Network Two-Sample Test

Yuan Zhang

ArXiv: 2208.07573

Joint work with Meijia Shao (OSU), Dong Xia (HKUST), Qiong Wu (UPenn) and Shuo Chen (UMD)

Introduction and motivation

Networks: relational data

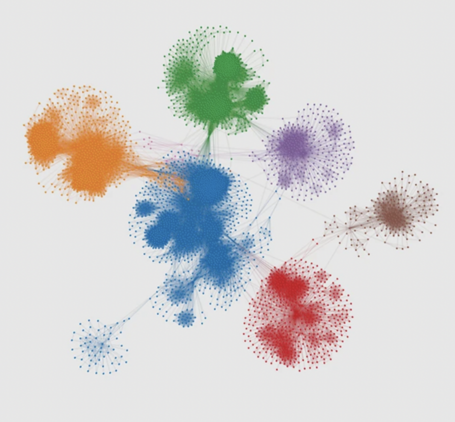


Figure: Plot of some socical network *(Fortunato & Newman: 20 years of network community detection, Nature, 2022)*

Networks: relational data

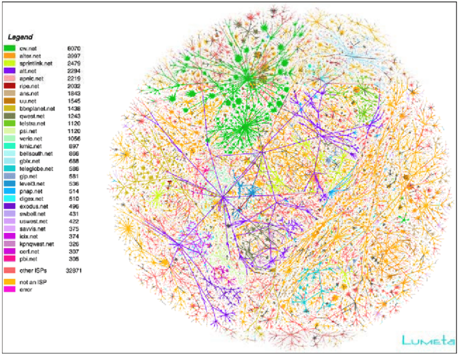


Figure: Internet network *(Zook et al., 2004)*

Introduction and motivation

- Relational nature: no direct observations on individuals
- Research interests (on one network)
  - Clustering nodes (community detection) *Zhao et al. (2012); Zhang & Zhou (2015); Gao et al. (2018)*
  - Learning node representation *Young & Scheinerman (2007); Cape et al. (2018); Lei (2021); Xie (2022+)*
  - Denoising and link prediction *Chatterjee (2015); Gao et al. (2015); Z et al. (2017); Xu (2018)*
  - And many, many more...

Introduction and motivation

Levin & Levina (2019) says well:

"A core problem in statistical network analysis is to develop network analogues of classical techniques."

Today's topic: network two-sample test

- Given two sets of networks $A^{(1)}, \ldots, A^{(N_A)}; B^{(1)}, \ldots, B^{(N_B)}$
- **Question:** Compare the models behind networks A and B
- Motivating applications:
  - Sociology: compare structural roles of two nodes through their ego-networks
  - Biomedical studies: compare brain imaging networks for diagnosis and disease sub-type discovery

Introduction and motivation

Two-sample test: $A^{(1)}, \ldots, A^{(N_A)}$ vs. $B^{(1)}, \ldots, B^{(N_B)}$

Problem difficulty depends on:

- 1. Multiple observations: $N_A, N_B \to \infty$?
  Ginestet et al. (2017); Ghoshdastidar & Von Luxburg (2018); Kolaczyk et al. (2020); Ghoshdastidar et al. (2020); Chen et al. (2022); Maugis et al. (2020); Bravo-Hermsdorff et al. (2021); Yuan & Wen (2021)

  With diverging N_A, N_B, one can:
  - Easily denoise A and B
  - Easily estimate the variability of your test statistic

  Issue:
  - Within-group heterogeneity (e.g., schizophrenia data)

Introduction and motivation

Two-sample test: $A^{(1)}, \ldots, A^{(N_A)}$ vs. $B^{(1)}, \ldots, B^{(N_B)}$

Problem difficulty depends on:

- 2. Known node registration?
  Ghoshdastidar and Von Luxburg (2018); Li and Li (2018); Ghoshdastidar et al. (2020); Chen et al. (2022).
  With known node registration, one can:
    - Perform edge-wise comparison

  But node correspondence...
    - may not exist (e.g., Google+ ego-network data)
    - may not be known (e.g., person $\leftrightarrow$ social media accounts)

Introduction and motivation

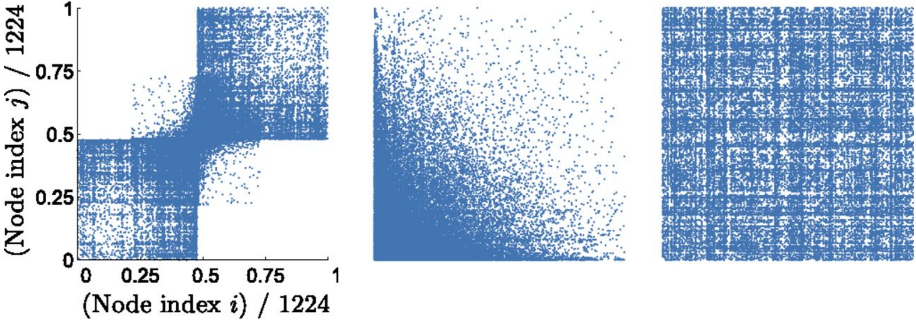


Figure: The same network plotted with three different node orders *(Olhede & Wolfe, 2014)*

Introduction and motivation

Two-sample test: $A^{(1)}, \ldots, A^{(N_A)}$ vs. $B^{(1)}, \ldots, B^{(N_B)}$

Assumption 1: Multiple observations
Assumption 2: Known node registration

- Many existing works assume both:
  Ghoshdastidar and Von Luxburg (2018); Ghoshdastidar et al. (2020); Chen et al. (2022).
- Some assume neither, but usually need other significant assumptions (low-rankness, degree monotonicity i.e. identification by degree):
  Tang et al. (2017); Agterberg et al. (2020); Yang et al. (2014); Sabanayagam et al. (2021)
  (*graph-matching*-based methods)

Introduction and motivation

Goals:

- No multiple observations: $N_A = N_B = 1$
- No/Unknown node registration
- Avoid strong structural assumptions
- Scalable and memory-parsimonious algorithm
- Handle potentially very different network sizes and sparsity levels
- Finite-sample higher-order accuracy

Problem set up

Input data:

- Adjacency matrices: $A \in \{0,1\}^{m\times m}$, $B \in \{0,1\}^{n\times n}$

Base model:

- Adjacency matrix: $A \in \{0,1\}^{m\times m}$

$$A_{ij} = A_{ji} = \begin{cases} 1 & i \leftrightarrow j \\ 0 & \text{otherwise} \end{cases}$$

- Symmetric edge probability matrix: $W^{(A)} \in \mathbb{R}^{m\times m}$:

$$A_{ij} | W^{(A)} \overset{\text{independent}}{\sim} \text{Bernoulli}(W_{ij}^{(A)})$$

Problem set up

Graphon model: two-stage data generation

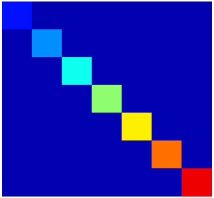
1. Latent node positions $X_1, \ldots, X_m \overset{\text{i.i.d.}}{\sim}$ Uniform$[0,1]$
2. Edge probabilities:
$$W_{ij}^{(A)} = \rho_A \cdot f_A(X_i, X_j)$$
where:
   - Latent graphon function $f_A : [0,1]^2 \to [0,1]$, s.t. $\int_0^1 \int_0^1 f_A(u,v)\, \mathrm{d}u\, \mathrm{d}v = 1$
   - ρ_A: sparsity multiplier

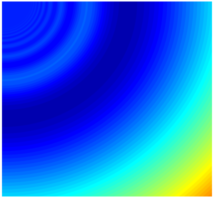
Problem set up

$$W_{ij}^{(A)} = \rho_A \cdot f_A(X_i, X_j)$$

Remarks:

- f_A encodes all network structures
- X_i encodes node's role
- Both are inestimable





Our method

Goal: compare f_A vs f_B, using just A and B

Is this even doable?

Idea: compare network summary statistics

- Question 1: which statistics to compare?
- Question 2: similar summary statistics $\overset{?}{\Leftrightarrow}$ similar network models

Our method

Question 1: which statistics to compare?

Network moment: frequency of the corresponding motif

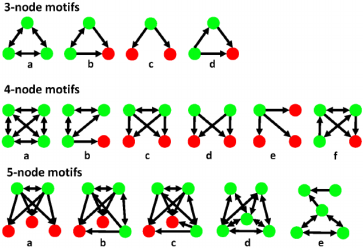


Figure: Network motifs *(Jayavelu & Bar, 2014)*

Our method

- Network motif: R with r nodes and s edges
- **Empirical network moment**

$$\widehat{U}_m := \binom{m}{r}^{-1} \sum_{1 \le i_1 < \ldots < i_r \le m} h(A_{i_1,\ldots,i_r})$$

where the kernel $h(\cdot)$ is

$$h(A_{i_1,\ldots,i_r}) := \begin{cases} 1, & \text{if } A_{i_1,\ldots,i_r} \text{ contains } R \\ 0, & \text{otherwise} \end{cases}$$

- **Example:** $R =$ Triangle:

$$\widehat{U}_m := \binom{m}{3}^{-1} \sum_{1 \le i_1 < i_2 < i_3 \le m} A_{i_1,i_2} A_{i_2,i_3} A_{i_3,i_1} \tag{1}$$

- Define population network moment: $\mu_m := \mathbb{E}[\widehat{U}_m]$
- Similarly define $\widehat{V}_n$ and ν_n for B

Our method

Question 2: similar summary statistics $\overset{?}{\Leftrightarrow}$ similar network models

- **Goal:** Horvitz–Thompson estimator (population version)

$$d_{m,n,\rho_A,\rho_B} := \rho_A^{-s} \cdot \mu_m - \rho_B^{-s} \cdot \nu_n$$

- Similar network moments $\approx$ similar network models!
  (Borgs et al, 2008)

Theorem 1 (Inverse Counting Lemma, [7]). *Suppose the graphon functions f_A and f_B are measurable and bounded by 1. For any k_0, if $|d_{m,n,\rho_A,\rho_B}| \leq 3^{-k_0^2}$ holds for all motifs with at most k_0 nodes, then the* cut distance, *denoted by $\delta_\square(\cdot,\cdot)$, satisfies*

$$\delta_\square(f_A, f_B) := \inf_\sigma \sup_{S,T \subset [0,1]} \left| \int_{S \times T} \left\{ f_A(x,y) - f_B(\sigma(x), \sigma(y)) \right\} \mathrm{d}x\, \mathrm{d}y \right| \leq \frac{22}{\sqrt{\log k_0}},$$

where $\sigma : [0,1] \to [0,1]$ ranges over all invertible measure-preserving maps.

Our method

- **Goal:** Horvitz–Thompson estimator (population version)

$$d_{m,n,\rho_A,\rho_B} := \rho_A^{-s} \cdot \mu_m - \rho_B^{-s} \cdot \nu_n$$

- Sample version (point estimator)

$$\widehat{D}_{m,n} := \widehat{\rho}_A^{-s} \cdot \widehat{U}_m - \widehat{\rho}_B^{-s} \cdot \widehat{V}_n$$

- Next steps:
  - i. estimate $\mathrm{Var}(\widehat{D}_{m,n})$ and studentize $\widehat{D}_{m,n}$ into $\widehat{T}_{m,n}$
  - ii. distribution approximation for $\widehat{T}_{m,n}$
- Let's analyze $\widehat{U}_m$

Our method

Analysis of $\widehat{U}_m$: take $R =$ Triangle as an example, assuming $\rho_A = 1$

$$\binom{m}{3}\widehat{U}_m = \underbrace{\sum_{i,j,k} W_{i,j}W_{j,k}W_{k,i}}_{:=U_m,\ \text{Randomness due to } X_i\text{'s}} \\ \underbrace{+\sum_{i,j}\Theta_{i,j}\eta_{i,j} + \sum_{i,j;k}\Theta_{i,j;k}\eta_{i,k}\eta_{j,k} + \sum_{i,j,k}\Theta_{i,j,k}\eta_{i,j}\eta_{j,k}\eta_{k,i}}_{:=(\widehat{U}_m - U_m),\ \text{Randomness due to observational error}} \qquad (2)$$

where:

- Recall $W_{i,j} = f(X_i, X_j)$, then U_m is a U-statistic
- $\eta_{i,j} = A_{i,j} - W_{i,j}^{(A)}$
- Θ's: coefficients, functions of X_i's

Our method

Decomposition of $\widehat{U}_m$:

$$\widehat{U}_m = U_m + (\widehat{U}_m - U_m)$$

where

- $U_m = U_m(X_1, \ldots, X_m)$: randomness in $W^{(A)}$, driven by random latent node positions
- $\widehat{U}_m - U_m$: observational errors in $A|W^{(A)}$, driven by Bernoulli edge realization
- "Non-degenerate" case: $\mathrm{Var}(U_m) \gg \mathrm{Var}(\widehat{U}_m - U_m)$, focus on U_m

Our method

- $U_m := \mathbb{E}[\widehat{U}_m | W^{(A)}]$ admits **Hoeffding's ANOVA decomposition**

$$U_m - \mu_m = \underbrace{\frac{r}{m} \sum_{i=1}^{n} g_{A;1}(X_i)}_{\text{Linear part, main term}} + \underbrace{\frac{r(r-1)}{m(m-1)} \sum_{1 \le i < j \le m} g_{A;2}(X_i, X_j)}_{\text{Quadratic part, correction}} + \text{(Remainder)} \quad (3)$$

  where all $g_{A;k}$'s are **uncorrelated**
- Other terms ($\widehat{\rho}_A, \widehat{V}_n, \widehat{\rho}_B$) can be analyzed similarly

Our method

- Recall:

$$\widehat{D}_{m,n} := \widehat{\rho}_A^{-s} \cdot \widehat{U}_m - \widehat{\rho}_B^{-s} \cdot \widehat{V}_n$$

- Variance estimation:

$$\widehat{S}^2_{m,n} := \frac{1}{m^2}\widehat{\alpha}_1^2(X_i) + \frac{1}{n^2}\widehat{\beta}_1^2(Y_j)$$

where

$$\widehat{\alpha}_1(X_i) := r\widehat{\rho}_A^{-s}\widehat{g}_{A;1}(X_i) - \underbrace{2s\widehat{\rho}_A^{-(s+1)}\widehat{U}_m\widehat{g}_{\rho_A;1}(X_i)}_{\text{Plug-in error}}$$

and β defined similarly

Our method

- Studentization:

$$\widehat{T}_{m,n} := \widehat{S}_{m,n}^{-1}\{\widehat{D}_{m,n} - (\rho_A^{-s} \cdot \mu_m - \rho_B^{-s} \cdot \nu_n)\}$$

- Distribution of $\widehat{T}_{m,n}$?

Our method

- Edgeworth expansion approximates $F_{\widehat{T}_{m,n}}(u)$ by

$$G_{m,n}(u) := \Phi(u) - \varphi(u) \cdot \left\{ Q_1 + Q_2(u^2+1) + I_0 \right\},$$

  where
  - Φ: CDF of $N(0,1)$
  - φ: PDF of $N(0,1)$
  - I_0, Q_1 and Q_2 are estimable coefficients :

$$I_0 \asymp Q_1 \asymp Q_2 \asymp (m \wedge n)^{1/2} \cdot (m^{-1} + n^{-1})$$

    (Lengthy details omitted, see paper ArXiv: 2208.07573)

If you're curious...

$$
\begin{aligned}
\mathcal{I}_0 &:= \sigma_{m,n}^{-1}(m^{-1}\alpha_0 - n^{-1}\beta_0), \\
Q_{m,n,\rho_A,\rho_B;1} &:= \frac{1}{2}\sigma_{m,n}^{-3}\Big\{ - m^{-2}\mathbb{E}[\alpha_4(X_1,X_2)\alpha_1(X_2)] - m^{-2}\mathbb{E}[\alpha_1(X_1)\alpha_3(X_1)] \\
&\quad + n^{-2}\mathbb{E}[\beta_1(Y_1)\beta_3(Y_1)] + n^{-2}\mathbb{E}[\beta_4(Y_1;Y_2)\beta_1(Y_2)]\Big\}, \\
Q_{m,n,\rho_A,\rho_B;2} &:= \sigma_{m,n}^{-3}\Big\{m^{-2}\big(\mathbb{E}[\alpha_1^3(X_1)/6 \\
&\quad +\alpha_1(X_1)\alpha_1(X_2)\alpha_2(X_1,X_2)]\big) - n^{-2}\big(\mathbb{E}[\beta_1^3(Y_1)/6+\beta_1(Y_1)\beta_1(Y_2)\beta_2(Y_1,Y_2)]\big)\Big\} \\
&\quad +\frac{1}{2}\sigma_{m,n}^{-5}\Big\{\big(- m^{-3}\xi_{\alpha;1}^2 - m^{-2}n^{-1}\xi_{\beta;1}^2\big)\cdot\mathbb{E}[\alpha_1(X_1)\alpha_3(X_1) + \alpha_4(X_1;X_2)\alpha_1(X_2)] \\
&\quad +\big(m^{-1}n^{-2}\xi_{\alpha;1}^2 + n^{-3}\xi_{\beta;1}^2\big)\cdot\mathbb{E}[\beta_1(Y_1)\beta_3(Y_1) + \beta_4(Y_1;Y_2)\beta_1(Y_2)]\Big\}.
\end{aligned}
$$

Our method

Inference procedures:

- Hypothesis testing:

$$H_0 : d_{m,n,\rho_A,\rho_B} = 0, \quad \text{vs} \quad H_a : d_{m,n,\rho_A,\rho_B} \neq 0. \tag{10}$$

The empirical p-value produced by our method is

$$\widehat{p}_{\text{val}} := 2 \cdot \min\left\{\widehat{G}_{m,n}(\widehat{T}^{(\text{obs})}_{m,n}), 1 - \widehat{G}_{m,n}(\widehat{T}^{(\text{obs})}_{m,n})\right\},$$

where we define the observed statistic by $\widehat{T}^{(\text{obs})}_{m,n} := \widehat{D}_{m,n}/\widehat{S}_{m,n} + \delta_T$. Given a significance level α,

- Cornish-Fisher CI:

$$\left(\widehat{D}_{m,n} - \left(\widehat{q}_{\widehat{T}_{m,n};1-\alpha/2} - \delta_T\right) \cdot \widehat{S}_{m,n},\ \ \widehat{D}_{m,n} - \left(\widehat{q}_{\widehat{T}_{m,n};\alpha/2} - \delta_T\right) \cdot \widehat{S}_{m,n}\right).$$

where $\widehat{q}_{\widehat{T}_{m,n};\alpha} := z_\alpha + \widehat{I}_0 + \widehat{Q}_1 + \widehat{Q}_2(z_\alpha^2 - 1)$ and z_α is $N(0,1)$ quantile

Higher-order accuracy (simplified version):

- Conditions:
  1. $\log(m \vee n)/(m \wedge n) \to 0$
  2. For acyclic R: $\rho_A m \to \infty$; for cyclic R: $\rho_A^{r/2} m \to \infty$;
     Similar conditions for B-index terms
- Define error bound:

$$M_A := \begin{cases} (\rho_A m)^{-1}, & \text{Acyclic } R \\ \rho_A^{-r/2} m^{-1}, & \text{Cyclic } R \end{cases} \tag{4}$$

- W.h.p., we have

$$\left\| F_{\widehat{T}_{m,n}}(u) - \widehat{G}_{m,n}(u) \right\|_\infty \lesssim (m \wedge n)(m^{-1} M_A + n^{-1} M_B), \tag{5}$$

  where $\widehat{G}_{m,n}(u)$: empirical Edgeworth expansion (EEE)
- Larger and more cyclic motifs requires denser networks

Theory

Power optimality:

- Under regularity conditions, our test achieves consistency:

$$\text{(Type-I error rate)} + \text{(Type-II error rate)} \to 0$$

  when

$$d_{m,n,\rho_A,\rho_B} \gg m^{-1/2} + n^{-1/2}$$

- No method is consistent when $d_{m,n,\rho_A,\rho_B} \lesssim m^{-1/2} + n^{-1/2}$

Theory

Our method simultaneously achieves:

- Higher-order accurate control of “risks”
- Rate-optimal “power”

It is not hard to achieve just one, but difficult to achieve both

Network hashing

Recall:

$$G_{m,n}(u) := \Phi(u) - \varphi(u) \cdot \left\{ Q_1 + Q_2(u^2+1) + I_0 \right\},$$

where

$$\mathcal{I}_0 := \sigma_{m,n}^{-1}(m^{-1}\alpha_0 - n^{-1}\beta_0),$$

$$Q_{m,n,\rho_A,\rho_B;1} := \frac{1}{2}\sigma_{m,n}^{-3}\Big\{ - m^{-2}\mathbb{E}[\alpha_4(X_1,X_2)\alpha_1(X_2)] - m^{-2}\mathbb{E}[\alpha_1(X_1)\alpha_3(X_1)] + n^{-2}\mathbb{E}[\beta_1(Y_1)\beta_3(Y_1)] + n^{-2}\mathbb{E}[\beta_4(Y_1;Y_2)\beta_1(Y_2)]\Big\},$$

$$Q_{m,n,\rho_A,\rho_B;2} := \sigma_{m,n}^{-3}\Big\{m^{-2}\big(\mathbb{E}[\alpha_1^3(X_1)/6 + \alpha_1(X_1)\alpha_1(X_2)\alpha_2(X_1,X_2)]\big) - n^{-2}\big(\mathbb{E}[\beta_1^3(Y_1)/6 + \beta_1(Y_1)\beta_1(Y_2)\beta_2(Y_1,Y_2)]\big)\Big\} + \frac{1}{2}\sigma_{m,n}^{-5}\Big\{\big(- m^{-3}\xi_{\alpha;1}^2 - m^{-2}n^{-1}\xi_{\beta;1}^2\big) \cdot \mathbb{E}[\alpha_1(X_1)\alpha_3(X_1) + \alpha_4(X_1;X_2)\alpha_1(X_2)] + \big(m^{-1}n^{-2}\xi_{\alpha;1}^2 + n^{-3}\xi_{\beta;1}^2\big) \cdot \mathbb{E}[\beta_1(Y_1)\beta_3(Y_1) + \beta_4(Y_1;Y_2)\beta_1(Y_2)]\Big\}.$$

No cross terms (e.g., sth. like $\mathbb{E}[\gamma(X_1,Y_1)]$)!

Network hashing

Suppose you "Google" a keyword network A in a huge network database $B_1, \ldots, B_K$:

1. Network hashing: each network $\rightarrow$ a few summary statistics
2. Fast query: compare A to each B_k using only summary statistics!

Benefits:

- Easy indexing and maintenance for network database curators, memory efficient
- Enhanced privacy protection
- Query is lightning fast, much faster than existing methods
- Our method can serve as a screening stage for other methods (to test exact equality of models, under additional assumptions)

Simulations

Benchmarks:

- Normal approximation
- Node sub-sampling (variant of *Bhattacharyya & Bickel, (2015)*)
- Node re-sampling (variant of *Green & Shalizi, (2022)*)

Motifs:

- Triangle (cyclic)
- V-shape (acyclic)

Performance evaluation criteria:

- Simulation 1: Accuracy: K-S distance $\|\widehat{F}_{\widehat{T}_{m,n}}(u) - F_{\widehat{T}_{m,n}}(u)\|_\infty$
- Simulation 2: Accuracy of CI level control
- Simulation 3: database query accuracy
- All simulations: Time cost

Simulation 1

A, B from two different smooth graphons

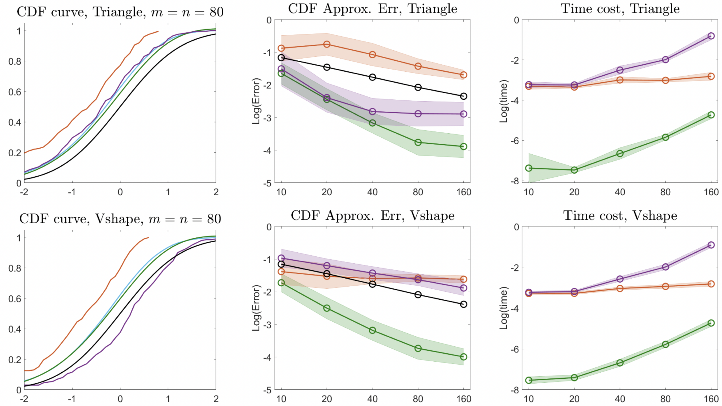
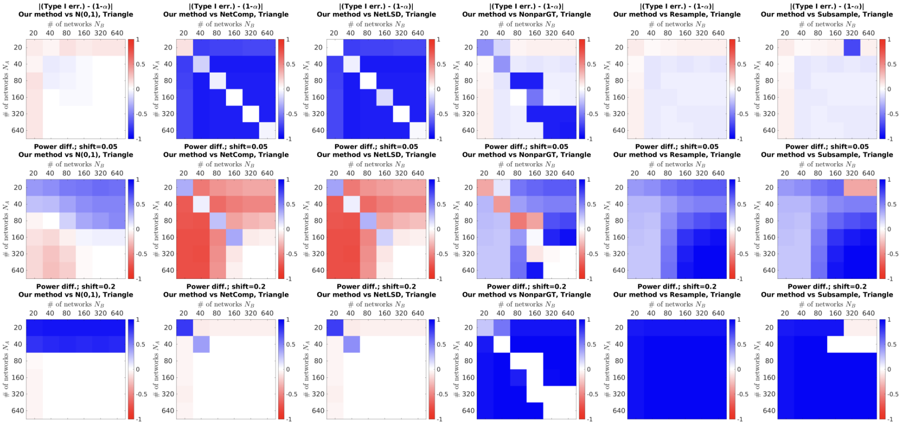


Figure: Green: our method; black: $N(0,1)$; orange: node sub-sampling; violet: node re-sampling; cyan (left panel): true distribution Monte Carlo emulation

Simulation 2



Simulation 3

Database: 10 different models, each generate 100 networks;
Keyword networks: one from a database model, one from outside

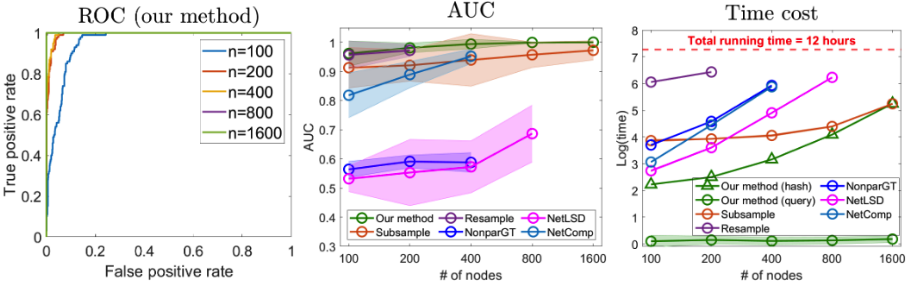


Figure: Comparison of query accuracy and speed. All networks are size n. Incomplete curves indicate benchmark went timeout.

Data examples

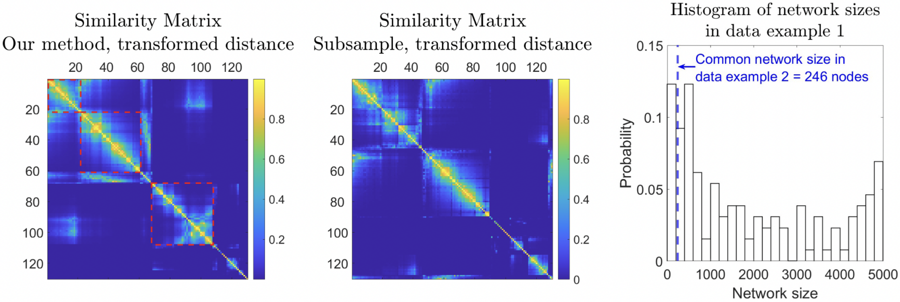


Figure: Data set 1: Google+ ego-networks

Data examples

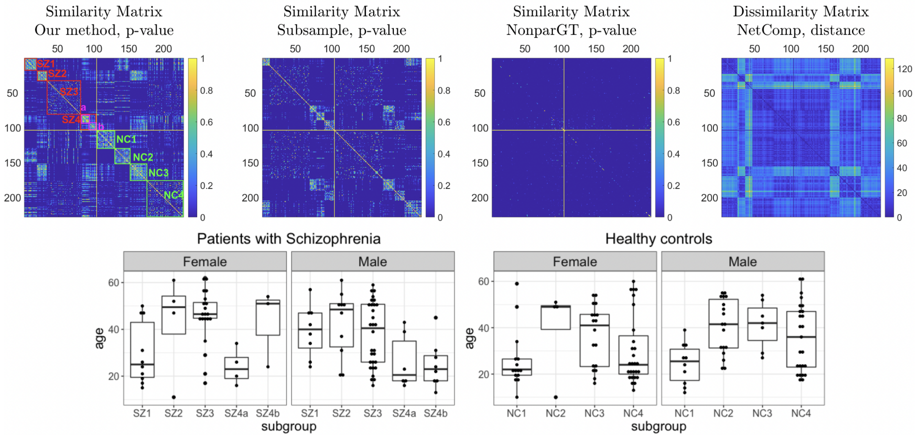


Figure: Data set 2: Schizophrenia brain image networks

Data examples

Table 2: Time cost comparison table. Unit is second. Timeout is 12 hours = 43200 seconds.

	Our method (hash)	Our method (test)	Subsample	Resample
Data example 1	116.39	18.81	10884.62	(Timeout)
Data example 2	3.60	64.36	2488.21	(Timeout)
	NonparGT	NetLSD	NetComp	
Data example 1	(Execution error)	(Timeout)	(Timeout)	
Data example 2	4327.09	(Execution error)	4304.51	

Figure: Time cost comparison

Extensions

What if I do have multiple observations? $A^{(1)}, \ldots, A^{(N_A)}; B^{(1)}, \ldots, B^{(N_B)}$, $N_A, N_B > 1$.

- If nodes are matched (same set of nodes):
  1. Pool (average over) all adjacency matrices in one group

$$A^{\text{pool}} := N_A^{-1} \sum_{\ell=1}^{N_A} A^{(\ell)}$$

  2. Apply our algorithm to pooled A, B with **no formula change**
- If nodes are independently generated:
  1. Compute $\widehat{U}_m^{(\ell)}$ for each network ℓ in Group A, then average:

$$\widehat{U}_m^{\text{pool}} := N_A^{-1} \sum_{\ell=1}^{N_A} \widehat{U}_m^{(\ell)}$$

     Do the same for $\widehat{\rho}_A$, $\widehat{V}_n$, $\widehat{\rho}_B$ and all emp. Edgeworth expansion coef.
  2. **Formula changes:** (assuming equal network sizes)
     - A new variance estimator
     - Edgeworth expansion formula: $m \to mN_A$, $n \to nN_B$

Extensions

Degenerate U-statistics:

- Reducing the U-statistic reinstates normality *(Weber, 1981; Chen & Kato, 2019; Shao et al, 2023)*
- **Example:** $Z_1, \ldots, Z_n \overset{\text{i.i.d.}}{\sim} N(0,1)$, then:

$$\sum_{1 \le i < j \le n} Z_i Z_j \overset{d}{\not\to} \text{Normal}$$

$$\text{However, } Z_1 Z_2 + Z_3 Z_4 + \cdots \overset{d}{\to} \text{Normal}$$

- Reduced statistic also computes faster, but inflates variance
- We proposed a novel test statistic, automatically adaptive to potential degeneracy (see paper)

Extensions

FDR control for query:

- Recall set up: test A vs each of $B_1, \ldots, B_K$
- Vector of test statistics: $\widehat{T} \approx \iota + \gamma W + \mathscr{K}$, where
  - $W \sim N(0,1)$
  - $\mathscr{K} \sim N(0, \Psi)$
- $\mathscr{K}$ has "weak dependence" *(Fan et al, 2012)*, i.e., $\Psi \approx \text{diag}(\Psi)$
- Conditional on W, $\widehat{T}$ produces nearly independent p-values

Theorem 6. *Suppose the conditions of Theorem 2 hold for all networks (keyword and database entries). Additionally, assume*

$$\lim_{K,m,n_{\min}\to\infty} \frac{\log(mK)}{m} + \frac{\log(n_{\min}K)}{n_{\min}} + \frac{\log^{1/2}(mK)}{\rho_A m} + \max_{j\in[1:K]} \frac{\log^{1/2}(n_j K)}{\rho_{B_j} n_j} = 0,$$

where $n_{\min} := \min_{k\in[1:K]} n_k$. *Then, we have*

$$\text{FDP}(t) - \frac{\sum_{k\in\{\text{true nulls}\}} \left[\Phi\big(a_k(z_{t/2}+\eta_k)\big) + \Phi\big(a_k(z_{t/2}-\eta_k)\big)\right]}{\sum_{k\in[1:K]} \mathbb{1}\big(\widehat{p}_k \le t\big)} \overset{\text{a.s.}}{\to} 0, \tag{16}$$

as $K, m, n_{\min} \to \infty$, *where* $a_k := \left(1-\gamma_k^2\right)^{-1/2}$, $z_t := \Phi^{-1}(t)$, *and* $\eta_k := \gamma_k W$ *with* $W \sim N(0,1)$.

Thank you!

Questions?