The Friendship Paradox
for Sparse Random Graphs

Frank den Hollander
Leiden University, The Netherlands

Challenges in Networks, 29 January – 02 February 2024,
International Centre for Theoretical Sciences, Bangalore, India.

§ THE FRIENDSHIP PARADOX

"On Average, Our Friends Have More Friends Than We Do!"

(Scott Feld, 1991)

friendship network

1. For $n \in \mathbb{N}$, let G_n be a finite undirected random graph with n vertices labeled by $[n] = \{1, \ldots, n\}$. Let d_i be the degree of vertex i.

2. Let $\Delta_{i,n}$ be the friendship-bias of vertex i, defined as the difference between the average degree of the neighbours of i and the degree of i itself, i.e.,

$$\Delta_{i,n} = \left[\frac{\sum_{j \in [n]} A_{ij} d_j}{d_i} - d_i\right] \mathbb{1}_{\{d_i \neq 0\}}, \qquad i \in [n],$$

where A is the adjacency matrix of G_n, i.e., A_{ij} is the number of edges between i and j.

3. The quenched and the annealed empirical friendship-bias distribution are

$$\mu_n = \frac{1}{n} \sum_{i \in [n]} \delta_{\Delta_{i,n}}, \qquad \tilde{\mu}_n = \mathbb{E}_n[\mu_n],$$

where $\mathbb{E}_n$ denotes expectation over G_n.

FRIENDSHIP PARADOX

For any graph G_n without self-loops, the average friendship bias is non-negative, i.e.,

$$\Delta_{[n]} = \frac{1}{n} \sum_{i \in [n]} \Delta_{i,n} = \int_{\mathbb{R}} x \mu_n(\mathrm{d}x) \geq 0.$$

Equality holds if and only if all the connected components of G_n are regular.

PROOF

Write

$$\begin{aligned}\Delta_{[n]} &= \frac{1}{n} \sum_{\substack{i\in[n] \\ d_i\neq 0}} \sum_{\substack{j\in[n] \\ d_j\neq 0}} A_{ij} \left(\frac{d_j}{d_i} - 1 \right) \\ &= \frac{1}{2n} \sum_{\substack{i\in[n] \\ d_i\neq 0}} \sum_{\substack{j\in[n] \\ d_j\neq 0}} A_{ij} \left(\sqrt{\frac{d_j}{d_i}} - \sqrt{\frac{d_i}{d_j}} \right)^2 \\ &\geq 0.\end{aligned}$$

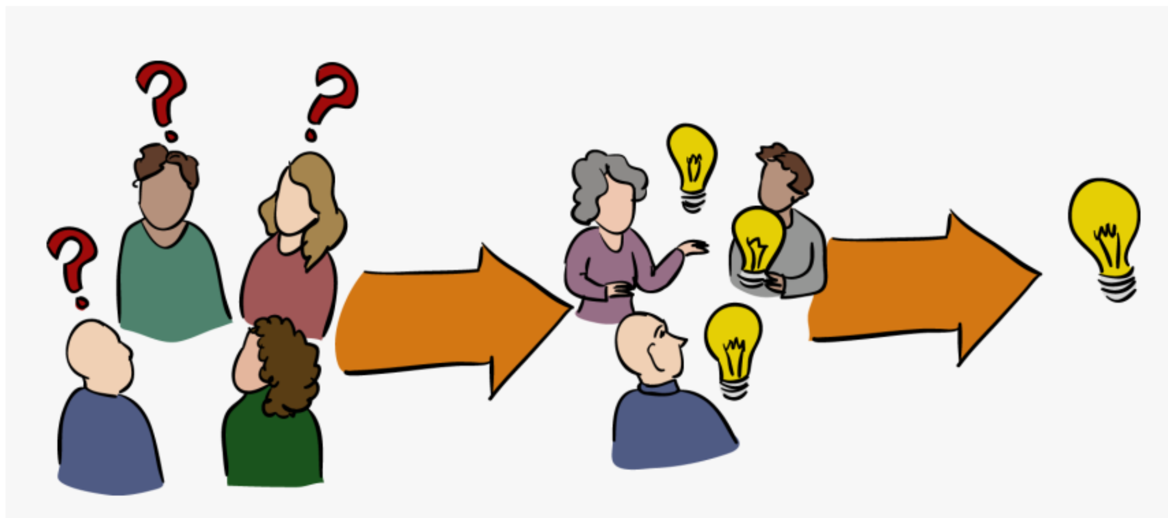
Here, the first equality uses that there are no self-loops, and the second equality uses that A is symmetric. Equality holds if and only if $i \mapsto d_i$ is constant on each connected component. □

§ KEY QUESTION

Can we quantify the friendship paradox and analyse the quantification for key examples of sparse random graphs?

R.S. Hazra, F. den Hollander, A. Parvaneh, preprint 22/12/2023
arXiv:2312.15105[math.PR]

I. CONVERGENCE



§ LOCAL CONVERGENCE

To quantify the friendship paradox we need the notion of local convergence.

> Informally, we require that what G_n looks like relative to a randomly drawn vertex converges as $n \to \infty$ to a random graph G_∞ with root vertex ϕ, denoted by the pair (G_∞, ϕ).

Let $\mathscr{G}$ be the set of connected locally finite rooted graphs, equipped with a topology suitable for local convergence.

Let $(G_n)_{n \in \mathbb{N}}$ be a sequence of finite random graphs in $\mathscr{G}$. For $n \in \mathbb{N}$, let $\mathscr{C}(G_n, U_n)$ denote the connected component of a uniformly chosen vertex U_n in G_n, viewed as a rooted graph with root vertex U_n.

DEFINITION

(a) G_n converges locally weakly to (G_∞, ϕ) with law $\tilde{\nu}$ if, for every bounded and continuous function $h\colon \mathscr{G} \to \mathbb{R}$,

$$\mathbb{E}_{\tilde{\nu}_n}\Big[h\Big(\mathscr{C}(G_n, U_n)\Big)\Big] \longrightarrow \mathbb{E}_{\tilde{\nu}}\Big[h\Big((G_\infty, \phi)\Big)\Big],$$

where $\mathbb{E}_{\tilde{\nu}_n}$ is expectation with respect to the random vertex U_n and the random graph G_n with joint law $\tilde{\nu}_n$, while $\mathbb{E}_{\tilde{\nu}}$ is expectation with respect to (G_∞, ϕ) with law $\tilde{\nu}$.

(b) G_n converges locally in probability to (G_∞, ϕ) with law $\tilde{\nu}$ if, for every bounded and continuous function $h\colon \mathscr{G} \to \mathbb{R}$,

$$\mathbb{E}_{\tilde{\nu}_n}\Big[h\Big(\mathscr{C}(G_n, U_n)\Big) \;\Big|\; G_n\Big] \xrightarrow{\mathbb{P}} \mathbb{E}_{\tilde{\nu}}\Big[h\Big((G_\infty, \phi)\Big)\Big].$$

§ CONVERGENCE THEOREMS

Consider an almost surely locally finite rooted random tree (G_∞, ϕ) with law $\bar{\mu}$. Let d_ϕ be the degree of ϕ, and let d_j be the size of the offspring of neighbour j of ϕ. Set

$$\Delta_\phi = \left[\frac{1}{d_\phi}\sum_{j=1}^{d_\phi}(d_j+1) - d_\phi\right]\mathbb{1}_{\{d_\phi \neq 0\}},$$

and let μ be the law of Δ_ϕ.

THEOREM 1 *If G_n converges locally in probability to (G_∞, ϕ), then $\mu_n \Rightarrow \mu$ in probability.*

THEOREM 2 *If G_n converges locally weakly to (G_∞, ϕ), then $\tilde{\mu}_n \Rightarrow \mu$.*

According to Theorem 1, for any locally tree-like random graph,

$$\mu_n([0,\infty)) \xrightarrow{\mathbb{P}} \mu([0,\infty)) \quad \text{when} \quad \mu(\{0\}) = 0.$$

The former equals the average number of vertices with a non-negative friendship-bias. We expect that in most cases the same is true when $\mu(\{0\}) \neq 0$.

DEFINITION

The friendship paradox is called

$$\begin{aligned} &\text{significant} && \text{when} \quad \mu([0,\infty)) \geq \tfrac{1}{2}, \\ &\text{insignificant} && \text{when} \quad \mu([0,\infty)) < \tfrac{1}{2}. \end{aligned}$$

II. FOUR EXAMPLES

R. van der Hofstad, *Random Graphs and Complex Networks*, Cambridge Series in Statistical and Probabilistic Mathematics, Volume 1 (2017), Volume 2 (2024).

§ HOMOGENEOUS Erdős-Rényi RANDOM GRAPH

For $\lambda \in (0,\infty)$ and $n \in \mathbb{N}$, let $\mathrm{HER}_n(\frac{\lambda}{n} \wedge 1)$ be the random graph in which each pair of distinct vertices in $i,j \in [n]$ is independently connected by an edge with probability $\frac{\lambda}{n} \wedge 1$.

It is well known that $\mathrm{HER}_n(\frac{\lambda}{n} \wedge 1)$ as $n \to \infty$ converges locally in probability to a Galton-Watson tree with offspring distribution Poisson(λ).

THEOREM 3a

(a) *For every* $\lambda \in (0,\infty)$,

$$\mu([0,\infty)) = \sum_{k \in \mathbb{N}_0} \frac{\mathrm{e}^{-\lambda}\lambda^k}{k!} \sum_{l \geq k(k-1)} \frac{\mathrm{e}^{-\lambda k}(\lambda k)^l}{l!}.$$

(b)

$$\lim_{\lambda \downarrow 0} \mu([0,\infty)) = 1, \qquad \lim_{\lambda \to \infty} \mu([0,\infty)) = \tfrac{1}{2}.$$

THEOREM 3b

$$\lim_{\lambda\downarrow 0} \mathbb{E}_{\bar\mu}[\Delta_\phi] = 0, \quad \lim_{\lambda\downarrow 0} \mathbb{E}_{\bar\mu}[\Delta_\phi^2] = 0,$$

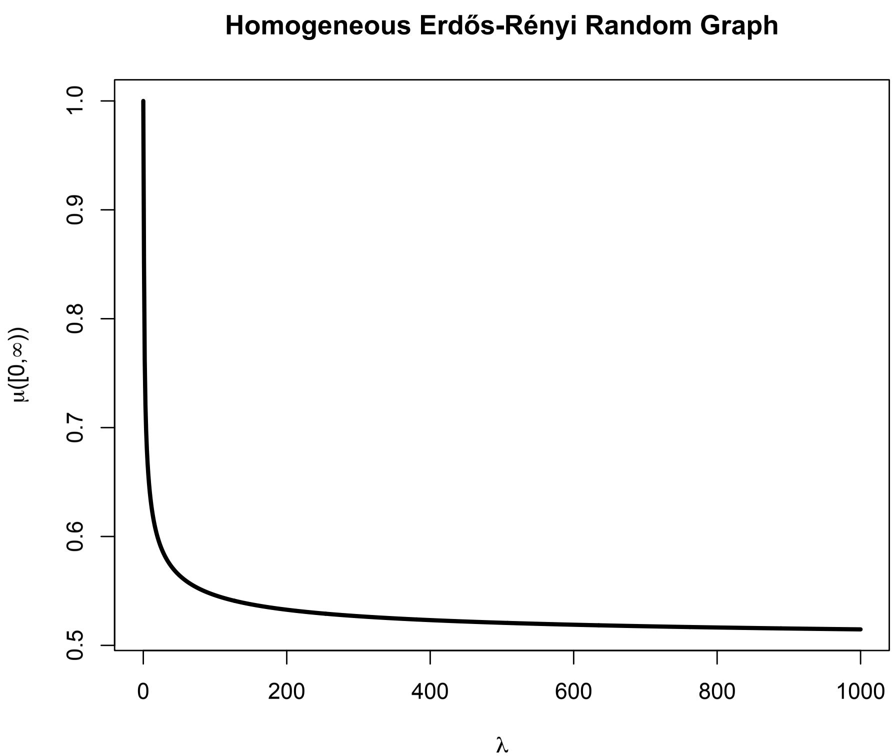
$$\lim_{\lambda\to\infty} \mathbb{E}_{\bar\mu}[\Delta_\phi] = 1, \quad \lim_{\lambda\to\infty} \mathbb{E}_{\bar\mu}[\Delta_\phi^2] = \infty.$$

THEOREM 3c

For every $\lambda \in (0,\infty)$,

$$\mathbb{P}_{\bar\mu}\{\Delta_\phi \geq x\} \sim \frac{\lambda\, e^{-2\lambda}}{\sqrt{2\pi x}} \exp\left\{ -x \log\left(\frac{x}{\lambda e}\right)\right\}, \quad x \to \infty.$$

Is $\mathrm{HER}_n(\frac{\lambda}{n} \wedge 1)$ significant for all $\lambda \in (0, \infty)$? Numerical computation suggests yes!



§ INHOMOGENEOUS Erdős-Rényi RANDOM GRAPH

Let $\mathcal{F}$ be the class of non-constant Riemann integrable functions $f\colon [0,1] \to (0,\infty)$ satisfying

$$M_- = \inf_{x\in[0,1]} f(x) > 0, \qquad M_+ = \sup_{x\in[0,1]} f(x) < \infty.$$

For $\lambda \in (0,\infty)$ and $n \in \mathbb{N}$, let $\mathrm{IER}_n(\lambda f)$ be the random graph in which each pair of distinct vertices $i,j \in [n]$ is independently connected by an edge with probability

$$\frac{\lambda}{n} f(\tfrac{i}{n}) f(\tfrac{j}{n}) \wedge 1.$$

It is well known that $\mathrm{IER}_n(\lambda f)$ as $n \to \infty$ converges locally in probability to a unimodular multi-type marked Galton-Watson tree.

THEOREM 4a

Let

$$\beta_x = \int_{[0,1]} \mathrm{d}y\, f^x(y), \qquad x \in (0,\infty).$$

Let $c_k(z)$ *be the probability that a* POISSON($\lambda\beta_1 z$) *random variable takes the value* $k \in \mathbb{N}_0$*.*

(a) *For every* $f \in \mathcal{F}$ *and* $\lambda \in (0,\infty)$*,*

$$\mu([0,\infty)) = \sum_{k\in\mathbb{N}_0} \left(\int_{[0,1]} \mathrm{d}x\, c_k(f(x)) \right)$$
$$\times \left[\beta_1^{-k} \prod_{j=1}^{k} \int_{[0,1]} \mathrm{d}x_j f(x_j) \sum_{l \geq k(k-1)} c_l \left(\sum_{j=1}^{k} f(x_j) \right) \right].$$

(b) *For every* $f \in \mathcal{F}$*,*

$$\lim_{\lambda\downarrow 0} \mu([0,\infty)) = 1, \qquad \lim_{\lambda\to\infty} \mu([0,\infty)) \geq \tfrac{1}{2}.$$

THEOREM 4b

For every $f \in \mathcal{F}$,

$$\lim_{\lambda \downarrow 0} \mathbb{E}_{\bar{\mu}}\big[\Delta_\phi\big] = 0, \qquad \lim_{\lambda \downarrow 0} \mathbb{E}_{\bar{\mu}}\big[\Delta_\phi^2\big] = 0,$$

$$\lim_{\lambda \to \infty} \mathbb{E}_{\bar{\mu}}\big[\Delta_\phi\big] = \infty, \qquad \lim_{\lambda \to \infty} \mathbb{E}_{\bar{\mu}}\big[\Delta_\phi^2\big] = \infty.$$

THEOREM 4c

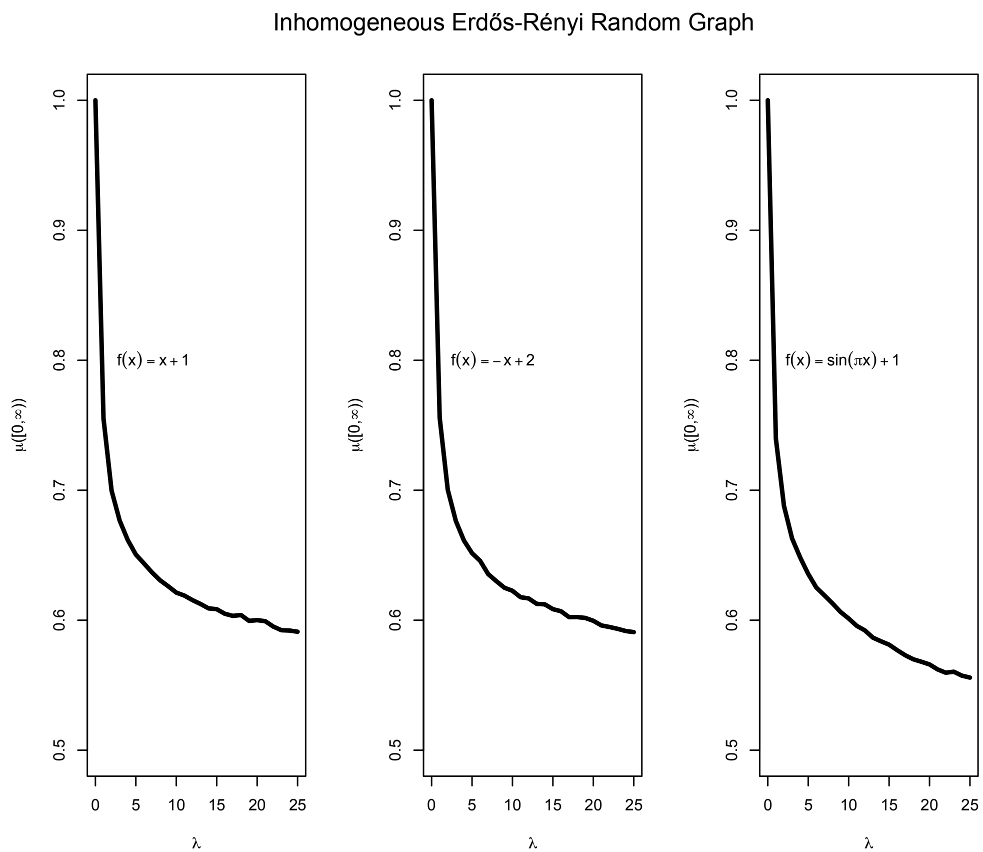
For every $f \in \mathcal{F}$ *and* $\lambda \in (0, \infty)$,

$$\mu([x, \infty)) \asymp \frac{1}{\sqrt{x}} \beta_x \exp\Big\{ -x \log\Big(\frac{x}{\lambda \beta_1 \mathrm{e}}\Big)\Big\}, \qquad x \to \infty.$$

(a) *If* $|\{y \colon f(y) = M_+\}| > 0$, *then* $\beta_x \asymp M_+^x$ *as* $x \to \infty$.

(b) *If* $f(y_\star) = M_+$ *and* $f(y_\star) - f(y) \asymp |y - y_\star|^\alpha$, $y \to y_\star$, *for some* $y_\star \in [0, 1]$ *and* $\alpha \in (0, \infty)$, *and* f *is bounded away from* M_+ *outside any neighbourhood of* $y_\star$, *then* $\beta_x \asymp x^{-1/\alpha} M_+^x$ *as* $x \to \infty$.

Is $\mathrm{IER}_n(\lambda f)$ significant for all $\lambda \in (0, \infty)$ irrespective of $f \in \mathcal{F}$? Numerical computation suggests yes!



§ CONFIGURATION MODEL

For $n \in \mathbb{N}$ and $\mathbf{d}_n \in \mathbb{N}_0^n$, let $\mathrm{CM}_n(\mathbf{d}_n)$ be the random graph drawn uniformly at random from the set of all graphs with degree sequence $\mathbf{d}_n$.

Let D_n be the degree of a uniformly chosen vertex. It is well known that if

$$D_n \Rightarrow D,\ \mathbb{E}[D_n] \to \mathbb{E}[D] \text{ with } \mathbb{P}\{D > 0\} = 1,\ \mathbb{E}[D] < \infty,$$

then $\mathrm{CM}_n(\mathbf{d}_n)$ as $n \to \infty$ converges locally in probability to a random tree that has offspring distribution $p = (p_k)_{k \in \mathbb{N}_0}$ at ϕ and $p^\star = (p_k^\star)_{k \in \mathbb{N}_0}$ elsewhere, where

$$p_k = \mathbb{P}\{D = k\}, \qquad p_k^\star = \frac{(k+1)p_{k+1}}{\mathbb{E}[D]}.$$

The configuration model has a version without self-loops that also converges locally in probability to the same limit.

A special case is

$$p_k = \frac{1}{\zeta(\tau)} k^{-\tau}, \quad k \in \mathbb{N}, \qquad \tau \in (2, \infty),$$

to which we refer as the Riemann offspring distribution. Here, the tail exponent τ modulates the presence of hubs.

THEOREM 5a

For Riemann offspring distribution,

$$\mu([0,\infty)) > \tfrac{1}{2} \quad \forall\, \tau \in (2,\infty),$$

$$\mu([0,\infty)) \to 1, \quad \tau \downarrow 2, \qquad \mu([0,\infty)) \to 1, \quad \tau \to \infty.$$

THEOREM 5b

For general offspring distribution,

$$\mathbb{E}_{\bar{\mu}}\big[\Delta_\phi\big] = \frac{\mathrm{Var}(D)}{\mathbb{E}[D]}.$$

If also $\mathbb{E}[D^2] < \infty$*, then*

$$\mathbb{E}_{\bar{\mu}}\big[\Delta_\phi^2\big] = \frac{\big(\mathbb{E}[D^3]\,\mathbb{E}[D] - (\mathbb{E}[D^2])^2\big)\,\mathbb{E}[D^{-1}] + \mathbb{E}[D^2]\,\mathrm{Var}(D)}{(\mathbb{E}[D])^2}$$

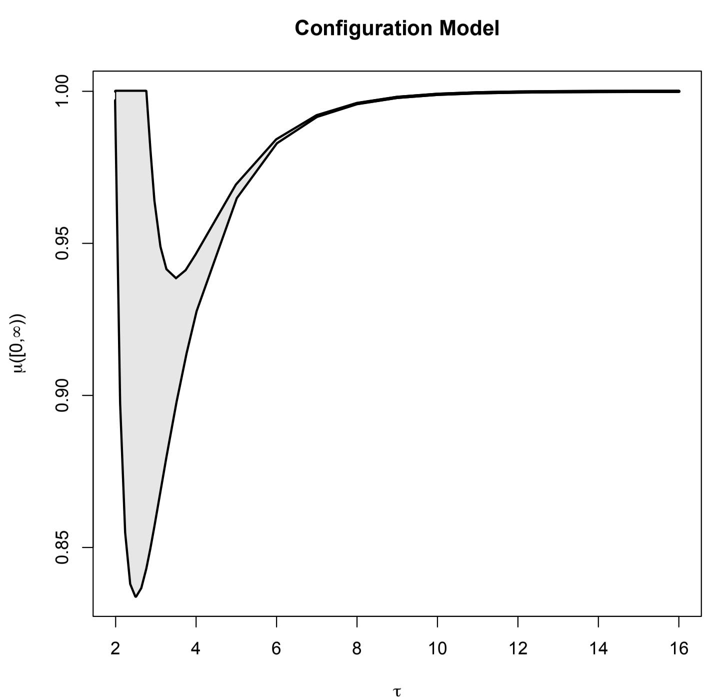
In particular, $\mathrm{Var}_{\bar{\mu}}(\Delta_\phi) \geq \mathrm{Var}(D)$.

THEOREM 5c

For Riemann offspring distribution,

$$\mathbb{P}_{\bar{\mu}}\{\Delta_\phi \geq x\} \asymp x^{-(\tau-2)}, \qquad x \to \infty.$$

Sharp analytic bounds for significance can be derived for Riemann offspring distribution.



§ PREFERENTIAL ATTACHMENT MODEL

For $\delta \geq -1$ and $n \in \mathbb{N}$, let $\mathsf{PAM}_n^{(\delta)}$ be the random graph obtained iteratively as follows.

- $\mathsf{PAM}_1^{(\delta)}$ consists of a single vertex v_1 with a single self-loop.
- For $n \in \mathbb{N}$, suppose that $v_1, \ldots, v_n$ are the vertices of $\mathsf{PAM}_n^{(\delta)}$ with degrees $d_{1,n}, \ldots, d_{n,n}$. Then $\mathsf{PAM}_{n+1}^{(\delta)}$ is obtained by adding a single vertex v_{n+1} with a single edge, and attaching this edge with probability

$$\mathbb{P}\Big\{v_{n+1} \to v_i \;\Big|\; \mathsf{PAM}_n^{(\delta)}\Big\} = \begin{cases} \dfrac{d_{i,n} + \delta}{n(2+\delta) + (1+\delta)}, & \text{if } i \in [n], \\[2ex] \dfrac{1 + \delta}{n(2+\delta) + (1+\delta)}, & \text{if } i = n+1. \end{cases}$$

It is well known that $\mathsf{PAM}_n^{(\delta)}$ converges locally in probability to a multi-type discrete-time branching process called the Pólya point tree.

The preferential attachment model has a version without self-loops that also converges locally in probability to the same limit.

THEOREM 5a

$$\mu([0,\infty)) \geq \tfrac{1}{2} \quad \forall\, \delta > -1, \qquad\qquad \lim_{\delta \downarrow -1} \mu([0,\infty)) = 1.$$

THEOREM 5b

Abbreviate

$$p_\delta = \mathbb{E}_{\bar{\mu}}[d_\phi^{-1}] = \sum_{k \in \mathbb{N}} \frac{(2+\delta)\Gamma(3+2\delta)\Gamma(k+\delta)}{k\Gamma(1+\delta)\Gamma(k+3+2\delta)}.$$

Then

$$\mathbb{E}_{\bar{\mu}}\Big[\Delta_\phi\Big] \begin{cases} \in \frac{2+\delta}{\delta}(\frac{1}{2}+p_\delta) + \Big[-(1-p_\delta), 0\Big], & \text{if } \delta \in (0,\infty), \\ = \infty, & \text{if } \delta \in (-1,0], \end{cases}$$

and $\mathbb{E}_{\bar{\mu}}[\Delta_\phi^2] < \infty$ *if and only if* $\delta \in (1,\infty)$*.*

For large n, the exponent of the degree distribution of $\mathrm{PAM}_n^{(\delta)}$ equals $\tau = 3 + \delta$.

THEOREM 5c

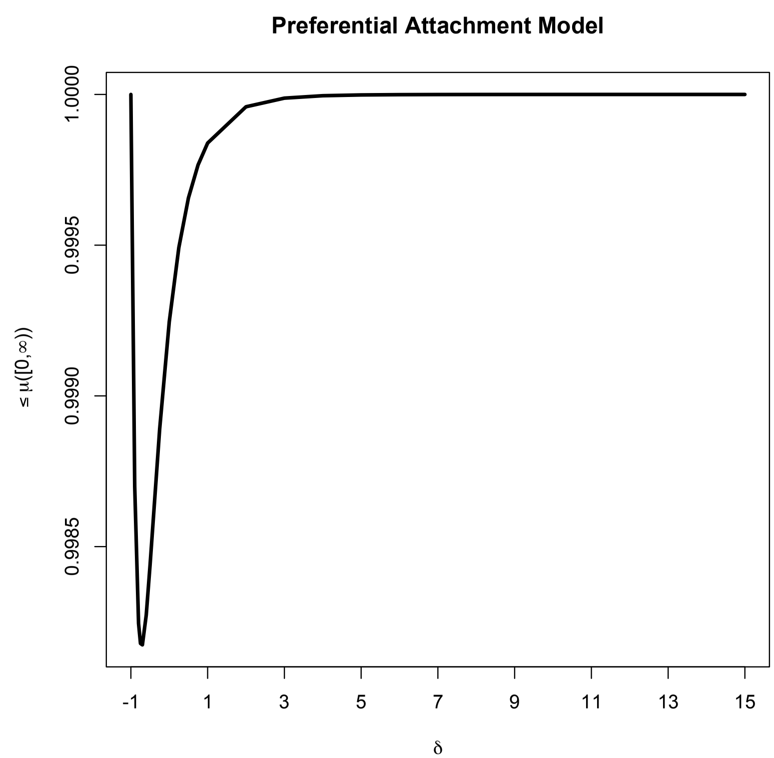
As $x \to \infty$,

$$\mu([x,\infty)) \gtrsim \begin{cases} x^{-(3\tau-4)}, & \text{if } \delta \in (-1,0), \\ x^{-(2\tau-1)}, & \text{if } \delta \in [0,\infty), \end{cases}$$

and

$$\mu([x,\infty)) \lesssim \begin{cases} x^{-(3-\tau)}, & \text{if } \delta \in [-\frac{1}{2},0), \\ x^{-1}, & \text{if } \delta = 0, \\ x^{-(\tau-3)}, & \text{if } \delta \in (0,\infty). \end{cases}$$

Numerical computation shows that, for large n, $\mathrm{PAM}_n^{(\delta)}$ is almost fully significant.



§ FUTURE CHALLENGES

▷ Derive necessary and sufficient conditions on random graph parameters guaranteeing significance.

▷ Provide full intuition behind the asymptotic results that were obtained.

▷ Study the multi-level friendship paradox.