A proof of the changepoint detection threshold conjecture in preferential attachment models

Hang Du [1] Shuyang Gong [2] Jiaming Xu [3]

[1]MIT [2]Peking University [3]Duke University

Motivation

- A statistical technique for identifying points in a data sequence where key properties change significantly, such as mean, variance, distribution, or correlation structure.
- Example: A time series $X_1, \dots, X_m, Y_{m+1}, \dots, Y_n$, where a change occurs at an unknown position m.
- Goal: Detect when and where such changes occur (i.e., determine m).
- Applications: - Finance (stock price shifts, volatility changes) - Bioinformatics (structural variations in DNA sequences) - Climate Science (shifts in temperature trends) - Quality Control (monitoring manufacturing defects)
- A convergence of real-world applications and theoretical foundations.

For the changepoint detection in networks, our observation is not real-valued data, but structure of networks. The following is some applications of changepoint detection in real-world network models.

- Cybersecurity and Fraud Detection
  - Detecting network intrusions and abnormal traffic patterns.
  - Identifying sudden shifts in transaction networks for fraud detection.
- Social Network Analysis
  - Monitoring evolving communities and detecting structural changes.
  - Identifying disinformation campaigns or viral content propagation.
- Financial Networks
  - Detecting market manipulation and insider trading activities.
  - Identifying systemic risks in interbank lending networks.
- Biological and Brain Networks
  - Tracking changes in brain connectivity (e.g., epilepsy detection).
  - Identifying mutations or structural variations in genetic networks.
- Transportation and Communication Networks - Monitoring traffic congestion and detecting network disruptions. - Identifying shifts in telecommunication patterns.

Mathematical Settings

- Preferential Attachment Model (PA model)
- Random growing graph model, evolving with time t.

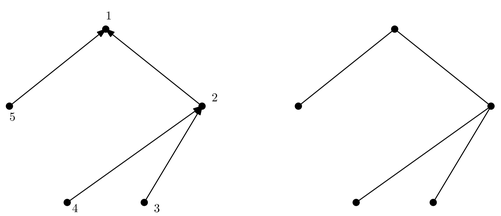


Figure 1. Preferential attachment model

- At time t, a new vertex v_t comes into the play and choose a vertex in G_{t-1} with probability proportional to the degrees in G_{t-1}.
- The observation is the unlabeled graph.

Why preferential attachment model?

- Degree distribution has power-law decay [Bollobás-Riordan-Spencer-Tusnády'01]

$$P(k) \propto \frac{1}{k^\gamma}.$$

  - Few vertices have large degree. rich-get-richer phenomenon.
  - For real-world networks, $2 < \gamma < 3$.
  - e.g. Social network (Twitter, Facebook), Biological network (protein-protein network), Citation network (some papers 1000+ citation, while others only a few)
- Exhibit "small world" phenomenon. [Hofstad-Zhu'25+] meaning that two "typical" vertices have small distance.
  - The average steps to reach from one node to another is small
  - 1. Social network. e.g., the Erdős number, 2. Epidemiology. e.g. COVID-19.

What causes structural change in this model? How to model the preferences?

- At each time t, δ_t is attachment parameter independent of the vertices in G_t
- Given G_t constructed at t, a new vertex v_{t+1} at time $t+1$

$$\mathbb{P}[v_{t+1} \to v] \propto \mathbf{deg}_{G_t}(v) + \delta_t$$

- Two extreme cases:
  - $\delta_t = \infty$ ignore the degree, attach uniformly.
  - $\delta_t = 0$ classical uniform attachment.
  - The smaller δ_t, the stronger the preference for high-degree vertices.
- **Question:** the detection problem (the arrival time is unknown, unlabeled graph):

$$H_0 : \delta_t = \delta, \text{ for some constant } \delta,$$
$$H_1 : \delta_t = \delta \mathbf{1}_{t \le \tau_n} + \delta' \mathbf{1}_{\tau_n + 1 \le t \le n}.$$

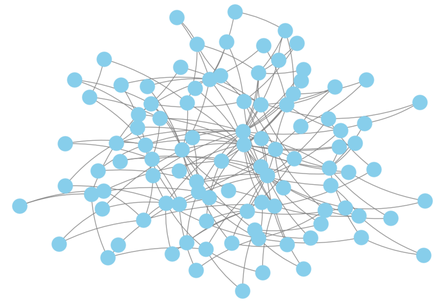
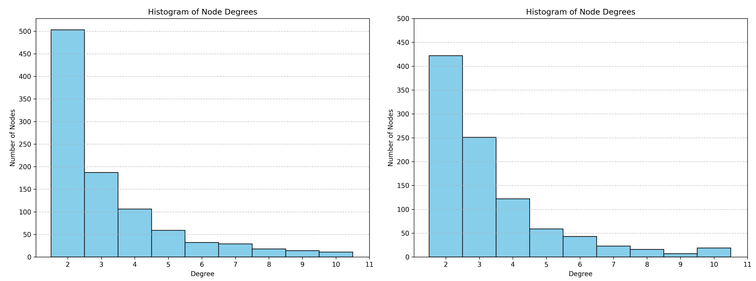


Figure 2. **Observation**: final snapshot of PA model (no access to arrival times)

Previous results



Positive results:

- [Bhamidi-Jin-Nobel'18] considered the changepoint recovery problem (under H_1):
  - Suppose that the changepoint $\tau_n \ge \varepsilon n$, there exists an estimator $\hat{\tau}_n$ that satisfies

$$|\hat{\tau}_n - \tau_n| = O_P\left(\log n \sqrt{n}\right).$$

  - They also established many other results, such as the power-law decay of degree distribution.
- [Bet-Bogerd-Castro-Hofstad'23+] considered the detection problem H_0 vs H_1:
  - They showed that if $\tau_n = n - \Delta_n$ where $\Delta_n/\sqrt{n} \to \infty$, then by counting the number of leaves, strong detection of H_0 and H_1 is achievable.

$$T(G_n) = N_1(n) - np_1(\delta).$$

  - They made the **conjecture** that when $\Delta_n = o(\sqrt{n})$, detection is impossible.
  1. All tests based on vertex degrees are powerless.
  2. All tests are powerless.

  The part 2 is particularly striking because it basically says neither vertex degrees nor high-level graph structures, e.g. cycles, trees... are useful for detection.

Negative results:

- [Kaddouri-Naulet-Gassiat'24+] showed that when $\Delta_n = o(n^{1/3})$, detection is impossible. They further prove that if the arrival times are revealed, then $\Delta \to \infty$ is the threshold for detection.

Our results

In [Du-G'-Xu'25+], we confirmed the conjecture by [Bet-Bogerd-Castro-Hofstad'23+] and showed that when $\Delta_n = o(\sqrt{n})$, detection is impossible. As a corollary, we demonstrated that the estimator in [Bhamidi-Jin-Nobel'18] is order-optimal.

Theorem (Du-G'-Xu'25+)

If changepoint τ_n satisfies $\Delta_n = n - \tau_n = o(\sqrt{n})$, then

$$\mathrm{TV}(\mathbb{P}_n, \mathbb{P}_{n,\tau_n}) = o(1),$$

where $\mathbb{P}_n$ is the law of G_n under H_0 (no changepoint), $\mathbb{P}_{n,\tau_n}$ is the law of G_n under H_1 (changepoint at τ_n)

The next theorem shows that the estimator in [Bhamidi-Jin-Nobel'18] is order-optimal.

Theorem (Du-G'-Xu'25+)

Assume that there exists $\varepsilon \in (0,1)$ such that $\tau_n \ge \varepsilon n$. Then there is no estimator $\hat{\tau}_n$ based on $G_n \sim \mathbb{Q}_{m,n,\delta,\delta',\tau_n}$ such that $|\hat{\tau}_n - \tau_n| = o(\sqrt{n})$ holds with non-vanishing probability uniformly for all $\tau_n \in [\varepsilon n, n]$.

Challenge of directly bounding second-moment

Define the Likelihood ratio

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

Then

$$\mathbb{E}_0[L^2(G)] = 1 + o(1) \Rightarrow \mathrm{TV}(\mathbb{P}_1, \mathbb{P}_0) = o(1).$$

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.

Proof ideas

Proof idea 1: interpolation

- Intermediate measures: construct a sequence of probability measures: $\{\mathbb{P}_{n,n-k}, 0 \le k \le \Delta_n\}$, $\mathbb{P}_{n,n-k}$ means changepoint occurs at time $n-k$.
- To show (by triangle inequality, the theorem follows)

$$\mathrm{TV}(\mathbb{P}_{n,n-k}, \mathbb{P}_{n,n-k+1}) = o\left(\frac{1}{\Delta}\right).$$

- Data processing inequality: it suffices to show

$$\mathrm{TV}(\mathbb{P}_{n,n-1}, \mathbb{P}_n) = o\left(\frac{1}{\Delta}\right).$$

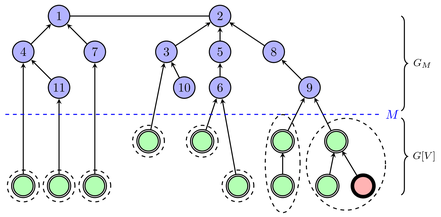
  simplifying a ton of analysis.....

Proof idea 2: the "easier" model

- Including the movies before time $M = n - N$ where $\Delta^2 \ll N = o(n)$ in $\mathbb{P}_n$ and $\mathbb{P}_{n,n-1}$, not only the final snapshot G_n.
- "Easier model" : probability measures on $(\overline{G}_M, G_n)$: $\mathcal{P}_n$ and $\mathcal{P}_{n-1}$.
- $\mathbb{P}_n$ and $\mathbb{P}_{n,n-1}$ are marginals of $\mathcal{P}_n$ and $\mathcal{P}_{n-1}$.

$$\begin{aligned}\mathrm{TV}(\mathbb{P}_n, \mathbb{P}_{n,n-1}) &\le \mathrm{TV}(\mathcal{P}_n, \mathcal{P}_{n-1}) \\ &\le \mathbb{E}_{\overline{G}_M}\left[\mathrm{TV}(\mathcal{P}_{n|\overline{G}_M}, \mathcal{P}_{n-1|\overline{G}_M})\right]\end{aligned}$$

- Final target: RHS is $o(1/\Delta)$.
- Similar technique has also been applied in [Kaddouri-Naulet-Gassiat'24+].



Bounding the likelihood ratio

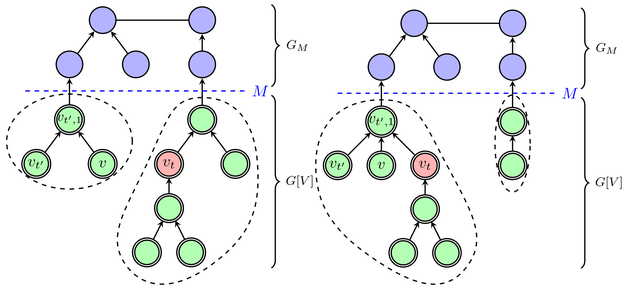
- Fix $\overline{G}_M$ $(L = \frac{d\mathcal{P}_{n-1|\overline{G}_M}}{d\mathcal{P}_{n|\overline{G}_M}})$

$$\mathrm{TV}(\mathcal{P}_{n|\overline{G}_M}, \mathcal{P}_{n-1|\overline{G}_M}) = \mathbb{E}_{\mathcal{P}_{n|\overline{G}_M}}|L-1| \le \sqrt{\mathrm{Var}_{\mathcal{P}_{n|\overline{G}_M}}[L]}.$$

- L has the form (Most last N vertices have $\mathcal{C}(v) = 1$)

$$L = \frac{C}{N} \sum_{v \in [n] \setminus [n-N]} |\mathcal{C}(v)| \lambda_v X_v,$$

  where $\sum_{w \in \mathcal{C}(v)} \lambda_w = 1$ and $c_1 < X_v < c_2$ uniformly bounded.
- Different components are "independent", then $\mathrm{Var}[L] = \Theta(\frac{1}{N})$.



Proof idea 3: coupling

- Encode the conditional laws $\mathcal{P}_{\overline{G}_M}$ by iid $\mathrm{Unif}[0,1]$ $(U_1, \dots, U_N)$.
- L is a function of $U_1, \dots, U_N$.

$$L = f(U_1, \dots, U_N).$$

- Use **Efron-Stein** to control variance

$$\begin{aligned}&\mathrm{Var}[L] \\ &\le \sum_{i=1}^N \mathbb{E}\left[\left(f(U_1, \dots, U_i, \dots, U_N) - f(U_1, \dots, \widetilde{U}_i, \dots, U_N)\right)^2\right] \\ &\le N \times \Theta\left(\frac{1}{N^2}\right) = \Theta(1/N),\end{aligned}$$

  where $\le$ in the last line is guaranteed by coupling, relying on the linear attachment rule.

References

[1] Hang Du, Shuyang Gong, and Jiaming Xu. A proof of the changepoint detection threshold conjecture in preferential attachment models. In Nika Haghtalab and Ankur Moitra, editors, *Proceedings of Thirty Eighth Conference on Learning Theory*, volume 291 of *Proceedings of Machine Learning Research*, pages 1559–1563. PMLR, 30 Jun–04 Jul 2025.