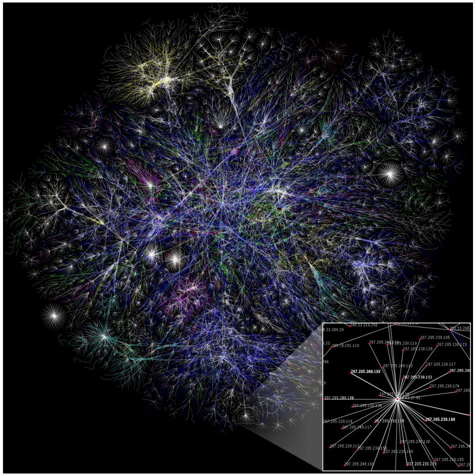


Network Science: Structure and Function

Remco van der Hofstad

Probability and Statistics Seminar,

IIT Bombay, November 24 and 25, 2025

TU/e
EINDHOVEN UNIVERSITY OF TECHNOLOGY

EURANDOM

NETWORKS

Lecture overview

⋆ Part 1: Complex networks and network science.

⋆ Part 2: Random graph models for complex networks.

⋆ Part 3: Local and ‘almost local’ structure of random graphs.

⋆ Part 4: Small- and ultra-small-world random graphs.

⋆ Part 5: Competition and fake news on scale-free random graphs.

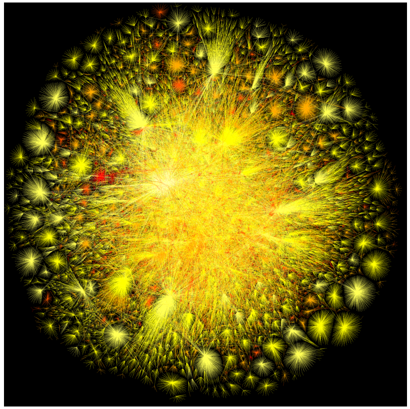
Will extensively use Mentimeter to keep all of you engaged. Do participate to improve your learning!

Part 1:

Complex networks and network science

Complex networks

Burst of activity in past 25 years.
See books Newman (2010) or Barabási (web book) for examples and theory.



Opte project, Barrett Lyon, 2010]

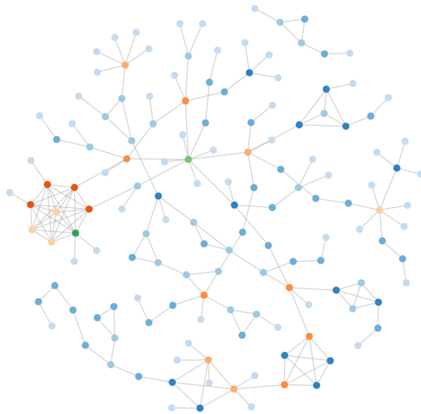
Networks come in different flavours:

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

Attention focussing on unexpected commonality.

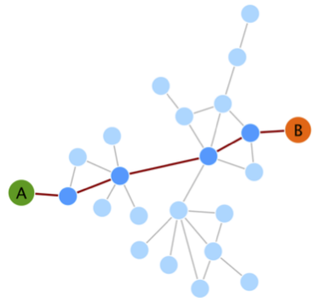
Graphs or networks

Network is another word for graph.
Graphs are mathematical constructs to study relations between objects.



Graph consists of vertices (= nodes, sites) and edges (= bonds).

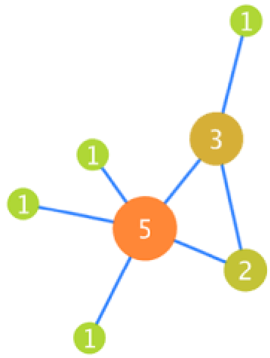
⋆ Vertices: elements of whom we study their relations;
⋆ Edges: relations between elements: cables, friendships, who eats who, hyperlink,...



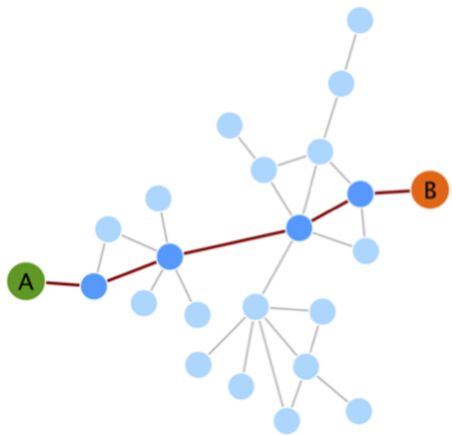
Edge is building block of relational data

Graph terminology

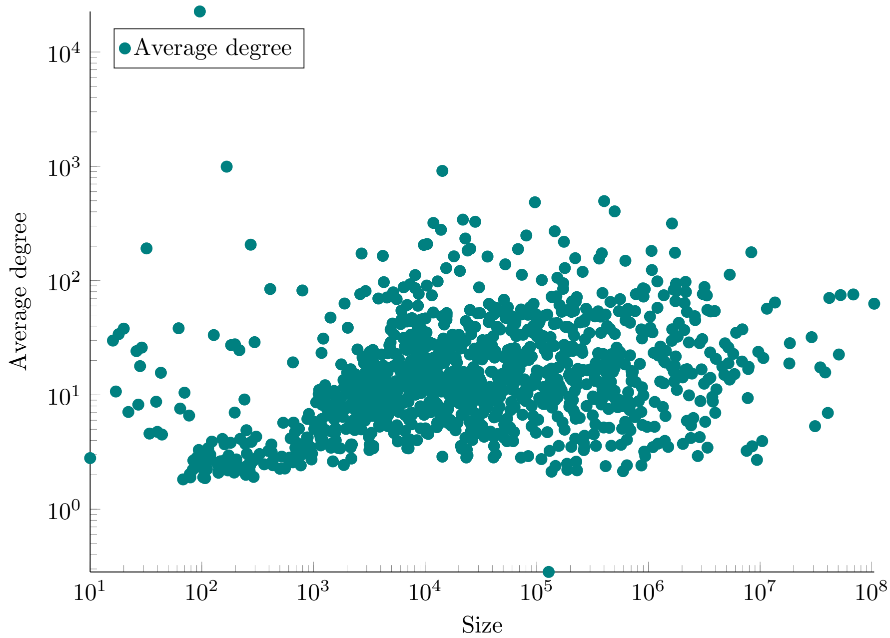
Degree vertex is number of edges it is in:
its number of 'friends'



Graph distance between pair of vertices is minimal number of edges needed to hop between them.

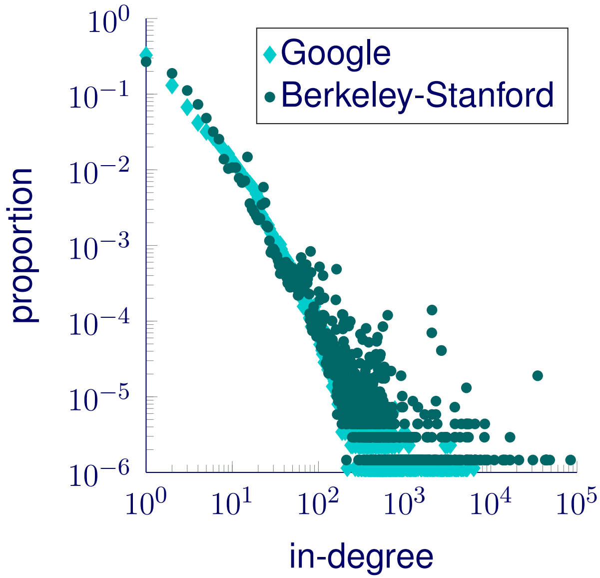


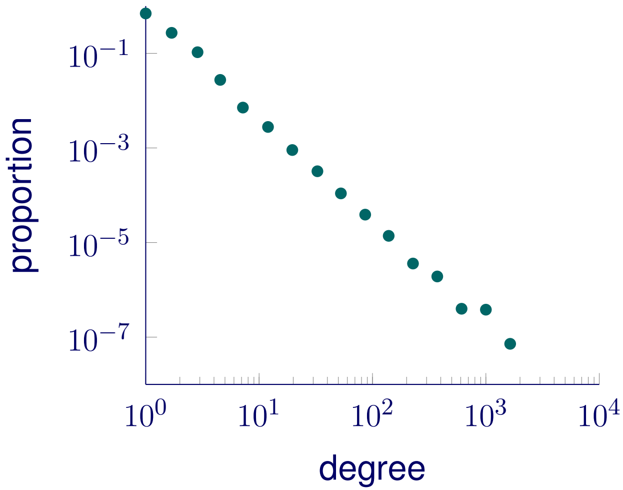
Networks are sparse



Average degrees of 1,203 networks in KONECT

Scale-free paradigm





Loglog plot (in-)degree sequences WWW and Internet (courtesy Krioukov)

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

▷ Empirical data finds $\tau \in (2, 3)$: highly-variable number of neighbours

Scale-free paradigm

Degree sequence $(p_1, p_2, p_3, \ldots)$ of graph:

p_1 is proportion of elements with degree 1,
p_2 is proportion of elements with degree 2,
...
p_k is proportion of elements with degree k.

Then

$$p_k \approx ck^{-\tau},$$

precisely when

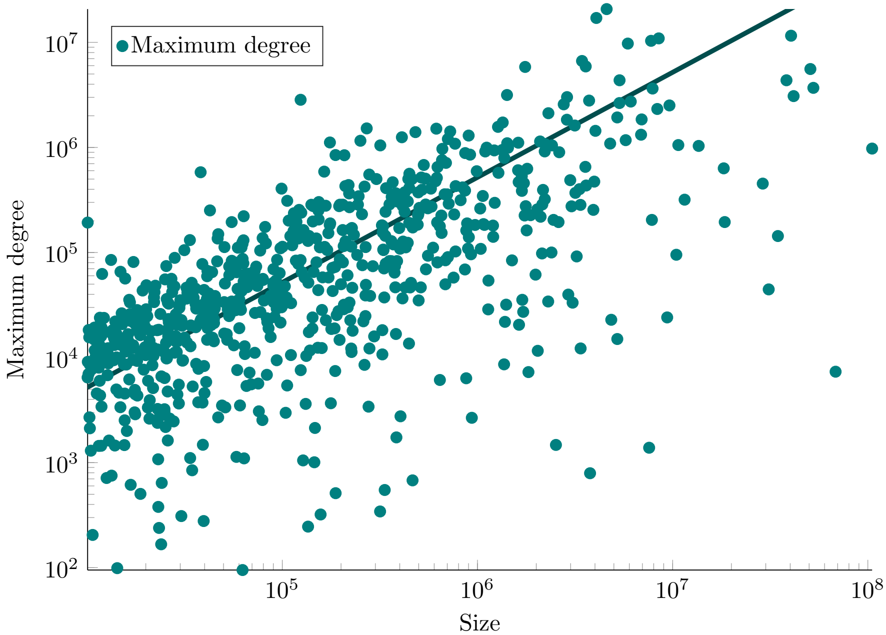
$$\log p_k \approx \log c - \tau \log k.$$

Approximate linearity $\log p_k$ and $\log k$

Maximal degree

How large is maximum of n i.i.d. random variables with probability mass function $p_k = ck^{-\tau}$?

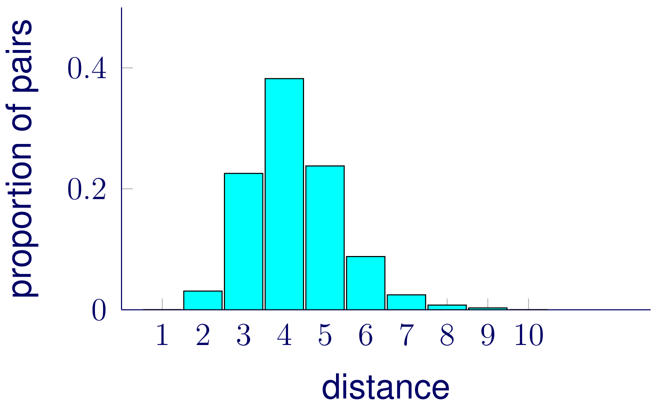
Network inhomogeneity

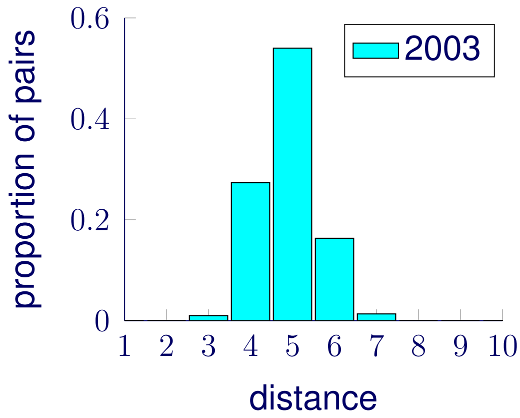


Maximal degrees in 727 networks larger than 10,000 from KONECT
Linear regression gives $\log d_{\max} = 0.742 + 0.519 \log n.$

Consistent with $\tau \approx 2.93$ on average.

Small-world paradigm





Distances in Strongly Connected Component WWW and IMDb in 2003.

Friendship paradox

Networking paradox [Scott L. Feld (1991)]:

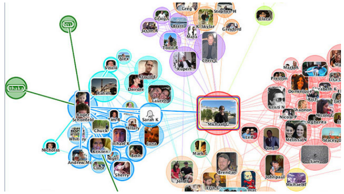
Why your friends have more friends than you do!

Wikipedia: In Twitter, the people a person follows almost certainly have more followers than they. This is because people are more likely to follow those who are popular than those who are not.

Random individual has k friends with probability equal to proportion of vertices with degree k.

Friendship paradox

Average number of friends random individual equals average degree network.



Wikipedia: The average number of friends that a typical friend has can be modeled by choosing, uniformly at random, an edge of the graph and an endpoint of that edge, and again calculating the degree of the selected endpoint.

▷ Paradox: Average degree person in random friendship turns out to be strictly larger than average degree.

In math formulas, with $D^\star$ degree of person in random friendship,

$$\mathbb{E}[D^\star] = \mathbb{E}[D] + \frac{\mathrm{Var}(D)}{\mathbb{E}[D]} > \mathbb{E}[D].$$

Your friends have more friends than you do!?

Friendship paradox

Are you convinced by above argument? I think of myself as random individual...

Friendship paradox

Take vertex uniformly at random, then take one of its neighbours and inspect its degree. Denote degrees at both sides (D_1, D_2). Then,

- $\triangleright$ D_1 has same distribution as D, but
- $\triangleright$ D_2 does not have same distribution as $D^\star$!

Still it can be shown that

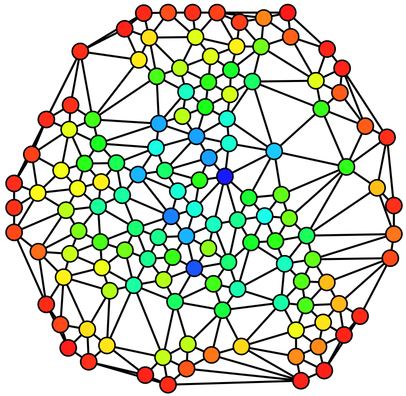
$$\mathbb{E}[D_2] > \mathbb{E}[D]!$$

Your friends have more friends than you do!

Centrality measures

▷ Closeness centrality:
Measures to what extent vertex can reach others using few hops.
Vertices with low closeness centrality are central in network.

▷ Betweenness centrality:
Measures extent to which vertex connects various parts of network.



Betweenness large for bottlenecks.

▷ PageRank:
Measures extent to which vertex is visited by random walk.
Used in Google to rank importance in web pages.

PageRank

▷ Page Rank. Solution $\vec{R}$ to

$$R_i = \alpha \sum_{j \to i} \frac{R_j}{D_j^{(\mathrm{out})}} + (1-\alpha).$$

Used in Google to rank importance in web pages:

Bringing order to the web!

Can also be seen as n times stationary distribution of bored surfer on network.

Network statistics

(Global) Clustering coefficient:

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

Proportion of friends that are friends of one another.

Assortativity coefficient:

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

Correlation between degrees at either end of edge.

Such measures are quite different across real-world networks.

Network science

▷ Complex networks modelled using

random graphs.

▷ Network functionality modelled by stochastic processes on, or algorithms for, them.

▷ A plethora of examples:

Disease spread	Synchronization
Information diffusion	Robustness to failures
Consensus reaching	Information retrieval

▷ Network algorithms: PageRank, community detection,...

▷ Prominent part of applied math for decades to come.

Part 2:

Random graph models for complex networks

Model jungle

Who of you have heard of following models?

Model jungle

Plethora of sparse random graph models have been invented:

$\star$ Static models:
Erdős-Rényi random graph, inhomogeneous random graph (IRG), configuration model (CM), exponential random graphs,...

$\star$ Dynamic models:
Growing models such as preferential attachment model (PAM), copying models, as well as dynamic versions of above models of fixed size...

Extensions:
$\star$ Directed random graphs: directed IRGs and CMs,...;
$\star$ Models with communities: stochastic block models, random intersection graphs, hierarchical CMs;
$\star$ Geometric random graphs: hyperbolic random graphs, geometric random graphs, geometric IRGs, geometric PAMs.

Model jungle

Plethora of sparse random graph models have been invented:

⋆ Static models:
Erdős-Rényi random graph, inhomogeneous random graph (IRG), configuration model (CM), exponential random graphs,...

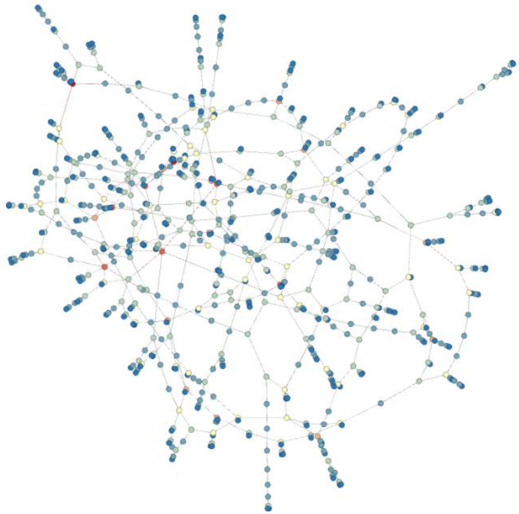
⋆ Dynamic models:
Growing models such as preferential attachment model (PAM), copying models, as well as dynamic versions of above models of fixed size...

Need techniques to deal with many models at once:
local convergence.

These talks

In this lecture series, we focus on the following models:

Static models:
Graph has fixed number of elements.
Configuration model and
Inhomogeneous random graphs



Dynamic models:
Graph has growing size:
Preferential attachment model

Erdős-Rényi

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

Erdős-Rényi random graph is random subgraph of complete graph on $[n]$ where each of pair of vertices is occupied independently with same probability.

Simplest imaginable model of a random graph.

$\triangleright$ Attracted tremendous attention since introduction 1959, mainly in combinatorics community.

Probabilistic method (Erdős et al).

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

Erdős-Rényi degrees

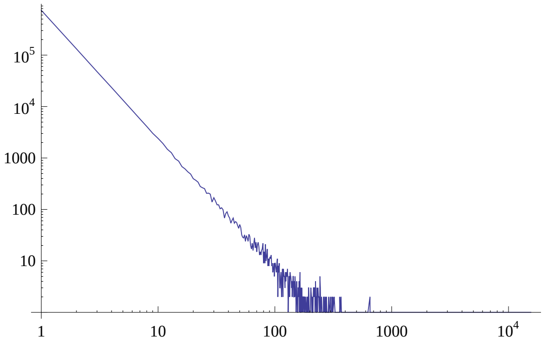
What is expected degree of Erdős-Rényi random graph with parameter p?

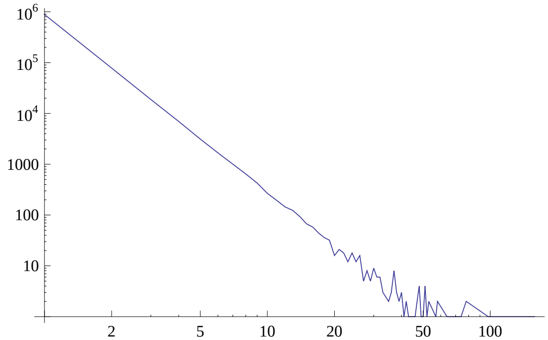
Configuration model

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

Here, chosen degrees could be those of real-world network, or generated by some random process.

▷ Highly-variable degrees: similar pictures as real-world networks:





Loglog plot degree sequence CM of $1,000,000$ vertices

Graph construction CM

▷ Assign d_j half-edges to vertex j.

Pair half-edges to create edges as follows:

a) Number half-edges from 1 to ℓ_n in any order.

b) Connect first half-edge at random with another half-edge.

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

▷ We denote resulting graph by $\mathrm{CM}_n(\boldsymbol{d})$.

Order?

Does the order in which we pair half-edges matter?

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,

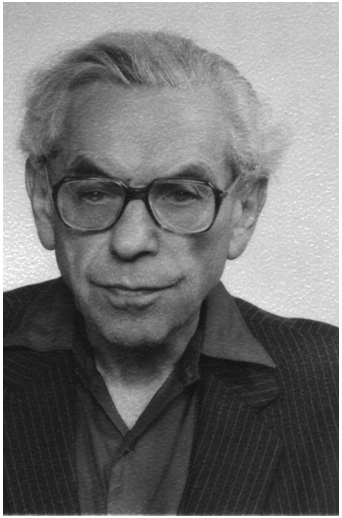
$\mathrm{CM}_n(\boldsymbol{d})$ is uniform random graph with degrees $\boldsymbol{d}$.

Probabilistic method

⋆ Configuration model conditioned on not having any self-loops and multiple edges is uniform random graph with prescribed degrees.

⋆ Probability that CM has no self-loops and multiple edges converges when

$$\frac{1}{n}\sum_{i\in[n]} d_i^2 \qquad \text{converges.}$$



Can be used to approximate how many such graphs exist

Probabilistic method
= Erdős magic!

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.

▷ Sparse GRG:

With W_n weight of uniform vertex, we assume that $W_n \xrightarrow{d} W$ for limiting random variable W.

Preferential attachment

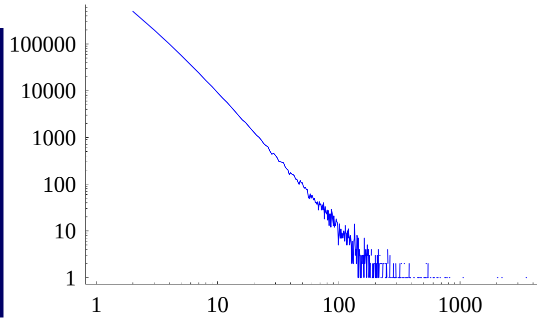
[Albert-Barabási (1999), (49596 citations Google scolar)]

At time n, single vertex is added with m edges emanating from it. Probability that edge connects to specific vertex is proportional to degree of vertex at that time, plus a constant, i.e., 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 $\tau > 2$, where value of τ depends on parameters model as $\tau = 3 + \delta/m$.



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

Preferential attachment

What makes preferential attachment model mathematically difficult?

Part 3:

Local and ‘almost local’ structure of random graphs

Local convergence

Has been proved for many sparse random graph models.

Local convergence implies

- $\triangleright$ one-sided law of large numbers $|\mathscr{C}_{\max}|/n$;
- $\triangleright$ convergence proportion neighborhoods of specific shape;
- $\triangleright$ convergence various other functionals:

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

Local convergence gives good starting point analysis
for many more graph properties.

Preliminaries

$\star$ Graph: $G = (V(G), E(G))$ with $V(G)$ vertex set, $E(G)$ edge set.

$\star$ Rooted graph: (G, v) with G graph and $v \in V(G)$ vertex or root.

$\star$ Neighbourhood: r-neighbourhood $B_r^{(G)}(v)$ of $v \in V(G)$ is rooted subgraph induced by all vertices at distance at most r from root v.

$\star$ Isomorphisms: Two rooted graphs $(G_1, v_1), (G_2, v_2)$ are isomorphic when there is bijection $\phi\colon V(G_1) \to V(G_2)$ mapping edges to edges and root to root. Denoted $(G_1, v_1) \simeq (G_2, v_2)$.

$\star$ Metric: Distance of rooted connected graphs $(G_1, v_1), (G_2, v_2)$ is

$$\mathrm{d}_{\mathscr{G}_\star}\big((G_1, v_1), (G_2, v_2)\big) = \frac{1}{1 + R^\star},$$

where $R^\star$ is largest value for which $B_r^{(G_1)}(v_1) \simeq B_r^{(G_2)}(v_2)$, and $\mathscr{G}_\star$ is space of rooted graphs modulo isomorphisms.

[Metric space well defined: $\mathscr{G}_\star$ Polish space under this metric [RGCNII, Chapter 2].]

Local convergence

$\star$ Random graph sequence is $(G_n)_{n\geq 1}$ satisfying $|V(G_n)| \to \infty$.
[Will often take $V(G_n) = [n] \equiv \{1, \ldots, n\}$.]

$\star$ Local weak convergence holds when, with o_n chosen uniformly from $V(G_n)$,

$$\mathbb{E}[h(G_n, o_n)] = \frac{1}{|V(G_n)|} \sum_{v \in V(G_n)} \mathbb{E}[h(G_n, v)] \to \mathbb{E}_{\bar{\mu}}\big[h(G, o)\big],$$

for any bounded and continuous functions $h \colon \mathscr{G}_\star \to \mathbb{R}$ and $\bar{\mu}$ some probability measure on $\mathscr{G}_\star$. [Benjamini-Schramm (2001), Aldous-Steele (2004).]

$\star$ Local convergence in probability holds when, instead,

$$\mathbb{E}[h(G_n, o_n) \mid G_n] = \frac{1}{|V(G_n)|} \sum_{v \in V(G_n)} h(G_n, v) \xrightarrow{\mathbb{P}} \mathbb{E}_{\mu}\big[h(G, o)\big],$$

for any bounded and continuous function $h \colon \mathscr{G}_\star \to \mathbb{R}$ and μ some probability measure on $\mathscr{G}_\star$.

Branching processes

Who of you have heard of branching processes?

Neighbourhoods CM

▷ Important ingredient in proof is description local neighbourhood of uniform vertex $o_n \in [n]$. Its degree has distribution $d_{o_n} \stackrel{d}{=} D_n$.

▷ Assume that $d_{o_n} \xrightarrow{d} D$ and $\mathbb{E}[D_n] \to \mathbb{E}[D]$.

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

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

$$\to \frac{(k+1)}{\mathbb{E}[D]} \mathbb{P}(D = k+1).$$

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

▷ Forward neighbours of neighbours of o_n are close to i.i.d. Also forward neighbours of forward neighbours approximately i.i.d. etc.

▷ Conclusion: Neighbourhood looks like branching process with offspring distribution $D^\star - 1$ (except root has offspring D).

Local convergence CM

Theorem 1. Consider $\mathrm{CM}_n(\boldsymbol{d})$ where degrees satisfy

$$D_n = d_{o_n} \xrightarrow{d} D, \qquad \text{and} \qquad \mathbb{E}[D_n] \to \mathbb{E}[D].$$

Then $\mathrm{CM}_n(\boldsymbol{d})$ converges in probability to unimodular branching process where root has offspring D, while every other vertex has independent offspring $D^\star - 1$.

Proof. Indicate idea on board.

Erdős-Rényi

What is local limit of the Erdős-Rényi random graph with $p = \lambda/n$?

Local convergence GRG

Theorem 2. Consider $\mathrm{GRG}_n(\boldsymbol{w})$ where weights satisfy

$$W_n = w_{o_n} \xrightarrow{d} W, \qquad \text{and} \qquad \mathbb{E}[W_n] \to \mathbb{E}[W].$$

Then $\mathrm{GRG}_n(\boldsymbol{w})$ converges in probability to unimodular branching process where root has offspring $D \sim \mathsf{Poi}(W)$, while every other vertex has independent offspring $\mathsf{Poi}(W^\star)$, where

$$\mathbb{P}(W^\star \leq w) = \frac{\mathbb{E}[W \mathbb{1}_{\{W \leq w\}}]}{\mathbb{E}[W]}.$$

▷ Implies result for Erdős-Rényi random graph.

Global clustering

Recall that the (global) clustering coefficient is defined by

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

Proportion of friends that are friends of one another.
Can be rewritten as

$$C_{G_n} = \frac{\frac{1}{n}\sum_{v\in[n]} \Delta_v}{\frac{1}{n}\sum_{v\in[n]} \binom{d_v}{2}},$$

where Δ_v is number of triangles containing v.

$$\boxed{\text{Is global clustering coefficient } C_{G_n} \text{ local?}}$$

Global clustering

$\star$ Let C_{G_n} denote global clustering coefficient. $[|V(G_n)| = n.]$

Theorem 3. Let $(G_n)_{n\geq 1}$ be graph sequence that converges locally in probability to $(G, o) \sim \mu$ as $|V(G_n)| = n \to \infty$. Then

$$C_{G_n} = \frac{\frac{1}{n}\sum_{v\in[n]} \Delta_v}{\frac{1}{n}\sum_{v\in[n]} \binom{d_v}{2}} \xrightarrow{\mathbb{P}} \frac{\mathbb{E}_\mu[\Delta_o]}{\mathbb{E}_\mu\Big[\binom{d_o}{2}\Big]},$$

when $D_n = d_{o_n}$ is such that $(D_n^2)_{n\geq 1}$ is uniformly integrable.

Proof. We note that

$$\frac{1}{n}\sum_{v\in[n]} \binom{d_v}{2} = \mathbb{E}\Big[\binom{d_{o_n}}{2} \mid G_n\Big], \qquad \frac{1}{n}\sum_{v\in[n]} \Delta_v = \mathbb{E}\Big[\Delta_{o_n} \mid G_n\Big].$$

Problem: $(G, o) \mapsto \binom{d_o}{2}$ and $(G, o) \mapsto \Delta_o$ unbounded functions in $\mathscr{G}_\star$.

However, when $(D_n^2)_{n\geq 1}$ is uniformly integrable, we can truncate the sum to bounded values, giving continuous functions. $\square$

Local clustering

Local clustering coefficient is defined by

$$\bar{C}_{G_n} = \frac{1}{n} \sum_{v \in [n]} \frac{\Delta_v}{\binom{d_v}{2}}.$$

Is local clustering coefficient $\bar{C}_{G_n}$ local?

Local clustering

$\star$ Let $\bar{C}_{G_n}$ denote local clustering coefficient. $[|V(G_n)| = n.]$

Theorem 4. Let $(G_n)_{n\geq 1}$ be graph sequence that converges locally in probability to $(G,o) \sim \mu$ as $|V(G_n)| = n \to \infty$. Then

$$\bar{C}_{G_n} = \frac{1}{n} \sum_{v\in[n]} \frac{\Delta_v}{\frac{1}{n}\sum_{v\in[n]} \binom{d_v}{2}} \xrightarrow{\mathbb{P}} \mathbb{E}_\mu\left[\frac{\Delta_o}{\binom{d_o}{2}}\right].$$

Proof. We note that

$$\frac{1}{n}\sum_{v\in[n]} \binom{d_v}{2} = \mathbb{E}\left[\frac{\Delta_{o_n}}{\binom{d_{o_n}}{2}} \mid G_n\right],$$

where

$$(G,o) \mapsto \frac{\Delta_o}{\binom{d_o}{2}}$$

is bounded continuous function in $\mathscr{G}_\star$. $\square$

Assortativity coefficient

Recall that assortativity coefficient is given by

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

Correlation between degrees at either end of edge.

Is assortativity coefficient ρ_{G_n} local?

Assortativity

$\star$ Let ρ_{G_n} denote assortativity coefficient. $[|V(G_n)| = n.]$

Theorem 5. Let $(G_n)_{n\geq 1}$ be graph sequence that converges locally in probability to $(G, o) \sim \mu$ as $|V(G_n)| = n \to \infty$. Then

$$\rho_{G_n} \xrightarrow{\mathbb{P}} \frac{\mathbb{E}_\mu[d_o^2 d_V] - \mathbb{E}_\mu[d_o^2]^2/\mathbb{E}_\mu[d_o]}{\mathbb{E}_\mu[d_o^3] - \mathbb{E}_\mu[d_o^2]^2/\mathbb{E}_\mu[d_o]},$$

when $D_n = d_{o_n}$ is such that $(D_n^3)_{n\geq 1}$ is uniformly integrable, and where V is a neighbor of o chosen uar.

Proof. Blackboard. $\square$

Number components

$\star$ Let K_n denote number of connected components. $[|V(G_n)| = n.]$

Is K_n/n local?

Number components

$\star$ Let K_n denote number of connected components. $[|V(G_n)| = n.]$

Theorem 6. Let $(G_n)_{n\geq 1}$ be graph sequence that converges locally in probability to $(G, o) \sim \mu$ as $|V(G_n)| = n \to \infty$. Then

$$\frac{K_n}{n} \xrightarrow{\mathbb{P}} \mathbb{E}_\mu\left[\frac{1}{|\mathscr{C}(o)|}\right].$$

Proof. We note that

$$\frac{K_n}{n} = \frac{1}{n}\sum_{v\in[n]} \frac{1}{|\mathscr{C}(v)|} = \mathbb{E}\left[\frac{1}{|\mathscr{C}(o_n)|} \mid G_n\right].$$

Since

$$(G, o) \mapsto \frac{1}{|\mathscr{C}(o)|}$$

is bounded continuous functional in $\mathscr{G}_\star$, the claim follows. $\square$

Giant component

$\star$ Let $\mathscr{C}_{\max}$ denote connected component of maximal size.

$[|V(G_n)| = n.]$

Is $|\mathscr{C}_{\max}|/n$ local?

Giant component

⋆ Let $\mathscr{C}_{\max}$ denote connected component of maximal size.

Theorem 7. Let $(G_n)_{n\geq 1}$ be graph sequence that converges locally in probability to $(G, o) \sim \mu$ as $|V(G_n)| = n \to \infty$. Then

$$\frac{|\mathscr{C}_{\max}|}{n} \xrightarrow{\mathbb{P}} \mu(|\mathscr{C}(o)| = \infty)$$

precisely when

$$\lim_{k\to\infty} \limsup_{n\to\infty} \frac{1}{n^2}\mathbb{E}\Big[\#\big\{(x, y)\colon |\mathscr{C}(x)| \geq k, |\mathscr{C}(y)| \geq k, x \not\longleftrightarrow y\big\}\Big] = 0.$$

Proof. **Hero** proof is number of vertices in clusters of size at least k

$$Z_{\geq k} = \sum_{v\in[n]} \mathbb{1}_{\{|\mathscr{C}(v)|\geq k\}}.$$

Giant 'almost local'

Note that

$$(G, o) \mapsto \mathbb{1}_{\{|\mathscr{C}(o)| \geq k\}}$$

is bounded continuous functional in $\mathscr{G}_\star$. Thus,

$$\frac{Z_{\geq k}}{n} = \mathbb{E}\Big[\mathbb{1}_{\{|\mathscr{C}(o_n)| \geq k\}} \mid G_n\Big] \xrightarrow{\mathbb{P}} \mu(|\mathscr{C}(o)| \geq k).$$

$\star$ Upper bound: Either $|\mathscr{C}_{\max}| < k$, or $|\mathscr{C}_{\max}| \leq Z_{\geq k}$. Thus,

$$\frac{|\mathscr{C}_{\max}|}{n} \leq \frac{k}{n} + \frac{Z_{\geq k}}{n} \xrightarrow{\mathbb{P}} \mu(|\mathscr{C}(o)| \geq k) \to \mu(|\mathscr{C}(o)| = \infty).$$

$\star$ Proves upper bound without any assumption other than

local convergence in probability!

$\square$

Giant 'almost local'

Key equation: $(|\mathscr{C}_{(i)}|)_{i\geq 1}$ are ordered connected component sizes:

$$\Big(\frac{Z_{\geq k}}{n}\Big)^2 = \sum_{i\geq 1} \frac{|\mathscr{C}_{(i)}|^2}{n^2}\mathbb{1}_{\{|\mathscr{C}_{(i)}|\geq k\}} + \frac{1}{n^2}\sum_{i\neq j} |\mathscr{C}_{(i)}||\mathscr{C}_{(j)}|\mathbb{1}_{\{|\mathscr{C}_{(i)}|,|\mathscr{C}_{(j)}|\geq k\}}.$$

$\star$ First term rhs is bounded above by

$$\frac{|\mathscr{C}_{\max}|}{n}\sum_{i\geq 1}\frac{|\mathscr{C}_{(i)}|}{n}\mathbb{1}_{\{|\mathscr{C}_{(i)}|\geq k\}} = \frac{|\mathscr{C}_{\max}|}{n}\frac{Z_{\geq k}}{n}.$$

$\star$ Second term rhs satisfies, by 'almost local' condition,

$$\frac{1}{n^2}\#\big\{(x,y)\colon |\mathscr{C}(x)|\geq k, |\mathscr{C}(y)|\geq k, x \not\longleftrightarrow y\big\} = o_{\mathbb{P},k}(1),$$

where latter denotes random variable $X_{n,k}$ s.t.

$$\lim_{k\to\infty}\limsup_{n\to\infty}\mathbb{P}(|X_{n,k}|\geq \varepsilon) = 0.$$

Giant 'almost local'

⋆ In terms of this notion, $Z_{\geq k}/n \xrightarrow{\mathbb{P}} \mu(|\mathscr{C}(o)| \geq k)$ implies

$$\frac{Z_{\geq k}}{n} = \mu(|\mathscr{C}(o)| = \infty) + o_{\mathbb{P},k}(1).$$

Thus, from key equation,

$$\Big[\mu(|\mathscr{C}(o)| = \infty) + o_{\mathbb{P},k}(1)\Big]^2 \leq \frac{|\mathscr{C}_{\max}|}{n}\Big[\mu(|\mathscr{C}(o)| = \infty) + o_{\mathbb{P},k}(1)\Big] + o_{\mathbb{P},k}(1),$$

so that

$$\frac{|\mathscr{C}_{\max}|}{n} \geq \mu(|\mathscr{C}(o)| = \infty) + o_{\mathbb{P},k}(1).$$

⋆ Taking $\lim_{k\to\infty} \limsup_{n\to\infty}$ proves lower bound on $|\mathscr{C}_{\max}|$. $\square$

Phase transition CM

Theorem 8 [Molloy-Reed (95), Janson-Luczak (09), Bollobás-Riordan (09), vdH (21).]
Let $\nu = \mathbb{E}[D(D-1)]/\mathbb{E}[D] > 1$. Then, largest component CM has size $\zeta n(1 + o_{\mathbb{P}}(1))$ with $\zeta \in (0, 1)$ for $\nu > 1$.

Here ν is expected number forward neighbours of side uniform edge.

▷ Giant is 'almost local': ζ has interpretation of

$\zeta = \mu($unimodular branching process root offspring D survives.$)$

▷ When $\mathbb{E}[D(D-1)] = \infty$, there always is a giant irrespective of how small $\mathbb{E}[D]$ is:

Robustness of the giant.

Erdős-Rényi

How large is the giant of the Erdős-Rényi random graph with $p = \lambda/n$?

Phase transition GRG

Theorem 9 [Bollobás-Janson-Riordan (07), vdH (24).]
Let $\nu = \mathbb{E}[W^2]/\mathbb{E}[W] > 1$. Then, largest component GRG has size $\zeta n(1 + o_{\mathbb{P}}(1))$ with $\zeta \in (0, 1)$ for $\nu > 1$.

$\triangleright$ Implies result for Erdős-Rényi random graph.

Giant is 'almost local': ζ is survival probability local limit.

$\triangleright$ When $\mathbb{E}[W^2] = \infty$, there always is a giant irrespective of how small $\mathbb{E}[W]$ is:

Robustness of the giant.

Part 4:

Small- and ultra-small-world random graphs

Graph distances CM

Study graph distances between uniform pair of vertices to study

small-world properties.

Theorem 10. [vdH-Hooghiemstra-Van Mieghem 05]
In lightly-inhomogeneous setting, graph distances grow

logarithmically.

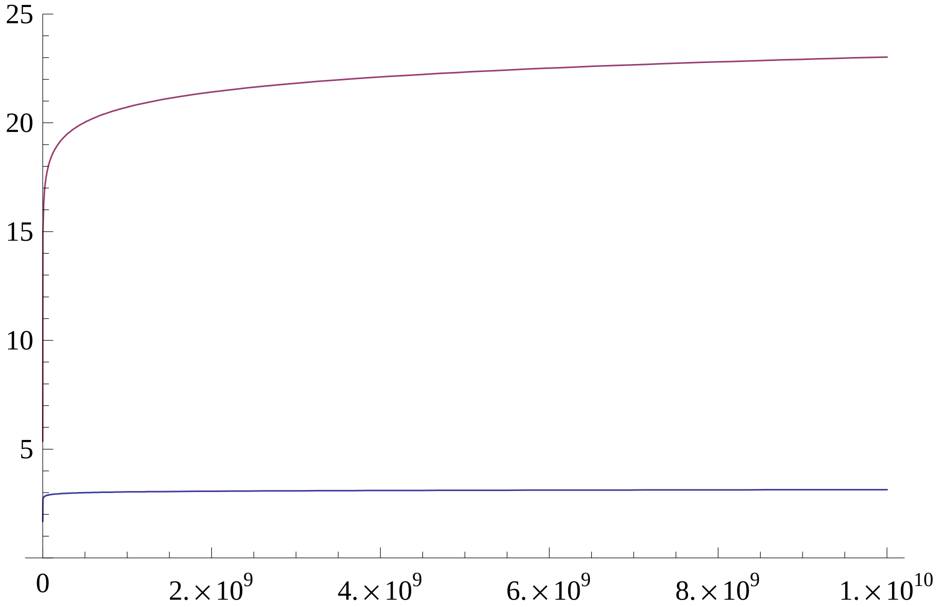
Theorem 11. [vdH-Hooghiemstra-Znamenski 07, Norros+Reittu 04]
In highly inhomogeneous setting, graph distances grow

doubly logarithmically.

▷ Technically: lightly-inhomogeneous setting means that

second moment of degrees is finite.

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



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

Graph distances CM

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

Theorem 10. [Theorem II.7.1]. When $D_n \xrightarrow{d} D, \mathbb{E}[D_n^2] \to \mathbb{E}[D^2]$ with $\nu = \mathbb{E}[D(D-1)]/\mathbb{E}[D] > 1$, conditionally on $H_n < \infty$,

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

$\triangleright$ For i.i.d. degrees having at most power-law tails, fluctuations are bounded and do not converge in distribution.

Theorem 11. [Thm II.7.2]. When $D_n \xrightarrow{d} D$, $\mathbb{E}[D_n^2] \to \mathbb{E}[D^2]$, and degree power-law exponent satisfies $\tau \in (2,3)$, conditionally on $H_n < \infty$,

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

$\triangleright$ vdH-Komjáthy16: For power-law tails, fluctuations are again bounded and do not converge in distribution.

Distances PA models

▷ Results CM and GRG are very alike, with CM having more general behavior (e.g., connectivity). Sign of 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.

▷ Signs point in this direction.

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

time dependence.

Distances PA models

Theorem 12 [Bol-Rio 04]. For $m \geq 2$ and $\tau = 3$,

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

Theorem 13 [Dom-vdH-Hoo 10, vdH+Zhu 25]. For $m \geq 2$, $\tau \in (3, \infty)$, exists $\nu > 1$,

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

Theorem 14 [Dommers-vdH-Hoo 10, Der-Mon-Mor 12, Car-Gar-vdH17]. For $m \geq 2, \tau \in (2, 3)$,

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

Universality!

Distances and neighbourhoods

When distances grow logarithmically, how quickly do you think sizes of k-neighbourhood from uniform vertex grow for k large? but not excessively large

Distances and neighbourhoods

When distances grow doubly logarithmically, how quickly do you think sizes of k-neighbourhood from uniform vertex grow for k large? but not excessively large

Neighbourhoods CM

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

▷ Take any of D_{o_n} neighbours a of o_n. Law of number of forward neighbours 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 $D^\star - 1$.

▷ Forward neighbours of neighbours of o_n are close to i.i.d. Also forward neighbours of forward neighbours approximately i.i.d...

▷ Conclusion: Neighbourhood looks like branching process with offspring distribution $D^\star - 1$ (except root has offspring D.)

Local convergence CM

$\triangleright$ Formal definition in terms of local convergence.

$\triangleright$ $\mathbb{E}[D^2] < \infty$: Finite-mean BP, which has exponential growth of generation sizes when $\nu = \mathbb{E}[D^\star - 1] = \mathbb{E}[D(D-1)]/\mathbb{E}[D] > 1$:

$$\nu^{-k} Z_k \xrightarrow{a.s.} Y \in (0, \infty).$$

Thus, graph distances grow like $\log_\nu n$.

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

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

$\triangleright$ Blackboard: Indicate ingredients proof.

Part 5:

Competition and fake news on scale-free random graphs

Competition

▷ Viral marketing aims to use social networks so as to accellerate adoption of novel products.

▷ Observation: Often one product takes almost complete market. Not always product of best quality:

Why?

▷ **Aim:** Explain this phenomenon, and relate it to network structure as well as spreading dynamics.

▷ Setting:
– Model social network as configuration model random graph;
– Model dynamics as competitions spreading through network. Vertices, once occupied by certain type, try to occupy their neighbours at (possibly) random and i.i.d. times.
▷ Speed of type might correspond to quality product.

Explosive setting

Theorem 15. [Deijfen-vdH (2013), vdH (2023)] Fix $\tau \in (2,3)$. Consider competition model, where types compete for territory with i.i.d. traversal times and both types are explosive. Then, each of types wins majority vertices with positive probability:

$$\frac{N_1(n)}{n} \xrightarrow{d} I \in \{0,1\}.$$

Number of vertices for losing type $N_{\rm los}(n)$ converges in distribution:

$$N_{\rm los}(n) \xrightarrow{d} N_{\rm los} \in \mathbb{N}.$$

The winner takes it all, the loser's standing small...

Who wins is determined by location of starting point types:
Location, location, location!

Explosion

⋆ In this setting, turns out that competition spread is explosive:

Both types reach 'infinitely many vertices' in bounded time.

How does this explain 'winner takes it all' phenomenon?

Fake news

▷ Fake news is huge societal problem, causing major disruptions.

▷ Observation: Fake news often spreads quickly through network.

▷ **Aim:** Investigate this phenomenon, and filter out dependence on network structure.

▷ Setting:
– Model social network as configuration model random graph;
– Model fake news dynamics as competing rumour spread through network. Vertices, after hearing fake or correct news, convince their neighbours at random and i.i.d. times.

▷ **Optimistic perspective:**

Once vertices hear correct news, they are immediately convinced by it.

Fake news model

▷ Assign i.i.d. traversal times $((L_e^{\mathcal{F}}, L_e^{\mathcal{R}}))_{e\in E(G)}$ to edges of graph.

▷ Fake news spreads using traversal times $(L_e^{\mathcal{F}})_{e\in E(G)}$.
▷ Correct news spreads from same source after delay d, using traversal times $(L_e^{\mathcal{R}})_{e\in E(G)}$.
▷ Upon hearing correct news, vertices are immediately convinced, and start spreading correct news.

▷ Strong survival: Fake news reaches positive proportion vertices with positive probability.

▷ Weak survival: Fake news reaches a growing number of vertices with positive probability.

What are conditions for fake news survival?

Strong survival

Theorem 16. [vdH-Shneer (2023)] Condition on news originating in giant component configuration model, and assume that degrees are i.i.d. with power-law distribution with $\tau \in (2,3)$.

Fake news survives strongly when age-dependent branching process with offspring distribution $D^{\star} - 1$ and lifetime $L^{\mathcal{F}}$ is explosive.

In explosive setting, fake news can reach positive proportion of vertices, even when on average it spreads slower.

It is hard to kill fake news!

Neighbourhoods CM

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

▷ Take any of D_{o_n} neighbours a of o_n. Law of number of forward neighbours 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 $D^\star - 1$.

▷ Forward neighbours of neighbours of o_n are close to i.i.d. Also forward neighbours of forward neighbours approximately i.i.d...

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

Local convergence CM

$\triangleright$ Formal definition in terms of local convergence.

$\triangleright$ $\mathbb{E}[D^2] < \infty$: Finite-mean BP, which has exponential growth of generation sizes when $\nu = \mathbb{E}[D^\star - 1] = \mathbb{E}[D(D-1)]/\mathbb{E}[D] > 1$:

$$\nu^{-k} Z_k \xrightarrow{a.s.} Y \in (0, \infty).$$

Thus, graph distances grow like $\log_\nu n$. Same true for rumour spread models with n replaced by time.

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

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

$\star$ Rumour spread models tend to be explosive. Type that explodes first wins.

Choice menu

⋆ Other almost local properties random graphs:

What would you like to know?

⋆ How to prove 'giant is almost local' condition?

⋆ Examples of local limits of random graphs:

What is your favourite model?

⋆ Extension to directed random graphs.

Conclusions

▷ Networks useful to interpret real-world phenomena:
centrality measures.

▷ Unexpected commonality networks:
scale free and small worlds.

▷ Random graph models explain properties networks.
Universality?

Example: Local limits often branching processes

Material Networks

▷ Random graphs and complex networks Volumes 1 and 2
`http://www.win.tue.nl/~rhofstad/NotesRGCN.html`
Aimed at master and PhD students in math.

▷ Network science by Albert-László Barabási.
Online book available at `http://networksciencebook.com`
For broad audience, requires critical reading.

▷ Networks: an introduction by Mark E. J. Newman.
For broad audience, gives most definitions of network notions.

▷ The atlas for the aspiring network scientist by Michele Costia.
`https://www.networkatlas.eu/`

▷ NetworkPages for articles on various applications of networks.
`https://www.networkpages.nl`

▷ Many other sources and popular books. Explore yourself!