Technische Universiteit Eindhoven
University of Technology

Small-worlds, complex networks and random graphs

Graphs and Randomness
IMPA, Rio de Janeiro,
February 19 – March 4 2018

Remco van der Hofstad

Joint work with:

- ▷ G. Hooghiemstra (Delft)
- ▷ P. Van Mieghem (Delft)
- ▷ D. Znamensky (Philips Research)
- ▷ H. van den Esker (Delft)
- ▷ S. Dommers (TU/e).

EURANDOM

Plan lectures

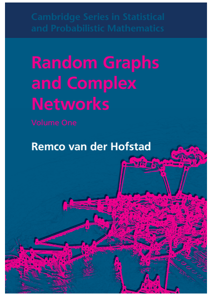
Lecture 1:
Real-world networks and random graphs

Lecture 2:
Small-world phenomena in random graphs

Lecture 3:
Information diffusion in random graphs

Material

▷ Intro random graphs:
Random Graphs and Complex Networks Volume 1
`http://www.win.tue.nl/~rhofstad/NotesRGCN.html`
Volume 2: in preparation on same site



Treat selected parts of Chapters I.1, I.6–I.8 and II.2–II.7.

▷ Slides available at
`http://www.win.tue.nl/~rhofstad/presentations.html`

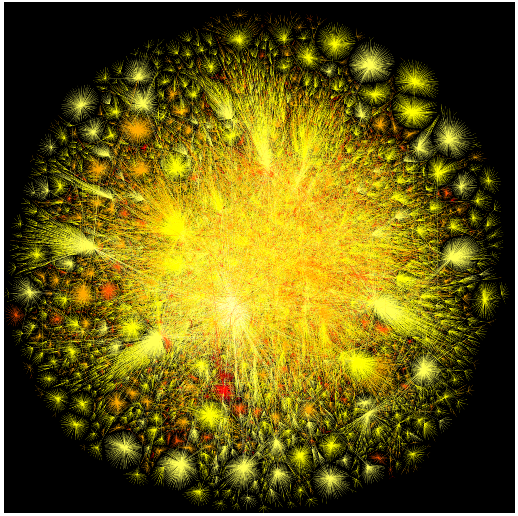
Argument are probabilistic, using
▷ first and second moment method;
▷ branching process approximations.

Lecture 1:

Real-world networks and random graphs

Complex networks

Yeast protein interaction network[a]



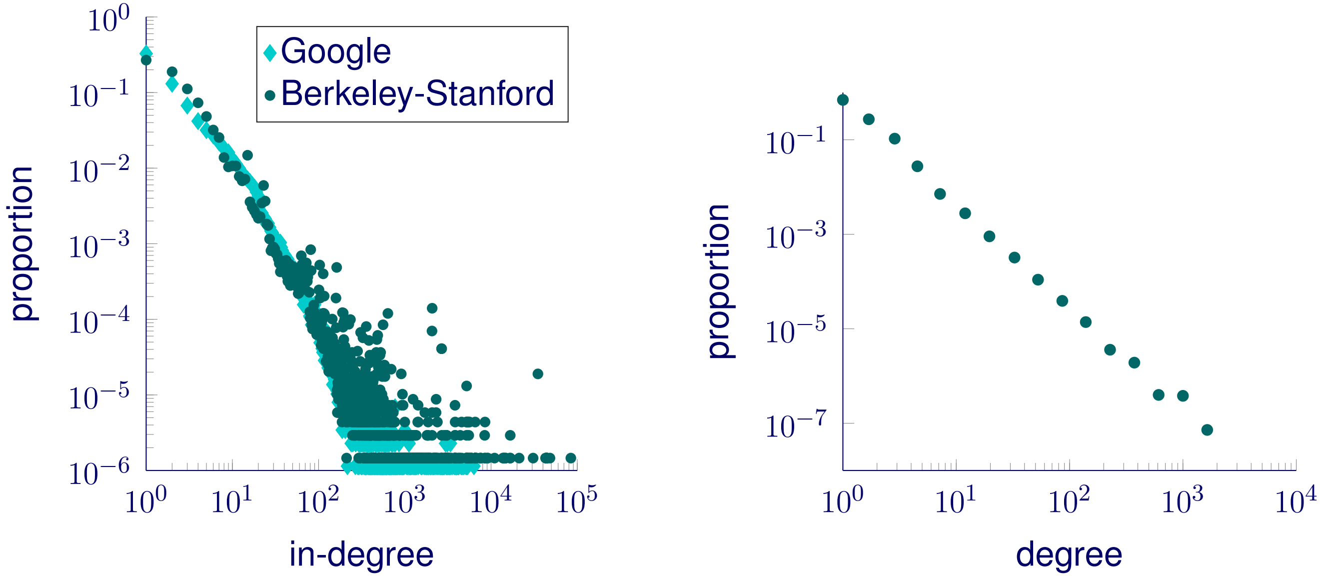
Internet 2010[b]

Attention focussing on unexpected commonality.

[a] Barabási & Óltvai 2004

[b] Opte project `http://www.opte.org/the-internet`

Scale-free paradigm

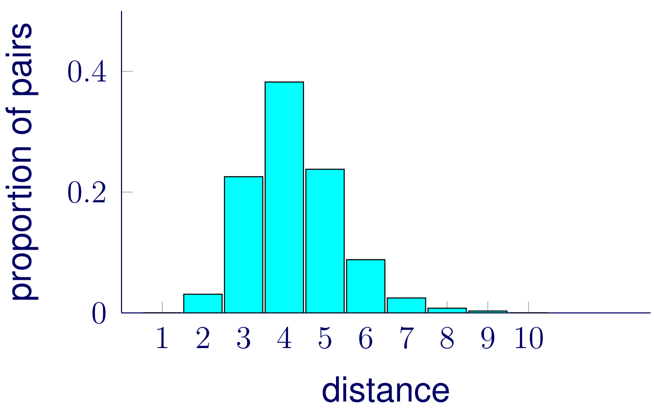


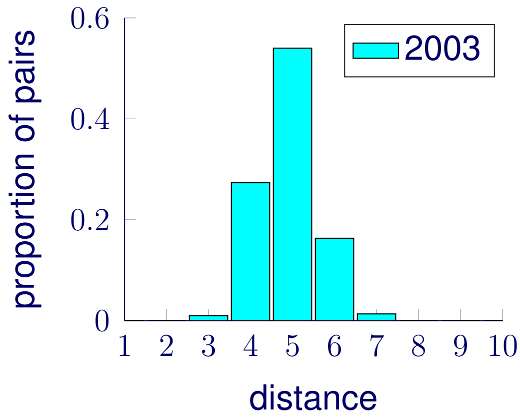
Loglog plot degree sequences WWW in-degree and Internet

▷ Straight line: proportion p_k of vertices of degree k satisfies $p_k = ck^{-\tau}$.

▷ Empirical evidence: Often $\tau \in (2, 3)$ reported.

Small-world paradigm





Distances in Strongly Connected Component WWW and IMDb.

Facebook

Largest virtual friendship network:
721 million active users,
69 billion friendship links.

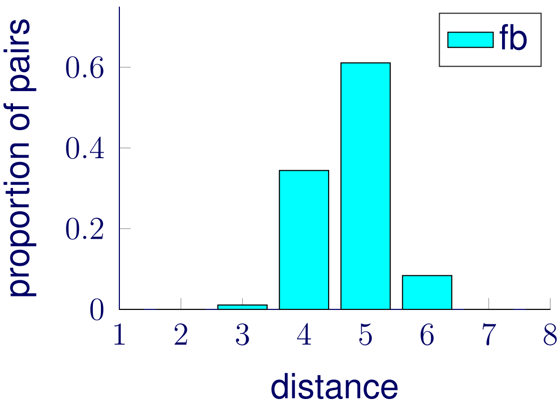
Typical distances on average four:

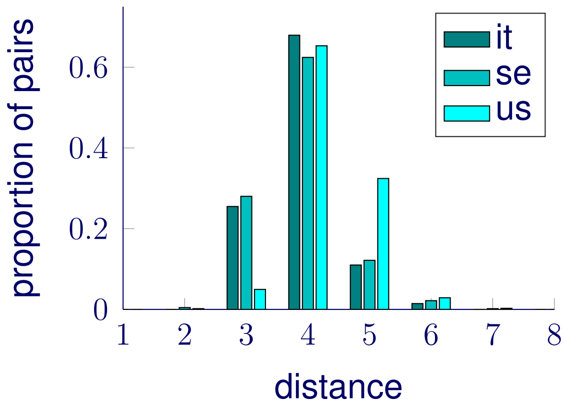
Four degrees of separation!

Fairly homogeneous (within countries, distances similar).

Recent studies:
Ugander, Karrer, Backstrom, Marlow (2011): topology
Backstrom, Boldi, Rosa, Ugander, Vigna (2011): graph distances.

Four degrees of separation





Distances in FaceBook in different subgraphs

Backstrom, Boldi, Rosa, Ugander, Vigna (2011)

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.”

Models complex networks

▷ Inhomogeneous Random Graphs:
Static random graph, independent edges with inhomogeneous edge occupation probabilities, yielding scale-free graphs.
(Chapters I.6, II.2 and II.5)

[Extensions of Erdős-Rényi random graphs Chapters I.4 and I.5.]

▷ Configuration Model:
Static random graph with prescribed degree sequence.
(Chapters I.7, II.3 and II.6)

▷ Preferential Attachment Model:
Dynamic model, attachment proportional to degree plus constant.
(Chapters I.8, II.4 and II.7)

Universality??

Erdős-Rényi

Erdős-Rényi random graph is random subgraph of complete graph on $[n] := \{1, 2, \ldots, n\}$ where each of $\binom{n}{2}$ edges is occupied independently with prob. p.

Simplest imaginable model of a random graph.

▷ Attracted tremendous attention since introduction 1959, mainly in combinatorics community:

Probabilistic method (Spencer, Erdős et al.).

▷ Average degree equals $(n-1)p \approx np$, so choose $p = \lambda/n$ to have sparse graph.

▷ Egalitarian: Every vertex has equal connection probabilities. Misses hub-like structure of real networks.

Inhomogeneous random graphs

▷ Extensions of Erdős-Rényi random graph with different vertices.

Chung-Lu: random graphs with prescribed expected degrees:
▷ Connected component structure (2002)
▷ Distance results (2002), PNAS
▷ Book (2006)

Most general:
▷ Bollobas, Janson and Riordan (2007)
▷ Söderberg (2007): Phys. Rev. E

Generalized random graph

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

$$p_{ij} = \frac{w_i w_j}{\ell_n + w_i w_j}, \qquad \text{with} \qquad \ell_n = \sum_{i\in[n]} w_i,$$

different edges being independent (Britton-Deijfen-Martin-Löf 05)
Resulting graph is denoted by $\mathrm{GRG}_n(\boldsymbol{w})$.

▷ Retrieve Erdős-Rényi RG with $p = \lambda/n$ when $w_i = n\lambda/(n-\lambda)$.

▷ Interpretation: w_i is close to expected degree vertex i.

▷ Related models:
Chung-Lu model: $p_{ij} = w_i w_j/\ell_n \wedge 1$;
Norros-Reittu model: $p_{ij} = 1 - \mathrm{e}^{-w_i w_j/\ell_n}$.
Janson (2010): General conditions for asymptotic equivalence.

Regularity vertex weights

Condition I.6.3. Denote empirical distribution function weight by

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

(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 choice 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 GRG

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

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

and $W \sim F$. In particular,

$$\sum_{l \geq k} p_l \sim c k^{-(\tau-1)} \qquad \text{iff} \qquad \mathbb{P}(W \geq k) \sim c k^{-(\tau-1)}.$$

Configuration model

$\triangleright$ Invented by Bollobás (80) EJC
to study number of graphs with given degree sequence.
Inspired by Bender+Canfield (78) JCT(A)
Giant component: Molloy, Reed (95)
Popularized by Newman-Strogatz-Watts (01)

$\triangleright$ In configuration model $\mathrm{CM}_n(\boldsymbol{d})$ degree sequence is prescribed:

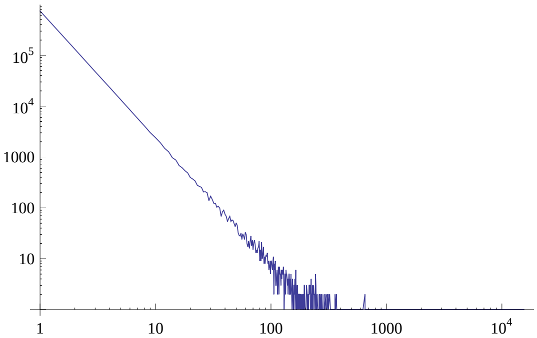
> $\triangleright$ n number of vertices;
> $\triangleright$ $\boldsymbol{d} = (d_1, d_2, \ldots, d_n)$ sequence of degrees is given.

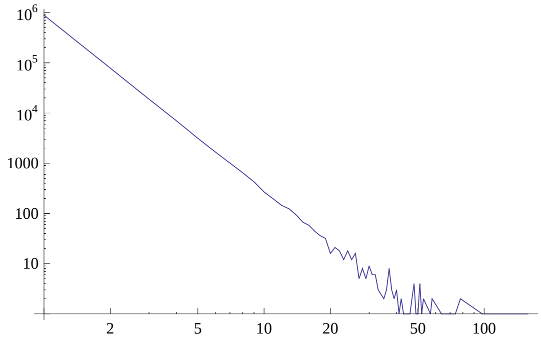
Often take $(d_i)_{i\in[n]}$ to be i.i.d. sequence.

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

$$\mathbb{P}(d_1 \geq k) = c_\tau k^{-\tau+1}(1 + o(1)).$$

Power laws 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.

Graph construction CM

▷ Assign d_j half-edges to vertex j. Assume 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 $\mathrm{CM}_n(\boldsymbol{d})$.

Regularity vertex degrees

Condition I.7.5. 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.

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

Canonical choice degrees

(A) Take $\boldsymbol{d} = (d_1, \ldots, d_n)$ as i.i.d. random variables with distribution function F.

Double randomness!

(B) Take $\boldsymbol{d} = (d_1, \ldots, d_n)$ such that

$$n_k = \#\{k\colon d_i = k\} = \lceil nF(k) \rceil - \lceil nF(k-1) \rceil.$$

Alternative: $d_i = [1-F]^{-1}(i/n)$, with F distribution function on $\mathbb{N}$.

$\triangleright$ Interpretation: Proportion of vertices i with $d_i = k$ is close to

$$F(k) - F(k-1) = p_k = \mathbb{P}(D = k),$$

where D has distribution function F.

Simple CMs

Proposition I.7.7. Let $G = (x_{ij})_{i,j\in[n]}$ be multigraph on $[n]$ s.t.

$$d_i = x_{ii} + \sum_{j\in[n]} x_{ij}.$$

Then, with $\ell_n = \sum_{v\in[n]} d_v$,

$$\mathbb{P}(\mathrm{CM}_n(\boldsymbol{d}) = G) = \frac{1}{(\ell_n - 1)!!} \frac{\prod_{i\in[n]} d_i!}{\prod_{i\in[n]} 2^{x_{ii}} \prod_{1\leq i\leq j\leq n} x_{ij}!}.$$

Consequently, number of simple graphs with degrees $\boldsymbol{d}$ equals

$$N_n(\boldsymbol{d}) = \frac{(\ell_n - 1)!!}{\prod_{i\in[n]} d_i!} \mathbb{P}(\mathrm{CM}_n(\boldsymbol{d}) \text{ simple}),$$

and, conditionally on $\mathrm{CM}_n(\boldsymbol{d})$ simple,

$\mathrm{CM}_n(\boldsymbol{d})$ is uniform random graph with degrees $\boldsymbol{d}$.

Relation GRG and CM

Theorem I.6.15. The $\mathrm{GRG}_n(\boldsymbol{w})$ with edge probabilities $(p_{ij})_{1\leq i<j\leq n}$ given by

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

conditioned on its degrees $\{d_i(X) = d_i \forall i \in [n]\}$ is uniform over all graphs with degree sequence $(d_i)_{i\in[n]}$.

Consequently, conditionally on degrees, $\mathrm{GRG}_n(\boldsymbol{w})$ has the same distribution as $\mathrm{CM}_n(\boldsymbol{d})$ conditioned on simplicity.

Allows properties of $\mathrm{GRG}_n(\boldsymbol{w})$ to be proved through $\mathrm{CM}_n(\boldsymbol{d})$ by showing that degrees $\mathrm{GRG}_n(\boldsymbol{w})$ satisfy right asymptotics.

Inspires Degree Regularity Condition.†

Cycles + self-loops

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

$\triangleright$ Let D_n denote degree of uniformly chosen vertex. Condition 7.5(a): D_n converges in distribution to limiting random variable D.

$\triangleright$ 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

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

[Theorem I.7.8, Prop. I.7.9]

$\triangleright$ Proof: moment method (Bollobás 80, Janson 09) or Chen-Stein method (Angel-Holmgren-vdH 16).

Lecture 2:

Small-world phenomenon on random graphs

Preferential attachment models

▷ Albert-Barabási (1999):
Emergence of scaling in random networks (Science).
30472 cit. (23-2-2017).
▷ Bollobás, Riordan, Spencer, Tusnády (2001):
The degree sequence of a scale-free random graph process (RSA)
766 cit. (23-2-2017).

[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

▷ Albert-Barabási (1999):
Emergence of scaling in random networks (Science).
30472 cit. (23-2-2017).
▷ Bollobás, Riordan, Spencer, Tusnády (2001):
The degree sequence of a scale-free random graph process (RSA)
766 cit. (23-2-2017).

[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.

Old-get-richer model.

Preferential attachment

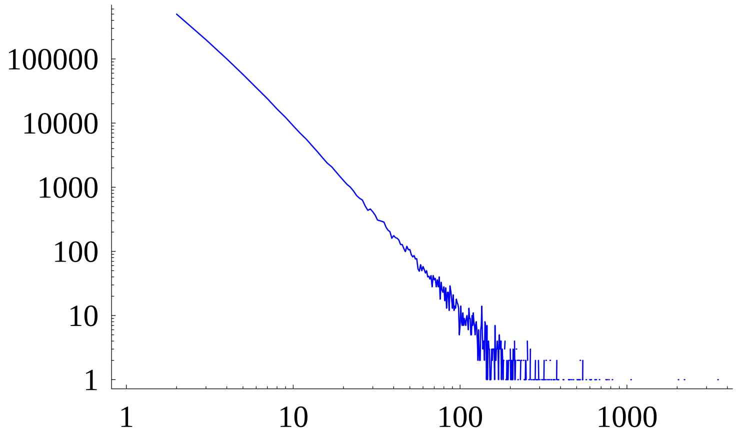
At time n, single vertex is added with m edges emanating from it. Probability that edge connects to ith vertex is proportional to

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

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

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

Bol-Rio-Spe-Tus 01 $\delta = 0$, DvdEvdHH09,...



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

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."

Degrees in PAM

Bollobás-Riordan-Spencer-Tusnády 01: First to give proof for $\delta = 0$. Tons of subsequent proofs, many of which follow same key steps:

▷ A clever Doob martingale:

$$M_n = \mathbb{E}[N_k(t) \mid \mathrm{PA}_n],$$

where $N_k(t)$ is number of vertices of degree k at time t, combined with Azuma-Hoeffding.

▷ Analysis of means: Identify asymptotics $\mathbb{E}[N_k(t)]$ and prove that

$$\frac{\mathbb{E}[N_k(t)]}{t} \to p_k.$$

This can be done in many different ways. We follow Section I.8.4.[†]

Network models I

▷ **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.

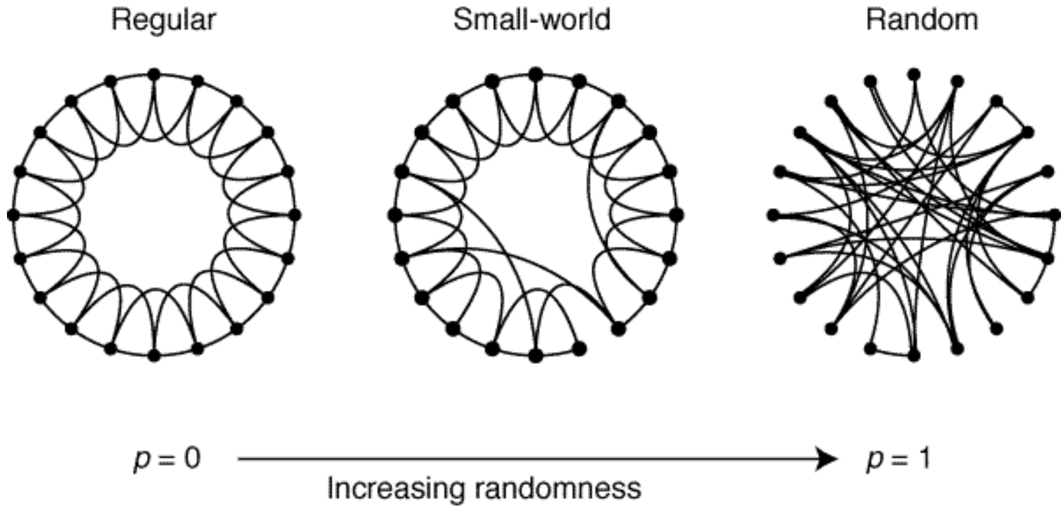
Application: Social networks?

▷ **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? Often used for brain.

Small-world model



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.

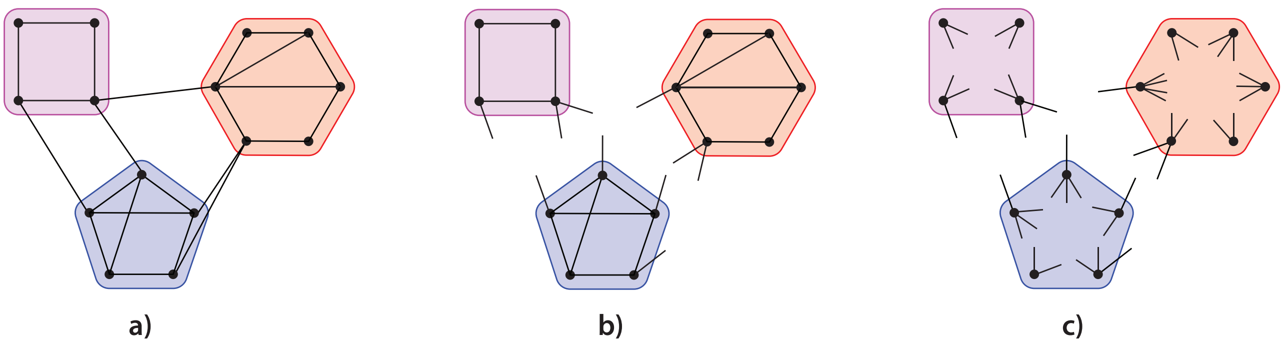
Application: Social networks, WWW?

Hierarchical CM

Vertex i is blown up to represent small community graph.
Connect inter-community half-edges uniformly at random.

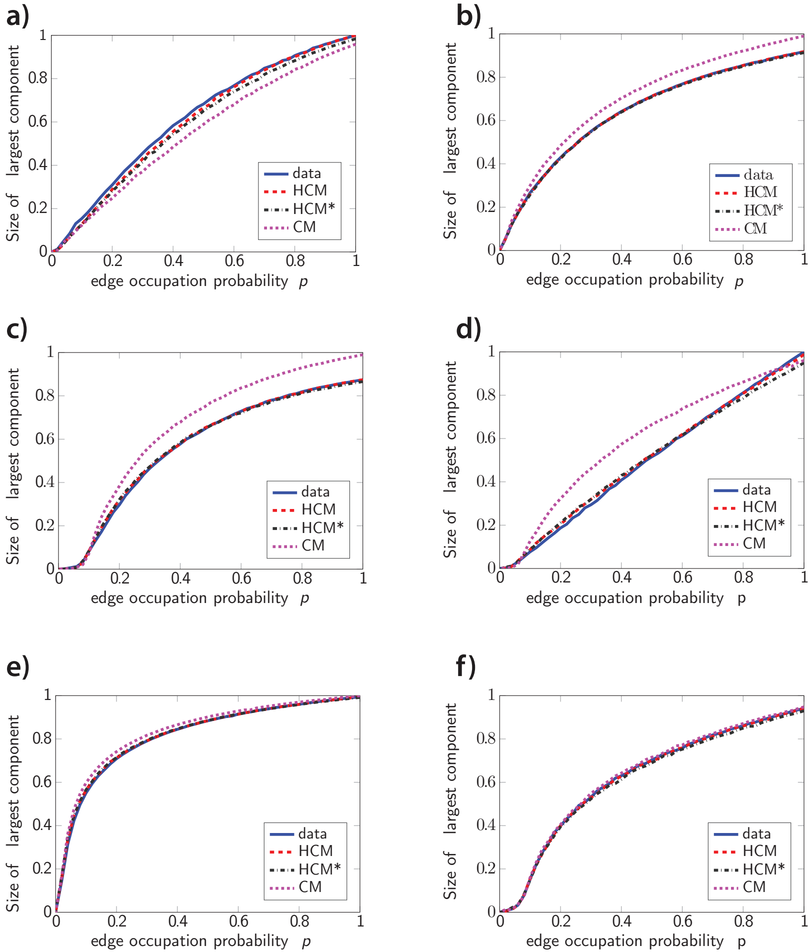
Result: Random graph with (roughly) specified communities.

Application: Many real-world networks on mesoscopic scale.
Stegehuis+vdH+vL16 Scientific Reports, Phys. Rev. E.



Network HCM HCM*

Percolation on HCM



Phase transition CM

Let $\mathcal{C}_{\max}$ denote largest connected component in $\mathrm{CM}_n(\boldsymbol{d})$.

Theorem 1. [Mol-Ree 95, Jan-Luc 07, Theorem II.3.4]. When Conditions 7.5(a-b) hold,

$$\frac{1}{n}|\mathcal{C}_{\max}| \xrightarrow{\mathbb{P}} \zeta,$$

where $\zeta > 0$ precisely when $\nu > 1$ with

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

$\triangleright$ Note: $\zeta > 0$ always true when $\nu = \infty$.

$\triangleright$ $d_{\min} = \min_{i\in[n]} d_i \geq 3 : \mathrm{CM}_n(\boldsymbol{d})$ with high probability connected. Wormald (81), Luczak (92).

$\triangleright$ $d_{\min} = \min_{i\in[n]} d_i \geq 2 : n - |\mathcal{C}_{\max}| \xrightarrow{d} X$ for non-trivial X. Luczak (92), Federico-vdH (17).

Phase transition for GRG

Let $\mathcal{C}_{\max}$ denote largest connected component in $\mathrm{GRG}_n(\boldsymbol{w})$.

Theorem 2. [Chu-Lu 03, Bol-Jan-Rio 07]. When Conditions 6.3(a-b) hold, there exists $\zeta < 1$ such that

$$\frac{1}{n}|\mathcal{C}_{\max}| \xrightarrow{\mathbb{P}} \zeta,$$

where $\zeta > 0$ precisely when $\nu > 1$, where

$$\nu = \frac{\mathbb{E}[W^2]}{\mathbb{E}[W]}.$$

▷ Note: $\zeta > 0$ always true when $\nu = \infty$.

▷ Bol-Jan-Rio 07 much more general.

Graph distances CM

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

Theorem 3. [vdHHVM05, Theorem II.6.1]. When Conditions I.7.5(a-c) hold and $\nu = \mathbb{E}[D(D-1)]/\mathbb{E}[D] > 1$, conditionally on $H_n < \infty$,

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

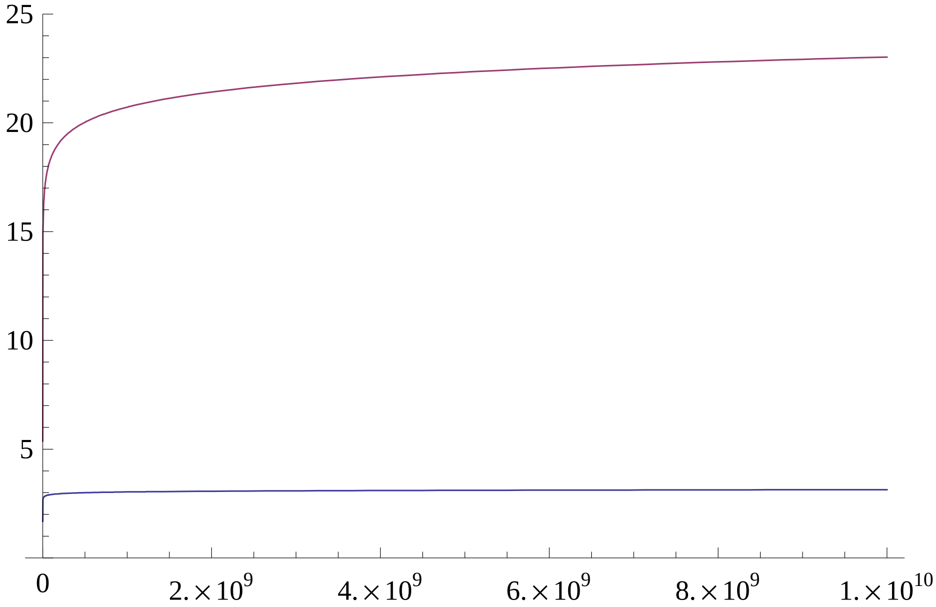
▷ For i.i.d. degrees having at most power-law tails, fluctuations are bounded.

Theorem 4. [vdHHZ07, Norros-Reittu 04, Theorem II.6.2]. Let Conditions I.7.5(a-b) hold. When $\tau \in (2,3)$, conditionally on $H_n < \infty$,

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

▷ vdH-Komjáthy16: For power-law tails, fluctuations are bounded and do not converge in distribution.

Six degrees of separation revisited



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

Diameter CM

Theorem 5. [Fernholz-Ramachandran 07, Theorem II.6.20]. Under Conditions I.7.5(a-b), there exists b s.t.

$$\frac{\mathrm{diam}(\mathrm{CM}_n(\boldsymbol{d}))}{\log n} \xrightarrow{\mathbb{P}} \frac{1}{\log(\nu)} + 2b.$$

Here $b > 0$ precisely when $\mathbb{P}(D \leq 2) > 0$.

Theorem 6. [Caravenna-Garavaglia-vdH 17, Theorem II.6.21]. Under Conditions I.7.5(a-b), when $\tau \in (2,3)$ and $\mathbb{P}(D \geq 3) = 1$,

$$\frac{\mathrm{diam}(\mathrm{CM}_n(\boldsymbol{d}))}{\log \log n} \xrightarrow{\mathbb{P}} \frac{2}{|\log(\tau-2)|} + \frac{2}{\log(d_{\min}-1)}.$$

Graph distances IRG

Theorem 7. [Chung-Lu 03, Bol-Jan-Rio 07, vdEvdHH08, Thm. II.5.2] When Conditions I.6.3(a-c) hold and $\nu = \mathbb{E}[W^2]/\mathbb{E}[W] > 1$, conditionally on $H_n < \infty$,

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

Under somewhat stronger conditions, fluctuations are bounded.

Theorem 8. [Chung-Lu 03, Norros-Reittu 06, Theorem II.5.3]. When $\tau \in (2,3)$, and Conditions I.6.3(a-b) hold, under certain further conditions on F_n, and conditionally on $H_n < \infty$,

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

▷ Similar extensions for diameter as for CM. Again Bol-Jan-Rio 07 prove Theorem 7 in much more general setting.†

Distances PA models

▷ Note that results CM and GRG are very alike, with CM having more general behavior (e.g., connectivity). Sign of the wished for universality.

> 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.

▷ In general, this question is still wide open, but certain indications are obtained.

▷ PAM tends to be much harder to analyze, due to

time dependence.

Connectivity PAM

Theorem 9. [Theorem II.4.13] Let $m \geq 2$. Then, there exists a random time $T < \infty$, such that the preferential attachment model is connected for all times after T.

▷ Not necessarily true when $m = 1$:
Depends on precise PA rule.

▷ Analogy: $\mathrm{CM}_n(\boldsymbol{d})$ with high probability connected when $d_{\min} \geq 3$.

Distances PA models

Theorem 10 [Bol-Rio 04]. For all $m \geq 2$ and $\tau = 3$,

$$\mathrm{diam}(\mathrm{PA}_{m,0}(n)) = \frac{\log n}{\log\log n}(1 + o_{\mathbb{P}}(1)), \qquad H_n = \frac{\log n}{\log\log n}(1 + o_{\mathbb{P}}(1)).$$

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

$$\mathrm{diam}(\mathrm{PA}_{m,\delta}(n)) = \Theta(\log n), \qquad H_n = \Theta(\log n).$$

Theorem 12 [Dommers-vdH-Hoo 10, Der-Mon-Mor 12, Car-Gar-vdH17]. For all $m \geq 2$ and $\tau \in (2, 3)$,

$$\frac{H_n}{\log\log n} \xrightarrow{\mathbb{P}} \frac{4}{|\log(\tau - 2)|}, \qquad \frac{\mathrm{diam}(\mathrm{PA}_{m,\delta}(n))}{\log\log n} \xrightarrow{\mathbb{P}} \frac{4}{|\log(\tau - 2)|} + \frac{2}{\log m}.$$

Proof: Neighborhoods CM

▷ Important ingredient in proof is description local neighborhood of uniform vertex $U_1 \in [n]$. Its degree has distribution $D_{U_1} \stackrel{d}{=} D$.

▷ Take any of D_{U_1} neighbors a of U_1. Law of number of forward neighbors of a, i.e., $B_a = D_a - 1$, is approximately

$$\mathbb{P}(B_a = k) \approx \frac{(k+1)}{\sum_{i \in [n]} d_i} \sum_{i \in [n]} \mathbb{1}_{\{d_i = k+1\}} \xrightarrow{\mathbb{P}} \frac{(k+1)}{\mathbb{E}[D]} \mathbb{P}(D = k+1).$$

Equals size-biased version of D minus 1. Denote this by $D^\star - 1$.

Lecture 3:

Small worlds and Information diffusion on random graphs

Graph distances CM

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

Theorem 3. [vdHHVM05, Theorem II.6.1]. When Conditions I.7.5(a-c) hold and $\nu = \mathbb{E}[D(D-1)]/\mathbb{E}[D] > 1$, conditionally on $H_n < \infty$,

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

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

Theorem 4. [vdHHZ07, Norros-Reittu 04, Theorem II.6.2]. Let Conditions I.7.5(a-b) hold. When $\tau \in (2,3)$, conditionally on $H_n < \infty$,

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

▷ vdH-Komjáthy16: For power-law tails, fluctuations are bounded and do not converge in distribution.

Proof: Neighborhoods CM

$\triangleright$ Important ingredient in proof is description local neighborhood of uniform vertex $U_1 \in [n]$. Its degree has distribution $D_{U_1} \overset{d}{=} D$.

$\triangleright$ Take any of D_{U_1} neighbors a of U_1. Law of number of forward neighbors of a, i.e., $B_a = D_a - 1$, is approximately

$$\mathbb{P}(B_a = k) \approx \frac{(k+1)}{\sum_{i\in[n]} d_i} \sum_{i\in[n]} \mathbb{1}_{\{d_i=k+1\}} \xrightarrow{\mathbb{P}} \frac{(k+1)}{\mathbb{E}[D]} \mathbb{P}(D = k+1).$$

Equals size-biased version of D minus 1. Denote this by $D^\star - 1$.

Local tree-structure CM

$\triangleright$ Forward neighbors of neighbors of U_1 are close to i.i.d. Also forward neighbors of forward neighbors have asymptotically same distribution...

$\triangleright$ Conclusion: Neighborhood looks like branching process with offspring distribution $D^\star - 1$ (except for root, which has offspring $D.)$

$\triangleright$ Tool to make this precise is

local weak convergence.

Local weak convergence

$\triangleright$ Key technique in analyzing sparse graphs is

local weak convergence.

Makes statement that local neighborhoods in CM are like BP exact. See Section II.1.4 for intro LWC and Section II.3.2 for LWC CM.[†]

$\triangleright$ Applies much more generally:
- General IRG: Section II.2.2.
- PAM: Berger-Borgs-Chayes-Saberi (14) and Section II.4.2.

$\triangleright$ LWC holds when

$$\frac{1}{n}\sum_{i\in[n]} \mathbb{1}_{\{B_r(i)\simeq(H,y)\}} \xrightarrow{\mathbb{P}} \mathbb{P}(B_r(\varnothing)\simeq(H,y)),$$

for any rooted graph (H,y), where $B_r(i)$ is r-neighborhood of $i\in[n]$ and $B_r(\varnothing)$ is r-neighborhood of $\varnothing$ in some limiting rooted random graph.

Local weak convergence

Local weak convergence implies that

$\triangleright |\mathcal{C}_{\max}|/n$ at most $\mathbb{P}(B_r(\varnothing) > 0 \forall r)$ (=one-sided LLN);
$\triangleright$ proportion neighborhoods of specific shape converges;
$\triangleright$ various continuous functionals in local weak convergence topology converge as well.

Examples include log partition function Ising model, PageRank distribution, and through somewhat more work and under more restrictions, densest subgraph.

$\triangleright$ Many global graph parameters, such as proportion vertices in giant component or graph distances do not directly converge, but

LWC gives good starting point analysis.

Local tree-structure CM

$\rhd$ $\mathbb{E}[D^2] < \infty$: Finite-mean BP, which has exponential growth of generation sizes:

$$\nu^{-k} Z_k \stackrel{a.s.}{\longrightarrow} M \in (0, \infty),$$

on event of survival.

$\rhd$ $\tau \in (2, 3)$: Infinite-mean BP, which has double exponential growth of generation sizes:

$$(\tau - 2)^k \log(Z_k \vee 1) \stackrel{a.s.}{\longrightarrow} Y \in (0, \infty),$$

on event of survival.

$\rhd$ Indication of proof...†

Local weak conv. PAM

▷ Pólya urn: Start with r_0, b_0 red and blue balls. Draw

red ball w.p. proportional to number of red balls plus a_r,
blue ball w.p. proportional to number of blue balls plus a_b.

Replace by two balls of same color. Then number of red balls at time n equals

$$R_n \sim r_0 + \mathsf{Bin}(n, U),$$

where U is Beta random variable with parameters $(r_0 + a_r, b_0 + a_b)$.

▷ Pólya urns: Can give a Pólya urn description of

ratio degree of vertex k compared to total degree vertices $[k]$.

▷ Gives Pólya urn description of PAM at time n that gives precise law in terms of n Beta variables and independent edges.

▷ Allows to give local weak limit of PAM in terms of multitype BP with continuous types (Ber-Bor-Cha-Sab 14)

Discussion small worlds

▷ Small worlds:
Results quantify small-world behavior random graphs. Random graphs are small worlds in general, ultra-small worlds when degrees have infinite variance.

▷ Many extensions to inhomogeneous random graphs, random intersection graphs, spatial scale-free graphs,...

Universality!

▷ Locally-tree like:
Random graphs studied here are locally tree-like. Much harder in general to move away from this. E.g., spatial models such as percolation, where critical behavior is of great interest...

Smallest-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 hop-count, which is number of edges on smallest-weight path.

How does weight structure influence
structure of smallest-weight paths?

▷ Assume that

edge weights are i.i.d. (continuous) random variables.

▷ Graph distances: weights = 1.

Choice of edge weights

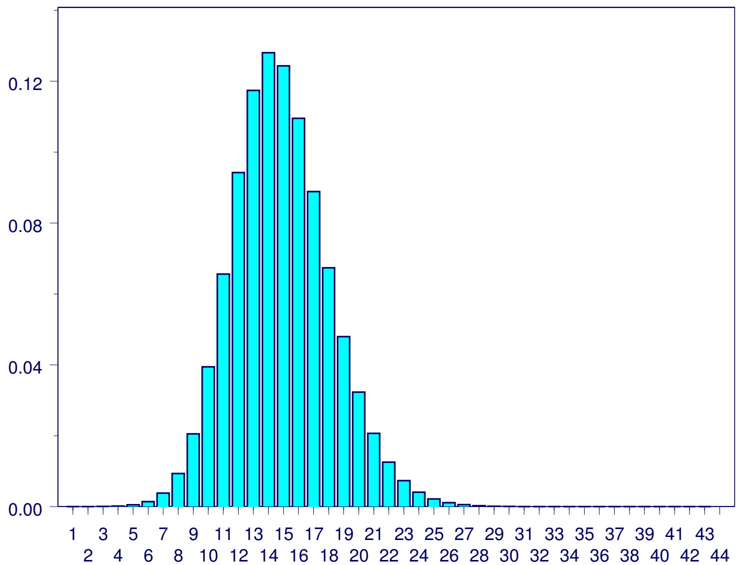
2000: CISCO recommended to use link weights that are proportional to inverse link capacity in Open Shortest Path First (OSPF). OSPF is interior routing protocol operating within a single autonomous system (AS).

CISCO recommendation: proportion $P_{\mathrm{Link}}[0, B]$ of link weights with value at most B is equal to proportion $P_{\mathrm{Cap}}[1/B, \infty)$ of links with capacity at least $1/B$.

$\triangleright$ Problem: No reliable data on empirical properties link capacities.

$\triangleright$ Solution: Use general continuous distribution link capacities. Thus, also edge weights have general continuous distribution.

Distances in IP graph



Poisson distribution??

Smallest-weight routing

▷ Smallest-weight routing problems fundamental for many related math and applied problems.

- Epidemic models;
- Rumor spread;
- Various randomized algorithms for communication (sensors);
- Competition processes,...

▷ One of the most basic information diffusion processes.

Setting

Graph denoted by $G = (V(G), E(G))$ with $|V(G)| = n$.
This talk: G configuration model, sometimes complete graph.

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

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

where π is path in G, while $(Y_e)_{e \in E(G)}$ are i.i.d. collection of weights with continuous law.

▷ Continuous weights: Optimal path π_n^* is a.s. unique. Then

$$H_n = |\pi_n^*|$$

denotes hopcount, i.e., number of edges in optimal path.

▷ Complete graph investigated in combinatorics (e.g., Janson 99) and theoretical physics (Havlin, Braunstein, Stanley, et al.).

Routing on sparse RGs

Theorem 13. [Bhamidi-vdH-Hooghiemstra 17]. Let $\mathrm{CM}_n(\boldsymbol{d})$ satisfy Condition 7.5 (a-b), and

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

Let weights be i.i.d. with general continuous distribution. Then, there exist $\alpha_n, \alpha, \beta, \gamma_n, \gamma > 0$ 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.

Universality!

Role hubs

Theorem 14. [Bhamidi-vdH-Hooghiemstra AoAP10]. Let degrees in $\mathrm{CM}_n(\boldsymbol{d})$ be i.i.d. with $\mathbb{P}(D \geq 2) = 1$ and power-law distribution with $\tau \in (2,3)$. Let weights be i.i.d. exponential r.v.'s. Then

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

for some limiting random variable $\mathcal{C}_\infty$, where Z is standard normal and $\alpha = 2(\tau-2)/(\tau-1) \in (0,1)$.

▷ Hopcount not order $\log n$: Weights $(1+E_e)_{e\in\mathcal{E}_n}$, where E_e i.i.d. exponential, and $\tau \in (2,3)$,

$$W_n, H_n = 2\log\log n/|\log(\tau-2)| \text{ plus tight.}$$

Weights $(1+X_e)_{e\in\mathcal{E}_n}$: know exactly when above tightness holds.

▷ Baroni, vdH, Komjáthy: Extension $\mathcal{C}_n \xrightarrow{d} \mathcal{C}_\infty$ to explosive case.

Discussion

Random weights have marked effect on optimal flow problem.

▷ Surprisingly universal behavior for FPP on configuration model. Even limiting random variables display large amount of universality.

▷ Universality is leading paradigm in statistical physics. Only few examples where universality can be rigorously proved. Extension to FPP on super-critical rank-1 IRGs.

▷ Proofs: Rely on coupling to continuous-time branching processes (CTBP). CTBP arises as FPP on Galton-Watson tree.

Proofs

Adding weights to branching process gives rise to

age-dependent branching process.

Is particular type of continuous-time branching process.

$\triangleright$ Let Z_t be number of alive individuals.

$\triangleright \mathbb{E}[D^2] < \infty$: CTBP is Malthusian: $\mathrm{e}^{-\alpha t} Z_t \xrightarrow{a.s.} W$ for some $W > 0$;

$$\mathcal{C}_n \approx \log n / \alpha ...$$

$\triangleright \tau \in (2, 3)$: CTBP can be explosive: $Z_t = \infty$ for some $t > 0$.
True for most weights...

$$\mathcal{C}_\infty = T_1 + T_2,$$

the sum of two i.i.d. explosion times.

Related results

Theorem 14. [Komjáthy and Kolossváry (15)]. Extension to FPP with exponential weights on inhomogeneous random graphs with fixed number of types and fixed expected degree.

Theorem 15. [Amini, Draief, Lelarge (13)]. Maximal weight from a uniform vertex for FPP on $\mathrm{CM}_n(\boldsymbol{d})$ with minimal degree $d_{\min} \geq 3$ and exponential weights satisfies

$$\frac{\max_{j\in[j]} \mathcal{C}_n(V,j)}{\log n} \xrightarrow{\mathbb{P}} \frac{1}{\nu-1} + \frac{1}{d_{\min}},$$

while weight diameter satisfies

$$\frac{\max_{i,j\in[j]} \mathcal{C}_n(i,j)}{\log n} \xrightarrow{\mathbb{P}} \frac{1}{\nu-1} + \frac{2}{d_{\min}}.$$

Similar results when $d_{\min} \in \{1,2\}$.

Related results

Theorem 16. [Amini-Peres (14)]. Maximal hopcount from single vertex on random d-regular graph with exponential weights satisfies

$$\max_{j\in[n]} \frac{H_n(1,j)}{\log n} \xrightarrow{\mathbb{P}} \alpha,$$

while hopcount diameter satisfies

$$\max_{i,j\in[n]} \frac{H_n(i,j)}{\log n} \xrightarrow{\mathbb{P}} \alpha^\star.$$

With

$$f(\alpha) = \alpha \log\Big(\frac{d-2}{d-1}\alpha\Big) - \alpha + \frac{1}{d-2},$$

$\alpha, \alpha^\star$ are solutions to $f(\alpha) = 0$ and $f(\alpha) = 1$, respectively.

Implications other models

FPP serves as a tool in many models:

▷ Direct interpretation as spread of gossip: information diffusion.

▷ Ding, Kim, Lubetzky, and Peres (2009, 2010):
Identification distance between two random vertices in two-core of slightly supercritical ERRG.

Key ingredient to determine diameter of slightly supercritical ERRG.

Winner takes it all!

FPP serves as a tool in many models:

Theorem 18. [Deijfen-vdH AoAP (2016)]
Consider competition model, where species compete for territory at unequal rates. For $\tau \in (2, 3)$, under conditions Theorem 13, each of species wins majority vertices with positive probability.
Number of vertices for losing species converges in distribution.

$\triangleright$ Antunovic, Dekel, Mossel, and Peres (2011): First passage percolation as competition model on random regular graphs.

$\triangleright$ Baroni, vdH, Komjáthy (2015): Extension to deterministic unequal weights
$\triangleright$ vdH, Komjáthy (2016): Deterministic equal weights: coexistence.

$\triangleright$ Alberg, Deijfen, Janson (2017): Extension to $\mathbb{E}[D^2] < \in$ and exponential edge weights.

Epidemic curve

FPP serves as a tool in many models:

Theorem 19. [Bhamidi-Komjáthy-vdH (2013)]
Under conditions Theorem 12, $P_n(t)$, proportion of vertices found by time t satisfies

$$P_n\big(t - \gamma_n \log n\big) \xrightarrow{d} P(t - S),$$

where S is random shift, and $t \mapsto P(t)$ distribution function.

Can interpret FPP as SIR epidemic with infinite infectious period: epidemic curve.

$\triangleright$ Barbour-Reinert (2012): Epidemic curve for more general epidemics, under stronger conditions on graph.

Conclusion routing on CM

Many results on FPP on random graphs.
Results show high amount of universality when degrees finite-variance. When degrees infinite variance, then unclear what universality classes are!

▷ Difficulty:

$\mathrm{CM}_n(\boldsymbol{d})$ contains sizeable complete graph when $\tau \in (2,3)$.

Brings us to study FPP on complete graph.

What are fluctuations diameter FPP?
Extension of Theorem 6 of Amini, Draief, Lelarge (2011)...
▷ Infinite variance degrees?
▷ General edge weights?

Complete graph

Theorem 20. [Janson 99]. Let $\mathcal{C}_n(i,j)$ be weight of smallest-weight path between $i, j \in [n]$ in K_n with exponential edge weights. Then,
(a) for every $i \neq j \in [n]$ fixed

$$\mathcal{C}_n(i,j)n/\log n \xrightarrow{\mathbb{P}} 1;$$

(b) for every $i \in [n]$ fixed

$$\max_{j\in[n]} \mathcal{C}_n(i,j)n/\log n \xrightarrow{\mathbb{P}} 2;$$

(c)

$$\max_{i,j\in[n]} \mathcal{C}_n(i,j)n/\log n \xrightarrow{\mathbb{P}} 3.$$

Complete graph

Theorem 21. [Janson 99, Bhamidi-vdH 17]. Let $\mathcal{C}_n(i,j)$ be weight of smallest-weight path between $i, j \in [n]$ in K_n with exponential edge weights. Then,

(a) for every $i \neq j \in [n]$ fixed

$$n\mathcal{C}_n(i,j) - \log n \xrightarrow{d} \Lambda_1 + \Lambda_2 - \Lambda_3,$$

where $\Lambda_1, \Lambda_2, \Lambda_3$ are three independent Gumbel random variables.

(b) for every $i \in [n]$ fixed, and with Λ_1, Λ_2 two independent Gumbel random variables,

$$\max_{j \in [n]} n\mathcal{C}_n(i,j) - 2\log n \xrightarrow{d} \Lambda_1 + \Lambda_2.$$

(c) for some limiting random variable Ξ,

$$\max_{i,j \in [n]} n\mathcal{C}_n(i,j) - 3\log n \xrightarrow{d} \Xi.$$

Open problems

Behavior FPP on complete graph with general edge weights.
Hard problem, many possible scalings for $H_n, \mathcal{C}_n$.
$\triangleright$ Exponential edge weights: Janson (1999).
Mean-field model may be harder than graph problem!
$\triangleright$ How is this related to $\mathrm{CM}_n(\boldsymbol{d})$ with infinite variance degrees?

Strong disorder regime and minimal spanning tree.
$\triangleright$ Closely related to critical percolation on random graph.
Diverse behavior depending on τ. (Braunstein, Havlin, Stanley,...)
$\triangleright$ Graph limit MST on complete graph:
Addario-Berry-Broutin-Goldschmidt-Miermont (2017).

Gold mine of interesting problems!

Literature CM

[1] Bollobás. A probabilistic proof of an asymptotic formula for the number of labelled regular graphs. *Eur. J. Combin.* **1**(4): 311–316, (1980).

[2] van der Hofstad, Hooghiemstra and Van Mieghem. Distances in random graphs with finite variance degrees. *RSA* **27**(1): 1042–9832, (2005).

[3] Fernholz and Ramachandran. The diameter of sparse random graphs. *RSA* **31**(4): 482–516, (2007).

[4] van der Hofstad, Hooghiemstra and Znamenskyi. Distances in random graphs with finite mean and infinite variance degrees. *Elec. J. Prob.* **12**(25): 703–766, (2007).

[5] Molloy and Reed. A critical point for random graphs with a given degree sequence. *RSA* **6**: 161–179, (1995).

[6] Newman, Strogatz and Watts. Random graphs with arbitrary degree distribution and their application. *Phys. Rev. E* **64**: 026118, (2000).

Literature IRG

[1] Bollobás, Janson and Riordan. The phase transition in inhomogeneous random graphs. *RSA* **31**(1): 3–122, (2007).

[2] Britton, Deijfen and Martin-Löf. Generating simple random graphs with prescribed degree distribution. *JSP* **124**(6): 1377–1397, (2006).

[3] Chung and Lu. The average distance in a random graph with given expected degrees. *Internet Math.* **1**(1): 91–113, (2003).

[4] Chung and Lu. Complex graphs and networks. CBMS Regional Conference Series in Mathematics. (2006).

[5] van den Esker, van der Hofstad and Hooghiemstra. Universality for the distance in finite variance random graphs. *J. Stat. Phys.* **133**(1): 169–202, (2006).

[6] Norros and Reittu. On a conditionally Poissonian graph process. *Adv. Appl. Probab.* **38**(1): 59–75, (2006).

Literature PAM

[1] Barabási and Albert. Emergence of scaling in random networks. *Science* **286**(2): 509–512, (1999).

[2] Bollobás and Riordan. The diameter of a scale-free random graph. *Combinatorica* **1**(2): 215–225, (2004).

[3] Bollobás, Riordan, Spencer and Tusnády. The degree sequence of a scale-free random graph process. *RSA* **18**(3): 279–290, (2001).

[4] Deijfen, van den Esker, van der Hofstad and Hooghiemstra. A preferential attachment model with random initial degrees. *Ark. Mat.* **47**(1): 41–72, (2009).

[5] Dereich, Mönch and Mörters. Typical distances in ultrasmall random networks. *Journ. Stat. Phys.* **139**(1): 72–107, (2010).

[6] Dommers, van der Hofstad and Hooghiemstra. Diameters in preferential attachment graphs. *Adv. Appl. Probab.* **44**(2): 583–601, (2012).

Literature FPP on graphs

[1] Bhamidi, Hooghiemstra and van der Hofstad. First passage percolation on random graphs with finite mean degrees. *AAP* **20**(5): 1907–1965, (2010).

[2] Bhamidi, Hooghiemstra and van der Hofstad. First passage percolation on sparse random graphs. *AoP* **45**(4): 2568–2630, (2017).

[3] Baroni, van der Hofstad and Komjáthy. Nonuniversality of weighted random graphs with infinite variance degree. *Journal of Applied Probability,* **54**(1): 146–164, (2017).

[4] Deijfen and van der Hofstad. The winner takes it all. *AAP,* **26**(4): 2419–2453, (2016).

[5] Janson. One, two and three times $\log n/n$ for paths in a complete graph with random weights. *CPC* **8**(04):347–361, (1999).

Literature flows on graphs

[1] Addario-Berry, Broutin, Goldschmidt and Miermont. The scaling limit of the minimum spanning tree. *AoP* **45**(5): 3075–3144, (2017).

[2] Amini and Peres. Shortest-weight paths in random regular graphs. *SIAM Journal on Discrete Mathematics* **28**: 656–672, (2014).

[3] Amini, Draief and Lelarge. Flooding in weighted sparse random graphs. *SIAM Journal on Discrete Mathematics* **27**: 1–26, (2013).

[4] Barbour and Reinert. Approximating the epidemic curve. *Electron. J. Probab.* **18**, no. 54, 30 pp, (2013).

[5] Ding, Kim, Lubetzky and Peres. Diameters in supercritical random graphs via first passage percolation. *CPC* **19**: 729–751, (2010).

[6] Kolossváry and Komjáthy. First passage percolation on inhomogeneous random graphs. *Adv. in Appl. Probab.* **47** (2): 589–610, (2015).