Testing Correlation of Unlabeled Random Graphs

Jiaming Xu

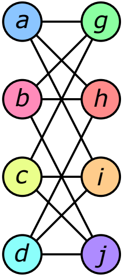
The Fuqua School of Business
Duke University

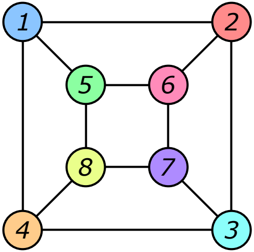
Joint work with
Yihong Wu (Yale) and Sophie H. Yu (Duke)

Computational and Methodological Statistics
December 21, 2020

Given two graphs A and B, decide whether $A \cong B$, i.e., there exists a bijection $\pi : V(A) \to V(B)$ such that

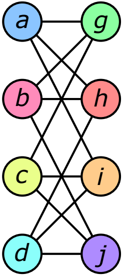
$$(u, v) \in E(A) \Leftrightarrow (\pi(u), \pi(v)) \in E(B)$$

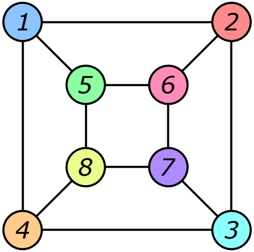




Given two graphs A and B, decide whether $A \cong B$, i.e., there exists a bijection $\pi : V(A) \to V(B)$ such that

$$(u, v) \in E(A) \Leftrightarrow (\pi(u), \pi(v)) \in E(B)$$





$\pi(a) = 1$

$\pi(b) = 6$

$\pi(c) = 8$

$\pi(d) = 3$

$\pi(g) = 5$

$\pi(h) = 2$

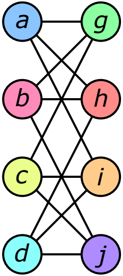
$\pi(i) = 4$

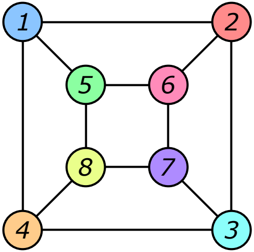
$\pi(j) = 7$

Graph isomorphism

Given two graphs A and B, decide whether $A \cong B$, i.e., there exists a bijection $\pi : V(A) \to V(B)$ such that

$$(u, v) \in E(A) \Leftrightarrow (\pi(u), \pi(v)) \in E(B)$$



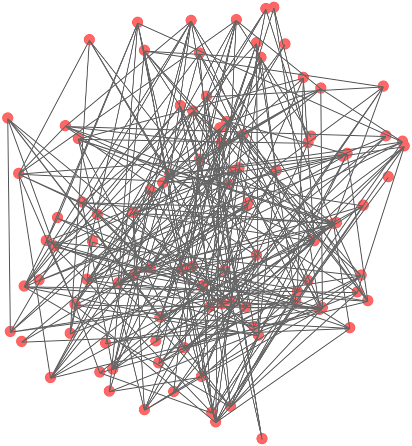


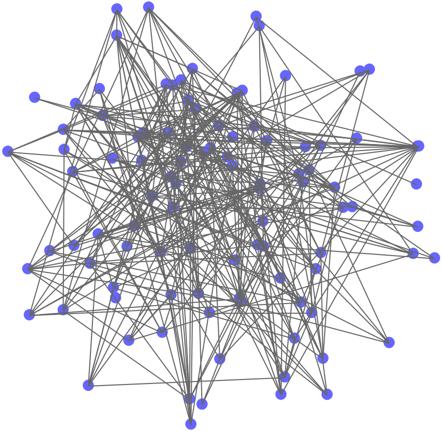
$$\pi(a) = 1$$
$$\pi(b) = 6$$
$$\pi(c) = 8$$
$$\pi(d) = 3$$
$$\pi(g) = 5$$
$$\pi(h) = 2$$
$$\pi(i) = 4$$
$$\pi(j) = 7$$

- Not known to be solvable in polynomial time in worst case
- In practice, two graphs are often not exactly isomorphic, but still want to tell whether their topologies are similar

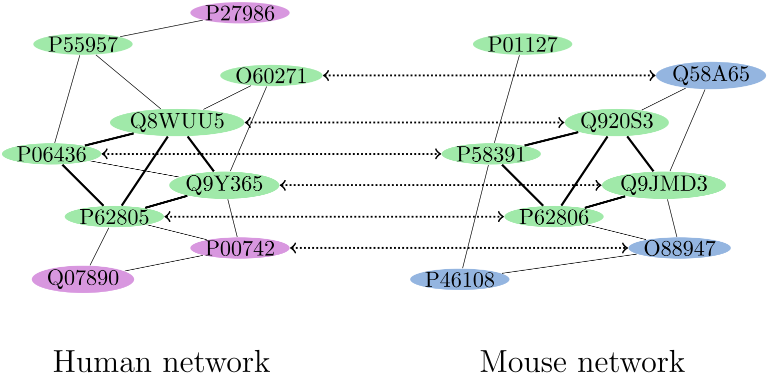
Two key challenges

- Statistical: two graphs may be correlated but not exactly isomorphic
- Computational: # of possible node mappings is $n!$ ($100! \approx 10^{158}$)





Motivation: Protein-protein interaction network

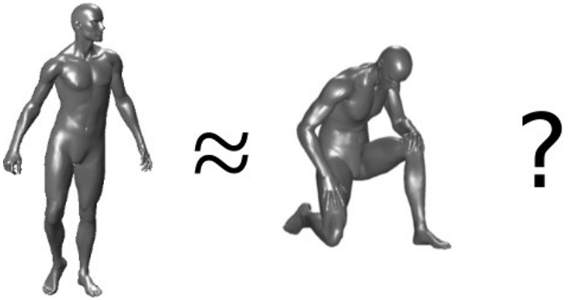


[Kazemi-Hassani-Grossglauser-Modarres '16]

Ontology: Assess the correlation of two biological networks in two different species based on network topology

Motivation: Deformable shape matching

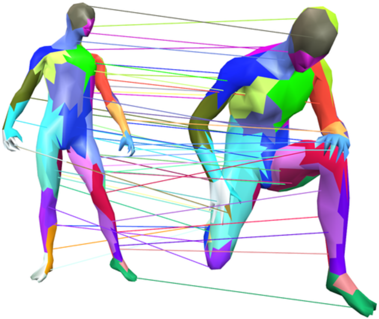
A fundamental problem in computer vision: Detect similar objects that undergo different deformations



Shape REtrieval Contest (SHREC) dataset

Motivation: Deformable shape matching

A fundamental problem in computer vision: Detect similar objects that undergo different deformations



Shape REtrieval Contest (SHREC) dataset

3-D shapes → geometric graphs (features → nodes, distances → edges)
Determine whether two graphs are similar in topologies

Beyond worst-case: Noisy graph isomorphism

Definition (Hypothesis testing)

Suppose we observe two random graphs A and B:

$$\mathcal{H}_0 : A \text{ and } B \text{ are independent}$$
$$\mathcal{H}_1 : A \text{ and } B^\pi = (B_{\pi(i)\pi(j)}) \text{ are edge-correlated}$$
$$\text{conditional on a uniform permutation } \pi$$

Goal: Test $\mathcal{H}_0$ versus $\mathcal{H}_1$

- Under $\mathcal{H}_1$, the inherent edge correlation is obscured by the latent node correspondence π
- Hypothesis testing aspect of graph matching (recover π under $\mathcal{H}_1$)
- The test needs to rely on graph invariants (invariant under graph isomorphism), such as
  - Subgraph counts (e.g. # of edges or triangles)
  - Spectrum (e.g. eigenvalues of adjacency matrices or Laplacians)

Gaussian setting

A and B are two Gaussian-weighted complete graphs

Definition (Gaussian Wigner model)

$$\mathcal{H}_0 : (A_{ij}, B_{ij}) \overset{\text{i.i.d.}}{\sim} \mathcal{N}\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \right)$$

$$\mathcal{H}_1 : \left(A_{ij}, B_{\pi(i)\pi(j)}\right) \overset{\text{i.i.d.}}{\sim} \mathcal{N}\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix} \right) \text{ conditional on uniform } \pi$$

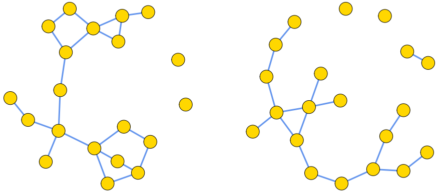
- Under both $\mathcal{H}_0$ and $\mathcal{H}_1$, A and B are Gaussian Wigner matrices marginally
- Under $\mathcal{H}_1$, ρ characterizes the correlation of edge weights

Erdős-Rényi setting

Definition (Erdős-Rényi graphs model [Barak-Chou-Lei-Schramm-Sheng'19])

$\mathcal{H}_0$: A and B are independent $\mathcal{G}(n, ps)$

$\mathcal{H}_1$: A and $B^\pi = (B_{\pi(i)\pi(j)})$ are independently edge-sampled from a common parent graph $\mathcal{G}(n, p)$ with sampling probability s



- Under both $\mathcal{H}_0$ and $\mathcal{H}_1$, A and B are $\mathcal{G}(n, ps)$ marginally
- Under $\mathcal{H}_1$, $\left(A_{ij}, B_{\pi(i)\pi(j)}\right)$ are correlated $\mathrm{Bern}(ps)$ with correlation coefficient $\rho \triangleq \frac{s(1-p)}{1-ps}$

Sharp threshold for detection: Gaussian

$\mathcal{Q}$ and $\mathcal{P}$: probability measure under $\mathcal{H}_0$ and $\mathcal{H}_1$

Theorem (Wu-X.-Yu '20)

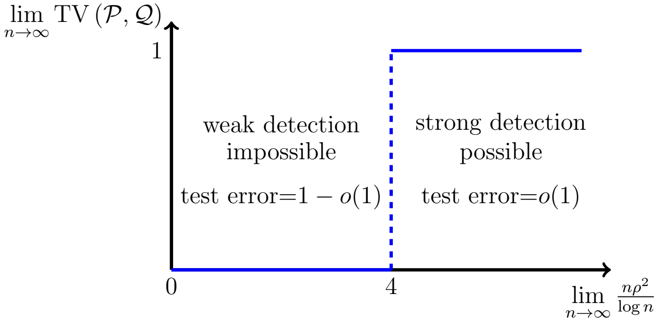
$$\mathrm{TV}(\mathcal{P}, \mathcal{Q}) = \begin{cases} 1 - o(1) & \textit{if } \rho^2 \geq 4\,\frac{\log n}{n-1} \\ o(1) & \textit{if } \rho^2 \leq (4-\epsilon)\,\frac{\log n}{n} \end{cases}$$

Sharp threshold for detection: Gaussian

$\mathcal{Q}$ and $\mathcal{P}$: probability measure under $\mathcal{H}_0$ and $\mathcal{H}_1$

Theorem (Wu-X.-Yu '20)

$$\mathrm{TV}(\mathcal{P}, \mathcal{Q}) = \begin{cases} 1 - o(1) & \textit{if } \rho^2 \geq 4\,\frac{\log n}{n-1} \\ o(1) & \textit{if } \rho^2 \leq (4-\epsilon)\,\frac{\log n}{n} \end{cases}$$



Sharp threshold for detection: dense Erdős-Rényi graphs

Theorem (Wu-X.-Yu '20)

$$\textit{If } s^2 \geq \frac{2\log n}{(n-1)p\left(\log\frac{1}{p} - 1 + p\right)}, \textit{ then } \mathrm{TV}\left(\mathcal{P}, \mathcal{Q}\right) = 1 - o\left(1\right)$$

- *(Dense regime)* $p = n^{-o(1)}$:

$$\textit{If } s^2 \leq \frac{(2-\epsilon)\log n}{np\left(\log\frac{1}{p} - 1 + p\right)}, \textit{ then } \mathrm{TV}\left(\mathcal{P}, \mathcal{Q}\right) = o(1)$$

Sharp threshold for detection: dense Erdős-Rényi graphs

Theorem (Wu-X.-Yu '20)

If $s^2 \geq \frac{2\log n}{(n-1)p\left(\log\frac{1}{p} - 1 + p\right)}$, *then* $\mathrm{TV}(\mathcal{P}, \mathcal{Q}) = 1 - o(1)$

- *(Dense regime)* $p = n^{-o(1)}$*:*

$$\textit{If } s^2 \leq \frac{(2-\epsilon)\log n}{np\left(\log\frac{1}{p} - 1 + p\right)}, \textit{ then } \mathrm{TV}(\mathcal{P}, \mathcal{Q}) = o(1)$$

Remark:

- Sharp phase transition at $\frac{ns^2p\left(\log\frac{1}{p}-1+p\right)}{\log n} = 2$
- $p \mapsto p(\log\frac{1}{p} - 1 + p)$ is *not* monotone and uniquely maximized at $p_* \approx 0.203 \Rightarrow$ edge density p_* is the "easiest" for detection

Threshold for detection: sparse Erdős-Rényi graphs

Theorem (Wu-X.-Yu '20)

If $s^2 \geq \frac{2\log n}{(n-1)p\left(\log\frac{1}{p} - 1 + p\right)}$, *then* $\mathrm{TV}(\mathcal{P}, \mathcal{Q}) = 1 - o(1)$

- *(Sparse regime)* $p = n^{-\Omega(1)}$*:*

$$\textit{If } s^2 < \frac{1}{np} \wedge 0.01, \ \textit{then } \mathrm{TV}(\mathcal{P}, \mathcal{Q}) = 1 - \Omega(1)$$

$$\textit{If } s^2 < \frac{1}{np} \wedge o(1), \ \ \textit{then } \mathrm{TV}(\mathcal{P}, \mathcal{Q}) = o(1)$$

Threshold for detection: sparse Erdős-Rényi graphs

Theorem (Wu-X.-Yu '20)

$$\textit{If } s^2 \geq \frac{2\log n}{(n-1)p\left(\log\frac{1}{p} - 1 + p\right)}, \textit{ then } \mathrm{TV}(\mathcal{P}, \mathcal{Q}) = 1 - o(1)$$

- *(Sparse regime)* $p = n^{-\Omega(1)}$:

$$\textit{If } s^2 < \frac{1}{np} \wedge 0.01, \textit{ then } \mathrm{TV}(\mathcal{P}, \mathcal{Q}) = 1 - \Omega(1)$$

$$\textit{If } s^2 < \frac{1}{np} \wedge o(1), \quad \textit{then } \mathrm{TV}(\mathcal{P}, \mathcal{Q}) = o(1)$$

$p = n^{-\alpha}$ for a constant α:
strong detection is possible if $s^2 > \frac{2}{np\alpha}$ and impossible if $s^2 < \frac{1}{np} \wedge 0.01$

What is known algorithmically

- Counting edges: achieve weak detection in linear time, if $s = \Omega(1)$

What is known algorithmically

- Counting edges: achieve weak detection in linear time, if $s = \Omega(1)$
- Counting "balanced" subgraphs [Barak-Chou-Lei-Schramm-Sheng'19]: correctly tell $\mathcal{H}_0$ vs. $\mathcal{H}_1$ w.p. at least 0.9 in poly-time, if

$$s = \Omega\left(1\right) \quad \text{and} \quad nps \in \left[n^{\epsilon}, n^{1/153}\right] \cup \left[n^{2/3}, n^{1-\epsilon}\right]$$

What is known algorithmically

- Counting edges: achieve weak detection in linear time, if $s = \Omega(1)$
- Counting "balanced" subgraphs [Barak-Chou-Lei-Schramm-Sheng'19]: correctly tell $\mathcal{H}_0$ vs. $\mathcal{H}_1$ w.p. at least 0.9 in poly-time, if

$$s = \Omega\left(1\right) \quad \text{and} \quad nps \in \left[n^{\epsilon}, n^{1/153}\right] \cup \left[n^{2/3}, n^{1-\epsilon}\right]$$

- Counting (weighted) trees – low-degree poly. approx. of $\frac{\mathcal{P}(A,B)}{\mathcal{Q}(A,B)}$ [Mao-Wu-X.-Yu 'forthcoming]: achieve strong detection in poly-time, if

$$s^2 > 1/2.956 \quad \text{and} \quad np \geq n^{-o(1)}$$

Order-optimal when $np = \Theta(1)$

What is known algorithmically

- Counting edges: achieve weak detection in linear time, if $s = \Omega(1)$
- Counting "balanced" subgraphs [Barak-Chou-Lei-Schramm-Sheng'19]: correctly tell $\mathcal{H}_0$ vs. $\mathcal{H}_1$ w.p. at least 0.9 in poly-time, if

$$s = \Omega(1) \quad \text{and} \quad nps \in \left[n^{\epsilon}, n^{1/153}\right] \cup \left[n^{2/3}, n^{1-\epsilon}\right]$$

- Counting (weighted) trees – low-degree poly. approx. of $\frac{\mathcal{P}(A,B)}{\mathcal{Q}(A,B)}$ [Mao-Wu-X.-Yu 'forthcoming]: achieve strong detection in poly-time, if

$$s^2 > 1/2.956 \quad \text{and} \quad np \geq n^{-o(1)}$$

Order-optimal when $np = \Theta(1)$

Polynomial-time test for $s = o(1)$ is open

Proof of positive results: QAP statistic

$$\mathcal{T}(A,B) \triangleq \max_{\pi \in \mathcal{S}_n} \sum_{i<j} A_{ij} B_{\pi(i)\pi(j)} \quad \text{(edge correlation)}$$

- This is known as Quadratic Assignment Problem
- Invariant to the node relabeling of both A and B
- Proof: First-moment calculation

Proof of negative results: second-moment method

$$\mathbb{E}_{\mathcal{Q}}\left[\left(\frac{\mathcal{P}(A,B)}{\mathcal{Q}(A,B)}\right)^2\right] = O(1) \qquad \Longrightarrow \mathrm{TV}(\mathcal{P},\mathcal{Q}) \le 1-\Omega(1)$$

Strong detection is impossible

$$\mathbb{E}_{\mathcal{Q}}\left[\left(\frac{\mathcal{P}(A,B)}{\mathcal{Q}(A,B)}\right)^2\right] = 1+o(1) \qquad \Longrightarrow \mathrm{TV}(\mathcal{P},\mathcal{Q}) = o(1)$$

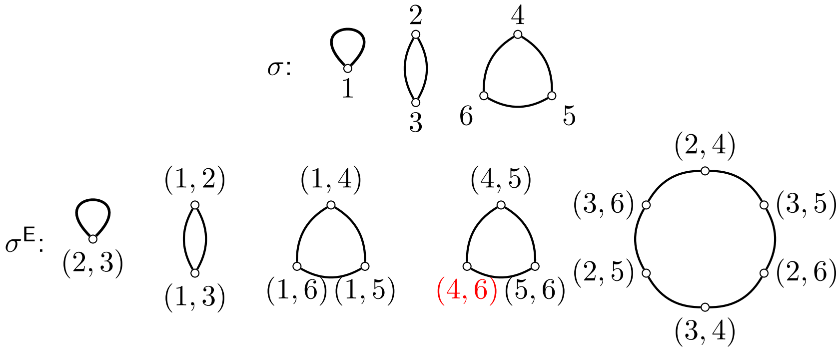
Weak detection is impossible

Second-moment calculation: cycle (orbit) decomposition

- Node permutation σ on $[n]$
- Edge permutation σ^{E} on $\binom{[n]}{2}$: $\sigma^{\mathsf{E}}((i,j)) \triangleq (\sigma(i), \sigma(j))$

- Node permutation σ on $[n]$
- Edge permutation σ^{E} on $\binom{[n]}{2}$: $\sigma^{\mathsf{E}}((i,j)) \triangleq (\sigma(i), \sigma(j))$

E.g. $n = 6$ and $\sigma = (1)(23)(456)$:



Cycles in node (edge) permutation are called node (edge) orbits

Second-moment calculation: cycle decomposition

$$\left(\frac{\mathcal{P}(A,B)}{\mathcal{Q}(A,B)}\right)^2 = \left(\mathbb{E}_\pi\left[\frac{\mathcal{P}(A,B|\pi)}{\mathcal{Q}(A,B)}\right]\right)^2$$

$$= \mathbb{E}_{\widetilde{\pi} \perp\!\!\!\perp \pi} \prod_{i<j} X_{ij} \quad X_{ij} \triangleq \frac{\mathcal{P}(B_{\pi(i)\pi(j)}|A_{ij})\mathcal{P}(B_{\widetilde{\pi}(i)\widetilde{\pi}(j)}|A_{ij})}{\mathcal{Q}(B_{\pi(i)\pi(j)})\mathcal{Q}(B_{\widetilde{\pi}(i)\widetilde{\pi}(j)})}$$

$$= \mathbb{E}_{\widetilde{\pi} \perp\!\!\!\perp \pi} \prod_{O \in \mathcal{O}} X_O \quad X_O \triangleq \prod_{(i,j)\in O} X_{ij}$$

$\mathcal{O}$: disjoint orbits of edge permutation σ^{E} with $\sigma \triangleq \pi^{-1} \circ \widetilde{\pi}$

Second-moment calculation: cycle decomposition

$$\left(\frac{\mathcal{P}(A,B)}{\mathcal{Q}(A,B)}\right)^2 = \left(\mathbb{E}_\pi\left[\frac{\mathcal{P}(A,B|\pi)}{\mathcal{Q}(A,B)}\right]\right)^2$$

$$= \mathbb{E}_{\widetilde{\pi} \perp\!\!\!\perp \pi} \prod_{i<j} X_{ij} \quad X_{ij} \triangleq \frac{\mathcal{P}(B_{\pi(i)\pi(j)}|A_{ij})\mathcal{P}(B_{\widetilde{\pi}(i)\widetilde{\pi}(j)}|A_{ij})}{\mathcal{Q}(B_{\pi(i)\pi(j)})\mathcal{Q}(B_{\widetilde{\pi}(i)\widetilde{\pi}(j)})}$$

$$= \mathbb{E}_{\widetilde{\pi} \perp\!\!\!\perp \pi} \prod_{O \in \mathcal{O}} X_O \quad X_O \triangleq \prod_{(i,j)\in O} X_{ij}$$

$\mathcal{O}$: disjoint orbits of edge permutation σ^{E} with $\sigma \triangleq \pi^{-1} \circ \widetilde{\pi}$

$$\mathbb{E}_{\mathcal{Q}}\left[\left(\frac{\mathcal{P}(A,B)}{\mathcal{Q}(A,B)}\right)^2\right] = \mathbb{E}_{\widetilde{\pi} \perp\!\!\!\perp \pi}\mathbb{E}_{\mathcal{Q}} \prod_{O\in\mathcal{O}} X_O = \mathbb{E}_{\widetilde{\pi} \perp\!\!\!\perp \pi} \prod_{O\in\mathcal{O}} \mathbb{E}_{\mathcal{Q}}\left[X_O\right]$$

Second-moment calculation: cycle decomposition

$$\left(\frac{\mathcal{P}(A,B)}{\mathcal{Q}(A,B)}\right)^2 = \left(\mathbb{E}_\pi\left[\frac{\mathcal{P}(A,B|\pi)}{\mathcal{Q}(A,B)}\right]\right)^2$$

$$= \mathbb{E}_{\widetilde{\pi} \perp\!\!\!\perp \pi} \prod_{i<j} X_{ij} \quad X_{ij} \triangleq \frac{\mathcal{P}(B_{\pi(i)\pi(j)}|A_{ij})\mathcal{P}(B_{\widetilde{\pi}(i)\widetilde{\pi}(j)}|A_{ij})}{\mathcal{Q}(B_{\pi(i)\pi(j)})\mathcal{Q}(B_{\widetilde{\pi}(i)\widetilde{\pi}(j)})}$$

$$= \mathbb{E}_{\widetilde{\pi} \perp\!\!\!\perp \pi} \prod_{O \in \mathcal{O}} X_O \quad X_O \triangleq \prod_{(i,j)\in O} X_{ij}$$

$\mathcal{O}$: disjoint orbits of edge permutation σ^{E} with $\sigma \triangleq \pi^{-1} \circ \widetilde{\pi}$

$$\mathbb{E}_{\mathcal{Q}}\left[\left(\frac{\mathcal{P}(A,B)}{\mathcal{Q}(A,B)}\right)^2\right] = \mathbb{E}_{\widetilde{\pi} \perp\!\!\!\perp \pi}\mathbb{E}_{\mathcal{Q}} \prod_{O \in \mathcal{O}} X_O = \mathbb{E}_{\widetilde{\pi} \perp\!\!\!\perp \pi} \prod_{O \in \mathcal{O}} \mathbb{E}_{\mathcal{Q}}[X_O]$$

$$\boxed{\mathbb{E}_{\mathcal{Q}}[X_O] = \begin{cases} \frac{1}{1-\rho^{2|O|}} & \text{Gaussian} \\ 1+\rho^{2|O|} & \text{Erdős-Rényi} \end{cases}}$$

Failure of second-moment

We show

$$\mathbb{E}_{\mathcal{Q}}\left[\left(\frac{\mathcal{P}(A,B)}{\mathcal{Q}(A,B)}\right)^2\right] = \begin{cases} 1+o(1) & \text{if } \rho^2 \le \frac{(2-\epsilon)\log n}{n} \\ +\infty & \text{if } \rho^2 \ge \frac{(2+\epsilon)\log n}{n} \end{cases}$$

- Gaussian: suboptimal by a factor of 2
- ER graphs: suboptimal by an unbounded factor when $p = o(1)$

Failure of second-moment

We show

$$\mathbb{E}_{\mathcal{Q}}\left[\left(\frac{\mathcal{P}(A,B)}{\mathcal{Q}(A,B)}\right)^2\right] = \begin{cases} 1+o(1) & \text{if } \rho^2 \le \frac{(2-\epsilon)\log n}{n} \\ +\infty & \text{if } \rho^2 \ge \frac{(2+\epsilon)\log n}{n} \end{cases}$$

- Gaussian: suboptimal by a factor of 2
- ER graphs: suboptimal by an unbounded factor when $p = o(1)$

Obstruction from short orbits

$$\mathbb{E}_{(A,B)\sim\mathcal{Q}}\left[\left(\frac{\mathcal{P}(A,B)}{\mathcal{Q}(A,B)}\right)^2\right] = \mathbb{E}_{\pi\perp\!\!\!\perp\widetilde{\pi}}\left[\prod_{O\in\mathcal{O}} \mathbb{E}_{\mathcal{Q}}\left[X_O\right]\right] \overset{\widetilde{\pi}=\pi}{\ge} \frac{1}{n!}\left(1+\rho^2\right)^{\binom{n}{2}}$$

Failure of second-moment

We show

$$\mathbb{E}_{\mathcal{Q}}\left[\left(\frac{\mathcal{P}(A,B)}{\mathcal{Q}(A,B)}\right)^2\right] = \begin{cases} 1+o(1) & \text{if } \rho^2 \le \frac{(2-\epsilon)\log n}{n} \\ +\infty & \text{if } \rho^2 \ge \frac{(2+\epsilon)\log n}{n} \end{cases}$$

- Gaussian: suboptimal by a factor of 2
- ER graphs: suboptimal by an unbounded factor when $p = o(1)$

Obstruction from short orbits

$$\mathbb{E}_{(A,B)\sim\mathcal{Q}}\left[\left(\frac{\mathcal{P}(A,B)}{\mathcal{Q}(A,B)}\right)^2\right] = \mathbb{E}_{\pi \perp\!\!\!\perp \tilde{\pi}}\left[\prod_{O\in\mathcal{O}} \mathbb{E}_{\mathcal{Q}}\left[X_O\right]\right] \overset{\tilde{\pi}=\pi}{\ge} \frac{1}{n!}\left(1+\rho^2\right)^{\binom{n}{2}}$$

Atypically large magnitude of $\prod_{O\in\mathcal{O}:|O|=k} X_O$ for short orbits of length $k \lesssim \log n \Rightarrow$ second-moment blows up

Truncated second-moment: dense regime

It suffices to consider $k = 1$:

$$\prod_{O \in \mathcal{O}:|O|=1} X_O \approx \left(\frac{1}{p}\right)^{2e_{A\wedge B^\pi}(F)}$$

F: set of fixed points of $\sigma \triangleq \pi^{-1} \circ \widetilde{\pi}$
$A \wedge B^\pi$: Intersection graph
$e_{A\wedge B^\pi}(F)$: # of edges of subgraph of $A \wedge B^\pi$ induced by F

- Under $\mathcal{P}$: $e_{A\wedge B^\pi}(S)$ concentrates uniformly over all S when $|S|$ is large
- On this typical event $\mathcal{E}$ under $\mathcal{P}$, when $|F|$ is large,

$$\mathbb{E}_{\mathcal{Q}}\left[\prod_{O \in \mathcal{O}:|O|=1} X_O \mathbf{1}_{\{\mathcal{E}\}}\right] \lesssim \mathbb{E}_{\mathcal{Q}}\left[\left(\frac{1}{p}\right)^{2e_{A\wedge B_\pi}(F)} \mathbf{1}_{\left\{e_{A\wedge B_\pi}(F) \le \binom{|F|}{2} ps^2\right\}}\right]$$

Truncated second-moment: sparse regime

Need to consider $k = \Theta(\log n)$. It can be shown

- Long orbits:

$$\mathbb{E}_{\mathcal{Q}}\left[\prod_{|O|>k} X_O\right] \le \left(1+\rho^k\right)^{\frac{n^2}{k}} = 1 + o(1)$$

- Short incomplete orbits:

$$\mathbb{E}_{\mathcal{Q}}\left[X_O \mid O \not\subset E\left(A \wedge B^{\pi}\right)\right] \le 1$$

- Short complete orbits:

$$X_O = \left(\frac{1}{p}\right)^{2|O|}, \quad \forall O \subset E\left(A \wedge B^{\pi}\right)$$

In the subcritical regime $nps^2 < 1$, $A \wedge B^{\pi} \sim \mathcal{G}(n, ps^2)$ is a pseudoforest
$\Rightarrow H_k \triangleq \cup_{O:|O|\le k, O\subset E(A\wedge B^{\pi})} O$ is a pseudoforest

Truncated second-moment: orbit pseudoforest

On $\mathcal{E} \triangleq \{(A, B, \pi) : A \wedge B^{\pi} \text{ is a pseudoforest}\}$ under $\mathcal{P}$:

$$\mathbb{E}_{\mathcal{Q}}\left[\prod_{O \in \mathcal{O}} X_O \mathbf{1}_{\{\mathcal{E}\}}\right] \le (1+o(1))\, \mathbb{E}_{\mathcal{Q}}\left[\left(\frac{1}{p}\right)^{2e(H_k)} \mathbf{1}_{\{H_k \text{ is a pseudoforest}\}}\right]$$

$$= (1+o(1)) \sum_{H \in \mathcal{H}_k} s^{2e(H)} \quad \text{(generating function)}$$

$\mathcal{H}_k$: pseudoforests assembled from edge orbits of length at most k

Truncated second-moment: orbit pseudoforest

On $\mathcal{E} \triangleq \{(A, B, \pi) : A \wedge B^\pi$ is a pseudoforest$\}$ under $\mathcal{P}$:

$$\mathbb{E}_{\mathcal{Q}}\left[\prod_{O \in \mathcal{O}} X_O \mathbf{1}_{\{\mathcal{E}\}}\right] \le (1+o(1))\, \mathbb{E}_{\mathcal{Q}}\left[\left(\frac{1}{p}\right)^{2e(H_k)} \mathbf{1}_{\{H_k \text{ is a pseudoforest}\}}\right]$$

$$= (1+o(1)) \sum_{H \in \mathcal{H}_k} s^{2e(H)} \quad \text{(generating function)}$$

$\mathcal{H}_k$: pseudoforests assembled from edge orbits of length at most k

Key challenge: enumerate orbit pseudoforests $\mathcal{H}_k$ using orbit structure

Concluding remarks

- Formulate the problem of testing network correlation and characterize the statistical detection limit
- The impossibility proof applies truncated second-moment method
- The sparse setting leverages the pseudoforest structure of subcritical Erdős-Rényi graphs
- A large computational gap may exist between the statistical and computational limits

Open problem

- Sharp detection threshold in the sparse regime
- Prove the existence of or close the computational gaps
- Sharp partial recovery threshold

Reference

- Y. Wu, J. X, & S. H. Yu *Testing correlation of unlabeled random graphs*. arXiv:2008.10097.