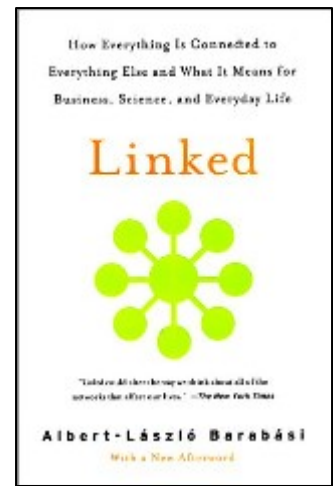


A review of Linked: The New Science of Networks
By Albert-Laszlo Barabasi¹
(summarized by Tom Paper)



Chapter 1 - Introduction

- Mafiaboy shuts down Yahoo
- Paul Spreads the gospel of Jesus
- Mafia boy went only to computers that let him in and that had access to Yahoo
- Paul went to larger cities instead of smaller cities
- Maps. Reductionism. Networks.
- The world has been taken apart and analyzed in all of its many parts but we haven't yet figured out how all the parts fit together. That is the point of Linked.

Chapter 2 – The Random Universe

- Euler, a mathematician, studied the Königsberg bridge puzzle. Turns out that the puzzle is really all about graph theory, the idea of nodes and links. The puzzle is: how can you cross all the bridges (links) and come back to the place where you started without crossing any bridge twice?
- Erdos & Renyi
 - Vintage port at a party. Random connections. At what point is there a transformation?
 - Percolation? All their math/graph theories were based on randomness, that the chances of any node having a link was a random occurrence. This means that a plot of the number of links per node would follow a Poisson distribution, like a bell curve with a steep peak and a narrow standard deviation.

Chapter 3 – Six Degrees of Separation

- Igy irtok ti: Written by Karinthy in 1929, extremely popular book. He later wrote a short story in which a character stated that he could connect to anyone through just five acquaintances.
- Stanley Milgram's "distance" test. Letters were sent to random people in Wichita and Omaha. The letters asked the recipients to send the original letter on to someone that they thought would know one of a small group of people living in

¹ Human Nature Review 2002 Volume 2: 312-316 (19 August)

<http://human-nature.com/nibbs/02/linked.html>

Linked: Barabasi ratifies Kauffman, demotes the social gene

By James Brody, A review of Linked: The New Science of Networks, By Albert-Laszlo Barabasi

NY: Perseus, 2002, 229 pp + 50 pp of notes & index.*

- Cambridge that Milgram knew. The median number of intermediaries between the random original recipients and the people in Cambridge was 5.5.
- 19 degrees of separation on the web – the average number of clicks from any page to any other. 800 million nodes exist on the web. While 19 degrees may seem like a high number for degrees of separation, it is small when you think about how big the web is. Why? Think about the power of doubling. See also logarithms.
 - Milgram’s subjects never knew what the shortest path actually was to their target. But a shorter path may have existed. A “social search engine” does not exist at this time.
 - Barabasi posits that degrees separating people on the planet has shrunk over time. The internet means that human networks are scale-free. Travel used to be harder.
 - When you discover a common acquaintance with a perfect stranger on the side of our planet, then you know that some other geometry is at work.

Chapter 4 – Small Worlds

- Granovetter – the strength of weak ties. He did a study that proved that you’re more likely to get a job from a weak connection than a close friend.
- Synchronization – unison clapping, fireflies, pacemaker cells, women’s menstrual cycles when they live together, cricket chirping.
- Duncan Watts, mathematician & sociologist, and his advisor, Steve Strogatz, developed the idea of a clustering coefficient for individuals in a friendship group = (# of actual links between friends / # of potential links).
- Erdos number. Erdos was a great and famous mathematician. Most scientific papers are co-published, which in networking theory implies a link. Scientists in the community would talk in terms of their “Erdos number,” which mean the number of links between the scientist in question and the great Erdos. The focus on Erdos numbers proves that scientists don’t pick their colleagues randomly.
- C. Elegans is a worm with 302 neurons. Watts and Strogatz proved that even with the neurons of a worm, clustering exists (> random).
- Small worlds vs. clustered worlds. Watts and Strogatz proved that a few random lines across a linked circle greatly reduces the average degrees of separation (which means the small world phenomena) but leaves the clustering coefficient unchanged.

Chapter 5 – Hubs and Connectors

- Malcolm Gladwell, author of The Tipping Point, writes about connectors, people with a knack of making friends and connections. He also wrote about mavens, people knowledgeable on a variety of subjects.
- Web pages do not exist with a random number of links. Certain pages are hubs. Amazon and Yahoo.
- Kevin Bacon game. Is he a hub in Hollywood. Actually, no. He’s 876th on the list of most connected. Rod Steiger is at the top, 2.53 links from everybody else. Being in the most movies does not mean being the most connected because porn stars like Ron Jeremy play in many movies.

- Hubs are important on the web, in Hollywood, in cells, in ecosystems and even in cancer.

Chapter 6 – The 80/20 Rule

- Vilfredo Pareto believed that laws of economics existed. He demonstrated that 80% of peas were produced by 20% of the pods, 80% of Italy's wealth was held by 20% of the population. Systems that follow Pareto's law can be explained by a power law, as opposed to a bell curve.
- US roadmap vs. airline routing map (p71). Random networks have scale because they have an average and even distribution around it. Networks with a few highly connected hubs are scale free and characterized by power laws.
- Phase transitions – when water turns to ice, when order turns to disorder or vice versa. Power laws are at work every time order spontaneously appears in complex systems. “Nature normally hates power laws.” (Maybe phase transitions are like perturbation.) Power laws are how complex systems self-organize.

Chapter 7 – The Rich Get Richer

- Two assumptions to the old models had to be abandoned: 1) The # of nodes is fixed; 2) Nodes randomly choose which other nodes to link to.
- Every network is either growing or shrinking. Static networks do not exist.
- People spend money advertising their websites because they know that viewers do not link randomly to websites.
- Preferential attachment. We tend to link to things that are already more connected. The result is that the more connected nodes grab a disproportionate share of new links.

Chapter 8 – Einstein's legacy

- Google overtook Inktomi; Boeing overtook DeHavilland.
- “New kid on the block effect” = fitness = the competitive ability of nodes to attract links (but close competitors will usually still exist)
- Bose-Einstein condensate – new material forms at low temperatures.
- Two types of scale-free networks:
 - 1. “Winner-take-all-networks” Microsoft in operating systems
 - 2. “Fit-get-rich-networks” PC's / Search engines.

Chapter 9 – Achilles' Heel

- Electric power grid is a man-made structure AND it is vulnerable due to its lack of interconnectedness.
- 1) Robustness = the ability of systems to survive internal failures
 - Scale-free networks tend to be very robust from random failure
- 2) Vulnerability – targeted attacks make scale-free networks vulnerable

- Cascades
 - Sea otter – it's comeback led to cascading vibrancy on the California West coast
 - Cascading failures – one failure leads to another – Watts showed that failures often go unnoticed for long periods of time before there is a landslide failure
- 3) Topology
 - The types of hubs and nodes in networks matters

Chapter 10 – Viruses and Fads

- Gaetan Dugas (gay, spread AIDS) & Mike Collins (cartoonist) are both examples of diffusion in a complex network.
- Hybrid corn, 1943 – 4 phases of adoption
 - 1) early adopters
 - 2) early majority
 - 3) late majority
 - 4) laggards
- Product introductions:
 - 1st: Innovators
 - 2nd: Hubs
 - 3rd: Average person
 - The hubs are the gatekeepers. If the hubs resist a product, it is doomed.
- Threshold diffusion model.
 - Spreading rate
 - Critical threshold
- Love bug – It was a deadly computer virus, but it stuck around.
 - Why? In scale-free networks, there is no epidemic threshold!
 - Why? Because of uneven topology, the presence of hubs
- Sexual network topology > is a power curve!
 - Wilt Chamberlain had 20,000 partners
- AIDS solution: treat the hubs. Research has shown that epidemics can be stopped when you treat the hubs.

Chapter 11 – The Awakening Internet

- Centralized / Decentralized / Distributed – Paul Baran proposed a distributed internet in 1964.
- The internet is routers communicating with each other through protocols. The internet is not fully mapped. www.peacockmaps.com We know that it follows a power curve.
- Internet model contains:
 - 1) Growth
 - 2) Preferential attachment
 - 3) Distance dependence
 - 4) Fractal structure

- Internet failures have followed cascading failures
- Parasitic computing – one computer forcing another to do computing
- The internet is like our skin. By 2010, there will be 10,000 telemetric devices for every human. When will the internet become self-aware?

Chapter 12 – The Fragmented Web

- The web is fragmented into continents and communities. The robots are the search engines mapping the territory.
- Size of the web:
 - Pre 1998: Common wisdom is that the size is known
 - Post 1998: Size is 320 million documents and that was 50% bigger than expected
 - 1999: 800 million documents and search engines only cover 40%
 - (Search engines don't try to find all the pages, just the best pages)
 - 2003: Only 24% of pages are reachable.
- Directed vs. non-directed networks
 - Directed means one-way; non-directed means you can go in any direction
 - Random and scale-free apply to both directed and non-directed networks
- The four continents of the web:
 - Core
 - In
 - Out
 - Islands
- Communities – what are they on the web? Hard to define, but I know it when I see it. A tendency to become isolated exists on the web.
 - Some communities have more sites and are better connected.
 - Pro-choice vs. pro-life
 - Interconnectedness can be managed.
- The web as a tool of control.
 - France tried to ban sales of Nazi memorabilia to French consumers.
 - Lawrence Lessig – Stanford Law professor – the web could become the perfect tool of control.
 - Code vs. Architecture – Barabasi's point is that code is a perfect tool of control, but as long as architecture remains free, then web will remain free.
 - Laws of large-scale topology.
- Brewster Kahle – founder of a search engine, he now runs Internet Archives in the Presidio of San Francisco. They are attempting to record /map what the web has looked like over time. In other words, if you want to understand what was on various websites right after the 9/11 attacks on America, you would access his archives. The point is that the web is fragmented and changing over time. Seeing it in a historical perspective offers tremendous insight and IT IS POSSIBLE.

Chapter 13 – The Map of Life

- Manic depression is viewed as a genetic disorder. However, no single gene is responsible, despite numerous attempts to prove it. The conclusion is that manic depression is caused by the network relationships of various genes.
- Mapping of the human genome. Involved 3 billion chemicals. While a terrific accomplishment, human behavior is not dependent upon components. Human behaviors depend upon the network relationships (systems) of the molecular components within humans.
- Cells are actually scale-free webs. A few molecules within each cell are involved in the majority of reactions. Genetic disorders result from the malfunctions of the nodes of the regulatory network.
- Cancer. In the late 1980's, the P53 gene was recognized as the "tumor suppressor." Healthy cells had some P53, but 50% of cancerous cells have a mutated P53 gene. Solving cancer was not as simple as making drugs that make sure that the P53 gene always does its job. The P53 gene is a hub in a proper functioning network of molecules and genes.
- The future of drugs involves understanding the cellular/molecular interactions within a cell. Mapping these relationships will take another decade. We will also need to monitor malfunctions in genes and proteins, but this is available today.
- Complexity. Humans are not that much more complex than a worm in terms of the number of genes (30,000 in humans vs. 20,000 in worms). However, we are vastly different when you factor in the greater level of complexity caused by the different # of genes.
 - Folding paper: Helps explain complexity. This is like the effect of folding a piece of paper over on itself.
 - 1 fold = 0.01 inches thick
 - 7 folds = 1 inch thick
 - 11 folds = 1 foot thick
 - 21 folds = 3 football fields thick
 - 25 folds = 3 miles thick
 - 38 folds = earth's circumference (25,000 miles)
 - 50 folds = distance earth to sun (93,000,000 miles)

Chapter 14 – Network Economy

- Traditional corporate organizations – based on a tree, with CEO at bottom and drone executives at the outer branches.
 - Problems:
 - 1) Information must be carefully filtered as it rises through the organization.
 - 2) Organizations become rigid with this organization
 - Best suited for mass production
- Need for flexibility is pushing organizations to a web/network organization: flat with cross-links between nodes.
- Board of Directors of Fortune 1000 companies
 - 10,100 directorships and 7,682 directors
 - 79% of the 7,682 directors sat on **one** board

- 14% of the 7,682 directors sat on **two** boards
- 7% of the 7,682 directors sat on **three or more** boards
- Vernon Jordon sat on 10 boards, the most of any director
 - He regularly meets with 106 other board members
 - His path to these boards was one step at a time, referred along the way by his co-directors
- The businesses within almost every industry form a vital network
 - Partnering (creating links between competitors) is a means of survival in many industries
 - Some countries and particular industries have made cooperation a way of life, between competitors and also between suppliers and subcontractors.
- Traditional economics would say that companies / competitors / customers / suppliers do not interact with each other, but that they mythically exist in their “market”
- The reality is that a market is nothing more than a directed network.
 - “...In markets the standard strategy is to drive the hardest possible bargain on the immediate exchange. In networks, the preferred option is often creating the indebtedness and reliance over the long haul.”²
 - Most of the time, when a business (node) fails, only the partners of node are damaged.
 - Sometimes, however, one failed node begets another failed node, which becomes a cascading failure.
- 1997 - small Thai company fails, leads to other failure and ultimately leads to a currency collapse.
- Outsourcing is the latest trend in the computer business. Compaq and Cisco have been major outsourcers.
 - Compaq got hurt because it’s overseas source couldn’t keep up with demand of their Pocket PC.
 - Cisco got hurt because it didn’t quickly shut off supply from its overseas supplier when its demand went away.
 - Point is that all companies, globally, are more and more linked today. A company that doesn’t consider its suppliers and partners as links in a network risks their long-term successful performance.
- Hotmail – Sabeer Bhatia – started hotmail in July , 1996. First major success of VIRAL MARKETING. Put a tag at the end of each e-mail, telling e-mail recipients that hotmail was private and free.
 - High spreading rate because:
 - 1) it is free (free!)
 - 2) it is easy to sign up (low learning path)
 - 3) every e-mail sent is an advertisement (rapid reach)
 - Attempts to replicate Hotmail’s success have not been very successful. Plenty of things are free and easy, but finding something that has a reason to have such a rapid reach is a challenge. EpidemicMarketing.com spent millions but didn’t make it.

² Walter W. Powell, *Neither Market nor Hierarchy: Network Forms of Organization*

- “The real legacy of the Internet is not the birth of thousands of new online companies but the transformation of existing businesses.”

Chapter 15 – Web Without A Spider



- - Real networks do not have a central organizing node, controlling and monitoring every link and node, in the same way that a spider manages every part of its web.
 - Real networks means:
 - No single node’s removal can destroy a network
 - There is no meticulous design of real networks
 - They are self-organized
 - Scale-free networks are seen over and over in situations, indicating common laws governing the emergence of complex networks
 - Al-Queda and the 9/11 attack are examples of a web without a spider. Taking out any one of the terrorists would not have stopped the attack. Valdis Krebs, a management consultant who normally uses network theory to analyze corporate communications, put together a map of the 19 hijackers, demonstrating that no one person was vital to the attack.

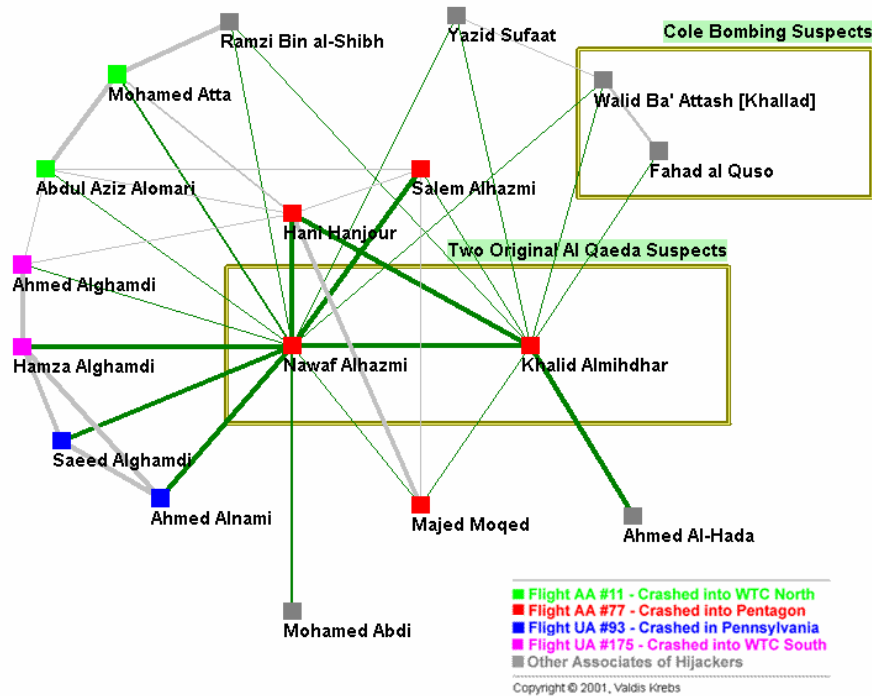


Figure 2 - All nodes within 1 step [direct link] of original suspects

-
- The strength of terrorism is that it operates without one node that is a hub.
- Assassination is most effective when a country depends upon a particular person for its well-being.
- Having multiple leaders is the best defense against terrorism.

Networking concepts:

- Terms
 - Nodes – any location on a network
 - Links – connections between two nodes
 - Hubs – a network location through which many connections pass
 - Scale – the importance (or “weight”) of any node
 - Scale-free = a network where there is great difference in the relative importance of various nodes
 - Hub & spoke airline system
 - Electric power grid
 - Companies and consumers linked by trade
 - Toyota and network of supporting companies (disruption when sole supplier of P-valves had a fire and other companies in the network, with no experience in this part, were in production of this part within 1 week).

- Cocktail party
 - Nervous system of living creatures
- Highways vs. airline systems
 - Highway system pre-Interstate
 - Each city (node) connected at random to other cities
 - A weather disruption in any one city would not affect ability to get to other cities
 - Airlines hub & spoke system
 - Certain cities are hubs (also nodes) and others are simply destinations (nodes)
 - A weather disruption that knocks out the service in a hub city causes far more damage than knocking out a particular node destination city
- Positives of networks
 - Good ideas can win acceptance more quickly
 - Hotmail – started 7/4/96 – after 1 year had 1 million users
 - After 2 years, when Microsoft purchased, 10 million users
- Negatives of networks
 - Achilles heel – knocking out a hub can have disastrous consequences
 - Cisco – pre-bubble bursting, thought it was impervious to overinvestment in inventory – but the bubble came and it had to write off billions of inventory – why? Cisco had a very self-centered view of its business.
 - A lack of cooperative spirit in a network leads to worse problems when a hub is knocked out – Toyota suppliers that responded so quickly were long-familiar with a cooperative environment
- Advanced concepts
 - Frequency of links for nodes on network
 - Normal curve in networks with scale – random networks
 - Power curve in scale-free networks
 - Power laws explain emergent networks like living cells and friendships
 - A few nodes (about 20%) have many links but most nodes (about 80%) have only 1 or 2 links
 - Hubs in scale-free networks make “...it possible to add massive numbers of nodes to the network with little effect on the number of connections needed to get from any one node to any other.”³
 - Google is a hub on the internet
 - Wilt Chamberlain was a hub
 - Fitness
 - Number of links connected to a node

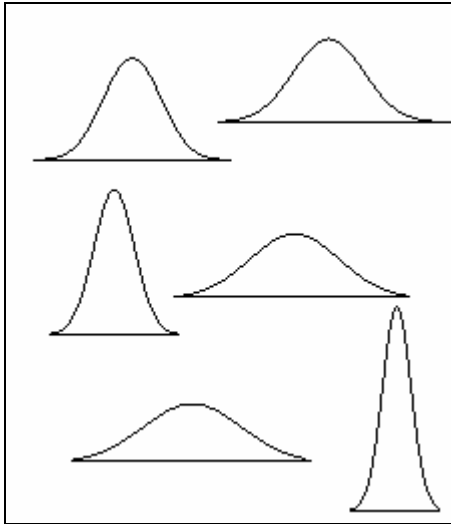
³ Brody

- “Fitness connectivity product” for nodes: “the number of links at a particular time multiplied by the rate at which the node recruits new links.”⁴
 - Evolvability
 - Ability of a hub to evolve in a way that maintains its connections to its various nodes
 - Spreading ideas or shutting them down
 - Hubs are useful for spreading ideas
 - Medicines are marketed to doctors who use them for their patients
 - Displacement
 - If a hub cannot continue to maintain its links with its various nodes, then it will cease to be a hub
 - Apple’s Newton >>> Palm Pilot
 - Disease and terrorism spread by attacks on hubs
 - AIDS is carried by many people but is sustained by hubs, people who have many sexual partners
 - Disease: a plague cannot spread if various cities are walled off and separated
 - Isolates vs. Hubs
 - Isolates are nodes that successfully exist without being a hub
 - Conversely, some hubs “eat” other hubs to the point that they become stagnant
 - Black holes eat everything in its path
 - Other examples: “...governments as well as health maintenance organizations, entrepreneurs, bullies, and rapists. Once, however, a hub is completely dominant, evolvability diminishes and evolution stumbles.”⁵
 - Modes
 - An attribute quality about a node

Normal Distributions

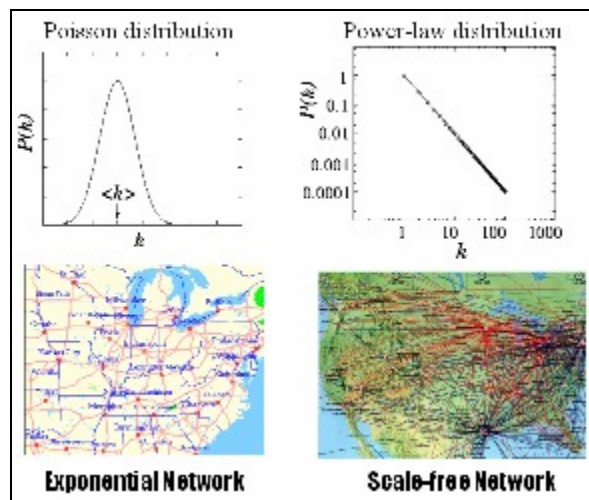
⁴ Brody

⁵ Brody



6

Normal distributions are a family of distributions that have the same general shape. They are symmetric with scores more concentrated in the middle than in the tails. Normal distributions are sometimes described as bell shaped. Examples of normal distributions are shown to the right. Notice that they differ in how spread out they are. The area under each curve is the same. The height of a normal distribution can be specified mathematically in terms of two parameters: the mean (μ) and the standard deviation (σ).⁷

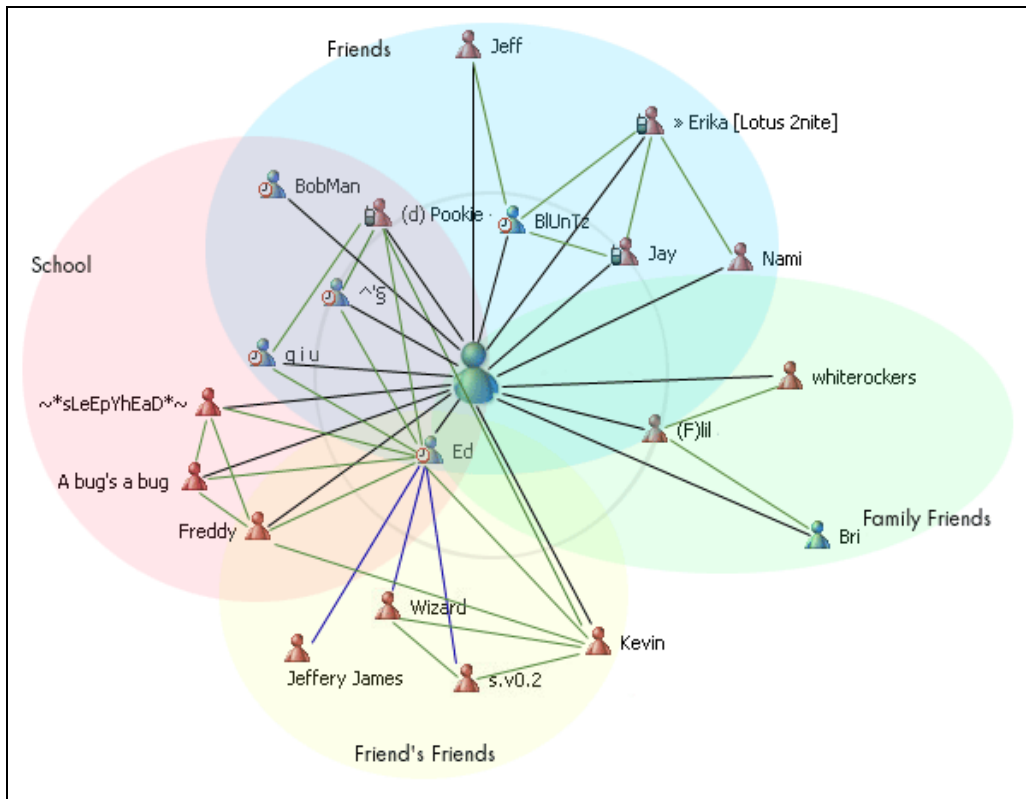
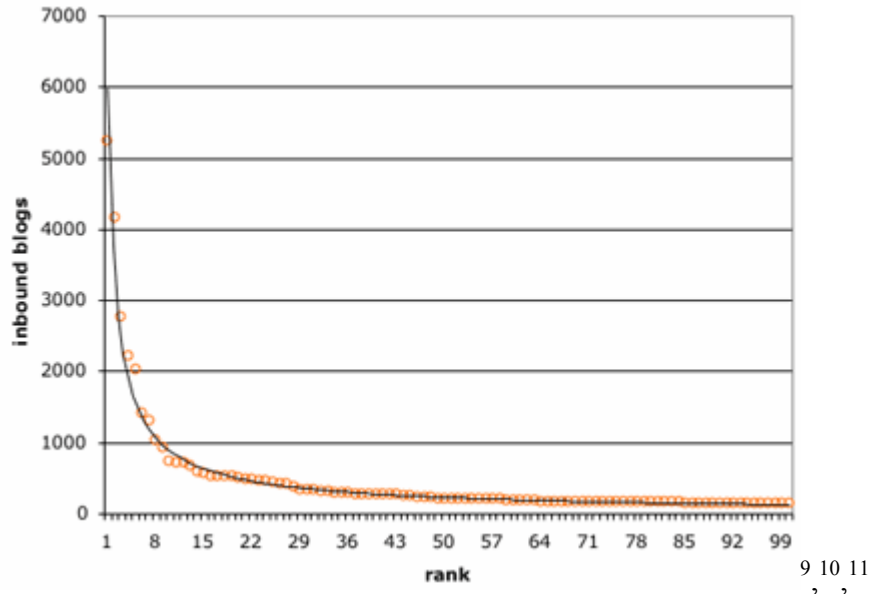


8

⁶ <http://davidmlane.com/hyperstat/A15211.html>

⁷ <http://davidmlane.com/hyperstat/A15211.html>

⁸ Barabasi



12

Comparing Random and Scale-Free Distribution

In the random network, the five nodes with the most links (in red) are connected to only 27% of all

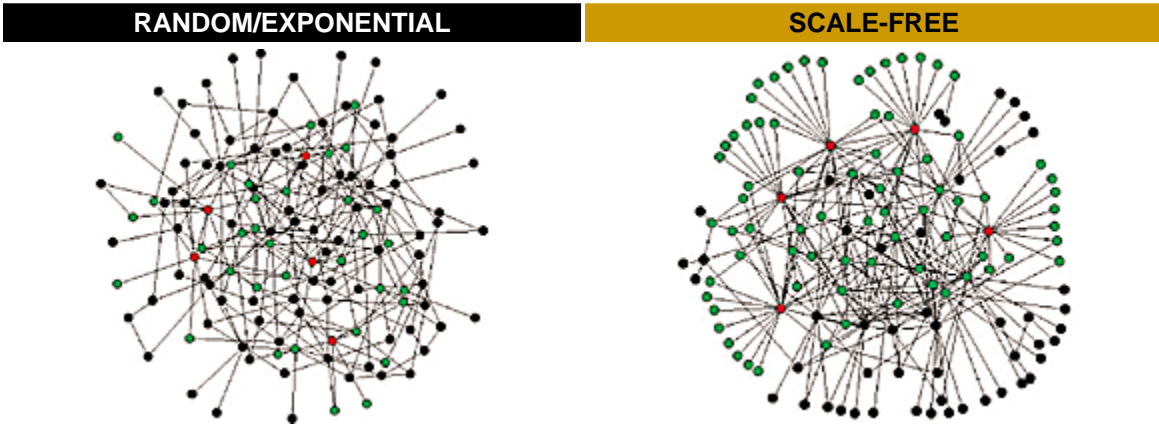
⁹ <http://www.kottke.org/03/02/weblogs-and-power-laws>

¹⁰ <http://www.michael-forman.com/language/network.html>

¹¹ <http://www.technorati.com/live/top100.html>

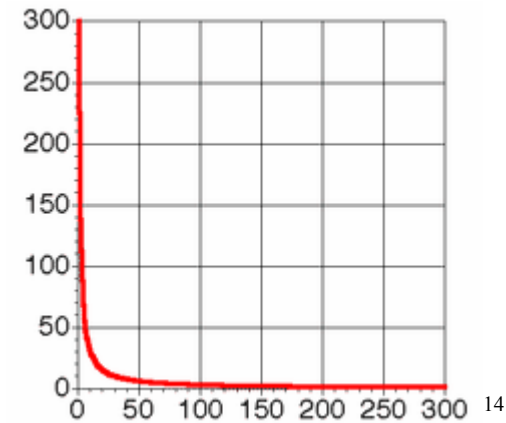
¹² <http://valiant.iat.sfu.ca/eculture/kjohnson/000432.html>

nodes (green). In the scale-free network, the five most connected nodes (red) are connected to 60% of all nodes (green).

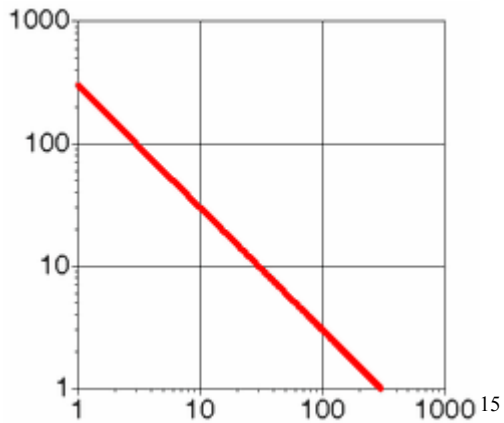


Source: the journal Nature¹³

 Note: both of the graphs below are for power curves in scale-free networks



Power curve

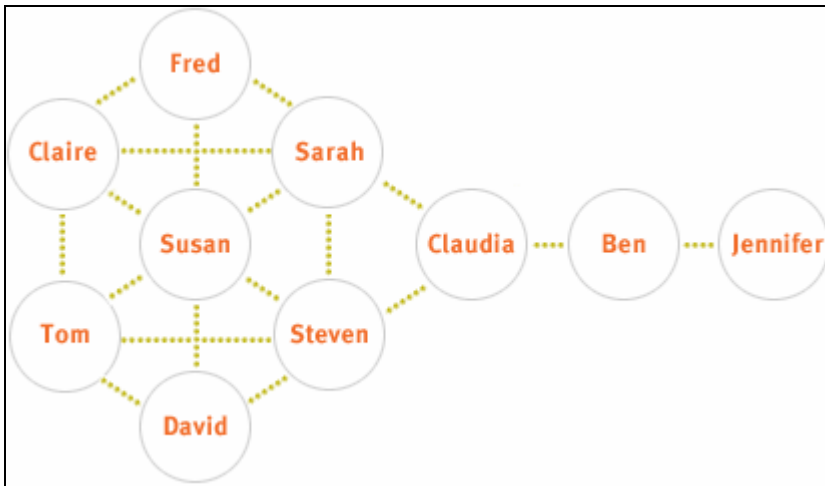


Same power curve graphed log-log scale

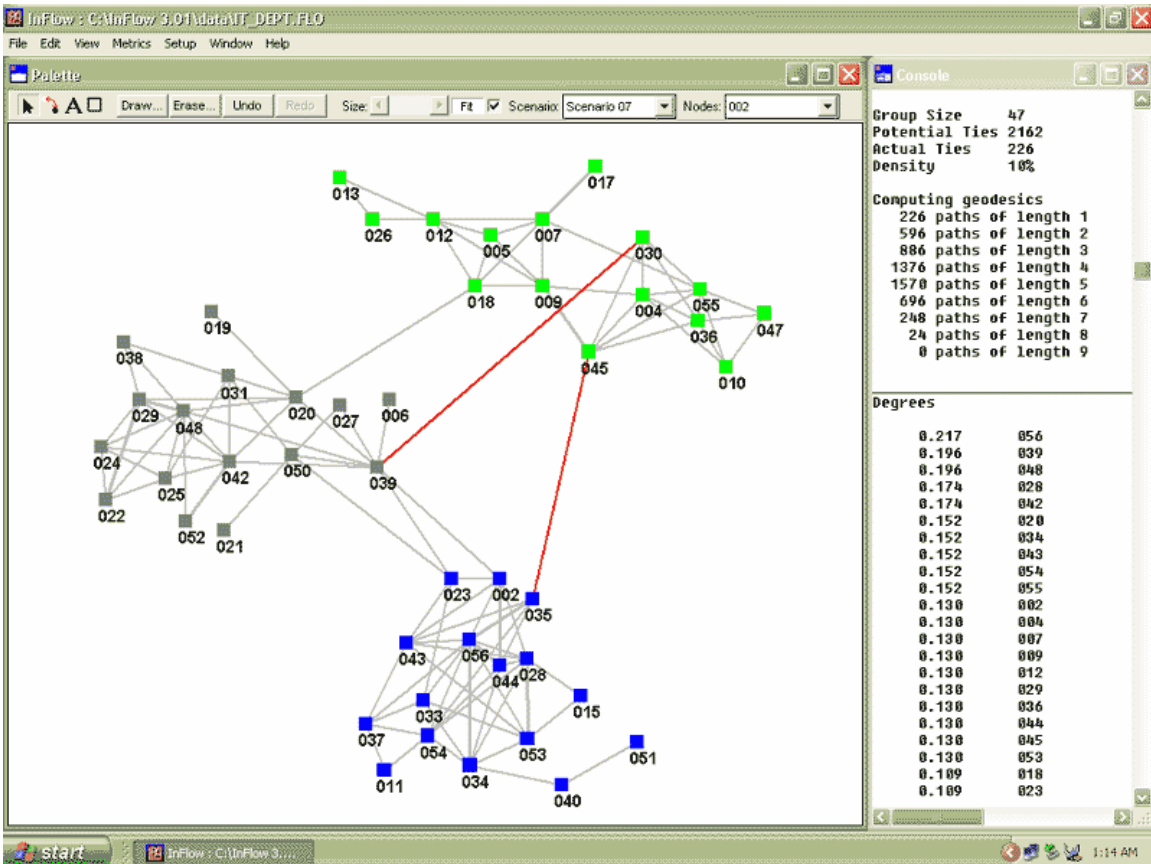
¹³ <http://www.computerworld.com/networkingtopics/networking/story/0,10801,75539,00.html>

¹⁴ <http://valiant.iat.sfu.ca/eculture/kjohnson/000244.html>

¹⁵ <http://valiant.iat.sfu.ca/eculture/kjohnson/000244.html>



16



17

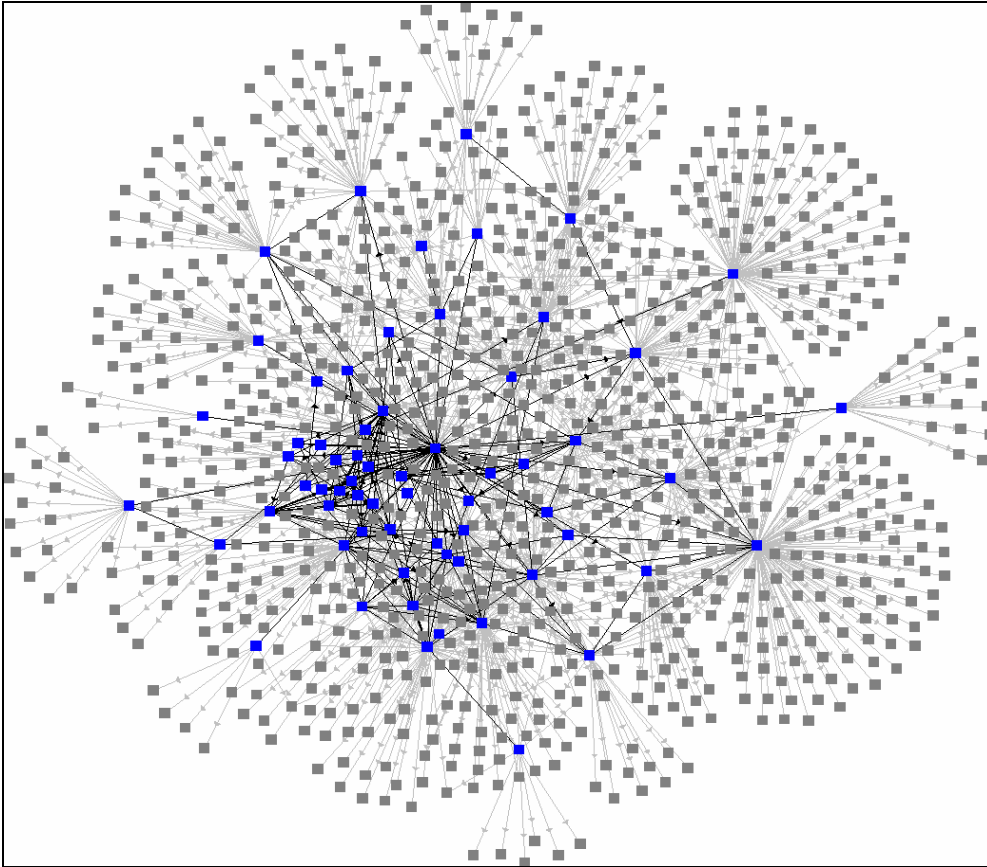
- <http://www.atlassian.com/software/confluence/default.jsp>
- <http://www.analytictech.com/netdraw.htm>
- <http://www.andrew.cmu.edu/user/krack/krackplot/krackindex.html>
- <http://www.research.att.com/sw/tools/graphviz/>

¹⁶ <http://semanticstudios.com/publications/semantics/000006.php>

¹⁷ <http://www.orgnet.com/inflow3.html>

http://www.metasight.co.uk/Knowledge_Management.htm

http://www.insna.org/INSNA/soft_inf.html - for a list of networking software



¹⁸

¹⁸ <http://radio.weblogs.com/0114726/categories/socialNetworks/2003/02/05.html>