An overview of modern network science

Real-world networks, network models, quantities of interest

AMSC808N/CMSC828V Maria Cameron Nov. 12, 2020

References for today:

- M. Newman, *The structure and function of complex networks*, SIAM Review, 45/2, 167—256, 2003
- D. Watts and S. Strogatz, *Collective dynamics of 'small-world' networks*, Nature, 393, 440–442, 1998
- A.-L. Barabasi and R. Albert, Emergence of scaling in random networks, Science, 286:509-512, 1999
- A.-L. Barabasi, Network science: http://networksciencebook.com/chapter/1#networks
- A. Broido and A. Clauset, Scale-free networks are rare, Nature Communications (2019) 10:1017

Examples of real-world network

- Social networks (Facebook, Twitter, friendships, actors-movies, directors-boards, coathorships, etc.)
- Information networks (World-Wide Web, academic citations, preference networks (movies, books))
- Technological networks (electric grids, airline routes, distribution networks (mail, deliveries, etc), internet)
- Biological networks (metabolic pathways, mechanistic interactions between proteins, gene regulatory networks, food web, neural networks, blood vessels)
- Physics (networks representing energy landscapes (vertices = local minima, edges = transition states (Morse index one saddles))
- Chemistry (in hydrocarbon molecules CkHm, C-atoms form networks)

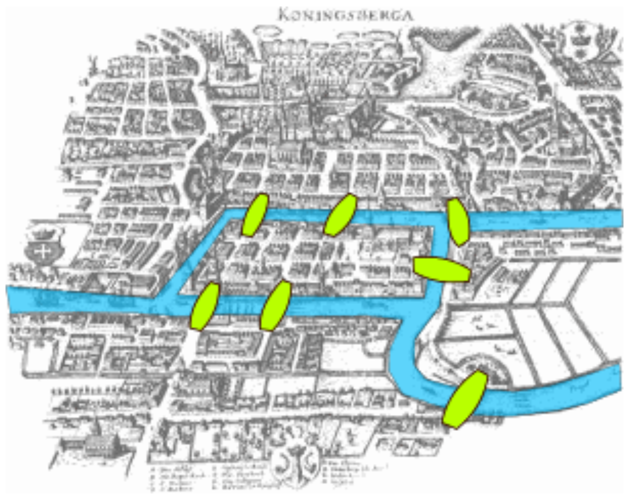
Onset of mathematical graph theory

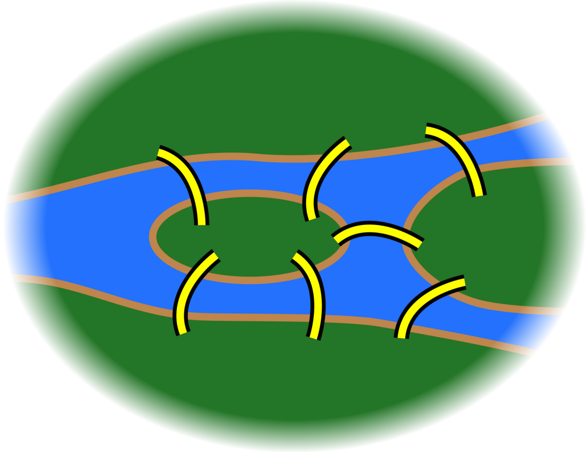
Euler, 1735

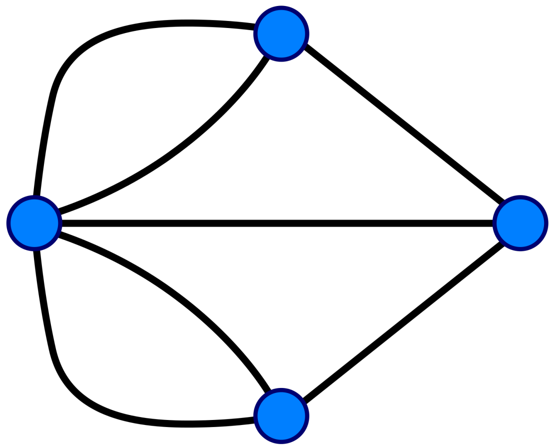
Seven bridges of Koenigsberg problem:
design a walk so that every bridge is crossed exactly once

The first mathematical proof for a graph problem

Wiki: Seven Bridges of Koenigsberg







The onset of mathematical graph theory

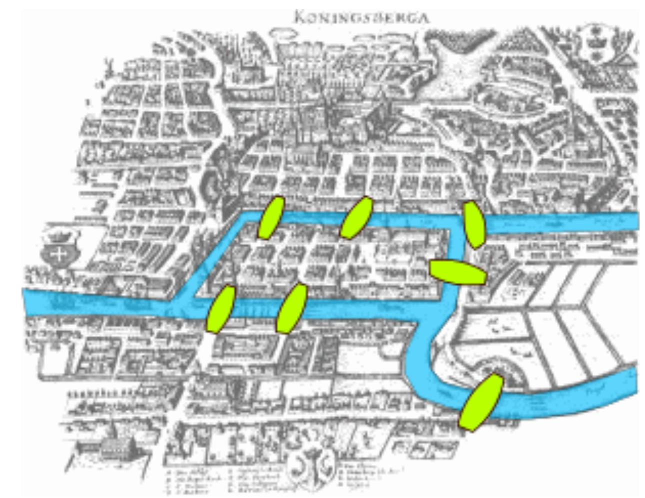
Seven bridges of Koenigsberg problem:
design a walk so that every bridge is crossed exactly once

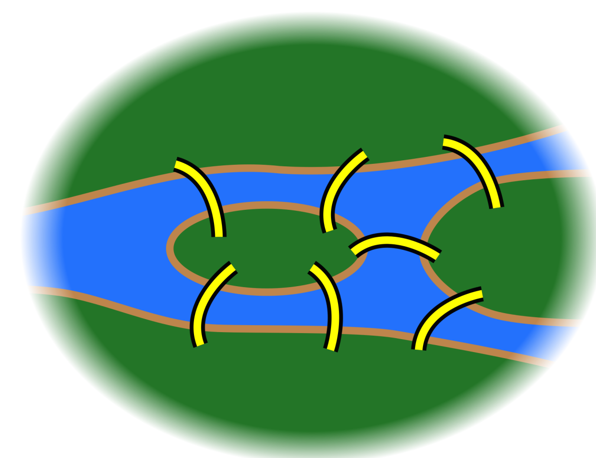
Euler's solution:

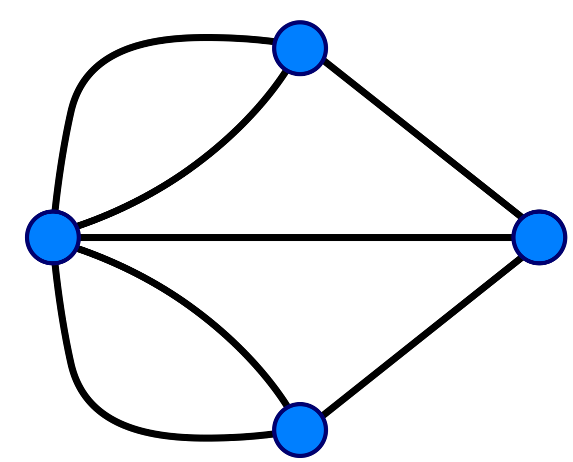
- The problem is equivalent to a problem on a graph.
- The requested walk exists if
    1. The graph is connected, and
    2. There are 0 or two vertices of odd degree.

In the latter case, the walk starts at one odd degree vertex and finishes at the other.

Wiki: Seven Bridges of Koenigsberg





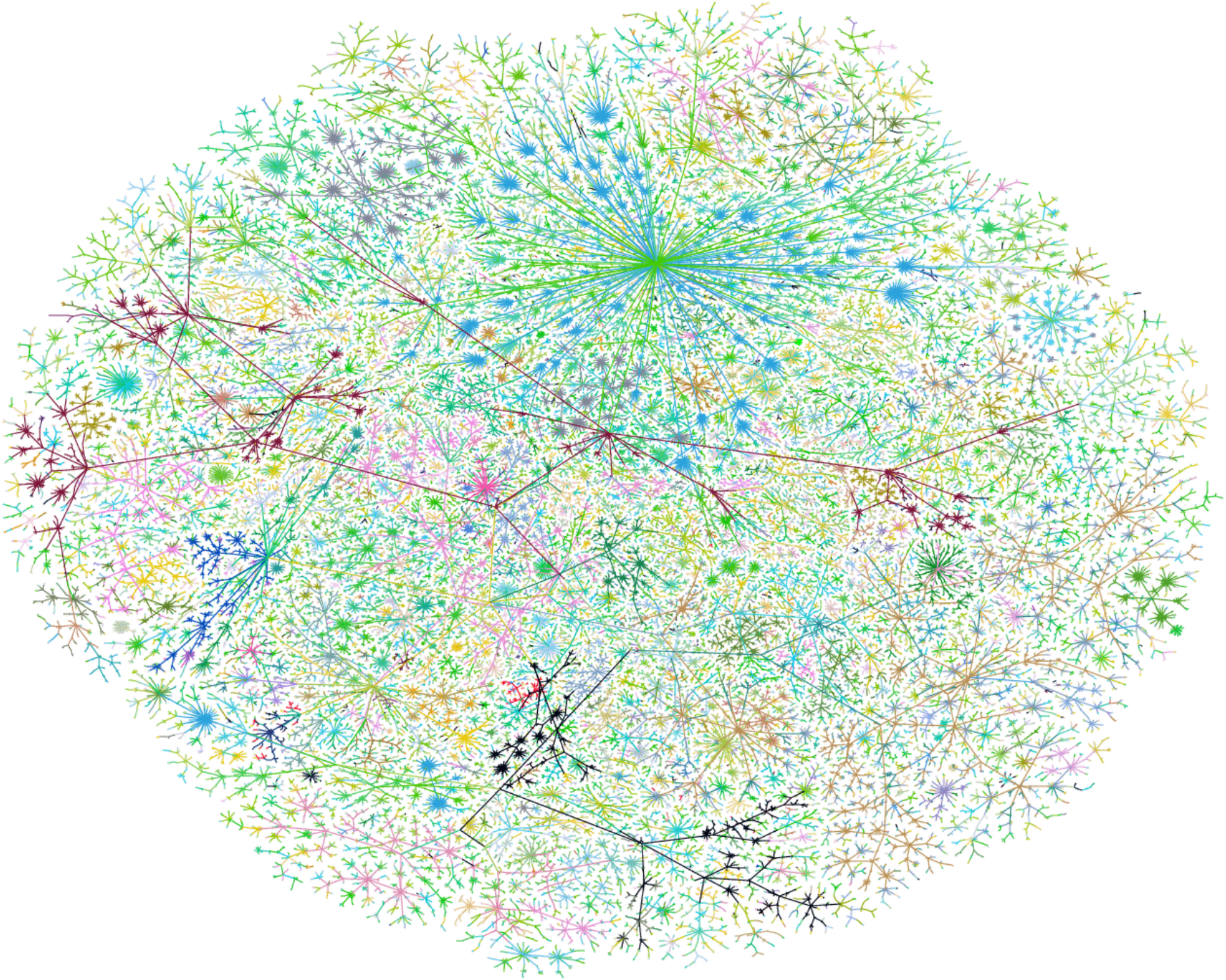


Studies of networks

- Mathematical graph theory
  - graphs are relatively small
  - properties of individual graphs are studied
  - depiction of graphs is very important
- Sociology (relationships between people)
- 21st century
  - Modern real-world networks are often huge (10^6—10^9 nodes)
  - Shift to analysis of large-scale statistical properties of graphs

Reasons for the shift to analysis of large-scale statistical properties of graphs

- Availability of computers and communication networks that allow us to gather data and analyze them
- Visual inspection is infeasible for analysis of huge graphs



A visualization of the network structure of the Internet at the level of “autonomous systems”—local groups of computers each representing hundreds or thousands of machines. Picture by Hal Burch and Bill Cheswick, courtesy of Lumeta Corporation.

Network Science

Albert Laszlo Barabasi

http://networksciencebook.com/chapter/0#introduction0

Vulnerability due to interconnectivity

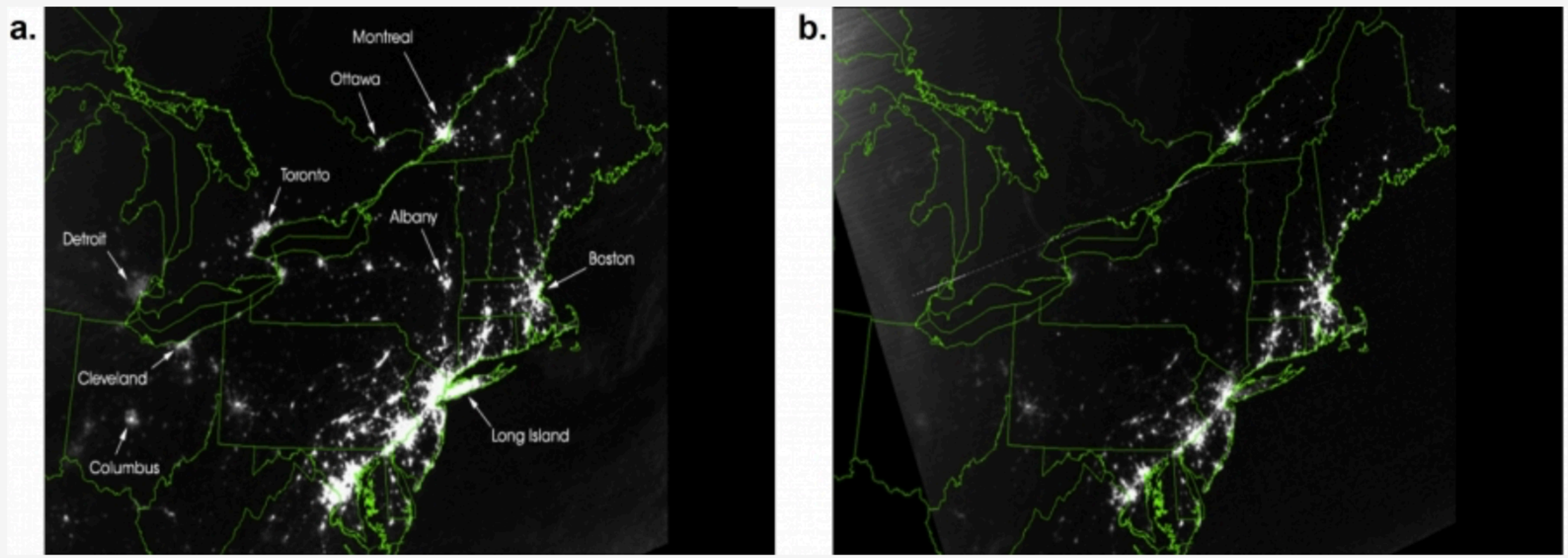


Image 1.1

2003 North American Blackout

- Satellite image on Northeast United States on August 13th, 2003,at 9:29pm (EDT), 20 hours *before* the 2003 blackout.
- The same as above, but 5 hours *after* the blackout.

- Cascading failure: 2003 blackout, 1997 and 2009-2011 financial crises.

- Artificially induced cascading failures: drying off financial supply to terrorist organizations, cancer treatment.

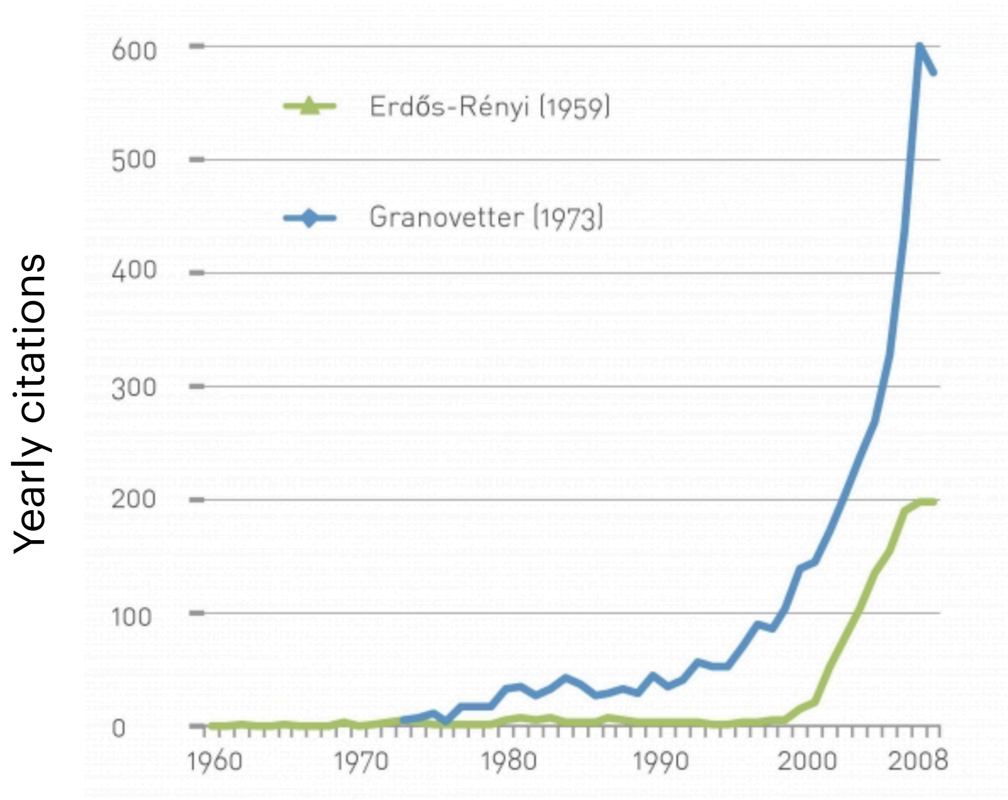
Networks at the heart of complex systems

"I think the next century will be the century of complexity", Stephen Hawking

- Cellular network (interactions between genes, proteins and metabolites) is a prerequisite of life.
- Neural network is a key to understand how our brain functions.
- Social network (the sum of professional, frienship, family ties) determines the spread of knowledge, behavior and resources.
- Communication networks (interconnected devices) are at the heart of modern communication systems.
- The power grid (generators and transmission lines) supply with energy all modern technology.
- Trade networks maintain our ability to exchange goods and services.
- Networks are at the heart of the most revolutionary technology of 21st century: Google, Facebook, Twitter.

Conclusion: to understand complex systems we need need to understand networks behind them.

Emergence of network science in 21st century



P. Erdős and A. Rényi. On random graphs. Publicationes Mathematicae, 6: 290, 1959.

M. S. Granovetter. The strength of weak ties. American Journal of Sociology, 78: 1360, 1973

- Emergence of network maps

Internet technology allows for fast data sharing and cheap digital storage.

- Universality of network characteristics

The architecture of networks emerging in various domains of science, nature, and technology are similar to each other, a consequence of being governed by the same organizing principles. Consequently we can use a common set of mathematical tools to explore these systems

The characteristics of network science

- Interdisciplinary nature: math, computer science, and numerous fields of science where network data are coming from
- Empirical, data driven nature. Predecessor: graph theory. Key difference: network science is empirical: focused on data, function, and utility. Each developed tool is tested on real data.
- Quantitative and mathematical nature. It is essential to master mathematical formalisms that allow us to analyze networks.
- Computational nature. Computational tools are essential for network science.

Societal impact

- ***Economic impact:*** from web search to social networking. The most successful companies of the 21st century, from Google to Facebook, Twitter, LinkedIn, Cisco, Apple and Akamai, base their technology and business model on networks.

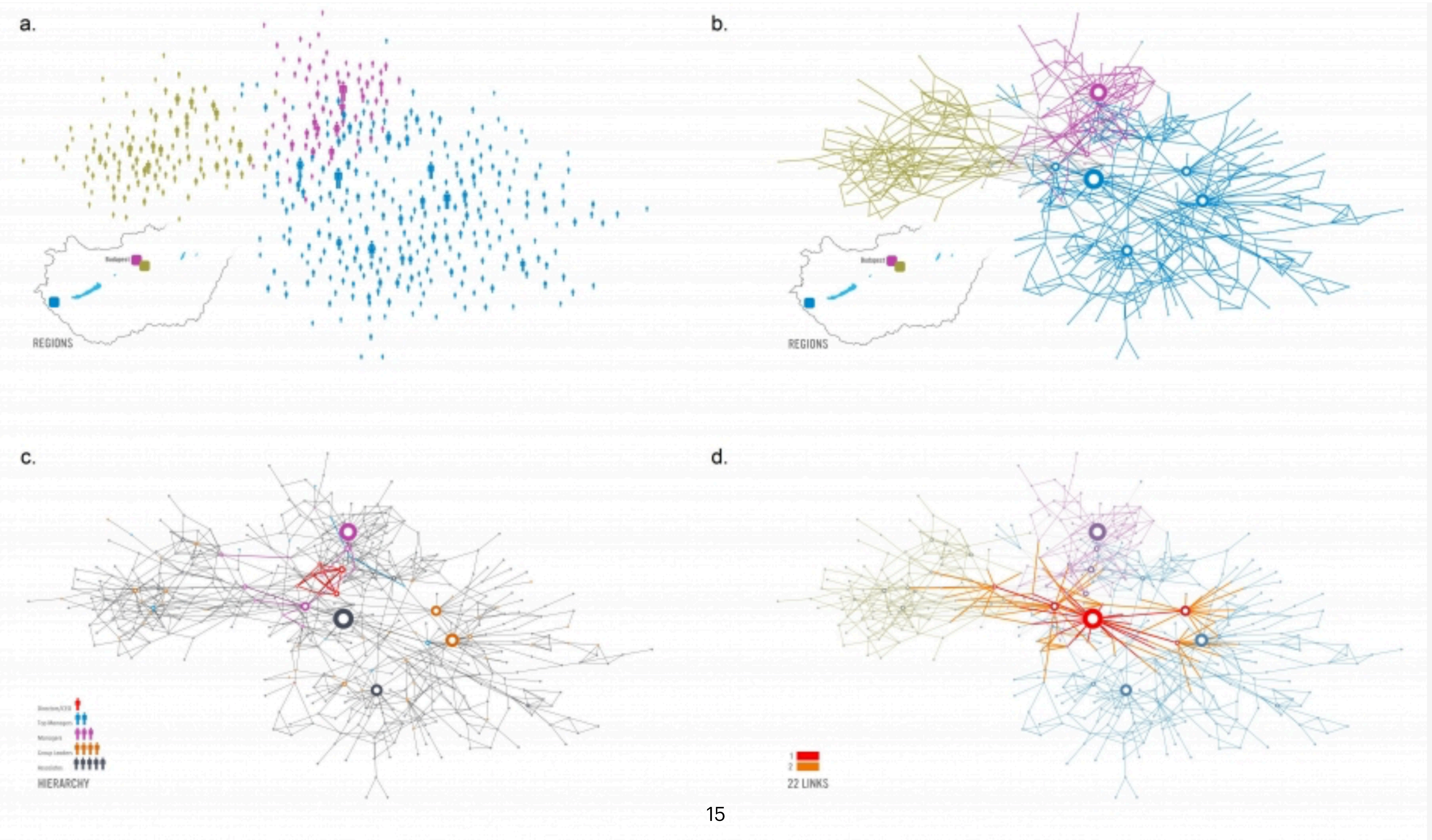
- ***Health:*** from drug design to metabolic engineering. The human genome project (completed in 2001) is the first comprehensive list of all human genes. It is insufficient per se: we need to understand how genes, proteins, metabolites, etc. interact with each other. Their breakdown is responsible for human disease. Network biology, network medicine. Drug development: network pharmacology.

- ***Security:*** fighting terrorism. Network-centric warfare (fight low intensity conflicts against terrorist and criminal networks that employ decentralized flexible organization). Spy museum in Washington D.C.

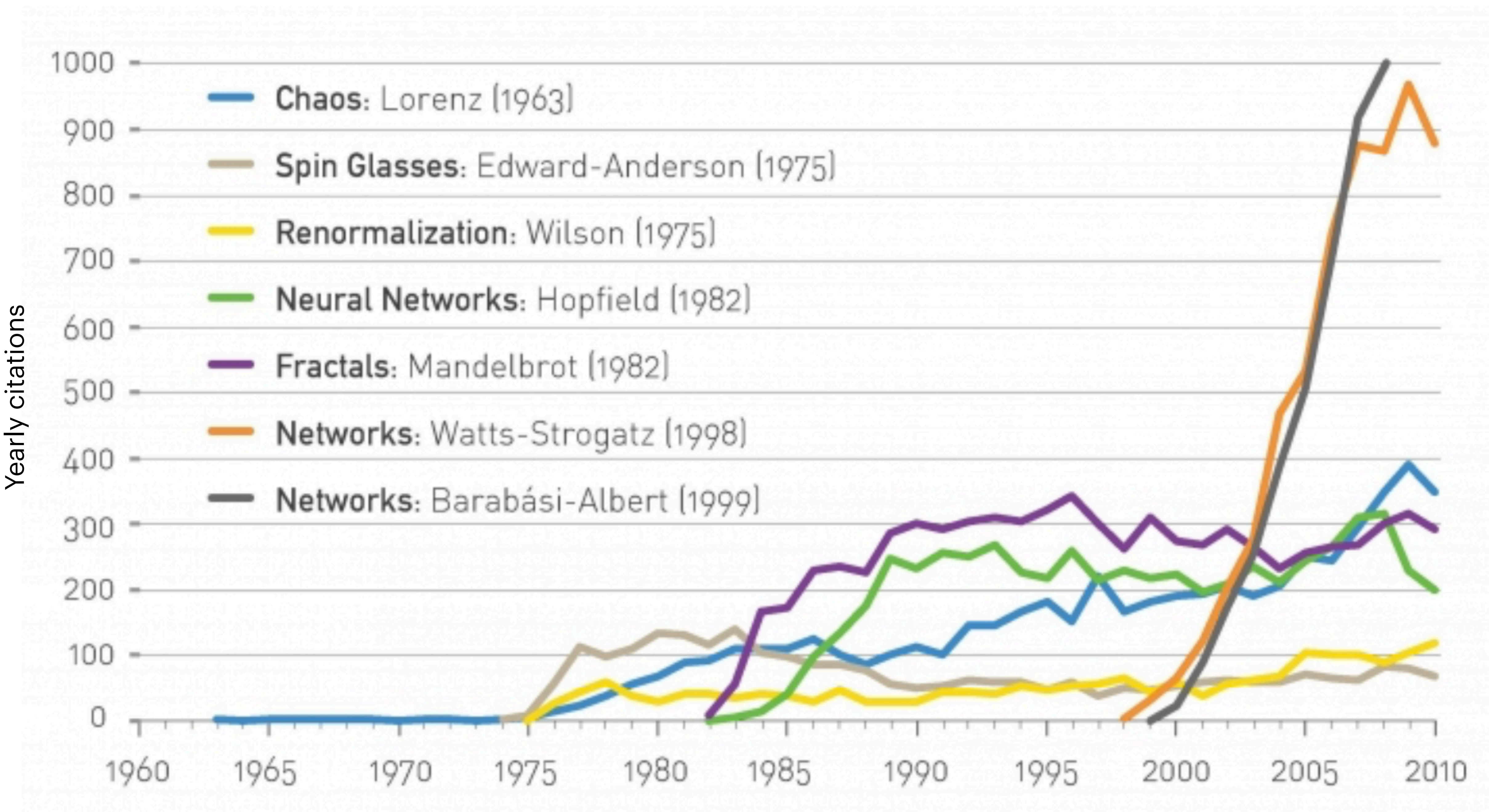
- ***Epidemics.*** Prediction of propagation. The first prediction was done for the H1N1 pandemic in 2009. Video.

- ***Neuroscience:*** mapping the brain. NIH Connectome project launched in 2010.

- ***Management: uncovering internal structure of organization.*** Companies like Maven 7, Activate Networks, and Urgent, offer tools and methodologies to map out the true structure of organizations.



Scientific impact



Most cited papers

- The 1998 paper by Watts and Strogatz in *Nature* on small world phenomena and the 1999 paper by Barabási and Albert in Science on scale-free networks were identified by Thompson-Reuters as being among *the top ten most cited papers in physical sciences during the decade after their publication*. Currently (2011) the Watts-Strogatz paper is the second most cited of all papers published in Nature in 1998 and the Barabási-Albert paper is the most cited paper among all papers published in *Science* in 1999.

- Four years after its publication the *SIAM* review by Mark Newman on network science became *the most cited paper of any journal published by the SIAM*.

- *Reviews of Modern Physics*, published since 1929, is the physics journal with the highest impact factor. Until 2012 the most cited paper of the journal was written by Nobel Prize winner Subrahmanyan Chandrasekhar, his classic 1944 review entitled *Stochastic Problems in Physics and Astronomy* [28]. During the 70 years since its publication, the paper gathered over 5,000 citations. Yet, in 2012 it was taken over by the first review of network science published in 2001 entitled *Statistical Mechanics of Complex Networks (Albert and Barabasi)*.

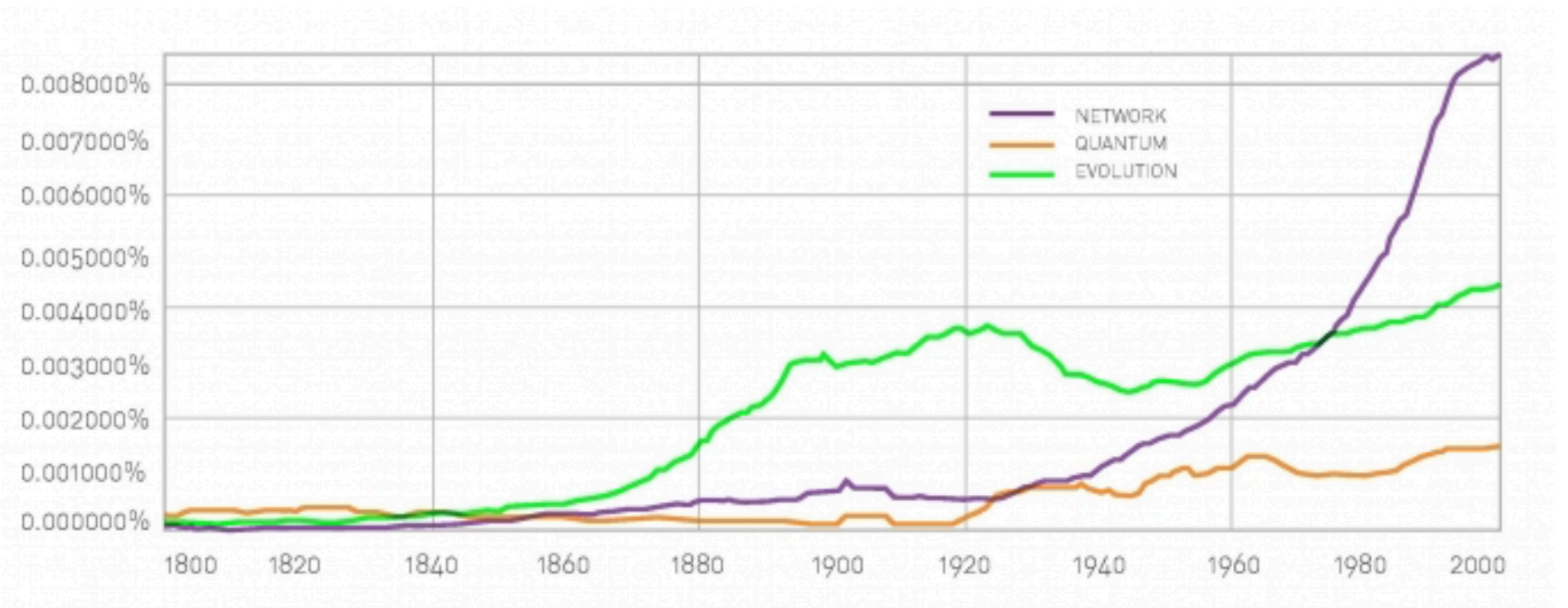
- The paper reporting the discovery that in scale-free networks the epidemic threshold vanishes, by Pastor-Satorras and Vespignani, is the most cited paper among the papers published in 2001 by *Physical Review Letters*, shared with a paper on quantum computing.

- The paper by Michelle Girvan and Mark Newman on community discovery in networks is the *most cited paper published in 2002 by Proceedings of the National Academy of Sciences*.

- The 2004 review entitled *Network Biology* (Oltvai and Barabasi) is the second most cited paper in the history of *Nature Reviews Genetics*, the top review journal in genetics.

Albert Laszlo Barabasi

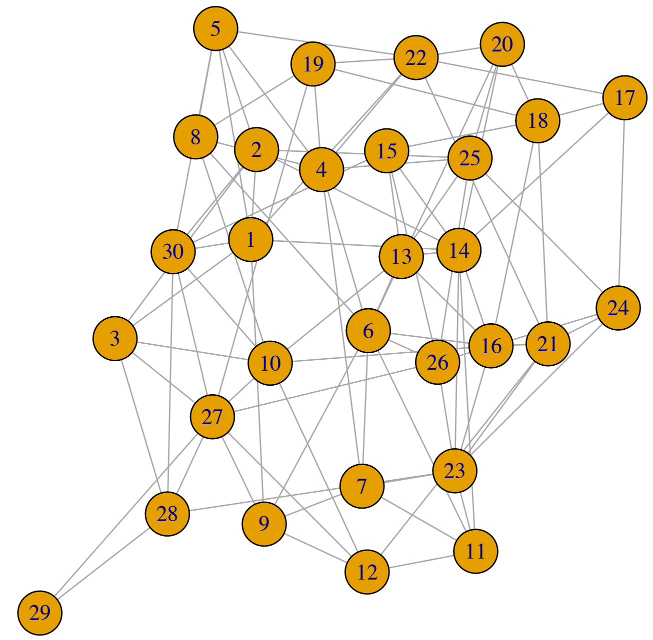
The societal awareness of networks



The frequency of use of words evolution, quantum, and networks in books published since 1880.

Network models

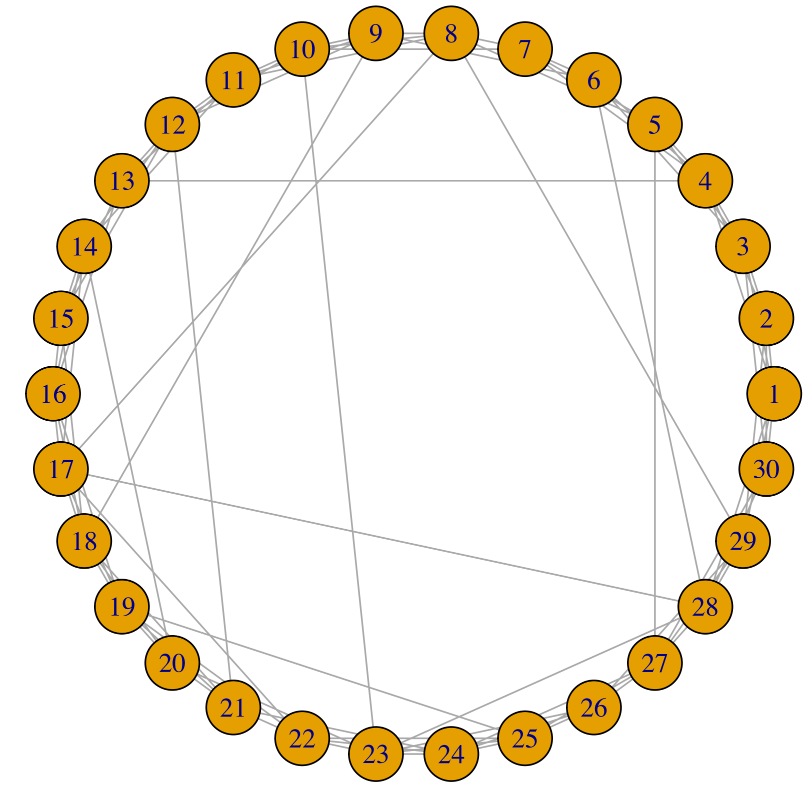
- Erdos-Renyi random graphs (1959): G(n,p)



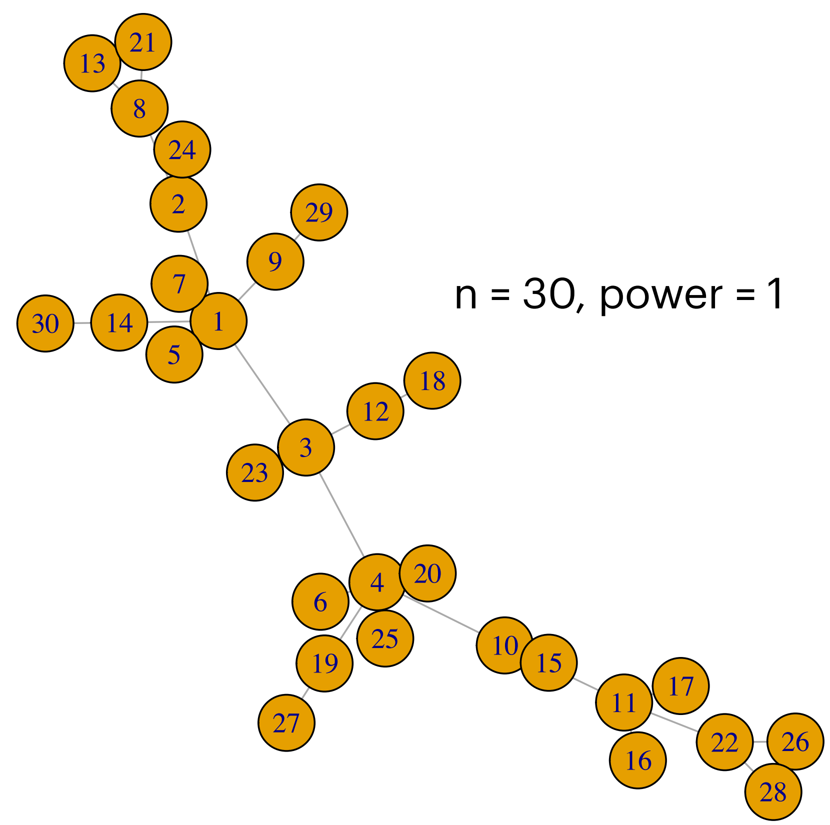
n = 30, p = 0.1

- Strogatz and Watts (1998): small-work network

n = 30, k = 3, p = 0.1

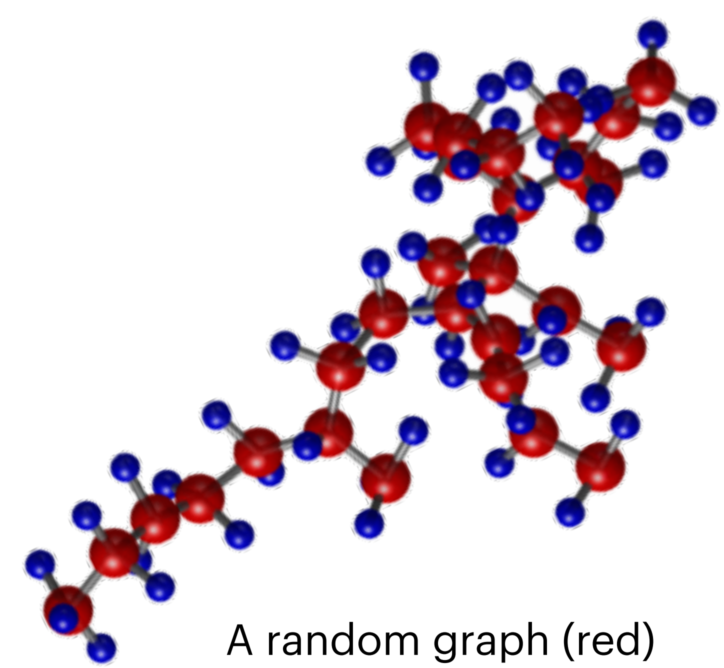


- Barabasi and Albert (1999): preferential attachment



n = 30, power = 1

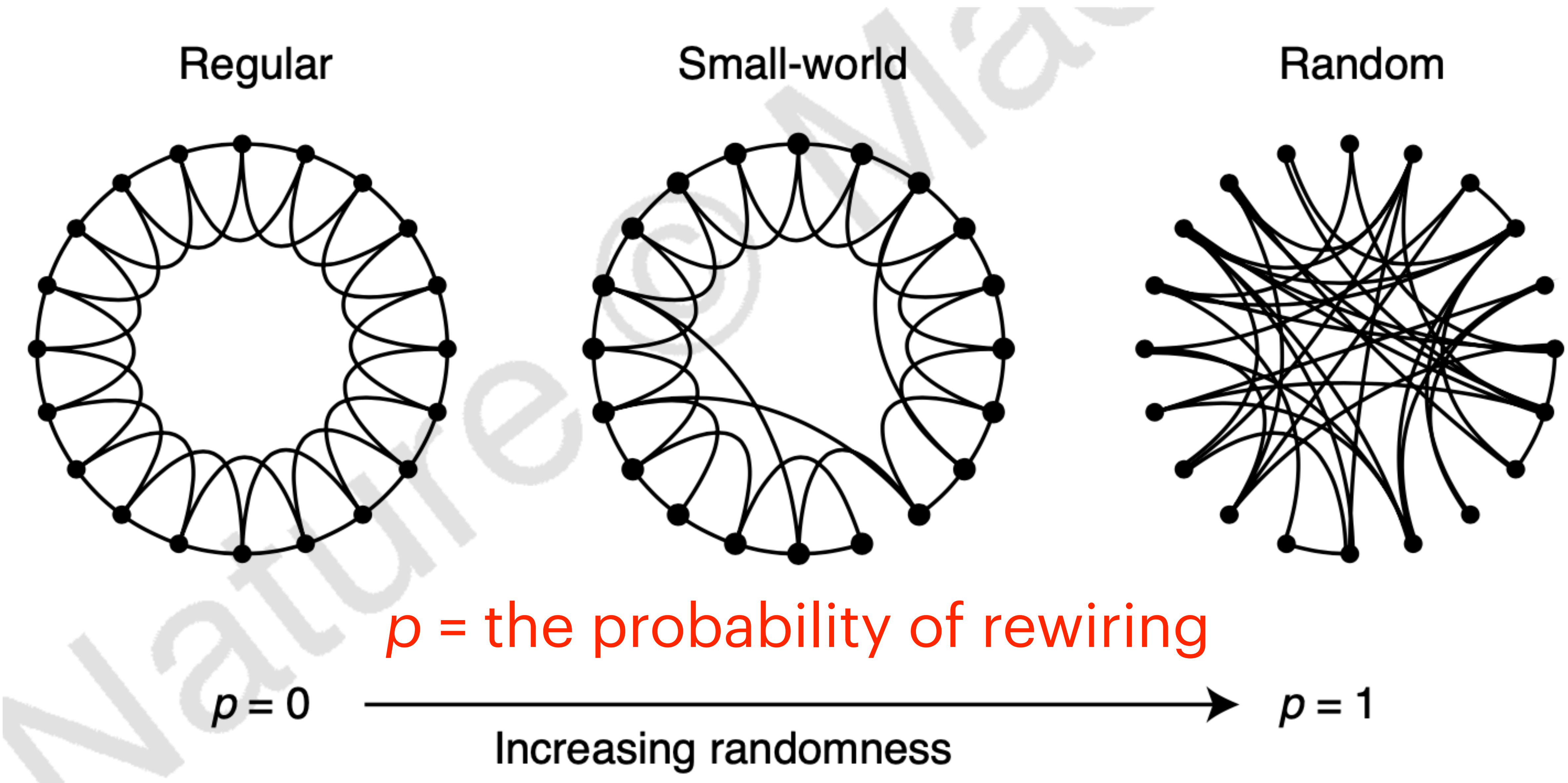
- Newman, Strogatz, Watts (2001): Random graphs with arbitrary degree distribution



A random graph (red)
with the given degree distribution
p_0, p_1, p_2, p_3, p_4

Collective dynamics of small-world networks

D. Watts and S. Strogatz, Nature, 1998



Two extreme cases:

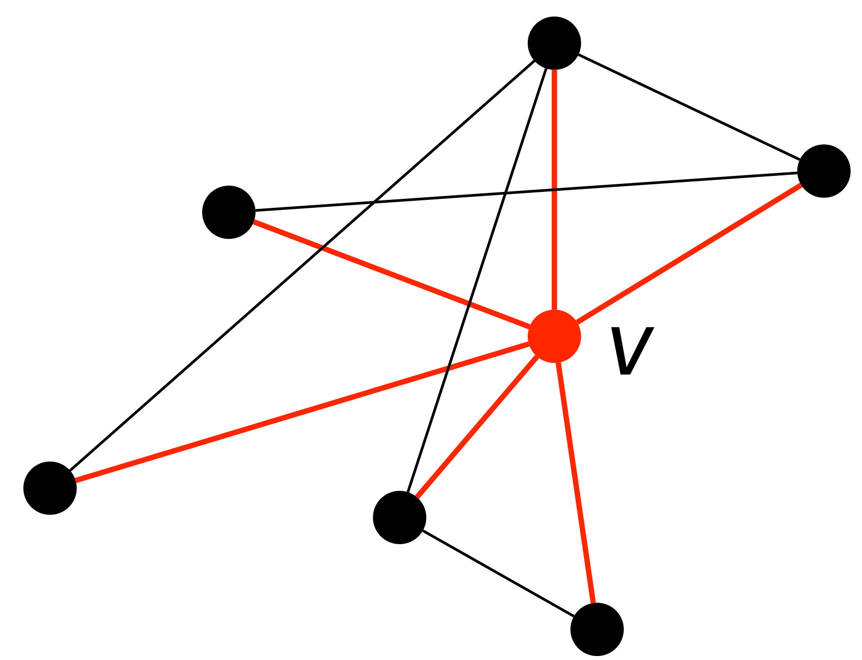
- Lattice-type network (ordered)
- Random graphs

Watts and Strogatz:

Real-world networks are somewhere in between

Two quantities of interest

- Characteristic path length $L(p)$ = mean length of the shortest path between two vertices
- Clustering coefficient C(p)



k_v nearest neighbors

$$C_v = \frac{N_{\text{edges between nearest neighbors of } v}}{(1/2)k_v(k_v - 1)}$$

$$C = \langle C_v \rangle$$

$L(p)$ and $C(p)$ at extreme cases

n = # of vertices, k = # of edges per vertex

Graphs of interest: $n \gg k \gg log(n) \gg 1$

Sparse

Large

Random graphs
will be connected

$$p \to 0: \quad L \sim \frac{n}{2k} \gg 1, \quad C \sim \frac{3}{4}$$

$$p \to 1: \quad L \approx L_{\text{random}} \sim \frac{\log(n)}{\log(k)}, \quad C \approx C_{\text{random}} \sim \frac{k}{n} \ll 1$$

Watts and Strogatz:

There is a large range of values of p where $L(p) \sim L_{random}$ while $C(p)$ is much larger than C_{random}

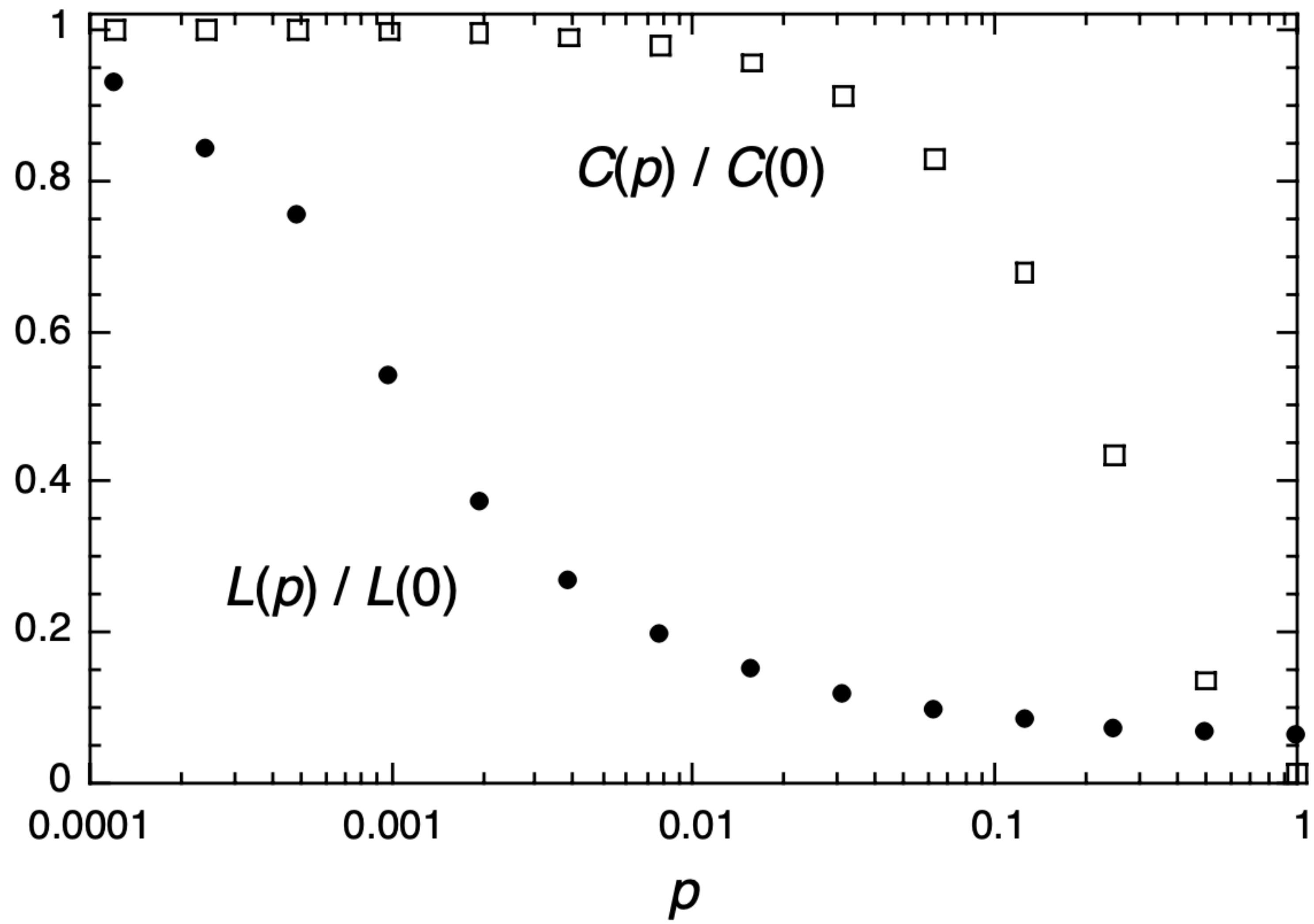
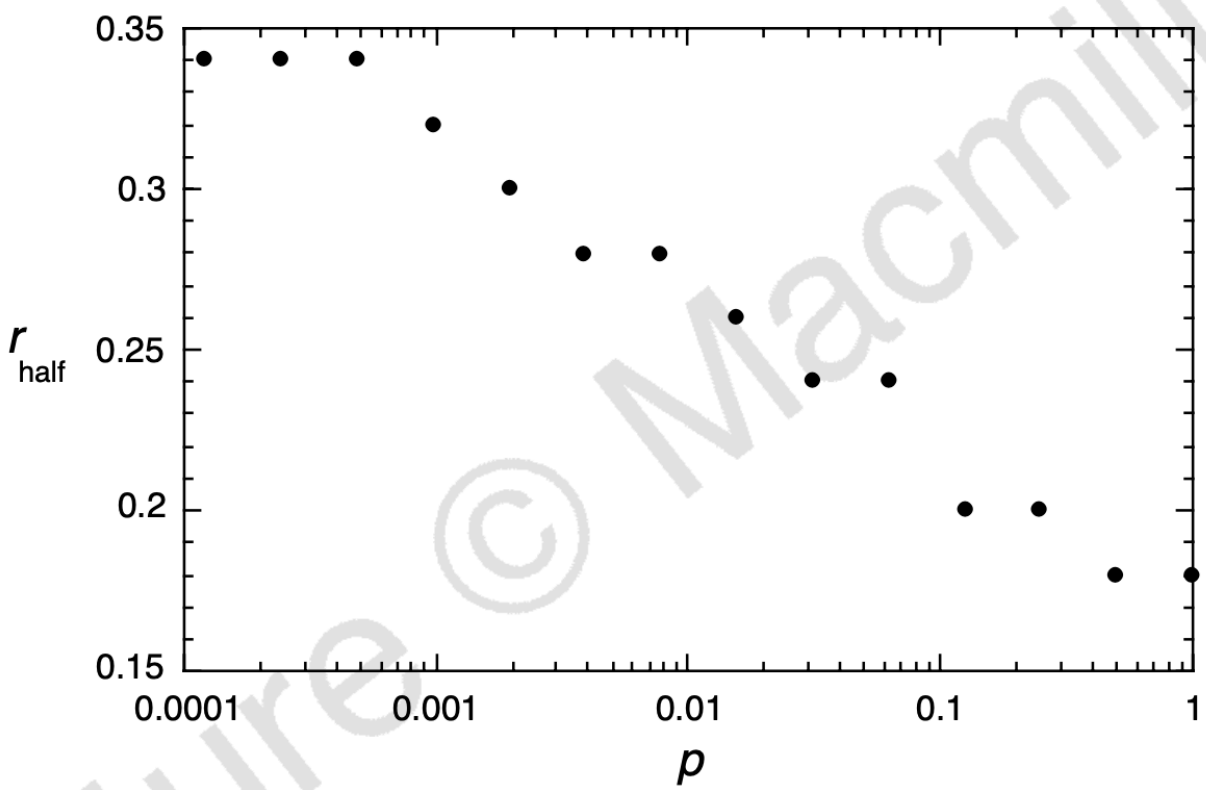


Table 1 Empirical examples of small-world networks

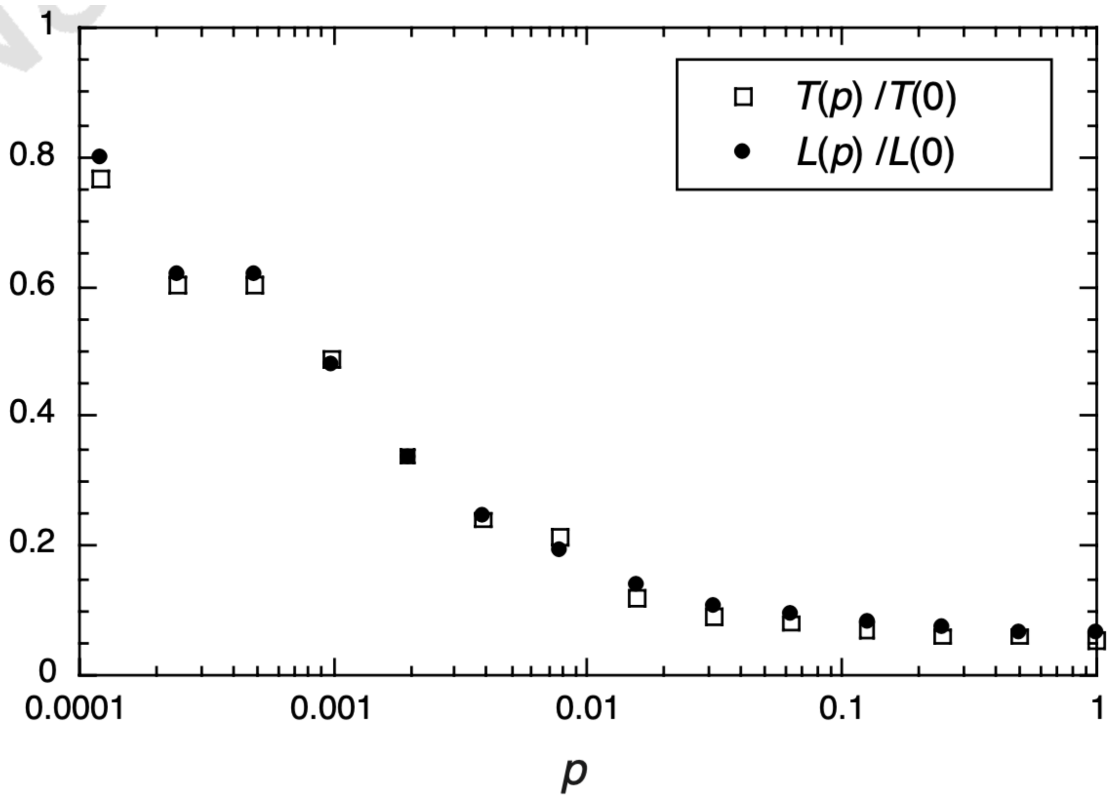
	L_{actual}	L_{random}	C_{actual}	C_{random}
Film actors	3.65	2.99	0.79	0.00027
Power grid	18.7	12.4	0.080	0.005
C. elegans	2.65	2.25	0.28	0.05

Dramatic effect of shortcuts

SIR model



The probability that half of the population gets infected



The time required for the maximally infectious disease to spread throughout the entire population