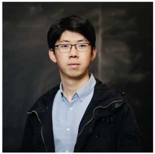
A Proof of The Changepoint Detection Threshold Conjecture in Preferential Attachment Models

Shuyang Gong

School of Mathematical Sciences
Peking University

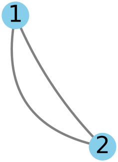



Hang Du (MIT) and Jiaming Xu (Duke)

38th Annual Conference on Learning Theory
Lyon, France

Preferential attachment models

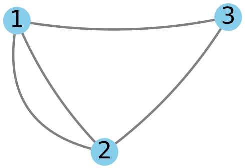
Initial graph G_2 consists of two vertices connected by m parallel edges



Preferential attachment models

At each time t, a new vertex t arrives and forms m edges, one at a time, to existing nodes $v \in [t-1]$:

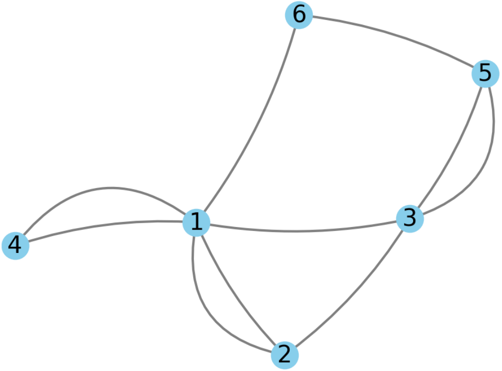
$$\mathbb{P}\{t \to v\} \propto \mathsf{deg}(v) + \delta_t$$



Preferential attachment models

At each time t, a new vertex t arrives and forms m edges, one at a time, to existing nodes $v \in [t-1]$:

$$\mathbb{P}\{t \to v\} \propto \deg(v) + \delta_t$$



Changepoint detection problem

Definition

$$\mathcal{H}_0 : \delta_t = \delta$$
$$\mathcal{H}_1 : \delta_t = \delta \mathbf{1}_{t \leq \tau_n} + \delta' \mathbf{1}_{\tau_n < t \leq n}$$

Changepoint detection problem

Definition

$$\mathcal{H}_0 : \delta_t = \delta$$
$$\mathcal{H}_1 : \delta_t = \delta \mathbf{1}_{t \le \tau_n} + \delta' \mathbf{1}_{\tau_n < t \le n}$$

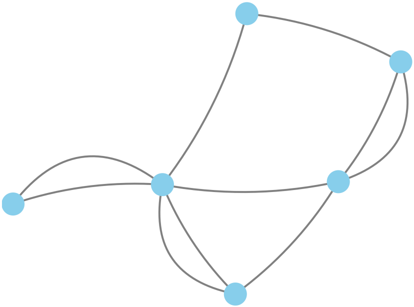
- $\delta \neq \delta' > -m$ are two fixed constants

Changepoint detection problem

Definition

$$\mathcal{H}_0 : \delta_t = \delta$$
$$\mathcal{H}_1 : \delta_t = \delta \mathbf{1}_{t \le \tau_n} + \delta' \mathbf{1}_{\tau_n < t \le n}$$

- $\delta \neq \delta' > -m$ are two fixed constants
- Only final network snapshot is observed (node arrival time unknown)



Changepoint detection problem

Definition

$$\mathcal{H}_0 : \delta_t = \delta$$
$$\mathcal{H}_1 : \delta_t = \delta \mathbf{1}_{t \le \tau_n} + \delta' \mathbf{1}_{\tau_n < t \le n}$$

- $\delta \neq \delta' > -m$ are two fixed constants
- Only final network snapshot is observed (node arrival time unknown)
- Problem gets harder with increasing τ_n

Changepoint detection problem

Definition

$$\mathcal{H}_0 : \delta_t = \delta$$
$$\mathcal{H}_1 : \delta_t = \delta \mathbf{1}_{t \le \tau_n} + \delta' \mathbf{1}_{\tau_n < t \le n}$$

- $\delta \neq \delta' > -m$ are two fixed constants
- Only final network snapshot is observed (node arrival time unknown)
- Problem gets harder with increasing τ_n
- Changepoint localization: estimate τ_n under $\mathcal{H}_1$ [Bhamidi-Jin-Nobel '18]

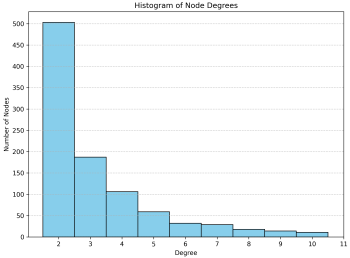
Changepoint detection problem

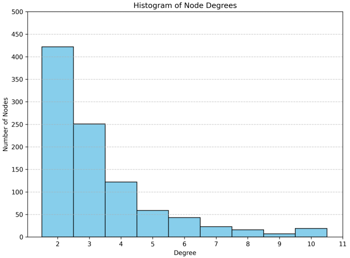
Definition

$$\mathcal{H}_0 : \delta_t = \delta$$
$$\mathcal{H}_1 : \delta_t = \delta \mathbf{1}_{t \le \tau_n} + \delta' \mathbf{1}_{\tau_n < t \le n}$$

- $\delta \neq \delta' > -m$ are two fixed constants
- Only final network snapshot is observed (node arrival time unknown)
- Problem gets harder with increasing τ_n
- Changepoint localization: estimate τ_n under $\mathcal{H}_1$ [Bhamidi-Jin-Nobel '18]
- Applications: detect structural changes in various settings, such as communication networks, social networks, financial networks, and biological networks [Cirkovic-Wang-Zhang '24].

A simple test based on minimum-degree





$n = 1000,\ m = 2,\ \delta(t) \equiv 0$ $\qquad$ $n = 1000,\ m = 2,\ \delta(t) = 10 \cdot \mathbf{1}\left(t > n - n^{0.8}\right)$

- [Bet-Bogerd-Castro-van der Hofstad '23] achieves strong detection by thresholding the number of degree m vertices.
- Intuition: Mean difference $= \Theta(n - \tau_n)$, while standard deviation $= O(\sqrt{n})$.

Changepoint detection conjecture

Conjecture (Bet-Bogerd-Castro-van der Hofstad '23)

Suppose $\tau_n = n - cn^\gamma$ *for a constant* c *and* $\gamma < 1/2$.

1. *All tests based on vertex degrees are powerless.*
2. *All tests are powerless.*

- Part 2 of the conjecture is particularly striking, because, if true, neither degree information nor any higher-level graph structure is useful for detection when $\gamma < 1/2$

Significant progress

Theorem (Kaddouri-Naulet-Gassiat '24)

Suppose $\tau_n = n - \Delta$. *If* $\Delta = o(n^{1/3})$ *for* $\delta > 0$ *or* $\Delta = o(n^{1/3}/\log n)$ *for* $\delta = 0$, *then*

$$\mathbb{P}_0(A_n) \to 0 \Longrightarrow \mathbb{P}_1(A_n) \to 0, \quad \textit{for all sequences of events } A_n$$

- As a consequence, $\mathrm{TV}(\mathbb{P}_0, \mathbb{P}_1) \le 1 - \Omega(1) \Rightarrow$ strong detection is impossible
- Does not cover the entire regime $\Delta = o(\sqrt{n})$ and the regime $\delta < 0$
- Does not rule out the possibility of weak detection

Our resolution

Theorem (Du-G.-Xu '25)

Suppose $\tau_n = n - \Delta$. *If* $\Delta = o(n^{1/2})$, *then*

$$\mathrm{TV}(\mathbb{P}_0, \mathbb{P}_1) = o(1)$$

Our resolution

Theorem (Du-G.-Xu '25)

Suppose $\tau_n = n - \Delta$. *If* $\Delta = o(n^{1/2})$, *then*

$$\mathrm{TV}(\mathbb{P}_0, \mathbb{P}_1) = o(1)$$

- As a consequence, all tests are powerless $\Rightarrow$ resolves the changepoint detection conjecture [Bet-Bogerd-Castro-van der Hofstad '23] in positive
- We prove a stronger statement: all tests remain powerless even if, in addition to G_n, the entire network history were observed up to time $n - N$ for $\Delta^2 \ll N \ll n$
- As a corollary, we prove no estimator can locate τ_n within $o(\sqrt{n})$ with $\Omega(1)$ probability $\Rightarrow$ the estimator in [Bhamidi-Jin-Nobel'18], which achieves $|\hat{\tau}_n - \tau_n| = O_P(\sqrt{n})$, is order-optimal

Proof ideas

Challenge of directly bounding second-moment

Define the Likelihood ratio

$$L(G) \triangleq \frac{\mathbb{P}_1(G)}{\mathbb{P}_0(G)}$$

Then

$$\mathbb{E}_{G_n \sim \mathbb{P}_0}\left[L^2(G_n)\right] = 1 + o(1) \implies \mathrm{TV}(\mathbb{P}_1, \mathbb{P}_0) = o(1)$$

Challenge of directly bounding second-moment

Define the Likelihood ratio

$$L(G) \triangleq \frac{\mathbb{P}_1(G)}{\mathbb{P}_0(G)}$$

Then

$$\mathbb{E}_{G_n \sim \mathbb{P}_0}\left[L^2(G_n)\right] = 1 + o(1) \implies \mathrm{TV}(\mathbb{P}_1, \mathbb{P}_0) = o(1)$$

- Widely used to prove impossibility of detection in high-dimensional statistics and network analysis (e.g. community detection)

Challenge of directly bounding second-moment

Define the Likelihood ratio

$$L(G) \triangleq \frac{\mathbb{P}_1(G)}{\mathbb{P}_0(G)}$$

Then

$$\mathbb{E}_{G_n \sim \mathbb{P}_0}\left[L^2(G_n)\right] = 1 + o(1) \implies \mathrm{TV}(\mathbb{P}_1, \mathbb{P}_0) = o(1)$$

- Widely used to prove impossibility of detection in high-dimensional statistics and network analysis (e.g. community detection)
- However, since only final network snapshot is observed, $L(G_n)$ involves an average over compatible network histories, making it hard to bound its second-moment directly

Our proof strategy

- Interpolation
  - $\mathbb{P}_{n,n-k}$: distribution of G_n with changepoint at time $n-k$
  - reduce to analyzing changepoint $\tau_n = n-1$:

$$\mathrm{TV}(\mathbb{P}_{n,n-1}, \mathbb{P}_{n,n}) = O\left(\frac{1}{\sqrt{N}}\right).$$

Our proof strategy

- Interpolation
  - $\mathbb{P}_{n,n-k}$: distribution of G_n with changepoint at time $n-k$
  - reduce to analyzing changepoint $\tau_n = n-1$:

$$\mathrm{TV}(\mathbb{P}_{n,n-1}, \mathbb{P}_{n,n}) = O\left(\frac{1}{\sqrt{N}}\right).$$

- "Easier" model: reveal the network history up to time $M = n - N$
  - The node arrival times are known.
  - Largely simplifies the form of likelihood ratio.

$$L = \frac{C_1}{N} \sum_{v \in V} |\mathcal{C}(v)| \lambda_v X_v,$$

Our proof strategy

- Interpolation
  - $\mathbb{P}_{n,n-k}$: distribution of G_n with changepoint at time $n-k$
  - reduce to analyzing changepoint $\tau_n = n-1$:

$$\mathrm{TV}(\mathbb{P}_{n,n-1}, \mathbb{P}_{n,n}) = O\left(\frac{1}{\sqrt{N}}\right).$$

- "Easier" model: reveal the network history up to time $M = n - N$
  - The node arrival times are known.
  - Largely simplifies the form of likelihood ratio.

$$L = \frac{C_1}{N} \sum_{v \in V} |\mathcal{C}(v)| \, \lambda_v X_v,$$

- Control the second moment
  - The components are approximately uncorrelated.
  - Tools: Combinatorial coupling, Efron-Stein inequality.

Concluding remarks

- We show changepoint detection threshold is $\tau_n = n - o(\sqrt{n})$, confirming a conjecture of [Bet-Bogerd-Castro-van der Hofstad '23]
- As by-product, we show changepoint localization threshold is also $\sqrt{n}$, matching upper bound in [Bhamidi-Jin-Nobel '18]
- Key proof ideas: reduces to bounding TV when changepoint occurs at $n-1$, reveal network history up to $n - o(n)$, and bound the second-moment of likelihood ratio.

Future directions

- General attachment rule: $\mathbb{P}(t \to v) \propto f(\text{deg}(v))$ [Banerjee-Bhamidi-Carmichael '22]
- Changepoint detection in general dynamic graph models
- Other related reconstruction and estimation problems in PA graphs

References

- Hang Du, Shuyang Gong, & Jiaming Xu. *A Proof of The Changepoint Detection Threshold Conjecture in Preferential Attachment Models*, arXiv:2502.00514, v3.