Challenges and Recent Advances in Network Inference

Jiaming Xu

The Fuqua School of Business
Duke University

INFORMS, APS Tutorial
October 26, 2025

Statistical inference on graphs

- Detecting or estimating hidden structures in large network data

$$\underbrace{X}_{\text{Hidden structure}} \mapsto \underbrace{G}_{\text{Network}} \mapsto \underbrace{\hat{X}}_{\text{estimate}}$$

Statistical inference on graphs

- Detecting or estimating hidden structures in large network data

$$\underbrace{X}_{\text{Hidden structure}} \mapsto \underbrace{G}_{\text{Network}} \mapsto \underbrace{\hat{X}}_{\text{estimate}}$$

- Key challenges: Understanding the fundamental limits:

Statistical inference on graphs

- Detecting or estimating hidden structures in large network data

$$\underbrace{X}_{\text{Hidden structure}} \mapsto \underbrace{G}_{\text{Network}} \mapsto \underbrace{\hat{X}}_{\text{estimate}}$$

- Key challenges: Understanding the fundamental limits:
  - Q1 Characterize statistical (information-theoretic) limit: What is possible/impossible?

Statistical inference on graphs

- Detecting or estimating hidden structures in large network data

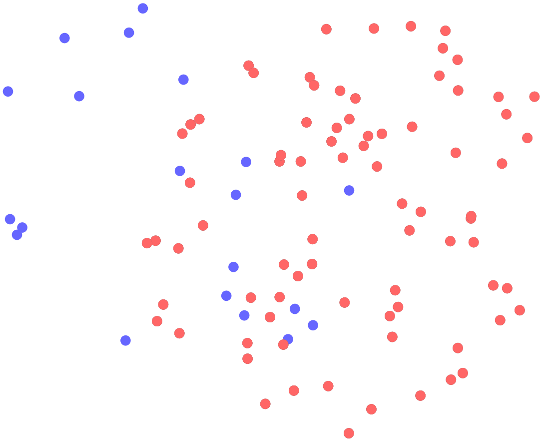
$$\underbrace{X}_{\text{Hidden structure}} \mapsto \underbrace{G}_{\text{Network}} \mapsto \underbrace{\hat{X}}_{\text{estimate}}$$

- Key challenges: Understanding the fundamental limits:
  - Q1 Characterize statistical (information-theoretic) limit: What is possible/impossible?
  - Q2 Can statistical limits be attained computationally efficiently, e.g., in polynomial time? If yes, how? If not, why?

Planted clique – graph view

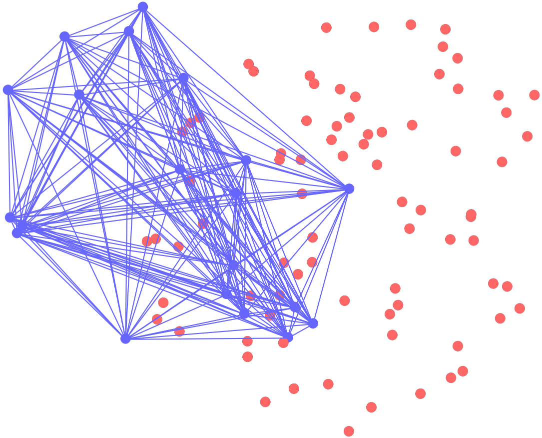
Planted clique – graph view

1. A set C of k vertices is chosen to form a clique



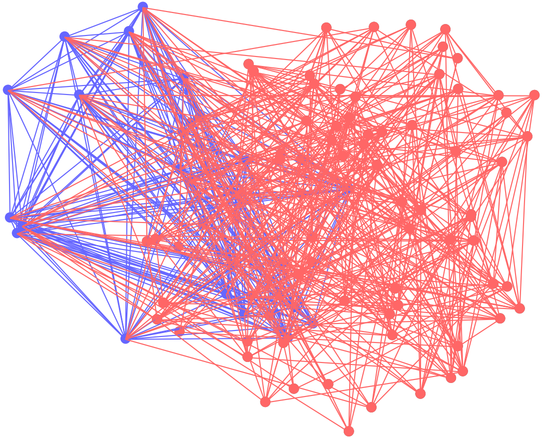
Planted clique – graph view

1. A set C of k vertices is chosen to form a clique



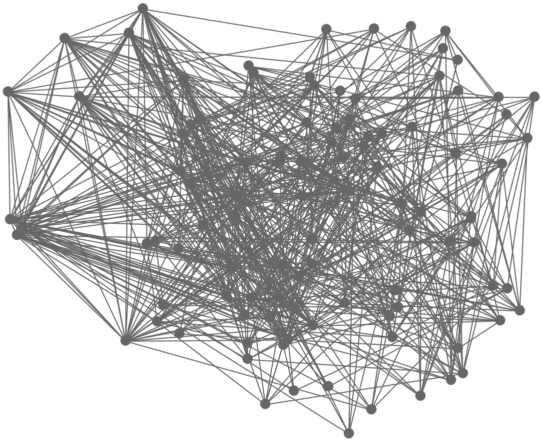
Planted clique – graph view

1. A set C of k vertices is chosen to form a clique
2. For every other pair of vertices, add an edge w.p. $\frac{1}{2}$

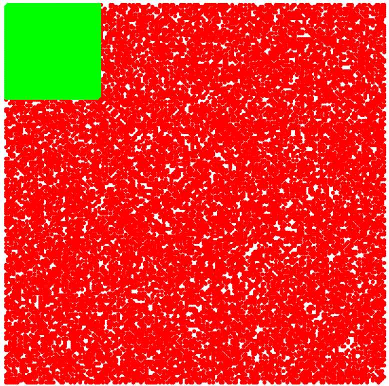


Planted clique – graph view

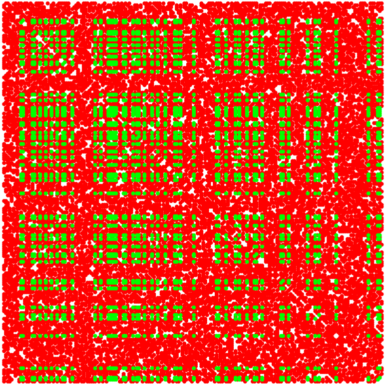
1. A set C of k vertices is chosen to form a clique
2. For every other pair of vertices, add an edge w.p. $\frac{1}{2}$



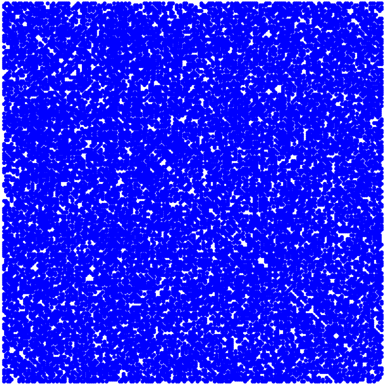
Planted clique – adjacency matrix view



Planted clique – adjacency matrix view



Planted clique – adjacency matrix view



Community detection in networks

- Networks with community structures arise in many applications

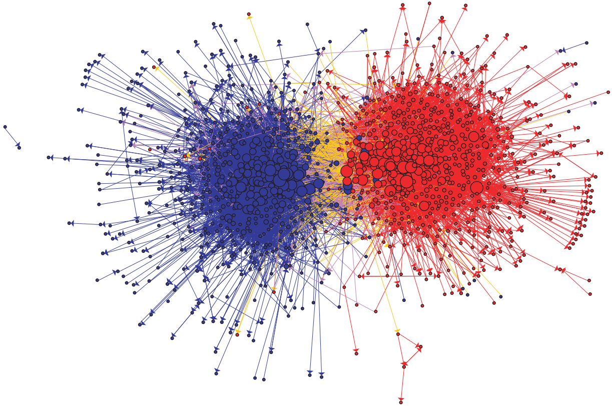


Figure: Political blogosphere and the 2004 U.S. election [Adamic-Glance '05]

Community detection in networks

- Networks with community structures arise in many applications
- Task: Discover underlying communities based on the network topology

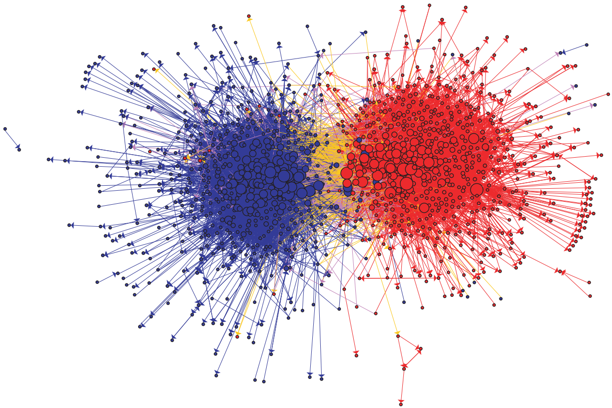
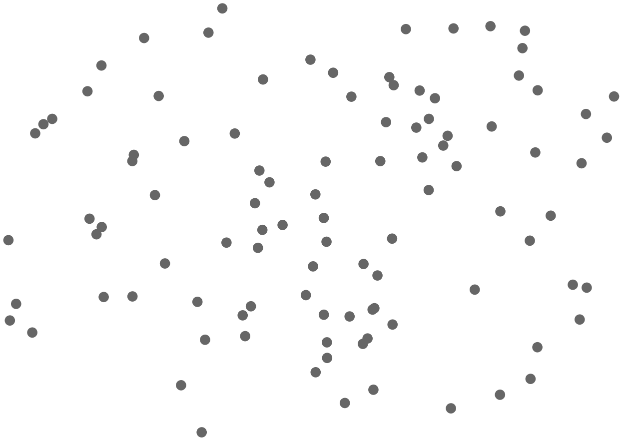


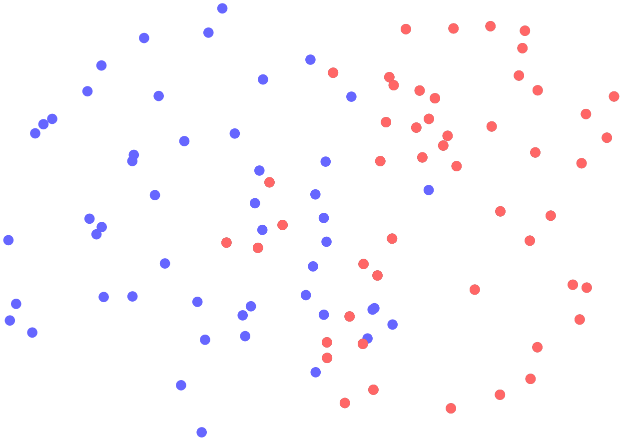
Figure: Political blogosphere and the 2004 U.S. election [Adamic-Glance '05]

Stochastic block model – graph view



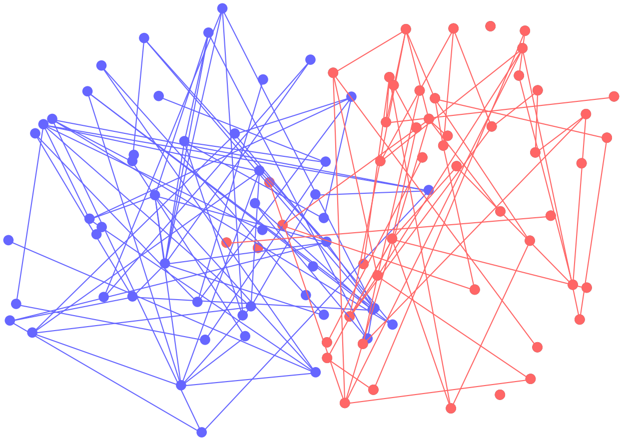
Stochastic block model – graph view

1. n nodes are assigned to 2 communities uniformly at random



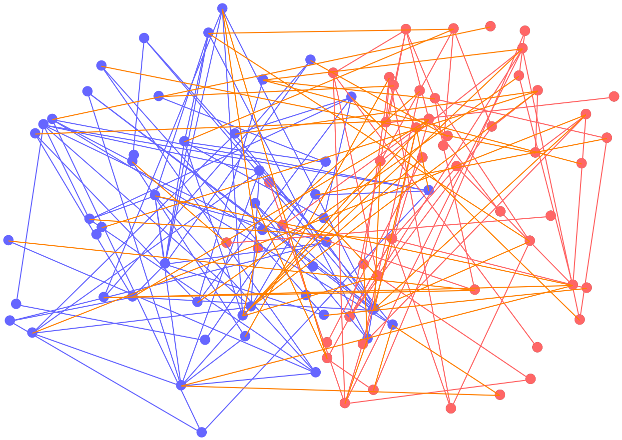
Stochastic block model – graph view

1. n nodes are assigned to 2 communities uniformly at random
2. For every pair of nodes in same community, add an edge w.p. p



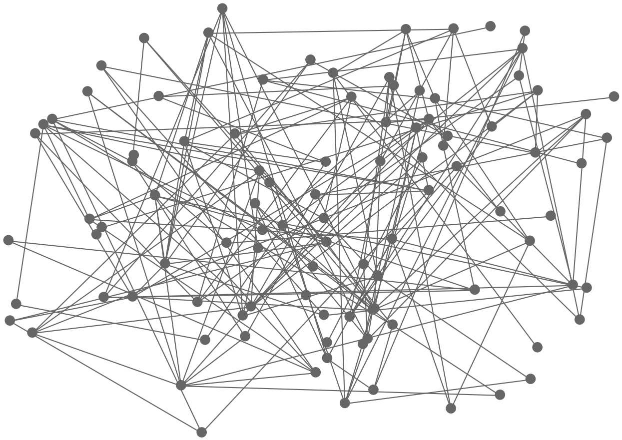
Stochastic block model – graph view

1. n nodes are assigned to 2 communities uniformly at random
2. For every pair of nodes in same community, add an edge w.p. p
3. For every pair of nodes in diff. community, add an edge w.p. q

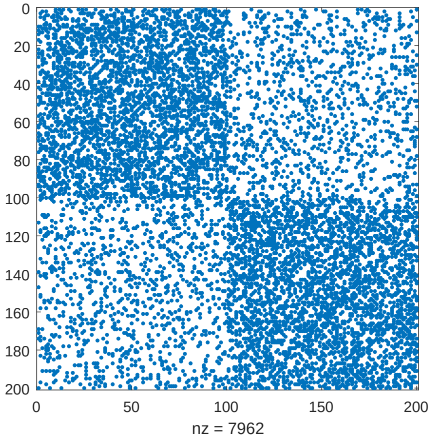


Stochastic block model – graph view

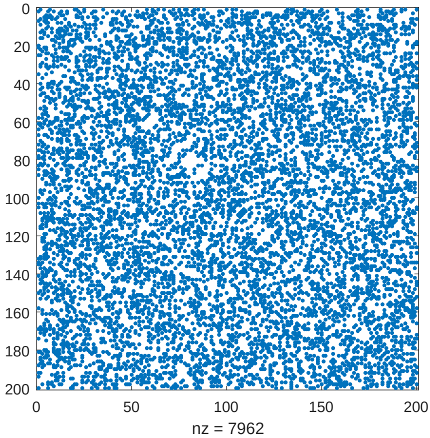
1. n nodes are assigned to 2 communities uniformly at random
2. For every pair of nodes in same community, add an edge w.p. p
3. For every pair of nodes in diff. community, add an edge w.p. q



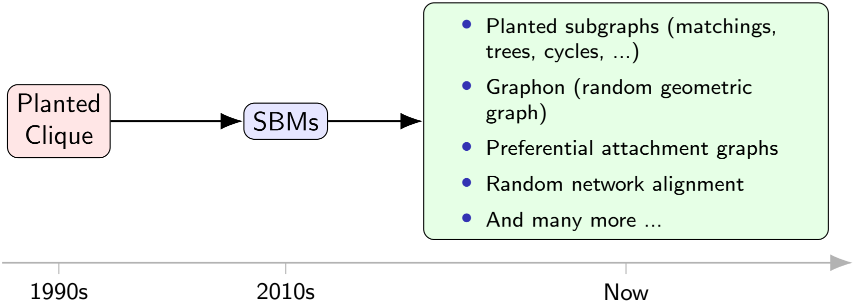
Stochastic block model – adjacency matrix view



Stochastic block model – adjacency matrix view



A flurry of network inference problems



Driven by both theoretical interests and practical applications

Significant methodological advances

Applied Probability

- Local weak convergence
- Random matrix & spectral methods

Optimization

- Relaxations (LP, SDP)
- Dual certificates & polyhedral combinatorics

Statistical Physics

- Belief propagation & message passing
- Interpolation method

Significant methodological advances

Applied Probability

- Local weak convergence
- Random matrix & spectral methods

Optimization

- Relaxations (LP, SDP)
- Dual certificates & polyhedral combinatorics

Statistical Physics

- Belief propagation & message passing
- Interpolation method

- These methods have led to sharp characterizations of information-theoretic and algorithmic phase transition thresholds.

Significant methodological advances

Applied Probability

- Local weak convergence
- Random matrix & spectral methods

Optimization

- Relaxations (LP, SDP)
- Dual certificates & polyhedral combinatorics

Statistical Physics

- Belief propagation & message passing
- Interpolation method

- These methods have led to sharp characterizations of information-theoretic and algorithmic phase transition thresholds.
- The proofs, however, often require substantial mathematical ingenuity.

Significant methodological advances

Applied Probability

- Local weak convergence
- Random matrix & spectral methods

Optimization

- Relaxations (LP, SDP)
- Dual certificates & polyhedral combinatorics

Statistical Physics

- Belief propagation & message passing
- Interpolation method

- These methods have led to sharp characterizations of information-theoretic and algorithmic phase transition thresholds.
- The proofs, however, often require substantial mathematical ingenuity.
- But what if I am not ingenious? Is there a simple, principled approach to try?

Significant methodological advances

Applied Probability

- Local weak convergence
- Random matrix & spectral methods

Optimization

- Relaxations (LP, SDP)
- Dual certificates & polyhedral combinatorics

Statistical Physics

- Belief propagation & message passing
- Interpolation method

- These methods have led to sharp characterizations of information-theoretic and algorithmic phase transition thresholds.
- The proofs, however, often require substantial mathematical ingenuity.
- But what if I am not ingenious? Is there a simple, principled approach to try?

This tutorial: Low-degree polynomial method
Analog of drift method in stochastic networks

Outline of tutorial

- Introduction to low-degree polynomial method
- Three prototypical examples
  - Planted clique
  - Stochastic block model
  - Random network alignment
- Concluding remarks

Polynomials on graph

- Given a graph G represented by adjacency vector $A = (A_{ij})_{1 \le i < j \le n}$
- A multivariate polynomial $f : \{0,1\}^{\binom{n}{2}} \to \mathbb{R}$

Polynomials on graph

- Given a graph G represented by adjacency vector $A = (A_{ij})_{1 \leq i < j \leq n}$
- A multivariate polynomial $f : \{0,1\}^{\binom{n}{2}} \to \mathbb{R}$

Example

- Edge count: $\sum_{i<j} A_{ij}$

Polynomials on graph

- Given a graph G represented by adjacency vector $A = (A_{ij})_{1 \leq i < j \leq n}$
- A multivariate polynomial $f : \{0,1\}^{\binom{n}{2}} \to \mathbb{R}$

Example

- Edge count: $\sum_{i<j} A_{ij}$
- Triangle count: $\sum_{i<j<k} A_{ij} A_{jk} A_{ik}$

Polynomials on graph

- Given a graph G represented by adjacency vector $A = (A_{ij})_{1 \le i < j \le n}$
- A multivariate polynomial $f : \{0,1\}^{\binom{n}{2}} \to \mathbb{R}$

Example

- Edge count: $\sum_{i<j} A_{ij}$
- Triangle count: $\sum_{i<j<k} A_{ij} A_{jk} A_{ik}$
- Subgraph-H count: $\sum_{S \cong H} \prod_{(i,j) \in S} A_{ij}$

Polynomials on graph

- Given a graph G represented by adjacency vector $A = (A_{ij})_{1 \le i < j \le n}$
- A multivariate polynomial $f : \{0,1\}^{\binom{n}{2}} \to \mathbb{R}$

Example

- Edge count: $\sum_{i<j} A_{ij}$
- Triangle count: $\sum_{i<j<k} A_{ij} A_{jk} A_{ik}$
- Subgraph-H count: $\sum_{S \cong H} \prod_{(i,j) \in S} A_{ij}$
- # of closed walks: $\mathrm{Tr}(A^D) = \sum_{i_1, i_2, \ldots, i_D} A_{i_1 i_2} A_{i_2 i_3} \cdots A_{i_D i_1}$

Polynomials on graph

- Given a graph G represented by adjacency vector $A = (A_{ij})_{1 \le i < j \le n}$
- A multivariate polynomial $f : \{0,1\}^{\binom{n}{2}} \to \mathbb{R}$

Example

- Edge count: $\sum_{i<j} A_{ij}$
- Triangle count: $\sum_{i<j<k} A_{ij} A_{jk} A_{ik}$
- Subgraph-H count: $\sum_{S \cong H} \prod_{(i,j) \in S} A_{ij}$
- # of closed walks: $\mathsf{Tr}(A^D) = \sum_{i_1, i_2, \ldots, i_D} A_{i_1 i_2} A_{i_2 i_3} \cdots A_{i_D i_1}$
- Message passing: $m_{j \to i} = h(\{m_{k \to j} : k \sim j, k \neq i\})$

Polynomials on graph

- Given a graph G represented by adjacency vector $A = (A_{ij})_{1 \leq i < j \leq n}$
- A multivariate polynomial $f : \{0,1\}^{\binom{n}{2}} \to \mathbb{R}$

Example

- Edge count: $\sum_{i<j} A_{ij}$
- Triangle count: $\sum_{i<j<k} A_{ij} A_{jk} A_{ik}$
- Subgraph-H count: $\sum_{S \cong H} \prod_{(i,j) \in S} A_{ij}$
- # of closed walks: $\mathsf{Tr}(A^D) = \sum_{i_1, i_2, \ldots, i_D} A_{i_1 i_2} A_{i_2 i_3} \cdots A_{i_D i_1}$
- Message passing: $m_{j \to i} = h(\{m_{k \to j} : k \sim j, k \neq i\})$
- Local algorithms: f depends on local neighborhood

Polynomial basis [Janson '90, '94]

- Consider the space $\mathcal{F}$ of real-valued functions on $\{0,1\}^{\binom{n}{2}}$ endowed with inner-product

$$\langle f, g \rangle \triangleq \mathbb{E}_{A_{ij} \overset{\text{iid}}{\sim} \mathsf{Bern}(q)}[f(A)g(A)]$$

Polynomial basis [Janson '90, '94]

- Consider the space $\mathcal{F}$ of real-valued functions on $\{0,1\}^{\binom{n}{2}}$ endowed with inner-product
$$\langle f, g\rangle \triangleq \mathbb{E}_{A_{ij} \overset{\text{iid}}{\sim} \mathsf{Bern}(q)}[f(A)g(A)]$$
- Fact: The orthogonal polynomial basis $\{\Phi_S : S \subset \binom{[n]}{2}\}$ spans entire $\mathcal{F}$

$$\Phi_S = \prod_{(i,j)\in S} \bar{A}_{ij}, \quad \bar{A}_{ij} = \frac{A_{ij} - q}{\sqrt{q(1-q)}}$$

Polynomial basis [Janson '90, '94]

- Consider the space $\mathcal{F}$ of real-valued functions on $\{0,1\}^{\binom{n}{2}}$ endowed with inner-product

$$\langle f, g \rangle \triangleq \mathbb{E}_{A_{ij} \overset{\text{iid}}{\sim} \mathsf{Bern}(q)}[f(A)g(A)]$$

- Fact: The orthogonal polynomial basis $\{\Phi_S : S \subset \binom{[n]}{2}\}$ spans entire $\mathcal{F}$

$$\Phi_S = \prod_{(i,j) \in S} \bar{A}_{ij}, \quad \bar{A}_{ij} = \frac{A_{ij} - q}{\sqrt{q(1-q)}}$$

- Quick check
  - Orthonormality: $\langle \Phi_S, \Phi_T \rangle = \mathbf{1}\{S = T\}$
  - Completeness: $\dim(\{\Phi_S : S \subset \binom{[n]}{2}\}) = \dim(\mathcal{F}) = 2^n$

Polynomial basis [Janson '90, '94]

- Consider the space $\mathcal{F}$ of real-valued functions on $\{0,1\}^{\binom{n}{2}}$ endowed with inner-product
$$\langle f, g\rangle \triangleq \mathbb{E}_{A_{ij} \overset{\text{iid}}{\sim} \mathsf{Bern}(q)}[f(A)g(A)]$$
- Fact: The orthogonal polynomial basis $\{\Phi_S : S \subset \binom{[n]}{2}\}$ spans entire $\mathcal{F}$
$$\Phi_S = \prod_{(i,j)\in S} \bar{A}_{ij}, \quad \bar{A}_{ij} = \frac{A_{ij} - q}{\sqrt{q(1-q)}}$$
- Quick check
  - Orthonormality: $\langle \Phi_S, \Phi_T\rangle = \mathbf{1}\{S = T\}$
  - Completeness: $\dim(\{\Phi_S : S \subset \binom{[n]}{2}\}) = \dim(\mathcal{F}) = 2^n$

Question

How to design polynomial-based estimator?

Polynomial approximation of likelihood ratio

$$\mathbb{H}_0 : A \sim \mathrm{Bern}(q)^{\otimes \binom{n}{2}} \triangleq Q \qquad \text{(Null model)}$$

$$\mathbb{H}_1 : A \sim P \qquad \text{(Planted model)}$$

Polynomial approximation of likelihood ratio

$$\mathbb{H}_0 : A \sim \mathrm{Bern}(q)^{\otimes \binom{n}{2}} \triangleq Q \qquad \text{(Null model)}$$
$$\mathbb{H}_1 : A \sim P \qquad \text{(Planted model)}$$

- By Neyman-Pearson Lemma, likelihood ratio test is optimal:

$$L(A) \triangleq \frac{P(A)}{Q(A)}$$

Polynomial approximation of likelihood ratio

$$\mathbb{H}_0 : A \sim \mathrm{Bern}(q)^{\otimes \binom{n}{2}} \triangleq Q \qquad \text{(Null model)}$$

$$\mathbb{H}_1 : A \sim P \qquad \text{(Planted model)}$$

- By Neyman-Pearson Lemma, likelihood ratio test is optimal:

$$L(A) \triangleq \frac{P(A)}{Q(A)}$$

- However, for many planted problems, P is a mixture over exponentially many components $\Rightarrow L$ is computationally hard to evaluate

Polynomial approximation of likelihood ratio

$$\mathbb{H}_0 : A \sim \text{Bern}(q)^{\otimes\binom{n}{2}} \triangleq Q \qquad \text{(Null model)}$$
$$\mathbb{H}_1 : A \sim P \qquad \text{(Planted model)}$$

- By Neyman-Pearson Lemma, likelihood ratio test is optimal:

$$L(A) \triangleq \frac{P(A)}{Q(A)}$$

- However, for many planted problems, P is a mixture over exponentially many components $\Rightarrow$ L is computationally hard to evaluate
- Instead, look for low-degree polynomial maximizing signal-to-noise ratio:

$$\max_{f:\deg(f)\leq D} \left(\frac{\mathbb{E}_P[f]}{\sqrt{\mathbb{E}_Q[f^2]}} = \frac{\langle L, f\rangle}{\sqrt{\langle f, f\rangle}} \right)$$

Polynomial approximation of likelihood ratio

$$\mathbb{H}_0 : A \sim \mathrm{Bern}(q)^{\otimes \binom{n}{2}} \triangleq Q \qquad \text{(Null model)}$$

$$\mathbb{H}_1 : A \sim P \qquad \text{(Planted model)}$$

- By Neyman-Pearson Lemma, likelihood ratio test is optimal:

$$L(A) \triangleq \frac{P(A)}{Q(A)}$$

- However, for many planted problems, P is a mixture over exponentially many components $\Rightarrow$ L is computationally hard to evaluate
- Instead, look for low-degree polynomial maximizing signal-to-noise ratio:

$$\max_{f:\deg(f)\le D} \left(\frac{\mathbb{E}_P[f]}{\sqrt{\mathbb{E}_Q[f^2]}} = \frac{\langle L, f\rangle}{\sqrt{\langle f, f\rangle}} \right)$$

- By Cauchy-Schwartz, optimum is $\|L_{\le D}\|$ and achieved by projection of L:

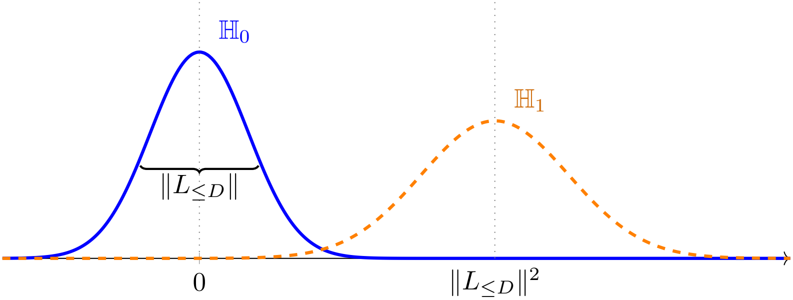
$$L_{\le D} = \underbrace{\sum_{S:|S|\le D} \langle L, \Phi_S\rangle \, \Phi_S}_{\text{weighted signed subgraph count}} \,, \text{ where } \Phi_S = \prod_{(i,j)\in S} \frac{A_{ij} - q}{\sqrt{q(1-q)}}$$

Low-degree polynomial prediction

Conjecture (Hopkins '18, informal)

For "sufficiently nice" planted problems,

- *If* $\|L_{\leq D}\| \to \infty$ *for* $D = O(\log n)$*, there exists degree-*D *polynomial succeeds in detecting or estimating the hidden structure*
- *If* $\|L_{\leq D}\| = O(1)$ *for* $D = O(\log n)$*, all polynomial-time algorithms fail in detection and estimation*



Low-degree polynomial prediction

Conjecture (Hopkins '18, informal)

For "sufficiently nice" planted problems,

- *If $\|L_{\leq D}\| \to \infty$ for $D = O(\log n)$, there exists degree-D polynomial succeeds in detecting or estimating the hidden structure*
- *If $\|L_{\leq D}\| = O(1)$ for $D = O(\log n)$, all polynomial-time algorithms fail in detection and estimation*

Remark

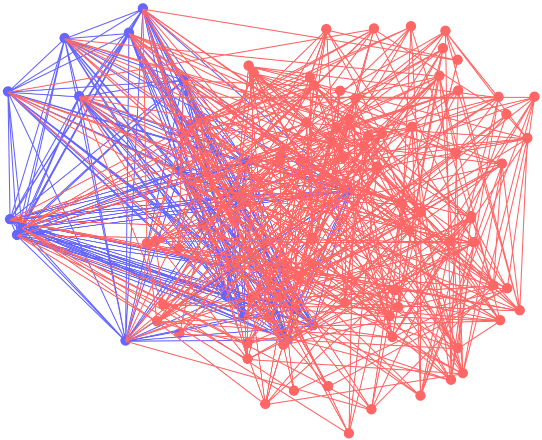
- Remarkably, this prediction aligns with many proven algorithmic upper and lower bounds for a wide class of planted problems
- Need $O(\log n)$ degree to cover spectral method, and many $O(\log n)$-polynomials can be computed in poly-time
- Significant progress on proving low-degree polynomial lower bounds [Wein '25]
- Focus of this tutorial: low-degree polynomial as an algorithmic tool

Outline of tutorial

- Introduction to low-degree polynomial method
- Three prototypical examples
  - Planted clique
  - Stochastic block model
  - Random network alignment
- Concluding remarks

Planted clique problem

1. A set C of k vertices is chosen to form a clique
2. For every other pair of vertices, add an edge w.p. $\frac{1}{2}$



Planted clique problem: testing

$$\mathbb{H}_0 : A \sim \mathcal{G}(n, 1/2) \qquad \text{(Null model)}$$

$$\mathbb{H}_1 : A \sim \mathcal{G}(n, 1/2, k) \qquad \text{(Planted model)}$$

Planted clique problem: testing

$$\mathbb{H}_0 : A \sim \mathcal{G}(n, 1/2) \qquad \text{(Null model)}$$
$$\mathbb{H}_1 : A \sim \mathcal{G}(n, 1/2, k) \qquad \text{(Planted model)}$$

- Orthogonal polynomial basis: $\Phi_S = \prod_{(i,j)\in S}(2A_{ij} - 1)$

Planted clique problem: testing

$$\mathbb{H}_0 : A \sim \mathcal{G}(n, 1/2) \qquad \text{(Null model)}$$
$$\mathbb{H}_1 : A \sim \mathcal{G}(n, 1/2, k) \qquad \text{(Planted model)}$$

- Orthogonal polynomial basis: $\Phi_S = \prod_{(i,j)\in S}(2A_{ij} - 1)$
- Likelihood-ratio projection:

$$\langle L, \Phi_S \rangle = \mathbb{E}_P[\Phi_S] = \mathbb{E}_C \mathbb{E}_{A|C}[\Phi_S] = \mathbb{E}_C[\mathbf{1}\{V(S) \subset C\}] \approx \left(\frac{k}{n}\right)^{|V(S)|}$$

where C is the vertex set of hidden clique and $V(S)$ is the vertex set of S

Planted clique problem: testing

$$\mathbb{H}_0 : A \sim \mathcal{G}(n, 1/2) \qquad \text{(Null model)}$$
$$\mathbb{H}_1 : A \sim \mathcal{G}(n, 1/2, k) \qquad \text{(Planted model)}$$

- Orthogonal polynomial basis: $\Phi_S = \prod_{(i,j)\in S}(2A_{ij} - 1)$
- Likelihood-ratio projection:

$$\langle L, \Phi_S \rangle = \mathbb{E}_P[\Phi_S] = \mathbb{E}_C \mathbb{E}_{A|C}[\Phi_S] = \mathbb{E}_C[\mathbf{1}\{V(S) \subset C\}] \approx \left(\frac{k}{n}\right)^{|V(S)|}$$

where C is the vertex set of hidden clique and $V(S)$ is the vertex set of S

- So, we get

$$\|L_{\leq D}\|^2 = \sum_{S:|S|\leq D} \langle L, \Phi_S \rangle^2 \approx \sum_{S:|S|\leq D} \left(\frac{k}{n}\right)^{2|V(S)|}$$

Planted clique problem: testing

$$\|L_{\leq D}\|^2 = \sum_{S:|S|\leq D} \langle L, \Phi_S \rangle^2 \approx \sum_{S:|S|\leq D} \left(\frac{k}{n}\right)^{2|V(S)|}$$

Planted clique problem: testing

$$\|L_{\leq D}\|^2 = \sum_{S:|S|\leq D} \langle L, \Phi_S \rangle^2 \approx \sum_{S:|S|\leq D} \left(\frac{k}{n}\right)^{2|V(S)|}$$

- If $D = 1$, $\|L_{\leq D}\|^2 \approx n^2 \left(\frac{k}{n}\right)^4 \gg 1$, if $k^2 \gg n$ $\Rightarrow$ counting edges succeeds

Planted clique problem: testing

$$\|L_{\le D}\|^2 = \sum_{S:|S|\le D} \langle L, \Phi_S \rangle^2 \approx \sum_{S:|S|\le D} \left(\frac{k}{n}\right)^{2|V(S)|}$$

- If $D = 1$, $\|L_{\le D}\|^2 \approx n^2 \left(\frac{k}{n}\right)^4 \gg 1$, if $k^2 \gg n \Rightarrow$ counting edges succeeds
- If restricting S to be D-cycles with $D \to \infty$,

$$\|L_{D-\text{cycle}}\| \approx n^D \left(\frac{k}{n}\right)^{2D} \gg 1,$$

if $k^2 > n$ (limit of spectral method [Alon-Krivelevich-Sudakov '98])

Planted clique problem: testing

$$\|L_{\le D}\|^2 = \sum_{S:|S|\le D} \langle L, \Phi_S\rangle^2 \approx \sum_{S:|S|\le D} \left(\frac{k}{n}\right)^{2|V(S)|}$$

- If $D = 1$, $\|L_{\le D}\|^2 \approx n^2 \left(\frac{k}{n}\right)^4 \gg 1$, if $k^2 \gg n \Rightarrow$ counting edges succeeds
- If restricting S to be D-cycles with $D \to \infty$,

$$\|L_{D-\text{cycle}}\| \approx n^D \left(\frac{k}{n}\right)^{2D} \gg 1,$$

  if $k^2 > n$ (limit of spectral method [Alon-Krivelevich-Sudakov '98])
- If restricting S to be D-trees with $D \to \infty$,

$$\|L_{D-\text{tree}}\| \approx \binom{n}{D+1}(D+1)^{D-1}\left(\frac{k}{n}\right)^{2(D+1)},$$

  if $k^2 > n/e$ (limit of message passing [Deshpande-Montanari '15])

Planted clique problem: testing

$$\|L_{\le D}\|^2 = \sum_{S:|S|\le D} \langle L, \Phi_S \rangle^2 \approx \sum_{S:|S|\le D} \left(\frac{k}{n}\right)^{2|V(S)|}$$

- If $D = 1$, $\|L_{\le D}\|^2 \approx n^2 \left(\frac{k}{n}\right)^4 \gg 1$, if $k^2 \gg n$ $\Rightarrow$ counting edges succeeds
- If restricting S to be D-cycles with $D \to \infty$,

$$\|L_{D-\text{cycle}}\| \approx n^D \left(\frac{k}{n}\right)^{2D} \gg 1,$$

  if $k^2 > n$ (limit of spectral method [Alon-Krivelevich-Sudakov '98])
- If restricting S to be D-trees with $D \to \infty$,

$$\|L_{D-\text{tree}}\| \approx (ne)^{D+1} \left(\frac{k}{n}\right)^{2(D+1)} \gg 1,$$

  if $k^2 > n/e$ (limit of message passing [Deshpande-Montanari '15])

Planted clique problem: testing

$$\|L_{\leq D}\|^2 = \sum_{S:|S|\leq D} \langle L, \Phi_S \rangle^2 \approx \sum_{S:|S|\leq D} \left(\frac{k}{n}\right)^{2|V(S)|}$$

- If $D = 1$, $\|L_{\leq D}\|^2 \approx n^2 \left(\frac{k}{n}\right)^4 \gg 1$, if $k^2 \gg n$ $\Rightarrow$ counting edges succeeds
- If restricting S to be D-cycles with $D \to \infty$,

$$\|L_{D-\text{cycle}}\| \approx n^D \left(\frac{k}{n}\right)^{2D} \gg 1,$$

  if $k^2 > n$ (limit of spectral method [Alon-Krivelevich-Sudakov '98])
- If restricting S to be D-trees with $D \to \infty$,

$$\|L_{D-\text{tree}}\| \approx (ne)^{D+1} \left(\frac{k}{n}\right)^{2(D+1)} \gg 1,$$

  if $k^2 > n/e$ (limit of message passing [Deshpande-Montanari '15])
- A complete proof of success also needs to bound the variance under $\mathbb{H}_1$

Planted clique problem: from testing to estimation

- Goal: estimate node i is in the planted clique or not

Planted clique problem: from testing to estimation

- Goal: estimate node i is in the planted clique or not
- Idea: count a family $\mathcal{H}$ of graphs rooted at i with D edges:

$$f_i = \sum_{S \in \mathcal{H}} \langle L, \Phi_S \rangle \, \Phi_S, \text{ where } \Phi_S = \prod_{(i,j) \in S} (2A_{ij} - 1)$$

Planted clique problem: from testing to estimation

- Goal: estimate node i is in the planted clique or not
- Idea: count a family $\mathcal{H}$ of graphs rooted at i with D edges:

$$f_i = \sum_{S \in \mathcal{H}} \langle L, \Phi_S \rangle \, \Phi_S, \text{ where } \Phi_S = \prod_{(i,j) \in S} (2A_{ij} - 1)$$

- E.g., choose $\mathcal{H}$ to be rooted D-trees. Conditional on planted clique C

Planted clique problem: from testing to estimation

- Goal: estimate node i is in the planted clique or not
- Idea: count a family $\mathcal{H}$ of graphs rooted at i with D edges:

$$f_i = \sum_{S \in \mathcal{H}} \langle L, \Phi_S \rangle \, \Phi_S, \text{ where } \Phi_S = \prod_{(i,j) \in S} (2A_{ij} - 1)$$

- E.g., choose $\mathcal{H}$ to be rooted D-trees. Conditional on planted clique C
  - Mean separation:

$$\mathbb{E}_P[f_i] \approx \sum_{S \in \mathcal{H}} \left(\frac{k}{n}\right)^{D+1} \mathbf{1}\{V(S) \subset C\} \approx \left(\frac{k^2 e}{n}\right)^{D} \mathbf{1}\{i \in C\}$$

Planted clique problem: from testing to estimation

- Goal: estimate node i is in the planted clique or not
- Idea: count a family $\mathcal{H}$ of graphs rooted at i with D edges:

$$f_i = \sum_{S \in \mathcal{H}} \langle L, \Phi_S \rangle \Phi_S, \text{ where } \Phi_S = \prod_{(i,j) \in S} (2A_{ij} - 1)$$

- E.g., choose $\mathcal{H}$ to be rooted D-trees. Conditional on planted clique C
  - Mean separation:

$$\mathbb{E}_P[f_i] \approx \sum_{S \in \mathcal{H}} \left(\frac{k}{n}\right)^{D+1} \mathbf{1}\{V(S) \subset C\} \approx \left(\frac{k^2 e}{n}\right)^D \mathbf{1}\{i \in C\}$$

  - Variance: need to show $\text{Var}_P[f_i] \ll \left(k^2 e/n\right)^{2D}$

Planted clique problem: from testing to estimation

- Goal: estimate node i is in the planted clique or not
- Idea: count a family $\mathcal{H}$ of graphs rooted at i with D edges:

$$f_i = \sum_{S \in \mathcal{H}} \langle L, \Phi_S \rangle \, \Phi_S, \text{ where } \Phi_S = \prod_{(i,j) \in S} (2A_{ij} - 1)$$

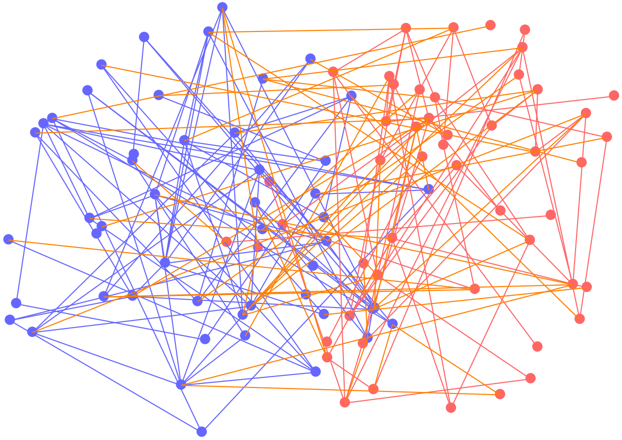
- E.g., choose $\mathcal{H}$ to be rooted D-trees. Conditional on planted clique C
  - Mean separation:

$$\mathbb{E}_P[f_i] \approx \sum_{S \in \mathcal{H}} \left(\frac{k}{n}\right)^{D+1} \mathbf{1}\{V(S) \subset C\} \approx \left(\frac{k^2 e}{n}\right)^D \mathbf{1}\{i \in C\}$$

  - Variance: need to show $\mathsf{Var}_P[f_i] \ll \left(k^2 e/n\right)^{2D}$
  - By Chebyshev's inequality, succeeds whp when $k^2 > n/e$ by choosing $D = \Theta(\log n)$

Community detection: Stochastic block model

1. n nodes are assigned to 2 communities uniformly at random
2. For every pair of nodes in same community, add an edge w.p. $\frac{a}{n}$
3. For every pair of nodes in diff. community, add an edge w.p. $\frac{b}{n}$



Stochastic block model: testing

$$\mathbb{H}_0 : A \sim \mathcal{G}\left(n, (a+b)/(2n)\right) \qquad \text{(Null model)}$$

$$\mathbb{H}_1 : A \sim \mathcal{G}\left(n, a/n, b/n\right) \qquad \text{(Planted model)}$$

Stochastic block model: testing

$$\mathbb{H}_0 : A \sim \mathcal{G}\left(n, (a+b)/(2n)\right) \qquad \text{(Null model)}$$
$$\mathbb{H}_1 : A \sim \mathcal{G}\left(n, a/n, b/n\right) \qquad \text{(Planted model)}$$

- Polynomial basis: $\Phi_S = \prod_{(i,j)\in S} \frac{A_{ij}-q}{\sigma}$ for $q = \frac{a+b}{2n}$ and $\sigma = \sqrt{q(1-q)}$

Stochastic block model: testing

$$\mathbb{H}_0 : A \sim \mathcal{G}\left(n, (a+b)/(2n)\right) \qquad \text{(Null model)}$$
$$\mathbb{H}_1 : A \sim \mathcal{G}\left(n, a/n, b/n\right) \qquad \text{(Planted model)}$$

- Polynomial basis: $\Phi_S = \prod_{(i,j)\in S} \frac{A_{ij}-q}{\sigma}$ for $q = \frac{a+b}{2n}$ and $\sigma = \sqrt{q(1-q)}$
- Likelihood-ratio projection:

$$\langle L, \Phi_S \rangle = \mathbb{E}_x \mathbb{E}_{A|x}[\Phi_S] = \mathbb{E}_x \left[\prod_{(i,j)\in S} \frac{r x_i x_j}{\sigma} \right] = \left(\frac{r}{\sigma}\right)^{|S|} \mathbf{1}\{S \text{ is even}\}$$

where $x \in \{\pm 1\}^n$ denotes the hidden community label and $r = \frac{a-b}{2n}$

Stochastic block model: testing

$$\mathbb{H}_0 : A \sim \mathcal{G}\left(n, (a+b)/(2n)\right) \qquad \text{(Null model)}$$
$$\mathbb{H}_1 : A \sim \mathcal{G}\left(n, a/n, b/n\right) \qquad \text{(Planted model)}$$

- Polynomial basis: $\Phi_S = \prod_{(i,j)\in S} \frac{A_{ij}-q}{\sigma}$ for $q = \frac{a+b}{2n}$ and $\sigma = \sqrt{q(1-q)}$
- Likelihood-ratio projection:

$$\langle L, \Phi_S \rangle = \mathbb{E}_x \mathbb{E}_{A|x}[\Phi_S] = \mathbb{E}_x \left[\prod_{(i,j)\in S} \frac{r x_i x_j}{\sigma} \right] = \left(\frac{r}{\sigma}\right)^{|S|} \mathbf{1}\{S \text{ is even}\}$$

where $x \in \{\pm 1\}^n$ denotes the hidden community label and $r = \frac{a-b}{2n}$

- So, we get

$$\|L_{\leq D}\|^2 = \sum_{S:|S|\leq D} \langle L, \Phi_S \rangle^2 = \sum_{\text{even } S:|S|\leq D} \left(\frac{r}{\sigma}\right)^{2|S|}$$

Stochastic block model: testing

$$\|L_{\leq D}\|^2 = \sum_{S:|S|\leq D} \langle L, \Phi_S \rangle^2 = \sum_{\text{even } S:|S|\leq D} \left(\frac{r}{\sigma}\right)^{2|S|}$$

Stochastic block model: testing

$$\|L_{\leq D}\|^2 = \sum_{S:|S|\leq D} \langle L, \Phi_S \rangle^2 = \sum_{\text{even } S:|S|\leq D} \left(\frac{r}{\sigma}\right)^{2|S|}$$

- Since S must be an even graph, it cannot be a tree.

Stochastic block model: testing

$$\|L_{\leq D}\|^2 = \sum_{S:|S|\leq D} \langle L, \Phi_S \rangle^2 = \sum_{\text{even } S:|S|\leq D} \left(\frac{r}{\sigma}\right)^{2|S|}$$

- Since S must be an even graph, it cannot be a tree.
- The number of vertices is at most D, achieved when S is a cycle

Stochastic block model: testing

$$\|L_{\leq D}\|^2 = \sum_{S:|S|\leq D} \langle L, \Phi_S \rangle^2 = \sum_{\text{even } S:|S|\leq D} \left(\frac{r}{\sigma}\right)^{2|S|}$$

- Since S must be an even graph, it cannot be a tree.
- The number of vertices is at most D, achieved when S is a cycle
- If restricting S to be D-cycles with $D \to \infty$,

$$\|L_{D-\text{cycle}}\|^2 \approx n^D \left(\frac{r}{\sigma}\right)^{2D} \approx \left(\frac{(a-b)^2}{2(a+b)}\right)^D \gg 1,$$

when $(a-b)^2 > 2(a+b)$ (detection threshold [Mossel-Neeman-Sly '15])

Stochastic block model: testing

$$\|L_{\le D}\|^2 = \sum_{S:|S|\le D} \langle L, \Phi_S \rangle^2 = \sum_{\text{even } S:|S|\le D} \left(\frac{r}{\sigma}\right)^{2|S|}$$

- Since S must be an even graph, it cannot be a tree.
- The number of vertices is at most D, achieved when S is a cycle
- If restricting S to be D-cycles with $D \to \infty$,

$$\|L_{D-\text{cycle}}\|^2 \approx n^D \left(\frac{r}{\sigma}\right)^{2D} \approx \left(\frac{(a-b)^2}{2(a+b)}\right)^D \gg 1,$$

  when $(a-b)^2 > 2(a+b)$ (detection threshold [Mossel-Neeman-Sly '15])
- To show $L_{D-\text{cycle}}$ succeeds, also need $\text{Var}_P[L_{D-\text{cycle}}] \ll (\mathbb{E}_P[L_{D-\text{cycle}}])^2$

Stochastic block model: from testing to estimation

- Goal: determine whether vertices u, v are in the same community or not

Stochastic block model: from testing to estimation

- Goal: determine whether vertices u, v are in the same community or not
- Idea: count set $\mathcal{H}$ of D-paths between u and v [Massoulié 13, Hopkins-Steurer '17, Mossel-Neeman-Sly '18, Abbe-Sandon '18]

$$f_{uv} = \frac{1}{n^{D/2-1}} \sum_{S \in \mathcal{H}} \Phi_S, \text{ where } \Phi_S = \prod_{(i,j) \in S} \frac{A_{ij} - q}{\sigma}$$

Stochastic block model: from testing to estimation

- Goal: determine whether vertices u, v are in the same community or not
- Idea: count set $\mathcal{H}$ of D-paths between u and v [Massoulié 13, Hopkins-Steurer '17, Mossel-Neeman-Sly '18, Abbe-Sandon '18]

$$f_{uv} = \frac{1}{n^{D/2-1}} \sum_{S \in \mathcal{H}} \Phi_S, \text{ where } \Phi_S = \prod_{(i,j) \in S} \frac{A_{ij} - q}{\sigma}$$

- Conditional on community label $x \in \{\pm 1\}$:
  - Mean separation:

$$\mathbb{E}_P[f_{uv}] = \frac{1}{n^{D/2-1}} \sum_{S \in \mathcal{H}} \left(\frac{r}{\sigma}\right)^D \prod_{(i,j) \in S} x_i x_j \approx \left(\frac{(a-b)^2}{2(a+b)}\right)^{D/2} x_u x_v$$

Stochastic block model: from testing to estimation

- Goal: determine whether vertices u, v are in the same community or not
- Idea: count set $\mathcal{H}$ of D-paths between u and v [Massoulié 13, Hopkins-Steurer '17, Mossel-Neeman-Sly '18, Abbe-Sandon '18]

$$f_{uv} = \frac{1}{n^{D/2-1}} \sum_{S \in \mathcal{H}} \Phi_S, \text{ where } \Phi_S = \prod_{(i,j) \in S} \frac{A_{ij} - q}{\sigma}$$

- Conditional on community label $x \in \{\pm 1\}$:
  - Mean separation:

$$\mathbb{E}_P[f_{uv}] = \frac{1}{n^{D/2-1}} \sum_{S \in \mathcal{H}} \left(\frac{r}{\sigma}\right)^D \prod_{(i,j) \in S} x_i x_j \approx \left(\frac{(a-b)^2}{2(a+b)}\right)^{D/2} x_u x_v$$

  - Variance: show $\mathsf{E}_P[f_{u,v}^2] = O(1) \times \left(\frac{(a-b)^2}{2(a+b)}\right)^D$

Stochastic block model: from testing to estimation

- Goal: determine whether vertices u, v are in the same community or not
- Idea: count set $\mathcal{H}$ of D-paths between u and v [Massoulié 13, Hopkins-Steurer '17, Mossel-Neeman-Sly '18, Abbe-Sandon '18]

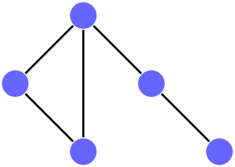
$$f_{uv} = \frac{1}{n^{D/2-1}} \sum_{S \in \mathcal{H}} \Phi_S, \text{ where } \Phi_S = \prod_{(i,j)\in S} \frac{A_{ij} - q}{\sigma}$$

- Conditional on community label $x \in \{\pm 1\}$:
  - Mean separation:

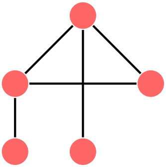
$$\mathbb{E}_P[f_{uv}] = \frac{1}{n^{D/2-1}} \sum_{S \in \mathcal{H}} \left(\frac{r}{\sigma}\right)^D \prod_{(i,j)\in S} x_i x_j \approx \left(\frac{(a-b)^2}{2(a+b)}\right)^{D/2} x_u x_v$$

  - Variance: show $\mathsf{E}_P[f_{u,v}^2] = O(1) \times \left(\frac{(a-b)^2}{2(a+b)}\right)^D$
  - Attain the sharp estimation threshold $(a-b)^2 > 2(a+b)$, by choosing $D = \Theta(\log n)$ [Hopkins-Steurer '17]

Network alignment: Correlated Erdős-Rényi graphs

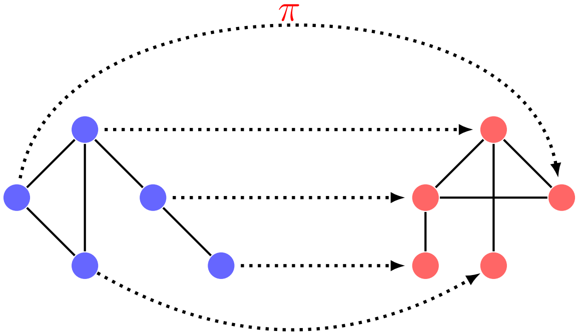


$A \sim \mathcal{G}(n, q)$



$B \sim \mathcal{G}(n, q)$

Network alignment: Correlated Erdős-Rényi graphs

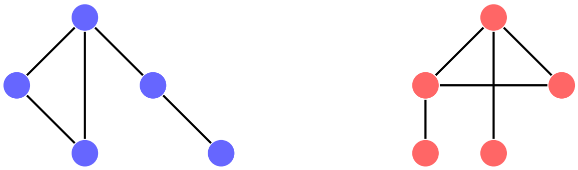


$A \sim \mathcal{G}(n, q)$ $B \sim \mathcal{G}(n, q)$

A and B are edge-wise correlated (ρ) under the hidden node correspondence π:

$\{A_{ij}, B_{\pi(i)\pi(j)}\}$ are i.i.d. pairs of $\mathrm{Bern}(q)$ with correlation ρ

Network alignment: Correlated Erdős-Rényi graphs



$A \sim \mathcal{G}(n,q)$ $\qquad\qquad$ $B \sim \mathcal{G}(n,q)$

A and B are edge-wise correlated (ρ) under the hidden node correspondence π:

$\{A_{ij}, B_{\pi(i)\pi(j)}\}$ are i.i.d. pairs of $\mathrm{Bern}(q)$ with correlation ρ

Goal: observe A and B, recover the hidden node correspondence π

Random network alignment: testing

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

$\mathbb{H}_1$: A and B are two ρ-correlated Erdős-Rényi graphs $\mathcal{G}(n, q, \rho)$

Random network alignment: testing

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

$\mathbb{H}_1$: A and B are two ρ-correlated Erdős-Rényi graphs $\mathcal{G}(n,q,\rho)$

- Orthogonal polynomial basis: for $S = (S_1, S_2)$,

$$\Phi_S(A,B) = \prod_{(i,j)\in S_1} \frac{A_{ij}-q}{\sqrt{q(1-q)}} \cdot \prod_{(i,j)\in S_2} \frac{B_{ij}-q}{\sqrt{q(1-q)}}$$

Random network alignment: testing

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

$\mathbb{H}_1$: A and B are two ρ-correlated Erdős-Rényi graphs $\mathcal{G}(n, q, \rho)$

- Orthogonal polynomial basis: for $S = (S_1, S_2)$,

$$\Phi_S(A, B) = \prod_{(i,j) \in S_1} \frac{A_{ij} - q}{\sqrt{q(1-q)}} \cdot \prod_{(i,j) \in S_2} \frac{B_{ij} - q}{\sqrt{q(1-q)}}$$

- Likelihood-ratio projection:

$$\begin{aligned} \langle L, \Phi_S \rangle = \mathbb{E}_\pi \mathbb{E}_{A,B|\pi}[\Phi_S] &= \mathbb{E}_\pi[\rho^{|S_1|} \mathbf{1}\{\pi(S_1) = S_2\}] \\ &= \rho^{|H|} \frac{1}{\mathsf{sub}(H)} \mathbf{1}\{S_1 \cong S_2 \cong H\} \end{aligned}$$

where $\cong$ means isomorphism and $\mathsf{sub}(H) = \#$ of copies of unlabeled H

Random network alignment: testing

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

$\mathbb{H}_1$: A and B are two ρ-correlated Erdős-Rényi graphs $\mathcal{G}(n, q, \rho)$

- Orthogonal polynomial basis: for $S = (S_1, S_2)$,

$$\Phi_S(A, B) = \prod_{(i,j)\in S_1} \frac{A_{ij} - q}{\sqrt{q(1-q)}} \cdot \prod_{(i,j)\in S_2} \frac{B_{ij} - q}{\sqrt{q(1-q)}}$$

- Likelihood-ratio projection:

$$\begin{aligned}\langle L, \Phi_S \rangle = \mathbb{E}_\pi \mathbb{E}_{A,B|\pi}[\Phi_S] &= \mathbb{E}_\pi[\rho^{|S_1|} \mathbf{1}\{\pi(S_1) = S_2\}] \\ &= \rho^{|H|} \frac{1}{\mathsf{sub}(H)} \mathbf{1}\{S_1 \cong S_2 \cong H\}\end{aligned}$$

  where $\cong$ means isomorphism and $\mathsf{sub}(H) = \#$ of copies of unlabeled H
- So, we get

$$\|L_{\le 2D}\|^2 = \sum_{S:|S|\le 2D} \langle L, \Phi_S \rangle^2 = \sum_{H:|H|\le D} \sum_{S_1,S_2 \cong H} \frac{\rho^{2|H|}}{\mathsf{sub}^2(H)} = \sum_{H:|H|\le D} \rho^{2|H|}$$

Random network alignment: testing

$$\|L_{\leq 2D}\|^2 = \sum_{S:|S|\leq 2D} \langle L, \Phi_S \rangle^2 = \sum_{H:|H|\leq D} \rho^{2|H|}$$

Random network alignment: testing

$$\|L_{\leq 2D}\|^2 = \sum_{S:|S|\leq 2D} \langle L, \Phi_S \rangle^2 = \sum_{H:|H|\leq D} \rho^{2|H|}$$

- If restricting H to be set $\mathcal{T}$ of unlabeled D-trees with $D \to \infty$,

$$\|L_{D-\mathsf{tree}}\|^2 = \rho^{2D} |\mathcal{T}| = \left(\frac{\rho^2}{\alpha}\right)^D \gg 1,$$

when $\rho^2 > \alpha,$ where $\alpha \approx 0.33833$ is Otter's constant [Otter '48]

Random network alignment: testing

$$\|L_{\leq 2D}\|^2 = \sum_{S:|S|\leq 2D} \langle L, \Phi_S \rangle^2 = \sum_{H:|H|\leq D} \rho^{2|H|}$$

- If restricting H to be set $\mathcal{T}$ of unlabeled D-trees with $D \to \infty$,

$$\|L_{D-\text{tree}}\|^2 = \rho^{2D}|\mathcal{T}| = \left(\frac{\rho^2}{\alpha}\right)^D \gg 1,$$

  when $\rho^2 > \alpha$, where $\alpha \approx 0.33833$ is Otter's constant [Otter '48]
- To show $L_{D-\text{tree}}$ succeeds, also need $\text{Var}_P[L_{D-\text{tree}}] \ll (\mathbb{E}_P[L_{D-\text{tree}}])^2$ [Mao-Wu-Yu-X.'24]

Random network alignment: from testing to estimation

- Goal: determine whether vertices u in A and v in B are true pair or not

Random network alignment: from testing to estimation

- Goal: determine whether vertices u in A and v in B are true pair or not
- Idea: count family $\mathcal{T}$ of rooted D-trees:

$$f_{uv} = \sum_{H \in \mathcal{T}} \frac{\rho^{|H|}}{\mathsf{sub}(H)} \sum_{S_1(u), S_2(v) \cong H} \prod_{(i,j) \in S_1} \frac{A_{ij} - q}{\sqrt{q(1-q)}} \cdot \prod_{(i,j) \in S_2} \frac{B_{ij} - q}{\sqrt{q(1-q)}}$$

Random network alignment: from testing to estimation

- Goal: determine whether vertices u in A and v in B are true pair or not
- Idea: count family $\mathcal{T}$ of rooted D-trees:

$$f_{uv} = \sum_{H \in \mathcal{T}} \frac{\rho^{|H|}}{\mathsf{sub}(H)} \sum_{S_1(u), S_2(v) \cong H} \prod_{(i,j) \in S_1} \frac{A_{ij} - q}{\sqrt{q(1-q)}} \cdot \prod_{(i,j) \in S_2} \frac{B_{ij} - q}{\sqrt{q(1-q)}}$$

- Conditional on latent node mapping π:
  - Mean separation (assuming H is uniquely rooted):

$$\mathbb{E}_P[f_{uv}] = \sum_{H \in \mathcal{T}} \rho^{2|H|} \mathbf{1}\{\pi(u) = v\} \sim \left(\frac{\rho^2}{\alpha}\right)^D \mathbf{1}\{\pi(u) = v\}$$

Random network alignment: from testing to estimation

- Goal: determine whether vertices u in A and v in B are true pair or not
- Idea: count family $\mathcal{T}$ of rooted D-trees:

$$f_{uv} = \sum_{H \in \mathcal{T}} \frac{\rho^{|H|}}{\mathsf{sub}(H)} \sum_{S_1(u), S_2(v) \cong H} \prod_{(i,j) \in S_1} \frac{A_{ij} - q}{\sqrt{q(1-q)}} \cdot \prod_{(i,j) \in S_2} \frac{B_{ij} - q}{\sqrt{q(1-q)}}$$

- Conditional on latent node mapping π:
  - Mean separation (assuming H is uniquely rooted):

$$\mathbb{E}_P [f_{uv}] = \sum_{H \in \mathcal{T}} \rho^{2|H|} \mathbf{1}\{\pi(u) = v\} \sim \left(\frac{\rho^2}{\alpha}\right)^D \mathbf{1}\{\pi(u) = v\}$$

  - To control the variance, we restrict to a special family $\mathcal{T}^*$ of unlabeled rooted trees–chandeliers, where $|\mathcal{T}^*| = (1/\alpha - o(1))^D$ [Mao-Wu-X.-Yu '23]

Outline of tutorial

- Introduction to low-degree polynomial method
- Three prototypical examples
  - Planted clique
  - Stochastic block model
  - Random network alignment
- Concluding remarks

A few remarks

- Tree- or cycle-based polynomials of degree D can be approximated in time $n^2 e^{O(D)}$ via color-coding [Alon-Yuster-Zwick '94]

A few remarks

- Tree- or cycle-based polynomials of degree D can be approximated in time $n^2 e^{O(D)}$ via color-coding [Alon-Yuster-Zwick '94]
- The polynomials often come from the low-degree projection of the likelihood ratio, though some extra "twists" may be needed for estimation

A few remarks

- Tree- or cycle-based polynomials of degree D can be approximated in time $n^2 e^{O(D)}$ via color-coding [Alon-Yuster-Zwick '94]
- The polynomials often come from the low-degree projection of the likelihood ratio, though some extra "twists" may be needed for estimation
- A major simplification comes from i.i.d. observations under $\mathbb{H}_0$, which allow us to explicitly construct an orthogonal polynomial basis

A few remarks

- Tree- or cycle-based polynomials of degree D can be approximated in time $n^2 e^{O(D)}$ via color-coding [Alon-Yuster-Zwick '94]
- The polynomials often come from the low-degree projection of the likelihood ratio, though some extra "twists" may be needed for estimation
- A major simplification comes from i.i.d. observations under $\mathbb{H}_0$, which allow us to explicitly construct an orthogonal polynomial basis
- A key step is to evaluate the projection coefficient $\langle L, \Phi_S \rangle$, which equals the mean of Φ_S under the planted model

A few remarks

- Tree- or cycle-based polynomials of degree D can be approximated in time $n^2 e^{O(D)}$ via color-coding [Alon-Yuster-Zwick '94]
- The polynomials often come from the low-degree projection of the likelihood ratio, though some extra "twists" may be needed for estimation
- A major simplification comes from i.i.d. observations under $\mathbb{H}_0$, which allow us to explicitly construct an orthogonal polynomial basis
- A key step is to evaluate the projection coefficient $\langle L, \Phi_S \rangle$, which equals the mean of Φ_S under the planted model
- To complete the analysis, we also need to bound the variance of the polynomial under $\mathbb{H}_1$. This can be quite challenging-sometimes special designs (e.g., counting chandeliers) help.

A few remarks

- Tree- or cycle-based polynomials of degree D can be approximated in time $n^2 e^{O(D)}$ via color-coding [Alon-Yuster-Zwick '94]
- The polynomials often come from the low-degree projection of the likelihood ratio, though some extra "twists" may be needed for estimation
- A major simplification comes from i.i.d. observations under $\mathbb{H}_0$, which allow us to explicitly construct an orthogonal polynomial basis
- A key step is to evaluate the projection coefficient $\langle L, \Phi_S \rangle$, which equals the mean of Φ_S under the planted model
- To complete the analysis, we also need to bound the variance of the polynomial under $\mathbb{H}_1$. This can be quite challenging-sometimes special designs (e.g., counting chandeliers) help.
- The low-degree polynomial method extends to many other high-dimensional inference settings. For example, with i.i.d. Gaussian null model, the orthogonal basis is given by Hermite polynomials.

A partial and ever-growing list of successes

- Planted dense subgraph [Sohn-Wein '25]
- Planted dense cycles [Mao-Wein-Zhang '23]
- Dense stochastic block models [Banerjee-Ma '17, Banerjee '18]
- Degree-corrected stochastic block models [Gao-Lafferty '17, Jin-Ke-Luo '19]
- Mixed-membership stochastic block models [Hopkins-Steurer '17]
- Random network alignment: correlated stochastic block models [Chen-Ding-Gong-Li '24,25, Chai-Rácz 24]
- Attributed network alignment [Wang-Wang-Wang'24]
- Testing random geometric graph vs. Erdős-Rényi [Bubeck-Ding-Eldan-Rácz '16]
- Planted submatrix [Sohn-Wein '25]
- Spiked Wigner model [Hopkins-Steurer '17, Sohn-Wein '25]
- Tensor PCA [Hopkins '18, Li '25]
- Shuffled linear regression [Li '25, Gong-Wu-X. '25]
- Procrustes-Wasserstein matching [Niu-Schramm-X. '25]
- And many more...

Challenges and open problems

- The likelihood ratio projection can be hard to compute
  - Example: random geometric graph. Suppose x_i's are i.i.d. on the unit sphere in $\mathbb{R}^d$, and conditional on x_i's, $A_{ij} \overset{\text{iid}}{\sim} \text{Bern}(\kappa(x_i, x_j))$.
  - In this case, $\langle L, \Phi_S \rangle = \mathbb{E}_P[\Phi_S]$ is hard to compute except for simple subgraphs such as cycles. See recent progress [Bangachev-Bresler '25]
  - A key obstacle in resolving the long-standing conjecture on detection threshold for RGG vs Erdős-Rényi graph [Liu-Mohanty-Schramm-Yang '21]

Challenges and open problems

- The likelihood ratio projection can be hard to compute
  - Example: random geometric graph. Suppose x_i's are i.i.d. on the unit sphere in $\mathbb{R}^d$, and conditional on x_i's, $A_{ij} \overset{\text{iid}}{\sim} \mathrm{Bern}(\kappa(x_i, x_j))$.
  - In this case, $\langle L, \Phi_S \rangle = \mathbb{E}_P[\Phi_S]$ is hard to compute except for simple subgraphs such as cycles. See recent progress [Bangachev-Bresler '25]
  - A key obstacle in resolving the long-standing conjecture on detection threshold for RGG vs Erdős-Rényi graph [Liu-Mohanty-Schramm-Yang '21]

- The null model may not have IID observations, so the nice orthogonality property is missing
  - Example: aligning random geometric graphs. Suppose y_i is correlated with $x_{\pi(i)}$, and conditional on y_i's, $B_{ij} \overset{\text{iid}}{\sim} \mathrm{Bern}(\kappa(y_i, y_j))$.
  - Under the null model, A and B are two independent random geometric graphs, but the orthogonal polynomial basis is unknown

Challenges and open problems

- The likelihood ratio projection can be hard to compute
  - Example: random geometric graph. Suppose x_i's are i.i.d. on the unit sphere in $\mathbb{R}^d$, and conditional on x_i's, $A_{ij} \overset{\text{iid}}{\sim} \text{Bern}(\kappa(x_i, x_j))$.
  - In this case, $\langle L, \Phi_S \rangle = \mathbb{E}_P[\Phi_S]$ is hard to compute except for simple subgraphs such as cycles. See recent progress [Bangachev-Bresler '25]
  - A key obstacle in resolving the long-standing conjecture on detection threshold for RGG vs Erdős-Rényi graph [Liu-Mohanty-Schramm-Yang '21]

- The null model may not have IID observations, so the nice orthogonality property is missing
  - Example: aligning random geometric graphs. Suppose y_i is correlated with $x_{\pi(i)}$, and conditional on y_i's, $B_{ij} \overset{\text{iid}}{\sim} \text{Bern}(\kappa(y_i, y_j))$.
  - Under the null model, A and B are two independent random geometric graphs, but the orthogonal polynomial basis is unknown

- Dynamic networks
  - Example: preferential attachment (PA) models. How can we design low-degree polynomial estimators for inference problems in PA graphs—such as community detection or network alignment?

Conclusions

- Network inference provides a rich family of problems that intertwine *applied probability, statistics, optimization, combinatorics, information theory, and more*.
- The low-degree polynomial method offers a simple yet principled framework for understanding the fundamental limits of high-dimensional inference.
- This tutorial has focused on low-degree "upper bounds"— showing how to design effective low-degree, polynomial-based estimators.
- A complementary perspective comes from low-degree "lower bounds", which characterize thresholds below which all low-degree polynomials fail. Under the low-degree conjecture, this further implies all polynomial-time algorithms fail

Further Reading

- A. Wein, *"Computational Complexity of Statistics: New Insights from Low-Degree Polynomials,"* June 2025.
- Y. Wu and J. Xu, *"Statistical Inference on Graphs: Selected Topics,"* https://people.duke.edu/~jx77/stats-graphs.pdf. Lecture notes