Long range dependence in evolving networks

Shankar Bhamidi
UNC Chapel Hill
Department of Statistics and OR

Outline

- Brief motivation and structure
- Seed detection in dynamic networks
- Change point detection
- Co-evolving networks

In collaboration with Sayan Banerjee, Iain Carmichael and Zoe Huang

Structure of talk

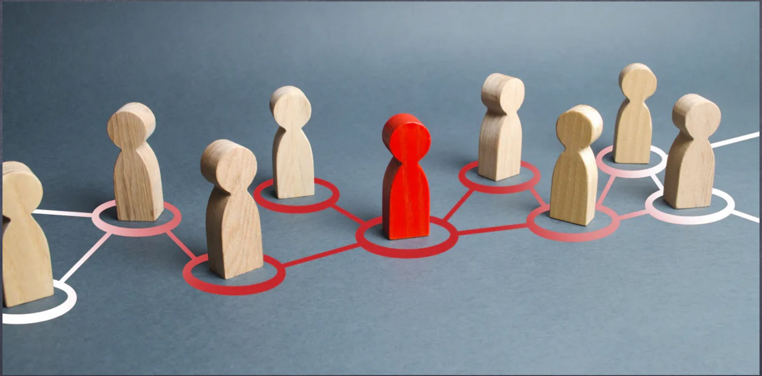
- For each problem area I will describe the motivation of the area in words

  ↪ Important

- I will describe our specific contributions

  ↪ Potentially Irrelevant

Seed detection in evolving networks

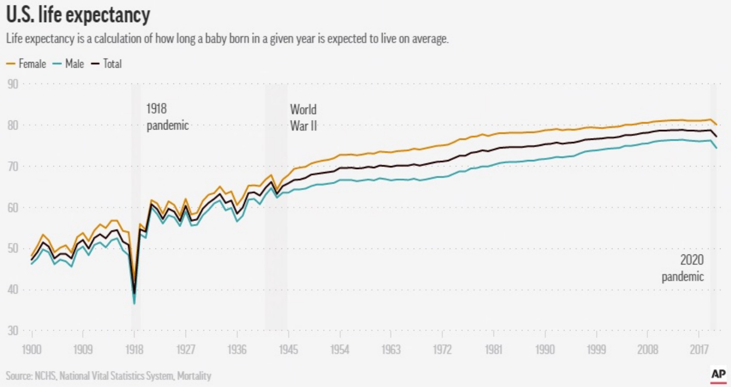


https://theconversation.com/patient-zero-why-its-such-a-toxic-term-134721

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.

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?

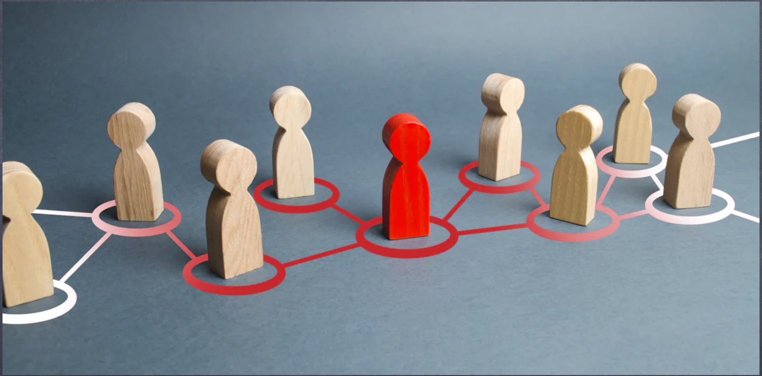
Network co-evolution: our motivation

- Most real world networks support some particular purpose (e.g. diffusion of information on Twitter)
- Co-evolution: Network influences individuals
  Individuals influence networks

- Till date majority of models deal either with
    - Dynamics on a fixed network (e.g. random walk or epidemics on a fixed network).
    - Dynamics "of" a network: Network itself changing in some fashion.
- However kept these two disciplines largely "separate". Most network practitioners believe co-evolving networks is the next frontier.

⟶ Goal: Understand conjectured phase transitions in one tractable model

Seed detection in evolving networks



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

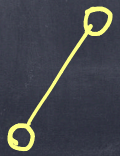
Our motivation in words

- 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.

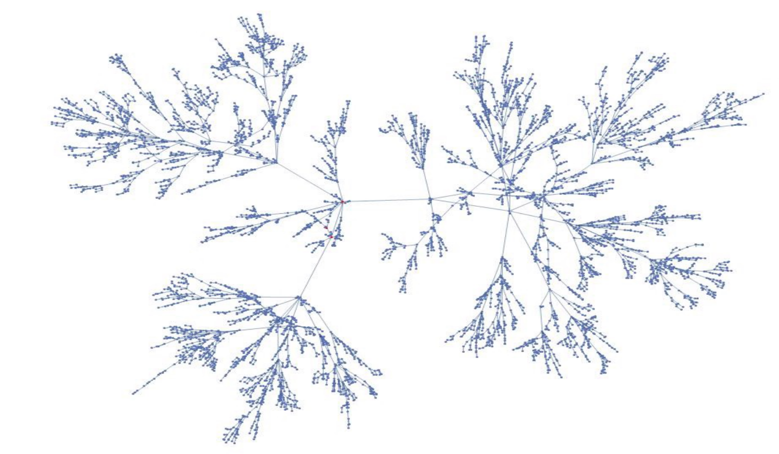
Probabilistic foundations

- 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)

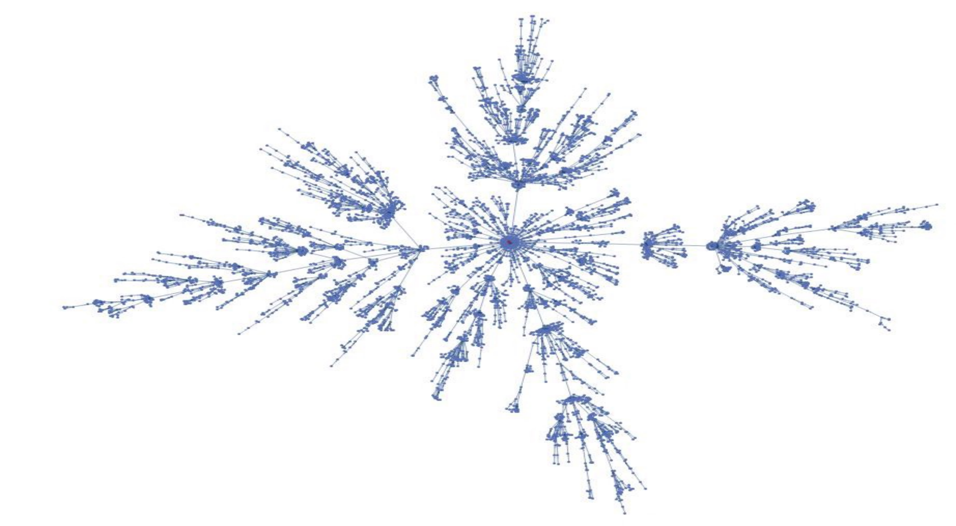


SIMULATION (n = 3000 ?)



Example $f(k) = k$ Preferential attachment

Simulation $(n = 5000)$



Setup:

- **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 **G**.
- Root finding algorithm: Fix $K \geq 1$ and a mapping H_K on **G** that takes an input finite unlabelled graph $\mathbf{g} \in \mathbf{G}$ and outputs a subset of K vertices from **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

$$\lim_{n \to \infty} \mathbb{P}(\text{seed} \in \text{outputted set of } \mathcal{T}_n) \geq 1 - \varepsilon$$

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)

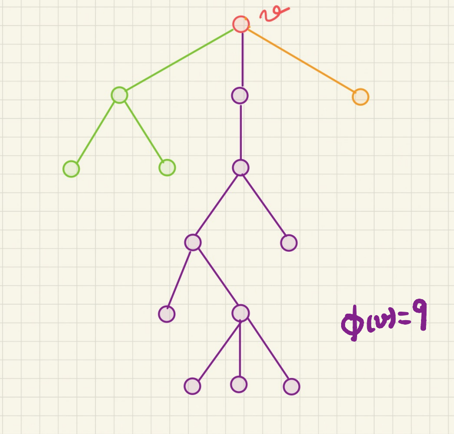
Persistence

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:** $f(k) = 1$

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

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

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

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

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

- **Sublinear preferential attachment:**

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

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?

- $f_* := \inf_{i \geq 0} f(i) > 0$; 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.

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

Lack of persistence

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:

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

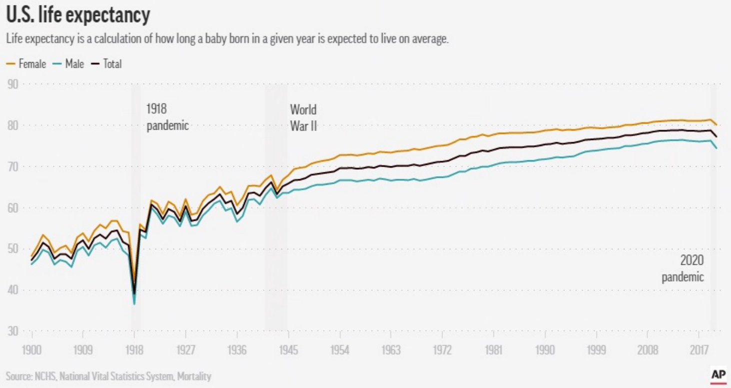
  where λ^* is the Malthusian rate of growth of the continuous time embedding.

- 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.

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

1:25 PM Fri Sep 3 79%

AA scholar.google.com

barabasi albert

Scholar About 47,400 results (0.03 sec) YEAR

User profiles for barabasi albert

Albert-László Barabási
Northeastern University, Harvard Medical School
Verified email at neu.edu
Cited by 269202

Statistical mechanics of complex networks [PDF] aps.org
R Albert, AL Barabási - Reviews of modern physics, 2002 - APS
Complex networks describe a wide range of systems in nature and society. Frequently cited examples include the cell, a network of chemicals linked by chemical reactions, and the Internet, a network of routers and computers connected by physical links. While traditionally ...
Cited by 24531 Related articles All 11 versions

Emergence of scaling in random networks [PDF] sciencemag.org
AL Barabási, R Albert - science, 1999 - science.sciencemag.org
Systems as diverse as genetic networks or the World Wide Web are best described as networks with complex topology. A common property of many large networks is that the vertex connectivities follow a scale-free power-law distribution. This feature was found to be ...
Cited by 40032 Related articles All 69 versions

Scale-free networks: a decade and beyond [PDF] sciencemag.org
AL Barabási - science, 2009 - science.sciencemag.org
For decades, we tacitly assumed that the components of such complex systems as the cell, the society, or the Internet are randomly wired together. In the past decade, an avalanche of research has shown that many real networks, independent of their age, function, and scope ...
Cited by 2040 Related articles All 14 versions

[BOOK] **The structure and dynamics of networks.**
ME Newman, ALE Barabási, DJ Watts - 2006 - psycnet.apa.org
From the Internet to networks of friendship, disease transmission, and even terrorism, the concept--and the reality--of networks has come to pervade modern society. But what exactly is a network? What different types of networks are there? Why are they interesting, and what ...
Cited by 3922 Related articles All 8 versions

Related searches

- barabasi albert **model**
- barabasi albert **graph**
- barabasi albert **preferential attachment**
- barabasi albert **lászló**
- barabasi albert **degree distribution**
- barabasi albert **emergence of scaling**
- barabasi albert **jeong**
- barabasi albert **erdös-rényi**

[HTML] **Interactome networks and human disease** [HTML] sciencedirect.c...
M Vidal, ME Cusick, AL Barabási - Cell, 2011 - Elsevier
Complex biological systems and cellular networks may underlie most genotype to

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

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

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

- $p_k \sim \dfrac{C}{k^{\alpha+3}}$ 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 !

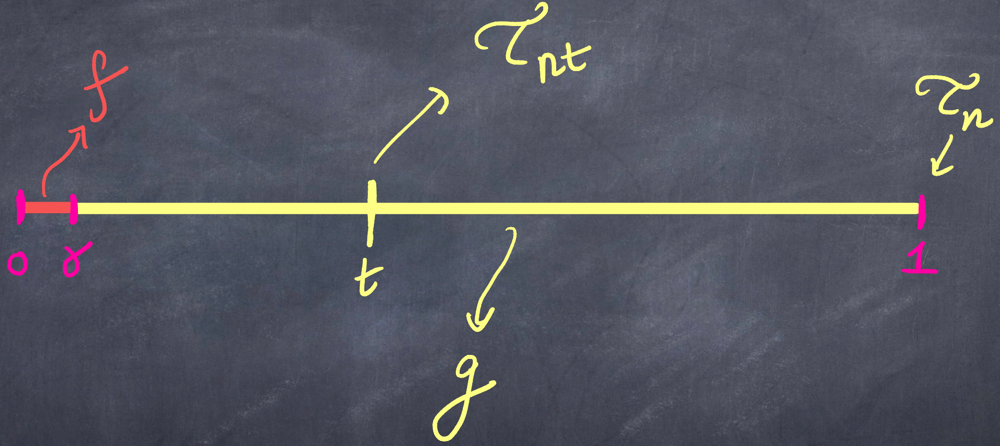
Standard change point model

- 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

$$g(k) = \text{general function}$$



τ_{nt}
τ_n
f
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

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

__Theorem__ [Banerjee, B, Carmichael]

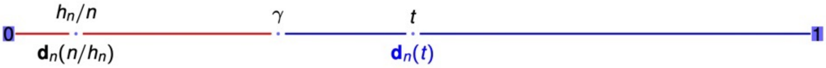
Under above technical conditions on f & g, irrespective of how small γ is f always wins!

- 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 distn $\left(\frac{N_k(nt)}{nt}\right)_{k \geq 1}$ 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 distn.

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

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

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

Lots of open problems

Simulations

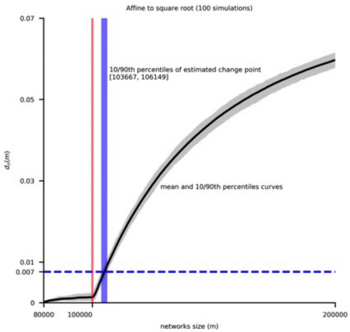


Figure: $n = 2 * 10^5$, $\gamma = 0.5$, $f_0(i) = i + 2$, $f_1(i) = \sqrt{i+2}$, $h_n = \log\log n$, $b_n = n^{1/\log\log n}$

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

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 $f_1(out - deg)$.

Result 1

- Here change point at n^{γ} (e.g. $\sqrt{n}$).
- Here
$$\frac{N_n(k)}{n} \xrightarrow{\mathrm{P}} p_k^1$$
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
$$\eta(\alpha, \beta) := \frac{\gamma(2+\beta) + (1-\gamma)(2+\alpha)}{(2+\alpha)(2+\beta)}. \qquad (5)$$

Motivation

- Most real world networks support some particular purpose (e.g. diffusion of information on Twitter)
- Co- evolution: Network influences individuals and vice- versa

Motivation 2: More sophisticated models for PA

Motivations Despite PA being heavily used, number of limitations

[I] Assumes global knowledge of network. Each new vertex needs complete knowledge of network

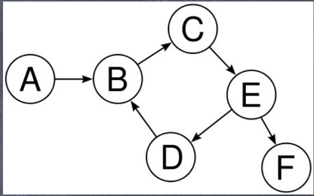
[II] In principle attractiveness should not depend ONLY on degree but potentially on "attenuated" neighborhood features.

Example: Page rank score attachment scheme.

Defn [Page rank Scores] Fix "damping factor" c.

For directed graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$, page rank score $(\pi_v : v \in \mathcal{V})$ is the stationary distr^n of a random walk that at each step

- with prob c does usual random walk using outgoing edges
- with prob $1-c$ jumps to a randomly selected vertex uniformly at random

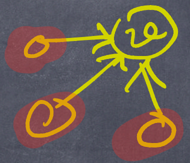


Computerscience wiki
Mr. Mckenty

Thus (π_v) satisfies linear system of equations

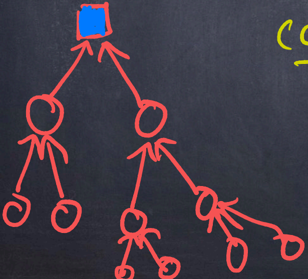
$$\pi_v = \frac{1-c}{n} + c \sum_{u \in N^-(v)} \pi_u / d^+(u)$$

out-degree of u



$N^-(v)$

Special case: Directed tree, directions to the root

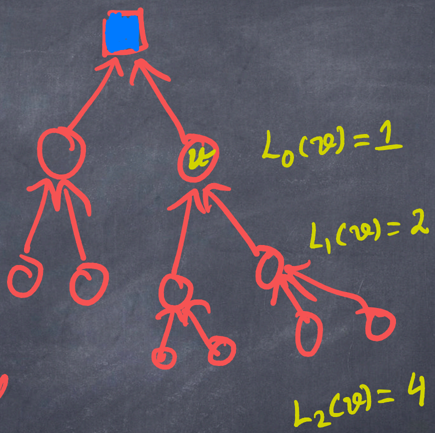


can check For $v \neq$ root

$$\pi_v = \frac{1-c}{n} \sum_{k=0}^{\infty} c^k L_k(v)$$

$L_k(v)$ = # of individuals at level k below v

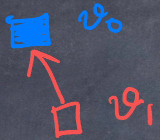
Thus if one does Preferential attachment using Page rank scores, then one does attachment using more global attractiveness function



III Motivation which might be contradictory to the previous motivations: Local exploration based attachment schemes

- Might want network evolution schemes where vertices decide to attach to a previous vertex after exploring the network "web-surfing" for some time.

Co-evolutionary network model ($\mathbf{P}$)

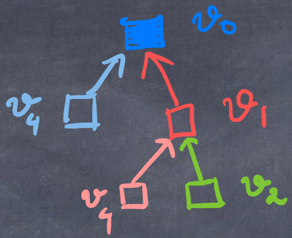
$T_1 =$ 

1. Having Constructed $\mathcal{T}_n$
2. At time $n+1$ a new vertex v_{n+1} enters system
3. Selects vertex V_n u.a.r in $\mathcal{T}_n$
4. Selects # of "exploration steps to root" variable

$$Z_{n+1} \sim \mathbf{P}$$

5. Goes up that many steps and attaches, stopping at root if need be.

$$\mathbf{P} = \text{pmf} = \{p_0, p_1, \ldots\} \qquad P(Z=i) = p_i \quad i \geq 0$$



Example

$C(2) = v_1, \quad Z_2 = 0$

$C(3) = v_2, \quad z_3 = 4$

$C(4) = v_2, \quad z_4 = 1$

$C(5) = v_0, \quad z_5 = 2$

$C(6) = v_2, \quad z_1 = 1$

$\vdots$

Special cases

① $p_0 \equiv 1$ $\rightarrow$ Random recursive tree
(Uniform Attachment)

② $p_0 = p$, $p_1 = 1-p$ $\rightarrow$ Preferential attachment
$$f(k) = k + \frac{(1-2p)}{p}$$

③ $p_0 = p$, $p_1 = p(1-p)$, $p_2 = p(1-p)^2$, ...

"Page rank model"

Theorem [Chebolu + Melsted 200x]

③ $\equiv$ Page rank attachment scheme with $1-c = p$

Theorems [Chebolut Melsted] Phase transition !

- If $P \leq \frac{1}{2}$

$$E(\text{degree of root}) = \tilde{\textcircled{H}}^{*}(n)$$

- If $P > \frac{1}{2}$

$$E(\text{degree of root}) = \tilde{\textcircled{H}}(n^{4pq})$$

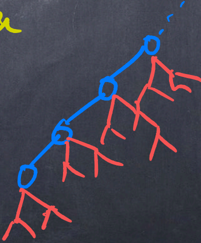
$* \; \tilde{\textcircled{H}} \Leftrightarrow O(n \log^{k}(n)) \qquad k \in \mathbb{Z}$

→ LOCAL WEAK CONVERGENCE

Theorem(s) [Banerjee, SB, Huang]

Let $Z \sim \mathbb{P}$

$\boxed{1}$ Assume $E(Z) < \infty$. Then the sequence of trees $\{\mathcal{T}_n\}_{n \geq 1}$ converge in the local weak convergene sense to a limiting infinite sin-tree



sin: tree with single infinite path to ∞

$\Rightarrow$ for example for every fixed $k \geq 0$

$$N_k(n) = \text{\# of vertices with } k \text{ children}$$

then $$\frac{N_k(n)}{n} \xmapsto{a.s} p_k \to \{p_k\}_{k \geq 0} = \text{PMF}$$

$\to$ If $E(Z) \leq 1$ $$\sum_{k=0}^{\infty} k p_k = 1$$

$\to$ If $E(Z) > 1$ $$\sum_{k=0}^{\infty} k p_k < 1$$

$\to$ Intuition for mass escaping to ∞. CONDENSAtion

$\boxed{2}$ Let $\{Z_i\}_{i\geq 1} \overset{IID}{=} \mathbb{P}$ $\quad f(s) = \text{pgf} = \sum_{k=0}^{\infty} p_k s^k$

Let $S_n = S_0 + \sum_{k=1}^{n} (Z_i - 1)$ = random walk started at S_0

Assumptions

① $p_0 \in (0,1)$, $p_0 + p_1 < 1$ [else if $p_0 + p_1 = 1 \Leftrightarrow$ preferential attachment regime]

② $f(s)$ is analytic at $s = 1$

Assumptions $\Rightarrow$ by work of [Daley 69] with a few more technical assumptions*

$\Rightarrow$ if we let $T_0 = \inf\{n \geq 1 : S_n = 0\}$ then

☐ If $E(Z) \neq 1$ then

$$P_1(n < T_0 < \infty) \approx e^{-n \log R}$$

i.e. $S_0 = 1$

for $R = R(P) > 1$

☐ If $E(Z) = 1$ then $R = 1$

* Aperiodicity + analyticity of pgf at $s = 1$.

Results [non-root degree]

Fix $k > 0$ and consider $D_{v_k}(n)$ = degree of v_k at time n

then $\forall \ \delta > 0$

$$\frac{D_{v_k}(n)}{n^{\frac{1}{k}-\delta}} \to \infty \qquad \frac{D_{v_k}(n)}{n^{1/k}(\log n)^{1+\delta}} \to 0$$

Results

If $E(Z) \le 1$

Let D = random variable with limiting degree distr^n

$$\lim_{k \to \infty} \frac{\log P(D \ge k)}{\log k} = -R$$

Intuitively $P(D \ge k) \simeq \dfrac{C}{k^{R}}$

↳ for $E(Z) > 1$ get upper + lower bounds for degree exponent

CONDENSATION

- Assume $E(Z) > 1$
- Few more technical conditions

Then

$$\frac{D_{v_0}(n)}{n} \xrightarrow{a.s.} \text{limit random variable} > 0$$

root

E.g. Random surfer model $p < \frac{1}{2}$ above is true!

No phase transition in Height

$$\varkappa_0 = \inf_{s \in (0,1)} \frac{f(s)}{s \log(1/s)}$$

$$\frac{H_n}{\log n} \xrightarrow{P} \varkappa_0$$

Connection between Random Walks + Trees

- what is the first step in studying such models?
- [Chebolu + Melsted idea for Page rank driven model]
- Fix a vertex $u \neq$ root
- Fix a vertex $t+1$ at time $\geq u$
- What is $P($ $t+1$ attaches to u / info till time $t)$?

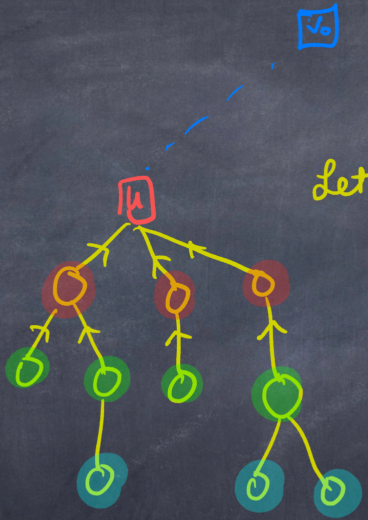
$L_0(t;u) = 1$

$L_1(t;u) = 3$

$L_2(t;u) = 4$

$L_3(t;u) = 3$

$\vdots$



Let $L_i(t;u)$ = # of vertices at distance i below vertex u at time t

$$P(t+1 \rightsquigarrow u \mid \text{information till time } t)$$

$$= \frac{L_0(t;u)}{t} p_0 + \frac{L_1(t;u)}{t} p_1 + \frac{L_2(t;u)}{t} p_2 + \cdots$$

<u>Check</u>: This leads to evolution equation for $\{L_k(t;u) : k \geq 0\}$ as

$L_k(t) = L_k(t;u)$

$$P\left(L_k(t+1) = L_k(t)+1 \mid \underset{\sim}{L}(t)\right)$$

$$= p_0 \frac{L_{k-1}(t)}{t} + p_1 \frac{L_k(t)}{t} + \cdots$$

$$= \frac{\left[A \cdot \underset{\sim}{L}(t)\right]_k}{t} \quad \text{where}$$

mass escaping above u.

$$A = \begin{pmatrix} 0 & 0 & 0 & \cdots \\ p_0 & p_1 & p_2 & \cdots \\ 0 & p_0 & p_1 & \cdots \\ 0 & 0 & p_0 & \cdots \\ 0 & \cdots & & \end{pmatrix}.$$

→ Easier to do things in continuous time. Continuous time version of what is happening below a vertex $u \neq$ root.

Let $\mathbb{T}$ denote the space of rooted, directed, labelled trees. Let $\mathcal{T}^*(\cdot)$ be the continuous time process of growing trees started with $\mathcal{T}^*(0) = \{v_0\}$,where v_0 is the root of the tree. The vertices in $\mathcal{T}^*(\cdot)$ are labelled $v_0, v_1, v_2, \ldots$ in order of appearance. $\mathcal{T}^*(\cdot)$ is generated by the following procedure:

"vertex u"

Each vertex reproduces at rate 1. When vertex v reproduces, a random variable Z following the law F is sampled independently.

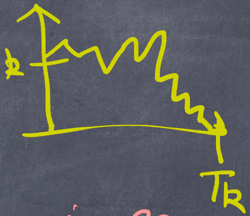
- If $Z \leq dist(v_0, v)$, then a new vertex $\tilde{v}$ is attached to the unique vertex u lying on the path between v and v_0 that satisfies $dist(v, u) = Z$ via a directed edge from $\tilde{v}$ to u.
- If $Z > dist(v_0, v)$, nothing occurs.

thus probability of new vertex being born to a current vertex = uniform distr^n

Two at first disjoint objects

Random walk: $S_n = S_0 + \sum_{i=1}^{n} (z_i - 1)$

$T_k = \inf\{n \geq 0 : S_n = 0 \mid S_0 = k\}$



Branching process: Let $L_k(t) = \#$ of vertices in generation k in the tree process described on previous page

Lemma:

$$E(L_k(t)) = \sum_{i=0}^{\infty} \frac{t^i}{i!} P(T_k = i)$$

Thank you for your attention!

ANY
QUESTIONS ?