Finding the Seed of Uniform Attachment Trees

Alan Pereira - UFGM
Gábor Lugosi - UPF

July 26, 2019

Introduction

Studies questions about old or extinct networks.

Introduction

Studies questions about old or extinct networks.

We want to find a source of a rumor/disease.

Introduction

Studies questions about old or extinct networks.

We want to find a source of a rumor/disease.

This problem was popularized by Shah-Zamah [6].

Setup

Setup

S_ℓ is a tree with $V(S_\ell) = \{1, \ldots, \ell\}$.

Setup

S_ℓ is a tree with $V(S_\ell) = \{1, \ldots, \ell\}$.

A random tree $T_n = T_n(S_\ell)$ with $V(T_n) = \{1, \ldots, n\}$ is a *uniform attachment tree with seed S_ℓ* if it is generated as follows:

Setup

S_ℓ is a tree with $V(S_\ell) = \{1, \dots, \ell\}$.

A random tree $T_n = T_n(S_\ell)$ with $V(T_n) = \{1, \dots, n\}$ is a *uniform attachment tree with seed* S_ℓ if it is generated as follows:

- $T_\ell = S_\ell$;

Setup

S_ℓ is a tree with $V(S_\ell) = \{1, \ldots, \ell\}$.

A random tree $T_n = T_n(S_\ell)$ with $V(T_n) = \{1, \ldots, n\}$ is a *uniform attachment tree with seed* S_ℓ if it is generated as follows:

- $T_\ell = S_\ell$;
- T_i is obtained by joining vertex i to a vertex of T_{i-1} chosen uniformly at random, independently of the past.

Influence of the seed

Bubeck, (Eldan), Mossel, and Rácz studied the influence of the seed in the growth of the random tree, first in preferential attachment [3] and after in uniform attachment [2].

Influence of the seed

Bubeck, (Eldan), Mossel, and Rácz studied the influence of the seed in the growth of the random tree, first in preferential attachment [3] and after in uniform attachment [2].

They did it by analysing

$$\delta(S^1, S^2) = \lim_{n\to\infty} TV(T_n(S^1), T_n(S^2))$$

Influence of the seed

Bubeck, (Eldan), Mossel, and Rácz studied the influence of the seed in the growth of the random tree, first in preferential attachment [3] and after in uniform attachment [2].

They did it by analysing

$$\delta(S^1, S^2) = \lim_{n\to\infty} TV(T_n(S^1), T_n(S^2))$$

Is δ a metric?

Influence of the seed

Bubeck, (Eldan), Mossel, and Rácz studied the influence of the seed in the growth of the random tree, first in preferential attachment [3] and after in uniform attachment [2].

They did it by analysing

$$\delta(S^1, S^2) = \lim_{n\to\infty} TV(T_n(S^1), T_n(S^2))$$

Is δ a metric?

Curien, Duquesne, Kortchemski and Manolescu: YES. [4]

The problem

The problem

Given $T_n(S_\ell)$ e want to find

- either a big set $H_1(T_n, \epsilon)$ such that

$$\mathbb{P}(H_1(T_n, \epsilon) \subset S_\ell) \geq 1 - \epsilon;$$

The problem

Given $T_n(S_\ell)$ e want to find

- either a big set $H_1(T_n, \epsilon)$ such that

$$\mathbb{P}(H_1(T_n, \epsilon) \subset S_\ell) \geq 1 - \epsilon;$$

- or a small a set $H_2(T_n, \epsilon)$ such that

$$\mathbb{P}(H_2(T_n, \epsilon) \supset S_\ell) \geq 1 - \epsilon.$$

Finding Adam

Finding Adam

Bubeck, Devroye, and Lugosi [1] considered the case $\ell = 1$ (in UA and PA).

Finding Adam

Bubeck, Devroye, and Lugosi [1] considered the case $\ell = 1$ (in UA and PA).

Jog and Loh [5] considered the same problem in non-linear preferential attachment.

Finding Adam

Bubeck, Devroye, and Lugosi [1] considered the case $\ell = 1$ (in UA and PA).

Jog and Loh [5] considered the same problem in non-linear preferential attachment.

We considered the cases

- Path P_ℓ.

Finding Adam

Bubeck, Devroye, and Lugosi [1] considered the case $\ell = 1$ (in UA and PA).

Jog and Loh [5] considered the same problem in non-linear preferential attachment.

We considered the cases

- Path P_ℓ.
- Star E_ℓ.

Finding Adam

Bubeck, Devroye, and Lugosi [1] considered the case $\ell = 1$ (in UA and PA).

Jog and Loh [5] considered the same problem in non-linear preferential attachment.

We considered the cases

- Path P_ℓ.
- Star E_ℓ.
- UART T_ℓ.

Theorem 1: $S_\ell = P_\ell$

Theorem 1: $S_\ell = P_\ell$

For $\ell \geq \max\left\{\frac{2e^2}{\gamma}\log\frac{1}{\epsilon}, \frac{2e^2}{\gamma}\log(4e^2)\right\}$ we have the following:

Theorem 1: $S_\ell = P_\ell$

For $\ell \geq \max\left\{ \frac{2e^2}{\gamma} \log \frac{1}{\epsilon}, \frac{2e^2}{\gamma} \log(4e^2) \right\}$ we have the following:
Given $T_n(P_\ell)$, $n >> 1$ we can find a set
$H_n = H_n(T_n, \varepsilon) \subset \{1, \ldots, n\}$ with $|H_n| \geq (1-\gamma)\ell$ such that

Theorem 1: $S_\ell = P_\ell$

For $\ell \geq \max\left\{\frac{2e^2}{\gamma}\log\frac{1}{\epsilon}, \frac{2e^2}{\gamma}\log(4e^2)\right\}$ we have the following:
Given $T_n(P_\ell)$, $n >> 1$ we can find a set
$H_n = H_n(T_n, \varepsilon) \subset \{1, \ldots, n\}$ with $|H_n| \geq (1-\gamma)\ell$ such that

$$\mathbb{P}\left\{H_n \subset P_\ell\right\} \geq 1 - \epsilon \, .$$

Theorem 2: $S_\ell = E_\ell$

For $\ell \geq \max\left\{C, \frac{8}{\gamma}\right\} \log \frac{1}{\epsilon}$ we have the following:
Given $T_n(E_\ell)$, $n >> 1$ we can find a set
$H_n = H_n(T_n, \varepsilon) \subset \{1, \ldots, n\}$ with $|H_n| \leq (1+\gamma)\ell$ such that

$$\mathbb{P}\left\{H_n \supset E_\ell\right\} \geq 1 - \epsilon \ .$$

Theorem 3: $S_\ell = T_\ell$

There exist c_1 and c_2 such that the following holds. Let
$\ell \geq c_1 \log^2 \frac{1}{\epsilon}$.
Given $T_n(T_\ell)$, $n >> 1$ we can find a set
$H_n = H_n(T_n, \varepsilon) \subset \{1, \ldots, n\}$ with $|H_n| \geq \ell/[c_2 \log(\ell/\epsilon)]$ such that

$$\mathbb{P}\{H_n \subset T_\ell\} \geq 1 - \epsilon\ .$$

Theorem 4

Theorem

Let $\epsilon \in (0, e^{-e^2})$*. Suppose that* T_n *is a uniform attachment tree with seed* $S_\ell = P_\ell$ *or* $S_\ell = E_\ell$ *for* $\ell \leq \frac{\log(1/\epsilon)}{\log\log(1/\epsilon)}$*. Then, for all* $n \geq 2\ell$*, any seed-finding algorithm that outputs a vertex set* H_n *of size* ℓ *has*

$$\mathbb{P}\left\{|H_n \cap S_\ell| \leq \frac{\ell}{2}\right\} \geq \epsilon \ .$$

How can we prove it?

How can we prove it?

The main idea is to prove that old vertices are more central than the new vertices (in some sense).

How can we prove it?

The main idea is to prove that old vertices are more central than the new vertices (in some sense).

The set H_n will be the set of the most central vertices in T_n.

How can we prove it?

The main idea is to prove that old vertices are more central than the new vertices (in some sense).

The set H_n will be the set of the most central vertices in T_n.

Let us define what means be more central.

Rooted tree and Induced subtree

Rooted tree and Induced subtree

A rooted tree (T, v) is the tree T with a distinguished vertex $v \in V(T)$.

Rooted tree and Induced subtree

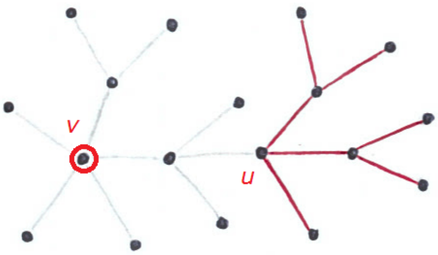
A rooted tree (T, v) is the tree T with a distinguished vertex $v \in V(T)$.

The subtree induced by u $(T, v)_{u\downarrow}$ is the subtree of (T, v) which grows from u in the opposite direction of v.

Rooted tree and Induced subtree

A rooted tree (T, v) is the tree T with a distinguished vertex $v \in V(T)$.

The subtree induced by u $(T, v)_{u\downarrow}$ is the subtree of (T, v) which grows from u in the opposite direction of v.



Centrality: Definition

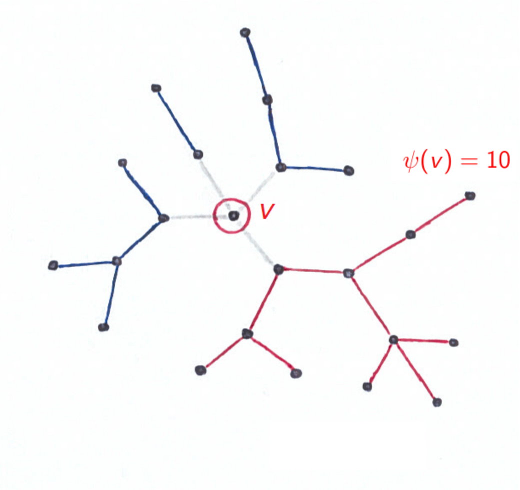
Given a tree T, the *anti-centrality* of a vertex $v \in V(T)$ is defined by

$$\psi(v) = \max_{u \in N(v)} |(T, v)_{u\downarrow}| \ .$$

Centrality: Definition

Given a tree T, the *anti-centrality* of a vertex $v \in V(T)$ is defined by

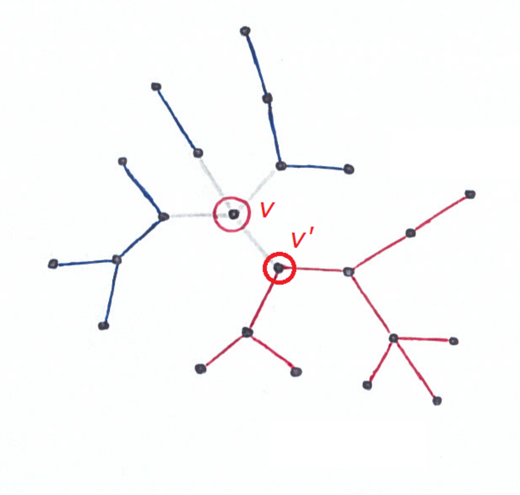
$$\psi(v) = \max_{u \in N(v)} |(T, v)_{u\downarrow}| \ .$$



Centrality

Given v, we denote v' to be some vertex in $N(v)$ such that

$$\psi(v) = \left|(T, v)_{v'\downarrow}\right| .$$



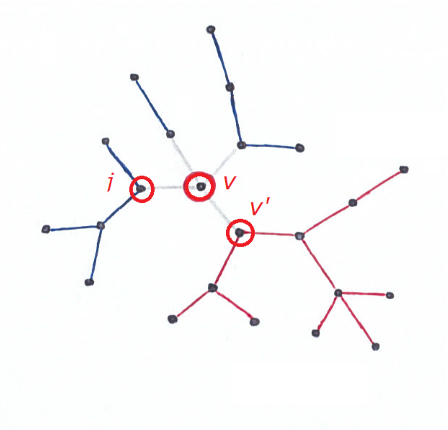
Comparing Centrality

Comparing Centrality

Case 1: When v is between v' and j we have $\psi(v) \leq \psi(j)$

Comparing Centrality

Case 1: When v is between v' and j we have $\psi(v) \leq \psi(j)$



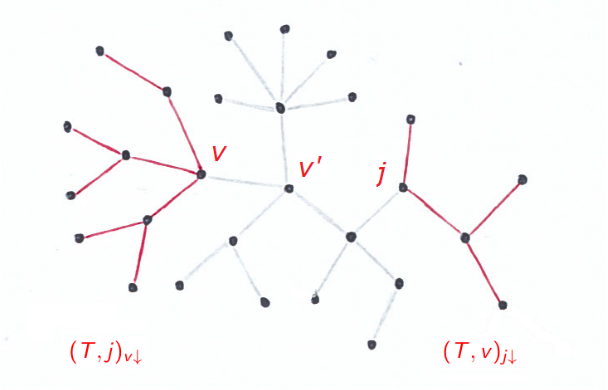
Comparing Centrality

Case 2: When v' is between v and j we have $\psi(v) \leq \psi(j)$ if

Comparing Centrality

Case 2: When v' is between v and j we have $\psi(v) \leq \psi(j)$ if

- $|(T,j)_{v\downarrow}| \geq |(T,v)_{j\downarrow}|$;



Sketch of the proof: Case $S_\ell = P_\ell$

Sketch of the proof: Case $S_\ell = P_\ell$

We will prove that
old central vertices are more central than new vertices.

Sketch of the proof: Case $S_\ell = P_\ell$

We will prove that
old central vertices are more central than new vertices.

More precisely

$$\mathbb{P}\left\{ \max_{\ell\gamma/2 \le j \le \ell(1-\gamma/2)} \psi(j) < \min_{\ell < i \le n} \psi(i) \right\} \ge 1 - \epsilon \ .$$

Sketch of the proof: Case $S_\ell = P_\ell$

We will prove that
old central vertices are more central than new vertices.

More precisely

$$\mathbb{P}\left\{ \max_{\ell\gamma/2 \le j \le \ell(1-\gamma/2)} \psi(j) < \min_{\ell < i \le n} \psi(i) \right\} \ge 1 - \epsilon \ .$$

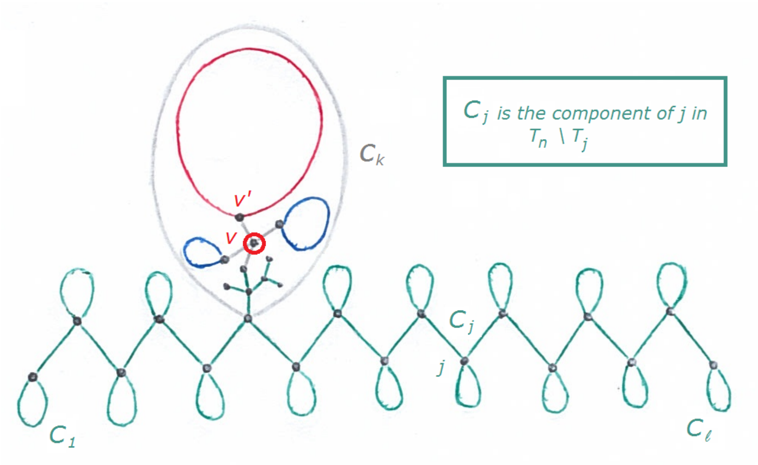
Let us prove that the complement has small probability.

Sketch of the proof

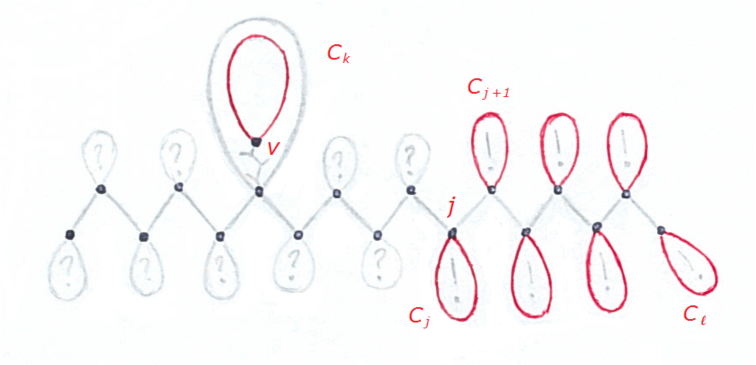
By Union Bound we have

$$\begin{aligned}\mathbb{P}\left\{\min_{\ell<i\leq n}\psi(i)\leq\max_{\ell\gamma/2\leq j\leq\ell(1-\gamma/2)}\psi(j)\right\} &\leq \sum_{j=\gamma\ell/2}^{(1-\gamma/2)\ell}\mathbb{P}\left\{\min_{\ell<i\leq n}\psi(i)\leq\psi(j)\right\}\\ &\leq \sum_{j=\gamma\ell/2}^{(1-\gamma/2)\ell}\sum_{k=1}^{\ell}\mathbb{P}\Big\{\exists v\in C_k\setminus\{k\}:\psi(v)\leq\psi(j)\Big\}\,.\end{aligned}$$

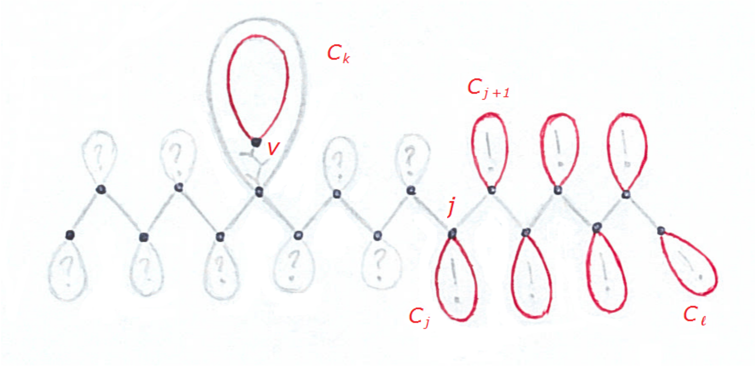
First case: (v', v, j)



Second case: (v, v', j)



Second case: (v, v', j)



The value of ℓ arise from a optimization of the bounds.

Sébastien Bubeck, Luc Devroye, and Gábor Lugosi.
Finding Adam in random growing trees.
Random Structures & Algorithms, 50(2):158–172, 2017.

Sébastien Bubeck, Ronen Eldan, Elchanan Mossel, and Miklós Rácz.
From trees to seeds: on the inference of the seed from large trees in the uniform attachment model.
Bernoulli, 23(4A):2887–2916, 2017.

Sébastien Bubeck, Elchanan Mossel, and Miklós Z Rácz.
On the influence of the seed graph in the preferential attachment model.
IEEE Transactions on Network Science and Engineering, 2(1):30–39, 2015.

Nicolas Curien, Thomas Duquesne, Igor Kortchemski, and Ioan Manolescu.
Scaling limits and influence of the seed graph in preferential attachment trees.

Journal de l'École Polytechnique–Mathématiques, 2:1–34, 2015.

Varun Jog and Po-Ling Loh.
Analysis of centrality in sublinear preferential attachment trees via the CMJ branching process.
IEEE Transactions on Network Science and Engineering, 4(1):1–12, 2017.

Devavrat Shah and Tauhid R. Zaman.
Rumors in a network: Who's the culprit?
IEEE Transactions on Information Theory, 57(8):5163–5181, 2011.


Thank you