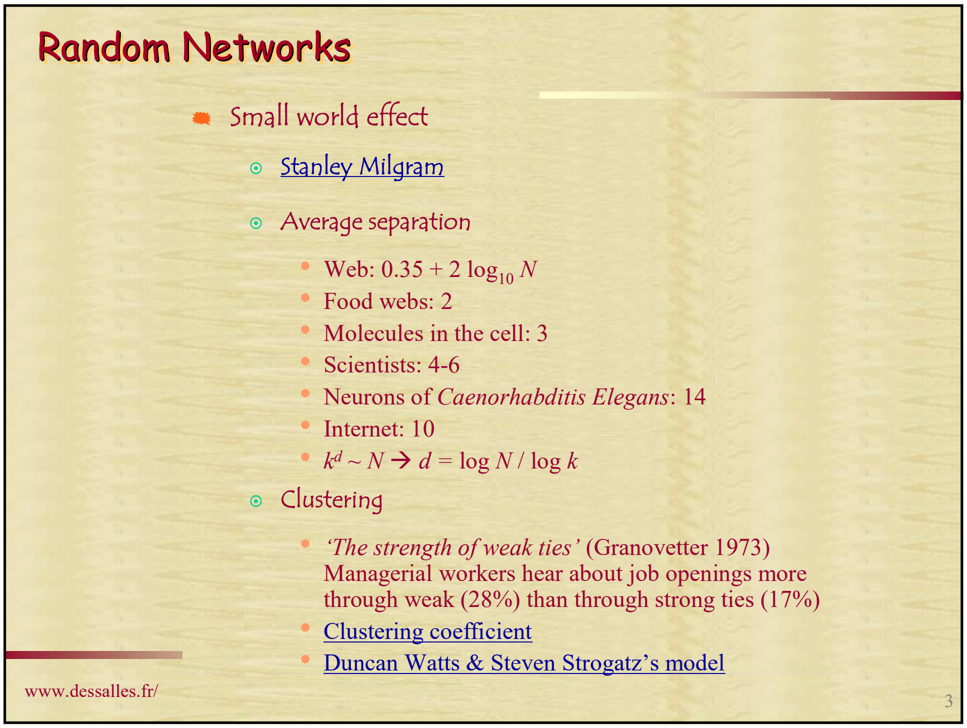
Small worlds

Jean-louis Dessalles

Telecom Paris – November 23

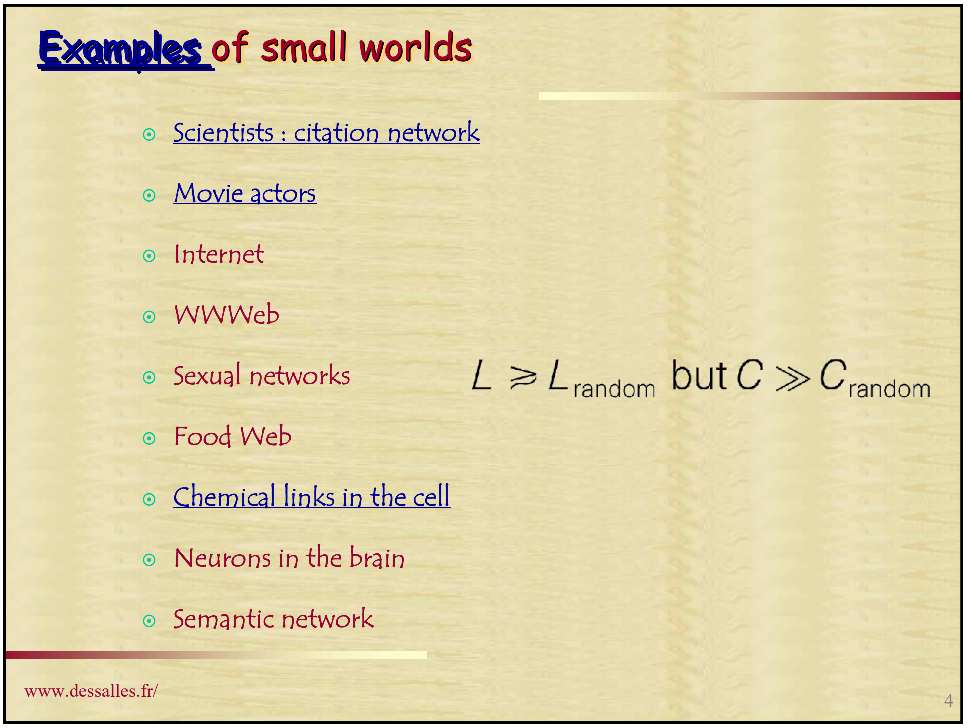
Random Networks

- Networks
- Ramdom networks (Paul Erdös & Albert Rényi)
    - $L \sim N$: giant cluster
        - Phase transition
        - Percolation
    - Poisson distribution
    - Clustering coefficient
        - a network of 71000 mathematicians: 0.1 instead of 10^{-5}



Random Networks

- Small world effect
  - Stanley Milgram
  - Average separation
    - Web: 0.35 + 2 $\log_{10} N$
    - Food webs: 2
    - Molecules in the cell: 3
    - Scientists: 4-6
    - Neurons of *Caenorhabditis Elegans*: 14
    - Internet: 10
    - $k^d \sim N \rightarrow d = \log N / \log k$
  - Clustering
    - *'The strength of weak ties'* (Granovetter 1973) Managerial workers hear about job openings more through weak (28%) than through strong ties (17%)
    - Clustering coefficient
    - Duncan Watts & Steven Strogatz's model



Examples of small worlds

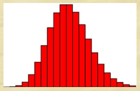
- Scientists : citation network
- Movie actors
- Internet
- WWWeb
- Sexual networks
- Food Web
- Chemical links in the cell
- Neurons in the brain
- Semantic network

$$L \geqslant L_{\text{random}} \text{ but } C \gg C_{\text{random}}$$

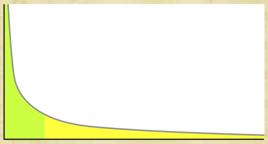
Hubs

- Highly connected nodes
  - The number of nodes with *k* links follows a power law
  - Microsoft messenger
  - Movie actors
- Models
  - Preferential attachment (rich get richer)
  - Edge copying model (cf. proteins)

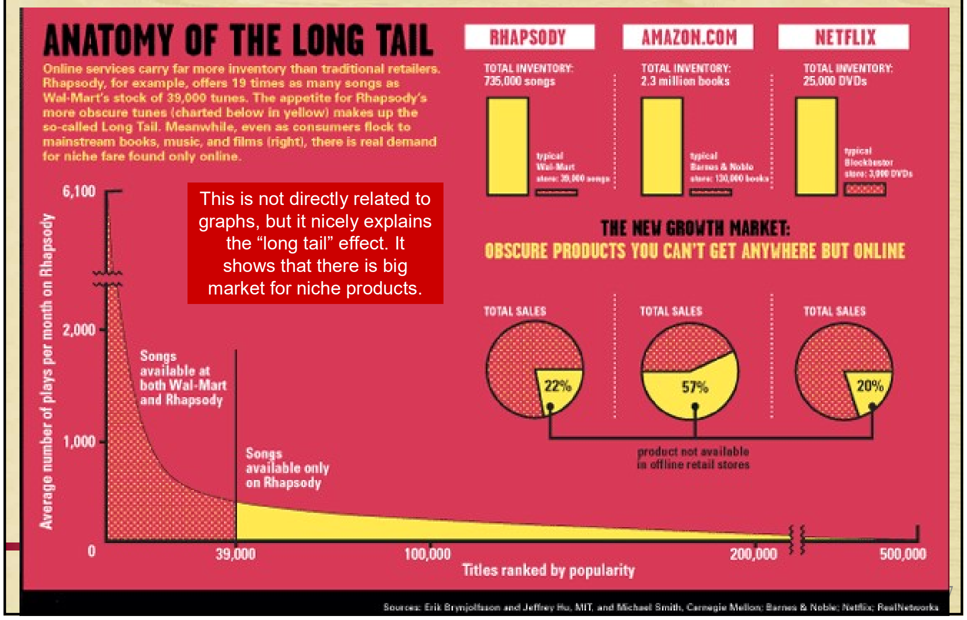
Degree distributions (1)



- Let p_k denote a fraction of nodes with degree k
- We can plot a histogram of p_k vs. k
- In a Erdös-Rényi random graph degree distribution follows Poisson distributi
- Degrees in real networks are heavily skewed to the right
- Distribution has a long tail of values that are far above the mean
- Heavy (long) tail:
  - Amazon sales
  - word length distribution, …



Detour: how long is the long tail?



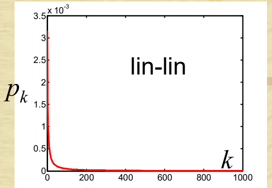
Online services carry far more inventory than traditional retailers. Rhapsody, for example, offers 19 times as many songs as Wal-Mart's stock of 39,000 tunes. The appetite for Rhapsody's more obscure tunes (charted below in yellow) makes up the so-called Long Tail. Meanwhile, even as consumers flock to mainstream books, music, and films (right), there is real demand for niche fare found only online.

This is not directly related to graphs, but it nicely explains the "long tail" effect. It shows that there is big market for niche products.

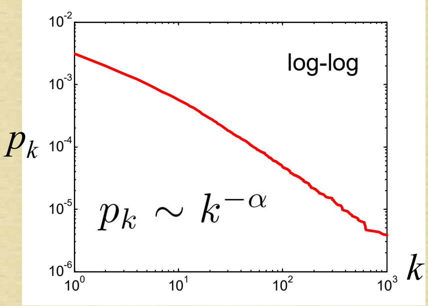
Sources: Erik Brynjolfsson and Jeffrey Hu, MIT, and Michael Smith, Carnegie Mellon; Barnes & Noble; Netflix; RealNetworks

---

Degree distributions (2)



- Many real world networks contain hubs: highly connected nodes
- We can easily distinguish between exponential and power-law tail by plotting on log-lin and log-log axis
- In scale-free networks maximum degree scales as $n^{1/(\alpha-1)}$



Degree distribution in a blog network

Preferential attachment

(Felix Richart in 2017)

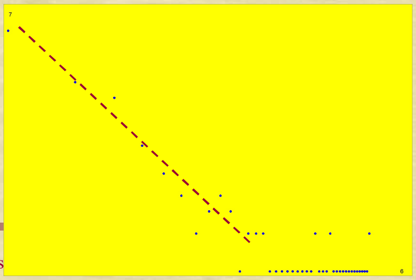
This program simulates a population of individuals who have to choose other individuals to link to.

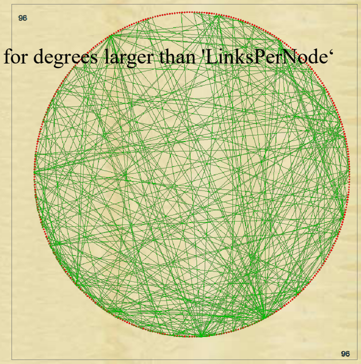
Strategies:

0- random: Individuals connect with randomly chosen individuals.
1- PA: (preferential attachment) Individuals pick a sample
and observe links internal to this sample.
They then connect with the most popular individual of the sample.
2- pick: individuals select an edge, and then connect with one
of the two extremities of that edge.

The "Network" window shows links between nodes.

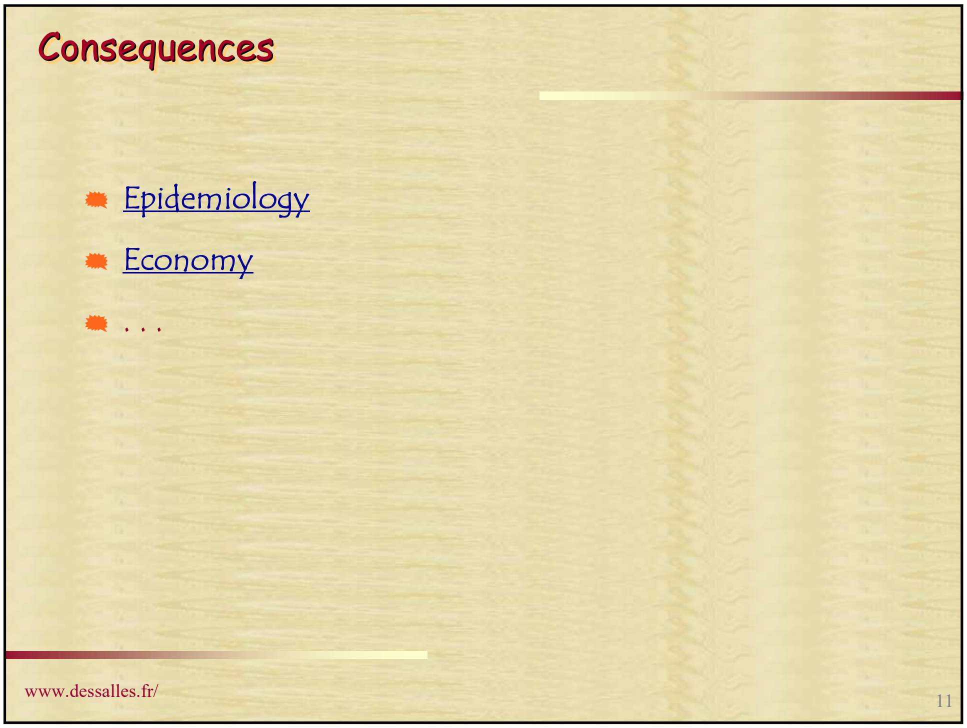
The "Degree distribution" window shows degree distribution for degrees larger than 'LinksPerNode'
in logarithmic scales.





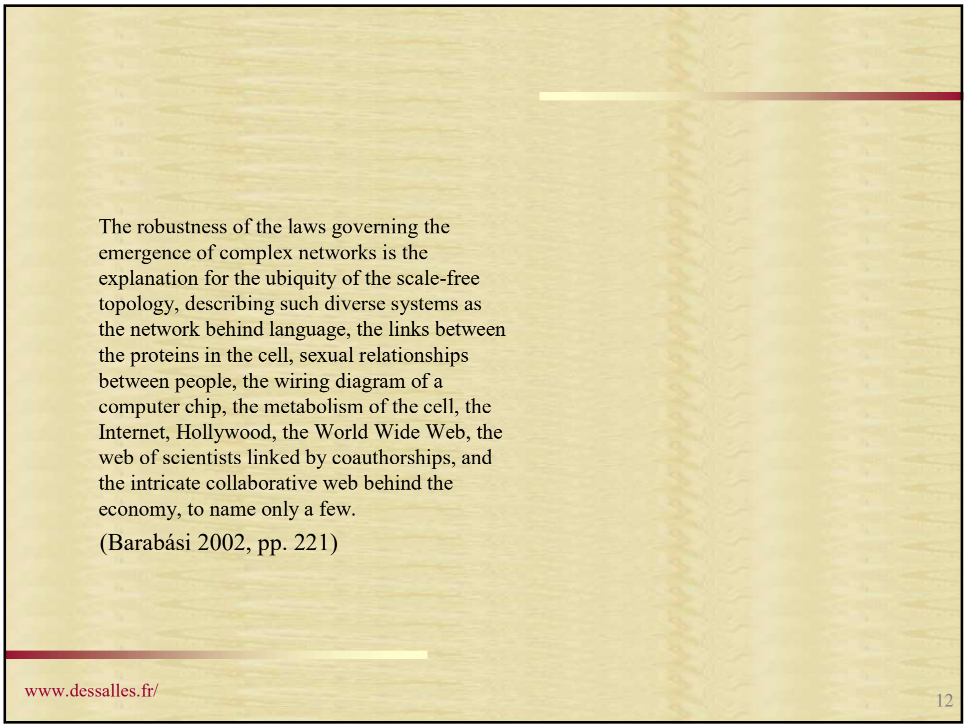
Properties

- Resilience
- Shrinking diameters
- Navigation



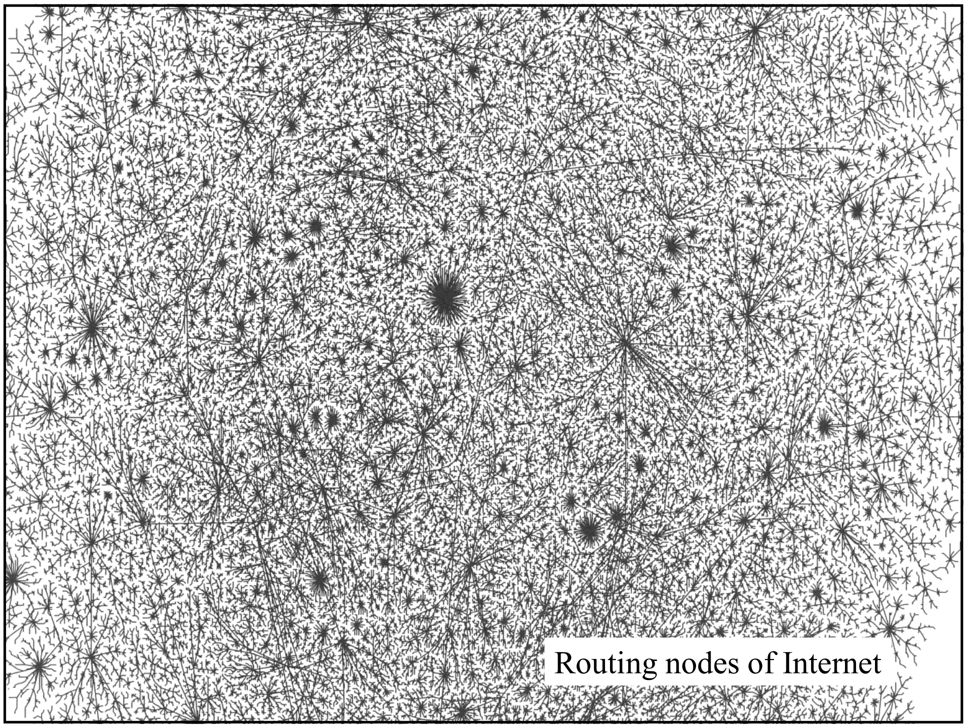
Consequences

- Epidemiology
- Economy
- . . .

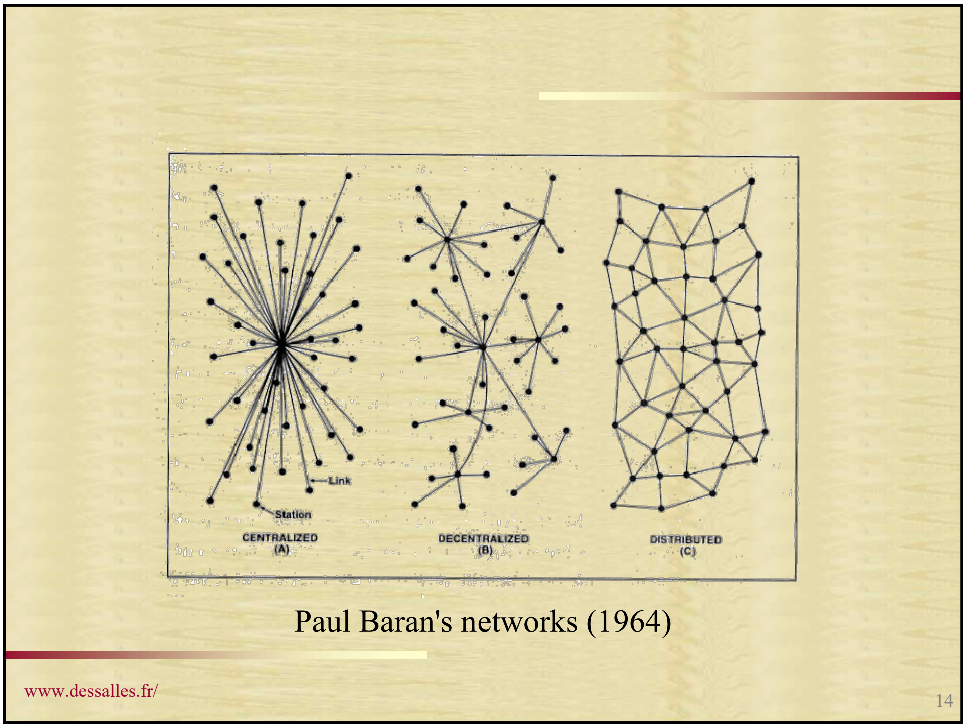


The robustness of the laws governing the emergence of complex networks is the explanation for the ubiquity of the scale-free topology, describing such diverse systems as the network behind language, the links between the proteins in the cell, sexual relationships between people, the wiring diagram of a computer chip, the metabolism of the cell, the Internet, Hollywood, the World Wide Web, the web of scientists linked by coauthorships, and the intricate collaborative web behind the economy, to name only a few.

(Barabási 2002, pp. 221)



Routing nodes of Internet



Paul Baran's networks (1964)

Measuring the separation between molecules is not an outgrowth of our obsession with six degrees of separation. The diameter of the network-or degree of separation between nodes-has biological significance. For instance, if we should find that the shortest chemical path between two molecules is one hundred, then any change in the concentration of the first molecule will have to go through one hundred intermediate reactions before reaching the second molecule. Any perturbation will decay and die along such a long path.

To our great surprise, the measurements indicated that the typical path lengths are much shorter than one hundred. In fact, cells are small worlds with *three degrees of separation*. That is, most pairs of molecules can be linked by a path of three reactions. Perturbations, therefore, are never localized: Any change in the concentration of a molecule will shortly reach most other molecules.

(Barabási 2002, pp. 185-186)

Gene duplication has a significant impact on the cellular network. It results in two identical genes, which produce identical proteins, that in turn interact with the same proteins. A new node thus has been created, the protein generated by the duplicated gene. Its neighbors, the proteins with which the duplicated protein interacts, will each now interact with both the parent and the identical offspring protein. Therefore, each protein in contact with the duplicated protein gains an extra link. In this game highly connected proteins have a natural advantage: They are more likely to have a link to the duplicating protein than their weakly connected cousins. It's not that hubs duplicate more often. Rather, since the hubs are in contact with more proteins, they are more likely to have a link to a duplicating node which offers them an extra link, a subtle version of preferential attachment.

(Barabási 2002, pp. 190-191)

Until recently economists viewed the economy as a set of autonomous and anonymous individuals interacting through the price system only, a model often called the standard formal model of economics. The individual actions of companies and consumers were assumed to have little consequence on the state of the market. Instead, the state of the economy was best captured by such aggregate quantities as employment, output, or inflation, ignoring the interrelated microbehavior responsible for these aggregate measures. Companies and corporations were seen as interacting not with each other but rather with "the market," a mythical entity that mediates all economic interactions.

In reality, the market is nothing but a directed network. Companies, firms, corporations, financial institutions, governments, and all potential economic players are the nodes. Links quantify various interactions between these institutions, involving purchases and sales, joint research and marketing projects, and so forth. The weight of the links captures the value of the transaction, and the direction points from the provider to the receiver. The structure and evolution of this weighted and directed network determine the outcome of all macroeconomic processes.

(Barabási 2002, pp. 208-209)

These events cannot be explained within a framework in which all organizations interact with a mythical market only. Cascading failures are a direct consequence of a network economy, of interdependencies induced by the fact that in a global economy no institution can work alone.

(Barabási 2002, pp. 211)