Random graphs and complex networks

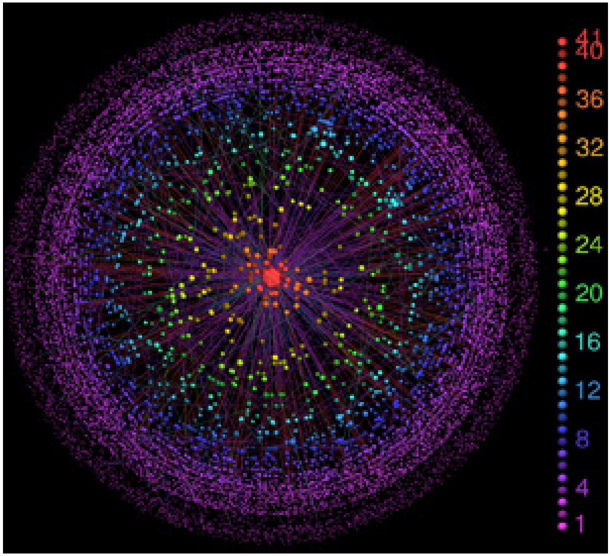
Remco van der Hofstad

TU/e technische universiteit eindhoven

EURANDOM

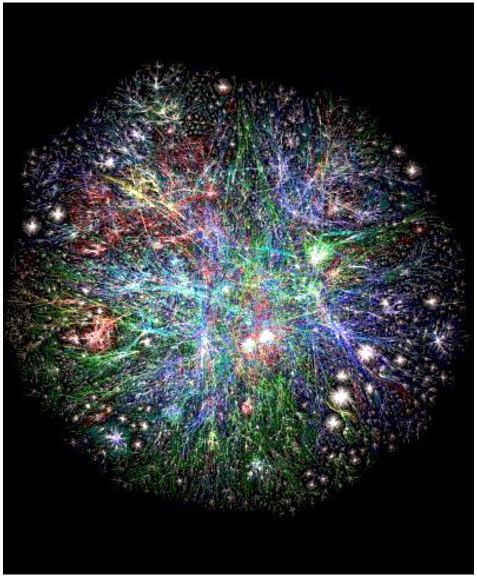
Random Graphs and Complex Networks (2WS12) course

25 years of network revolutions



A spherical representation of Internet (Shavitt et al.)

25 years of network revolutions



3D representation of the World-Wide Web
(www.vlib.us/web/worldwideweb3d.html)

Next network revolution?

Next network revolution?

Network functions

Internet: e-mail

WWW: Information gathering

Friendship networks: gossiping, spread of information and disease

Network functions

Internet: e-mail

Routing on networks

WWW: Information gathering

Searching on networks

Friendship networks: gossiping, spread of information and disease

Spread of diseases and motion on networks

Network functions

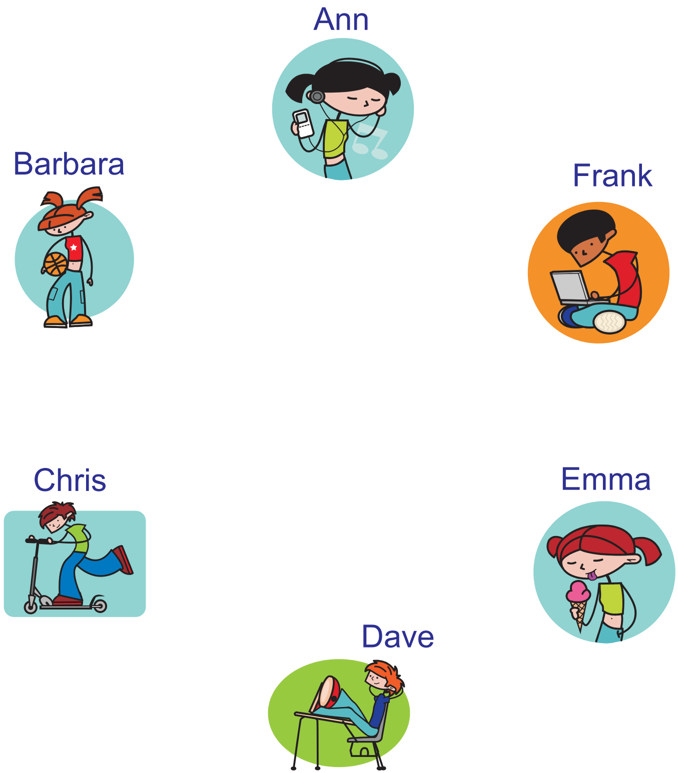
Internet: e-mail
Routing on networks

WWW: Information gathering
Searching on networks

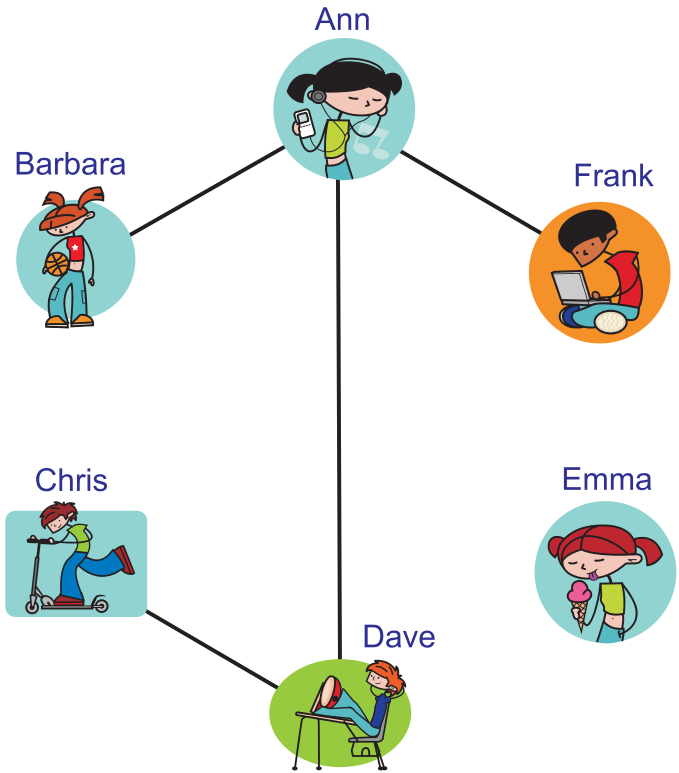
Friendship networks: gossiping, spread of information and disease
Spread of diseases and motion on networks

Fundamental understanding of networks still very limited!

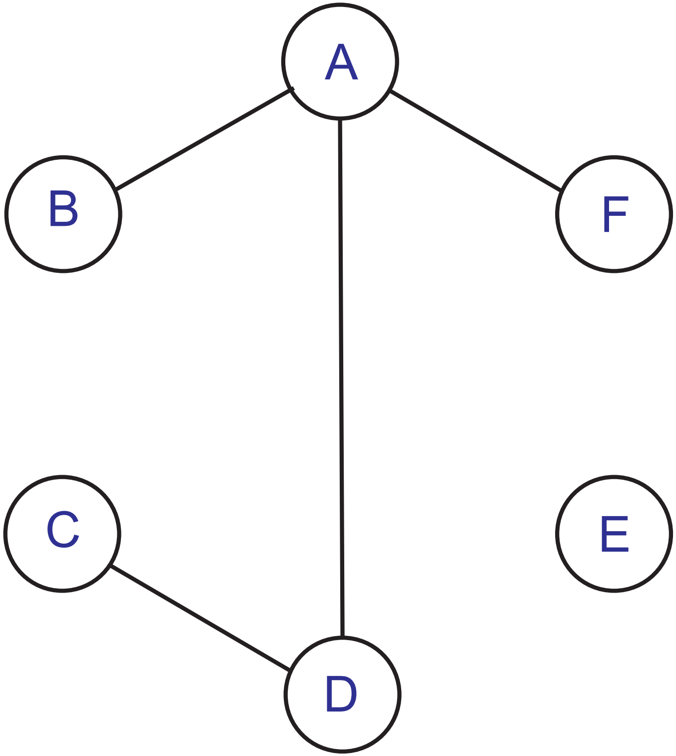
Networks and Graphs



Networks and Graphs



Networks and Graphs



Complex networks

Burst of activity in past decade.
See reviews Newman (2003) and Albert-Barabási (2002) for examples.

Networks come in different flavors:

- Social networks: Acquaintances, sexual relations, film actors,...
- Information networks: Collaboration graphs, WWW,...
- Technological networks: Internet, power/telephone grids,...
- Biological networks: Food webs, neurons, protein interactions,...

Attention focussing on unexpected commonality.

Structure affects function

- Structure affects function

Structure social relations affects spread of disease.

- Topology due to underlying mechanism

Description of topology sheds light on underlying mechanism (e.g., biological and social networks).

- Analysis of data bases

Internet has made data bases readily available, fast computers have made their analysis possible.

- Testing ground for network efficiency

Testing new Internet protocols should be done on good Internet models. Dito for testing ground of WWW search engines.

Internet and WWW

• Vertices Internet are Internet routers, edges physical links between routers. Edges have weights. E-mail is sent along path with smallest sum of weights.

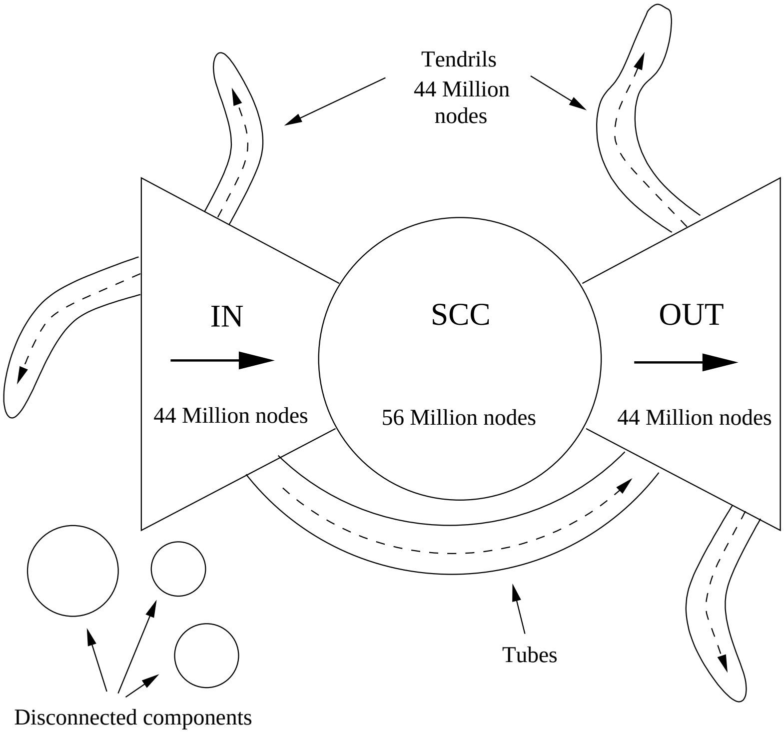
• Vertices WWW are web pages, edges are (directed) hyper links between pages.

Properties:

• Internet is large, chaotic, (fairly) homogeneous, connected,...
• WWW is huge, directed, hard to measure, non-connected,...

• Levels Internet: Router (IP) and autonomous system (AS) levels.
• Decomposition of WWW due to Broder et al.

The Web



The WWW according to Broder et al.

Collaboration graph in mathematics

• Vertices are mathematicians, two mathematicians are connected by an edge when they have co-authored a paper.

• Size: $\pm$ 401.000 mathematicians, $\pm$ 270.000 in one component.
• Edges: $\pm$ 676.000, so that average degree: $\pm$ 3.4.

`http://www.oakland.edu/enp`
`http://www.ams.org/mathscinet`

Six degrees of separation

"Everybody on this planet is separated only by six other people. Six degrees of separation. Between us and everybody else on this planet. The president of the United States. A gondolier in Venice… It's not just the big names. It's anyone. A native in the rain forest. (…) An Eskimo. I am bound to everyone on this planet by a trail of six people. It is a profound thought."

Six degrees of separation

In 1967, Milgram performed an experiment, where people in the Mid-West were asked to send a package to a person in Massachusetts, via intermediaries they know on first name basis.

Experimental fact: (almost always) at most six intermediaries:

Six degrees of separation.

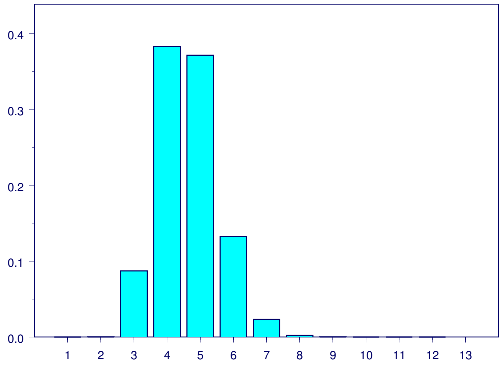
Experiment received enormous attention, but was later heavily criticized due to the small data set.

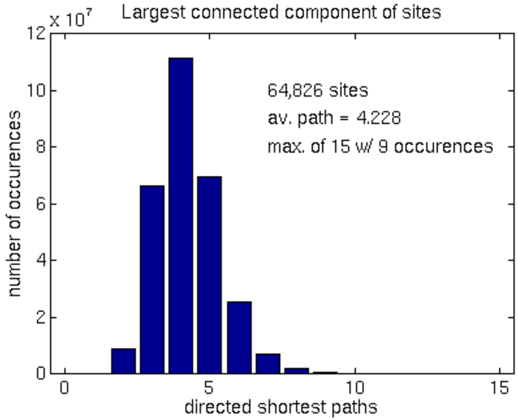
In 2000, Duncan Watts performed an e-mail version of the experiment on a huge scale, with roughly the same outcome:

Small-world phenomenon.

`http://smallworld.columbia.edu`

Small-world phenomenon





Distances in AS graph and WWW (Adamic 99)

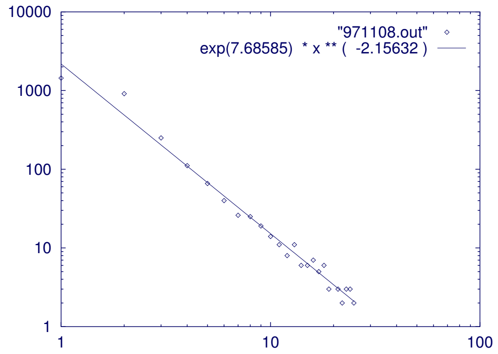
Power-law degree sequences

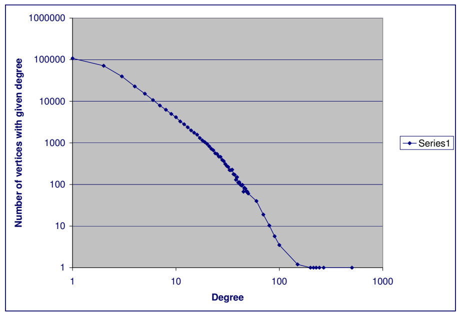
Degree sequence $(N_1, N_2, N_3, \ldots)$ of graph:

N_1 is number of elements with degree equal to 1,
N_2 is number of elements with degree equal to 2,
N_3 is number of elements with degree equal to 3...

$$N_k \approx Ck^{-\tau}.$$

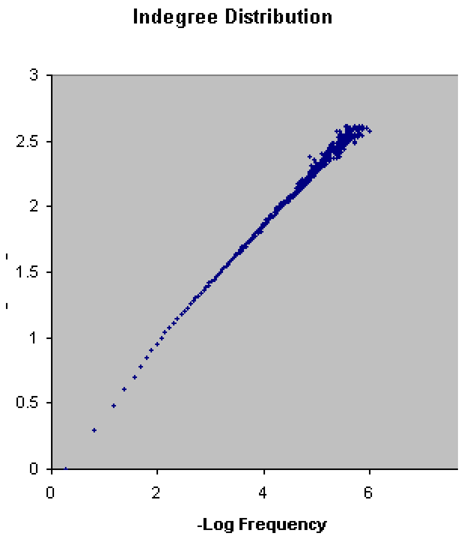
Scale-free phenomenon





Loglog plot of degree sequences in AS graph in Internet in 1997 (FFF97) and in the collaboration graph among mathematicians (`http://www.oakland.edu/enp`)

Scale-free nature in-degrees Web



Power-law in-degrees in the WWW (Kumar et al).

Modeling complex networks

- Inhomogeneous Random Graphs:
Static random graph, independent edges with inhomogeneous edge occupation probabilities, yielding scale-free graphs.
(BJR07, CL02, CL03, BDM-L05, CL06, NR06, EHH06,...)

- Configuration Model:
Static random graph with prescribed degree sequence.
(MR95, MR98, RN04, HHV05, EHHZ06, HHZ07, JL07, FR07,...)

- Preferential Attachment Model:
Dynamic random graph, attachment proportional to degree plus constant.
(BA99, BRST01, BR03, BR04, M05, B07, HH07,...)

Configuration model

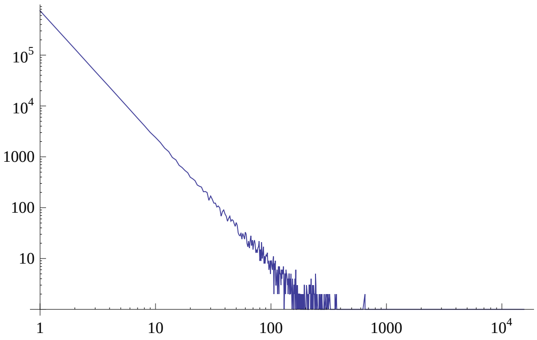
N is number of vertices. Consider i.i.d. sequence of degrees $D_1, D_2, \ldots, D_n$, with

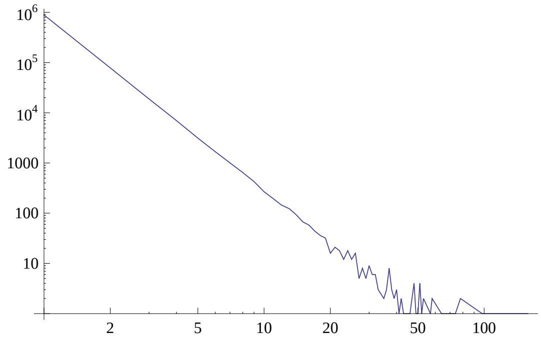
$$\mathbb{P}(D_1 \geq k) = c_\tau k^{-\tau+1}(1 + o(1)),$$

where c_τ is normalizing constant and $\tau > 1$.

Will later explain how edges are formed such that degrees are correct.

Scale-free nature CM





Loglog plot of degree sequence CM with $n = 1.000.000$ and $\tau = 2.5$ and $\tau = 3.5$, respectively.

Configuration model: graph construction

How to construct graph with above degree sequence?

- Assign to vertex j degree D_j.

$$L_n = \sum_{i=1}^{n} D_i$$

is total degree. Assume L_n is even.
Incident to vertex i have D_i 'stubs' or half edges.

Configuration model: graph construction

How to construct graph with above degree sequence?

• Assign to vertex j degree D_j.

$$L_n = \sum_{i=1}^{n} D_i$$

is total degree. Assume L_n is even.
Incident to vertex i have D_i 'stubs' or half edges.

• Connect stubs to create edges as follows:
Number stubs from 1 to L_n in any order.
First connect first stub at random with one of *other* $L_n - 1$ stubs.
Continue with second stub (when not connected to first) and so on, until all stubs are connected...

Properties configuration model

- Total number of edges is L_n, so average degree is $L_n/n \to \mathbb{E}[D]$.

- CM has giant component when

$$\nu = \frac{\mathbb{E}[D(D-1)]}{\mathbb{E}[D]} > 1.$$

Note that this is weaker than $\mathbb{E}[D] > 2$.

- Local neighborhood of vertex in CM is close to tree. Cycles occur, but are quite long.

Preferential attachment

In preferential attachment models, network is growing in time, in such a way that new vertices are more likely to be connected to vertices that already have high degree.

Preferential attachment

In preferential attachment models, network is growing in time, in such a way that new vertices are more likely to be connected to vertices that already have high degree.

• At time t, a single vertex is added to the graph with m edges emanating from it. Probability that an edge connects to the i^{th} vertex is proportional to

$$D_i(t-1)+\delta,$$

where $D_i(t)$ is degree vertex i at time t, $\delta > -m$ is parameter model.

Preferential attachment

In preferential attachment models, network is growing in time, in such a way that new vertices are more likely to be connected to vertices that already have high degree.

• At time t, a single vertex is added to the graph with m edges emanating from it. Probability that an edge connects to the i^{th} vertex is proportional to

$$D_i(t-1)+\delta,$$

where $D_i(t)$ is degree vertex i at time t, $\delta > -m$ is parameter model.

• Different edges can attach with different updating rules:
(a) intermediate updating degrees with self-loops (BA99, BR04, BRST01)
(b) intermediate updating degrees without self-loops;
(c) without intermediate updating degrees. i.e., independently.

Preferential attachment

Properties PA model

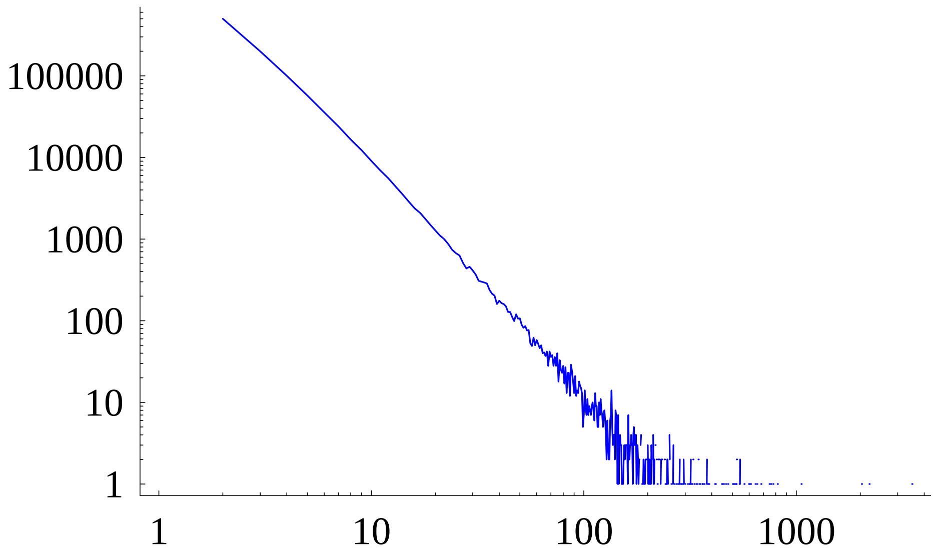
Total number of edges is $2mt$, so average degree is $2m$.

PA model is with high probability connected when $m \geq 2$.
When $m = 1$, have a tree.

Structure of graph is quite insensitive to precise construction graph:
- Attachment rules;
- Fixed number of edges per vertex or random.

Scale-free nature PA

Yields power-law degree sequence with exponent $\tau = 3 + \delta/m \in (2, \infty)$.



$$(m = 2, \delta = 0, \tau = 3 + \frac{\delta}{m} = 3)$$

"...the scale-free topology is evidence of organizing principles acting at each stage of the network formation. (...) No matter how large and complex a network becomes, as long as preferential attachment and growth are present it will maintain its hub-dominated scale-free topology."

Distances in configuration model

H_n is graph distance between uniform pair of connected vertices in graph of size n.

Distances in configuration model

H_n is graph distance between uniform pair of connected vertices in graph of size n.

- When $\tau > 3$ and $\nu > 1$ (HHV03)

$$H_n \approx \log_\nu n, \qquad \text{with} \qquad \nu = \frac{\mathbb{E}[D(D-1)]}{\mathbb{E}[D]}.$$

Distances in configuration model

H_n is graph distance between uniform pair of connected vertices in graph of size n.

- When $\tau > 3$ and $\nu > 1$ (HHV03)

$$H_n \approx \log_\nu n, \qquad \text{with} \qquad \nu = \frac{\mathbb{E}[D(D-1)]}{\mathbb{E}[D]}.$$

- When $\tau \in (2,3)$, (Norros+Reittu04, HHZ07)

$$H_n \approx 2\frac{\log\log n}{|\log(\tau-2)|}.$$

Distances in configuration model

H_n is graph distance between uniform pair of connected vertices in graph of size n.

- When $\tau > 3$ and $\nu > 1$ (HHV03)

$$H_n \approx \log_\nu n, \qquad \text{with} \qquad \nu = \frac{\mathbb{E}[D(D-1)]}{\mathbb{E}[D]}.$$

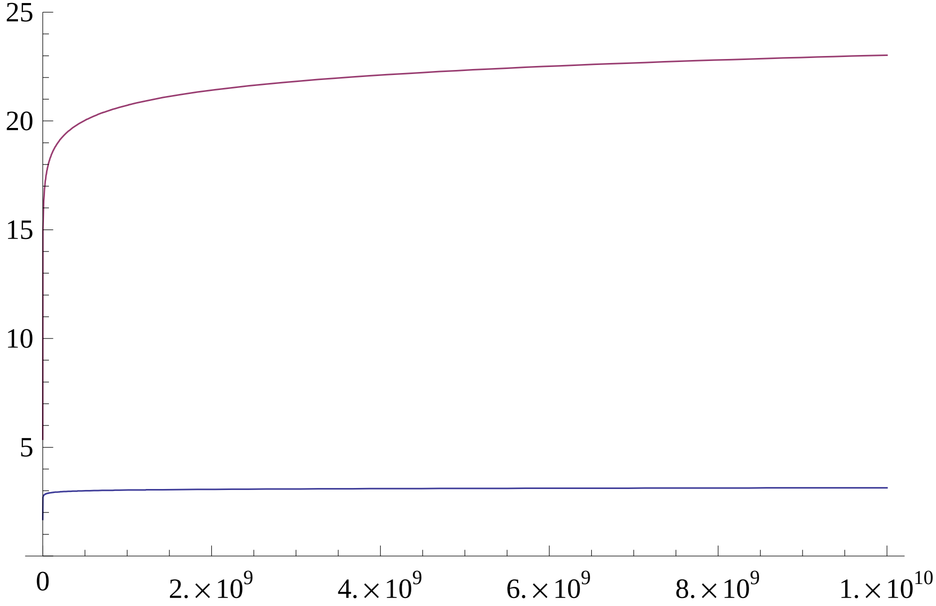
- When $\tau \in (2,3)$, (Norros+Reittu04, HHZ07)

$$H_n \approx 2\frac{\log\log n}{|\log(\tau-2)|}.$$

- When $\tau \in (1,2)$, (EHHZ06)

$$H_n \quad \text{uniformly bounded.}$$

$x \mapsto \log\log x$ grows extremely slowly



Plot of $x \mapsto \log x$ and $x \mapsto \log\log x$.

Consequences and proof

Proof relies on coupling of neighborhood of vertices to branching process.

Extensions:

• Fluctuations around leading order are uniformly bounded, and we compute its 'limiting distribution'.

• Diameter of graph is maximal distance between any pair of connected vertices.
Diameter CM of order $\log n$ when $\mathbb{P}(D_i \geq 3) < 1$ (FR07),
while of order $\log \log n$ when $\tau \in (2, 3)$ and $\mathbb{P}(D_i \geq 3) = 1$ (HHZ07).

Consequences and proof

Proof relies on coupling of neighborhood of vertices to branching process.

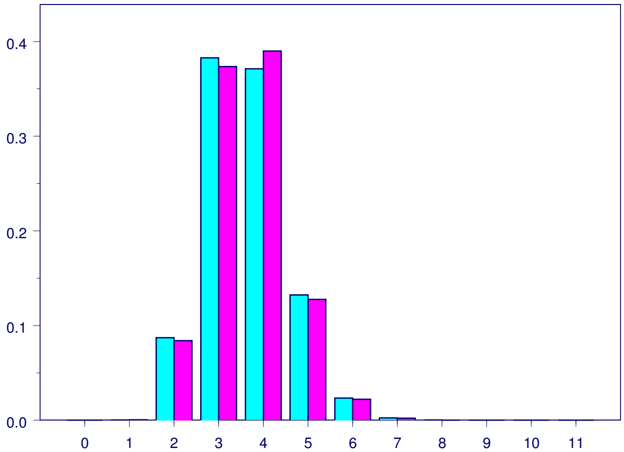
Extensions:

• Fluctuations around leading order are uniformly bounded, and we compute its 'limiting distribution'.

• Diameter of graph is maximal distance between any pair of connected vertices.
Diameter CM of order $\log n$ when $\mathbb{P}(D_i \geq 3) < 1$ (FR07),
while of order $\log \log n$ when $\tau \in (2, 3)$ and $\mathbb{P}(D_i \geq 3) = 1$ (HHZ07).

• Similar methodology applies to various inhomogeneous random graphs, with similar (but generally weaker) results.

Comparison Internet data



Number of AS traversed in hopcount data (blue) compared to the model (purple) with $\tau = 2.25, N = 10,940.$

Distances PA models of size n

- Diameter PA bounded above by $C\log n$ for all $m \geq 1$ and $\delta > -m$ (HH07);

- Diameter PA $\Theta(\log\log n)$ for $m \geq 2$ and $\delta \in (-m, 0)$ for which power-law exponent $\tau \in (2,3)$ (DHH09);

- Diameter PA bounded above by $(1+\varepsilon)\frac{\log t}{\log\log n}$ for $m \geq 2$ and $\delta = 0$ for which power-law exponent $\tau = 3$ (a) (BR04);

- Diameter PA bounded below by $(1+\varepsilon)\frac{\log t}{\log\log n}$ for $m \geq 2$ and $\delta = 0$ for which power-law exponent $\tau = 3$ (a-c) (BR04, DHH09).

- Diameter PA bounded below by $\varepsilon\log n$ for $m \geq 2$ and $\delta > 0$ for which power-law exponent $\tau \in (3, \infty)$ (a-c) (DHH09).

Universality PA models

First evidence of strong form of universality:
random graphs with similar degree structure share similar behavior.

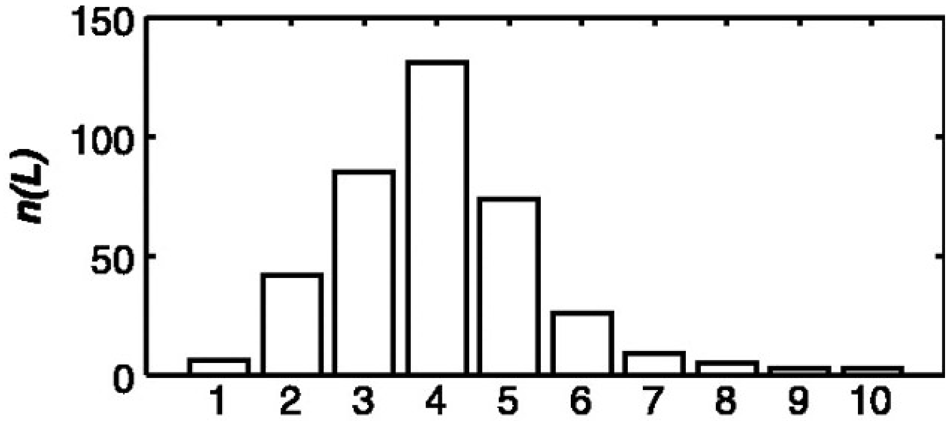
For random graphs, universality predicted by physics community.

Universality is leading paradigm in statistical physics.
Only few examples where universality can be rigorously proved.

More information on random graphs:

`www.win.tue.nl/~rhofstad/NotesRGCN.pdf`

Small-world phenomenon



Distances in social network (Small-World Project Watts (2003))