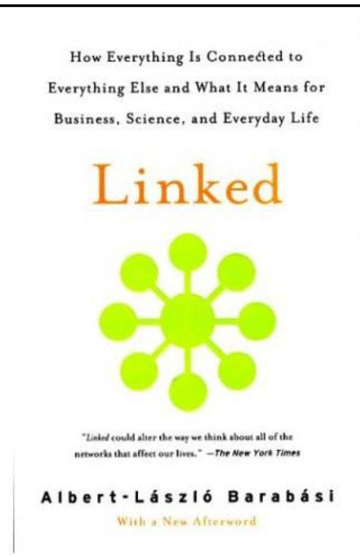
Scale Free Networks

Julita Vassileva
Social Computing Class 2009

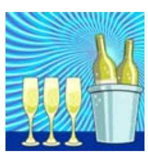
Based on

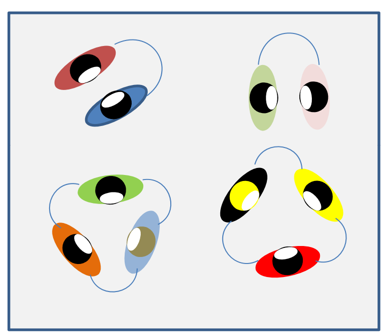


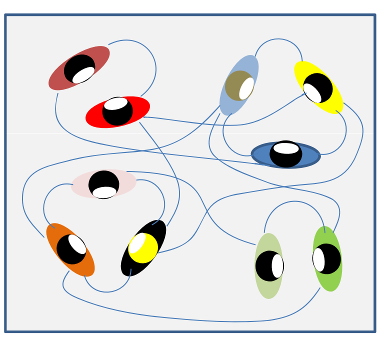
- Available on Amazon.ca, for $16.00 CDN

The cocktail party puzzle

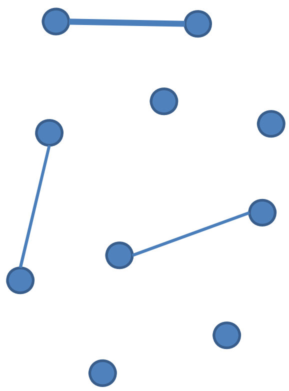
- You are hosting a party with 100 guests.
- You have one bottle of really good and very expensive wine (the bottle without a label).
- You tell one of the guests about this and ask him not to tell anyone, but only the new acquaintances she made at the party about it.
- Assuming that everyone who learns about the precious bottle follows this rule, what is the chance that there is still some wine left by the end of the party?







Random graphs - **Paul Erdős**  **and**  **Alfred Renyi**  **1959**

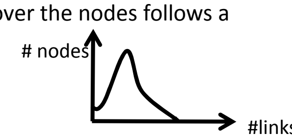


In a random graph some nodes can be lucky and have many links, some may have none, or few.

But when there are enough connections, a giant cluster emerges and a path can be found between every two nodes. Having 1 connection is enough to be part of the cluster.

In such a connected graph most nodes will have the approximately same number of links – “egalitarian” .

The number of links over the nodes follows a Poisson distribution



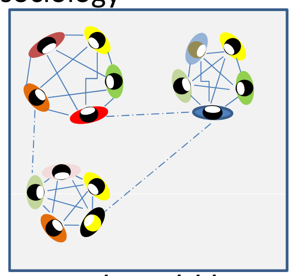
Questions:
- Is the Web a random graph?
- Is society a random graph?

Six degrees of separation

- The Stanley Milgram experiment, 1967
  - find the average “distance” between any two people in the US
  - 42 of the 160 letter made it to the destination
  - Average path length: 5.5
- Experiment repeated with email by Watts and Strogatz
  - resulting in average path length about 6.8
- So human networks (social networks) form a **small world**…
- How about the Web?
  - Analysis of 800,000,000 nodes (documents) found that the average path length is 18.59…. Still a small world.
- Huge networks, with billions of nodes collapse – most of the average path lengths are between 2 and 14! Why ??
- Does this knowledge make it easy to find what you look for?
  - No, we need cues to search, some extra information

Small worlds

- Mark Granovetter, 1960 dissertation:
  - How do people find jobs?
  - Started the study of **social networks** in sociology
  - The "strength of weak ties"
- Watts and Strogatz, 1998
  - Clustering coefficient: how many links exist out of all possible links
  - Random NW have low clustering
  - But real social NW have high clustering



- A small world can be obtained from clustered world by adding links connecting random nodes (weak ties)

Hubs and connectors

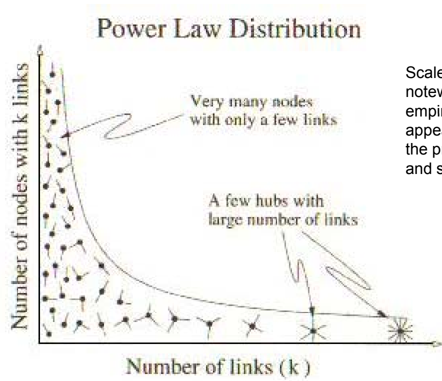
- Examples:
  - How many actors have played in a movie with Kevin Bacon? (1,800) His average separation from any Hollywood actor is 2.79
  - What is the number of people you know?
  - Number of links pointing to websites?
  - Number of daily visits of websites?
  - Number of reactions done by different molecules in a cell?
  - Number of phone calls placed by telephones in the network?
- Great inequality in distributions -- there are connectors (hubs)
- But the random universe does not support connectors.

If the Web was a random network, the probability of a site with 500 incoming links would be 10^-99, practically 0. But there were 400 such pages in a study of 20% of the web in 2000. There was also a site with 2 million incoming links.
Both the model of Erdos & Renyi and the Watts & Strogatz model can not explain this.

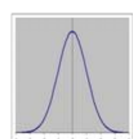
The 80/20 rule

- Vilfredo Pareto, 1910
  - 80% of the peas are produced by 20% of the plants
  - 80% of the wealth in Italy was owned by 20% of the population
  - 80% of profits are generated by 20% of the employees (Murphy's law)
  - 80% of the complaints are generated by 20% of the customers
  - 80% of the decisions are made in 20% of the meeting time…
  - 80% of the links point to 15% of the websites
  - 80% of the citations go to 38% of the scientists…
  - 4/5 of our efforts are largely irrelevant
  - → POWER LAWS

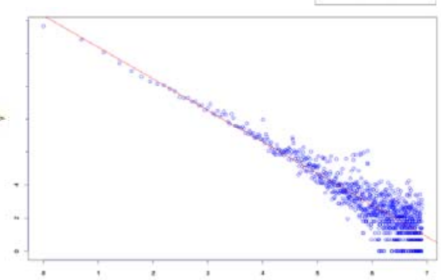
A **scale-free network** is a network whose degree distribution follows a power law. That is, the fraction $P(k)$ of nodes in the network having k connections to other nodes goes for large values of k as $P(k) \sim k^{-\gamma}$ where γ is a constant whose value is typically in the range $2<\gamma<3$, although occasionally it may lie outside these bounds.



Compare with Normal Distribution (Bell-curve)



Scale-free networks are noteworthy because many empirically observed networks appear to be scale-free, including the protein networks, citation networks, and some social networks.

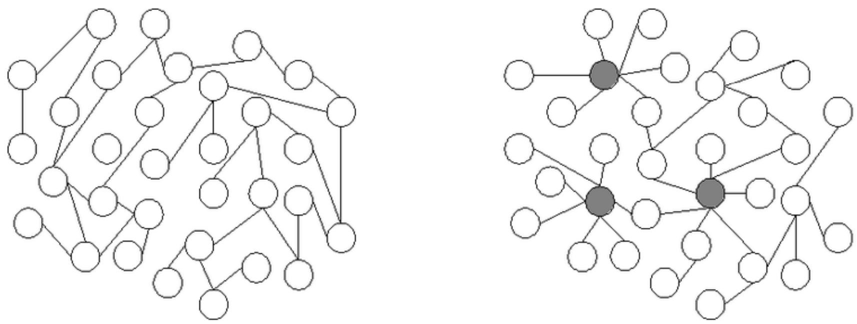


N(k) - # pages with K incoming links

$N(k) \sim k^{-\gamma}$, where γ – degree exponent, in this case $\gamma = 2.5$

From: **www2002.org/CDROM/poster/164/**

Comparing a random and power-law network



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

Power-laws signal self-organization

- Power law networks are far from random
- In physics, power-laws indicate phase-transitions, order emerging from disorder:
  - Water molecules arranging themselves in a crystal lattice (freezing)
  - Ferromagnets becoming magnets at a critical temp.
  - Near the critical point, elements of order and disorder mix; several critical quantities follow power laws
  - In physics there are theories that explain this (Kenneth Wilson, 1971, awarded Nobel Prize in 1982)
  - But what is the mechanism behind self-organization in complex networks?

The rich get richer

- **Albert-László Barabási**, **Réka Albert,** 1999
- How do networks grow?
  - In a random network, new nodes connect to randomly picked existing nodes
  - But in reality, new nodes are more likely to connect to nodes which already have a lot of connections
  - **Preferential attachment**
  - Feedback loops → power laws
- **Growth** and **preferential attachment** explain the hubs and power laws in complex networks.
- But this model predicts that the oldest nodes would have most links
- Newcomers will have no chance.

How to explain successful newcomers?

- For example, Google?
- All nodes are not equal, some are "better", more attractive, offer a better quality of service, or are better at making friends, or have better looks (younger actors/actresses ☺)
- **Fitness** of a node in a competitive environment
- The "Fit get rich" model (borrowing formalisms from quantum mechanics) predicts a phenomenon called Einstein-Bose condensation
- In some networks (under special conditions) all links will ultimately point to one node: "The winner takes it all"

Google or Microsoft ?

Robust Scale Free Networks

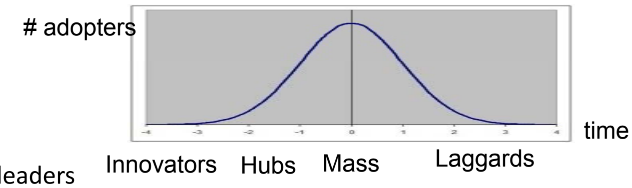
- Scale-free networks are extremely robust in case of random failures
- Studying network resilience
  - In random networks, some node failures can easily break a network into isolated, non-communicating frameworks – e.g. closing simultaneously highways in and out of Jacksonville and Lake City, FL, will close Florida to the rest of the US.
  - Yet, a study of the Internet resilience showed that we can remove 80% of all nodes, and the remaining 20% will still remain connected
  - The key to this is the presence of hubs, removing nodes randomly is not likely to affect them, and they hold the NW

Vulnerable Scale Free Networks

- Yet, scale-free NW are very vulnerable to **targeted attacks** and to **cascading failures**
- In case of targeted attack on a critical number of hubs, the network disintegrates very quickly
- Cascading failures – examples
  - Power grid black outs (1996, 2003)
  - Cascades of malfunctioning routers on the Internet
  - Cascading East Asian economic crisis in 1997
  - Cascades in ecological habitats

Spreading Viruses and Innovation

- Viruses
- Innovation



  - Hubs:
    - Opinion leaders
    - Power users
    - Influencers
  - Are not necessarily innovators, but they are key to spreading an innovation, launching an idea….
  - Yet, not all innovations catch on (e.g. Apple's Newton). Why some do and some do not?
  - Diffusion models

Disease diffusion models

- Threshold model: Each innovation has
  - **spreading rate** – the likelihood that it will be adopted by a person introduced to it, and
  - **critical threshold** – defined by the properties of the NW in which the information spreads
  - If **spreading rate** < **critical threshold**, it will die,

    Else, the number of people adopting the innovation will increase exponentially.

This model has been used by epidemiologists, marketers, sociologists, political scientists. However, it doesn't explain the persistence of some viruses like AIDS. Why?

Because it assumes a random network.

Swedish study of sexual contacts

AIDS rising…

Region	Adults & children living with HIV/AIDS	Adults & children newly infected	Adult prevalence*	Deaths of adults & children
Sub-Saharan Africa	22.0 million	1.9 million	5.0%	1.5 million
North Africa & Middle East	380,000	40,000	0.3%	27,000
Asia	5 million	380,000	0.3%	380,000
Oceania	74,000	13,000	0.4%	1,000
Latin America	1.7 million	140,000	0.5%	63,000
Caribbean	230,000	20,000	1.1%	14,000
Eastern Europe & Central Asia	1.5 million	110,000	0.8%	58,000
North America, Western & Central Europe	2.0 million	81,000	0.4%	31,000
Global Total	33.0 million	2.7 million	0.8%	2.0 million

Possible solution?

- …?
- Treat the Hubs!
- But
  - How to identify them?
  - Is this ethical?

Run-away internet

- Started as decentralized network (military communication tool)
- Escaped and grew into a large scale-free network that can be studied but hardly managed or influenced…
  - Growth and preferential attachment (based on bandwidth available, distance dependence, underlying fractal structure of router distribution…)
  - Vulnerable to simple errors, coordinated attacks, can be exploited for parasitic computing
  - With its own "senses" – cameras, sensors, processors

Some food for thought…

"While entirely of human design, the Internet now lives a life of its own. It has all the characteristics of a complex evolving system, making it more similar to a cell than a computer chip. Many diverse components, developed separately, contribute to the functioning of a system that is far more than the sum of its parts. Therefore Internet researchers are increasingly morphing from designers into explorers. They are like biologists or ecologists who are faced with an incredibly complex system that, for all practical purposes, exists independently of them." (pp.149-150)

Albert-László Barabási, Linked, Plume Publ. 2003.