Processes on random graphs: routing and attack vulnerability

Remco van der Hofstad

TU/e technische universiteit eindhoven

EURANDOM

BMS Summer School 2011 at TU Berlin 'Random Motions and Random Graphs', September 26 - October 7, 2011

Exercise sessions by Robert Fitzner

Random graphs

Plan of the lectures:

Lecture 1:
Real-world networks and inhomogeneous random graphs

Lecture 2:
Inhomogeneous random graphs: small-world phenomenon

Lecture 3:
Configuration model and its properties

Lecture 4:
Real-world networks and routing on random graphs

Material

Lecture notes in preparation:

Random Graphs and Complex Networks

http://www.win.tue.nl/∼rhofstad/NotesRGCN.html

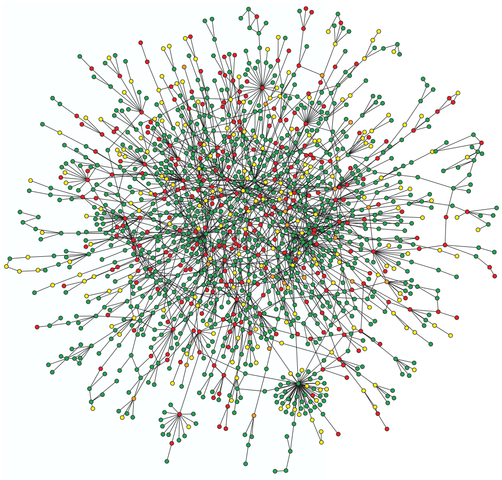
http://www.win.tue.nl/∼rhofstad/NotesRGCN.pdf

Material from Chapters 6-11.
Exercises taken from lecture notes.

Lecture 1:

Real-world networks and inhomogeneous random graphs

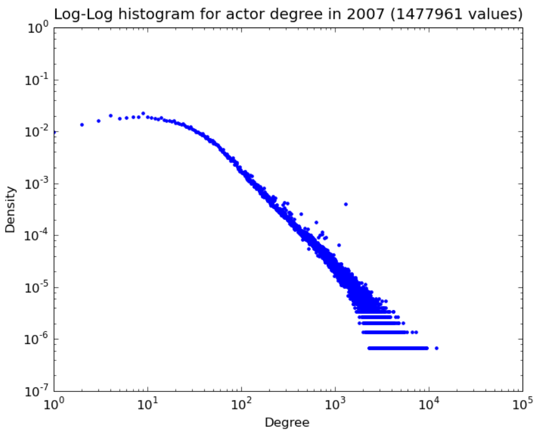
Complex networks

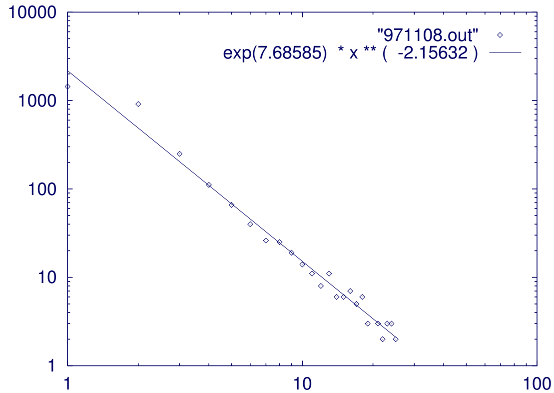


Yeast protein interaction network

Internet topology in 2001

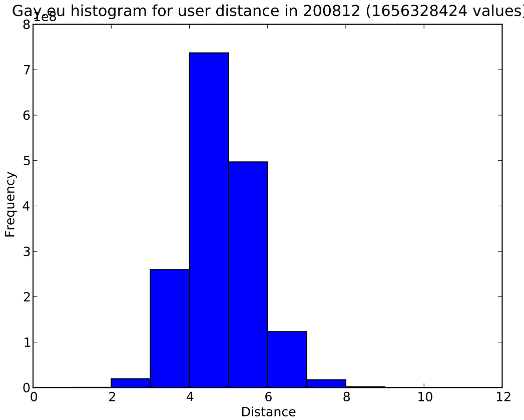
Scale-free paradigm

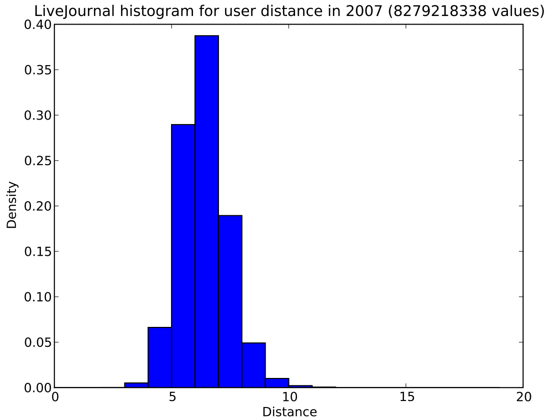




Loglog plot of degree sequences in Internet Movie Data Base (2007) and in the AS graph (FFF97)

Small-world paradigm





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

Modeling real networks

- Inhomogeneous Random Graphs:
Static random graph, independent edges with inhomogeneous edge occupation probabilities, yielding scale-free graphs.

- Configuration Model:
Static random graph with prescribed degree sequence.

- Preferential Attachment Model:
Dynamic random graph, attachment proportional to degree plus constant.

Modeling real networks

- Inhomogeneous Random Graphs:
Static random graph, independent edges with inhomogeneous edge occupation probabilities, yielding scale-free graphs.

- Configuration Model:
Static random graph with prescribed degree sequence.

- Preferential Attachment Model:
Dynamic random graph, attachment proportional to degree plus constant.

Universality??

Erdős-Rényi random graph

Vertex set $[n] := \{1, 2, \ldots, n\}$.

Erdős-Rényi random graph is random subgraph of complete graph on $[n]$ where each of $\binom{n}{2}$ edges is occupied with probab. p.

Simplest imaginable model of a random graph.

• Attracted tremendous attention since introduction 1959, mainly in combinatorics community.

Probabilistic method (Erdős et al).

Egalitarian: Every vertex has equal probability of being connected to. Misses hub-like structure of real networks.

Rank-1 inhomogeneous random graphs

Attach edge with probability p_{ij} between vertices i and j, where

$$p_{ij} = \frac{w_i w_j}{\ell_n + w_i w_j},$$

and

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

and different edges are independent.
Interpretation: w_i is close to expected degree vertex i.

When $w_i = n\lambda/(n-\lambda)$, we retrieve Erdős-Rényi random graph with $p = \lambda/n$.

Condition 6.3: Regularity vertex weights

Denote empirical distribution function weight by

$$F_n(x) = \frac{1}{n} \sum_{i \in [n]} \mathbb{1}_{\{w_i \le x\}}, \qquad x \ge 0.$$

(a) Weak convergence of vertex weight. There exists F s.t.

$$W_n \xrightarrow{d} W,$$

where W_n and W have distribution functions F_n and F.
(b) Convergence of average vertex weight.

$$\lim_{n\to\infty} \mathbb{E}[W_n] = \mathbb{E}[W] > 0.$$

(c) Convergence of second moment vertex weight.

$$\lim_{n\to\infty} \mathbb{E}[W_n^2] = \mathbb{E}[W^2].$$

Canonical choices of weights

(a) Take $\boldsymbol{w} = (w_1, \ldots, w_n)$ as i.i.d. random variables with distribution function F.

(b) Take $\boldsymbol{w} = (w_1, \ldots, w_n)$ as

$$w_i = [1 - F]^{-1}(i/n).$$

Interpretation: proportion of vertices i with $w_i \leq x$ is close to $F(x)$.[†]

Power-law example:

$$F(x) = \begin{cases} 0 & \text{for } x < a, \\ 1 - (a/x)^{\tau - 1} & \text{for } x \geq a, \end{cases}$$

in which case

$$[1 - F]^{-1}(u) = a(1/u)^{-1/(\tau-1)}, \qquad \text{so that} \qquad w_j = a(n/j)^{1/(\tau-1)}.$$

Degree structure graph

Denote proportion of vertices with degree k by

$$P_k^{(n)} = \frac{1}{n} \sum_{i \in [n]} \mathbb{1}_{\{D_i = k\}},$$

where D_i denotes degree of vertex i.

Let Condition 6.3(a) hold. Then, model is sparse, i.e., there exists probability distribution $(p_k)_{k=0}^{\infty}$ s.t.

$$P_k^{(n)} \xrightarrow{\mathbb{P}} p_k \qquad \text{where} \qquad p_k = \mathbb{E}\Big[\mathrm{e}^{-W} \frac{W^k}{k!}\Big],$$

where $W \sim F$.

In particular, $\sum_{l \geq k} p_k \sim ck^{-(\tau-1)}$ iff $\mathbb{P}(W \geq k) \sim ck^{-(\tau-1)}$.†

Critical value IRG

Bollobás-Janson-Riordan (07), Chung-Lu (02), [Theorem 9.2]:
Let $W \sim F$, then

- largest component $\sim \zeta n$ with $\zeta \in (0,1)$ for $\nu = \mathbb{E}[W^2]/\mathbb{E}[W] > 1$;
- largest component $o(n)$ for $\nu = \mathbb{E}[W^2]/\mathbb{E}[W] \leq 1$.

Identifies critical value IRG as

$$\nu = \mathbb{E}[W^2]/\mathbb{E}[W] = 1,$$

where ν is asymptotic expected number of forward neighbors, and W is asymptotic weight of uniform vertex.

Robustness of networks

Random attack: Remove vertices uniformly at random with probability p. Obtain rank-1 IRG where now probability of edge ij between kept vertices equals

$$\frac{w_i w_j}{\ell_n + w_i w_j},$$

and otherwise equals 0.

Giant component exists whenever

$$(1-p)\nu > 1.$$

In particular, when $\nu = \infty$, always giant component:

Robust to random failure.

Robustness of networks

Deliberate attack: Remove proportion p of vertices with highest weight. Obtain rank-1 IRG where probability of edge ij for $i, j > np$ equals

$$\frac{w_i w_j}{\ell_n + w_i w_j},$$

while otherwise probability equals 0.

Thus, giant component exists whenever

$$\frac{\sum_{i>np} w_i^2}{\ell_n} > 1.$$

In particular, even when $\nu = \infty$, for p large, no giant component:

Fragile to deliberate attacks.

Exercises Monday afternoon

Exercises:

6.1, 6.2, 6.3, 6.8, 6.15;

9.1, 9.2, 9.3, 9.4, 9.5, 9.6, 9.7, 9.8, 9.9, 9.11, 9.12, 9.13.

[Numbers refer to October 3, 2011 version of the lecture notes.]

Lecture 2:

Inhomogeneous random graphs: small-world phenomenon and its proofs

Graph distance inhomogeneous random graphs

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

Graph distance inhomogeneous random graphs

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

Theorem 1. [Theorem 9.3] When $\nu > 1$ and Condition 6.3(a-c) holds, conditionally on $H_n < \infty$,

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

Under stronger conditions, fluctuations are bounded (vdEvdHH08).

Graph distance inhomogeneous random graphs

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

Theorem 1. [Theorem 9.3] When $\nu > 1$ and Condition 6.3(a-c) holds, conditionally on $H_n < \infty$,

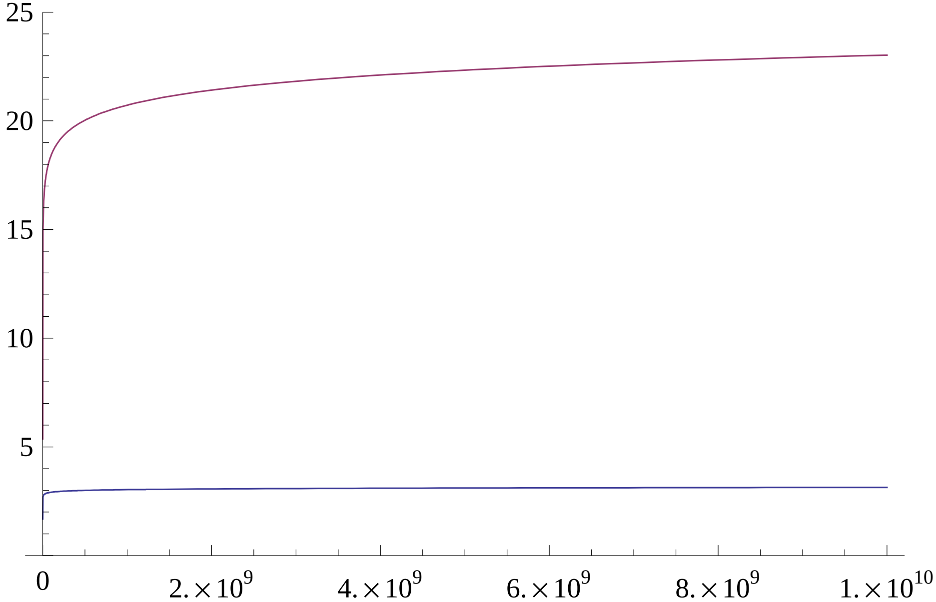
$$\frac{H_n}{\log_\nu n} \xrightarrow{\mathbb{P}} 1.$$

Under stronger conditions, fluctuations are bounded (vdEvdHH08)

Theorem 2. [Theorem 9.4] (CL03, Norros+Reittu 06). When $\tau \in (2,3)$, and Condition 6.3(a-b) hold, under certain further conditions on F_n, and conditionally on $H_n < \infty$,

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

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



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

Digression 1: $\tau = 3$

Theorem 3. [Theorem 9.22]. When $\tau = 3$ in the form

$$w_i = c(n/i)^{1/2},$$

then, conditionally on $H_n < \infty$,

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

Lecture 3:

Configuration model and its properties

Configuration model

Invented by Bollobás (1980), EJC: 285 cit. to study
number of graphs with given degree sequence.
Inspired by Bender+Canfield (1978), JCT(A): 300 cit.
Giant component studied by Molloy, Reed (1995), RSA: 664 cit.
Popularized by Newman, Strogatz, Watts (2001), Psys. Rev. E: 1190 cit.

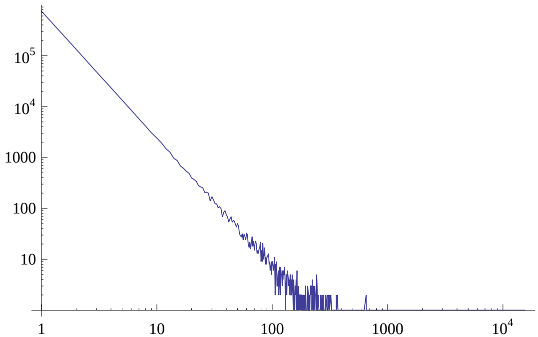
Let n be number of vertices and $d_1, d_2, \ldots, d_n$ sequence of degrees. Often take $(d_i)_{i\in[n]}$ to be sequence of independent and identically distributed (i.i.d.) random variables with certain distribution.

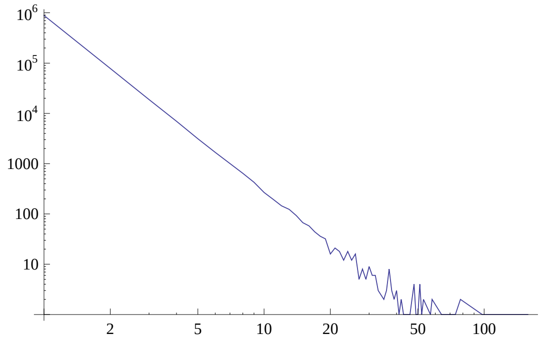
Special attention for power-law degrees, i.e., when

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

where c_τ is constant and $\tau > 1$.

Power-law degree sequence CM





Loglog plot of degree sequence CM with i.i.d. degrees $n = 1,000,000$ 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.

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

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.

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

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* $\ell_n - 1$ stubs.
Continue with second stub (when not connected to first) and so on, until all stubs are connected...

Condition 7.2: Regularity vertex degrees

Denote empirical distribution function degrees by

$$F_n(x) = \frac{1}{n} \sum_{i \in [n]} \mathbb{1}_{\{d_i \leq x\}}, \qquad x \geq 0.$$

(a) Weak convergence of vertex degrees. There exists F s.t.

$$D_n \xrightarrow{d} D,$$

where D_n and D have distribution functions F_n and F.
(b) Convergence of average vertex weight.

$$\lim_{n \to \infty} \mathbb{E}[D_n] = \mathbb{E}[D] > 0.$$

(c) Convergence of second moment vertex degrees.

$$\lim_{n \to \infty} \mathbb{E}[D_n^2] = \mathbb{E}[D^2].$$

Properties configuration model

CM can have cycles and multiple edges, but these are relatively scarce compared to the number of edges. [Theorem 7.6]

Let D_n denote degree of uniformly chosen vertex. Condition 7.2(a): D_n converges in distribution to limiting random variable D.

[Theorem 7.8, Prop. 7.9] When $\mathbb{E}[D_n^2] \to \mathbb{E}[D^2] < \infty$, then numbers of self-loops and multiple edges converge in distribution to two independent Poisson variables with parameters $\nu/2$ and $\nu^2/4$, respectively, where

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

Configuration model (CM) is locally tree-like.

Properties configuration model (Cont.)

Can interpret parameter ν as mean of size-biased distribution of D minus one.

This distribution is asymptotic distribution of forward degree of neighbor of uniformly chosen vertex.

$\nu > 1$ is equivalent to branching process approximation of connected components being supercritical, and giant component existing.

Graph distances in configuration model

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

Graph distances in configuration model

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

Theorem 4. (vdHHVM03). When $\tau > 3$ and $\nu > 1$, 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.

Graph distances in configuration model

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

Theorem 4. (vdHHVM03). When $\tau > 3$ and $\nu > 1$, 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 5. (vdHHZ07, Norros+Reittu 04). When $\tau \in (2,3)$, conditionally on $H_n < \infty$,

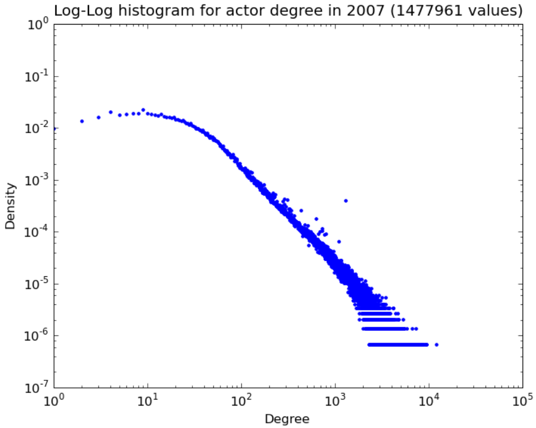
$$\frac{H_n}{\log\log n} \xrightarrow{\mathbb{P}} \frac{2}{|\log(\tau-2)|}.$$

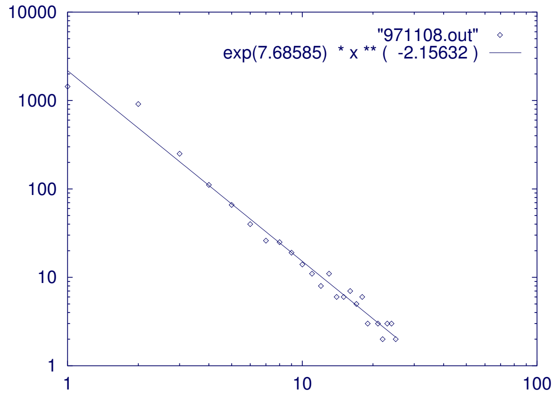
For i.i.d. degrees having power-law tails, fluctuations are bounded.

Lecture 4:

Real-world networks and routing on random graphs

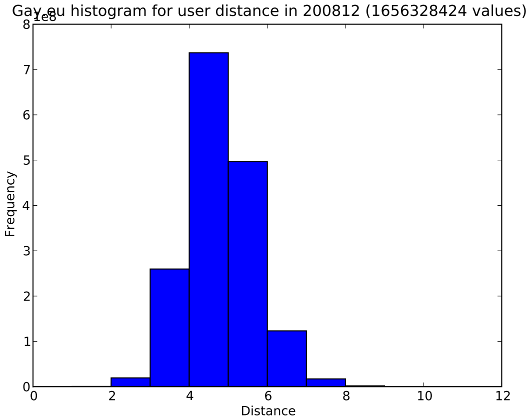
Scale-free paradigm

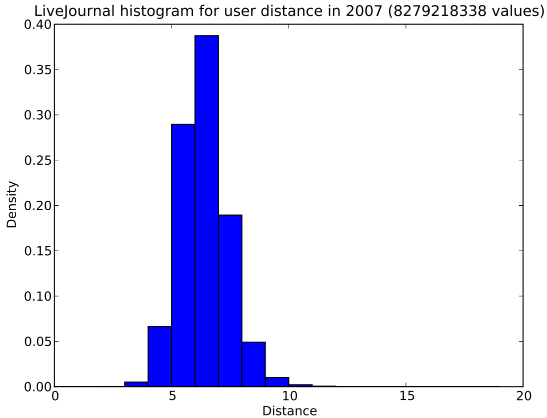




Loglog plot of degree sequences in Internet Movie Data Base (2007) and in the AS graph (FFF97)

Small-world paradigm





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

Network functions

Internet: e-mail

WWW: Information gathering

Friendship networks: gossiping, spread of information and disease

Power grids: reliability

Network functions

Internet: e-mail

Routing on networks, congestion, network failure

WWW: Information gathering

Crawling networks, motion on networks

Friendship networks: gossiping, spread of information and disease

Spread of diseases, motion on networks, consensus reaching

Power grids: reliability

Robustness to (random and deliberate) attacks

Network functions

Internet: e-mail
Routing on networks, congestion, network failure

WWW: Information gathering
Crawling networks, motion on networks

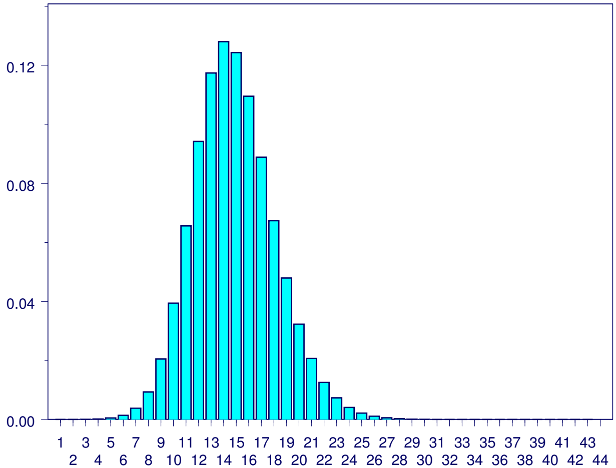
Friendship networks: gossiping, spread of information and disease
Spread of diseases, motion on networks, consensus reaching

Power grids: reliability
Robustness to (random and deliberate) attacks

Processes on networks!

Movies

Distances in IP graph



Poisson distribution??

Digression 2: Preferential attachment models

Albert-Barabási (1999):
Emergence of scaling in random networks (Science)
8737 citations on April 4, 2011.
Bollobas, Riordan, Spencer, Tusnády (2001):
The degree sequence of a scale-free random graph process (RSA)
371 citations in April 4, 2011.

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.

Digression 2: Preferential attachment models

At time n, 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

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

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

Digression 2: Preferential attachment models

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

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

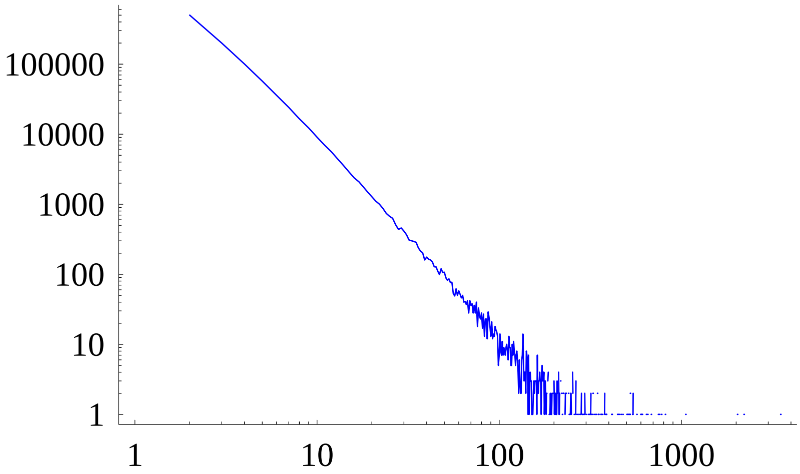
where $D_i(n)$ is degree vertex i at time n, $\delta > -m$ is parameter model.

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 $\tau = 3 + \delta/m \in (2, \infty)$.
(Bollobás, Riordan, Spencer, Tusnády (01) $\delta = 0$, Deijfen, vdE, vdH, Hoo (09),...) [Theorem 8.2]



$$(m = 2, \delta = 0, \tau = 3 + \frac{\delta}{m} = 3, n = 1,000,000)$$

Albert-László Barabási

"...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 PA models

Non-rigorous physics literature predicts that scaling distances in preferential attachment models similar to the one in configuration model with equal power-law exponent degrees.

Distances PA models

Diam_n is diameter in PA model of size n.

Theorem 6 (Dommers-vdH-Hoo 10). For all $m \geq 2$ and $\tau \in (3, \infty)$,

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

Theorem 7 (Dommers-vdH-Hoo 10, DerMonMor 11). For all $m \geq 2$ and $\tau \in (2, 3)$,

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

and

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

[Talk Christian Mönch.]

Distances PA models

Theorem 8 (Bol-Rio 04). For all $m \geq 2$ and $\tau = 3$, for model (a),

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

Universality!

Shortest-weight problems

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:
Aldous' stochastic mean-field model of distance.

Problem with exponential edge weights received tremendous attention on complete graph, here extend to random graphs.
Graph distances: weights = 1.

Results

Theorem 9. (BvdHH10). Let H_n be number of edges between two uniformly chosen vertices on CM with i.i.d. exponential edge weights.

Assume $D \geq 2$ a.s. and $\nu = \frac{\mathbb{E}[D(D-1)]}{\mathbb{E}[D]} > 1$.
For $\tau > 3$ or $\tau \in (2,3)$,

$$\frac{H_n - \alpha \log n}{\sqrt{\alpha \log n}} \xrightarrow{d} Z,$$

where Z is standard normal, and

$$\alpha = \frac{\nu}{\nu - 1} > 1 \qquad \text{for} \quad \tau > 3,$$
$$\alpha = \frac{2(\tau - 2)}{\tau - 1} \in (0,1) \qquad \text{for} \quad \tau \in (2,3).$$

Results

Theorem 10. (BvdHH10). Let $\mathcal{C}_n$ be weight of shortest path between two uniformly chosen vertices on CM with i.i.d. exponential edge weights.
Assume $D \geq 2$ a.s. and $\nu = \frac{\mathbb{E}[D(D-1)]}{\mathbb{E}[D]} > 1.$

Then, for some limiting random variable $\mathcal{C}_\infty$, and for $\tau > 3$ or $\tau \in (2,3)$,

$$\mathcal{C}_n - \gamma \log n \xrightarrow{d} \mathcal{C}_\infty,$$

where

$$\begin{aligned} \gamma &= \frac{1}{\nu - 1} > 0 \qquad && \text{for} \quad \tau > 3, \\ \gamma &= 0 && \text{for} \quad \tau \in (2,3). \end{aligned}$$

Discussion Theorems 9-10

Random weights have marked effect on shortest-weight problem.

Proof Theorems 9-10: Comparison neighborhood uniform vertex to branching process, and use wealth of results on FPP on trees.

Surprisingly universal behavior for FPP on configuration model.
Universality is leading paradigm in statistical physics.
Only few examples where universality can be rigorously proved.
Extension to FPP on super-critical Erdős-Rényi random graph.

Key question:
To what extent is universality true for processes on random graphs models?

Cool application by Ding, Kim, Lubetzky, and Peres identifying distance between two random vertices in two-core of slightly supercritical ERRG.

Digression 3: FPP on complete graph

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.

Theorem 11. (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

$$\mathcal{C}_n - \frac{1}{\lambda} \log n \xrightarrow{d} \mathcal{C}_\infty,$$

while

$$\frac{H_n - s \log n}{\sqrt{s^2 \log n}} \xrightarrow{d} Z,$$

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

What are universality classes FPP on complete graph?
[Talk Jesse Goodman.]

Topology matters

Theorem 13. (BvdHH in progress). For configuration model with degree exponent $\tau > 3$, there exist $\alpha, \beta > 0$ such that

$$\frac{H_n - \alpha \log n}{\sqrt{\beta \log n}} \xrightarrow{d} Z.$$

Hopcount not always of order $\log n$:
Weights $(1 + E_e)_{e \in \mathcal{E}_n}$ and $\tau \in (2, 3)$, $H_n = \Theta(\log \log n)$.

What are universality classes FPP on random graph, and are they related to ones for FPP on complete graph?