Universality of distances in random graphs

Remco van der Hofstad

TU/e technische universiteit eindhoven

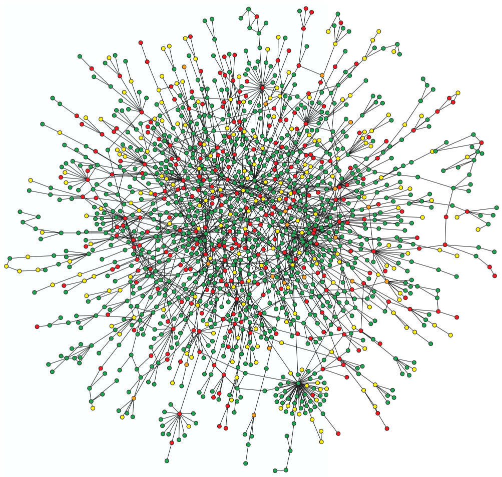
EURANDOM

Third La Pietra Week in Probability, June 23-27, 2008

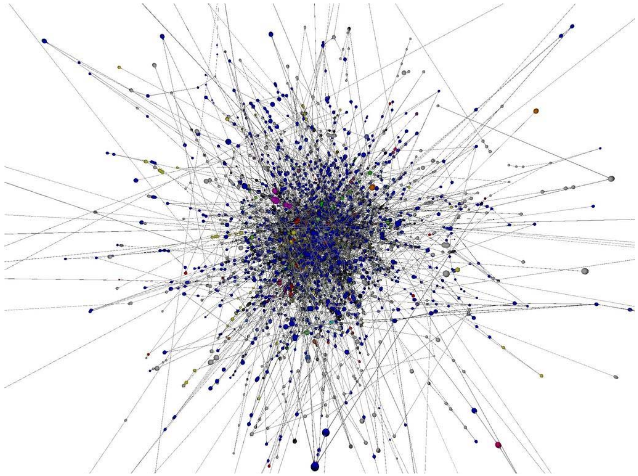
Joint work with:

- Gerard Hooghiemstra (TU Delft)
- Henri van den Esker (TU Delft)
- Piet Van Mieghem (TU Delft)
- Dmitri Znamenski (EURANDOM, now Philips Research)

Complex networks

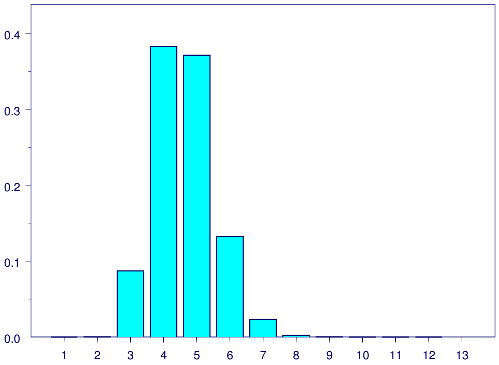


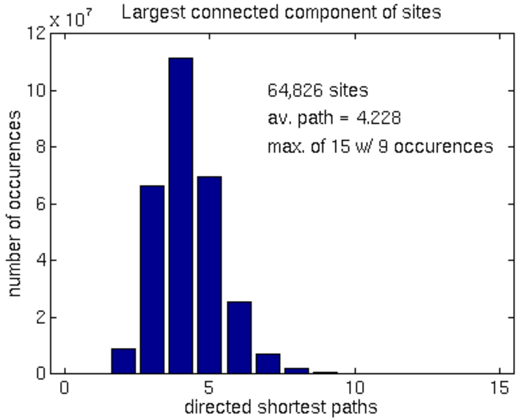
Yeast protein interaction network



Internet topology in 2001

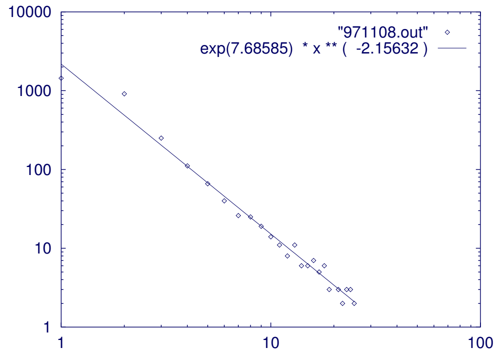
Small-world phenomenon

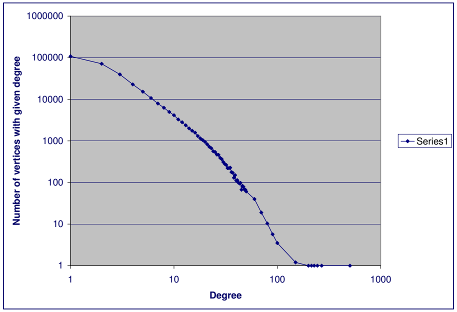




Distances in AS graph and WWW (Adamic 99)

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`)

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,...)

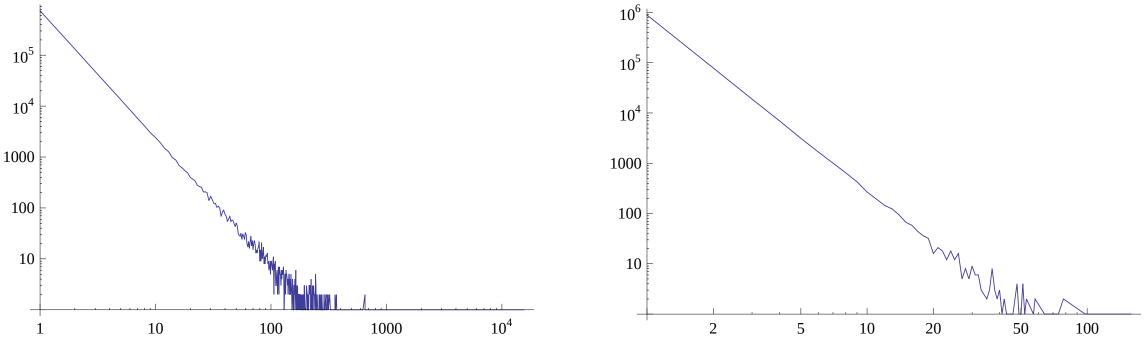
Configuration model

Let n be the 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 $\tau > 1$.

Power law degree sequence 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...

Distances in configuration model

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

Distances in configuration model

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

- 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.

- 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)

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

Distances in configuration model

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

- 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)

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

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

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

Consequences and proof

Proof relies on
(a) extreme value theory when $\tau \in [1, 2)$;
(b) coupling of neighborhood of vertices to branching process when $\tau > 2$,

When $\tau > 2$, CM is locally tree-like, and $\nu > 1$ is equivalent to branching process being supercritical, and giant component existing.

Consequences and proof

Proof relies on
(a) extreme value theory when $\tau \in [1, 2)$;
(b) coupling of neighborhood of vertices to branching process when $\tau > 2$,

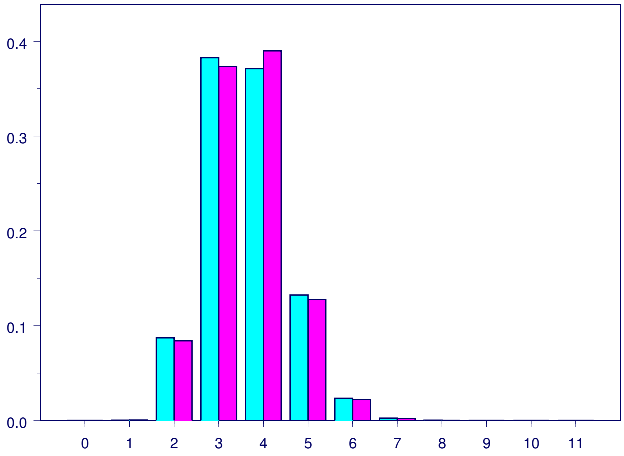
When $\tau > 2$, CM is locally tree-like, and $\nu > 1$ is equivalent to branching process being supercritical, and giant component existing.

Extensions:

• Fluctuations around leading order are uniformly bounded, and 'limiting distribution' computed in terms of martingale limit branching process.

• Diameter of graph is maximal distance between any pair of connected vertices.
Diameter CM is $\Theta(\log n)$ 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).

Comparison Internet data



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

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

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

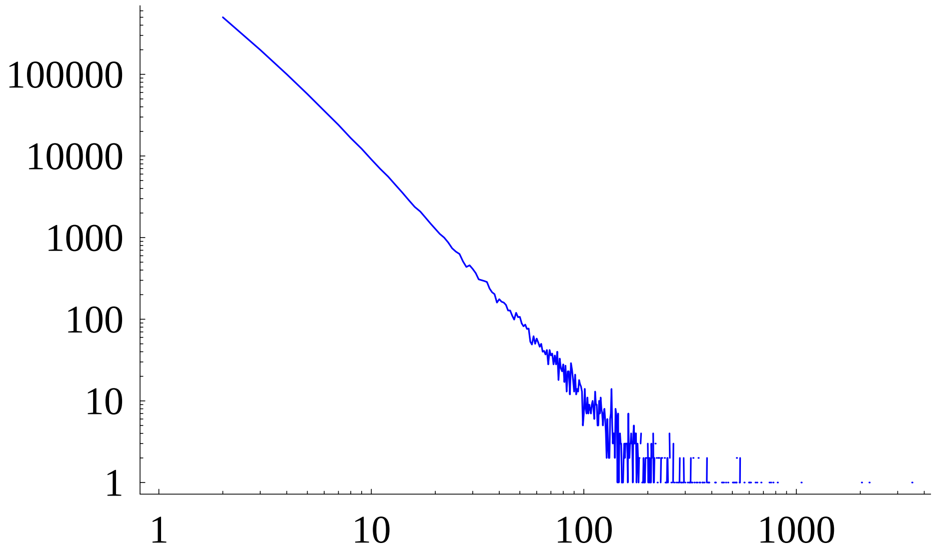
where $d_i(t)$ is degree vertex i at time t, $\delta > -m$ is parameter model.

• 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.

(Graphs in cases (b-c) have advantage of being connected.)

Scale-free nature PA

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



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

Distances PA models

Diam_n is diameter in PA model of size n. Then

- For all $m \geq 2$ and $\tau \in (3, \infty)$ (HH07)

$$\mathrm{Diam}_n = \Theta(\log n).$$

- For all $m \geq 2$ and $\tau = 3$ (BR04, HH07)

$$\mathrm{Diam}_n \geq \frac{\log n}{\log\log n},$$

while, for model (a), matching upper bound exists (BR04).

- For all $m \geq 2$ and $\tau \in (2, 3)$ (HH07)

$$\mathrm{Diam}_n \leq C \log\log n.$$

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...

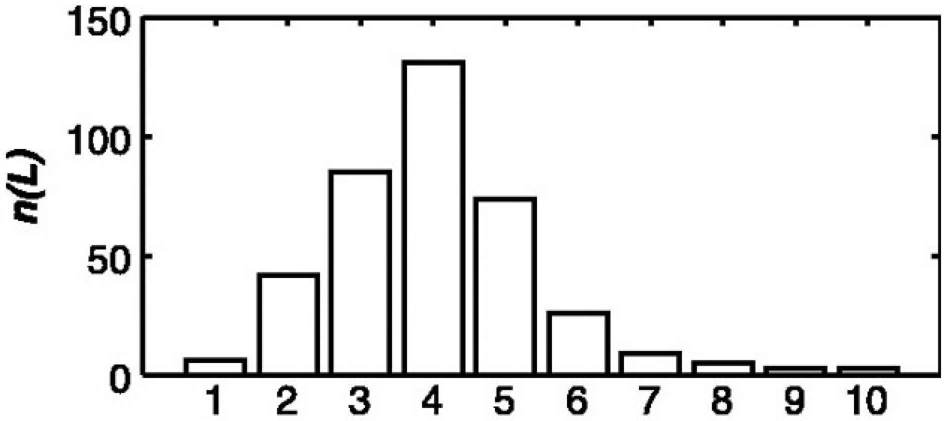
Universality is leading paradigm in statistical physics.
Only few examples where universality can be rigorously proved.

Key question: Can universality be proved for processes such as Ising model or contact process on random graphs?

More information on power-law and Erdős-Rényi random graphs:

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

Small-world phenomenon



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