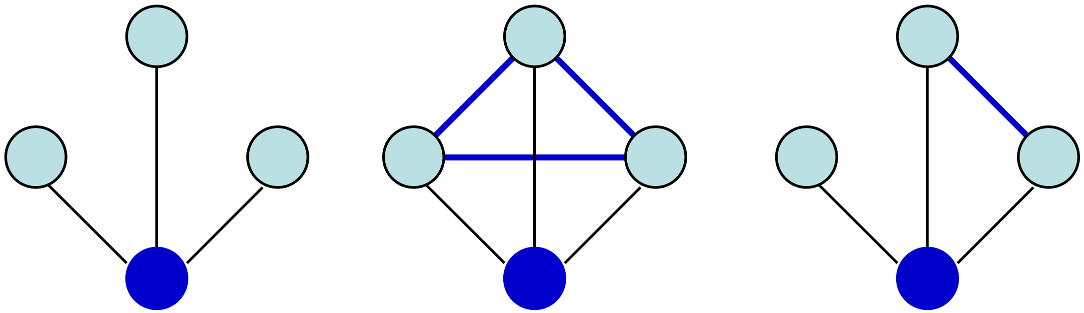
- $E|\Gamma(i)|$ is the number of edges in neighborhood of vertex i
- k_i (degree of vertex i): the number of edges incident to the vertex i
- $k_i(k_i-1)/2$ is the maximum possible number of edges between neighbors of vertex i (max. possible number of edges that could exist among the vertices within the neighbourhood of vertex i)

Clustering (2)

- If two neighbors of a node are connected, then these three nodes together form a triangle (3-cycle)

  ⇒ Local clustering measure of average number of 3-cycles





c=0 c=1 c=1/3

Clustering (3)

Definition: Clustering coefficient $\overline{C}$ of the graph G=(V,E) is the average of the local clustering coefficients of all the vertices |V| = n

$$\overline{C} = \frac{1}{n}\sum_{i=1}^{n} c(i)$$

- Clustering coefficient C_{coeff} measure of the percentage of 3-cycles among all connected node triplets in the graph

Interpretation

- Clustering is a measure of local robustness in the graph
- Implications
  - The higher the local clustering of a node, the more interconnected are its neighbors, thus increasing path diversity locally around the node
  - Networks with strong clustering are likely to be chordal or of low chordality, 4 which makes certain routing strategies perform better
  - Clustering used as litmus test for verifying the accuracy of a topology model or generator

Clustering (4)

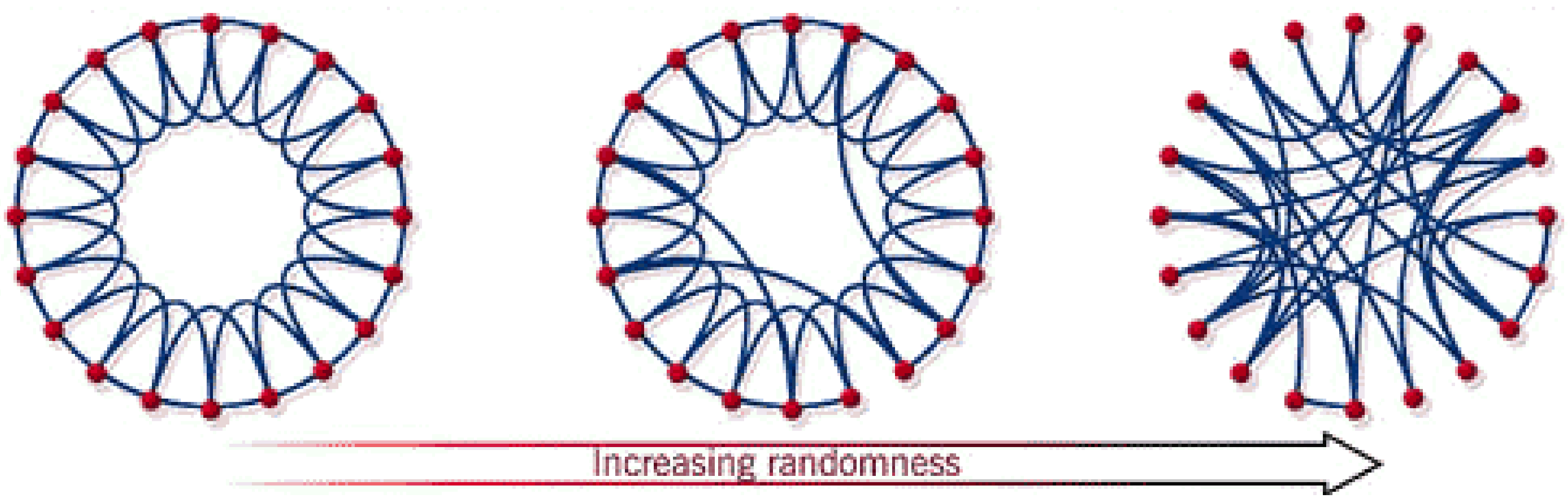
A graph is considered small-world

- if its clustering coefficient C is significantly higher than a random graph constructed on the same vertex set
- if the graph has approximately the same mean-shortest path length as its corresponding random graph

Regular graph
high *clustering*
large diameter

Small-world graph
high *clustering*
small diameter

Random graph
small *clustering*
small diameter

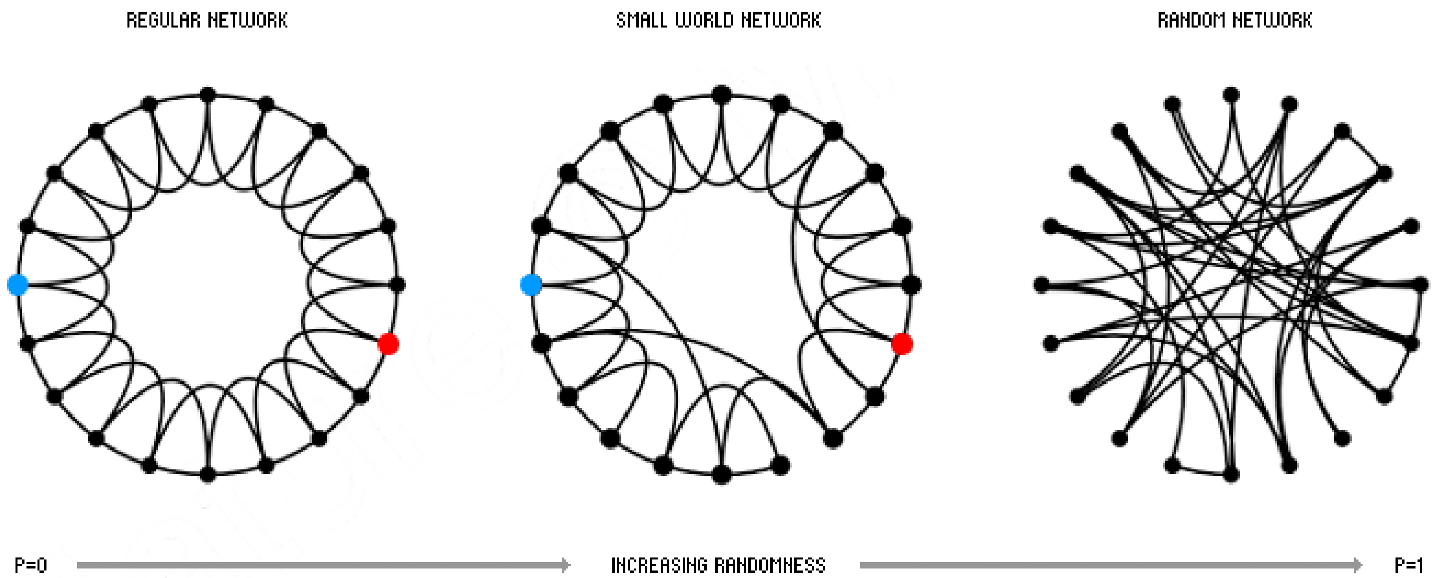


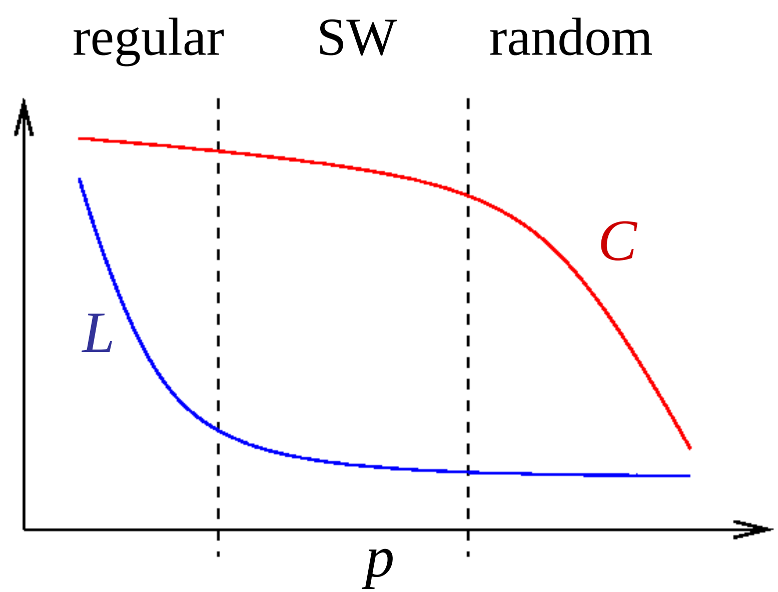
N = 1000 k =10
D = 100 L = 49.51
C = 0.67

N =1000 k = 8-13
D = 14 L = 11.1
C = 0.63

N = 1000 (k =5-18)
D = 5 L = 4.46
C = 0.01

Clustering (5)





C(p) = clustering coeff.

L(p) = characteristic path length

Betweeness (1)

Betweenness

- Most common metric to measure centrality
- Measures the number of shortest paths traversing a vertex(node) or edge(link) if each individuals send a message to all other individuals
- Estimation of the potential traffic load (flow of information) on this node/link assuming uniformly distributed traffic following shortest paths

Definition

- σ_{ij} : number of shortest paths between nodes i and j
- z : either a node or link
- $\sigma_{ij}(z)$: number of shortest paths between i and j going through z
  Betweeness B(z): $B_c(z) = \sum_{i,j} \frac{\sigma_{ij}(z)}{\sigma_{ij}}$
- The maximum possible value for node and link betweenness is n(n – 1) → to compare betweenness in graphs of different sizes, normalization by n(n – 1)

Betweeness (2)

Definition

- σ_{ij} : number of shortest paths between nodes i and j
- x : node
- $\sigma_{ij}(x)$: number of shortest paths between i and j going through node x

  Betweeness: $B_c(x) = \sum_{\substack{i \neq x \neq j \\ i,x,j \in V}} \frac{\sigma_{ij}(x)}{\sigma_{ij}}$

Definition

- σ_{ij} : number of shortest paths between nodes i and j
- y : link
- $\sigma_{ij}(y)$: number of shortest paths between i and j going through link y

  Betweeness: $B_c(y) = \sum_{\substack{i \neq j \in V \\ y \in E}} \frac{\sigma_{ij}(y)}{\sigma_{ij}}$

Betweeness (3)

Interpretation

- Important metric for traffic engineering applications that try to estimate potential traffic load on nodes/links and potential congestion points in a given topology

- Critical for evaluating the accuracy of topology sampling by tree-like probes (e.g. BGP)
  - The broader the betweenness distribution, the higher the statistical accuracy of the sampled graph
  - Note: exploration process statistically focuses on nodes/links with high betweenness thus providing an accurate sampling of the distribution tail and capturing relevant statistical information

- Note: link betweenness is not a measure of centrality but a measure of a certain combination of link centrality and radiality

Spectrum (1)

Definition

- A : n × n adjacency matrix of a graph constructed by setting the value of its element as
  - $a_{ij} = a_{ji} = 1$ if there is a link between nodes i and j
  - all other elements have value 0
- Scalar λ are the eigenvalue and vector v the eigenvector of A if A v = λ v
- Spectrum of a graph is the set of eigenvalues λ of its adjacency matrix A

Interpretation: (one of the) most important global characteristics of the topology

- Provides bounds for critical graph characteristics such as distance-related parameters, expansion properties, and values related to separator problems estimating graph resilience under node/link removal
- Most networks with high values as eigenvalues have small diameter, expand faster, and are more robust

Spectrum (2)

Example of spectrum-related metrics

- Robustness of network
  - Critical metric for topology comparison analysis
  - Measure of network robustness under link removal (equals minimum balanced cut size of a graph)
  - Relation to spectrum: graph's largest eigenvalues provide bounds on network robustness with respect to both link and node removals
- Performance: Maximum traffic throughput of network
  - Relation to spectrum: network conductance can be tightly estimated by the gap between the first and second largest eigenvalues

Spectrum (3)

Application to Traffic engineering

- Graphs with larger eigenvalues have, in general, more node- and link-disjoint paths to choose from

Spectral analysis

- Powerful tool for detailed investigation of network structure
- Example: discovering clusters of highly interconnected nodes and revealing AS hierarchy

Example: Five networks with the same node degree distribution

(a) Node degree distribution (degree versus rank on log-log scale)

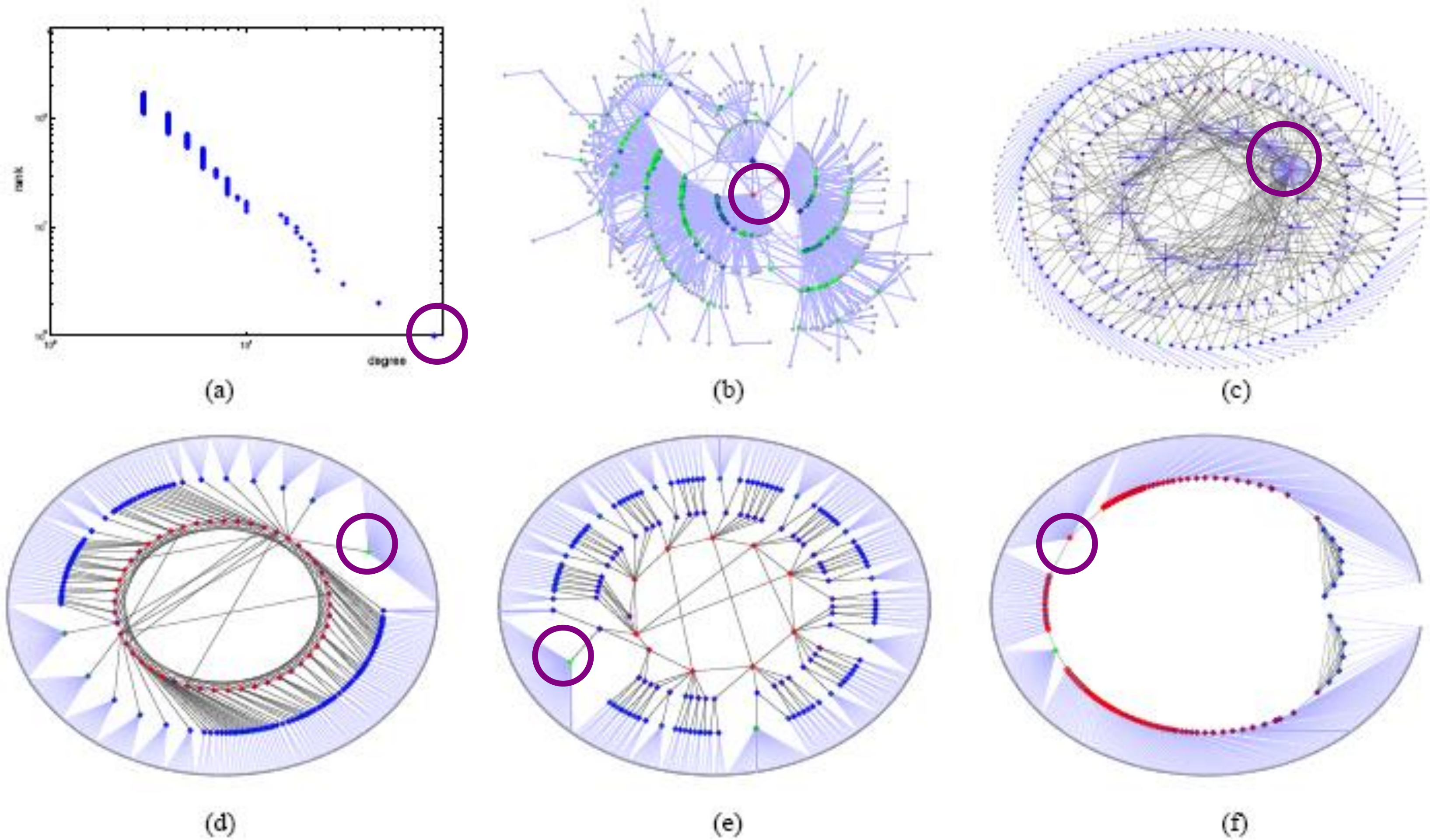
(b) Network resulting from PA

(c) Network resulting from the general model of random graphs (GRG) method with a given expected degree sequence

(d) Heuristically optimal topology (HOT) using Power Law Random Graph (PLRG)

(e) Abilene-inspired topology

(f) Sub-optimally designed topology



(a) (b) (c)

(d) (e) (f)

Scaling dependency on Topology

Internet topological properties characterized by

- Node degree (k) distribution: approximated by long tail power law distribution $P(k) \sim k^{-\gamma}$, $\gamma = 2.254$ (scaling index)

  The Internet is characterized by a fewer nodes with a large degree a large number of nodes with a low degree

- Node degree correlation: negative correlation between a node's degree k and its nearest-neighbors average degree

  Disassortative mixing ($r = -0.236 < 0$): high-degree nodes tend to connect with low-degree nodes and visa versa

- Clustering coefficient: characterizes the extent to which vertices adjacent to any vertex v are adjacent to each other) = 0.4

  Strong clustering means large number of triangular sub-graphs (>< regular tree structure)

- Characteristic path length: median of the means of shortest path lengths connecting each node to all other nodes ~ 3.7

- Average distance between nodes grows proportionally to log(n), where n is number of nodes

Backup Material

Power-law distributions

Power-laws are laws of the form: $P(k) = C\,k^{-\gamma}$

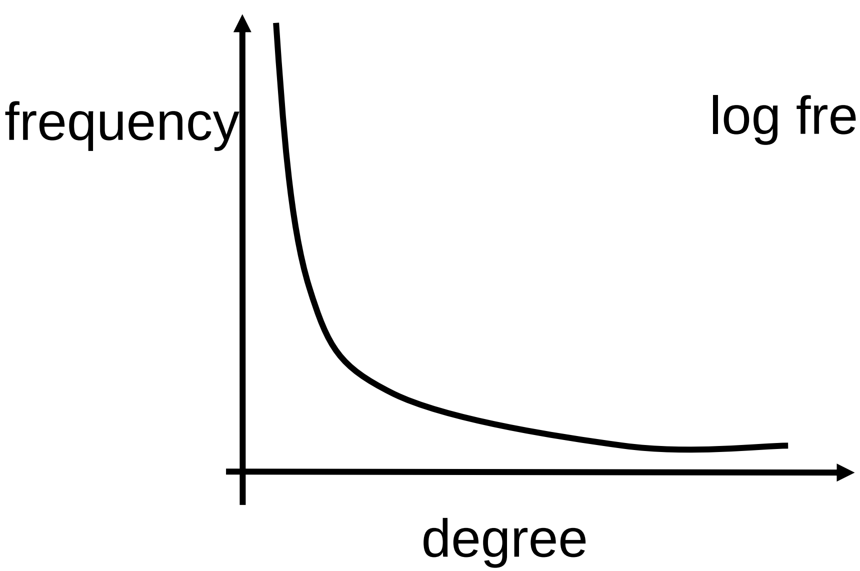
where

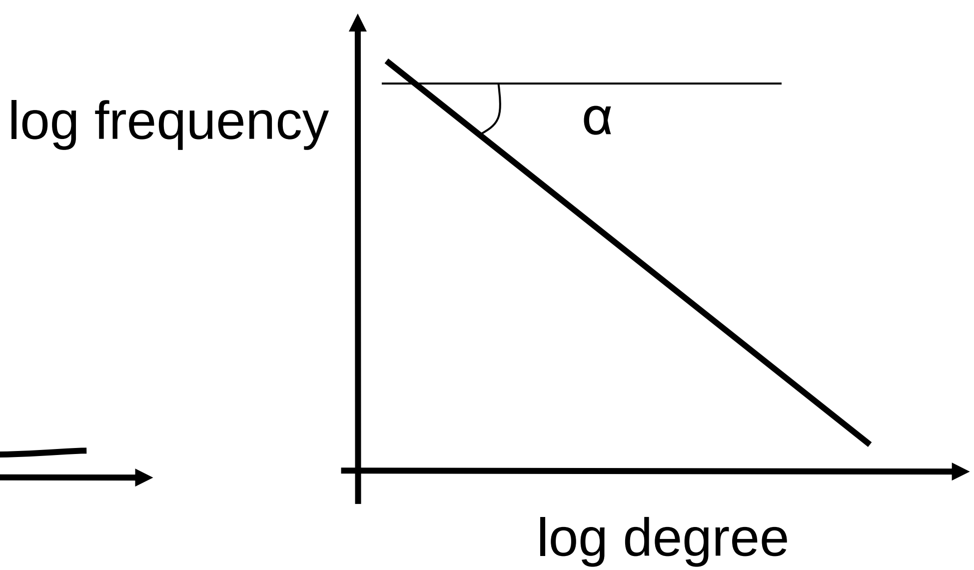
- γ : scale index (or power law exponent, typically $2 \le \gamma \le 3$)
- C : constant

Properties of power laws

$$P(k) = C\,k^{-\gamma} \Leftrightarrow \log(P(k)) = -\gamma \log(k) + \log C$$

Power-law distribution gives a line in log-log plot

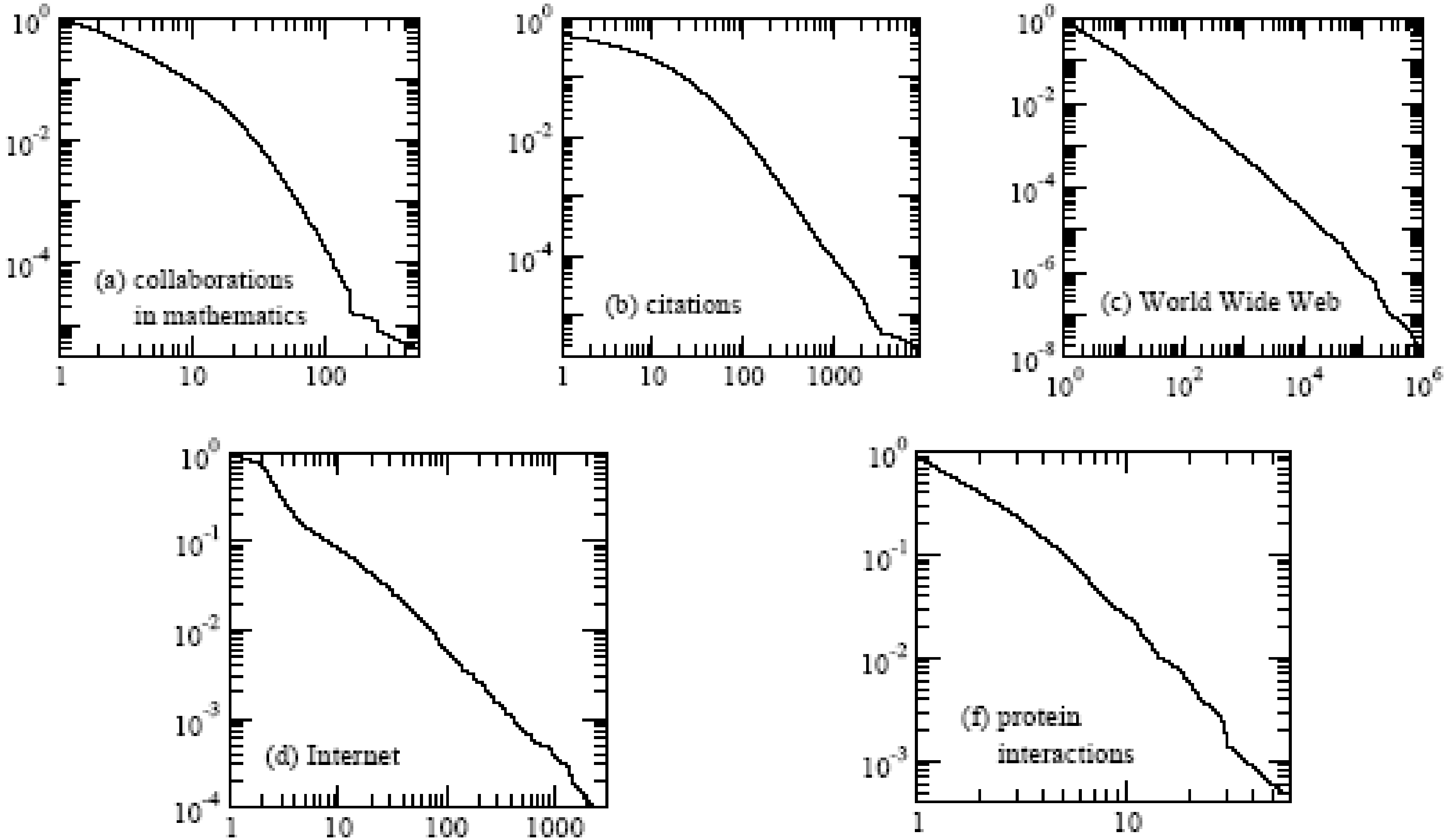




Power-law distributions: Examples

Heavy-tail distribution

- non-negligible fraction of nodes has very high degree (hubs)



Source [Newman 2003]

References

References

[Aiello00] W. Aiello, F. Chung, and L. Lu, *A Random Graph Model for Massive Graphs*, Proceedings of ACM Symposium on Theory of Computing, pp.171–180, 2000.

[Aiello01] W. Aiello, F. Chung, and L. Lu, *A Random Graph Model for Power Law Graphs*, Experimental Math. 10, pp.53–66, 2001.

[Aiello02] W. Aiello, F. Chung, and L. Lu, *Random Evolution in MassiveGraphs*, In Handbook on Massive Data Sets, Vol.2, pp.97-122, Spinger, 2002.

[Albert00a] R. Albert, H. Jeong, and A.-L. Barabasi, *Attack and error tolerance of complex networks*, Nature 406, pp.378-382, 2000.

[Albert00b] R. Albert and A.-L. Barabasi, *Topology of Evolving Networks: Local Events and Universality*, Physical Review Letters, Vol.85, No.24, pp.5234-5237, 2000.

[Albert02] R. Albert and A.-L. Barabási, *Statistical mechanics of complex networks*, Reviews of Modern Physics, Vol.74, pp.47-97, 2002.

[Alderson05a] D. Alderson, L. Li, W. Willinger, and J.C. Doyle, *Understanding Internet Topology: Principles, Models, and Validation*, ACM/IEEE Transactions on Networking, Vol.13, No.6, 2005.

[Alderson05b] D. Alderson, and W. Willinger, *A contrasting look at self-organization in the Internet and next-generation communication networks*, IEEE Communications Magazine, July 2005.

[Barabasi99a] A.-L. Barabasi, and R. Albert, *Emergence of scaling in random networks*, Science, Vol.286, No.5439, pp.509-512, 1999.

[Barabasi99b] A.-L. Barabasi, R. Albert, and H. Jeong, *Mean-field theory for scale-free random networks*, Physica A 272, pp.173-187, 1999.

[Barabasi01] A.-L. Barabasi, E. Ravasz, and T. Vicsek, *Deterministic scale-free networks*, Physica A, Vol.299, No.599, 2001.

[Barabasi03a] A.-L. Barabasi, *Linked: How Everything Is Connected to Everything Else and What It Means*, Plume, 2003.

[Barabasi03b] A.-L.Barabasi, Z. Dezso, E. Ravasz, S.H.Yook, and Z. Oltvai, *Scale-free and Hierarchical Structures in Complex Networks*, AIP Conference Proceedings, 2003.

References

[Bollobas01] B. Bollobas, O. Riordan, J. Spencer, and G. Tusnady, *The Degree Sequence of a Scale-Free Random Process*, Random Structures and Algorithms, Vol.18, No.3, pp.279–290, 2001.

[Bollobas03a] B. Bollobas, and O. Riordan, *Robustness and Vulnerability of Scale-Free Random Graphs, Internet Mathematics*, Vol.1, No.1, pp.1-35, 2003.

[Bollobas03b] B. Bollobas, and O. Riordan, *Coupling Scale-Free and Classical Random Graphs, Internet Mathematics*, Vol.1, No.2, pp.215-225, 2003.

[Bollobas04] B. Bollobas, and O. Riordan, *The Diameter of a Scale-Free Random Graph*, Combinatorica, Vol.24, No.1, pp.5–34, 2004.

[Bu02] T. Bu, and D. Towsley, *On distinguishing between Internet power law topology generators*, Proc. of IEEE INFOCOM 2002, Dec. 2002.

[Carlson99] J. M. Carlson, and J. Doyle, *Highly optimized tolerance: A mechanism for power laws in designed systems*, Physical Review E, Vol.60, pp.1412-1427, 1999.

[Chang03] H. Chang, S. Jamin, and W. Willinger, *Internet connectivity at the AS-level: An optimization-driven modeling approach*, Proc. of MoMeTools, 2003.

[Chang04] H. Chang, R. Govindan, S. Jamin, S.J. Shenker, and W. Willinger, *Towards capturing representative AS-level Internet topologies*, Computer Networks Journal, Vol.44, pp.737-755, April 2004.

[Chang06] H. Chang, S. Jamin, and W. Willinger, *To peer or not to peer: Modeling the evolution of the Internet's AS-level topology*, Proc. of IEEE INFOCOM, 2006.

[Chakrabarti06] D. Chakrabati, and C. Faloutsos, *Graph mining: Laws, generators, and algorithms*, ACM Comput. Surv., Vol.38, No.1, 2006.

[Chen02] Q. Chen, H. Chang, R. Govindan, S. Jamin, S.J. Shenker, and W. Willinger, *The origin of power laws in Internet topologies revisited*, Proc. of IEEE INFOCOM 2002, June 2002.

[Chung03a] F. Chung, and L. Lu, *The Average Distance in a Random Graph with Given Expected Degrees*, Internet Mathematics, Vol.1, No.1, pp.91–114, 2003.

References

[Chung03b] F. Chung, L. Lu, and V. Vu, *The Spectra of Random Graphs with Given Expected Degrees*, Internet Mathematics, Vol.1, No.3, pp.257-275, 2003.

[Chung04] F. Chung, and L. Lu, *Coupling Online and Offline Analyses for Random Power Law Graphs*, Internet Mathematics, Vol.1, No.4, pp.409-461, 2004.

[Cooper04] C. Cooper, A. Frieze, and J. Vera, *Random Deletion in a Scale-Free Random Graph Process*, Internet Mathematics, Vol.1, No.4, pp.463-483, 2004.

[DallAsta05] L. Dall'Asta, I. Alvarez-Hamelin, A. Barrat, A. Vazquez, and A. Vespignani, *Exploring networks with traceroute-like probes: Theory and simulations*, Theoretical Computer Science, Special Issue on Complex Networks, 2005.

[Dimitropoulos06] X. Dimitropoulos, D. Krioukov, G. Riley, and K. Claffy, *Revealing the Autonomous System taxonomy: The machine learning approach*, Proc. of PAM 2006, March 2006.

[Dimitropoulos07] X. Dimitropoulos, D. Krioukov, M. Fomenkov, B. Huffaker, Y. Hyun, K. Claffy, and G. Riley, *AS relationships: Inference and validation*, Computer Communication Review, Vol.37, No.1, 2007

[Dorogovstev03] S. N. Dorogovtsev, and J.F.F. Mendes, *Evolution of Networks*, Oxford:England, Oxford University Press, 2003.

[Doyle05] J.C. Doyle, D. Alderson, L. Li, S. Low, M. Roughan, S. Shalunov, R. Tanaka, and W. Willinger, *The "robust yet fragile" nature of the Internet*, PNAS 102(41), 2005.

[Erdos61] P. Erdos and A. Renyi, *On the strength of connectedness of random graphs*, Acta Math. Acad. Sci. Hungar, Vol.12, pp.261-267, 1961.

[Erdos60a] P. Erdos, and A. Rényi, *On the evolution of random graphs*, Publ. Math. Inst. Hung. Acad. Sci, Vol.5, pp.17-61, 1960.

[Erdos60b] P. Erdos, and T. Gallai, *Graphs with prescribed degrees of vertices*, Mat. Lapok (Hungarian), Vol.11, pp.264-274, 1960.

[Erdos59] P. Erdos, and A. Renyi, *On random graphs*, I Publ. Math. (Debrecen), Vol.9, pp.290-297, 1959.

[Fabrikant02] A. Fabrikant, E. Koutsoupias, and C.H. Papadimitriou, *Heuristically Optimized Tradeoffs: A New Paradigm for Power Laws in the Internet*, Proceedings of ICALP 2002, Lecture Notes in Computer Science 2380, pp.110–122, Springer–Verlag, 2002.

References

[Faloutsos99] M. Faloutsos, P. Faloutsos, and C. Faloutsos, *On power-law relationships of the Internet topology*, Proc. ACM SIGCOMM 1999 and in ACM Computer Communication Review, Vol.29, pp.251-263, 1999.

[Jin00] C. Jin, Q. Chen, and S. Jamin, *INET topology generator*, Technical Report, EECS Department, University of Michigan, 2000.

[Jin02] S. Jin, and A. Bestavros, *Small-world internet topologies: Possible causes and implications on scalability of end-system multicast*, Technical Report, Boston University, 2002.

[Krapivsky01] P. L. Krapivsky, and S. Redner, *Organization of growing random networks*, Physical Review E, Vol.63, No.6, May 2001.

[Krioukov04] D. Krioukov, K. Fall, and X. Yang, *Compact routing on Internet-like graphs*, Proc. IEEE INFOCOM, 2004.

[Li04] L. Li, D. Alderson, W. Willinger, and J. Doyle, *A first-principles approach to understanding the Internet's router-level topology*, Proc. of ACM SIGCOMM, 2004.

[Li05] L. Li, D. Alderson, J.C. Doyle, and W. Willinger, *Toward a Theory of Scale-Free Networks: Definition, Properties, and Implications*, Internet Mathematics, Vol.2, No.4, pp.431-523, 2005.

[Mahadevan06a] P. Mahadevan, D. Krioukov, M. Fomenkov, B. Huffaker, X. Dimitropoulos, kc claffy, and A. Vahdat, *The Internet AS-level topology: Three data sources and one definitive metric*, ACM Computer Communication Review, Vol.36, No.1, 2006.

[Mahadevan06b] P. Mahadevan, D. Krioukov, K. Fall, and A. Vahdat, *Systematic topology analysis and generation using degree correlations*, Proc. of ACM SIGCOMM, 2006.

[Medina00] A. Medina, I. Matta, and J. Byers, *On the origin of power laws in Internet topologies*, ACM Computer Communication Review, Vol.30, No.2, pp.18-28, April 2000.

[Mihail02] M. Mihail, C.H. Papadimitriou, *On the Eigenvalue Power Law*, RANDOM, pp.254-262, 2002.

References

[Newman02] M.E.J. Newman, *Assortative Mixing in Networks*, Phys. Rev. Lett. Vol.89, p.208701, 2002.

[Newman03] M.E.J. Newman, *The Structure and Function of Complex Networks*, SIAM Reviews, vol.45, no.2, pp.167-256, 2003.

[Newman04] Juyong Park and M. E. J. Newman, Statistical mechanics of networks, PHYSICAL REVIEW E 70, 066117 (2004)

[Newman05] M.E.J. Newman, *Power laws, Pareto distributions and Zipfs law*, Contemporary Physics, vol.46, no.5, pp.323–351, January 2005.

[Oliveira03] R. Oliveira, and J. Spencer, *Connectivity Transitions in Networks with Super-Linear Preferential Attachment*, Internet Mathematics, vol.2, no.2, pp.121-163, 2003.

[Palmer01] C. Palmer, G. Siganos, M. Faloutsos, C. Faloutsos and P. Gibbons, *The connectivity and fault-tolerance of the Internet topology*, NRDM 2001, Santa Barbara, CA, May 2001.

[Radoslavov00] P. Radoslavov, H. Tangmunarunkit, H. Yu, R. Govindan, S. Shenker, D. Estrin, *On characterizing network topologies and analyzing their impact on protocol design*, USC-CS-TR-00-731, March 2000.

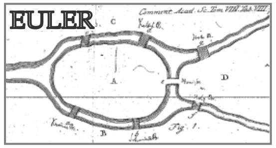
[Serrano06] M. A. Serrano, M. Boguna, and A. Daz-Guilera, *Modeling the Internet*, The European Physical Journal B, vol.50, pp.249-254, 2006.

[Siganos03] G. Siganos, M. Faloutsos, P. Faloutsos, and C. Faloutsos, *Power Laws and the AS-level Internet topology*, IEEE/ACM Transactions on Networking, vol.11, no.4, pp.514-524, August 2003.

[Tangmunarunkit02] H. Tangmunarunkit, R. Govindan, S. Jamin, S. Shenker, W. Willinger, *Network Topology Generators: Degree-Based vs. Structural*, Proc. of ACM SIGCOMM 2002.

[Vukadinovic02] D. Vukadinovic, P. Huang, T. Erlebach, *On the Spectrum and Structure of Internet Topology Graphs*, Proc. of I2CS 2002.

[Waxman98] Waxman, *Routing of multipoint connections*, IEEE JSAC, 1988.



References

[Willinger04a] W. Willinger, D Alderson, J.C. Doyle, and L. Li, *More "normal" than Normal: scaling distributions in complex systems*, Proc. of Winter Simulation Conference (WSC), 2004.

[Willinger04b] W. Willinger, D Alderson, and L. Li, *A pragmatic approach to dealing with high-variability in network measurements*, Proc. of ACM SIGCOMM 2004.

[Xang06] X. Wang and D. Loguinov, *Wealth-based evolution model for the Internet AS-level topology*, Proc. of IEEE INFOCOM, 2006.

[Yook02] S.-H. Yook, H. Jeong, and A.-L. Barabasi, *Modeling the Internets large-scale topology*, PNAS, Vol.99, pp.13382-86, 2002.

[Zegura97] E. W. Zegura, K.L. Calvert and M.J. Donahoo, *A Quantitative Comparison of Graph-based Models for Internet Topology*, IEEE/ACM Transactions on Networking, vol.5, no.6, Dec.1997.

[Zhou04] S. Zhou and R. J. Mondragon, *Accurately modeling the Internet topology*, Physical Review E, vol.70, p.066108, 2004.