

# Detecting Cliques in Inhomogeneous Networks

---

Srijan Sengupta  
North Carolina State University

With Subhankar Bhadra (Pennsylvania State University)

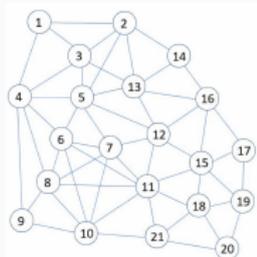
IMSI workshop on Recent Advances in Random Networks, Chicago, Jan 2026

# Introduction

---

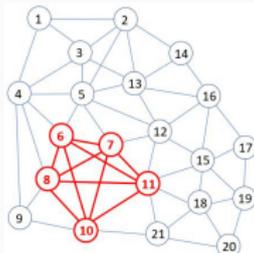
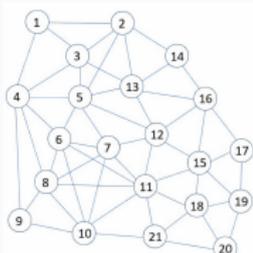
# Problem Statement

- Data:  $A$  is the (symmetric) adjacency matrix of a network with  $n$  nodes, where  $A_{ij} \sim \text{Bernoulli}(P_{ij})$  indep for all  $1 \leq i < j \leq n$
- A **clique** is a *fully connected* subgraph. Can form randomly, e.g.,  $2 \log_{\frac{1}{p}}(n)$ -clique for Erdős-Renýi graphs where  $P_{ij} = p$  for all  $i < j$
- Given a network, we wish to answer the following questions:



# Problem Statement

- Data:  $A$  is the (symmetric) adjacency matrix of a network with  $n$  nodes, where  $A_{ij} \sim \text{Bernoulli}(P_{ij})$  indep for all  $1 \leq i < j \leq n$
- A **clique** is a *fully connected* subgraph. Can form randomly, e.g.,  $2 \log_{\frac{1}{p}}(n)$ -clique for Erdős-Renýi graphs where  $P_{ij} = p$  for all  $i < j$
- Given a network, we wish to answer the following questions:



1. **Detection:** Does the network contain an **anomalous clique**?

$H_0 : A_{ij} \sim \text{Bernoulli}(P_{ij})$  for all  $i < j$ , vs.

$H_1 : \exists S \subset \{1, \dots, n\}$  such that  $A_{ij} \sim \begin{cases} 1 & \text{if } i, j \in S, \\ P_{ij} & \text{otherwise} \end{cases}$

2. **Localization:** If yes, which nodes form the clique?  $\Leftrightarrow$  Estimate  $S$

- Financial fraud rings
- Social and professional cliques (e.g., citations? :)
- Cybersecurity: adversarial attacks and botnet detection
- Protein complexes in protein-protein interaction (PPI) networks

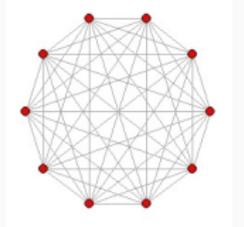
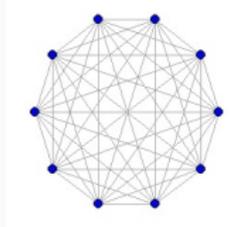
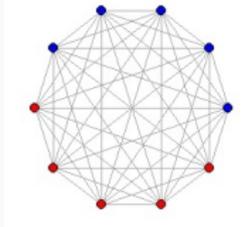
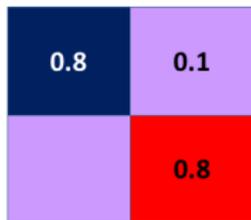
- Homogeneous Erdős-Rényi model:  $P_{ij} = p$  for all  $i < j$
- Very well-studied for 30+ years and a lot is known about algorithmic and statistical limits for detection/ localization.
- The celebrated **planted clique conjecture**: no polynomial-time algorithm can find a clique of size  $k = o(\sqrt{n})$

## The inhomogeneous case is fundamentally different

- Inhomogeneous case:  $P_{ij}$  varies across  $i, j$
- It is not only the **size** of a clique that makes it anomalous
- Whether a clique is anomalous depends also on its **members**

# The inhomogeneous case is fundamentally different

- Inhomogeneous case:  $P_{ij}$  varies across  $i, j$
- It is not only the **size** of a clique that makes it anomalous
- Whether a clique is anomalous depends also on its **members**



- Example: stochastic blockmodel with  $25 \times 2$  nodes.
  - The “balanced” 10-clique (second from left), with five blue and five red nodes, is *anomalous* (probability  $\approx 0\%$ )
  - But, a pure blue or pure red 10-clique is *not anomalous* (probability  $\approx 87\%$ ).
- So we cannot simply find the largest clique and check whether it is anomalous

**Scan statistic:** look at all (or nearly all) subgraphs to see if there is an anomalous clique

- Theoretical bounds: Hajek et al (2015), Verzelen and Arias-Castro (2015), and so on
- Bogerd et al (2021) established theoretical bounds under the Chung-Lu model
- Computationally infeasible because there are  $O(2^n)$  subgraphs to scan

## Proposed methodology

---

# Egonets

The  $i^{th}$  egonet is the **subgraph** spanned by **neighbors** of the  $i^{th}$  node.

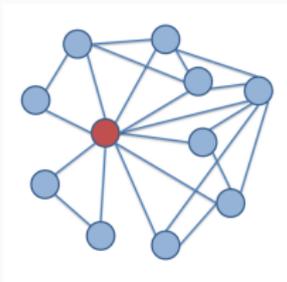
# Egonets

The  $i^{\text{th}}$  egonet is the **subgraph** spanned by **neighbors** of the  $i^{\text{th}}$  node.



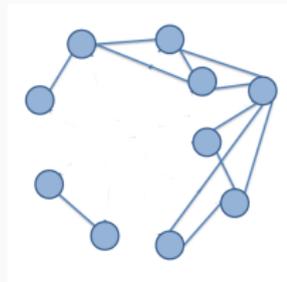
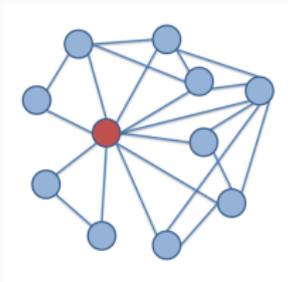
# Egonets

The  $i^{\text{th}}$  egonet is the **subgraph** spanned by **neighbors** of the  $i^{\text{th}}$  node.



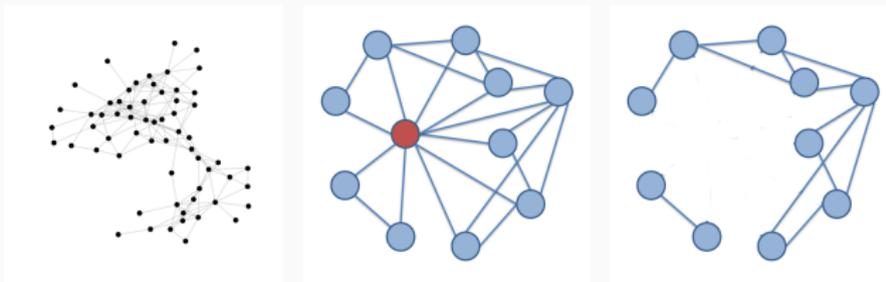
# Egonets

The  $i^{\text{th}}$  egonet is the **subgraph** spanned by **neighbors** of the  $i^{\text{th}}$  node.



# Egonets

The  $i^{\text{th}}$  egonet is the **subgraph** spanned by **neighbors** of the  $i^{\text{th}}$  node.



Two crucial (tautological) points to note:

1. There are exactly  $n$  egonets, computationally feasible!
2. If the network contains a clique (anomalous or not), and the  $i^{\text{th}}$  node is the member of a clique, then the egonet contains the rest of the clique

# Heuristics

$H_0 : A_{ij} \sim \text{Bernoulli}(P_{ij})$  for all  $i < j$ , vs.

$H_1 : \exists S \subset \{1, \dots, n\}$  such that  $A_{ij} \sim \begin{cases} 1 & \text{if } i, j \in S, \\ P_{ij} & \text{otherwise} \end{cases}$

# Heuristics

$H_0 : A_{ij} \sim \text{Bernoulli}(P_{ij})$  for all  $i < j$ , vs.

$H_1 : \exists S \subset \{1, \dots, n\}$  such that  $A_{ij} \sim \begin{cases} 1 & \text{if } i, j \in S, \\ P_{ij} & \text{otherwise} \end{cases}$

**Egonet degree:**  $E_i = \sum_{j < k} A_{ij} A_{ik} A_{jk}$

$$\cdot \mathbb{E}_{H_0}(E_i | A_{i,\cdot}) = \sum_{j < k} \sum A_{ij} A_{ik} P_{jk}$$

# Heuristics

$H_0 : A_{ij} \sim \text{Bernoulli}(P_{ij})$  for all  $i < j$ , vs.

$H_1 : \exists S \subset \{1, \dots, n\}$  such that  $A_{ij} \sim \begin{cases} 1 & \text{if } i, j \in S, \\ P_{ij} & \text{otherwise} \end{cases}$

**Egonet degree:**  $E_i = \sum_{j < k} A_{ij} A_{ik} A_{jk}$

$$\cdot \mathbb{E}_{H_0}(E_i | A_{i,\cdot}) = \sum_{j < k} \sum A_{ij} A_{ik} P_{jk}$$

$$\cdot \mathbb{E}_{H_1}(E_i | A_{i,\cdot}) = \sum_{j < k} \sum A_{ij} A_{ik} P_{jk} + \begin{cases} \sum_{j < k \in S} (1 - P_{jk}), & \text{if } i \in S \\ \sum_{j < k \in S} A_{ij} A_{ik} (1 - P_{jk}) & \text{if } i \notin S \end{cases}$$

# Heuristics

$H_0 : A_{ij} \sim \text{Bernoulli}(P_{ij})$  for all  $i < j$ , vs.

$H_1 : \exists S \subset \{1, \dots, n\}$  such that  $A_{ij} \sim \begin{cases} 1 & \text{if } i, j \in S, \\ P_{ij} & \text{otherwise} \end{cases}$

**Egonet degree:**  $E_i = \sum_{j < k} A_{ij} A_{ik} A_{jk}$

$$\cdot \mathbb{E}_{H_0}(E_i | A_{i,\cdot}) = \sum_{j < k} \sum A_{ij} A_{ik} P_{jk}$$

$$\cdot \mathbb{E}_{H_1}(E_i | A_{i,\cdot}) = \sum_{j < k} \sum A_{ij} A_{ik} P_{jk} + \begin{cases} \sum_{j < k \in S} (1 - P_{jk}), & \text{if } i \in S \\ \sum_{j < k \in S} A_{ij} A_{ik} (1 - P_{jk}) & \text{if } i \notin S \end{cases}$$

**Summary:** Define  $\Delta_i := \mathbb{E}_{H_1}(E_i | A_{i,\cdot}) - \mathbb{E}_{H_0}(E_i | A_{i,\cdot})$ . Then,

1.  $\Delta_i > 0$  for all  $i = 1, \dots, n$
2. However, crucially,  $\Delta_{i,i \in S} > \Delta_{i,i \notin S}$ .

# Detection algorithm

- Consider

$$E_i - \mathbb{E}_{H_0}(E_i | A_{i,\cdot}) = \sum_{j < k} A_{ij} A_{ik} (A_{jk} - P_{jk})$$

- For any  $i$ , the (conditional) expectation of this statistic is zero under  $H_0$  and positive under  $H_1$
- Since  $P$  is unknown (more on this later), we replace it by its estimated version. Egonet statistic:

$$T_i := E_i - \widehat{\mathbb{E}}_{H_0}(E_i | A_{i,\cdot}) = \sum_{j < k} A_{ij} A_{ik} (A_{jk} - \hat{P}_{jk})$$

- Reject if  $T_i > C$  for any  $i = 1, \dots, n$
- $C$  is a function of  $A$

## Localization via EgoRank algorithm

$$\mathbb{E}_{H_1}(T_i | A_{i,\cdot}) \approx \begin{cases} \sum_{j < k \in S} (1 - P_{jk}), & \text{if } i \in S \\ \sum_{j < k \in S} A_{ij} A_{ik} (1 - P_{jk}) & \text{if } i \notin S \end{cases}$$

**Heuristics:** Under  $H_1$ , the largest  $T_i$ 's are associated with  $i \in S$ .

# Localization via EgoRank algorithm

$$\mathbb{E}_{H_1}(T_i | A_{i,\cdot}) \approx \begin{cases} \sum_{j < k \in S} (1 - P_{jk}), & \text{if } i \in S \\ \sum_{j < k \in S} A_{ij} A_{ik} (1 - P_{jk}) & \text{if } i \notin S \end{cases}$$

**Heuristics:** Under  $H_1$ , the largest  $T_i$ 's are associated with  $i \in S$ .

---

**Algorithm 1** Egonet ranking algorithm

---

**Require:** adjacency matrix  $A$ , probability estimate  $\hat{P}$

- 1: for each  $i \in [n]$ , compute egonet statistic  $T_i \leftarrow \sum \sum_{j < k} A_{ij} A_{ik} (A_{jk} - \hat{P}_{jk})$
  - 2: arrange  $T_i$ 's in descending order, let  $T_{(i)}$  be the  $i^{\text{th}}$  ordered egonet statistic and  $v_i$  be the corresponding node index, that is,  $T_{v_i} = T_{(i)}$
  - 3: set  $\hat{S} \leftarrow \{v_1\}$ ,  $i \leftarrow 1$
  - 4: **while**  $v_{i+1}$  is connected to all nodes in  $S$  **do**
  - 5:      $S \leftarrow \hat{S} \cup \{v_{i+1}\}$
  - 6:      $i \leftarrow i + 1$
  - 7: **end while**
  - 8: **return**  $\hat{S}$
-

Recall that the egonet statistic is defined as

$$T_i = \sum_{j < k} A_{ij} A_{ik} (A_{jk} - \hat{P}_{jk})$$

So we need to obtain  $\hat{P}$ .

**Strategy:** assume that  $P$  belongs to some **known class** of models (e.g., Chung-Lu, Stochastic Blockmodel, Random Dot Product Graph model, etc.) and use a  $\hat{P}$  which is known to be consistent for that class.

## Theoretical results

---

For some  $P$  in the chosen class of models,

$H_0 : A_{ij} \sim \text{Bernoulli}(P_{ij})$  for all  $i < j$ , vs.

$H_1 : \exists S \subset \{1, \dots, n\}$  such that  $A_{ij} \sim \begin{cases} 1 & \text{if } i, j \in S, \\ P_{ij} & \text{otherwise} \end{cases}$

- **Consistency for detection:** We want to show that
  - Under  $H_0$ ,  $\text{Prob}[\text{at least one egonet test is rejected}] \rightarrow 0$
  - Under  $H_1$ ,  $\text{Prob}[\text{at least one egonet test is rejected}] \rightarrow 1$
- **Strong consistency for localization:** We want to show that under the alternative model,  $\text{Prob}[\hat{S} = S] \rightarrow 1$ .

Under  $H_0$ ,

$$\begin{aligned} T_i &= \sum_{j < k} A_{ij} A_{ik} (A_{jk} - \hat{P}_{jk}) \\ &= \underbrace{\sum_{j < k} A_{ij} A_{ik} (A_{jk} - P_{jk})}_{\text{random deviation}} + \underbrace{\sum_{j < k} A_{ij} A_{ik} (P_{jk} - \hat{P}_{jk})}_{\text{estimation noise}} \end{aligned}$$

To show that  $P[\cup_i \{T_i > C\}] \rightarrow 0$ .

Under  $H_1$ :  $A \sim Q$  where  $Q := P + B$  such that

$$B_{ij} = \begin{cases} 1 - P_{ij} & \text{if } i, j \in S \text{ and } i < j, \\ 0 & \text{if } i \notin S \text{ or } j \notin S, \text{ and } i < j. \end{cases}$$

# Proof strategy

Under  $H_1$ :  $A \sim Q$  where  $Q := P + B$  such that

$$B_{ij} = \begin{cases} 1 - P_{ij} & \text{if } i, j \in S \text{ and } i < j, \\ 0 & \text{if } i \notin S \text{ or } j \notin S, \text{ and } i < j. \end{cases}$$

$$T_i = \sum_{j < k} A_{ij} A_{ik} \left[ \underbrace{(A_{jk} - Q_{jk})}_{\text{random deviation}} + \underbrace{(\tilde{P}_{jk} - \hat{P}_{jk})}_{\text{estimation "bias"}} + \underbrace{(Q_{jk} - \tilde{P}_{jk})}_{\text{signal}} \right]$$

# Proof strategy

Under  $H_1$ :  $A \sim Q$  where  $Q := P + B$  such that

$$B_{ij} = \begin{cases} 1 - P_{ij} & \text{if } i, j \in S \text{ and } i < j, \\ 0 & \text{if } i \notin S \text{ or } j \notin S, \text{ and } i < j. \end{cases}$$

$$T_i = \sum_{j < k} A_{ij} A_{ik} \left[ \underbrace{(A_{jk} - Q_{jk})}_{\text{random deviation}} + \underbrace{(\tilde{P}_{jk} - \hat{P}_{jk})}_{\text{estimation "bias"}} + \underbrace{(Q_{jk} - \tilde{P}_{jk})}_{\text{signal}} \right]$$

To show that

- $P[\cup_i \{T_i > C\}] \rightarrow 1$
- With probability going to one,  $\min_{i \in S} T_i > \max_{i \in S^c} T_i$

# Proof strategy

Under  $H_1$ ,  $\hat{P}$  is the result of applying the estimator on a network generated from  $Q = P + B$ , and  $Q \notin \mathcal{P}$ . How does  $\hat{P}$  behave?

- If  $B$  is too small cannot be distinguished from the null case
- If  $B$  is too large, it dominates  $P$  so that  $\hat{P}$  is close to  $Q$ , not  $P$
- The middle ground is achieved when  $B$  is large enough but not too large. In this case, there is some  $\tilde{P}$  such that  $\|\hat{P} - \tilde{P}\|$  is small and  $\|Q - \tilde{P}\|$  is large

# Proof strategy

Under  $H_1$ ,  $\hat{P}$  is the result of applying the estimator on a network generated from  $Q = P + B$ , and  $Q \notin \mathcal{P}$ . How does  $\hat{P}$  behave?

- If  $B$  is too small cannot be distinguished from the null case
- If  $B$  is too large, it dominates  $P$  so that  $\hat{P}$  is close to  $Q$ , not  $P$
- The middle ground is achieved when  $B$  is large enough but not too large. In this case, there is some  $\tilde{P}$  such that  $\|\hat{P} - \tilde{P}\|$  is small and  $\|Q - \tilde{P}\|$  is large

$$T_i = \sum_{j < k} A_{ij} A_{ik} \left[ \underbrace{(A_{jk} - Q_{jk})}_{\text{random deviation}} + \underbrace{(\tilde{P}_{jk} - \hat{P}_{jk})}_{\text{estimation "bias"}} + \underbrace{(Q_{jk} - \tilde{P}_{jk})}_{\text{signal}} \right]$$

When the signal is larger than the estimation bias, we achieve detection power and strong consistency for localization.

# Summary of results

The egonet method achieves detection and localization consistency under the following conditions for each model class:

Model	Known $P$	Erdos Renyi	Chung-Lu	RDPG- $d$
	-	$P_{jk} = \rho$	$P_{jk} = \theta_j \theta_k$	$P_{jk} = \mathbf{x}_j^T \mathbf{x}_k$
$\hat{P}_{jk}$	-	$\frac{\sum_{i, k \neq i} (1 - A_{ij} A_{ik}) A_{jk}}{\binom{n-1}{2} - \binom{D_j}{2}}$	$\frac{D_j D_k}{2M}$	$\hat{\mathbf{x}}_j^T \hat{\mathbf{x}}_k$ (from ASE)
$C$	$n \rho_n (\log(n))^{\frac{1}{2} + \epsilon}$	$np (\log n)^{\frac{1}{2}}$	$n (\rho_n \log n)^{\frac{1}{2}}$	$\delta(P) \log n$
Cond. 1	$b^{(1)} \gg n \rho_n \sqrt{\log n}$	$b^{(1)} \gg np (\log n)^{\frac{1}{2}}$	$b^{(1)} \gg n (\rho_n \log n)^{\frac{1}{2}}$	$\ B\ ^2 \gg \delta(P) \log n$
Cond. 2	NONE	$b^{(1)} = o(n^2)$	$\max_{j \in S} \sum_{l \in S, l \neq j} (1 - \theta_j \theta_l) \ll \sum_l \theta_l,$ $\sum_{j \neq k; j, k \in S} \sum (1 - \theta_j \theta_k) \ll \left( \sum_l \theta_l \right)^2$	$\ B\  \ll \delta(P)$

$$b^{(1)} = \sum_{i,j} B_{ij} \text{ where } B_{ij} = \begin{cases} 1 - P_{ij} & \text{if } i, j \in S \text{ and } i < j, \\ 0 & \text{if } i \notin S \text{ or } j \notin S, \text{ and } i < j, \end{cases}$$

and  $\delta(P) = \max_i \sum_j P_{ij}$  is the maximum expected degree.

# Numerical Results

---

# Applications

- British MP network: 360 nodes with 2 communities
- Political blogs network: 1222 nodes with 2 communities
- Existing SOTA method (Deshpande and Montanari, 2015): anomalous clique **detected** for both (homogeneous ER model)
- Egonet method: anomalous clique **not detected** in either network (with RDPG-4 for British MP and DCBM-2 for political blogs)

# Applications

- British MP network: 360 nodes with 2 communities
- Political blogs network: 1222 nodes with 2 communities
- Existing SOTA method (Deshpande and Montanari, 2015): anomalous clique **detected** for both (homogeneous ER model)
- Egonet method: anomalous clique **not detected** in either network (with RDPG-4 for British MP and DCBM-2 for political blogs)
- *Which inference makes more sense?*

	Largest clique	ER: $\sum_{i,j \in \text{clique}} \hat{P}_{ij}$	Inhom: $\sum_{i,j \in \text{clique}} \hat{P}_{ij}$
British MP	$\binom{19}{2} = 171$	18.08	158.11
Political blogs	$\binom{20}{2} = 190$	4.25	157.33

The ER model used in DM15 **does not account for heterogeneity** of the nodes forming the clique, it is purely a function of overall network sparsity.

But both networks have community structure and degree heterogeneity, which the egonet method accounts for.

## What if we “planted” a clique?

We planted cliques of sizes 40, 50, and 60 on the polblogs network such that the clique nodes are equally distributed in the two communities (like the balanced case in the toy SBM).

$k$	Power	$ S \cap \hat{S} / S $	$ S^c \cap \hat{S} / S^c $
40	83	98.32	0.00
50	100	96.86	0.00
60	100	84.68	0.00

**Table 1:** Results for the political blogs network with a planted clique averaged over 100 replications.

## Chung-Lu model: null model

We generated 1000 networks with  $n = 500, 1000$  nodes, and density parameter,  $\delta = 0.05, 0.1$  and  $0.2$ . The  $\{\theta_i\}$  are sampled from the  $\text{Uniform}(0, \rho_n)$ , so that  $\rho_n$  is the maximum edge probability and  $\delta = \rho_n/4$  is the expected network density.

	$n = 500$	$n = 1000$
$\delta$	Type-1 error	Type-1 error
0.05	0.00	0.00
0.1	0.00	0.00
0.2	0.00	0.00

## Chung-Lu model: alternative model

Under  $H_1$ , recall that the condition is:  $b^{(1)} \gg n(\rho_n \log n)^{\frac{1}{2}}$

We planted a clique of  $k = c\sqrt{2n\rho_n\sqrt{\log n}}$  nodes, where the nodes were selected in three ways: (i) **random uniform**, (ii) the  $k$  largest  $\theta_i$ , (iii) the  $k$  smallest  $\theta_i$

$\delta$	$c$	$n = 500$				$n = 1000$			
		$k$	Power	$ S \cap \hat{S} / S $	$ S^c \cap \hat{S} / S^c $	$k$	Power	$ S \cap \hat{S} / S $	$ S^c \cap \hat{S} / S^c $
0.05	1	22	0.00	100.00	0.00	32	0.00	100.00	0.00
0.05	2	44	2.50	100.00	0.00	64	99.20	100.00	0.00
0.05	4	89	100.00	100.00	0.00	129	100.00	100.00	0.00
0.1	1	31	0.00	99.96	0.00	45	0.00	100.00	0.00
0.1	2	63	92.00	100.00	0.00	91	100.00	100.00	0.00
0.1	4	126	100.00	100.00	0.00	183	100.00	100.00	0.00
0.2	1	44	0.00	42.72	1.10	64	0.00	43.16	0.90
0.2	2	89	100.00	97.55	0.00	129	100.00	100.00	0.00
0.2	4	178	100.00	100.00	0.00	259	100.00	100.00	0.00

# Chung-Lu model: alternative model

Under  $H_1$ , recall that the condition is:  $b^{(1)} \gg n(\rho_n \log n)^{\frac{1}{2}}$

We planted a clique of size where  $k = c\sqrt{2n\rho_n\sqrt{\log n}}$ , in three ways:

(i) random uniform, (ii) **the  $k$  largest  $\theta_i$** , (iii) the  $k$  smallest  $\theta_i$

$\delta$	$c$	$n = 500$				$n = 1000$			
		$k$	Power	$ S \cap \hat{S} / S $	$ S^c \cap \hat{S} / S^c $	$k$	Power	$ S \cap \hat{S} / S $	$ S^c \cap \hat{S} / S^c $
0.05	1	22	0.00	99.95	0.00	32	0.00	100.00	0.00
0.05	2	44	0.00	100.00	0.00	64	0.00	100.00	0.00
0.05	4	89	0.00	100.00	0.00	129	100.00	100.00	0.00
0.1	1	31	0.00	69.14	10.69	45	0.00	85.18	4.20
0.1	2	63	0.00	100.00	0.00	91	0.00	100.00	0.00
0.1	4	126	0.00	100.00	0.00	183	100.00	100.00	0.00
0.2	1	44	0.00	43.78	41.56	64	0.00	42.59	41.04
0.2	2	79	0.00	74.55	12.42	129	0.00	67.81	14.61
0.2	4	178	0.00	100.00	0.00	259	0.00	100.00	0.00

## Chung-Lu model: alternative model

Under  $H_1$ , recall that the condition is:  $b^{(1)} \gg n(\rho_n \log n)^{\frac{1}{2}}$

We planted a clique of  $k = c\sqrt{2n\rho_n\sqrt{\log n}}$  nodes, where the nodes were selected in three ways: (i) random uniform, (ii) the  $k$  largest  $\theta_i$ , (iii) **the  $k$  smallest  $\theta_i$**

$\delta$	$c$	$n = 500$				$n = 1000$			
		$k$	Power	$ S \cap \hat{S} / S $	$ S^c \cap \hat{S} / S^c $	$k$	Power	$ S \cap \hat{S} / S $	$ S^c \cap \hat{S} / S^c $
0.05	1	22	0.00	100.00	0.00	32	0.00	100.00	0.00
0.05	2	44	100.00	100.00	0.00	64	100.00	100.00	0.00
0.05	4	89	100.00	100.00	0.00	129	100.00	100.00	0.00
0.1	1	31	0.00	100.00	0.00	45	0.00	100.00	0.00
0.1	2	63	100.00	100.00	0.00	91	100.00	100.00	0.00
0.1	4	126	100.00	100.00	0.00	183	100.00	100.00	0.00
0.2	1	44	0.00	100.00	0.00	64	0.00	100.00	0.00
0.2	2	89	100.00	100.00	0.00	129	100.00	100.00	0.00
0.2	4	178	100.00	100.00	0.00	259	100.00	100.00	0.00

## RDPG model

We generated 100 networks with  $n = 2000$ ,  $\{X_i\} \sim \text{Dirichlet}(1, 1, 1)^T$ , and scaled  $P$  so that the average network density is  $\delta = 0.2$ .

We randomly selected  $k$  nodes to plant the clique. We reject when  $T_i > \delta(P) \log n$ .

$k$	$\ B\ $	$\lambda_d(P)$	$\ B\ /\lambda_d(P)$	Power	$ S \cap \hat{S} / S $	$ S^c \cap \hat{S} / S^c $
60	47.21	97.68	0.48	0.00	97.90	0.00
70	55.2	98.37	0.56	9.00	100.00	0.00
80	63.21	97.31	0.65	100.00	100.00	0.00
90	71.26	96.65	0.74	100.00	100.00	0.00
100	79.2	96.66	0.82	100.00	100.00	0.00
110	87.19	97.27	0.9	100.00	100.00	0.00
120	95.18	96.94	0.98	100.00	98.01	0.00
130	103.19	96.92	1.07	100.00	53.35	0.00
140	111.2	97.35	1.14	97.00	2.86	1.04
150	119.21	97.8	1.22	95.00	2.53	2.09

# Conclusion

- To the best of our knowledge, this is the first method to achieve consistent clique detection and localization for inhomogeneous networks in a computationally feasible manner
- **Opinion:** This is one example of many problems in network science that have not been extended (in a statistical sense) beyond ER models, leading to an apparent **ubiquity** (e.g., cliques, small-world property, core-periphery...)
- Next steps: the upper bound on  $B$  (condition C2) is somewhat unsatisfactory. Is there a way to circumvent this?



Scan for arXiv preprint