Testing network correlation efficiently via counting trees

Jiaming Xu

The Fuqua School of Business
Duke University

Joint work with
Cheng Mao (GTech), Yihong Wu (Yale), and Sophie H. Yu (Duke)

JSM, August 11, 2022

Graph isomorphism

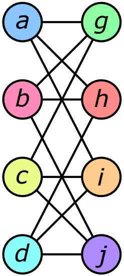
Given two graphs A and B, decide whether $A \cong B$, i.e., there exists a bijection $\pi : V(A) \to V(B)$ such that

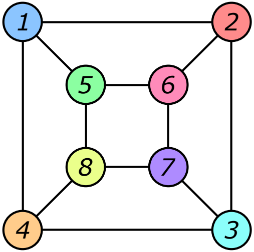
$$(u, v) \in E(A) \Leftrightarrow (\pi(u), \pi(v)) \in E(B)$$

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

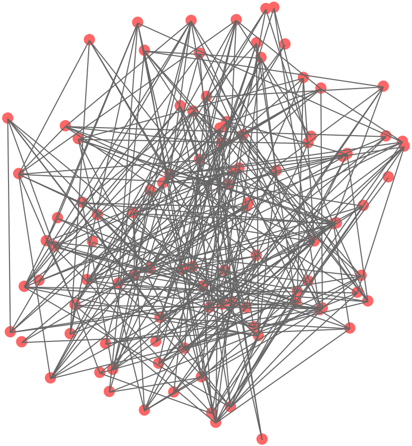
$$(u, v) \in E(A) \Leftrightarrow (\pi(u), \pi(v)) \in E(B)$$

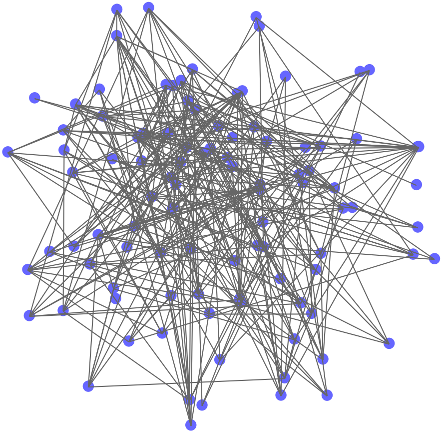




- Not known to be solvable in polynomial time in worst case

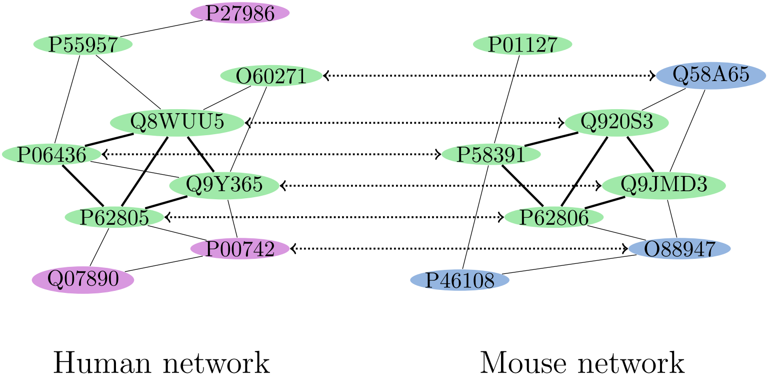
Two key challenges





- 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

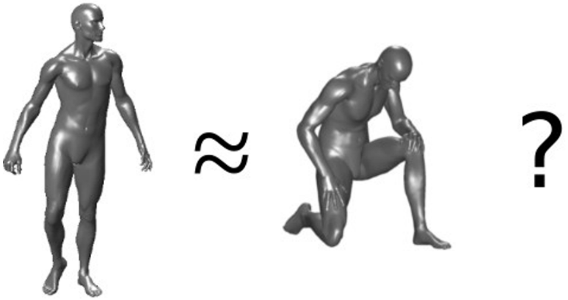


[Kazemi-Hassani-Grossglauser-Modarres '16]

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

Motivation: Deformable shape matching

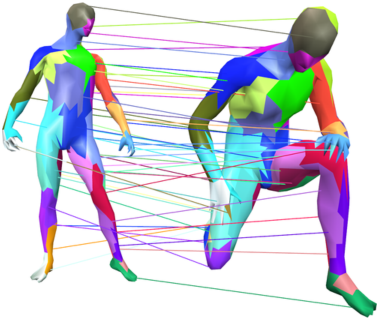
A fundamental problem in computer vision: Detect similar objects that undergo different deformations



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

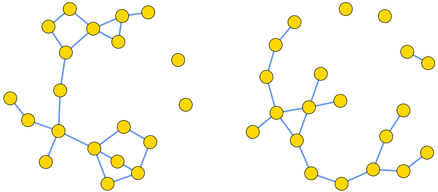
3-D shapes $\rightarrow$ geometric graphs (features $\rightarrow$ nodes, distances $\rightarrow$ edges)
Determine whether two graphs are similar in topologies

Beyond worst-case: noisy graph isomorphism test

Definition

$\mathbb{H}_0$: A and B are independent Erdős-Rényi graphs $\mathcal{G}(n, q)$

$\mathbb{H}_1$: A and $B^\pi = (B_{\pi(i)\pi(j)})$ are correlated $\mathcal{G}(n, q)$ with correlation parameter ρ, conditional on a uniform permutation π

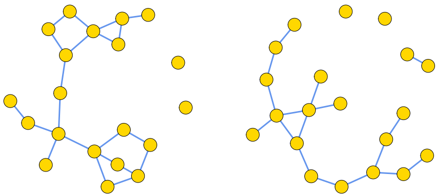


Beyond worst-case: noisy graph isomorphism test

Definition

$\mathbb{H}_0$: A and B are independent Erdős-Rényi graphs $\mathcal{G}(n, q)$

$\mathbb{H}_1$: A and $B^\pi = (B_{\pi(i)\pi(j)})$ are correlated $\mathcal{G}(n, q)$ with correlation parameter ρ, conditional on a uniform permutation π



- Hypothesis testing aspect of graph matching (recover π under $\mathbb{H}_1$)
- The inherent edge correlation is obscured by latent node mapping π
- The test needs to rely on graph invariants (e.g. # of edges or triangles)

Key questions

Goal: Design a test statistic $f(A, B)$ with threshold $\tau \in \mathbb{R}$ to achieve consistent detection:

$$\lim_{n\to\infty} \left[\mathbb{P}_{\mathbb{H}_1}\left(f(A,B) < \tau\right) + \mathbb{P}_{\mathbb{H}_0}\left(f(A,B) \geq \tau\right)\right] = 0$$

Two key questions

- What is the statistical detection limit?
- Can we attain the limit in polynomial-time?

Statistical detection limit

Consider an equivalent model: under $\mathbb{H}_1$, A and B are independently subsampled from a common parent graph $\mathcal{G}(n, p)$ with probability s: $q = ps$ and $\rho = \frac{s(1-p)}{1-ps}$

Theorem (Wu-X.-Yu '20)

Consistent detection is possible if and only if

$$s^2 \geq \begin{cases} \frac{2 \log n}{np\left(\log \frac{1}{p} - 1 + p\right)} & \text{if } p = n^{-o(1)} \text{ (dense)} \\ \Theta\left(\frac{1}{np}\right) & \text{if } p = n^{-\Omega(1)} \text{ (sparse)} \end{cases}$$

Statistical detection limit

Consider an equivalent model: under $\mathbb{H}_1$, A and B are independently subsampled from a common parent graph $\mathcal{G}(n,p)$ with probability s: $q = ps$ and $\rho = \frac{s(1-p)}{1-ps}$

Theorem (Wu-X.-Yu '20)

Consistent detection is possible if and only if

$$s^2 \geq \begin{cases} \frac{2\log n}{np\left(\log\frac{1}{p}-1+p\right)} & \text{if } p = n^{-o(1)} \text{ (dense)} \\ \Theta\left(\frac{1}{np}\right) & \text{if } p = n^{-\Omega(1)} \text{ (sparse)} \end{cases}$$

- [Ding-Hu' 22] determined the sharp constant in the sparse graph case
- Achieved by computationally intractable statistic:

$$\mathsf{MCS}(A,B) \triangleq \max_{\pi\in\mathcal{S}_n} \sum_{i<j} A_{ij}B_{\pi(i)\pi(j)} \quad \text{(max common subgraph)}$$

- Question: How to test efficiently?

Statistical detection limit

Consider an equivalent model: under $\mathbb{H}_1$, A and B are independently subsampled from a common parent graph $\mathcal{G}(n,p)$ with probability s: $q = ps$ and $\rho = \frac{s(1-p)}{1-ps}$

Theorem (Wu-X.-Yu '20)

Consistent detection is possible if and only if

$$s^2 \geq \begin{cases} \frac{2\log n}{np\left(\log\frac{1}{p}-1+p\right)} & \text{if } p = n^{-o(1)} \text{ (dense)} \\ \Theta\left(\frac{1}{np}\right) & \text{if } p = n^{-\Omega(1)} \text{ (sparse)} \end{cases}$$

- [Ding-Hu' 22] determined the sharp constant in the sparse graph case
- Achieved by computationally intractable statistic:

$$\mathsf{MCS}(A,B) \triangleq \max_{\pi \in \mathcal{S}_n} \sum_{i<j} A_{ij} B_{\pi(i)\pi(j)} \quad \text{(max common subgraph)}$$

- Question: How to test efficiently? subgraph count

Testing via subgraph counts

Let $X_H(G)$ denote the number of copies of H in G

- Under $\mathbb{H}_0$: $X_H(A)$ and $X_H(B)$ are independent
- Under $\mathbb{H}_1$: $X_H(A)$ and $X_H(B)$ are correlated

Implication: Test based on the correlation between $X_H(A)$ and $X_H(B)$

Testing via subgraph counts

Let $X_H(G)$ denote the number of copies of H in G

- Under $\mathbb{H}_0$: $X_H(A)$ and $X_H(B)$ are independent
- Under $\mathbb{H}_1$: $X_H(A)$ and $X_H(B)$ are correlated

Implication: Test based on the correlation between $X_H(A)$ and $X_H(B)$

First attempt: H is a single edge [Wu-X.-Yu '20]

$$\frac{X_H(A) - X_H(B)}{\sqrt{n(n-1)q(1-q)}} \approx \begin{cases} \mathcal{N}(0,1) & \text{Under } \mathbb{H}_0 \\ \mathcal{N}(0,1-\rho) & \text{Under } \mathbb{H}_1 \end{cases}$$

$\Rightarrow$ consistent detection demands $\rho \to 1$

Testing via subgraph counts

Let $X_H(G)$ denote the number of copies of H in G

- Under $\mathbb{H}_0$: $X_H(A)$ and $X_H(B)$ are independent
- Under $\mathbb{H}_1$: $X_H(A)$ and $X_H(B)$ are correlated

Implication: Test based on the correlation between $X_H(A)$ and $X_H(B)$

First attempt: H is a single edge [Wu-X.-Yu '20]

$$\frac{X_H(A) - X_H(B)}{\sqrt{n(n-1)q(1-q)}} \approx \begin{cases} \mathcal{N}(0,1) & \text{Under } \mathbb{H}_0 \\ \mathcal{N}(0,1-\rho) & \text{Under } \mathbb{H}_1 \end{cases}$$

$\Rightarrow$ consistent detection demands $\rho \to 1$

Lesson: need to count more "distinguishing" subgraph

Counting a flock of black swans

Counting a family $\mathcal{H}$ of rare subgraphs [Barak-Chou-Lei-Schramm-Sheng'19]

- H is rare: $\mathbb{E}\left[X_H(A)\right] = O(1) \Rightarrow$ unlikely to co-occur under $\mathbb{H}_0$
- $\mathcal{H}$ needs to be "rich" to ensure co-occurrence under $\mathbb{H}_1$
- co-occurrences of different H need to be almost uncorrelated

Counting a flock of black swans

Counting a family $\mathcal{H}$ of rare subgraphs [Barak-Chou-Lei-Schramm-Sheng'19]

- H is rare: $\mathbb{E}[X_H(A)] = O(1) \Rightarrow$ unlikely to co-occur under $\mathbb{H}_0$
- $\mathcal{H}$ needs to be "rich" to ensure co-occurrence under $\mathbb{H}_1$
- co-occurrences of different H need to be almost uncorrelated

Construction of $\mathcal{H}$ based on random d-regular graphs of size $K = \rho^{-C}$
$\Rightarrow$ correctly tell $\mathbb{H}_0$ vs. $\mathbb{H}_1$ w.p. 0.9 in time n^K, if

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

Counting a flock of black swans

Counting a family $\mathcal{H}$ of rare subgraphs [Barak-Chou-Lei-Schramm-Sheng'19]

- H is rare: $\mathbb{E}[X_H(A)] = O(1) \Rightarrow$ unlikely to co-occur under $\mathbb{H}_0$
- $\mathcal{H}$ needs to be "rich" to ensure co-occurrence under $\mathbb{H}_1$
- co-occurrences of different H need to be almost uncorrelated

Construction of $\mathcal{H}$ based on random d-regular graphs of size $K = \rho^{-C}$
$\Rightarrow$ correctly tell $\mathbb{H}_0$ vs. $\mathbb{H}_1$ w.p. 0.9 in time n^K, if

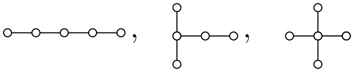
$$\rho = \Omega(1) \quad \text{and} \quad nq \in \left[n^{\epsilon}, n^{1/153}\right] \cup \left[n^{2/3}, n^{1-\epsilon}\right]$$

Limitations:

- nq needs to fall into specific range and does not cover sparse graphs
- The test error is not vanishing
- The runtime grows polynomially at rate of $\rho^{-C} \Rightarrow$ too expensive for large networks

Our idea: counting large trees

Our strategy: count a family $\mathcal{T}$ of trees of K edges with K growing in n

e.g. $K = 4, \quad \mathcal{T} = \Big\{$  $\Big\}$

- $\#$ of vertices $= \#$ of edges $+1 \Rightarrow$ abundant even in sparse graphs
- $\mathcal{T}$ is "rich": $|\mathcal{T}|$ grows exponentially in K
- Tree structure allows for efficient counting even when K grows in n

An obstacle: subgraph counts for different trees are still too correlated, especially when $nq \geq 1$

Our solution: counting signed trees

Counting signed trees

- Centered adjacency matrices: $\overline{A} = A - \mathbb{E}[A]$ and $\overline{B} = B - \mathbb{E}[B]$

Counting signed trees

- Centered adjacency matrices: $\overline{A} = A - \mathbb{E}[A]$ and $\overline{B} = B - \mathbb{E}[B]$
- Counting signed trees

$$f_{\mathcal{T}}(A, B) \triangleq \sum_{H \in \mathcal{T}} \mathsf{aut}(H) W_H(\overline{A}) W_H(\overline{B}),$$

where $\mathsf{aut}(H)$ is the number of automorphism of H, and

$$W_H(M) \triangleq \sum_{S \cong H} \prod_{(i,j) \in S} M_{ij} \quad \text{(signed subgraph count)}$$

Counting signed trees

- Centered adjacency matrices: $\overline{A} = A - \mathbb{E}[A]$ and $\overline{B} = B - \mathbb{E}[B]$
- Counting signed trees

$$f_{\mathcal{T}}(A, B) \triangleq \sum_{H \in \mathcal{T}} \mathsf{aut}(H) W_H(\overline{A}) W_H(\overline{B}),$$

where $\mathsf{aut}(H)$ is the number of automorphism of H, and

$$W_H(M) \triangleq \sum_{S \cong H} \prod_{(i,j) \in S} M_{ij} \quad \text{(signed subgraph count)}$$

- Crucial orthogonal property:

$$\mathbb{E}\left[W_H(\overline{A}) W_{H'}(\overline{A})\right] = 0 \Longleftrightarrow H \ncong H'$$

Statistical guarantee

Theorem (Mao-Wu-X.-Yu '21)

Suppose

$$nq \geq n^{-o(1)}, \quad \rho^2 > \alpha, \quad \omega(1) \leq K = O\left(\frac{\log n}{\log \log n}\right).$$

Then the testing error satisfies

$$\mathbb{P}_{\mathbb{H}_1}\left(f_{\mathcal{T}}(A,B) \leq \tau\right) + \mathbb{P}_{\mathbb{H}_0}\left(f_{\mathcal{T}}(A,B) \geq \tau\right) = o(1),$$

where $\tau = C\mathbb{E}_{\mathbb{H}_1}\left[f_{\mathcal{T}}(A,B)\right]$ *for any fixed constant* $0 < C < 1$.

- $\alpha \approx 0.33833$ is Otter's constant given by $|\mathcal{T}| \approx (1/\alpha)^K$
- $nq \geq n^{-o(1)}$ applies to very sparse regime and is necessary for existence of trees with $K = \omega(1)$ edges

Proof sketch: mean and variance calculations

$$f_{\mathcal{T}}(A,B) \triangleq \sum_{H \in \mathcal{T}} \mathsf{aut}(H) W_H(\overline{A}) W_H(\overline{B}), \; W_H(M) \triangleq \sum_{S \cong H} \prod_{(i,j) \in S} M_{ij}$$

- $\mathbb{E}_{\mathbb{H}_0}[f_{\mathcal{T}}(A,B)] = 0$ and by the orthogonal property,

$$\text{(SNR)} \quad \frac{(\mathbb{E}_{\mathbb{H}_1}[f_{\mathcal{T}}(A,B)])^2}{\mathrm{Var}_{\mathbb{H}_0}[f_{\mathcal{T}}(A,B)]} = \rho^{2K} |\mathcal{T}| \approx \left(\frac{\rho^2}{\alpha} \right)^K$$

Proof sketch: mean and variance calculations

$$f_{\mathcal{T}}(A,B) \triangleq \sum_{H\in\mathcal{T}} \mathsf{aut}(H) W_H(\overline{A}) W_H(\overline{B}),\ W_H(M) \triangleq \sum_{S\cong H} \prod_{(i,j)\in S} M_{ij}$$

- $\mathbb{E}_{\mathbb{H}_0}[f_{\mathcal{T}}(A,B)] = 0$ and by the orthogonal property,

$$\text{(SNR)}\quad \frac{(\mathbb{E}_{\mathbb{H}_1}[f_{\mathcal{T}}(A,B)])^2}{\mathrm{Var}_{\mathbb{H}_0}[f_{\mathcal{T}}(A,B)]} = \rho^{2K}|\mathcal{T}| \approx \left(\frac{\rho^2}{\alpha}\right)^K$$

- To drive the SNR to ∞, need $\rho^2 > \alpha \approx 0.33833$

Proof sketch: mean and variance calculations

$$f_{\mathcal{T}}(A,B) \triangleq \sum_{H \in \mathcal{T}} \mathsf{aut}(H) W_H(\overline{A}) W_H(\overline{B}),\ W_H(M) \triangleq \sum_{S \cong H} \prod_{(i,j) \in S} M_{ij}$$

- $\mathbb{E}_{\mathbb{H}_0}[f_{\mathcal{T}}(A,B)] = 0$ and by the orthogonal property,

$$\text{(SNR)} \quad \frac{(\mathbb{E}_{\mathbb{H}_1}[f_{\mathcal{T}}(A,B)])^2}{\mathrm{Var}_{\mathbb{H}_0}[f_{\mathcal{T}}(A,B)]} = \rho^{2K}|\mathcal{T}| \approx \left(\frac{\rho^2}{\alpha}\right)^K$$

- To drive the SNR to ∞, need $\rho^2 > \alpha \approx 0.33833$
- Further show

$$\frac{(\mathbb{E}_{\mathbb{H}_1}[f_{\mathcal{T}}(A,B)])^2}{\mathrm{Var}_{\mathbb{H}_1}[f_{\mathcal{T}}(A,B)]} \to \infty$$

  Needs delicate analysis of correlations between subgraph counts

Computational guarantees

Efficient tree counts via color coding

How to compute $W_H(M) \triangleq \sum_{S \cong H} \prod_{(i,j) \in S} M_{ij}$ efficiently?

Efficient tree counts via color coding

How to compute $W_H(M) \triangleq \sum_{S \cong H} \prod_{(i,j) \in S} M_{ij}$ efficiently?

- Exhaustive search takes n^K time: super poly-time when $K \to \infty$

Efficient tree counts via color coding

How to compute $W_H(M) \triangleq \sum_{S \cong H} \prod_{(i,j)\in S} M_{ij}$ efficiently?

- Exhaustive search takes n^K time: super poly-time when $K \to \infty$
- Our solution: color coding [Alon-Yuster-Zwick '95]
  1. Assign random color μ to each vertex from color set $[K+1]$ uniformly
  2. Define *colorful* subgraph count

$$Y_H(M,\mu) \triangleq \sum_{S \cong H} \mathbf{1}_{\{S \text{ has } K+1 \text{ distinct colors}\}} \prod_{(i,j)\in S} M_{ij}$$

     Key implication: $\mathbb{E}_\mu \left[Y_H(M,\mu) \right] = r W_H(M)$, where $r = \frac{(K+1)!}{(K+1)^{K+1}}$

  3. Generate $1/r$ independent random colorings μ_t so that

$$W_H(M) \approx \sum_{t=1}^{1/r} Y_H(M,\mu_t) \triangleq \widehat{W}_H(M)$$

$Y_H(M,\mu)$ can be computed via dynamic programming in $n^2 e^{O(K)}$

Computational guarantees

Define an approximate statistic:

$$\widetilde{f}_{\mathcal{T}} \triangleq \sum_{H \in \mathcal{T}} \mathsf{aut}(H) \widehat{W}_H(\overline{A}) \widehat{W}_H(\overline{B})$$

Theorem (Mao-Wu-X.-Yu '21)

Under both $\mathbb{H}_0$ and $\mathbb{H}_1$,

$$\frac{\widetilde{f}_{\mathcal{T}} - f_{\mathcal{T}}}{\mathbb{E}_{\mathbb{H}_1}[f_{\mathcal{T}}]} \xrightarrow{L_2} 0.$$

Moreover, $\widetilde{f}_{\mathcal{T}}$ can be computed in time $n^{2+o(1)}$.

Implication: $\widetilde{f}_{\mathcal{T}}$ satisfies the same statistical guarantee as $f_{\mathcal{T}}$

Statistical-computational gap

Statistical-computational gap

- Statistical limit: $\rho \approx \frac{1}{nq}$
- Tree counting limit: $\rho^2 > \alpha \approx 0.33833$
- There is a large computational gap, especially when the graphs are dense

Key question: Can we close the computational gap?

Statistical-computational gap

- Statistical limit: $\rho \approx \frac{1}{nq}$
- Tree counting limit: $\rho^2 > \alpha \approx 0.33833$
- There is a large computational gap, especially when the graphs are dense

Key question: Can we close the computational gap?

No, based on the low-degree hardness conjecture [Hopkins '18, Kunisky-Wein-Bandeira '19]: computationally hard when $\rho \leq \frac{1}{\text{polylog}(n)}$

Signed tree counting as a low-degree projection

- Space of functions on the hypercube $\{0,1\}^{2\binom{n}{2}}$ endowed with

$$\langle f, g \rangle \triangleq \mathbb{E}_{\mathbb{H}_0}\left[f(A,B)g(A,B)\right]$$

Signed tree counting as a low-degree projection

- Space of functions on the hypercube $\{0,1\}^{2\binom{n}{2}}$ endowed with

$$\langle f, g \rangle \triangleq \mathbb{E}_{\mathbb{H}_0}\left[f(A,B) g(A,B) \right]$$

- Orthonormal polynomial basis [O'Donnell '4]

$$\phi_{\mathsf{S}}(A,B) \triangleq \sigma^{-|S_1|-|S_2|} \prod_{(i,j)\in S_1} \overline{A}_{ij} \prod_{(k,\ell)\in S_2} \overline{B}_{k\ell}, \quad \mathsf{S} \triangleq (S_1, S_2)$$

where $\sigma^2 = q(1-q)$

Signed tree counting as a low-degree projection

- Space of functions on the hypercube $\{0,1\}^{2\binom{n}{2}}$ endowed with

$$\langle f, g\rangle \triangleq \mathbb{E}_{\mathbb{H}_0}\left[f(A,B)g(A,B)\right]$$

- Orthonormal polynomial basis [O'Donnell '4]

$$\phi_{\mathsf{S}}(A,B) \triangleq \sigma^{-|S_1|-|S_2|} \prod_{(i,j)\in S_1} \overline{A}_{ij} \prod_{(k,\ell)\in S_2} \overline{B}_{k\ell}, \quad \mathsf{S} \triangleq (S_1, S_2)$$

where $\sigma^2 = q(1-q)$

- Low-degree projection of likelihood ratio $L(A,B) = \frac{\mathbb{P}_{\mathbb{H}_1}(A,B)}{\mathbb{P}_{\mathbb{H}_0}(A,B)}$

$$f_{\mathcal{H}}(A,B) \triangleq \sum_{H\in\mathcal{H}} \sum_{S_1: S_1\cong H} \sum_{S_2: S_2\cong H} \langle L, \phi_{\mathsf{S}}\rangle \, \phi_{\mathsf{S}}$$

for given collection $\mathcal{H}$ of unlabeled graphs

Low-degree hardness

- "Optimal" degree-$2K$ polynomial of (A, B):

$$f^* = \underset{f:\ \text{degree}(f) \le 2K}{\arg\max} \frac{\left(\mathbb{E}_{\mathbb{H}_1}\left[f\right]\right)^2}{\mathbb{E}_{\mathbb{H}_0}\left[f^2\right]}$$

Low-degree hardness

- "Optimal" degree-$2K$ polynomial of (A, B):

$$f^* = \underset{f\colon \text{degree}(f) \le 2K}{\arg\max} \frac{\left(\mathbb{E}_{\mathbb{H}_1}[f]\right)^2}{\mathbb{E}_{\mathbb{H}_0}[f^2]} = f_{\mathcal{H}^*}(A, B)$$

where $\mathcal{H}^*$ consists of all unlabeled graphs induced by K edges

- "Optimal" degree-$2K$ polynomial of (A, B):

$$f^* = \arg\max_{f\colon \text{degree}(f) \le 2K} \frac{(\mathbb{E}_{\mathbb{H}_1}[f])^2}{\mathbb{E}_{\mathbb{H}_0}[f^2]} = f_{\mathcal{H}^*}(A, B)$$

where $\mathcal{H}^*$ consists of all unlabeled graphs induced by K edges

- $$\frac{(\mathbb{E}_{\mathbb{H}_1}[f])^2}{\mathbb{E}_{\mathbb{H}_0}[f^2]} = \sum_{H \in \mathcal{H}^*} \rho^{2|E(H)|} \le \sum_{k=1}^{K} (4ek)^k \rho^{2k} = O(1)$$

provided that $\rho^2 \lesssim \frac{1}{K}$

Low-degree hardness

- “Optimal” degree-$2K$ polynomial of (A, B):

$$f^* = \underset{f:\ \text{degree}(f) \le 2K}{\arg\max} \frac{(\mathbb{E}_{\mathbb{H}_1}[f])^2}{\mathbb{E}_{\mathbb{H}_0}[f^2]} = f_{\mathcal{H}^*}(A, B)$$

  where $\mathcal{H}^*$ consists of all unlabeled graphs induced by K edges

- $$\frac{(\mathbb{E}_{\mathbb{H}_1}[f])^2}{\mathbb{E}_{\mathbb{H}_0}[f^2]} = \sum_{H \in \mathcal{H}^*} \rho^{2|E(H)|} \le \sum_{k=1}^{K} (4ek)^k \rho^{2k} = O(1)$$

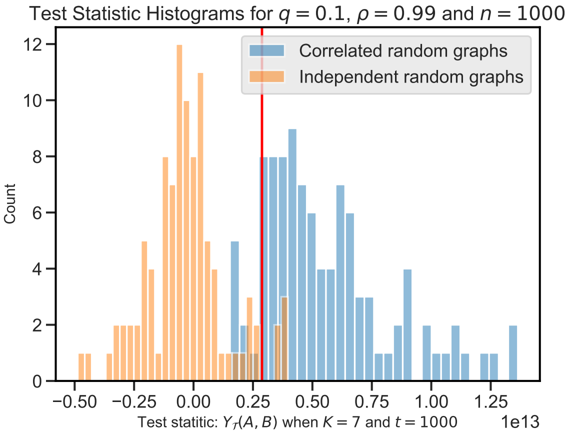
  provided that $\rho^2 \lesssim \frac{1}{K}$

- Low-degree hardness conjecture [Hopkins '18, Kunisky-Wein-Bandeira '19]: All degree-$\log n$ polynomials fail when $\rho \le \frac{1}{\text{polylog}(n)}$

Simulation study

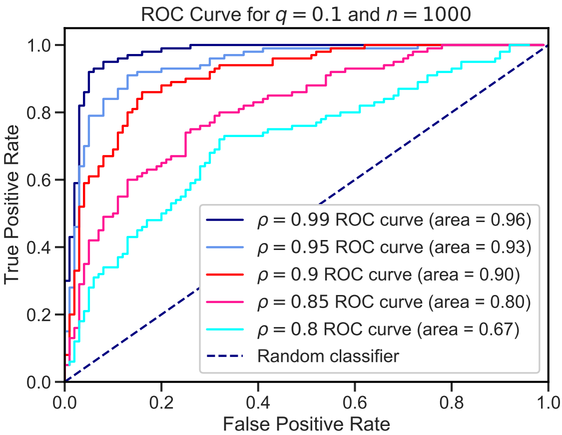
Simulation results

$n = 1000,\ K = 7,\ \rho = 0.99$



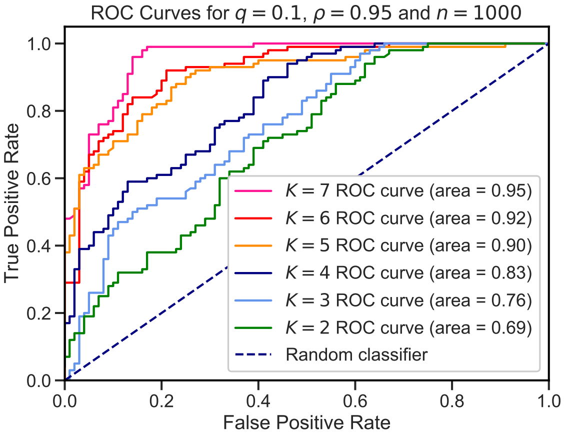
Type-I and type-II error are around 5% and 9%

$n = 1000$ and $K = 6$



Consistent with theoretical prediction: need $\rho \gtrsim |T|^{-\frac{1}{2K}} \approx 0.82$ for SNR exceeding 1, as $|\mathcal{T}| = 11$ when $K = 6$.

$n = 1000,\ q = 0.1,\ \rho = 0.95$



Performance of tree counting statistic improves as tree size K increases

Concluding remarks

- Efficient test via counting signed trees up to Otter's constant
- Low-degree hardness conjecture: no efficient test when $\rho \leq \frac{1}{\mathrm{polylog}(n)}$

Concluding remarks

- Efficient test via counting signed trees up to Otter's constant
- Low-degree hardness conjecture: no efficient test when $\rho \leq \frac{1}{\text{polylog}(n)}$

Future work:

- Polynomial-test up to $\rho = o(1)$: bounded tree-width graph
- Recovery of node correspondence via signed tree signatures
- Testing network correlation in the presence of degree heterogeneity

Concluding remarks

- Efficient test via counting signed trees up to Otter's constant
- Low-degree hardness conjecture: no efficient test when $\rho \leq \frac{1}{\text{polylog}(n)}$

Future work:

- Polynomial-test up to $\rho = o(1)$: bounded tree-width graph
- Recovery of node correspondence via signed tree signatures
- Testing network correlation in the presence of degree heterogeneity

Reference

- C. Mao, Y. Wu, J. X, & S. H. Yu *Testing network correlation efficiently via counting trees*. arXiv:2110.11816
- Y. Wu, J. X, & S. H. Yu *Testing correlation of unlabeled random graphs*. arXiv:2008.10097. To appear in *Annals of Applied Probability*