A Tutorial on the Method of Orthogonal Polynomials for Random Graphs via the Planted Clique Problem

Cheng Mao

Georgia Institute of Technology

November 24, 2025

Abstract

In recent years, there has been a surge of interest in applying the method of orthogonal polynomials to study inference problems on random graphs. This approach has been successful in obtaining both positive results for new algorithms via subgraph counts and negative results for computationally efficient algorithms in the framework of low-degree polynomials. To introduce the method of orthogonal polynomials on random graphs, this self-contained tutorial focuses on the planted clique model, which has become an iconic model in the study of statistical-to-computational gaps. It is based on lectures given by the author in a topics course in Fall 2023 and a tutorial in Spring 2025 at Georgia Tech.

Contents

1 Introduction to the planted clique model

The goal of this tutorial is to provide a self-contained introduction to the method of orthogonal polynomials on random graphs and the statistical-to-computational gap for the planted clique model, assuming a minimal background in this area. For a comprehensive review of this topic, see

the recent survey [Wei25] on the method of low-degree polynomials for average-case computational complexity.

1.1 Random graph models

A *graph* or an *undirected graph* G consists of a pair (V, E) where V is the vertex set and $E \subset \binom{V}{2}$ is the edge set. We take $V = [n] := \{1, 2, \dots, n\}$, and thus $E \subset \binom{[n]}{2} = \{(i, j) : 1 \le i < j \le n\}$. The *adjacency matrix* $A \in \{0, 1\}^{n \times n}$ of the graph G is a symmetric matrix defined by $A_{ij} = 1$ if $(i, j) \in E$ and $A_{ij} = 0$ if $(i, j) \notin E$. The diagonal entries A_{ii} are assumed to be 0 (i.e., no self-loops). For simplicity, we often identify the graph with its adjacency matrix and write "graph A" and "edge A_{ij}". If $(i, j) \in E$, then vertex j is a *neighbor* of vertex i. The number of neighbors of vertex i is called the *degree* of vertex i.

In statistical inference, the observed graphs are typically assumed to be random graphs from certain probabilistic models.

Definition 1.1 (Erdős–Rényi model)**.** *We say that a graph A is an* Erdős–Rényi graph *and write $A \sim G(n, p)$ if A is a graph on n vertices with i.i.d. edges $A_{ij} \sim \mathsf{Ber}(p)$, where $\mathsf{Ber}(p)$ denotes the Bernoulli distribution with parameter $p \in [0, 1]$.*

A *clique* is a complete subgraph of a graph. The *planted clique* model refers to one of the following [Jer92, AKS98].

Definition 1.2 (Planted clique model)**.** *For a fixed $k \in [n]$, take a uniformly random subset $\mathcal{C} \subset [n]$ of size $|\mathcal{C}| = k$. Conditional on $\mathcal{C}$, we observe a graph A with $A_{ij} = 1$ if $i, j \in \mathcal{C}$ and $A_{ij} \sim \mathsf{Ber}(1/2)$ otherwise, where the edges are assumed to be independent. We write $A \sim G(n, 1/2, k)$.*

Definition 1.3 (Planted clique model with independent vertices)**.** *For a fixed $k \in [n]$, let $\mathcal{C}$ be a random subset of $[n]$ that contains each vertex $i \in [n]$ independently with probability k/n. Conditional on $\mathcal{C}$, we observe a graph A with $A_{ij} = 1$ if $i, j \in \mathcal{C}$ and $A_{ij} \sim \mathsf{Ber}(1/2)$ otherwise, where the edges are assumed to be independent. We write $A \sim \tilde{G}(n, 1/2, k)$.*

We mainly tackle the model $G(n, 1/2, k)$ but also consider the second model $\tilde{G}(n, 1/2, k)$ for technical convenience. Note that in $\tilde{G}(n, 1/2, k)$, the size of $\mathcal{C}$ concentrates around its expectation k with high probability by the Chernoff bound if $k \to \infty$ as $n \to \infty$, so the two models are quite similar.

1.2 Testing and estimation

Given a graph A from the planted clique model, we would like to recover the location of the clique. An even more basic question is to tell whether a given random graph A contains a planted clique at all. We now formalize these questions using the language of *statistical hypothesis testing* and *estimation*.

Testing To test whether the observed graph A contains a planted clique, it amounts to test between two hypotheses H_0 and H_1 defined as follows.

- *Null hypothesis* H_0: $A \sim G(n, 1/2)$.
- *Alternative hypothesis* H_1: $A \sim G(n, 1/2, k)$.

A *test* $\phi : \mathbb{R}^{n\times n} \to \{0, 1\}$ takes in the observation A and outputs a binary number indicating the hypothesis. A criterion for a "good" test is the following. A test ϕ is *consistent* if it makes an error with a vanishing probability as the size of the problem grows, i.e.,

$$\mathbb{P}_0\{\phi(A) = 1\} + \mathbb{P}_1\{\phi(A) = 0\} \to 0 \tag{1}$$

as $n \to \infty$, where $\mathbb{P}_i$ denotes the probability under the hypothesis H_i. We are generally interested in a set of conditions in terms of the problem parameters under which there exists a consistent test. We say that the testing problem is "easy" and we achieve an "upper bound" or a "positive result" when there is a consistent test; on the other hand, if there is no consistent test in a certain regime of parameters, we say that the problem is "hard" and we achieve a "lower bound" or a "negative result".

Estimation Assuming that the observed graph A is from the planted clique model $G(n, 1/2, k)$, i.e., H_1 is true, we are interested in estimating the location of the clique. More precisely, we aim to construct an estimator $\hat{\mathcal{C}} = \hat{\mathcal{C}}(A) \subset [n]$ that is close to the true subset $\mathcal{C} \subset [n]$. Suppose that the difference between $\mathcal{C}$ and $\hat{\mathcal{C}}$ is measured by a certain *loss* $L(\mathcal{C}, \hat{\mathcal{C}})$, e.g., $L(\mathcal{C}, \hat{\mathcal{C}}) = |\hat{\mathcal{C}} \triangle \mathcal{C}|$. Then the goal is to characterize the loss in high probability or the *risk* $\mathbb{E}[L(\mathcal{C}, \hat{\mathcal{C}})]$ as $n \to \infty$.

We have *exact recovery* if $\hat{\mathcal{C}} = \mathcal{C}$ (with high probability as $n \to \infty$). We have *almost exact recovery* if $L(\mathcal{C}, \hat{\mathcal{C}}) = o(k)$ and *partial recovery* if $L(\mathcal{C}, \hat{\mathcal{C}})$ is nontrivially small. If there is an estimator $\hat{\mathcal{C}}$ such that $L(\mathcal{C}, \hat{\mathcal{C}})$ can be bounded from above, then we have an "upper bound" or a "positive result". If $L(\mathcal{C}, \hat{\mathcal{C}})$ is bounded from below for any estimator in a class of functions, then we have a "lower bound" or a "negative result". We aim for upper and lower bounds of the same order as $n \to \infty$, and if so, we have achieved the optimal *statistical rate of estimation* for the problem.

Note that whether a testing problem is easy or hard may depend on what class of tests we consider. The rates of estimation also depend on the function class of the estimator. *Information-theoretic* bounds refer to those obtained when considering all tests or estimators that are measurable with respect to A, while *computational* bounds refer to those obtained when considering a certain class of computationally efficient tests or estimators. For example, a procedure that involves searching over all possible subsets $\mathcal{C} \subset [n]$ may not give an efficient test or estimator because $\binom{n}{|\mathcal{C}|}$ is not polynomial in n if $|\mathcal{C}| \to \infty$ as $n \to \infty$.

1.3 Positive results for testing

We now consider the problem of detecting the planted clique, i.e., testing between H_0 and H_1. The difficulty of this testing problem is clearly related to the size k of the planted clique:

- If k is too small, then it is impossible to distinguish H_1 from H_0. If k is sufficiently large, then it is easy to distinguish H_1 from H_0.
- What is the threshold k above which we can distinguish H_1 from H_0 with probability $1 - o(1)$ given infinite computational power? Here $o(1)$ denotes a vanishing quantity as $n \to \infty$. We call this threshold the information-theoretic or statistical threshold.
- What is the threshold k above which we can distinguish H_1 from H_0 with probability $1-o(1)$ using a polynomial-time algorithm (from a certain class)? We call this threshold the computational threshold.

- Are the above two thresholds the same? If they are not the same, the gap between them is referred to as the *statistical-to-computational* gap.

Similar questions can be asked for the estimation or recovery problem too. In the sequel, we will discuss informally the information-theoretic and computational thresholds for detecting or recovering a planted clique.

Information-theoretic threshold First, we provide evidence that the information-theoretic threshold of k scales logarithmically in n.

For any k vertices, there are at most $\binom{k}{2}$ possible edges. Therefore, under H_0, the induced subgraph of A on these vertices is a clique with probability $2^{-\binom{k}{2}}$. If k is a constant and $n \to \infty$, there are a growing number of sets of k vertices in A, each forming a clique with a constant probability. Hence A contains a clique of size k with high probability for any constant k. As a result, if under H_1 a clique of constant size is planted, there is no significant difference between H_0 and H_1, and so we do not expect to be able to consistently detect the planted clique.

The above discussion suggests that the clique size k needs to grow in n if we aim for a positive result. It also motivates us to study the *clique number* $\omega(A)$ which is defined to be the size of the largest clique in A. If $\omega(A) \le k_0$ with high probability under H_0 and k is larger than k_0, then it is possible to detect and locate the planted clique of size k under H_1.

Proposition 1.4. *Let $A \sim G(n, 1/2)$. For any constant $\epsilon > 0$, we have*

$$\mathbb{P}\{\omega(A) \le (2+\epsilon)\log_2 n\} \to 1 \quad \text{as } n \to \infty.$$

Proof. For any fixed subset $\mathcal{C} \subset [n]$ of size k, it holds that

$$\mathbb{P}\{A_{ij} = 1 \text{ for all distinct } i, j \in \mathcal{C}\} = 2^{-\binom{k}{2}}.$$

Since there are $\binom{n}{k}$ subsets of $[n]$ of size k, we obtain

$$\mathbb{P}\{\omega(A) \ge k\} \le \mathbb{P}\{A \text{ contains a clique of size } k\} \le \binom{n}{k} \cdot 2^{-\binom{k}{2}} \le n^k \, 2^{-\frac{k(k-1)}{2}}.$$

For $k := \lfloor (2+\epsilon)\log_2 n \rfloor$, we have

$$\log_2(n^k 2^{-\frac{k(k-1)}{2}}) = k \log_2 n - k(k-1)/2 \to -\infty$$

as $n \to \infty$, so the conclusion holds. □

This result immediately implies that there is a consistent test for distinguishing H_1 from H_0.

Corollary 1.5. *Suppose that $k > (2+\epsilon)\log_2 n$ for a constant $\epsilon > 0$. Then*

$$\mathbb{P}_0\{\omega(A) > (2+\epsilon)\log_2 n\} + \mathbb{P}_1\{\omega(A) \le (2+\epsilon)\log_2 n\} \to 0 \quad \text{as } n \to \infty.$$

In other words, the test $\phi(A) := \mathbb{1}\{\omega(A) > (2+\epsilon)\log_2 n\}$ is consistent in the sense of (1).

Later we will show that if $k \le (2-\epsilon)\log_2 n$ for a constant $\epsilon > 0$, then there is no consistent test. Thus the information-theoretic threshold for planted clique detection is tightly characterized to be $k \approx 2\log_2 n$.

The threshold for exact recovery of the planted clique under H_1 turns out to be the same as the detection threshold. In particular, the upper bound is achieved by taking $\hat{\mathcal{C}}$ to be the vertex set of the largest clique in A. It is not difficult to show that $\hat{\mathcal{C}} = \mathcal{C}$ with high probability under H_1 if $k \ge (2+\epsilon)\log_2 n$, although this is not immediately implied by Proposition 1.4.

Computational threshold The issue with the above procedure based on the clique number $\omega(A)$ is that it cannot be efficiently computed. In the worst case, finding the largest clique in A entails an exhaustive search which takes exponential time. If we restrict our attention to polynomial-time algorithms, the threshold of k in n for testing or recovering the planted clique may be much larger than $2\log_2 n$. To see how large the computational threshold should be, let us consider computationally efficient procedures based on the vertex degrees of A.

The following is an immediate consequence of Hoeffding's inequality and the union bound.

Proposition 1.6. *Let $d_1, \dots, d_n$ denote the vertex degrees of A. Denote the total number of edges in A by $d(A) = \sum_{(i,j)\in\binom{[n]}{2}} A_{ij}$. There is an absolute constant $C > 0$ such that the following holds.*

- *For $A \sim G(n, 1/2)$, we have*

$$\mathbb{P}\left\{\left|d(A) - \frac{1}{2}\binom{n}{2}\right| > Cn\sqrt{\log n}\right\} \le n^{-10}.$$

- *For $A \sim G(n, 1/2, k)$ as in Definition 1.2, we have*

$$\mathbb{P}\left\{\left|d(A) - \frac{1}{2}\left[\binom{n}{2} + \binom{k}{2}\right]\right| > Cn\sqrt{\log n}\right\} \le n^{-10}.$$

  Moreover,

$$\mathbb{P}\left\{\max_{i\in[n]\setminus\mathcal{C}}\left|d_i - \frac{n-1}{2}\right| > C\sqrt{n\log n}\right\} \le n^{-10}$$

  and

$$\mathbb{P}\left\{\max_{i\in\mathcal{C}}\left|d_i - \frac{n+k}{2} + 1\right| > C\sqrt{n\log n}\right\} \le n^{-10}.$$

We then readily obtain the following corollary by simple union bounds.

Corollary 1.7. *There is an absolute constant $C > 0$ such that the following holds.*

- *Suppose that $k > 4\sqrt{Cn}\,(\log n)^{1/4}$. Then*

$$\mathbb{P}_0\left\{d(A) > \frac{1}{2}\binom{n}{2} + Cn\sqrt{\log n}\right\} + \mathbb{P}_1\left\{d(A) \le \frac{1}{2}\binom{n}{2} + Cn\sqrt{\log n}\right\} \le 2n^{-10}.$$

- *Suppose that $k > 4C\sqrt{n\log n}$. If we define $\hat{\mathcal{C}} := \{i \in [n] : d_i > \frac{n-1}{2} + C\sqrt{n\log n}\}$, then*

$$\mathbb{P}_1\{\hat{\mathcal{C}} \ne \mathcal{C}\} \le n^{-9}.$$

Therefore, if $k \gtrsim \sqrt{n\log n}$, we have a consistent test between H_0 and H_1 and also an estimator that achieves exact recovery with high probability under H_1. The condition can be improved to $k \gtrsim \sqrt{n}$ for several slightly more sophisticated algorithms. Note that this condition on k is still much worse than the information-theoretic threshold $k \approx 2\log_2 n$. In fact, it is widely conjectured that $k \asymp \sqrt{n}$ is the computational threshold for detecting or recovering a planted clique. That is to say, if $k \ll \sqrt{n}$, then there may be no polynomial-time algorithm that can distinguish H_1 from H_0 consistently. We will prove a version of this lower bound for a certain class of efficient algorithms later.

2 Orthogonal polynomials and the second moment method

2.1 Polynomial basis of functions on a random graph

We start with the following general setup. Let $A \in \{0,1\}^N$ be a random vector with independent Bernoulli entries $A_i \sim \mathsf{Ber}(p_i)$ for $i \in [N]$. For brevity, we write $A \sim \mathsf{Ber}(p)$. Given the observation A, suppose that we would like to do testing or estimation about the parameter vector $p \in [0,1]^N$. Note that this model can be specialized to a random (undirected) graph by taking $N = \binom{n}{2}$ and viewing A as a matrix in $\{0,1\}^{n \times n}$.

Since a test or an estimator is simply a function of $A \in \{0,1\}^N$, it is crucial to understand the space of real-valued functions on $\{0,1\}^N$, denoted by

$$\mathcal{F} := \left\{ f : \{0,1\}^N \to \mathbb{R} \right\}.$$

It is easily seen that $\mathcal{F}$ is a vector space isomorphic to $\mathbb{R}^{2^N}$. Furthermore, since we are studying $A \sim \mathsf{Ber}(p)$, it is natural to endow $\mathcal{F}$ with the inner product

$$\langle f, g \rangle := \mathbb{E}[f(A)\, g(A)]$$

for $f, g \in \mathcal{F}$ and the canonical norm $\|f\| := \sqrt{\langle f, f \rangle}$.

Next, we construct an orthonormal basis of the inner product space $\mathcal{F}$. Towards this end, define

$$\bar{A}_i := \frac{A_i - p_i}{\sqrt{p_i(1-p_i)}}$$

which is a standardized version of A_i, i.e., $\mathbb{E}[\bar{A}_i] = 0$ and $\mathsf{Var}(\bar{A}_i) = 1$. For each $\alpha \subset [N]$, define

$$\phi_\alpha(A) := \prod_{i \in \alpha} \bar{A}_i$$

and in particular, $\phi_\varnothing(A) := 1$.

Theorem 2.1. *The set $\{\phi_\alpha : \alpha \subset [N]\}$ is an orthonormal basis of the inner product space $\mathcal{F}$.*

Proof. First, we check the orthonormality of $\{\phi_\alpha : \alpha \subset [N]\}$: for $\alpha, \beta \subset [N]$,

$$\mathbb{E}[\phi_\alpha(A)\, \phi_\beta(A)] = \mathbb{E}\left[\prod_{i \in \alpha} \bar{A}_i \cdot \prod_{j \in \beta} \bar{A}_j \right] = \prod_{i \in \alpha \cap \beta} \mathbb{E}[\bar{A}_i^2] \cdot \prod_{i \in \alpha \triangle \beta} \mathbb{E}[\bar{A}_i] = \begin{cases} 1 & \text{if } \alpha = \beta, \\ 0 & \text{if } \alpha \neq \beta. \end{cases}$$

Moreover, $\mathcal{F}$ has dimension 2^N and there are 2^N choices of $\alpha \subset [N]$, so we reach the conclusion. □

In the orthonormal basis, each function $\phi_\alpha(A)$ is in fact a polynomial in the entries $(A_i)_{i \in [N]}$, and the degree of $\phi_\alpha(A)$ is $|\alpha|$. Therefore, the space $\mathcal{F}$ of all real-valued functions on A is spanned by polynomials of degrees at most N by the above theorem. The following theorem provides a more general result.

Theorem 2.2. *For any integer $0 \le D \le N$, let $\mathcal{F}_{\le D}$ denote the set of polynomials in $(A_i)_{i \in [N]}$ that have degrees at most D. Then $\{\phi_\alpha : \alpha \subset [N], |\alpha| \le D\}$ is an orthonormal basis of $\mathcal{F}_{\le D}$.*

Proof. Since we already have the orthonormality of $\{\phi_\alpha : \alpha \subset \binom{[n]}{2}, |\alpha| \le D\}$, it suffices to show that this set of polynomials spans $\mathcal{F}_{\le D}$. This can be proved by induction on D.

The base case $D = 0$ is trivial. For $D \ge 1$, it suffices to note that for every $\alpha \subset [N]$ with $|\alpha| = D$, the monomial $\prod_{i\in\alpha} A_i$ can be expressed as a linear combination of $\phi_\alpha(A) = \prod_{i\in\alpha} \bar{A}_i$ and polynomials in $(A_i)_{i\in[N]}$ of degrees at most $D-1$. □

We now specialize the above results to the setting of random graph [Jan94], in which case we will see that the basis function $\phi_\alpha(A)$ becomes more meaningful. To be more precise, we let $N = \binom{n}{2}$, use the double index $(i,j) \in \binom{[n]}{2}$ in place of the single index $i \in [N]$, and view $A \in \{0,1\}^{n\times n}$ as the adjacency matrix of the observed undirected graph. Any subset $\alpha \subset \binom{[n]}{2}$ can be identified as the subgraph of the complete graph K_n induced by the set of edges α. Fix a template graph H on d vertices where $d \in [n]$. Then the number of copies of H as subgraphs in the graph A, known as the *subgraph count*, is equal to

$$\sum_{\alpha\subset\binom{[n]}{2}, \alpha\cong H} \prod_{i\in\alpha} A_{ij},$$

where $\cong$ denotes *graph isomorphism*. Analogously, the quantity

$$\sum_{\alpha\subset\binom{[n]}{2}, \alpha\cong H} \phi_\alpha(A) = \sum_{\alpha\subset\binom{[n]}{2}, \alpha\cong H} \prod_{i\in\alpha} \bar{A}_{ij}$$

is known as the *signed subgraph count* of copies of H in A. As the above theorems suggest, there is essentially no loss of generality in focusing on (signed) subgraph counts, because they span the entire space of real-valued functions on the given graph.

2.2 Lower bounds for testing

Continuing with the setup of the previous section, we now introduce general frameworks for proving information-theoretic and computational lower bounds for testing.

One way to prove an information-theoretic lower bound is via the *second moment method* which studies the χ^2-divergence between P_0 and P_1, introduced in Section A.1. The following result follows immediately from Lemma A.1 and Theorem A.2.

Theorem 2.3. *For testing between $H_0 : A \sim P_0$ and $H_1 : A \sim P_1$, we have*

$$\inf_\phi \left(\mathbb{P}_0\{\phi(A) = 1\} + \mathbb{P}_1\{\phi(A) = 0\}\right) \ge 1 - \frac{1}{2}\sqrt{\chi^2(P_1, P_0)},$$

where the infimum is over all possible tests ϕ from the sample space to $\{0,1\}$. In particular, if $\chi^2(P_1,P_0) = o(1)$, then $\inf_\phi \left(\mathbb{P}_0\{\phi(A) = 1\} + \mathbb{P}_1\{\phi(A) = 0\}\right) = 1 - o(1)$.

Moreover, if we have $\chi^2(P_1,P_0) \le C$ for a constant $C > 0$, then

$$\inf_\phi \left(\mathbb{P}_0\{\phi(A) = 1\} + \mathbb{P}_1\{\phi(A) = 0\}\right) \ge c$$

for a constant $c = c(C) > 0$. In other words, there does not exist a consistent test.

By the above result, to establish an information-theoretic lower bound against all tests, it suffices to control the χ^2-divergence

$$\chi^2(P_1, P_0) = \mathbb{E}_0\, L^2 - 1 = \mathsf{Var}_0(L) = \|L\|^2 - 1 = \|L - 1\|^2,$$

where we let $L(A) = p_1(A)/p_0(A)$ denote the likelihood ratio and recall that $\|L\|^2 = \langle L, L\rangle$ is the squared norm of L.

Next, we show that, if the goal is to establish a lower bound against polynomial tests of degrees at most D, it suffices to study the projected likelihood ratio $L_{\le D}$. To be more precise, recall that $\mathcal{F}_{\le D}$ denotes the set of polynomials in $(A_i)_{i\in[N]}$ that have degrees at most D. Define the function $L_{\le D}(A)$ to be the projection of the likelihood ratio $L(A)$ onto $\mathcal{F}_{\le D}$, i.e.,

$$L_{\le D} = \sum_{\alpha:|\alpha|\le D} \langle L, \phi_\alpha\rangle\, \phi_\alpha \tag{2}$$

in the notation of Theorem 2.2. We also define the *degree-D χ^2-divergence* between P_1 and P_0 as

$$\chi^2_{\le D}(P_1, P_0) := \|L_{\le D}\|^2 - 1 = \|L_{\le D} - 1\|^2, \tag{3}$$

where the equality holds because $\phi_\varnothing = 1$ and $\langle L_{\le D}, 1\rangle = \langle L, 1\rangle = 1$ for any integer $D \ge 0$. It follows that

$$\begin{aligned}\sqrt{\chi^2_{\le D}(P_1, P_0)} &= \max_{f\in\mathcal{F}_{\le D}\cap 1^\perp,\, \|f\|\le 1} \langle L, f\rangle \\ &= \max_{f\in\mathcal{F}_{\le D},\, \mathbb{E}_0[f]=0,\, \mathsf{Var}_0(f)\le 1} \mathbb{E}_0[L(A)f(A)] = \max_{f\in\mathcal{F}_{\le D},\, \mathbb{E}_0[f]=0,\, \mathsf{Var}_0(f)\le 1} \mathbb{E}_1[f(A)]. \end{aligned} \tag{4}$$

A polynomial $f(A)$ in $(A_i)_{i\in[N]}$ is said to

- *strongly separate* P_0 and P_1 if

$$\sqrt{\max\{\mathsf{Var}_0(f(A)), \mathsf{Var}_1(f(A))\}} = o\Big(\Big|\,\mathbb{E}_1[f(A)] - \mathbb{E}_0[f(A)]\Big|\Big) \quad \text{as } n\to\infty;$$

- *weakly separate* P_0 and P_1 if

$$\sqrt{\max\{\mathsf{Var}_0(f(A)), \mathsf{Var}_1(f(A))\}} = O\Big(\Big|\,\mathbb{E}_1[f(A)] - \mathbb{E}_0[f(A)]\Big|\Big) \quad \text{as } n\to\infty.$$

Note that if $f(A)$ strongly separates P_0 and P_1, say, with $\mathbb{E}_1[f(A)] > \mathbb{E}_0[f(A)]$, then we can take $\tau := \frac{1}{2}\big(\mathbb{E}_1[f(A)] + \mathbb{E}_0[f(A)]\big)$, and by Chebyshev's inequality,

$$\begin{aligned}\mathbb{P}_0\{f(A) > \tau\} &\le \mathbb{P}_0\left\{\big|f(A) - \mathbb{E}_0[f(A)]\big| > \frac{1}{2}\big(\mathbb{E}_1[f(A)] - \mathbb{E}_0[f(A)]\big)\right\} \\ &\le \frac{4\,\mathsf{Var}_0(f(A))}{(\mathbb{E}_1[f(A)] - \mathbb{E}_0[f(A)])^2} = o(1).\end{aligned}$$

Similarly, $\mathbb{P}_1\{f(A) \le \tau\} = o(1)$. Therefore, the test $\phi(A) = \mathbb{1}\{f(A) > \tau\}$ is consistent.

With the above definitions, it is straightforward to prove the following result.

Theorem 2.4. *If $\chi^2_{\le D}(P_1, P_0) \le C$ for a constant $C > 0$, then there exists no polynomial $f \in \mathcal{F}_{\le D}$ that strongly separates P_0 and P_1.*

Moreover, if $\chi^2_{\le D}(P_1, P_0) = o(1)$, then there exists no polynomial $f \in \mathcal{F}_{\le D}$ that weakly separates P_0 and P_1.

Proof. Suppose there is a polynomial $f(A)$ that strongly (respectively, weakly) separates P_0 and P_1. Without loss of generality, we can standardize $f(A)$ under P_0, i.e., $\mathbb{E}_0[f(A)] = 0$ and $\mathsf{Var}_0(f(A)) = 1$. Then we have $|E_1[f(A)]| \to \infty$ as $n \to \infty$ by the strong separation (respectively, $|E_1[f(A)]| \ge c > 0$ by the weak separation). This contradicts the assumption on $\chi^2_{\le D}(P_1, P_0)$, in view of the formula (4). □

A couple of remarks follow. First, we never considered $\mathsf{Var}_1(f(A))$ in the above lower bounds. Nevertheless, $\mathsf{Var}_1(f(A))$ will play an important role when we prove upper bounds. Second, it is also possible to prove lower bounds directly on the type I and type II errors for low-degree polynomials, at the cost of a more involved formulation which we do not present here.

2.3 Statistical-to-computational gap for planted clique detection

Using tools from the last section, we provide evidence supporting the conjectured statistical-to-computational gap for the detection of a planted clique. Before proving the lower bounds, we remark that taking $D = \text{polylog}(n)$ typically yields a good prediction of the conjectured computational threshold of a problem. Therefore, we mean a scaling of degree D polylogarithmic in n when speaking of a "low-degree" polynomial.

In view of Theorems 2.3 and 2.4, to prove information-theoretic and computational lower bounds for detecting a planted clique, it suffices to bound $\chi^2(P_1, P_0)$ and $\chi^2_{\le D}(P_1, P_0)$ respectively. Recall that $\{\phi_\alpha : \alpha \subset \binom{[n]}{2}, |\alpha| \le D\}$ is an orthonormal basis of $\mathcal{F}_{\le D}$ by Theorem 2.2. By (2) and (3),

$$\chi^2_{\le D}(P_1, P_0) = \sum_{\alpha: 1 \le |\alpha| \le D} \langle L, \phi_\alpha \rangle^2 = \sum_{\alpha: 1 \le |\alpha| \le D} \mathbb{E}_0[L(A)\, \phi_\alpha(A)]^2 = \sum_{\alpha: 1 \le |\alpha| \le D} \mathbb{E}_1[\phi_\alpha(A)]^2. \quad (5)$$

Moreover, for a random graph A on n vertices, the degree D is at most $\binom{n}{2}$, and $\{\phi_\alpha : \alpha \subset \binom{[n]}{2}\}$ is an orthonormal basis of the set of all functions of A. Hence we have $\chi^2(P_1, P_0) = \chi^2_{\le \binom{n}{2}}(P_1, P_0)$. Theorem 2.5 below consists of the main negative results for testing in the planted clique model. (The low-degree calculation for second result in Theorem 2.5 can be found in [Hop18], and I am not aware of a reference for proving the first result in this particular way.)

Theorem 2.5. *Let $P_0 = G(n, 1/2)$ and $P_1 = G(n, 1/2, k)$. Denote the likelihood ratio by $L = p_1/p_0$. Then we have the following results:*

- *If $k \le (2 - \epsilon) \log_2 n$ for a fixed constant $\epsilon \in (0, 2)$, then $\chi^2(P_1, P_0) = o(1)$.*
- *If $k \le n^{1/2-\epsilon}$ for a fixed constant $\epsilon \in (0, 1/2)$ and $D = o\big((\frac{\log n}{\log \log n})^2\big)$, then $\chi^2_{\le D}(P_1, P_0) = o(1)$.*

Proof. In view of (5), we need to bound $\mathbb{E}_1[\phi_\alpha(A)]$ for $\alpha \subset \binom{[n]}{2}$. Recall that under P_1, the vertex set of the clique, denoted by $\mathcal{C}$, is a uniformly random subset of $[n]$ with a fixed size $|\mathcal{C}| = k$. If

either i or j is not in $\mathcal{C}$, then conditionally $A_{ij} \sim \mathsf{Ber}(1/2)$ and $\mathbb{E}[\bar{A}_{ij} \mid \mathcal{C}] = 0$; otherwise, $A_{ij} = 1$ and $\bar{A}_{ij} = \frac{A_{ij} - 1/2}{\sqrt{(1/2)\cdot(1/2)}} = 1$. Therefore, by the independence of A_{ij} conditional on $\mathcal{C}$, we have

$$\mathbb{E}_1[\phi_\alpha(A)] = \mathbb{E}\left[\prod_{(i,j)\in\alpha} \mathbb{E}_1[\bar{A}_{ij} \mid \mathcal{C}]\right] = \mathbb{P}\{i, j \in \mathcal{C} \text{ for all } (i,j) \in \alpha\}$$

which is the probability that $\mathcal{C}$ contains all vertices of α viewed as a graph. Let $v(\alpha)$ denote the number of vertices of α. If $v(\alpha) > k$, the above probability is obviously zero. If $v(\alpha) \le k$, we have

$$\mathbb{E}_1[\phi_\alpha(A)] = \frac{\binom{n-v(\alpha)}{k-v(\alpha)}}{\binom{n}{k}} = \frac{k(k-1)\cdots(k-v(\alpha)+1)}{n(n-1)\cdots(n-v(\alpha)+1)} \le (k/n)^{v(\alpha)}. \tag{6}$$

- First, consider the case $D = \binom{n}{2}$ so that $\chi^2(P_1, P_0) = \chi^2_{\le D}(P_1, P_0)$. Then we have

$$\begin{aligned}\chi^2(P_1, P_0) &= \sum_{\alpha: 1 \le |\alpha| \le \binom{n}{2}} \mathbb{E}_1[\phi_\alpha(A)]^2 \\ &\le \sum_{\alpha: 2 \le v(\alpha) \le k} (k/n)^{2v(\alpha)} = \sum_{m=2}^{k} \sum_{\alpha: v(\alpha)=m} (k/n)^{2m} \le \sum_{m=2}^{k} n^m 2^{km/2} (k/n)^{2m},\end{aligned}$$

  where the last step holds because there are at most $\binom{n}{m} 2^{\binom{m}{2}} \le n^m 2^{m^2/2} \le n^m 2^{km/2}$ graphs α with $v(\alpha) = m$. Furthermore, if $k \le (2-\epsilon)\log_2 n$, then

$$n 2^{k/2} (k/n)^2 \le n\, n^{1-\epsilon/2} \Big(\frac{2\log_2 n}{n}\Big)^2 = \frac{(2\log_2 n)^2}{n^{\epsilon/2}} = o(1).$$

  We conclude that

$$\chi^2(P_1, P_0) \le \sum_{m=2}^{k} (o(1))^m = o(1).$$

- Next, consider the low-degree case where $D = o\big((\frac{\log n}{\log\log n})^2\big)$. For brevity, we assume that $\sqrt{D}$ is an integer. For $m \le 2\sqrt{D}$, there are at most $\binom{n}{m} 2^{\binom{m}{2}} \le n^m 2^{m^2} \le n^m 2^{m\sqrt{D}}$ graphs α such that $v(\alpha) = m$. For $2\sqrt{D} < m \le 2D$, there are at most $\binom{n}{m}\binom{m}{2}^D \le n^m m^{2D}$ graphs α such that $v(\alpha) = m$ and $|\alpha| \le D$. It follows that

$$\begin{aligned}\chi^2_{\le D}(P_1, P_0) &\le \sum_{\alpha: 1 \le |\alpha| \le D} (k/n)^{2v(\alpha)} = \sum_{m=2}^{2D} \sum_{v(\alpha)=m,\, |\alpha| \le D} (k/n)^{2m} \\ &\le \sum_{m=2}^{2\sqrt{D}} n^m 2^{m^2} (k/n)^{2m} + \sum_{m=2\sqrt{D}}^{2D} n^m m^{2D} (k/n)^{2m}.\end{aligned}$$

  For the first term, note that for $D = o\big((\frac{\log n}{\log\log n})^2\big)$ and $k \le n^{1/2-\epsilon}$, we have

$$n 2^{2\sqrt{D}} (k/n)^2 \le n e^{o(\log n)} (k/n)^2 \le n^{1+o(1)} n^{-1-2\epsilon} = o(1).$$

Therefore,

$$\sum_{m=2}^{2\sqrt{D}} n^m 2^{m^2} (k/n)^{2m} \le \sum_{m=2}^{2\sqrt{D}} (n 2^{2\sqrt{D}} (k/n)^2)^m = o(1).$$

For the second term, note that for $m = 2\sqrt{D}$, we have

$$\begin{aligned} n^{2\sqrt{D}} (2\sqrt{D})^{2D} (k/n)^{4\sqrt{D}} &= \big(n(k/n)^2 (2\sqrt{D})^{\sqrt{D}}\big)^{2\sqrt{D}} \\ &\le \big((k^2/n)(\log n)^{o(\frac{\log n}{\log\log n})}\big)^{2\sqrt{D}} \le \big(n^{-2\epsilon} n^{o(1)}\big)^{2\sqrt{D}} = o(1). \end{aligned}$$

Moreover, for $2\sqrt{D} \le m < 2D$, we have

$$\begin{aligned} \frac{n^{m+1}(m+1)^{2D}(k/n)^{2(m+1)}}{n^m m^{2D}(k/n)^{2m}} &\le n(k/n)^2 \Big(1 + \frac{1}{2\sqrt{D}}\Big)^{2D} \\ &\le (k^2/n)\, e^{\sqrt{D}} \le n^{-2\epsilon} e^{o(\log n)} \le n^{-2\epsilon} n^{o(1)} = o(1). \end{aligned}$$

We conclude that

$$\sum_{2\sqrt{D}}^{2D} n^m m^{2D} (k/n)^{2m} = o(1).$$

The two terms combined yield that $\chi^2_{\le D}(P_1, P_0) = o(1)$.

□

It is then straightforward to combine Theorem 2.5 with Theorems 2.3 and 2.4 to obtain information-theoretic and computational lower bounds. Recall that Corollary 1.5 provides an information-theoretic upper bound. Moreover, in Corollary 1.7, the total number of edges in A, denoted by $d(A)$, is used as the test statistic to give a computational upper bound. It is easily seen that the degree-1 polynomial $d(A)$ satisfies

$$\sqrt{\max\{\mathsf{Var}_0(d(A)), \mathsf{Var}_1(d(A))\}} \le \sqrt{\binom{n}{2} \cdot \frac{1}{2} \cdot \frac{1}{2}} = o\left(\frac{1}{2}\binom{k}{2}\right) = o\Big(\Big|\, \mathbb{E}_1[d(A)] - \mathbb{E}_0[d(A)]\Big|\Big)$$

if $k \gg \sqrt{n}$, so $d(A)$ strongly separates P_0 and P_1 by definition in this regime. The statistical-to-computational gap for planted clique detection is summarized as follows.

Theorem 2.6. *Consider testing between $H_0 : A \sim G(n, 1/2)$ and $H_1 : A \sim G(n, 1/2, k)$.*

- *If $k \ge (2+\epsilon)\log_2 n$ for a fixed $\epsilon > 0$, then there is a consistent test.*
- *If $k \le (2-\epsilon)\log_2 n$ for a fixed $\epsilon > 0$, then there is no consistent test.*
- *If $k \gg n^{1/2}$, then there is a polynomial in $(A_{ij})_{i<j}$ of degree 1 that strongly separates P_0 and P_1.*
- *If $k \le n^{1/2-\epsilon}$ for a fixed $\epsilon > 0$, then there is no polynomial in $(A_{ij})_{i<j}$ of degree $o\big((\frac{\log n}{\log\log n})^2\big)$ that weakly separates P_0 and P_1.*

Remark 2.7. *The information-theoretic part of Theorem 2.5, i.e., the bound on $\chi^2(P_1, P_0)$, can be proved by analyzing the likelihood ratio directly. Let $p_1(A \mid \mathcal{C})$ denote the density of $A \sim P_1$ conditional on $\mathcal{C}$, so that*

$$p_1(A) = \frac{1}{\binom{n}{k}} \sum_{\mathcal{C} \subset [n], |\mathcal{C}|=k} p_1(A \mid \mathcal{C}).$$

Then we have

$$L(A) = \frac{p_1(A)}{p_0(A)} = \frac{\frac{1}{\binom{n}{k}} \cdot 2^{-\binom{n}{2}+\binom{k}{2}} \cdot T_k(A)}{2^{-\binom{n}{2}}} = \frac{T_k(A)}{\binom{n}{k} \cdot 2^{-\binom{k}{2}}} = \frac{T_k(A)}{\mathbb{E}_0[T_k(A)]}, \tag{7}$$

where $T_k(A)$ denotes the number of cliques of size k in the observed graph A. It follows that

$$\chi^2(P_1, P_0) = \mathbb{E}_0[(L(A) - 1)^2] = \frac{\mathsf{Var}(T_k(A))}{(\mathbb{E}_0[T_k(A)])^2}.$$

Therefore, it remains to understand the first two moments of the polynomial $T_k(A)$ of degree $\binom{k}{2}$. We omit the rest of the proof. This proof is in fact more "standard" and somewhat simpler than that of Theorem 2.5 because we only need to analyze one polynomial rather than all of them.

A Information-theoretic tools

A.1 Divergences between probability distributions

Let p and q denote the probability density (or mass) functions of distributions P and Q respectively. Suppose that P is absolutely continuous with respect to Q. Then p/q is the Radon–Nikodym derivative. For any convex function $f : (0, \infty) \to \mathbb{R}$ such that $f(1) = 0$ and f is strictly convex at 1 (i.e., for any $s, t > 0$ and $\lambda \in (0, 1)$ such that $\lambda s + (1-\lambda)t = 1$, we have $\lambda f(s) + (1-\lambda) f(t) > f(1)$), the *$f$-divergence* between P and Q is defined as

$$D_f(P \| Q) := \mathbb{E}_0 \left[f\Big(\frac{p(X)}{q(X)} \Big) \right] = \int q \cdot f(p/q).$$

Using Jensen's inequality, it is easy to check that $D_f(P\|Q) \geq 0$ where equality holds if and only if $P = Q$. We consider two special f-divergences:

- *Total variation distance*: $\mathsf{TV}(P, Q) = \frac{1}{2} \int |p - q|$, i.e., $f(x) = \frac{1}{2}|x - 1|$.
- *χ^2-divergence*: $\chi^2(P, Q) = \int (p - q)^2/q$, i.e., $f(x) = (x - 1)^2$.

Note that a "divergence" is not necessarily symmetric, so we only use the term "distance" when the divergence is indeed symmetric. The following result is well-known and frequently used in the literature.

Lemma A.1. *We have*

$$\mathsf{TV}(P, Q) \leq \frac{1}{2} \sqrt{\chi^2(P, Q)}.$$

Moreover, if $\chi^2(P, Q) \leq C$ for a constant $C > 0$, then there is a constant $c = c(C) > 0$ such that

$$\mathsf{TV}(P, Q) \leq 1 - c.$$

Proof. For the first bound, it suffices to note that

$$2\,\mathsf{TV}(P,Q) = \mathbb{E}_0\left|\frac{p}{q}-1\right| \le \sqrt{\mathbb{E}_0\left[\left(\frac{p}{q}-1\right)^2\right]} = \sqrt{\chi^2(P,Q)}.$$

For the second result, note that $1-\mathsf{TV}(P,Q) = \int \min(p,q)$. We have

$$\begin{aligned}\left(\int\sqrt{pq}\right)^2 &= \left(\int\sqrt{\min(p,q)\cdot\max(p,q)}\right)^2 \\ &\le \int\min(p,q)\cdot\int\max(p,q) \le \int\min(p,q)\cdot\int(p+q) = 2\int\min(p,q),\end{aligned}$$

where we used the Cauchy–Schwarz inequality. Moreover,

$$\begin{aligned}\left(\int\sqrt{pq}\right)^2 &= \exp\left(2\log\int\sqrt{pq}\right) = \exp\left(2\log\int_{pq>0}p\sqrt{\frac{q}{p}}\right) \\ &\ge \exp\left(2\int_{pq>0}p\log\sqrt{\frac{q}{p}}\right) = \exp\left(-\int_{pq>0}p\log\frac{p}{q}\right),\end{aligned}$$

where we used Jensen's inequality. Applying Jensen's inequality again, we obtain

$$\int_{pq>0}p\log\frac{p}{q} \le \log\int_{pq>0}p\frac{p}{q} = \log\int q\Big(\frac{p}{q}\Big)^2 = \log(\chi^2(P,Q)+1).$$

Combining everything, if $\chi^2(P,Q)\le C$, then $\int\sqrt{pq}\ge c'$ and so $\int\min(p,q)\ge c$ for constants $c,c'>0$. We conclude that $\mathsf{TV}(P,Q)\le 1-c$. □

We remark that the χ^2-divergence enjoys a property called *tensorization*: The χ^2-divergence between two product distributions can be easily calculated from the χ^2-divergences between individual pairs of component distributions. To be more precise, let $P=\otimes_{i=1}^n P_i$ (i.e., P is the joint distribution of $(X_1,\dots,X_n)$, where $X_i\sim P_i$ independently for $i=1,\dots,n$) and $Q=\otimes_{i=1}^n Q_i$. It is easy to verify that

$$\chi^2(P,Q)+1 = \prod_{i=1}^n\big(\chi^2(P_i,Q_i)+1\big).$$

There is no such tensorization property for the total variation distance. Therefore, when bounding the total variation distance between two product distributions, it is often convenient to bound it by the χ^2-divergence and use the tensorization property.

A.2 Neyman–Pearson lemma

Consider testing between two hypotheses $H_0: A\sim P_0$ and $H_1: A\sim P_1$. Let p_0 and p_1 denote their respective densities (or masses).

Theorem A.2 (Neyman–Pearson)**.** *For testing between $H_0: A\sim P_0$ and $H_1: A\sim P_1$, we have*

$$\inf_{\phi}\big(\mathbb{P}_0\{\phi(A)=1\}+\mathbb{P}_1\{\phi(A)=0\}\big) = 1-\mathsf{TV}(P_0,P_1),$$

where the infimum is over all possible tests ϕ from the sample space to $\{0,1\}$ and TV denotes the total variance distance

$$\mathsf{TV}(P_0, P_1) = \frac{1}{2}\int |p_0 - p_1| = 1 - \int \min(p_0, p_1).$$

Moreover, the infimum is achieved by the likelihood ratio test $\phi^* := \mathbb{1}\{p_1/p_0 \geq 1\}$.

Proof. The fact $\mathsf{TV}(P_0, P_1) = 1 - \int \min(p_0, p_1)$ is easy to check. Moreover, we have

$$\begin{aligned}
\mathbb{P}_0\{\phi^* = 1\} + \mathbb{P}_1\{\phi^* = 0\} &= \int_{\{\phi^*=1\}} p_0 + \int_{\{\phi^*=0\}} p_1 \\
&= \int_{\{p_1 \geq p_0\}} p_0 + \int_{\{p_1 < p_0\}} p_1 \\
&= \int_{\{p_1 \geq p_0\}} \min(p_0, p_1) + \int_{\{p_1 < p_0\}} \min(p_0, p_1) \\
&= \int \min(p_0, p_1) = 1 - \mathsf{TV}(P_0, P_1).
\end{aligned}$$

For any test ϕ, define $R := \{\phi = 1\}$. Let $R^* := \{p_1 \geq p_0\}$. Then we have

$$\begin{aligned}
\mathbb{P}_0\{\phi = 1\} + \mathbb{P}_1\{\phi = 0\} &= \mathbb{P}_0\{R\} + 1 - \mathbb{P}_1\{R\} \\
&= 1 + \int_R (p_0 - p_1) \\
&= 1 + \int_{R \cap R^*} (p_0 - p_1) + \int_{R \cap (R^*)^c} (p_0 - p_1) \\
&= 1 - \int_{R \cap R^*} |p_0 - p_1| + \int_{R \cap (R^*)^c} |p_0 - p_1| \\
&= 1 - \int |p_0 - p_1| \big(\mathbb{1}\{R \cap R^*\} - \mathbb{1}\{R \cap (R^*)^c\} \big),
\end{aligned}$$

which is minimized at $R = R^*$. □

References

[AKS98] Noga Alon, Michael Krivelevich, and Benny Sudakov. Finding a large hidden clique in a random graph. *Random Structures & Algorithms*, 13(3-4):457–466, 1998.

[Hop18] Samuel Hopkins. *Statistical inference and the sum of squares method.* Cornell University, 2018.

[Jan94] Svante Janson. *Orthogonal decompositions and functional limit theorems for random graph statistics*, volume 534. American Mathematical Soc., 1994.

[Jer92] Mark Jerrum. Large cliques elude the metropolis process. *Random Structures & Algorithms*, 3(4):347–359, 1992.

[Wei25] Alexander S Wein. Computational complexity of statistics: New insights from low-degree polynomials. *arXiv preprint arXiv:2506.10748*, 2025.