# Overlay (and P2P) Networks

## Part II

Samu Varjonen

**Ashwin Rao**

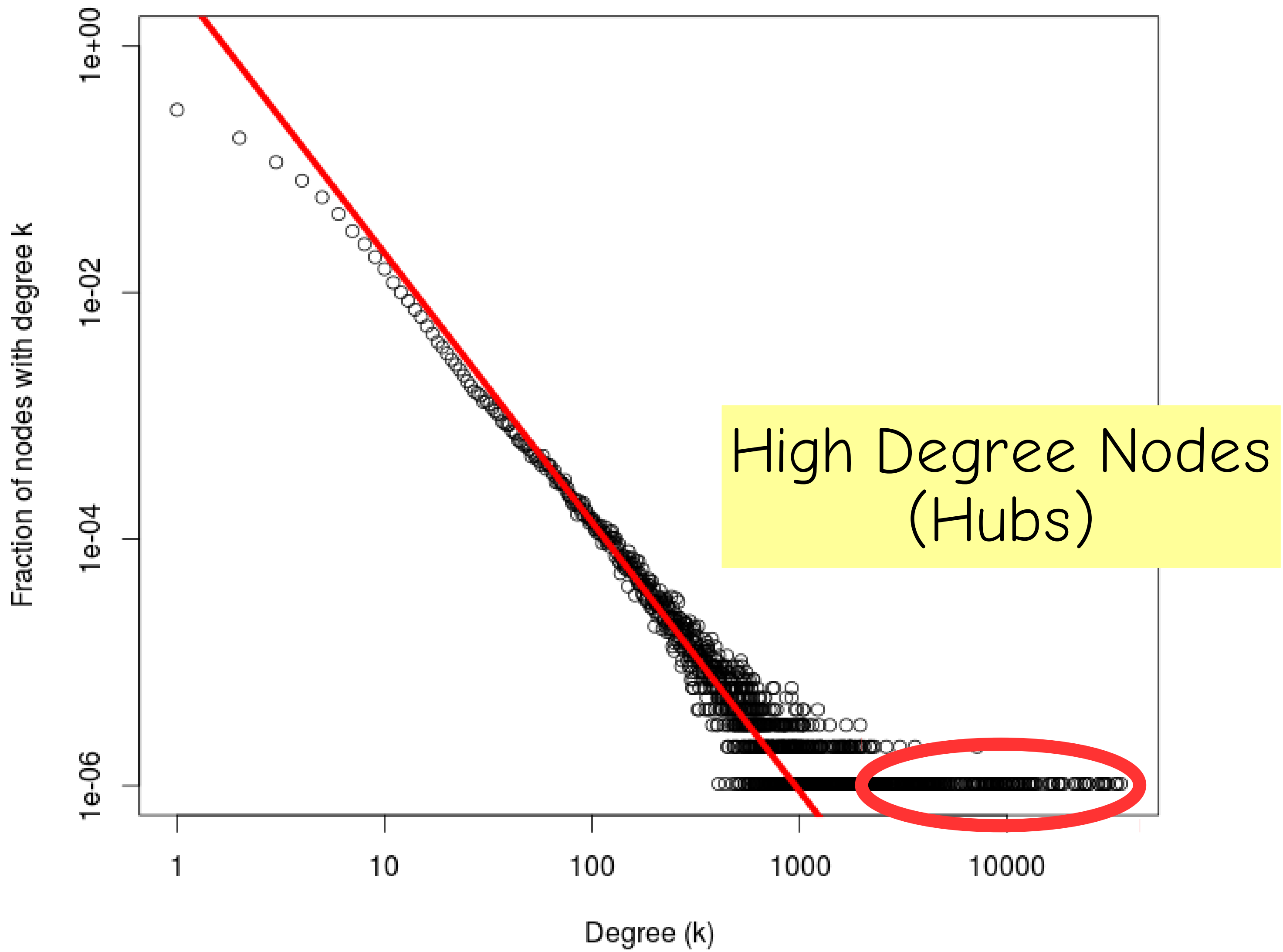
# Models of Complex Networks to Model Overlay Networks

- Milgram's Experiment
- Duncan Watts Random Rewiring Model
- Scale Free Networks (Power-Law Networks)
  - Preferential attachment
  - Evolving Copying Model (Copying Generative Model)
- Navigation in Small World

# Outline for this lecture

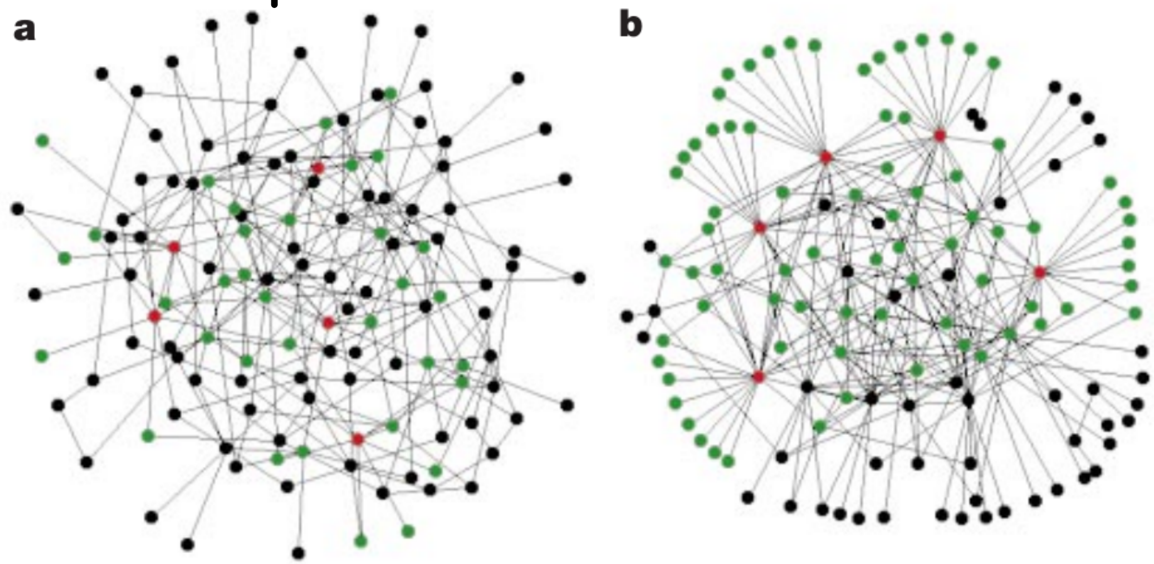
- Error and Attack Tolerance of Complex Networks
- Navigation in Complex Networks
- Mathematics and the Internet: A Source of Enormous Confusion and Great Potential
- Summary on Modeling Overlay Networks

# Scale-Free Model for AS-Graph




AS Topology of skitter dataset parsed by SNAP team
http://snap.stanford.edu/data/as-skitter.html

# Importance of Hubs




(Random Graph) Exponential

Scale-free

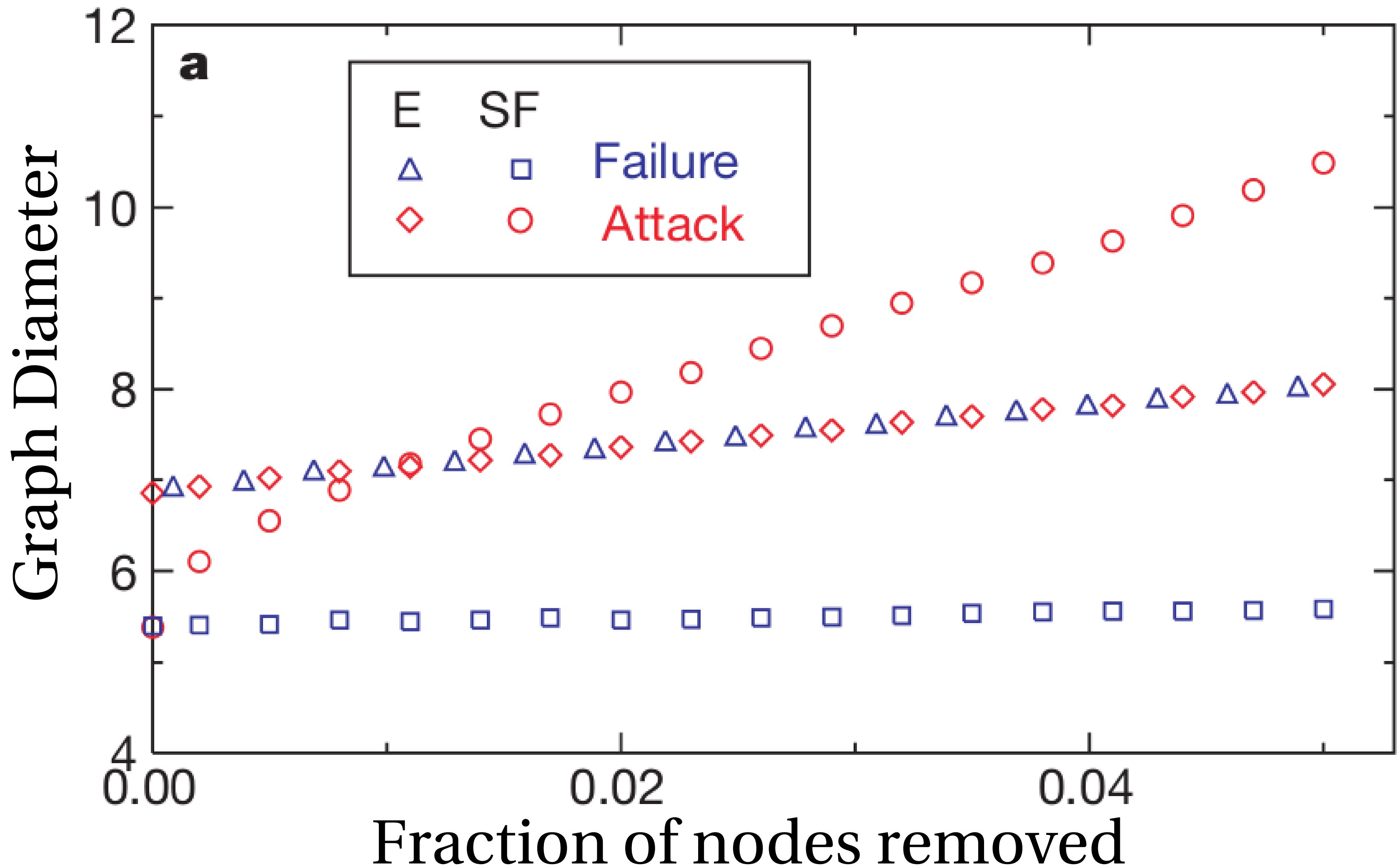
**Figure 1** Visual illustration of the difference between an exponential and a scale-free network. **a**, The exponential network is homogeneous: most nodes have approximately the same number of links. **b**, The scale-free network is inhomogeneous: the majority of the nodes have one or two links but a few nodes have a large number of links, guaranteeing that the system is fully connected. Red, the five nodes with the highest number of links; green, their first neighbours. Although in the exponential network only 27% of the nodes are reached by the five most connected nodes, in the scale-free network more than 60% are reached, demonstrating the importance of the connected nodes in the scale-free network Both networks contain 130 nodes and 215 links ($\langle k \rangle = 3.3$). The network visualization was done using the Pajek program for large network analysis: ⟨http://vlado.fmf.uni-lj.si/pub/networks/pajek/pajekman.htm⟩.

Albert, Réka, et al. **"Error and attack tolerance of complex networks."** nature 406, no. 6794 (2000): 378-382.

# Error vs Attack

- Error (Node Failure)
  - random node fails (malfunction)
- Attack
  - Selected node with a given property is made to fail
  - Which nodes would you target if you knew the network is a scale-free network?
    - *Nodes with the highest degree*

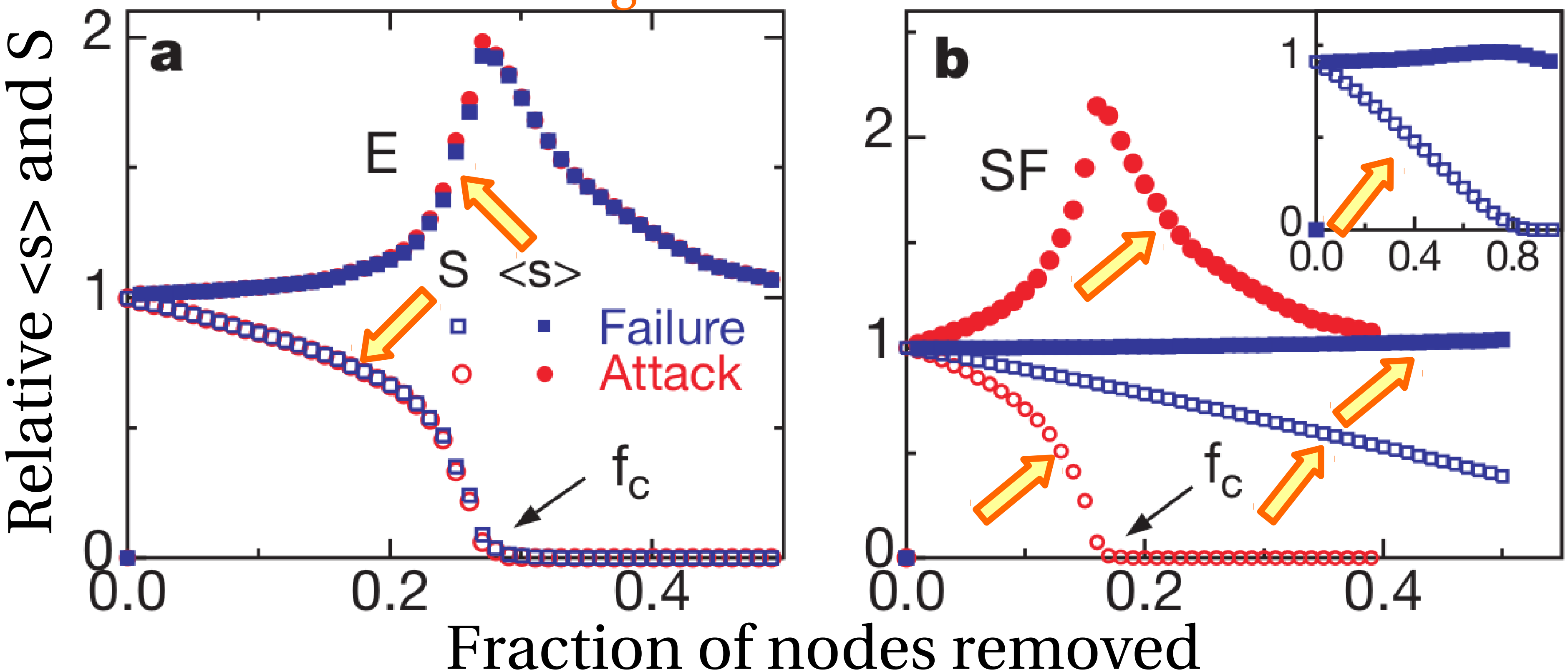
# Impact of Errors and Attacks (Graph Diameter)




Albert, Réka, et al. **"Error and attack tolerance of complex networks."** nature 406, no. 6794 (2000): 378-382.

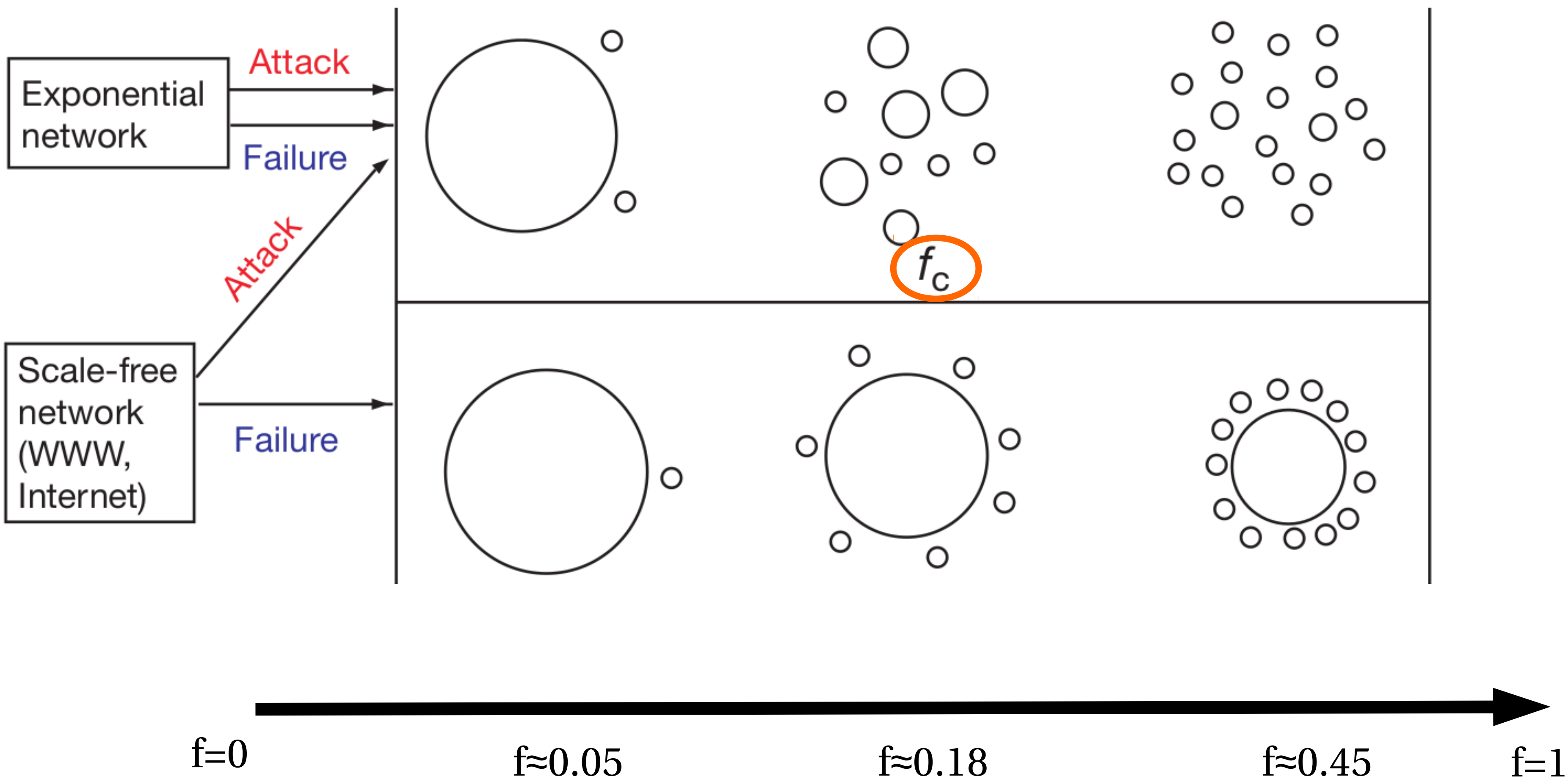
# Impact of Errors and Attacks (Size of Largest Cluster)

S: Fraction of nodes in largest cluster
<s>: average size of isolated clusters




Albert, Réka, et al. **"Error and attack tolerance of complex networks."** nature 406, no. 6794 (2000): 378-382.

# Network Response to Attacks and Failures

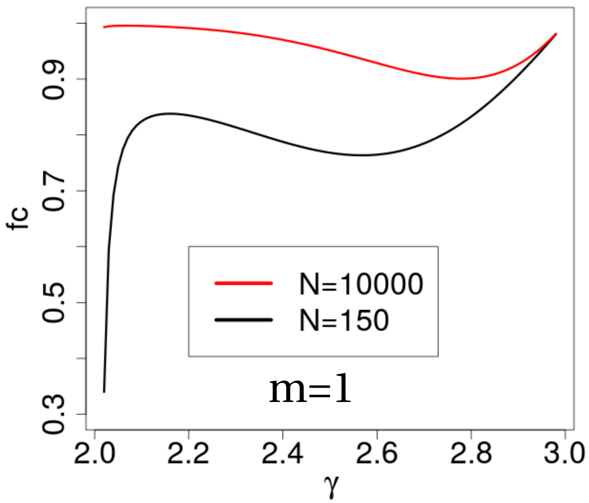



Albert, Réka, et al. **"Error and attack tolerance of complex networks."** nature 406, no. 6794 (2000): 378-382.

# Critical Threshold

(random node failures)

$$f_c = 1 - \frac{1}{\beta - 1} \quad \begin{cases} \gamma : \text{ exponent of power-law} \\ m : \text{ smallest degree} \\ N : \text{ number of nodes in the graph} \\ K : \text{ largest degree , } K \approx mN^{\frac{1}{\gamma-1}} \end{cases}$$




where

$$\beta = \frac{|2-\gamma|}{|3-\gamma|} \times \begin{cases} m & \text{if } \gamma > 3 \\ m^{\gamma-2}K^{3-\gamma} & \text{if } 2 < \gamma < 3 \\ K & \text{if } 1 < \gamma < 2 \end{cases}$$

for $2 < \gamma < 3$

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

Cohen's technique can be extended to errors

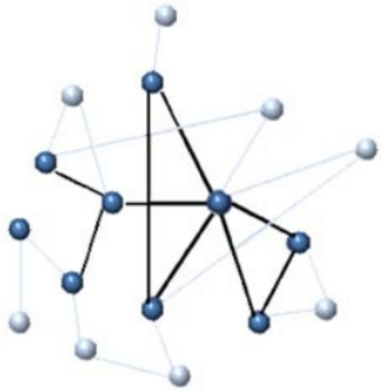
(No closed form for $f_c$ for errors )

Cohen, Reuven et al. "**Resilience of the Internet to random breakdowns.**" Physical review letters 85, no. 21 (2000): 4626.

# Summary on Attack and Error Tolerance of Complex Networks

Scale-free networks resilient to random failures but vulnerable to targetted attacks

# Clustering of Nodes



$k_T$: extract subgraph $G(k_T)$ with nodes having degree $k > k_T$

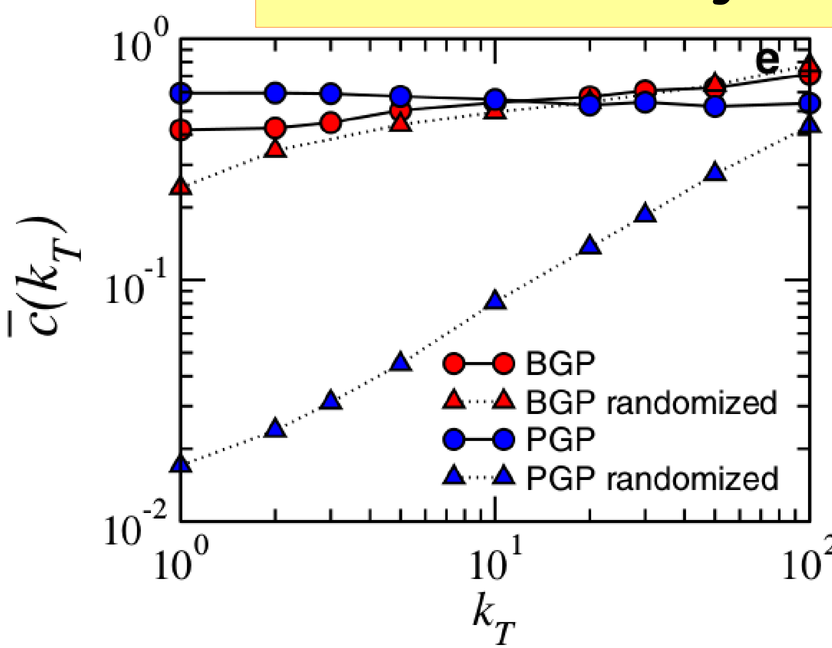
$< k_i(k_T) >$ : average degree of this subgraph $G(k_T)$

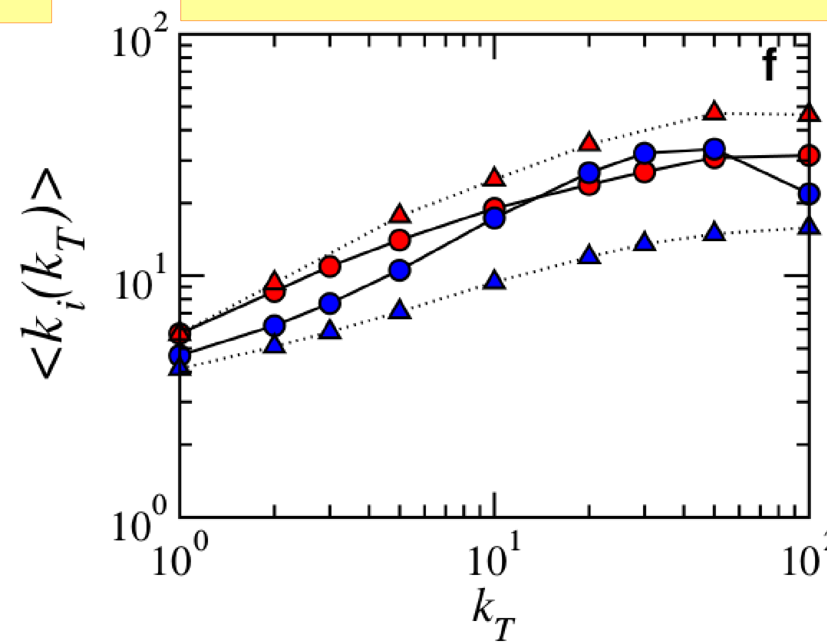
$\bar{c}(k_T)$: clustering coefficient of subgraph $G(k_T)$

$k_T = 2$

**Constant clustering coeff.**

**Average Degree Increases**




**How to Generate Scale-Free Graphs with Strong Clustering**

Serrano, M. Angeles, Dmitri Krioukov, and Marián Boguná. **"Self-similarity of complex networks and hidden metric spaces."** Physical review letters 100, no. 7 (2008): 078701.

# Generating Scale-Free Graphs with Strong Clustering

Take all nodes and distribute them within an underlying circle

Assign each node an expected degree k where P(k) $\sim k^{-\gamma}$

Connect each pair of nodes with a connection probability r(d; k,k′)

d is the distance between these two nodes in the circle

$\mathrm{d}_c = kk'$ is also called the characteristic distance

$$\mathrm{r}\ (\mathrm{d};\ \mathrm{k},\ \mathrm{k}') = \left(1 + \frac{d}{d_c}\right)^{-\alpha}$$

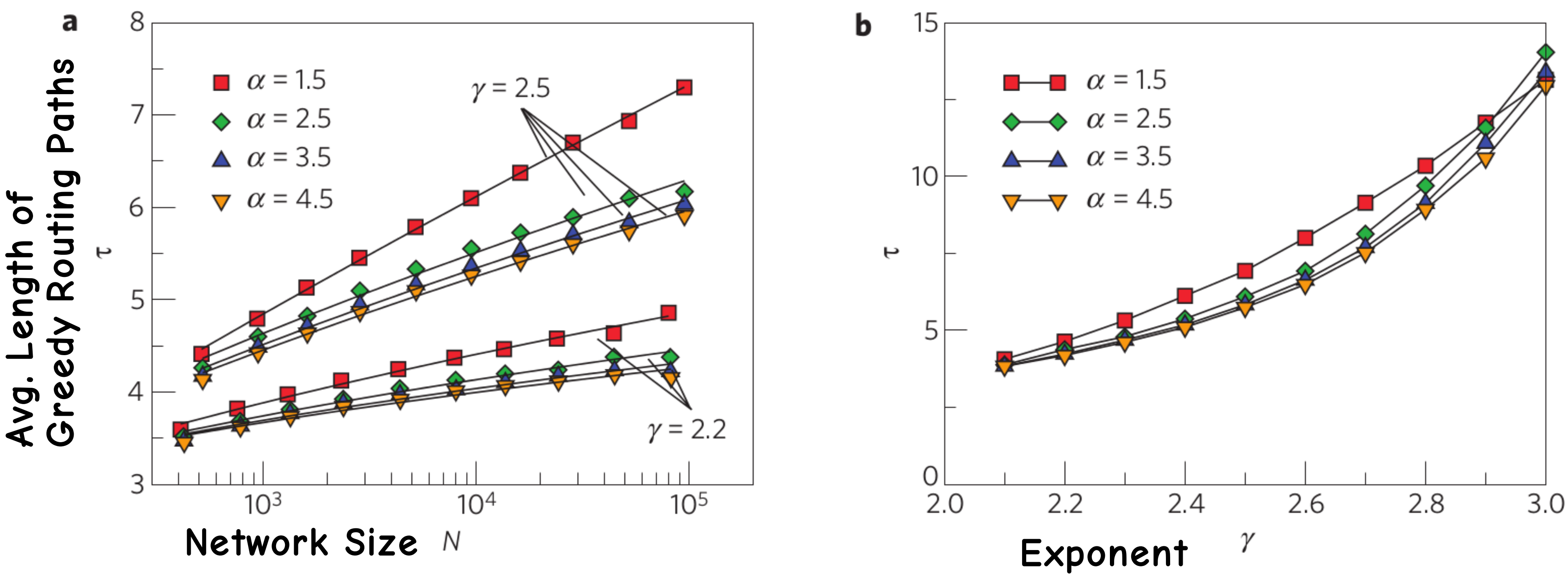
Hubs will be connected with a high probability because of large $\mathrm{d}_c$

Low degree nodes connected only if (hidden distance) d is small

Hubs connected to low degree nodes at moderate hidden distance

$\alpha$ importance of hidden distance

# Path Length
## (Greedy Routing)




Path length grows polylogarithmically with the network size

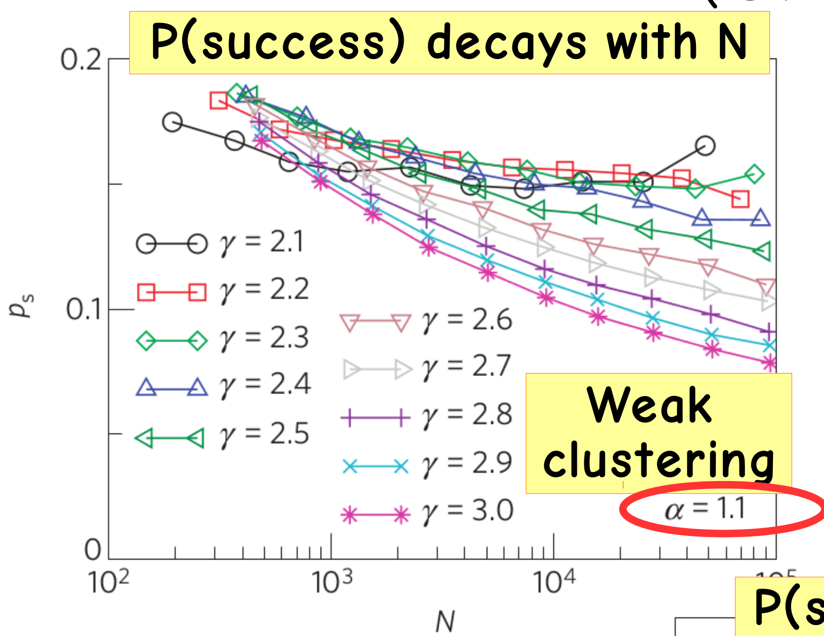
Paths shorter for smaller exponents and stronger clustering

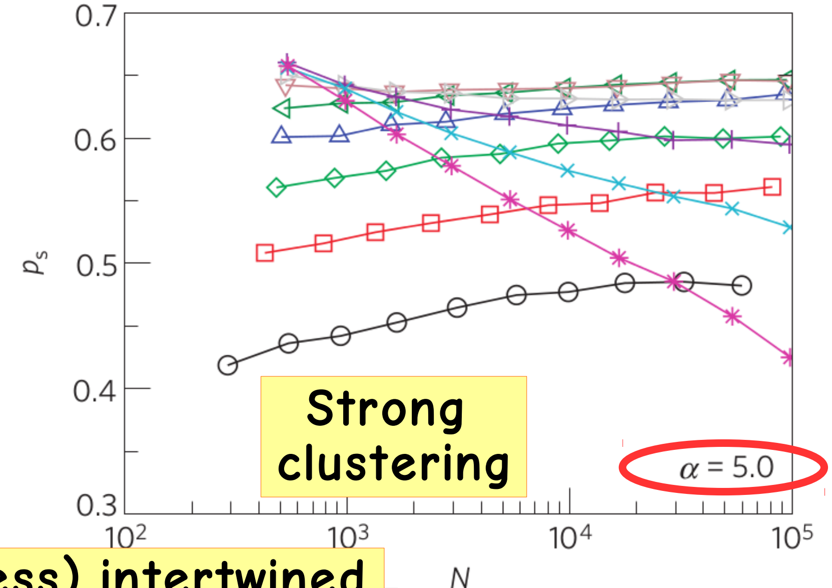
Boguna, Marian et al. **"Navigability of complex networks."** Nature Physics 5, no. 1 (2009): 74-80.

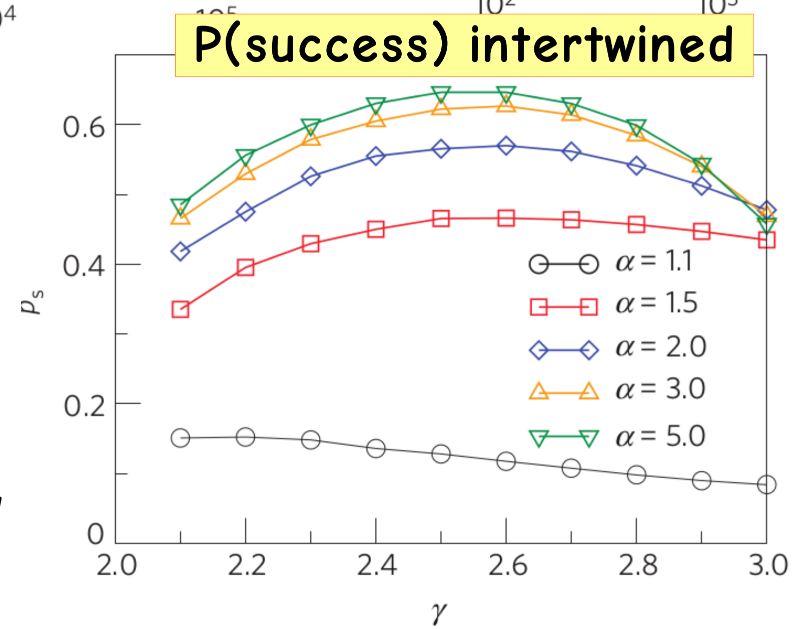
# Greedy Routing

- Hidden Space as the coordinate space
  - *Hidden space is circle in this example*
- Greedy Routing: Send to neighbor who is closer to the destination (in hidden space)
- Unsuccessful Paths: None of your neighbors are closer to the destination in the hidden space

# Success Probability
## (Greedy Routing)

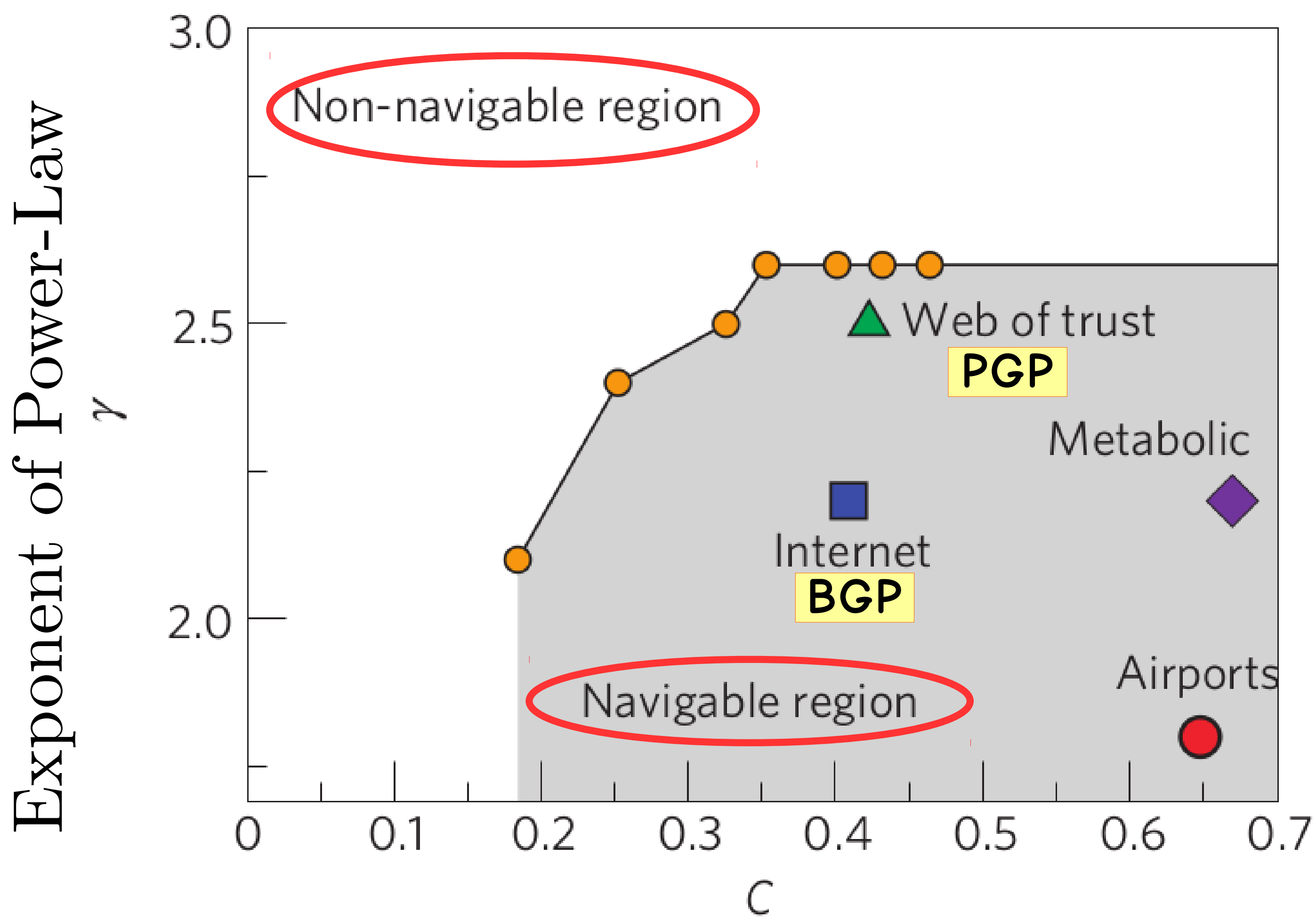



Boguna, Marian et al.
**"Navigability of complex networks."**
Nature Physics 5, no. 1 (2009): 74-80.

# Navigation in Scale Free Networks

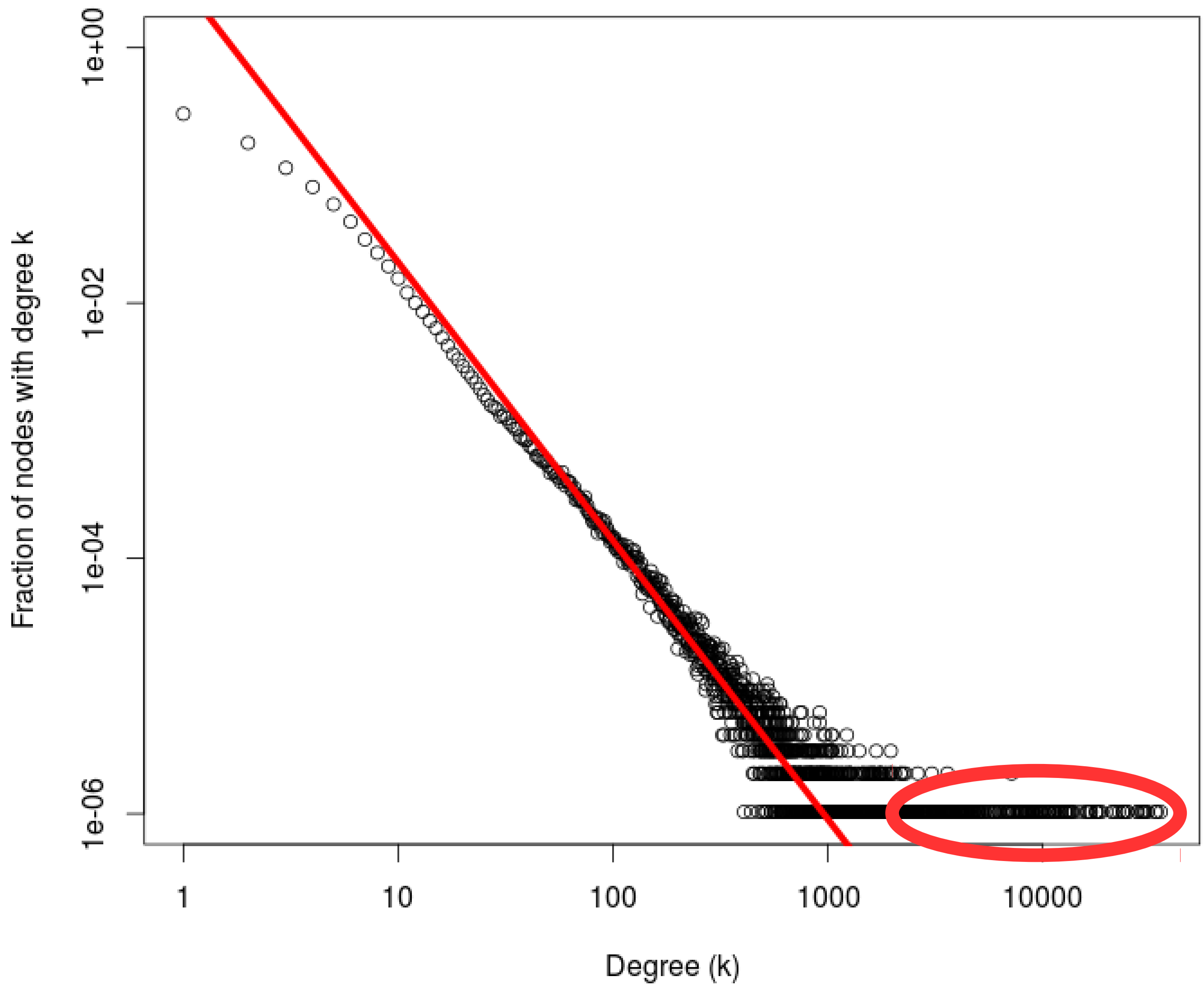



Clustering Coefficient $C = f(\gamma, \alpha)$

# Implications of Result

- Internet Routing
  - Routers currently exchange signals to keep cohenrent view of network
  - Network size increasing with time
  - Hidden metric space eliminates the need for control signals exchanged to notify changes in network
- How to proceed to discover the hidden metric space
- Does Shortest Path imply Shortest Time to destination?
  - What happens in case of congestion at hubs?

# Scale-Free Model for AS-Graph




AS Topology of skitter dataset parsed by SNAP team
http://snap.stanford.edu/data/as-skitter.html

# Is the Scale-Free Internet A Myth?

- What we have seen till now wrt to Preferential Attachment
  - Preferential attachment results in Hubs
  - Hubs vulnerable to coordinated attacks
  - Why is the Internet still up and running
- Is the Scale-Free modeling paradigm consistent with the engineered nature of the Internet and the design constraints imposed by existing technology?
  - Is the simplistic toy model too generic?
  - Do the available measurements, their analysis, and their modeling efforts support the claims made by "Error and Attack Tolerance" paper?

# Importance of Measurements

- Tool for measurement study for AS-measurements
  - Traceroute
- Biases of traceroute
  - Uses IPv4 Protocol
    - What about non-IPv4 protocols like MPLS?
    - Entry points to non-IPv4 regions can aggregate to Hubs
  - Only reports the interfaces traversed by the packet
    - Routers can have multiple interfaces and appear on different routes with different IP addresses

# Leverage Domain Knowledge

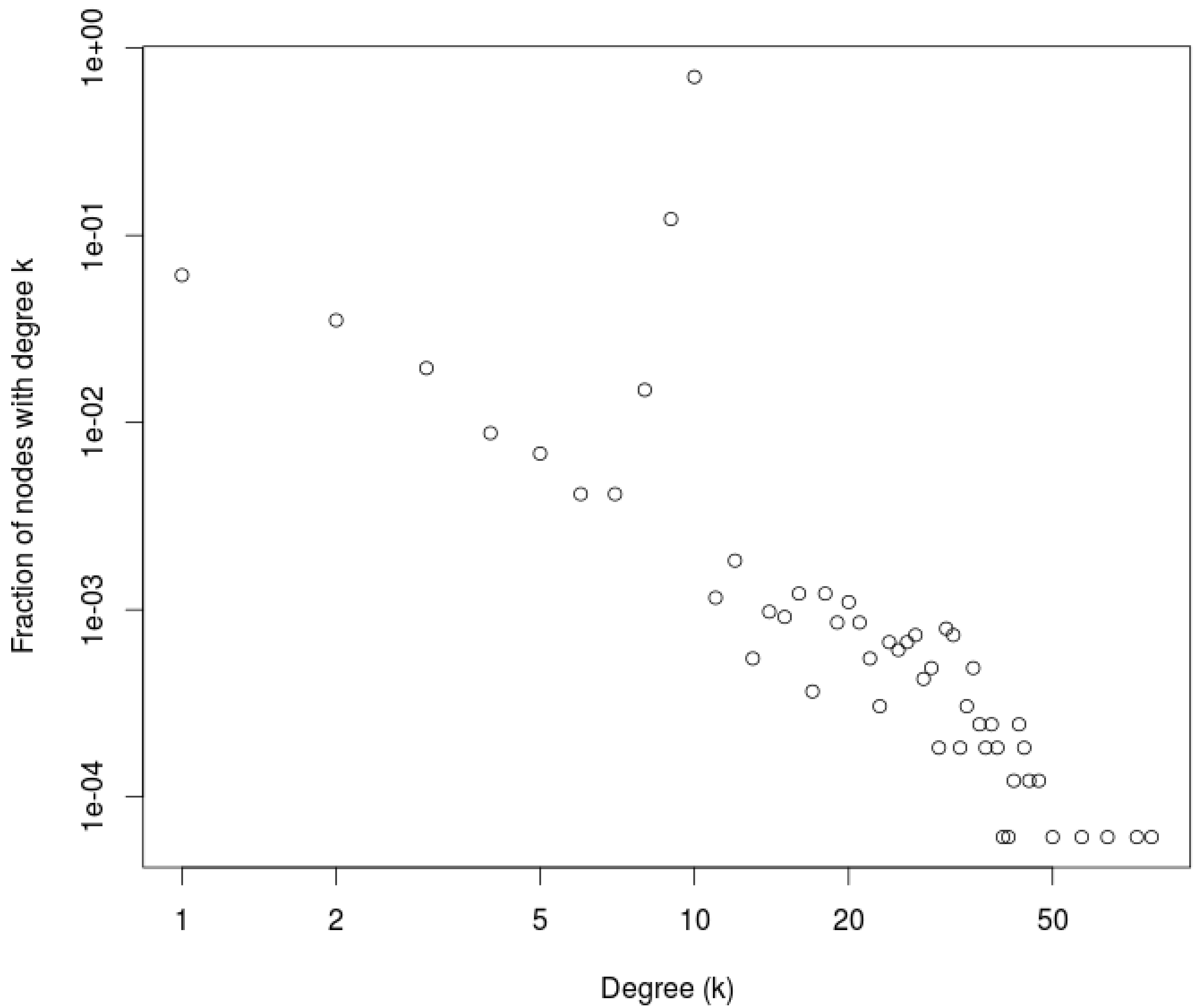
- Device Constraints
  - Finite number of interfaces on routers
  - Finite capacity of routers
- Placement of High Degree Nodes
  - Edge vs Core

**What about Overlay Networks?**

How would you deploy the network if you are a network engineer?

- Leverage domain knowledge to identify driving forces behind the design of high engineered systems such as the Internet

# Scale Free for Gnutella?




Gnutella August 2002 dataset parsed by SNAP team
http://snap.stanford.edu/data/p2p-Gnutella31.html

# Recap of Modeling Overlay Networks

- Milgram's Experiment
- Duncan Watts Random Rewiring Model
- Scale-Free Networks
  - Preferential attachment
  - Evolving Copying Model (Copying Generative Model)
  - Scale-Free with Strong Clustering
- Error and Fault Tolerance of Complex Networks
- Navigation (Greedy Routing)
  - In Small World (Kleinberg's Small World)
  - In Complex Networks (Scale-Free with Strong Clustering)
- Mathematics and the Internet: A Source of Enormous Confusion and Great Potential

# Commonly used metrics

- Clustering Coefficient
- Diameter
- Degree Distribution

# Methodology

1) Make observations (conduct measurement studies)
2) Build model to explain observations
   - Choose the right level of granularity (zoom level)
   - Strip the problem to a simple form
   - Attempt to formulate the problem and model the system
3) Validate model
   - Reproduce observations/measurements
   - Explain observations
4) Revisit step 2 (and 1) to improve understanding

# Important Articles


- Milgram, Stanley. "**The small world problem.**" Psychology today 2.1 (1967): 60-67
- Watts, Duncan and Strogatz, Steven. "**Collective dynamics of 'small-world' networks.**" Nature 393.6684 (1998): 440-442.
- Barabási, Albert-László, and Albert, Réka. "**Emergence of scaling in random networks.**" Science 286, no. 5439 (1999): 509-512.
- Kleinberg, Jon. "**The small-world phenomenon: An algorithmic perspective.**" In ACM Symposium on Theory of computing, pp. 163-170. 2000.
- Ravi Kumar et al. "**Stochastic models for the web graph.**" In Annual Symposium on Foundations of Computer Science, 2000.
- Albert, Réka, and Barabási, Albert-László. "**Statistical mechanics of complex networks.**" Reviews of modern physics 74.1 (2002): 47.
- Newman, Mark. **"The structure and function of complex networks."** SIAM review 45, no. 2 (2003): 167-256.
- Mitzenmacher, M. (2004). "**A brief history of generative models for power law and lognormal distributions**." Internet mathematics, 1(2), 226-251.
- Mark Newman. **"Power laws, Pareto distributions and Zipf's law."** Contemporary physics 46, no. 5 (2005): 323-351
- Jure Leskovec et al. "**Graphs over time: densification laws, shrinking diameters and possible explanations.**" In ACM SIGKDD, pp. 177-187. 2005.
- Boguna, Marian et al. **"Navigability of complex networks."** Nature Physics 5, no. 1 (2009): 74-80.
- W Willinger et al. "**Mathematics and the internet: A source of enormous confusion and great potential.**" In Notices of the AMS. 2009.

# Sources for these slides

- Sasu Tarkoma ”**Overlay and P2P Networks**”, 2015
- Datasets from Stanford Network Analysis Project (SNAP)