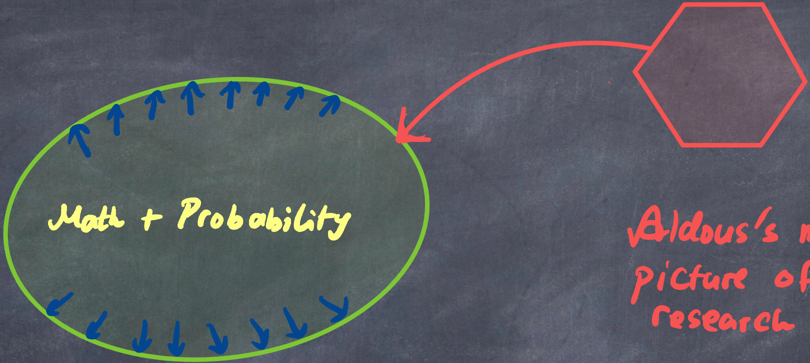
Living discreetly but thinking continuously

Dynamic network models and stochastic approximation

Oxford discrete math and probability seminar, June 2024.

Shankar Bhamidi
UNC Chapel Hill
Department of Statistics and OR



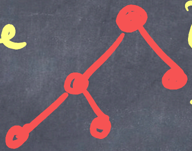
Aldous's mental picture of research

Bottom line: Area of dynamic networks needs mathematicians!

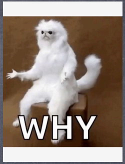
(One) Math Punchline

- Consider a sequence of growing network models $\{\mathcal{T}_n : n \geq 1\}$ in discrete time
- Fix your favorite empirical quantity of interest

  e.g. $\dfrac{\#\text{ of vertices of degree} = 10}{n}$

− {# of vertices whose distance 2 neighborhood looks like  }

n



Turns out: In many many network models Continuous time branching processes naturally describe the limits of such objects.

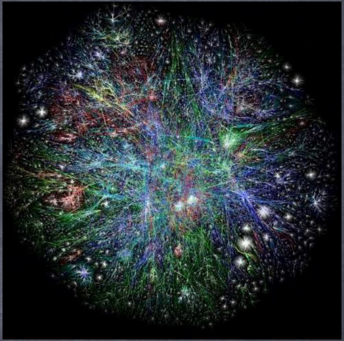
OUTLINE

1. Motivation from one area: Attributed network models
   - Fundamental questions and hypothesis
   - "News you can use"
   - Propogation of chaos $\rightarrowtail$ CTBP $\rightarrowtail$ Math understanding
2. Seed detection in dynamic networks
3. Change point detection

SUMMARY FINDINGS

1. Dynamic network models are truly complicated beasts. Simple rules give rise to complex phenomenon, quite often hard to predict even from simulation

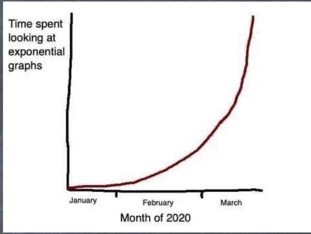
2. Owing either explicitly (construction of model) or implicitly (propogation of chaos) dynamics often driven towards evolution mechanisms in continuous time branching processes.



↳ "Internet" by Opte Project

SUMMARY FINDINGS CONTINUED

$\boxed{3}$ Continuous time branching processes grow exponentially (at some rate λ) while functionals of interest (e.g. degree distribution, Page rank scores) grow at a different rate ($\alpha_{\text{functional}}$). Asymptotics emerges from the interplay of these two rates.

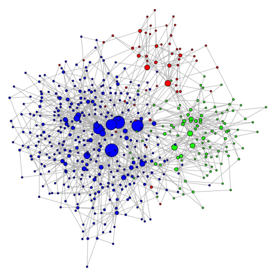


Found on Twitter
e.g. Michael Reuter

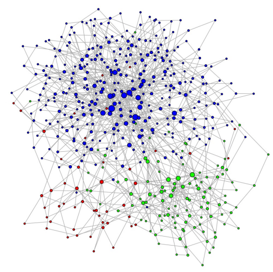
I. ATTRIBUTED NETWORK MODELS

Motivation

- Most social networks consist of vertices with attributes.
- $\mathcal{S}$ = attribute space. For talk $\mathcal{S} = \{1, 2, \ldots, k\}$
- Typically these networks are
  - Dynamic
  - Connections modulated by factors such as
    - heterogeneity of connection propensities across attributes
    - time and path dependent
    - Popularity bias



$\gamma = 1$



$\gamma = 0.2$

- Corresponding social networks play major role in diffusion of information
- Used by Companies via ranking/centrality algorithms to find influential nodes and pay such nodes to direct flow of information, effect perception of specific groups etc
- **Number of FOLK THEOREMS**

<u>Example</u>: Most centrality scores have similiar behavior for such networks

Related important question

- In many settings cannot directly observe network. Need to sample from network
- Perhaps interested in a "rare" minority
- e.g. Asian Immigrant populations in Research triangle and Impact of COVID etc in early 2020 [GIOVANNA MERLI } Duke + Carolina
  TED MOUW } Population Center]

WebRDS from University of Michigan Ann Arbor

Punchline - Has motivated a detailed development of network models that incorporate important functionals in their evolution

- Derive insight about various phenomenon from these models

"Mechanistic network models from domains of complexity science can enable researchers to consider various hypothetical scenarios ... This allows to evaluate robustness of algorithms with regards to different aspects concerning minorities, for example fairness or discrimination."

F. Karimi, M. Oliveira, and M. Strohmaier: arxiv 2206:07113

Main model in town

- Latent space $\mathcal{S} = [K] = \{1, 2, \ldots, K\}$.
- Fix a probability measure $\boldsymbol{\pi}$ on $\mathcal{S}$ (density of different types).
- Potentially asymmetric function $\kappa : \mathcal{S} \times \mathcal{S} \to \mathbb{R}_+$ (propensities of pairs of nodes to connect, based on their attributes).
- Preferential attachment parameter $\gamma \in [0, 1]$. [$\gamma \equiv$ for this talk]

Model class $\mathscr{P}(\gamma, \boldsymbol{\pi}, \kappa)$

- Vertices enter the system sequentially for $n \geq 1$ starting with a base connected graph $\tilde{\mathcal{G}}_0$. Write v_n for the vertex that enters at time n; every vertex v_n has attribute distribution $a(v_n) \sim \boldsymbol{\pi}$ independent of $\left\{\tilde{\mathcal{G}}_s : 0 \leq s \leq n-1\right\}$.
- For $v \in \tilde{\mathcal{G}}_n$, let $\deg(v, n) =$ degree of v at time n.
- Conditional on $\tilde{\mathcal{G}}_n$ the probability that v_{n+1} connects to $v \in \tilde{\mathcal{G}}_n$ is given by:

$$\mathbb{P}(v_{n+1} \rightsquigarrow v | \tilde{\mathcal{G}}_n, a(v_{n+1}) = a^\star) = \frac{\kappa(a(v), a^\star)[\deg(v, n)]^\gamma}{\sum_{v' \in \tilde{\mathcal{G}}_n} \kappa(a(v'), a^\star)[\deg(v', n)]^\gamma}$$

Will restrict to $\gamma = 1$ in this talk. Will view as directed graphs with edges pointing from children to parent.

Interpretation of Kernel

$K($  ,  $)$ = Propensity of new vertex (type = CORGI) to connect to existing vertex (type = Shibu)

Functionals of interest

- **Degree distribution of the graph:** Fix $k \geq 1$. $N_n(k) = \#$ of vertices of degree k in $\mathcal{G}_n$. $\mathbf{p}_n = \{N_n(k)/n : k \geq 0\}$ = empirical probability mass function.
- **Joint distribution of attributes and types:** $\varkappa_n(\cdot) = \frac{1}{n}\sum_{v \in \mathcal{V}_n} \delta_{(\deg(v), a(v))}$.
- **Page rank scores for directed graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ with damping factor $c \in (0,1)$=** stationary distribution $(\mathfrak{R}_{v,c} : v \in \mathcal{G})$ of following random walk: at each step, with probability c, follow an outgoing edge (uniform amongst available choices) from current location in the graph. With probability $1-c$, restart at uniformly selected vertex in entire graph. Given by linear system of equations:

$$\mathfrak{R}_{v,c} = \frac{1-c}{n} + c \sum_{u \in \mathcal{N}^-(v)} \frac{\mathfrak{R}_{u,c}}{d^+(u)} \tag{1}$$

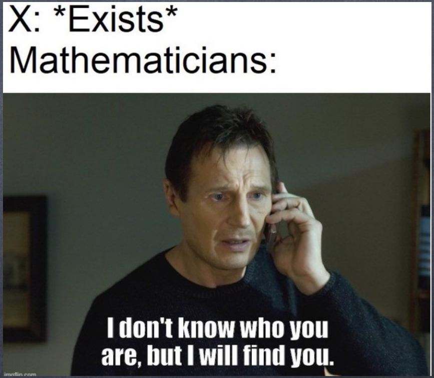
where $\mathcal{N}^-(v)$ is the set of vertices with edges pointed at v and $d^+(u)$ is the out-degree of vertex u. [Can similiarly look at joint distn between attribute + page rank]

Methodological questions: how do centrality measures (degree centrality; page rank scores) vary by attribute type?

Main issue: math tractability for functional of interest

$\mathcal{P}$ = "Please analyze this model"

$\mathcal{U}$ = "Useful (maybe)"



Basic assumption and setup for results

- Assume $\boldsymbol{\pi}(\{a\}) > 0 \ \forall a \in \mathcal{S}$ and $\kappa_{a,b} > 0 \ \forall a, b \in \mathcal{S}$.
- For talk assume $\gamma = 1$ (*Linear preferential attachment*).

Model inputs

Kernel κ and weight measure $\boldsymbol{\nu}$.

Attributed network model $\{\tilde{\mathcal{G}}_n : n \geq 0\}$

$$\mathbb{P}\left(a(v_{n+1}) = a^\star, v_{n+1} \rightsquigarrow v | \tilde{\mathcal{G}}_n\right) := \frac{\kappa(a(v), a^\star)\nu(a^\star)[\deg(v,n)]^\gamma}{\sum_{a\in[K]} \sum_{v'\in\tilde{\mathcal{G}}_n} \nu(a)\kappa(a(v'), a)[\deg(v',n)]^\gamma}.$$

YUCK!

Seems like a mess: types of new vertices tightly coupled with the evolution of the entire process.

Rationale and math curiosity question

- *$\mathscr{U}$ can be simulated via dynamics where every vertex essentially* **behaves independently**
- Suppose one wanted to simulate model class $\mathscr{U}$ starting from one vertex of type a, then:
- Every vertex v that enters the system (starting with the root of type a) gives birth in continuous time independently to child nodes with attributes, connected to the vertex.
- For a node of type a, conditional on its degree d, the rate of reproduction of a child node of type a' is $\nu(a)\kappa(a, a')d^{\alpha}$.

Write $\{\mathrm{BP}(t) : t \geq 0\}$ for the (continuous time) process. For $n \geq 1$, T_n be the (random) time such that the size $|\mathrm{BP}(T_n)| = n$. Then easy to check that $\{\mathrm{BP}(T_n) : 1 \leq n \leq N\}$ has the same distribution as $\left\{\tilde{\mathcal{G}}_n : 1 \leq n \leq N\right\} \sim \mathscr{U}(\gamma, \nu, \kappa)$.

Math curiosity question

Suppose we can choose $\boldsymbol{\nu}$ such that "asymptotically" composition of population is approximately $\boldsymbol{\pi}$. Are the two model classes $\mathscr{P}$ and $\mathscr{U}$ "similar"?

Basic punchline of the entire Attributed network models

Answer to math curiosity question = YES. Can carry out the entire program, so that asymptotics of all functionals of interest derivable from the "easier to simulate" model class $\mathscr{U}$.

$$\begin{tikzcd}
\mathscr{P}(\gamma, \boldsymbol{\pi}, \kappa) \arrow[r] \arrow[d, dashed, "??"'] & \boldsymbol{\nu}(\gamma, \boldsymbol{\pi}, \kappa) \arrow[d] \\
\text{Asymptotics} & \mathscr{U}(\gamma, \boldsymbol{\nu}, \kappa) \arrow[l]
\end{tikzcd}$$

 Goal model class $\mathcal{P}$

Inputs: π and κ

Let $\mathcal{P}(\mathcal{S})$ denote the space of all probability measures on $\mathcal{S}$. Define (in the interior of $\mathcal{P}(\mathcal{S})$) the function:

$$V_{\pi}(\mathbf{y}) := 1 - \frac{1}{2}\sum_{j\in\mathcal{S}} \pi_j \left(\log(y_j) + \log(\sum_{k\in\mathcal{P}} y_k \kappa_{k,j}) \right)$$

Fundamental Lemma (Jordan (2013), EJP)

Under above Assumptions, $V_{\pi}(\cdot)$ has a *unique* minimizer $\boldsymbol{\eta} := \boldsymbol{\eta}(\boldsymbol{\pi}) = (\eta_1(\boldsymbol{\pi}), \ldots, \eta_K(\boldsymbol{\pi}))$ in the interior of $\mathcal{P}(\mathcal{S})$.

$$\nu_b := \frac{\pi_b}{\sum_{l=1}^{K} \kappa_{l,b}\eta_l}, \qquad \phi_{a,b} := \kappa_{a,b}\nu_b, \qquad \phi_a := \sum_{b=1}^{K} \phi_{a,b} = 2 - \frac{\pi_a}{\eta_a},$$

Algorithm

- Consider Model class $\mathcal{U}$ with parameters ν and K
- Easy to simulate as a branching process (in continuous time). Individuals behave independently
- Anything else ??

Main result 1 (all joint work with Nelson Antunes, Sayan Banerjee, Vladas Pipiras)



Theorem (2023) for $\gamma = 1$

Asymptotics for all "local" functionals of model class $\mathscr{P}$ can be obtained from model class $\mathscr{U}$ with above choice of $\boldsymbol{\nu}$. For example, pick a vertex at random in $\mathcal{G}_n \sim \mathscr{P}$ and consider the descendant subtree of that vertex. Then the distribution of this descendant subtree converges to the following:

- Pick $A \sim \boldsymbol{\pi}$.
- Start a branching process simulating model class $\mathscr{U}(1, \boldsymbol{\nu}, \kappa)$ starting from a single vertex of type A.
- Run this simulation for $\tau =$ Exponential random variable with rate $= 2$.

Under the hood: associated branching process $\mathscr{U}$ grows at rate $\lambda = 2$: Simulation takes $\approx \frac{1}{2}\log n$ in the computer to generate network of size n.

Branching process grows like e^{2t}. For a vetex of type a, Number of children = degree+1 grows like $e^{\phi_a t}$. Interplay gives the following:

Degree distribution

For each $a \in [K]$, $\mathbf{p}_n^a \to \mathbf{p}_\infty^a$ where the tail pmf is given by

$$\bar{\mathbf{p}}_\infty^a(k) = \frac{\Gamma\left(1 + \frac{2}{\phi_a}\right)\Gamma(k+1)}{\Gamma\left(k+1+\frac{2}{\phi_a}\right)}, \qquad k \geq 0.$$

In particular $\mathbf{p}_\infty^a(k) \sim k^{1+2/\phi_a}$ as $k \to \infty$.

Previous derived in 2013 by Jordan using stochastic approximation techniques. Part of the methodological contribution of our work is to show, stochastic approximation techniques can be used to track evolution of motif counts. → e.g. trees of any shape and attribute structure

Degree distribution tails does depend on the attribute type. Thus potentially, degree centrality scores depend in a non-trivial manner on the type of a vertex.

Implications 2: Page rank scores for model class $\mathscr{P}$

- Recall $\mathcal{G}_n$ is directed with edges from child to parent. For $v \in \mathcal{G}_n$, let $P_l(v,n)$ denote the number of *directed* paths of length l that end at v in $\mathcal{G}_n$. Since $\mathcal{G}_n$ is a directed tree, easy to check PageRank scores have the explicit formulae:

$$\mathfrak{R}_{v,c}(n) = \frac{(1-c)}{n}\left(1 + \sum_{l=1}^{\infty} c^l P_l(v,n)\right).$$

- *Stare at this formula: suggests connection to percolation, where each edge retained with probability c, deleted with probability* $1-c$*.*
- Easier to formulate results in terms of the *graph normalized* PageRank scores $\{R_{v,c}(n) : v \in \mathcal{G}_n\} = \{n\mathfrak{R}_{v,c}(n) : v \in \mathcal{G}_n\}$.
- Empirical distribution of normalized PageRank scores,

$$\hat{\mu}_{n,\mathrm{PR}} := n^{-1} \sum_{v \in \mathcal{G}_n} \delta\{R_{v,c}(n)\}.$$

Algorithm

→ Go back to model class $\mathcal{U}$

→ Consider percolation on $\mathcal{U}$

→ Turns out: This can again be viewed as a different Branching process. "Easy" to analyze.

→ Punchline: Asymptotics about $\wp$ follow from $\mathcal{U}$.

- Consider $\mathrm{BP}_a(\cdot)$, branching process started with one vertex of type a.
- $\mathcal{R}_{\emptyset,c}(t) = (1-c)\left(1 + \sum_{l=1}^{\infty} c^l P_{l,\emptyset}(t)\right)$.
- Define "limit" $\mathcal{R}_{\emptyset,c} = \mathcal{R}_{\emptyset,c}(\tau) = (1-c)\left(1 + \sum_{l=1}^{\infty} c^l P_{l,\emptyset}(\tau)\right)$.
- As before τ is an exponential rate two random variable.

Weird matrix associated with $\mathscr{U}$

$$\mathbf{M}^{(c)} = \left(\mathbf{M}^{(c)}_{(a,b)} := c\phi_{a,b} + \phi_a \mathbb{1}\{a=b\}\right)_{a,b\in[K]} .$$

λ_c= Perron-Frobenius eigen-value of $\mathbf{M}^{(c)}$.

 [again Comes from model class $\mathcal{U}$]

Fix $a \in [K]$ and damping factor $c \in (0,1)$. For any $t \geq 0$, write $\mathrm{BP}_a^c(t)$ for the connected cluster of the root (which is also a tree) when we retain each edge $e \in \mathrm{BP}_a(t)$ with probability c and delete with probability $(1-c)$, independently across edges. Write $\{\mathrm{BP}_a^c(t) : t \geq 0\}$ for the corresponding non-decreasing rooted tree value process. Let $\mathfrak{Z}_a^c(t) = |\mathrm{BP}_a^c(t)|$ for the size of the cluster at time t.

Turns out: $\mathrm{BP}_a^c(\cdot)$ *is also a branching process.* λ_c *is the rate of growth of* $\mathrm{BP}_a^c(\cdot)$ *i.e.* $|\mathrm{BP}_a^c(\cdot)| \approx e^{\lambda_c t}$

Page rank asymptotics

For every continuity point r of the distribution of $\mathcal{R}_{\emptyset,c}$ under $\mathbb{P}_a$

$$n^{-1} \sum_{v \in \mathcal{G}_n} 1\{a(v) = a, R_{v,c}(n) > r\} \xrightarrow{\mathrm{P}} \pi_a \mathbb{P}_a(\mathcal{R}_{\emptyset,c} > r).$$

Further there exists constants $B_1 < B_2 < \infty$ such that for any attribute:

$$B_1 r^{-2/\lambda_c} \leq \mathbb{P}_a(\mathcal{R}_{\emptyset,c} > r) \leq B_2 r^{-2/\lambda_c}$$

News you can use: Page rank score distributions do **not** *depend on the attribute type. Negates some of the standard assumptions in social networks.*

1. **Uniform node sampling** ($\mathfrak{U}$)**:** Here one picks a vertex uniformly at random from $\mathcal{G}_n$.
2. **Sampling proportional to degree** ($\mathfrak{D}$)**:** Pick a vertex uniformly at random and then pick a neighbor of this vertex uniformly at random.
3. **Sampling proportional to in-degree** ($\mathfrak{ID}$)**:** Pick a vertex at random and then select the parent; by convention, if the root is picked (which happens with probability $o_{\mathbb{P}}(1)$ as $n \to \infty$) then select the root.

4. **Sampling proportional to Page rank ($\mathfrak{PR}_c$):** Fix a damping factor c and sample a vertex with probability proportional to the page rank scores $\{\mathfrak{R}_{v,c} : v \in \mathcal{G}_n\}$. In the context of the (tree) network model $\{\mathcal{G}_n : n \geq 1\}$ starting with a single root at time zero, by work of Chebolu+Melsted: this can be accomplished by the following "local" algorithm:
   1. Pick a vertex uniformly V at random from $\mathcal{G}_n$.
   2. Independently let $G \sim \text{Geom}(1-c) - 1$ (here $\text{Geom}(\cdot)$ is a Geometric random variable with prescribed parameter with support starting at one).
   3. Starting from V Traverse G steps towards the root (i.e. using the directions of edges in $\mathcal{G}_n$ from child to parent), stopping at the root, if the root is reached before G steps.
4. **Fixed length sampling ($\mathfrak{PR}_M$):** Fix $M \geq 0$. Consider the same implementation of the page rank scheme but here the halting distribution is taken to be $G \equiv M$. Abusing notation, we use $\mathfrak{PR}_M$ to denote this sampling scheme.

Will skim next two slides

Bottom line: Get explicit formulae for bias of various network sampling schemes.

→ All using $\mathcal{U}$

Define matrix

$$\mathbf{M} = \left(\mathbf{M}_{(a,b)} := \frac{\phi_{a,b}}{2 - \phi_a} \right)_{a,b \in [K]}.$$

Turns out this has Perron-Frobenius eigen-value =1. Let $\Psi = (\Psi_1, \Psi_2, \ldots, \Psi_K)$ denote the corresponding right eigen-vector, normalized so that $\sum_{a \in [K]} \pi_a \Psi_a = 1$. Consider the Markov chain $\mathbf{S} := \{S_n : n \geq 0\}$ on $[K]$ with transition probability matrix

$$\mathbb{P}_i^{\mathbf{S}}(S_1 = j) := \mathbb{P}^{\mathbf{S}}(S_1 = j | S_0 = i) = \frac{\mathbf{M}_{i,j} \Psi_j}{\Psi_i}, \qquad j \in [K].$$

Write $\mathbb{E}_i^{\mathbf{S}}$ for the expectation operator under $\mathbb{P}_i^{\mathbf{S}}$.

Implications for network sampling from $\mathscr{P}$

1. Under uniform sampling $\mathbb{P}_{\mathfrak{U}}(a(V_n) = b|\mathcal{G}_n) \xrightarrow{a.s.} \pi_b$.
2. Under sampling proportional to degree $\mathbb{P}_{\mathfrak{D}}(a(V_n) = b|\mathcal{G}_n) \xrightarrow{a.s.} \eta_b$.
3. Under sampling proportional to in-degree,
$$\mathbb{P}_{\mathfrak{I}\mathfrak{D}}(a(V_n) = b|\mathcal{G}_n) \xrightarrow{a.s.} \eta_b\phi_b = \pi_b \frac{\phi_b}{2-\phi_b} = \pi_b \Psi_b \mathbb{E}_b^{\mathbf{S}}\left[\frac{1}{\Psi_{S_1}}\right].$$
4. Under sampling proportional to Page-Rank, letting $G \sim \text{Geom}(1-c) - 1$ independent of $\mathbf{S}$,
$$\mathbb{P}_{\mathfrak{P}\mathfrak{R}_c}(a(V_n) = b|\mathcal{G}_n) \xrightarrow{a.s.} \pi_b \Psi_b \mathbb{E}_b^{\mathbf{S}}\left[\frac{1}{\Psi_{S_G}}\right].$$
Since $\mathbf{S}$ has stationary distribution $\{\pi_a \Psi_a : a \in [K]\}$,
$$\lim_{c\uparrow 1} \lim_{n\to\infty} \mathbb{P}_{\mathfrak{P}\mathfrak{R}_c}(a(V_n) = b|\mathcal{G}_n) \xrightarrow{a.s.} \pi_b \Psi_b.$$
5. Under fixed length walk sampling,
$$\mathbb{P}_{\mathfrak{P}\mathfrak{R}_M}(a(V_n) = b|\mathcal{G}_n) \xrightarrow{a.s.} \pi_b \Psi_b \mathbb{E}_b^{\mathbf{S}}\left[\frac{1}{\Psi_{S_M}}\right].$$

Consider the specific case of model class $\mathscr{P}$ with two classes $1, 2$ with,

$$\boldsymbol{\kappa} = (\kappa(i,j))_{1\leq i,j\leq 2} = \begin{pmatrix} 1 & 1 \\ a & 1 \end{pmatrix}, \qquad \boldsymbol{\pi} = \frac{1}{1+\theta}(\theta, 1). \tag{2}$$

We will be interested in the specific case where $\theta \to 0$, more specifically in the setting

$$\theta := \theta(a) = D\sqrt{a},$$

where $D > 0$ is a fixed constant and where $a \downarrow 0$. Thus,

1. Type 1 vertices are relatively *rare* compared to type 2 vertices; we will often refer to type 1 vertices as minorities and type 2 as majorities.
2. Newly entering majority vertices into the population have equal propensity to connect to minority or majority vertices. Minorities have (relatively) **much higher** propensity to connect to other minority vertices, as compared to majority vertices.

Implications

As $a \downarrow 0$:

1. Under uniform node sampling,

$$\mathbb{P}_{\mathfrak{U}}(a(V_n) = 1|\mathcal{G}_n) \xrightarrow{a.s.} D\sqrt{a} + O(a).$$

2. For sampling proportional to degree,

$$\mathbb{P}_{\mathfrak{D}}(a(V_n) = 1|\mathcal{G}_n) \xrightarrow{a.s.} 2D\sqrt{a} - (4D^2 + \frac{1}{2})a + O(a^{3/2}).$$

3. For random in-degree based sampling,

$$\mathbb{P}_{\mathfrak{I}\mathfrak{D}}(a(V_n) = 1|\mathcal{G}_n) \xrightarrow{a.s.} 3D\sqrt{a} + O(a).$$

4. For Page-rank based sampling (both Geometric and fixed node implementations):

$$\lim_{c\uparrow 1} \lim_{n\to\infty} \mathbb{P}_{\mathfrak{P}\mathfrak{R}_c}(a(V_n) = 1|\mathcal{G}_n) = \frac{2D^2 - \frac{1}{2} + \sqrt{\left((2D^2 - \frac{1}{2})^2 + 4D^2\right)}}{2D^2 + \frac{1}{2} + \sqrt{\left((2D^2 - \frac{1}{2})^2 + 4D^2\right)}} + O(\sqrt{a})$$

$$= \lim_{M\uparrow\infty} \lim_{n\to\infty} \mathbb{P}_{\mathfrak{P}\mathfrak{R}_M}(a(V_n) = 1|\mathcal{G}_n)$$

Insight gleaned from above analysis

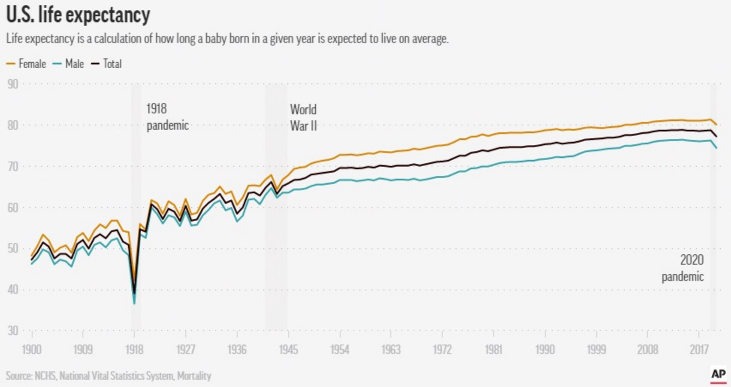
- In the "natural" time-scale of the above models, processes grow exponentially
- "Should imply": Signature of the seed of the network should "persist" for a long time.

→ Should make "estimating" initial seed when one has no temporal information "doable"

→ Should make change point detection "harder."

Long range dependence!

Change Point Detection



Source: Associated Press

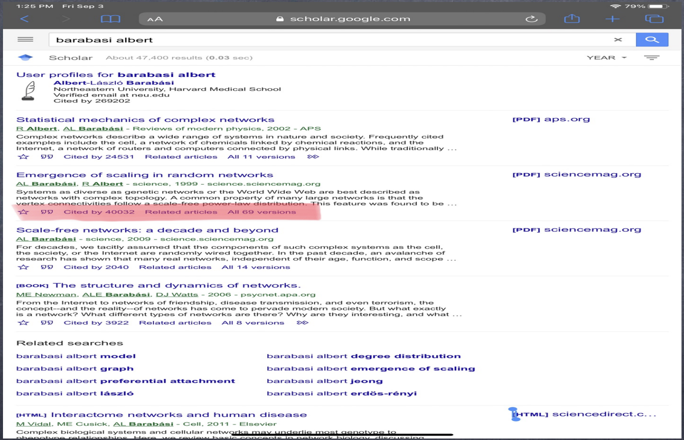
Our motivation in words

- Suppose you have temporal network data.
  - Ex: Adjacency matrix at all or sub-sample of time points
  - Ex: Time series observations at each node etc
- Suppose network experiences a shock at some point.
- Can we detect this change point from observations?
- Changes in structural properties of the system?

Recall : Probabilistic foundations

- Network model : Fix attachment function f. Start with single seed.
- At each stage new vertex v enters system. Connects to one pre-existing vertex
- Probability connecting to a vertex u in the system proportional to $f(\text{degree}(u))$.
- $\mathcal{T}_n$ = network of size n

Example $f(k) = k + \alpha$ Preferential attachment



Known results for $f(k) = k + \alpha$

- $N_k(n)$ = # of vertices of degree k in $\mathcal{T}_n$

$$\frac{N_k(n)}{n} \longmapsto p_k$$

- $$p_k \sim \frac{C}{k^{\alpha+3}}$$ Degree exponent $= \alpha + 3$

- max-degree $= M_n \sim n^{\frac{1}{\alpha+2}}$

Example of Standard change point model

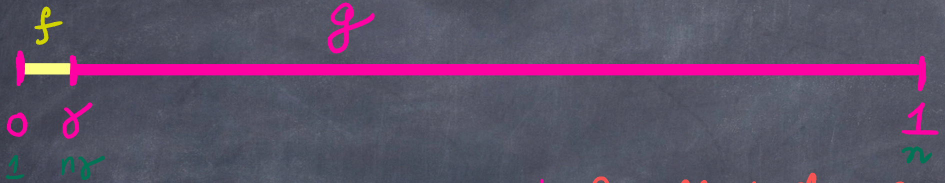
- Fix $\gamma \in (0,1)$.
- For $t \in [1, n\gamma]$, network uses attachment function

$$f(k) = k + \alpha$$

- For $t \in [n\gamma+1, n]$, network uses

$$g(k) = k + \beta$$

Any guesses on the degree exponent?



Guesses?

Recall under no change

$f(k) = k + \alpha$

degree exponent $= \alpha + 3$

$g(k) = k + \beta$

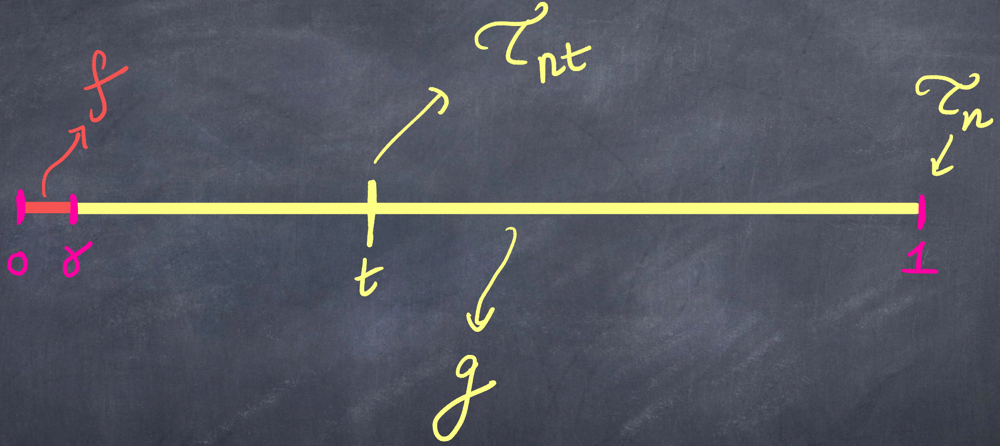
degree exponent $= \beta + 3$

Punchline of the Theorems

Irrespective of how small γ is (e.g $\gamma = .01$ or $\gamma = .00000001$), the initializer function Always wins !

Standard change point model

- Fix $\gamma \in (0,1)$.
- For $t \in [1, n\gamma]$, network uses attachment function
  $$f(k) = \text{general function}$$
- For $t \in [n\gamma+1, n]$, network uses
  $$g(k) = \text{general function}$$



f
τ_{nt}
τ_n
0
γ
t
1
g

Fix $t \in [0,1]$. Let $N_k(nt)$ = # of vertices of degree k in $\mathcal{T}_{nt}$

Theorem [Banerjee, B, Carmichael]
Under conditions on f and g $\exists$ explicit probability mass functions $\{(p_k(t))_{k\geq 1} : t \in [0,1]\}$ such that

$$\sup_{t\in[0,1]} \left| \frac{N_k(nt)}{nt} - p_k(t) \right| \longmapsto 0$$

Theorem [Banerjee, B, Carmichael]

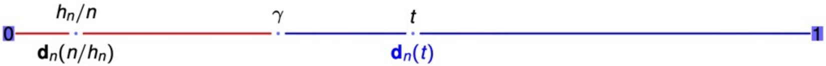
Under above technical conditions on f & g, irrespective of how small γ is f always wins!

- So if degree exponent with f and no change point is δ so is the model with change point.

Change point estimator : For each $t \in (0,1)$ Compare degree distrn $\left(\frac{N_k(nt)}{nt}\right)_{k \geq 1}$ with the degree distribution when network is of size $\frac{n}{\ln n}$ (recall change point at $n\gamma$)

and become alarmed the first time there seems to be a big change in degree distrn.

Change point estimator



Nonparametric change point estimator

Fix any two sequences $h_n \to \infty, b_n \to \infty$: $\frac{\log h_n}{\log n} \to 0$, $\frac{\log b_n}{\log n} \to 0$. Define

$$\hat{T}_n = \inf \left\{ t \geq \frac{1}{h_n} : \sum_{k=0}^{\infty} 2^{-k} \left| \frac{D_n(k, \mathcal{T}^{\boldsymbol{\theta}}_{\lfloor nt \rfloor})}{nt} - \frac{D_n(k, \mathcal{T}^{\boldsymbol{\theta}}_{\lfloor n/h_n \rfloor})}{n/h_n} \right| > \frac{1}{b_n} \right\}.$$

Then $\hat{T}_n \xrightarrow{\mathrm{P}} \gamma$.

Lots of open problems

Simulations

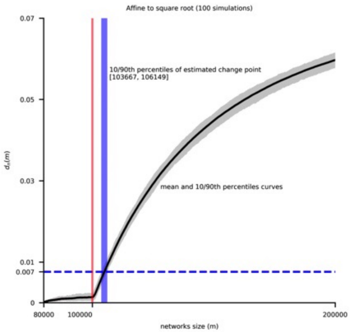


Figure: $n = 2 * 10^5$, $\gamma = 0.5$, $f_0(i) = i + 2$, $f_1(i) = \sqrt{i+2}$, $h_n = \log\log n$, $b_n = n^{1/\log\log n}$

$$d_n(m) := \sum_{k=0}^{\infty} 2^{-k} \left| \frac{D_n(k, \mathcal{T}_m^{\theta})}{m} - \frac{D_n(k, \mathcal{T}_{\lfloor n/h_n \rfloor}^{\theta})}{n/h_n} \right|, \qquad \frac{n}{\log\log n} < m \le n.$$

The big bang model: What if the change happened very early in the system?

Figure: Big Bang: Getty images

Fix functions $f_0, f_1 : \{0, 1, 2, \ldots\} \to \mathbb{R}_+$ and $\gamma \in (0, 1)$. Let $\boldsymbol{\theta} = (f_0, f_1, \gamma)$.

Model

- **Time** $1 \leq m \leq n^\gamma$ Vertices perform attachment with probability proportional to $f_0(out - deg)$.
- **Time** $n^\gamma < m \leq n$ Vertices perform attachment with probability probability proportional to $f_1(out - deg)$.

Change point detection: Quick big bang

Result 1

- Here change point at n^γ (e.g. $\sqrt{n}$).
- Here

$$\frac{N_n(k)}{n} \xrightarrow{\mathbb{P}} p_k^1$$

namely the degree distribution of the model run purely with attachment function f_1

So what changes?

1. **Uniform $\rightsquigarrow$ Linear:** $f_0 \equiv 1$ whilst $f_1(k) = k + 1 + \alpha$ for fixed $\alpha > 0$. Then for $\omega_n \uparrow \infty$,

$$\frac{n^{\frac{1-\gamma}{2+\alpha}} \log n}{\omega_n} \ll M_n(1) \ll n^{\frac{1-\gamma}{2+\alpha}} (\log n)^2.$$

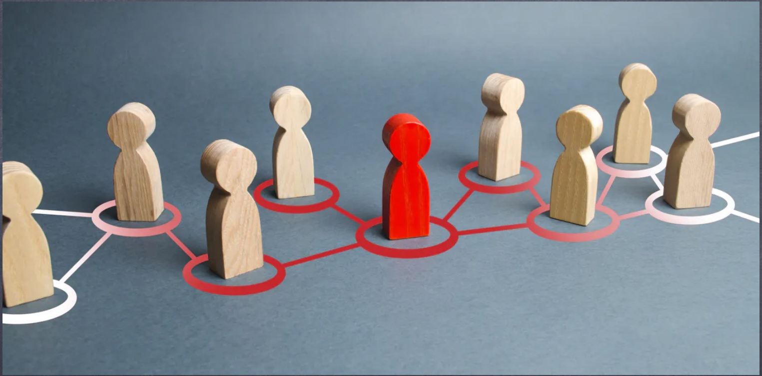
2. **Linear $\rightsquigarrow$ Uniform:** $f_0(k) = k + 1 + \alpha$ whilst $f_1(\cdot) \equiv 1$.

$$\frac{n^{\frac{\gamma}{2+\alpha}} \log n}{\omega_n} \ll M_n(1) \ll n^{\frac{\gamma}{2+\alpha}} (\log n)^2.$$

3. **Linear $\rightsquigarrow$ Linear:** $f_0(k) = k + 1 + \alpha$ whilst $f_1(k) = k + 1 + \beta$ where $\alpha \neq \beta$. Then $M_n(1)/n^{\eta(\alpha,\beta)}$ is tight where

$$\eta(\alpha, \beta) := \frac{\gamma(2+\beta) + (1-\gamma)(2+\alpha)}{(2+\alpha)(2+\beta)}. \quad (5)$$

Seed detection in evolving networks



https://theconversation.com/patient-zero-why-its-such-a-toxic-term-134721

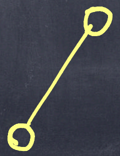
Our motivation in words

- Dynamic network started with a single node ("patient zero") or seed graph at time zero.
- observe network when it is of large size e.g. $n = 10^6$. with no temporal information only network topology (adjacency matrix)
- Have a fixed budget say $K = 30$.
- GOAL: Output 30 vertices such that with high prob. seed is in the output.

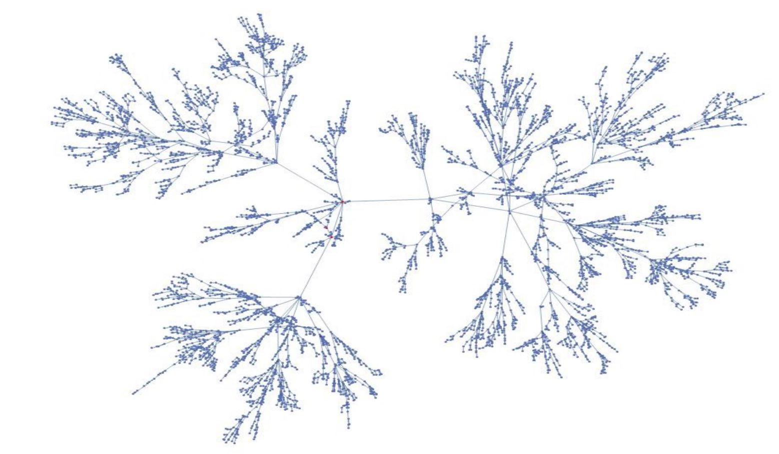
Probabilistic foundations

- Network model: Fix attachment function f. Start with single seed.
- At each stage new vertex v enters system. Connects to one pre-existing vertex
- Probability connecting to a vertex u in the system proportional to $f(\text{degree}(u))$.
- $\mathcal{T}_n$ = network of size n

Example: $f = 1$ (Random recursive tree)

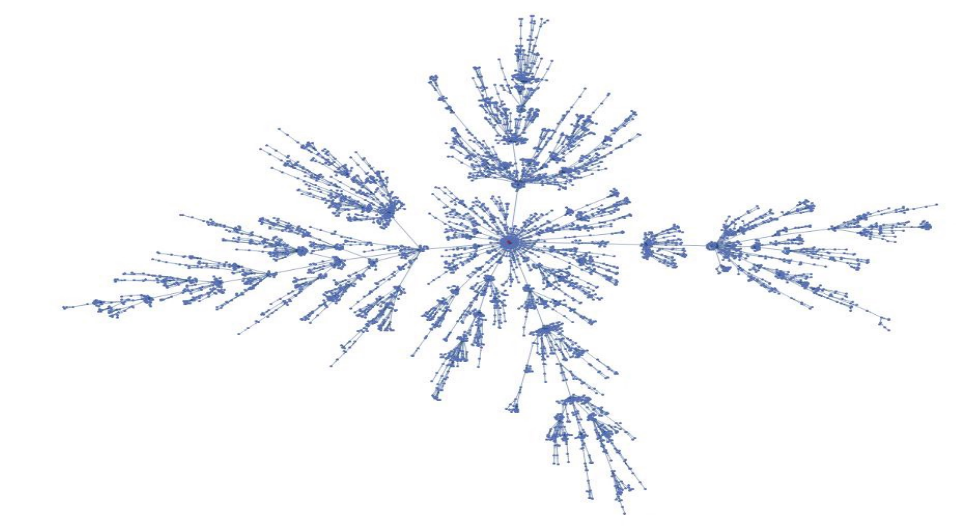


SIMULATION (n = 3000 ?)



Example $f(k) = k$ Preferential attachment

Simulation (n = 5000)



Setup:

- $\mathbf{G}$: space of equivalence classes (upto isomorphisms) of finite unlabelled graphs.
- For finite labelled graph $\mathcal{G}$: $\mathcal{G}^\circ$ for the isomorphism class of $\mathcal{G}$ in $\mathbf{G}$.
- Root finding algorithm: Fix $K \geq 1$ and a mapping H_K on $\mathbf{G}$ that takes an input finite unlabelled graph $\mathbf{g} \in \mathbf{G}$ and outputs a subset of K vertices from $\mathbf{g}$.

Root finding algorithms

Let $\{\mathcal{G}_n : n \geq 0\}$ be a sequence of growing random networks. Fix $0 < \varepsilon < 1$ and $K \geq 1$. A mapping H_K is called a budget K root finding algorithm with error tolerance ε for the sequence of networks if,

$$\liminf_{n\to\infty} \mathbb{P}(1 \in H_K(\mathcal{G}_n^\circ)) \geq 1 - \varepsilon.$$

Question: can we choose K independent of n? Dependence on ε?

Class of seed detection algorithms

- Centrality based measures
- For each vertex obtain some measure of centrality so collection of numbers $\{\phi(u) : u = \text{vertex in } \mathcal{T}_n\}$
- Example :
  - Degree centrality : $\phi(u) =$ degree of u
  - Eigen-vector centrality
  - Centroid or Jordan centrality

ALGORITHM

- Suppose budget $= K$
- Output the "top" K vertices(Could be smallest or largest depending on the measure)
- Say that above has error tolerance ε if

$$\lim_{n \to \infty} \mathbb{P}(\text{seed} \in \text{outputed set of } T_n) \geq 1 - \varepsilon$$

Fundamental questions

- For given error tolerance ε (e.g. $\varepsilon = .01$)
  Can we select K independent of n = size of network?

- How does $K = K(\varepsilon)$ depend on ε?

  $\frac{1}{\varepsilon}$? $\frac{1}{\varepsilon^{100}}$? $\frac{1}{\varepsilon^{10000}}$?

Related notion: Robustness (Morters-Dietrich; Jog-Loh)

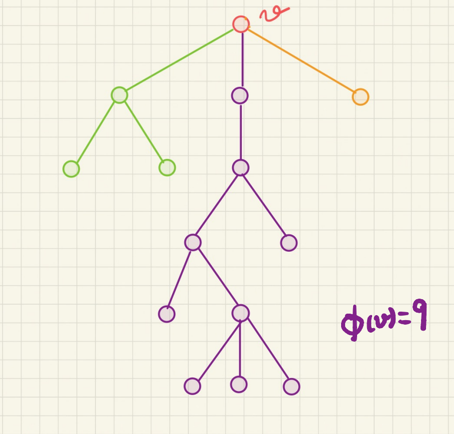
Persistence

Fix $K \geq 1$ and a network centrality measure Ψ. For a family of network models $\{\mathcal{G}_n : n \geq 1\}$ say that this sequence is (Ψ, K) **persistent** if $\exists\, n^* < \infty$ a.s. such that for all $n \geq n^*$ the optimal K vertices $(v_{1,\Psi}(\mathcal{G}_n^\circ), v_{2,\Psi}(\mathcal{G}_2^\circ), \ldots, v_{K,\Psi}(\mathcal{G}_n^\circ))$ remain the same and further the relative ordering amongst these K optimal vertices remains the same.

Example: If degree centrality was persistent this implies, the *identity* of the maximal degree vertex becomes fixed within finite time and no other vertex can overtake the degree of this vertex after this time.

Such phenomenon once again a hallmark of long range dependence.

Jordan or centroid Centrality*



$\phi(v)$ = size of the largest subtree of a child of v

* Only works for trees. First analyzed by Bubeck - Devroye - Lugosi.

Centroid centrality sufficiency bounds

technical

Banerjee and B(2020)

Under ~~above~~ assumptions:

1. Suppose for some $\overline{C}_f > 0$, $\beta \geq 0$, f satisfies $f_* \leq f(i) \leq \overline{C}_f \cdot i + \beta$ for all $i \geq 1$. Then $\exists$ positive constants C_1, C_2 such that for any error tolerance $0 < \varepsilon < 1$, the budget requirement satisfies,
$$K_\Psi(\varepsilon) \leq \frac{C_1}{\varepsilon^{(2\overline{C}_f+\beta)/f_*}} \exp(\sqrt{C_2 \log 1/\varepsilon}).$$
2. If further the attachment function f is in fact bounded with $f(i) \leq f^*$ for all $i \geq 1$ then one has for any error tolerance $0 < \varepsilon < 1$,
$$K_\Psi(\varepsilon) \leq \frac{C_1}{\varepsilon^{f^*/f_*}} \exp(\sqrt{C_2 \log 1/\varepsilon}).$$

Centroid centrality necessary bounds

- If $\exists\ \underline{C}_f > 0$ and $\beta \geq 0$ such that $f(i) \geq \underline{C}_f \cdot i + \beta$ for all $i \geq 1$ then $\exists$ a positive constant C_1' such that for any error tolerance $0 < \varepsilon < 1$,

$$K_\Psi(\varepsilon) \geq \frac{C_1'}{\varepsilon^{(2\underline{C}_f + \beta)/f(1)}}.$$

- For general f one has for any error tolerance $0 < \varepsilon < 1$,

$$K_\Psi(\varepsilon) \geq \frac{C_1'}{\varepsilon^{f_*/f(1)}}.$$

Special cases

- **Uniform attachment:** $f(k) = 1$

$$\frac{C_1'}{\varepsilon} \le K_\Psi(\varepsilon) \le \frac{C_1}{\varepsilon} \exp(\sqrt{C_2 \log \frac{1}{\varepsilon}})$$

- **Pure Preferential attachment:** $f(k) = k$

$$\frac{C_1'}{\varepsilon^2} \le K_\Psi(\varepsilon) \le \frac{C_1}{\varepsilon^2} \exp(\sqrt{C_2 \log \frac{1}{\varepsilon}}).$$

- **Affine preferential attachment:** $f(k) = k + \beta$

$$\frac{C_1'}{\varepsilon^{\frac{2+\beta}{1+\beta}}} \le K_\Psi(\varepsilon) \le \frac{C_1}{\varepsilon^{\frac{2+\beta}{1+\beta}}} \exp(\sqrt{C_2 \log \frac{1}{\varepsilon}}).$$

- **Sublinear preferential attachment:**

$$\frac{C_1'}{\varepsilon} \le K_\Psi(\varepsilon) \le \frac{C_1}{\varepsilon^2} \exp(\sqrt{C_2 \log \frac{1}{\varepsilon}}).$$

Disadvantages of Centroid centrality

- Essentially need quite precise information of entire network
- *Natural question:* How do more local measures like degree centrality perform? Does there exist a *persistent hub* (i.e. maximal degree vertex fixates within finite time)?
- *Fake popularity:* Suppose i-th vertex enters the system with m_i edges that it attaches to the current existing system (again with popularity of vertices measured via some function f). How quickly does $m_i \uparrow \infty$ to break persistence phenomenon?

Assumptions and notation

- $f_* := \inf_{i \geq 0} f(i) > 0$; further at most linear growth $f(i) \leq C_f(i)$.
- $\sum_{i=0}^{\infty} \frac{1}{f(i)} = \infty$.
- $\Phi_k(x) = \int_0^x \frac{1}{f^k(z)} dz$.
- $\mathcal{K}(t) = \Phi_2 \circ \Phi_1^{-1}(t), t \geq 0$.
- $d_{max}(n) := \max_{0 \leq k \leq n} d_k(n)$.
- *Index of the maximal degree:*

$$\mathcal{I}_n^* := \inf\{0 \leq i \leq n : d_i(n) \geq d_j(n) \text{ for all } j \leq n\}.$$

Banerjee + B(2020)

Under a few technical assumptions on f and *f is increasing*:

- Suppose $\Phi_2(\infty) < \infty$ (e.g. $f(k) = k^\alpha$ for $\alpha \in (1/2, 1]$) and that $\limsup_{n\to\infty} \frac{\Phi_1(m_n)}{\log s_n} \leq \frac{1}{8C_f}$. Then a persistent hub emerges almost surely in the random graph sequence

Do not need increasing assumption for trees.

$s_n = \sum_{i=1}^{n} m_i$

Banerjee + B(2020)

- Assume $\Phi_2(\infty) = \infty$ (e.g. $f(k) = k^\alpha$ for $\alpha \in (0, 1/2)$) and (we are working in the tree case) and $f(k) \to \infty$ as $k \to \infty$. Then index of maximal degree satisfies:

$$\frac{\log \mathcal{I}_n^*}{\mathcal{K}\left(\frac{1}{\lambda^*} \log n\right)} \xrightarrow{P} \frac{\lambda^{*2}}{2}, \quad \text{as } n \to \infty.$$

  where λ^* is the Malthusian rate of growth of the continuous time embedding.

- For $f(k) = k^\alpha$ for $\alpha \in (0, 1/2)$,

$$\frac{\log \mathcal{I}_n^*}{(\log n)^{\frac{1-2\alpha}{1-\alpha}}} \xrightarrow{P} \frac{(\lambda^*)^{\frac{1}{1-\alpha}}}{2}, \quad \text{as } n \to \infty.$$

Inspired by Morters and Dietrich who proved similar results for a different evolving network model.

Thank you for your attention!

ANY QUESTIONS ?