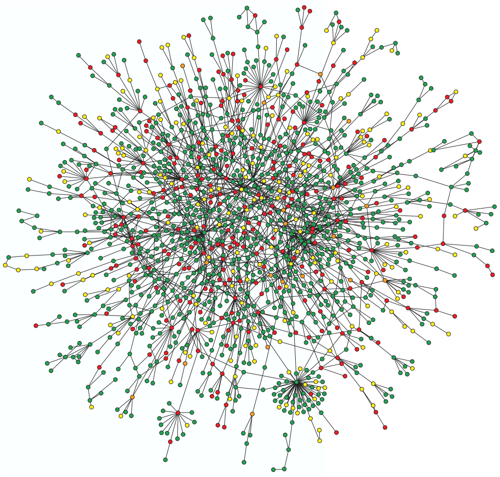
Random network models and routing on weighted networks

Remco van der Hofstad

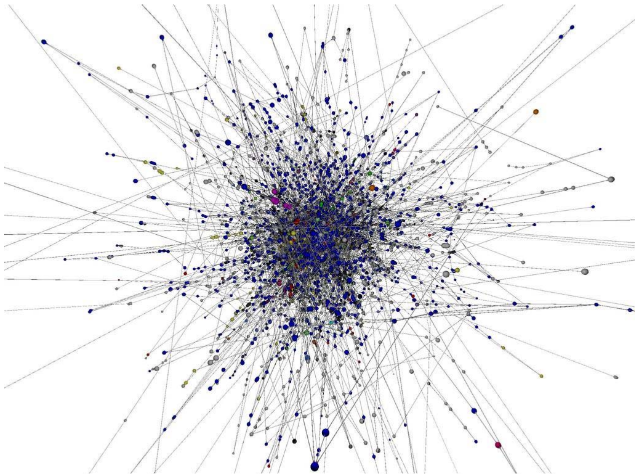
Mathematics and Neuroscience: A Dialogue,
September 3-5, 2013

Joint work with:
G. Hooghiemstra, P. Van Mieghem (Delft)
S. Bhamidi (North Carolina).

Complex networks

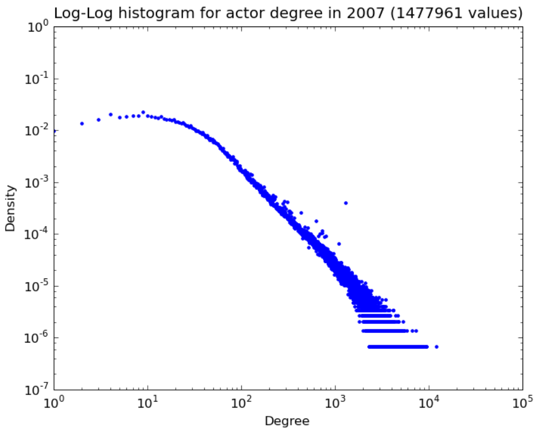


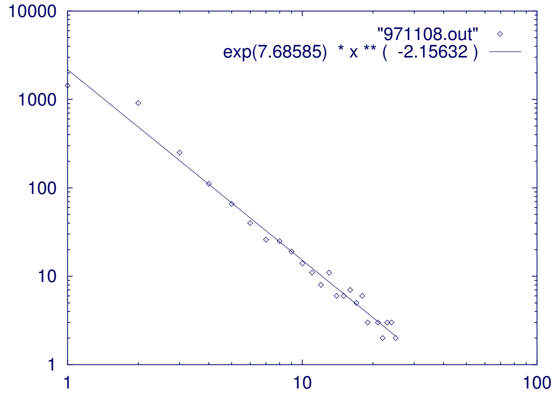
Yeast protein interaction network



Internet topology in 2001

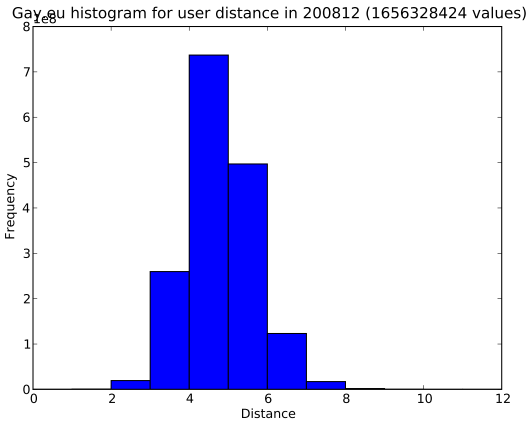
Scale-free paradigm

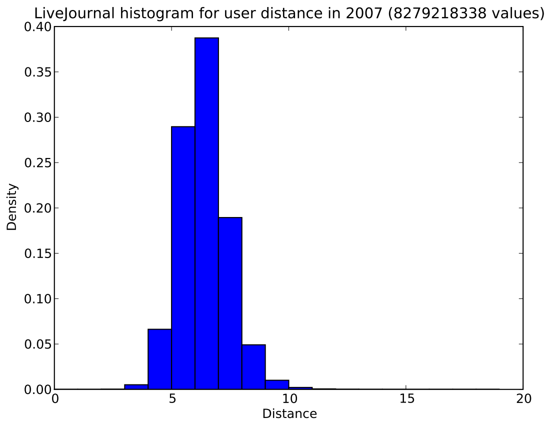




Loglog plot of degree sequences in Internet Movie Data Base (2007) and in the AS graph (FFF97)

Small-world paradigm





Distances in social networks `gay.eu` on December 2008 and `livejournal` in 2007.

Network statistics I

▷ Clustering:

$$C = \frac{3 \times \text{ number of triangles}}{\text{number of connected triplets}}.$$

Proportion of friends that are friends of one another.

▷ Assortativity:

$$\rho = \frac{\frac{1}{|E_n|}\sum_{ij\in E_n} d_i d_j - \left(\frac{1}{|E_n|}\sum_{ij\in E_n} d_i\right)^2}{\frac{1}{|E_n|}\sum_{ij\in E_n} d_i^2 - \left(\frac{1}{|E_n|}\sum_{ij\in E_n} d_i\right)^2}.$$

Correlation between degrees at either end of edge.

[Recent work vdH-Litvak (2013): assortivity coefficient flawed. Proposes rank correlations instead.]

Network statistics II

▷ Closeness centrality:

$$\ell_i = \frac{1}{n} \sum_{j \in [n]} \mathrm{dist}(i, j).$$

Vertices with low closeness centrality are central in network.

▷ Betweenness centrality:

$$b_i = \frac{1}{n^2} \sum_{s,t \in [n]} \frac{n_{st}^i}{n_{st}},$$

where n_{st} is number of shortest paths between s, t, and n_{st}^i is number of shortest paths between s, t that pass through i.
Betweenness large for bottlenecks.

Modeling complex networks

Use random graphs to model uncertainty in how

connections between elements are formed.

Two settings:

▷ Static models:
Graph has fixed number of elements.

▷ Dynamic models:
Graph has evolving number of elements.

Universality??

Configuration model

▷ Invented by Bollobás (1980), EJC: 441 cit. (19-5-2013)
to study number of graphs with given degree sequence.
Inspired by Bender+Canfield (1978), JCT(A): 493 cit. (19-5-2013)
Giant component: Molloy, Reed (1995), RSA: 1208 cit. (19-5-2013)
Popularized by Newman, Strogatz, Watts (2001), Psys. Rev. E: 2074 cit. (19-5-2013).

▷ n number of vertices;
▷ $\boldsymbol{d} = (d_1, d_2, \ldots, d_n)$ sequence of degrees.
Often take $(d_i)_{i\in[n]}$ to be sequence of independent and identically distributed (i.i.d.) random variables with certain distribution.

▷ Special attention to power-law degrees, i.e., for $\tau > 1$ and c_τ

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

Configuration model: graph construction

▷ Assign d_j half-edges to vertex j. Asume total degree

$$\ell_n = \sum_{i \in [n]} d_i$$

is even.

▷ Pair half-edges to create edges as follows:
Number half-edges from 1 to ℓ_n in any order.
First connect first half-edge at random with one of other $\ell_n - 1$ half-edges.

▷ Continue with second half-edge (when not connected to first) and so on, until all half-edges are connected.

▷ Resulting graph is denoted by $\mathrm{CM}_n(\boldsymbol{d})$.

Graph distances in CM

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

Theorem 1. (vdHHVM03). When $\nu = \mathbb{E}[D(D-1)]/\mathbb{E}[D] \in (1,\infty)$ and $\mathbb{E}[D_n^2] \to \mathbb{E}[D^2]$, conditionally on $H_n < \infty$,

$$\frac{H_n}{\log_\nu n} \xrightarrow{\mathbb{P}} 1.$$

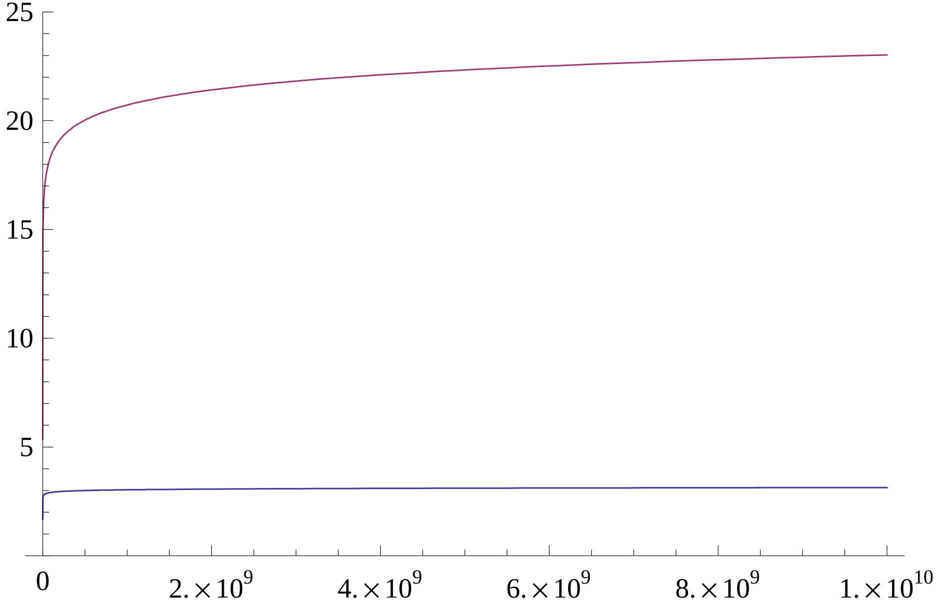
For i.i.d. degrees having power-law tails, fluctuations are bounded.

Theorem 2. (vdHHZ07, Norros+Reittu 04). When $\tau \in (2,3)$, conditionally on $H_n < \infty$,

$$\frac{H_n}{\log\log n} \xrightarrow{\mathbb{P}} \frac{2}{|\log(\tau-2)|}.$$

For i.i.d. degrees having power-law tails, fluctuations are bounded.

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



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

Preferential attachment models

Albert-Barabási (1999):
Emergence of scaling in random networks (Science).
16850 cit. (19-5-2013).
Bollobás, Riordan, Spencer, Tusnády (2001):
The degree sequence of a scale-free random graph process (RSA)
506 cit. (19-5-2013).

[In fact, Yule 25 and Simon 55 already introduced similar models.]

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.

Rich-get-richer model.

Preferential attachment models

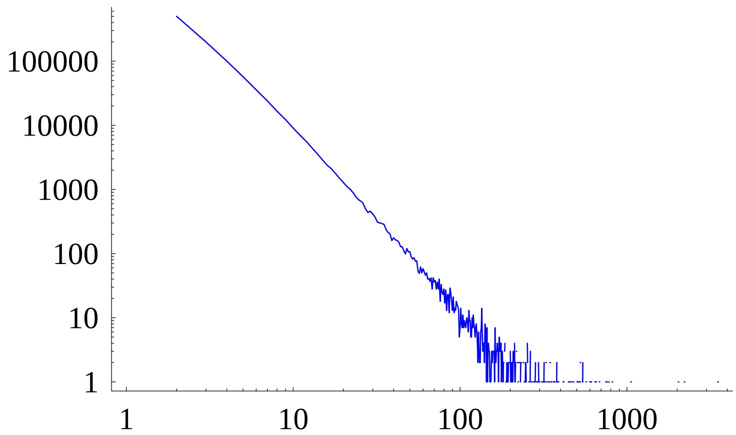
At time n, single vertex is added with m edges emanating from it. Probability that edge connects to i^{th} vertex is proportional to

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

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

Yields power-law degree sequence with exponent $\tau = 3 + \delta/m > 2$.

BRST01 $\delta = 0$,
DvdEvdHH09,...



$m = 2, \delta = 0, \tau = 3, n = 10^6$

Distances PA models

Theorem 3 (Bol-Rio 04). For all $m \geq 2$ and $\tau = 3$,

$$H_n = \frac{\log n}{\log \log n}(1 + o_{\mathbb{P}}(1)).$$

Theorem 4 (Dommers-vdH-Hoo 10). For all $m \geq 2$ and $\tau \in (3, \infty)$,

$$H_n = \Theta(\log n).$$

Theorem 5 (Dommers-vdH-Hoo 10, DerMonMor 11). For all $m \geq 2$ and $\tau \in (2, 3)$,

$$\frac{H_n}{\log \log n} \xrightarrow{\mathbb{P}} \frac{4}{|\log (\tau - 2)|}.$$

Network modeling mayhem

Models:

- ▷ Configuration Model
- ▷ Inhomogeneous Random Graphs
- ▷ Preferential Attachment Model

What is bad about these models?

- ▷ Low clustering and few short cycles (unlike social networks);
- ▷ No communities (unlike collaboration networks and WWW);
- ▷ No attributes (geometry, gender,...);

Models are caricature of reality!

Network models I

▷ Small-world model:
Start with d-dimensional torus (=circle $d = 1,$ donut $d = 2,$ etc).
Put in nearest-neighbor edges. Add few edges between uniform vertices, either by rewiring or by simply adding.

Result: Spatial random graph with high clustering, but degree distribution with thin tails.

Application: None?

▷ Configuration model with clustering:
Input per vertex i is number of simple edges, number of triangles, number of squares, etc. Then connect uniformly at random.

Result: Random graph with (roughly) specified degree, triangle, square, etc distribution over graph.

Application: Social networks?

Network models II

▷ Random intersection graph:
Specify collection of groups. Vertices choose group memberships. Put edge between any pairs of vertices in same group.

Result: Flexible collection of random graphs, with high clustering, communities by groups, tunable degree distribution.

Application: Collaboration graphs?

▷ Spatial preferential attachment model:
First give vertex uniform location. Let it connect to close by vertices with probability proportionally to degree.

Result: Spatial random graph with scale-free degrees and high clustering.

Application: Social networks, WWW?

Network models III

$\triangleright$ Scale-free percolation:
Vertex set $\mathbb{Z}^d$. Each vertex x has a weight W_x, which form a collection of independent and identically distributed random variables.

Put edge between x and y with probability, conditionally on weights, equal to

$$p_{xy} = 1 - \mathrm{e}^{-W_x W_y / \|x-y\|^\alpha},$$

where $\alpha > 0$ is parameter model.

Result: Spatial random graph with scale-free degrees when weights obey power-law, high clustering and small-world.

Application: Social networks, WWW, brain?

Distances other models

Similar results (though often weaker) proved for related models:

▷ Random intersection graphs;

▷ Small-world model;

▷ Scale-free percolation.

Full extent of universality paradigm still unclear.

Work in progress!

Weighted graphs

▷ In many applications, edge weights represent cost structure graph, such as economic or congestion costs across edges.

▷ Time delay experienced by vertices in network is given by hopcount H_n, which is number of edges on shortest-weight path.

How does weight structure influence hopcount and weight SWP?

▷ Assume that

edge weights are i.i.d. random variables.

Graph distances: weights = 1.

Setting

▷ Central objects of study:
$\mathcal{C}_n$ is smallest-weight two uniform connected vertices, i.e.,

$$\mathcal{C}_n = \min_{\pi : V_1 \to V_2} \sum_{e \in \pi} X_e,$$

where X_e is edge-weight of edge e, $V_1, V_2 \in [n]$ chosen uniformly.
Hopcount H_n is number of edges in smallest-weight path $|\pi^*|$, where π^* is unique minimizing path.

▷ Restrict ourselves to complete graph K_n or configuration model, weights are i.i.d. with continuous distribution.

▷ Problem on complete graph received tremendous attention in theoretical physics community in works by Havlin, Braunstein, Stanley, et al.

Weighted sparse random graph

H_n number of edges in shortest-weight path two uniform connected vertices, $\mathcal{C}_n$ its weight.

Theorem 6. (BvdHH 12). Let configuration model satisfy $D_n \xrightarrow{d} D$, and

$$\lim_{n\to\infty} \mathbb{E}[D_n^2 \log(D_n \vee 1)] = \mathbb{E}[D^2 \log(D \vee 1)].$$

Then, there exist $\alpha_n, \beta, \gamma_n > 0$ with $\alpha_n \to \alpha, \gamma_n \to \gamma$ s.t.

$$\frac{H_n - \alpha_n \log n}{\sqrt{\beta \log n}} \xrightarrow{d} Z, \qquad \mathcal{C}_n - \gamma_n \log n \xrightarrow{d} \mathcal{C}_\infty,$$

where Z is standard normal, $\mathcal{C}_\infty$ is some limiting random variable.

Weighted complete graphs

Consider complete graph $K_n = ([n], E_n)$ with edge weights E_e^s, where $(E_e)_{e \in E_n}$ are i.i.d. exponentials.
Janson (1999): Scaling weight, flooding, diameter for $s = 1$.

Theorem 7. (BvdH10). Let $\mathcal{C}_n$ and H_n be weight and number of edges of shortest path between two uniformly chosen vertices in K_n. Then, with

$$\lambda = \lambda(s) = \Gamma(1 + 1/s)^s,$$

there exists a limiting random variable $\mathcal{C}_\infty$, such that

$$\frac{H_n - s \log n}{\sqrt{s^2 \log n}} \xrightarrow{d} Z, \qquad \mathcal{C}_n - \frac{1}{\lambda} \log n \xrightarrow{d} \mathcal{C}_\infty,$$

where Z is standard normal.

Weights matter: $s < 0$

Not always CLT, even when weights have density:
Consider complete graph $K_n = ([n], \mathcal{E}_n)$ with edge weights E_e^s, where $(E_e)_{e\in\mathcal{E}_n}$ are i.i.d. exponentials and $s < 0$.

Theorem 8. (BvdHH10b). H_n converges in distribution. Limit is constant $k = k(s)$ for most s...

Minimal spanning tree

Recent interest in minimal spanning tree on complete graph:

Theorem 9. (AB-B-G13). Minimal spanning tree is no small-world:

$$H_n/n^{1/3} \xrightarrow{d} H_\infty.$$

MST on graph is closely related to critical percolation on graph.

Explains $n^{1/3}$ behavior as this also appears for critical Erdős-Rényi random graphs. Are such distances observed in brain networks?

▷ Clustering: Tree is poor network. For example, tree has zero clustering.

Networks of the brain

Several levels:

▷ Neuronal level: 10^{11} vertices of average degree 10^4;

▷ Functional level: much smaller, modular structure.

What is meaning network?

Features:

▷ Short time scales: stochastic process on network (non-linear?);

▷ Long time scales: network is changed by functionality brain (learning, pruning,...);

▷ Strong dependence between different regions network.

Big question:

What is a good network model for brain functionality?

Weighted brain graphs

▷ Brain: data has weight (e.g., correlation between data signals) between any pair of vertices. Yields weighted complete graph.

Big question:

How to obtain informative network data from collection of weights?

Thresholding?
Comparing networks with different average edge weights?
Union of smallest-weight paths?

▷ Weight distribution: Edge weights are likely dependent.
How robust are results to dependencies?

▷ Application to brain: Interpretation edge weights?
Negative edge weights?