Testing Correlation of Unlabeled Random Graphs

Jiaming Xu

The Fuqua School of Business
Duke University

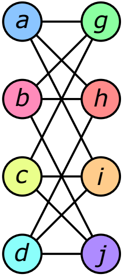
Joint work with
Yihong Wu (Yale) and Sophie H. Yu (Duke)

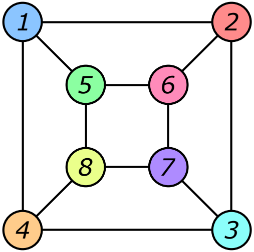
Computational Phase Transitions Workshop, Simons Institute
September 21, 2020

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

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

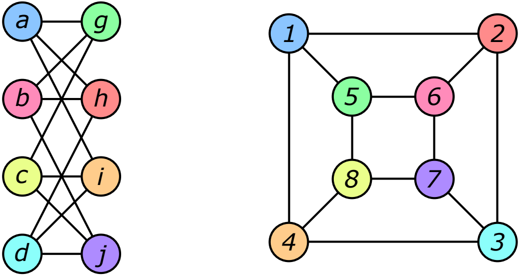




Graph isomorphism

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

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



$\pi(a) = 1$

$\pi(b) = 6$

$\pi(c) = 8$

$\pi(d) = 3$

$\pi(g) = 5$

$\pi(h) = 2$

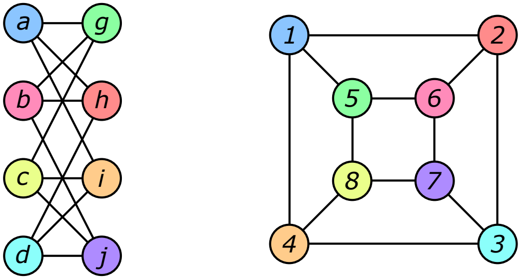
$\pi(i) = 4$

$\pi(j) = 7$

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

Beyond worst-case: Noisy graph isomorphism

Definition

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 $\mathcal{H}_1$

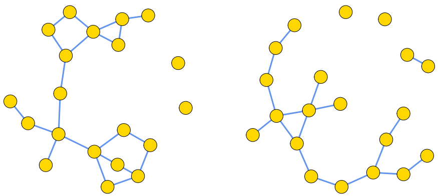
- Under $\mathcal{H}_1$, the inherent edge correlation is obscured by the latent node correspondence π
- The test needs to rely on graph invariants (invariant under graph isomorphism), such as
  - Subgraph counts (e.g. # of edges or triangles)
  - Spectrum (e.g. eigenvalues of adjacency matrices or Laplacians)

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-subsampled from a common parent graph $\mathcal{G}(n, p)$ with subsampling 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 $\rho \triangleq \frac{s(1-p)}{1-ps}$
- Hypothesis testing aspect of graph matching (recover π under $\mathcal{H}_1$)

Sharp threshold for detection: dense Erdős-Rényi graphs

Theorem (Wu-X.-Yu '20)

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

Sharp threshold for detection: dense Erdős-Rényi graphs

Theorem (Wu-X.-Yu '20)

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

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

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

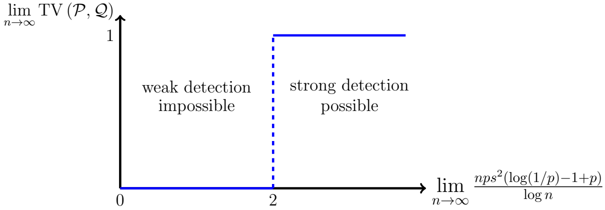
Sharp threshold for detection: dense Erdős-Rényi graphs

Theorem (Wu-X.-Yu '20)

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

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

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



Threshold for detection: sparse Erdős-Rényi graphs

Theorem (Wu-X.-Yu '20)

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

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

$$\textit{If } s^2 \leq \frac{1 - \omega(n^{-1/3})}{np} \wedge 0.01, \ \textit{then } \mathrm{TV}\left(\mathcal{P}, \mathcal{Q}\right) = 1 - \Omega(1)$$

$$\textit{If } s^2 \leq \frac{1 - \omega(n^{-1/3})}{np} \wedge o(1), \ \ \textit{then } \mathrm{TV}\left(\mathcal{P}, \mathcal{Q}\right) = o(1)$$

Threshold for detection: sparse Erdős-Rényi graphs

Theorem (Wu-X.-Yu '20)

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

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

$$\textit{If } s^2 \leq \frac{1-\omega(n^{-1/3})}{np} \wedge 0.01, \textit{ then } \mathrm{TV}\left(\mathcal{P}, \mathcal{Q}\right) = 1 - \Omega(1)$$

$$\textit{If } s^2 \leq \frac{1-\omega(n^{-1/3})}{np} \wedge o(1), \quad \textit{then } \mathrm{TV}\left(\mathcal{P}, \mathcal{Q}\right) = o(1)$$

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

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

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

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

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

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

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

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

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

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

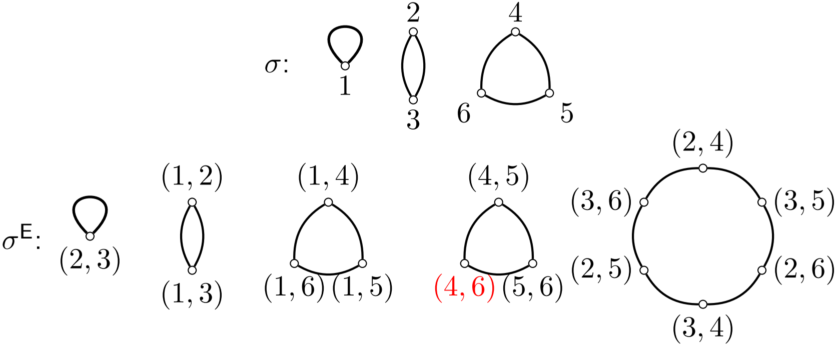
Second-moment calculation: cycle (orbit) decomposition

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

Second-moment calculation: cycle (orbit) decomposition

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

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 $\sigma \triangleq \pi^{-1}\circ\widetilde{\pi}$

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

$$\mathbb{E}_{\mathcal{Q}}[X_O] = 1+\rho^{2|O|}, \quad \rho \triangleq \frac{s(1-p)}{1-ps}$$ (use Egorychev method)

Failure of second-moment

We show

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

- Suboptimal by an unbounded factor when $p = o(1)$

Failure of second-moment

We show

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

- 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

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

- Suboptimal by an unbounded factor when $p = o(1)$

Obstruction from short orbits

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

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

Conditional 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
- Conditional on this typical event $\mathcal{E}$ under $\mathcal{P}$, when $|F|$ is large,

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

Conditional 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

Conditional second-moment: orbit pseudoforest

Conditional 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

Conditional second-moment: orbit pseudoforest

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

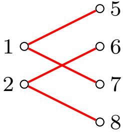
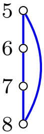
$$\mathbb{E}_{\mathcal{Q}}\left[\prod_{O \in \mathcal{O}} X_O \mathbf{1}_{\{\mathcal{E}\}}\right] \leq (1+o(1))\, \mathbb{E}_{\mathcal{Q}}\left[\left(\frac{1}{p}\right)^{2e(H_k)} \mathbf{1}_{\{H_k \text{ is a pseudoforest}\}}\right]$$

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

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

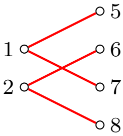
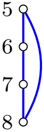
Key challenge: enumerate orbit pseudoforests $\mathcal{H}_k$ using orbit structure

Classification of edge orbits

Type	Edge orbit	Orbit graph
M	(13,24)	
B	(15,26,17,28)	
C	(56,67,78,85)	
S	(57,68)	

Assume $\sigma = (12)(34)(5678)$

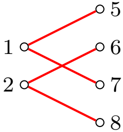
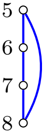
Classification of edge orbits

Type	Edge orbit	Orbit graph
M	(13,24)	
B	(15,26,17,28)	
C	(56,67,78,85)	
S	(57,68)	

Assume $\sigma = (12)(34)(5678)$

Type C edge orbit is a cycle

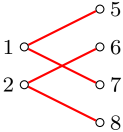
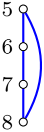
Classification of edge orbits

Type	Edge orbit	Orbit graph
M	(13,24)	
B	(15,26,17,28)	
C	(56,67,78,85)	
S	(57,68)	

Assume $\sigma = (12)(34)(5678)$

Type C edge orbit is a cycle

Type $\mathrm{B}_{m,\ell}$ ($\ell < m$) edge orbit is cycle-free if ℓ divides m; otherwise, it contains a component with at least two cycles

Type	Edge orbit	Orbit graph
M	(13,24)	
B	(15,26,17,28)	
C	(56,67,78,85)	
S	(57,68)	

Assume $\sigma = (12)(34)(5678)$

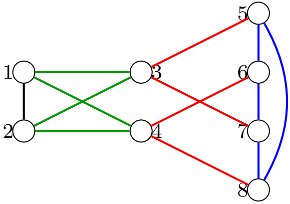
Type C edge orbit is a cycle

Type $\mathrm{B}_{m,\ell}$ ($\ell < m$) edge orbit is cycle-free if ℓ divides m; otherwise, it contains a component with at least two cycles

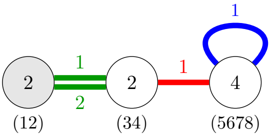
Key: identify the forbidden co-occurrence patterns under pseudoforest constraint

Backbone graph representation

- Node orbit in H (splits) $\Leftrightarrow$ giant node in Γ (shaded)
- Edge orbit in H $\Leftrightarrow$ giant edge in Γ with label



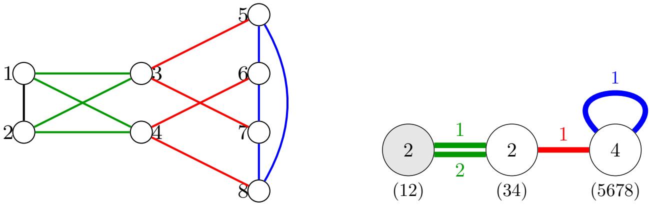
(a) Orbit graph H



(b) Backbone graph Γ

Backbone graph representation

- Node orbit in H (splits) $\Leftrightarrow$ giant node in Γ (shaded)
- Edge orbit in H $\Leftrightarrow$ giant edge in Γ with label



(a) Orbit graph H (b) Backbone graph Γ

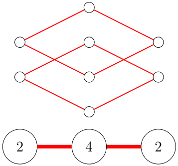
Enumeration Prototype: For $1 \le m \le k$,

1. Matching stage. Construct Γ_m induced by n_m giant nodes (node orbits) of size m
2. Splitting stage. Add splits to some components of Γ_m
3. Bridging stage. Add bridges between Γ_m and Γ_ℓ for ℓ dividing m

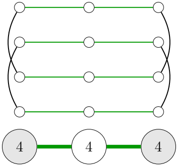
Warm-up: Enumerate orbit forest

When orbit graph H is a forest, its corresp. backbone graph Γ satisfies

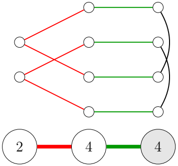
- Γ_m is a forest with no self-loop and parallel edges
- No bridge between Γ_m and Γ_ℓ unless ℓ divides m
- Each component of Γ_m contains at most 1 split or is incident to 1 bridge to Γ_ℓ, but not both.



(a) A component in Γ_4 is incident to 2 bridges



(b) A component in Γ_4 contains 2 splits



(c) A component in Γ_4 contains 1 split and is incident to 1 bridge

For $1 \le m \le k$,

1. Matching stage. Construct a rooted forest Γ_m consisting of n_m giant nodes and a_m giant edges
2. Splitting stage. Choose b_m components from $n_m - a_m$ tree components of Γ_m and add a split to the root
3. Bridging stage. Choose c_m out of the remaining $n_m - a_m - b_m$ tree components of Γ_m, add a bridge connecting its root to Γ_ℓ for some ℓ dividing m

$$\sum_{H \in \mathcal{F}_k} s^{2e(H)} \le \prod_{1 \le m \le k} \left(1 + \underbrace{s^{2m} m n_m}_{\text{matching}} + \underbrace{s^m \mathbf{1}_{\{m:\text{even}\}}}_{\text{splitting}} + \underbrace{s^{2m} \sum_{\ell < m} \ell n_\ell}_{\text{bridging}} \right)^{n_m}$$

Enumerate orbit pseudoforest

When orbit graph H is a forest, its corresp. backbone graph Γ satisfies

- Γ_m is a psedudoforest with self-loops and parallel edges counted as cycles
- No bridge between Γ_m and Γ_ℓ unless ℓ divides m
- Each unicyclic component of Γ_m is plain (contains no split and is not incident to bridge in Γ_ℓ)
- Each tree component of Γ_m contains at most 2 splits

Enumerate orbit pseudoforest

When orbit graph H is a forest, its corresp. backbone graph Γ satisfies

- Γ_m is a psedudoforest with self-loops and parallel edges counted as cycles
- No bridge between Γ_m and Γ_ℓ unless ℓ divides m
- Each unicyclic component of Γ_m is plain (contains no split and is not incident to bridge in Γ_ℓ)
- Each tree component of Γ_m contains at most 2 splits
- If there are two bridges incident to a common tree component in Γ_m, then the ending points of the two bridges must belong to distinct plain tree components in $\Gamma_{m/2}$
- If there is a bridge incident to a tree component that contains a split in Γ_m, then the ending point of the bridge must belong to a plain tree component in $\Gamma_{m/2}$

Enumerate orbit pseudoforest

For $1 \le m \le k$,

1. Matching stage. Construct a rooted pseudoforest Γ_m consisting of n_m giant nodes and a_m giant edges
2. Splitting stage. Choose b_m components from $n_m - a_m$ tree components of Γ_m and add either 1 or 2 splits
3. Forward bridging stage. Choose c_m out of the remaining $n_m - a_m - b_m$ tree components of Γ_m, add a bridge connecting its root to Γ_ℓ for some ℓ dividing m
4. Backward bridging stage. Choose d_m from the remaining $n_m - a_m - b_m - c_m$ tree components of Γ_m, add a bridge connecting its root to Γ_{2m}

$$\sum_{H \in \mathcal{H}_k} s^{2e(H)} \le \prod_{m=1}^{k} \Bigg(1 + \underbrace{s^{2m} 2mn_m}_{\text{matching}} + \underbrace{s^m n_m}_{\text{splitting}} + \underbrace{s^{2m} \sum_{\ell<m} \ell n_\ell}_{\text{FB}} + \underbrace{s^{4m} mn_{2m}}_{\text{BB}}\Bigg)^{n_m}$$

Averaging over orbit lengths

Theorem (Wu-X.-Yu '20)

Suppose $k(\log k)^4 = o(n)$. *If* $s \le 0.1$,

$$\mathbb{E}_{\pi \perp\!\!\!\perp \widetilde{\pi}} \left[\prod_{m=1}^{k} \left(1 + s^m n_m + 2s^{2m} \sum_{\ell \le m} \ell n_\ell + s^{4m} m n_{2m} \right)^{n_m} \right] = O(1),$$

where n_m *is the number of* m*-node orbits in* $\sigma = \pi^{-1} \circ \widetilde{\pi}$

- Poisson approximation: n_m's are approximately i.i.d. $\mathrm{Pois}(\frac{1}{m})$
- Partition the product into disjoint parts and recursively peel off the expectation backwards

Concluding remarks

- Formulate the problem of testing network correlation and characterize the statistical detection limit
- The impossibility proof applies conditional 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

Reference

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