Computer Science 2: Advanced Algorithms Fall 2026

Lecture 6 — Thursday, September 10th, 2026

Prof. Jarosław Błasiok *Scribe: Davide Secco*

1 Overview

In the last lecture we studied undirected s-t connectivity in logarithmic space, using expander graphs and derandomized squaring as the key tools.

In this lecture we study *average-case complexity*, using the *planted clique problem* as the guiding example. We show that the problem is statistically solvable as soon as $k > \lceil 2\log_2 n \rceil$, and we give two polynomial-time algorithms that work for much larger cliques:

- an elementary *degree-based* algorithm, which succeeds for $k \geq C\sqrt{n \log n}$;
- a *spectral* algorithm, which succeeds for $k \geq C\sqrt{n}$.

Along the way we develop the tools needed for the analysis: concentration inequalities (Markov, Chebyshev, moment-generating functions, subgaussian random variables and Hoeffding's inequality), ε-nets of the unit sphere, and a bound of order $\sqrt{n}$ on the operator norm of a random sign matrix.

2 Worst-case and average-case complexity

Classically, a computational problem specifies

- a set of valid inputs;
- a set of acceptable outputs for each input.

A worst-case algorithm must return an acceptable answer for *every* valid input, so a single pathological input can determine the worst-case complexity.

In many situations this is not quite satisfying. A problem may be hard on its worst-case inputs while being much easier on the inputs that actually arise in applications, on inputs drawn from a specified random model, or on inputs with some additional structure. Conversely, if we care about *hardness*, for instance to build a cryptographic system, keys are generated at random: it is not enough that breaking the system is hard for the worst key, we need it to be hard for a random key.

In the first lecture we looked at worst-case inputs and introduced randomness in the *algorithm*. Here the *input* itself is random.

Definition 1 (Distributional problem)**.** *A distributional problem specifies, together with the problem, a probability distribution $\mathcal{D}_n$ over inputs of size n. An algorithm A solves it if it succeeds on a typical draw:*

$$\mathop{\mathbb{P}}_{X\sim\mathcal{D}_n,\ \text{randomness of } A}\big[A(X) \text{ is acceptable}\big] \geq \frac{2}{3}.$$

Remark 2. *It is instructive to compare Definition 1 with a worst-case randomized algorithm, where the input x is fixed first (possibly by an adversary) and the only randomness is in the coins of A:*

$$\forall x: \qquad \mathop{\mathbb{P}}_{\text{randomness of } A}\big[A(x) \text{ is acceptable}\big] \geq \frac{2}{3}.$$

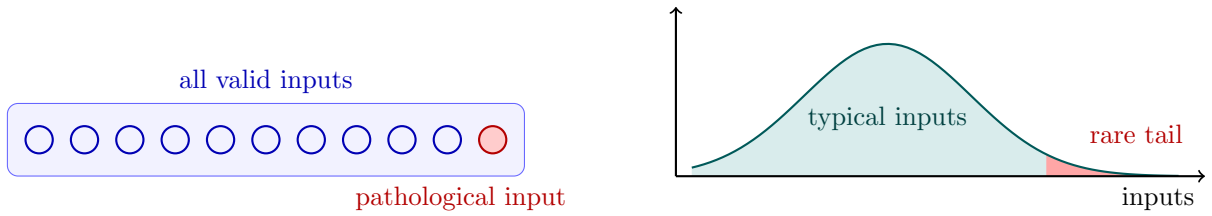
In Definition 1 there is no "for all x": the input $X \sim \mathcal{D}_n$ is random too, and only the overall *success probability is required to be at least $2/3$. Hence the algorithm may fail on some inputs, even always, as long as these inputs have small total probability under $\mathcal{D}_n$ (Figure 1).*

For example, suppose that under $\mathcal{D}_n$ 90% of the inputs are "good", and on them A is correct with probability 0.8, while on the remaining 10% of "bad" inputs A is always wrong. Then

$$\mathop{\mathbb{P}}_{X\sim\mathcal{D}_n,\ \text{randomness of } A}\big[A(X) \text{ is acceptable}\big] = 0.9 \cdot 0.8 + 0.1 \cdot 0 = 0.72 \geq \frac{2}{3},$$

so A satisfies Definition 1, but not the worst-case requirement, since on bad inputs its success probability is 0.

The example also shows an important difference. For worst-case randomized algorithms (say, for a decision problem) the constant $2/3$ is arbitrary: repeating A independently and taking the majority answer boosts the success probability to almost 1 on every *input. In the example, repetition boosts the success probability on good inputs to almost 1, but on bad inputs A is wrong every time, so the overall success probability can never exceed 0.9. Errors due to the coins of A can be removed by repetition; errors due to bad inputs cannot.*



(a) worst case (b) average case: $X \sim \mathcal{D}_n$

Figure 1: (a) A worst-case algorithm must be correct on every valid input, so a single pathological input determines its complexity. (b) An average-case algorithm only needs to succeed on a typical draw from $\mathcal{D}_n$; it may fail on a set of inputs of small total probability.

Worst-case analysis remains essential, but it can hide useful structure: realistic inputs may follow a specific random model, pathological instances may have exponentially small probability, and typical instances may contain statistical patterns that an algorithm can exploit. Average-case analysis asks two separate questions:

1. Can the distributions be distinguished *at all*?
2. Can they be distinguished *in polynomial time*?

The planted clique problem makes the gap between these two questions especially clear.

3 The planted clique problem

3.1 The two distributions

Null distribution. Under the null hypothesis

$$H_0 : \qquad G \sim G(n, 1/2),$$

i.e. G is an Erdős–Rényi graph: for every unordered pair $\{i, j\}$ of distinct vertices we flip an independent fair coin, and $\{i, j\}$ is an edge if it comes up heads.

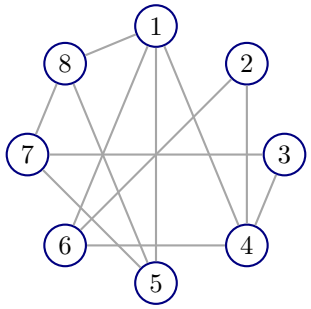


Figure 2: A sample of $G(8, 1/2)$: each of the $\binom{8}{2} = 28$ pairs is an edge independently with probability $1/2$.

Planted distribution. Under the planted hypothesis

$$H_1 : \qquad G \sim \mathcal{P}_1(n, k),$$

where G is generated as follows:

1. sample $G_0 \sim G(n, 1/2)$;
2. choose $S \subseteq [n]$ uniformly at random among all k-element sets, independently of G_0;
3. add every missing edge inside S, so that S becomes a clique.

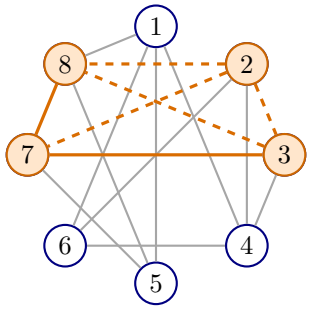


Figure 3: The planted distribution with $k = 4$, starting from the graph G_0 of Figure 2 and $S = \{2, 3, 7, 8\}$: solid orange edges inside S were already present in G_0, dashed orange edges are added to make S a clique.

3.2 Decision and recovery

There are two associated computational questions.

Decision problem. Construct a polynomial-time algorithm A such that

$$\mathop{\mathbb{P}}_{G\sim G(n,1/2)}[A(G)=0]\geq\frac{2}{3}\qquad\text{and}\qquad\mathop{\mathbb{P}}_{G\sim\mathcal{P}_1(n,k)}[A(G)=1]\geq\frac{2}{3}.$$

In simple words: we are shown a single graph, without being told how it was generated, and we must say which of the two random processes produced it. We answer 0 if we think it is a completely random graph, and 1 if we think that a clique of size k has been hidden in it. The algorithm has to be right with probability at least $2/3$ in *both* cases. Both conditions are needed: the algorithm that always answers 1 is always right on planted graphs but always wrong on random ones, while a random guess is right only with probability $1/2$.

Recovery problem. Given $G\sim\mathcal{P}_1(n,k)$, output the planted set S with probability at least $2/3$.

In simple words: now we know that a clique has been hidden in the graph, and we want to find it, i.e. to point out exactly which k vertices form S. Intuitively, this is harder than the decision problem: finding the clique is more than just noticing that there is one.

3.3 Statistical detectability

We first ask whether the two distributions can be told apart *at all*, allowing the distinguishing algorithm to use an unlimited amount of computation (even exponential time). This is a purely information-theoretic question: does a graph sampled from H_0 carry enough of a statistical signature that no graph from H_1 could plausibly share it, regardless of how hard that signature is to compute? We answer this question first, and only afterwards, in Sections 4 and 7, do we ask for a polynomial-time algorithm.

If we do not care about running time, the two distributions are very far from each other as soon as k is a bit larger than $2\log_2 n$. The reason is that large cliques are extremely rare under the null distribution.

Here is where the number $2\log_2 n$ comes from. The next few paragraphs give a slower, more intuitive derivation of this threshold; a direct proof is Lemma 3 below.

Fix a candidate set A of r vertices. It is a clique exactly when all $\binom{r}{2}=\frac{r(r-1)}{2}$ pairs inside it are edges: choose the first vertex in r ways and the second, distinct from it, in $r-1$ ways, giving $r(r-1)$ ordered pairs, then divide by 2 since $\{i,j\}$ and $\{j,i\}$ are the same edge.

Each of these pairs is an edge independently with probability $\frac{1}{2}$, so all of them are edges together with probability

$$\mathbb{P}[A\text{ is a clique}]=\left(\frac{1}{2}\right)^{\binom{r}{2}}=2^{-\binom{r}{2}}=2^{-\frac{r(r-1)}{2}}\underset{\substack{(r-1)\approx r\\ \text{for large } r}}{\approx}2^{-\frac{r^2}{2}}.$$

There are $\binom{n}{r}$ candidate sets A of size r. Expanding the binomial coefficient gives the crude count (Lemma 3 below keeps track of $r!$ exactly, without this shortcut):

$$\binom{n}{r} = \frac{n!}{r!\,(n-r)!} = \frac{\overbrace{n(n-1)(n-2)\cdots(n-r+1)}^{\substack{r \text{ factors, each} \\ n-i \approx n \ (r \ll n)}}}{\underbrace{r!}_{\substack{\text{drop: negligible} \\ \text{only on log. scale}}}} \approx n^r.$$

Let X_r be the number of r-vertex cliques in G:

$$\mathbb{E}[X_r] = \underbrace{\binom{n}{r}}_{\text{number of sets } A} \cdot \underbrace{2^{-\binom{r}{2}}}_{\mathbb{P}[A \text{ is a clique}]} \approx n^r\, 2^{-\frac{r^2}{2}}.$$

Since:

$$n = 2^{\log_2 n}, \longrightarrow n^r = \left(2^{\log_2 n}\right)^r = 2^{r \log_2 n}$$

Then:

$$\mathbb{E}[X_r] \approx n^r\, 2^{-\frac{r^2}{2}} = 2^{r \log_2 n} \cdot 2^{-\frac{r^2}{2}} = 2^{f(r)}, \qquad f(r) := r \log_2 n - \frac{r^2}{2},$$

$$\begin{cases} r \log_2 n & \text{reward: more candidate sets} \\ \dfrac{r^2}{2} & \text{penalty: harder to be a clique} \end{cases}$$

This is a competition between two terms that both grow with r, but at different rates: the reward $r \log_2 n$ grows *linearly* in r (more candidate sets), while the penalty $r^2/2$ grows *quadratically* in r (exponentially more edges are required). Since $x \mapsto 2^x$ is increasing and $2^0 = 1$, the sign of this competition alone tells us whether $\mathbb{E}[X_r]$ is large or small:

- if $f(r) > 0$ (reward wins), then $\mathbb{E}[X_r] \approx 2^{f(r)} > 1$: many r-cliques are expected;
- if $f(r) < 0$ (penalty wins), then $\mathbb{E}[X_r] \approx 2^{f(r)} < 1$: essentially no r-clique is expected.

A linear term is eventually overtaken by a quadratic one, no matter its slope, so $f(r)$ must turn negative once r is large enough. Factoring out r,

$$f(r) = r \log_2 n - \frac{r^2}{2} = r\left(\log_2 n - \frac{r}{2}\right),$$

shows that f is a downward-opening parabola in r (leading coefficient $-\frac{1}{2}$), with roots at $r = 0$ and at the point where reward and penalty exactly balance:

$$r \log_2 n = \frac{r^2}{2} \iff \log_2 n = \frac{r}{2} \iff r = 2 \log_2 n.$$

A downward parabola is positive strictly between its two roots and negative outside them, so $f(r) > 0$ for $0 < r < 2 \log_2 n$ (reward wins: many cliques expected) and $f(r) < 0$ for $r > 2 \log_2 n$ (penalty wins: essentially none expected).

So $r = 2 \log_2 n$ is the threshold: below it, $f(r) > 0$ and cliques of that size are expected in abundance; above it, $f(r) < 0$ and they essentially never occur. The lemma below makes this precise, replacing the approximations $\binom{n}{r} \approx n^r$ and $\binom{r}{2} \approx \frac{r^2}{2}$ with exact bounds.

Lemma 3. *Let $\omega(G)$ denote the size of the largest clique of G. Let $G \sim G(n, 1/2)$ and $r = \lceil 2\log_2 n \rceil + 1$. Then $\mathbb{P}[\omega(G) \geq r] \leq \frac{1}{n}$. In particular, with high probability $\omega(G) \leq \lceil 2\log_2 n \rceil$.*

Proof. Let X_r be the number of r-vertex cliques in G. A fixed r-element set is a clique if and only if all its $\binom{r}{2}$ internal edges are present, which happens with probability $2^{-\binom{r}{2}}$. By linearity of expectation (summing over all $\binom{n}{r}$ possible positions of the clique),

$$\mathbb{E}[X_r] = \binom{n}{r} 2^{-\binom{r}{2}}.$$

By Markov's inequality (Lemma 7 below),

$$\mathbb{P}[\omega(G) \geq r] = \mathbb{P}[X_r \geq 1] \leq \mathbb{E}[X_r] = \binom{n}{r} 2^{-\binom{r}{2}} \leq \frac{1}{r!} \left(n\, 2^{-\frac{r-1}{2}} \right)^r \leq \frac{1}{r!} \leq \frac{1}{n},$$

where we used $\binom{n}{r} \leq \frac{n^r}{r!}$, then $r - 1 \geq 2\log_2 n$, and finally $r! \geq 2^{r-1} \geq n^2$. □

Under the planted distribution the graph always contains a clique of size k. Therefore, for

$$k > \lceil 2\log_2 n \rceil,$$

the following exhaustive algorithm distinguishes the two distributions: compute a largest clique of G, and output H_1 if its size is at least k (and H_0 otherwise). Under H_1 it is always correct, and under H_0 it errs with probability at most $1/n$ by Lemma 3.

This establishes *statistical* (information-theoretic) distinguishability, but maximum clique search takes exponential time in the worst case. The main question is whether the same can be done *efficiently*.

Remark 4. *Similarly, if running time does not matter, recovery can be attempted by listing all k-cliques of a planted instance: if the planted clique is the unique k-clique of G, this identifies S. The bound of Lemma 3 only gives a sufficient condition for decision; by itself it does not prove uniqueness of the planted clique.*

Remark 5. *The NP-completeness of CLIQUE is not relevant here: it is a statement about worst-case inputs, while here the inputs follow two precise distributions.*

3.4 The statistical and computational thresholds are far apart

Planted clique is one of the central open problems in average-case complexity: statistical distinguishability follows from the simple argument above, but no polynomial-time algorithm is known well below $\sqrt{n}$.

Conjecture 6 (Planted clique conjecture)**.** *For every fixed $0 < \delta < \frac{1}{2}$, there is no polynomial-time algorithm solving the decision problem for $k = n^{\frac{1}{2}-\delta}$.*

In this regime the problem is information-theoretically easy but conjectured to be computationally hard. If planted clique is hard, then suitable reductions transfer this hardness to other average-case problems, which is why it is used as a basic hardness assumption. The conjecture goes back to the works of Jerrum and Kučera.

Today we discuss two algorithms: an elementary degree algorithm for $k \geq C\sqrt{n \log n}$ and a spectral algorithm for $k \geq C'\sqrt{n}$, with sufficiently large constants C, C'.

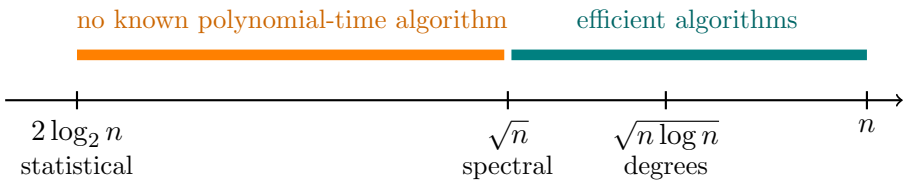


Figure 4: Thresholds for the planted clique problem, as a function of the clique size k.

4 A degree-based algorithm

A very simple first idea: planting a clique creates more edges. One could count all edges and ask when the increase caused by the clique exceeds the random fluctuations of the null model. We can do better by using the same idea *locally*: vertices of the clique should have larger degree, so we can look at the vertices of largest degree.

Expected degrees.
Under the null distribution, for every vertex v,

$$\deg(v) \sim \mathrm{Bin}(n-1, 1/2), \qquad \mu_0 := \mathbb{E}[\deg(v)] = \frac{n-1}{2}.$$

Under the planted distribution:

- if $v \notin S$, its degree distribution is still $\mathrm{Bin}(n-1, 1/2)$;
- if $v \in S$, it has $k-1$ certain edges inside S and $n-k$ possible edges outside S, each present with probability $1/2$, so

$$\deg(v) = \underbrace{(k-1)}_{\substack{\text{certain edges} \\ \text{inside the clique}}} + \underbrace{\mathrm{Bin}(n-k, 1/2)}_{\substack{\text{random edges} \\ \text{towards } [n]\setminus S}},$$

$$\mu_1 := \mathbb{E}[\deg(v)] = k - 1 + \frac{n-k}{2} = \frac{n+k-2}{2}.$$

Comparing to the null mean $\mu_0 = \frac{n-1}{2}$, this is

$$\mu_1 = \mu_0 + \frac{k-1}{2}.$$

So the mean shift is $(k-1)/2$, which grows with k. However, different means alone are not enough: the two degree distributions overlap unless the signal is large (Figure 5). To justify a degree threshold we need a high-probability bound on the *width* of each distribution, i.e. we need to show that degrees are close to their expectations. This is what concentration inequalities do.

5 Concentration inequalities

Concentration inequalities concern random variables that are sums of independent random variables (or depend on them in a weak form), whose expectation is typically easy to compute. They say that with high probability such a random variable is very close to its expectation.

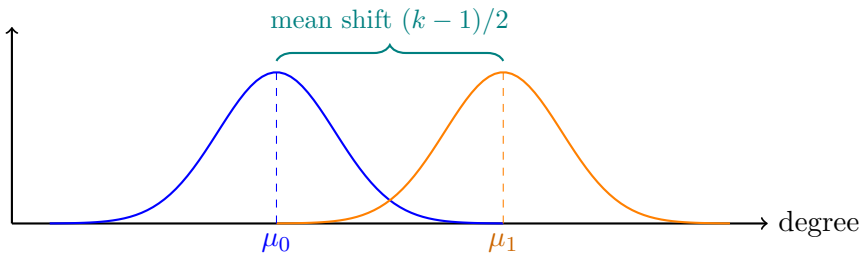


Figure 5: Degree distribution of a vertex outside S (left) and inside S (right).

5.1 Markov and Chebyshev

Lemma 7 (Markov's inequality)**.** *If $X \geq 0$ and $a > 0$, then*

$$\mathbb{P}[X \geq a] \leq \frac{\mathbb{E}[X]}{a}.$$

Meaning of the symbols:

- $X \geq 0$*: the random variable X takes only nonnegative values;*
- $a > 0$*: any threshold we choose;*
- $\mathbb{P}[X \geq a]$*: the probability that X is at least a;*
- $\mathbb{E}[X]$*: the mean (expected value) of X.*

In words: *the probability that X exceeds the threshold a cannot be larger than the mean of X divided by a.*

Intuition: *if the mean of X is small, X cannot often take very large values.*

Proof. Define $Y := a\,\mathbf{1}_{\{X \geq a\}}$. Pointwise $0 \leq Y \leq X$, therefore

$$\mathbb{E}[X] \geq \mathbb{E}[Y] = a\,\mathbb{E}[\mathbf{1}_{\{X \geq a\}}] = a\,\mathbb{P}[X \geq a].$$

Dividing by a proves the inequality. □

The bound is useful only when a is sufficiently larger than $\mathbb{E}[X]$.

Lemma 8 (Chebyshev's inequality)**.** *For every random variable Y with finite variance and every $a > 0$,*

$$\mathbb{P}\left[|Y - \mathbb{E}[Y]| \geq a\right] \leq \frac{\mathrm{Var}(Y)}{a^2}.$$

Meaning of the symbols:

- Y*: any random variable with finite variance (not necessarily nonnegative);*
- $a > 0$*: any threshold we choose, this time measured as a distance from the mean;*

- $|Y - \mathbb{E}[Y]| \geq a$: *$Y$ deviates from its own mean by at least a, on either side;*
- $\mathrm{Var}(Y)$*: the variance of Y, i.e. how spread out Y is around its mean.*

In words: *the probability that Y deviates from its mean by at least a cannot be larger than the variance of Y divided by a^2.*

Intuition: *if Y is tightly concentrated around its mean (small variance), large deviations are unlikely; and since a appears squared, the bound decays much faster as a grows than Markov's bound did.*

Proof. Apply Markov's inequality to the nonnegative random variable $Z := (Y - \mathbb{E}[Y])^2$, whose expectation is $\mathbb{E}[Z] = \mathrm{Var}(Y)$:

$$\mathbb{P}\left[|Y - \mathbb{E}[Y]| \geq a\right] = \mathbb{P}[Z \geq a^2] \leq \frac{\mathbb{E}[Z]}{a^2} = \frac{\mathrm{Var}(Y)}{a^2}. \qquad \square$$

Polynomial tails are often too weak.

- Chebyshev gives a tail of order $\dfrac{1}{a^2}$: only *polynomial* decay in a.
- For sums of independent bounded variables, the true tail often behaves like e^{-ca^2}: *exponential* decay in a^2.
- Equivalently, a deviation of λ times the standard deviation has probability of order $e^{-\frac{\lambda^2}{2}}$ — like the tail of a Gaussian random variable.

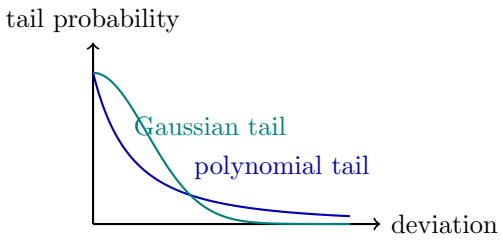


Figure 6: A polynomial tail (blue) decays much more slowly than a Gaussian-type tail (teal), even though both start from the same value.

The moment-generating function will let us turn Markov's inequality into bounds that see these exponentially small tails.

Remark 9 (Relation with the CLT)**.** *This is related to the central limit theorem: a suitably centered and normalized sum of independent random variables converges in distribution to a Gaussian. However, the CLT does not by itself give the bounds we need. Convergence in distribution controls probabilities at fixed thresholds, while we need quantitative, finite-n bounds at thresholds that move into the tails as n grows. We prove concentration through moment-generating functions instead.*

5.2 The moment-generating function

Introduction before defining the moment-generating function. Fix a real random variable Z and a real parameter θ. Applying the ordinary function $x \mapsto e^{\theta x}$ to Z produces a new random variable, written $e^{\theta Z}$; for discrete Z, its expectation is computed exactly as for any other function of Z, $\mathbb{E}[e^{\theta Z}] = \sum_z e^{\theta z}\, \mathbb{P}(Z = z)$. Think of θ as a dial we are free to turn: for $\theta > 0$ the map $x \mapsto e^{\theta x}$ is increasing, so it emphasizes the large, positive values of Z, while for $\theta < 0$ it is decreasing, so it instead emphasizes the large negative values of Z. Turning θ lets us shine a light on either tail of Z at will.

Three properties make $e^{\theta Z}$ the right transform to pair with Markov's inequality (Lemma 7):

- *Nonnegativity:* $e^{\theta Z} \geq 0$ always, even though Z itself need not be. This lets Markov's inequality, which requires a nonnegative variable, apply to $e^{\theta Z}$.
- *Monotonicity:* for $\theta > 0$, $u \mapsto e^{\theta u}$ is strictly increasing, so

  $$Z \geq t \iff e^{\theta Z} \geq e^{\theta t}.$$

  This turns a statement about Z into an equivalent statement about $e^{\theta Z}$.
- *Multiplicativity:* the exponential turns sums into products, $e^{\theta(x+y)} = e^{\theta x}\, e^{\theta y}$; combined with independence, this will turn the MGF of a sum of independent random variables into a product of MGFs (Lemma 11 below) — a product being far easier to control than a sum.

Definition 10. *For a real random variable Z, its* moment-generating function *(MGF) is*

$$\varphi_Z(\theta) := \mathbb{E}[e^{\theta Z}], \qquad \theta \in \mathbb{R},$$

where we allow the value $+\infty$: for some θ, this expectation may fail to be finite.

Exponential Markov. For every $\theta > 0$, the monotonicity property above gives

$$Z \geq t \iff e^{\theta Z} \geq e^{\theta t},$$

and Markov's inequality (Lemma 7), applied to the nonnegative variable $e^{\theta Z}$, gives

$$\mathbb{P}[Z \geq t] = \mathbb{P}[e^{\theta Z} \geq e^{\theta t}] \leq \frac{\mathbb{E}[e^{\theta Z}]}{e^{\theta t}} = \varphi_Z(\theta)\, e^{-\theta t}, \qquad \theta > 0. \tag{1}$$

This bound holds for *every* choice of $\theta > 0$, and each choice gives a valid, but generally different, upper bound on $\mathbb{P}[Z \geq t]$. Since we want the tightest possible bound, later on we minimize the right-hand side of (1) over $\theta > 0$; this optimization is exactly what will produce the Gaussian-shaped tail bounds we are after.

MGF of independent sums. The multiplicativity of the exponential, combined with independence, gives the crucial algebraic property of the MGF.

Lemma 11. *If Z_1 and Z_2 are independent, then $\varphi_{Z_1+Z_2} = \varphi_{Z_1}\varphi_{Z_2}$.*

Proof.

$$\begin{aligned}\varphi_{Z_1+Z_2}(\theta) &= \mathbb{E}[e^{\theta(Z_1+Z_2)}] \\ &= \mathbb{E}[e^{\theta Z_1} e^{\theta Z_2}] \\ &= \mathbb{E}[e^{\theta Z_1}]\ \mathbb{E}[e^{\theta Z_2}] \\ &= \varphi_{Z_1}(\theta)\varphi_{Z_2}(\theta),\end{aligned}$$

where the first equality uses $e^{\theta(Z_1+Z_2)} = e^{\theta Z_1} e^{\theta Z_2}$, and the third uses the independence of $e^{\theta Z_1}$ and $e^{\theta Z_2}$. □

Independence converts addition of random variables into multiplication of their MGFs — the same computation works for characteristic functions, discussed below.

Relation to moments, and a technical caveat. The name "moment-generating function" comes from the following fact. Suppose φ_Z is finite on some open interval containing 0. Then φ_Z is infinitely differentiable at 0, and its derivatives recover the moments of Z:

$$\varphi_Z^{(k)}(0) = \mathbb{E}[Z^k], \qquad k = 0, 1, 2, \ldots$$

(this comes from differentiating under the expectation, which is justified precisely because φ_Z is finite in a neighbourhood of 0; without this condition the moments need not even be finite). Equivalently, on that interval, φ_Z agrees with its Taylor series at 0,

$$\varphi_Z(\theta) = \sum_{k\geq 0} \frac{\theta^k}{k!}\ \mathbb{E}[Z^k].$$

Everything stated so far in this paragraph is fairly immediate from the definitions (differentiating under the expectation sign). The next fact is of a different nature: it is a genuinely more serious statement, whose proof requires a fair amount of complex analysis. The same finiteness condition guarantees that φ_Z determines the distribution of Z uniquely: two random variables with the same MGF, finite near 0, have the same distribution. Without this condition the MGF can fail to determine the distribution, and can even fail to be finite anywhere except at $\theta = 0$.

Remark 12. *The characteristic function $\psi_Z(t) := \mathbb{E}[e^{itZ}]$ avoids this issue entirely: since $|e^{itZ}| = 1$, it is always finite, for every distribution and every $t \in \mathbb{R}$, and it always determines the distribution uniquely, via the inverse Fourier transform (we essentially saw the formula for it, in the discrete case, in Lecture 2). This fact, while not completely trivial, also is not terribly deep, unlike the corresponding statement for the MGF above. The MGF and the characteristic function are restrictions of the same complex function $F_Z(w) := \mathbb{E}[e^{wZ}]$ (wherever it is defined) to the real axis and to the imaginary axis, respectively: $\varphi_Z(\theta) = F_Z(\theta)$ and $\psi_Z(t) = F_Z(it)$.*

5.3 Subgaussian random variables

Definition 13 (Subgaussian random variable)**.** *A mean-zero random variable Z is* σ-subgaussian *if*

$$\mathbb{E}[e^{\theta Z}] \leq \exp\left(\frac{\sigma^2\theta^2}{2}\right) \qquad \text{for every } \theta \in \mathbb{R}.$$

If $Z \sim N(0, \sigma^2)$ then equality holds: $\mathbb{E}[e^{\theta Z}] = \exp(\sigma^2\theta^2/2)$. So a subgaussian variable is one whose MGF is dominated by the MGF of a Gaussian; this is equivalent to a number of other conditions (e.g. Gaussian-type tails), but the MGF definition is the most useful one for us. One should think of σ as a scale parameter: it bounds the standard deviation, $\mathrm{Var}(Z) \le \sigma^2$ (compare the quadratic terms in θ), although it need not be equal to it.

Lemma 14 (Parameters add in squares)**.** *Suppose $Z_1, \dots, Z_m$ are independent, mean-zero, and Z_i is σ_i-subgaussian. Then $\sum_i Z_i$ is $\sqrt{\sum_i \sigma_i^2}$-subgaussian.*

Proof. By independence,

$$\mathbb{E}\exp\Big(\theta \sum_{i=1}^m Z_i\Big) = \prod_{i=1}^m \mathbb{E}[e^{\theta Z_i}] \le \prod_{i=1}^m \exp\Big(\frac{\sigma_i^2\theta^2}{2}\Big) = \exp\Big(\frac{\theta^2}{2}\sum_{i=1}^m \sigma_i^2\Big). \qquad \square$$

Bounded random variables have no tail at all beyond their support, so it is not surprising that they are subgaussian. The quantitative statement is Hoeffding's lemma.

Lemma 15 (Hoeffding's lemma)**.** *If $\mathbb{E}[Z] = 0$ and $Z \in [a, b]$ almost surely, then*

$$\mathbb{E}[e^{\theta Z}] \le \exp\Big(\frac{\theta^2(b-a)^2}{8}\Big) \qquad \forall \theta \in \mathbb{R},$$

i.e. Z is $\frac{b-a}{2}$-subgaussian. In particular, if $|Z| \le b$ then $\mathbb{E}[e^{\theta Z}] \le e^{b^2\theta^2/2}$: every mean-zero variable supported in $[-b, b]$ is b-subgaussian.

Proof. If $a = b$ then $Z = 0$ and there is nothing to prove, so assume $a < 0 < b$ (as $\mathbb{E}\, Z = 0$). For every $z \in [a, b]$, convexity of the exponential gives

$$e^{\theta z} \le \frac{b-z}{b-a}e^{\theta a} + \frac{z-a}{b-a}e^{\theta b}.$$

Taking expectations and using $\mathbb{E}[Z] = 0$,

$$\mathbb{E}[e^{\theta Z}] \le \frac{b}{b-a}e^{\theta a} - \frac{a}{b-a}e^{\theta b}.$$

The right-hand side is the MGF of a suitable two-point random variable, so extremal bounded variables behave no worse than a two-point distribution. It remains a one-variable calculus argument: set $p = -a/(b-a) \in (0, 1)$ and $u = \theta(b-a)$, so that the right-hand side equals $e^{L(u)}$ with

$$L(u) = -pu + \log\left(1 - p + pe^u\right).$$

Then $L(0) = 0$, $L'(0) = 0$ and $L''(u) = q(1-q) \le 1/4$, where $q = pe^u/(1-p+pe^u) \in [0, 1]$. By Taylor's theorem $L(u) \le u^2/8 = \theta^2(b-a)^2/8$. $\square$

5.4 From the MGF to exponential tails: Hoeffding's inequality

Lemma 16. *If Z is σ-subgaussian with $\sigma > 0$, then for every $t > 0$*

$$\mathbb{P}[Z \ge t] \le \exp\Big(-\frac{t^2}{2\sigma^2}\Big) \qquad \text{and} \qquad \mathbb{P}[|Z| \ge t] \le 2\exp\Big(-\frac{t^2}{2\sigma^2}\Big).$$

Proof. By (1) and the definition, for every $\theta > 0$,

$$\mathbb{P}[Z \geq t] \leq \mathbb{E}[e^{\theta Z}]e^{-\theta t} \leq \exp\left(\frac{\sigma^2\theta^2}{2} - \theta t\right).$$

The exponent $f(\theta) = \sigma^2\theta^2/2 - \theta t$ is minimized at $\theta^* = t/\sigma^2$ (where $f'(\theta^*) = 0$ and $f'' = \sigma^2 > 0$), and substitution gives $f(\theta^*) = -t^2/(2\sigma^2)$. Since $-Z$ is also σ-subgaussian, the same argument bounds $\mathbb{P}[Z \leq -t]$, and a union bound gives the two-sided statement. □

For example, a deviation of 10σ has probability at most $2e^{-50}$. Combining Lemmas 14, 15 and 16 we obtain the main tool of this lecture, due to Hoeffding.

Theorem 17 (Hoeffding's inequality)**.** *Let $X_1, \ldots, X_m$ be independent random variables with*

$$\mathbb{E}[X_i] = 0, \qquad |X_i| \leq b_i,$$

and set $B^2 := \sum_{i=1}^m b_i^2 > 0$. Then for every $t > 0$

$$\mathbb{P}\left[\left|\sum_{i=1}^m X_i\right| \geq t\right] \leq 2\exp\left(-\frac{t^2}{2B^2}\right).$$

Proof. Each X_i is b_i-subgaussian by Hoeffding's lemma, so $\sum_i X_i$ is B-subgaussian by Lemma 14; conclude with Lemma 16. □

Writing $t = \lambda B$, Hoeffding's inequality says that normalized deviations have Gaussian tails:

$$\mathbb{P}\left[\left|\sum_i X_i\right| \geq \lambda B\right] \leq 2e^{-\lambda^2/2}, \qquad B = \sqrt{\textstyle\sum_i b_i^2}.$$

Note that the scale B is deterministic, and that independence is essential: without it, perfectly correlated bounded variables can add coherently and produce fluctuations of order m instead of order $\sqrt{m}$.

6 Concentration of graph degrees

A fixed null degree. Fix $v \in [n]$ and let $G \sim G(n, 1/2)$. Then

$$\deg(v) = \sum_{u \neq v} Z_{uv},$$

where Z_{uv} indicates whether $\{u, v\}$ is an edge; the Z_{uv} are independent Bernoulli(1/2) variables. We center them: define $X_{uv} := Z_{uv} - \frac{1}{2}$, so that

$$\mathbb{E}[X_{uv}] = 0, \qquad |X_{uv}| \leq \frac{1}{2}, \qquad \text{and} \qquad \deg(v) = \frac{n-1}{2} + \sum_{u \neq v} X_{uv}.$$

Each X_{uv} is $\frac{1}{2}$-subgaussian, so the centered degree is subgaussian with parameter

$$B = \sqrt{\sum_{u \neq v} \left(\frac{1}{2}\right)^2} = \frac{\sqrt{n-1}}{2}.$$

Hence a fixed degree fluctuates by $O(\sqrt{n})$: by Hoeffding's inequality,

$$\mathbb{P}\left[\left|\deg(v) - \frac{n-1}{2}\right| \geq t\right] \leq 2\exp\left(-\frac{2t^2}{n-1}\right). \tag{2}$$

All vertices simultaneously. Taking $t = C\sqrt{n \log n}$ in (2) makes the failure probability at most

$$2\exp\left(-\frac{2C^2 n \log n}{n-1}\right) \leq 2n^{-2C^2}.$$

Let $\mathcal{B}_v = \left\{|\deg(v) - \frac{n-1}{2}| > C\sqrt{n \log n}\right\}$. By the union bound,

$$\mathbb{P}[\exists v : \mathcal{B}_v] \leq \sum_{v=1}^{n} \mathbb{P}[\mathcal{B}_v] \leq 2n^{1-2C^2}.$$

So for a sufficiently large constant C (already $C = 1$ gives $2/n$), with high probability *every* null degree lies in

$$\frac{n-1}{2} \pm C\sqrt{n \log n}.$$

The same holds in the planted model for every vertex $v \notin S$, whose degree is still $\mathrm{Bin}(n-1, 1/2)$.

Planted clique degrees. For $v \in S$,

$$\deg(v) = (k-1) + \sum_{u \notin S} Z_{uv}, \qquad Z_{uv} \sim \text{Bernoulli}(1/2) \text{ independently},$$

so $\mathbb{E}[\deg(v)] = k - 1 + \frac{n-k}{2} = \frac{n+k-2}{2}$. The random part is a sum of at most n bounded independent variables, so the same argument shows that with high probability every planted vertex satisfies

$$\deg(v) = \frac{n+k-2}{2} \pm C\sqrt{n \log n}.$$

6.1 The degree algorithm

The two high-probability degree ranges are disjoint once the mean gap $(k-1)/2$ exceeds twice the fluctuation radius $C\sqrt{n \log n}$. We use the midpoint threshold

$$\tau := \frac{\mu_0 + \mu_1}{2} = \frac{n-1}{2} + \frac{k-1}{4}.$$

With high probability,

$$\begin{aligned} &\text{under } H_0: && \max_v \deg(v) \leq \frac{n-1}{2} + C\sqrt{n \log n}, \\ &\text{under } H_1: && \deg(v) \geq \frac{n-1}{2} + \frac{k-1}{2} - C\sqrt{n \log n} \quad \text{for every } v \in S. \end{aligned}$$

These intervals lie on opposite sides of τ when $k \geq C_0\sqrt{n \log n}$, for a sufficiently large constant C_0.

Degree test.

1. Compute every vertex degree.
2. Output H_1 if some vertex has degree at least $\tau = \frac{n-1}{2} + \frac{k-1}{4}$.
3. Otherwise output H_0.

Theorem 18. *For $k \geq C_0\sqrt{n \log n}$, the degree test distinguishes the two distributions with probability tending to one. Selecting all vertices with degree at least τ also recovers the planted clique in the same parameter range.*

For recovery, note that with high probability all vertices of S are above the threshold, while all vertices outside S (whose degrees are $\mathrm{Bin}(n-1, 1/2)$) are below it. This algorithm is due to Kučera.

7 The spectral planted clique algorithm

7.1 The signed adjacency matrix

Definition 19. *The* signed adjacency matrix $M \in \mathbb{R}^{n \times n}$ *of G is defined by $M_{ii} = 0$ and, for $i \neq j$,*

$$M_{ij} = \begin{cases} +1, & \{i,j\} \in E, \\ -1, & \{i,j\} \notin E. \end{cases}$$

Under $G(n, 1/2)$ the entries above the diagonal are independent Rademacher random variables, $\mathbb{P}[M_{ij} = 1] = \mathbb{P}[M_{ij} = -1] = \frac{1}{2}$, so $\mathbb{E}[M_{ij}] = 0$: the matrix is centered around zero. Note that M is symmetric, so only the $\binom{n}{2}$ entries with $i < j$ are independent, and $M_{ji} = M_{ij}$.

Planting a clique creates a coherent positive block: after permuting rows and columns so that S comes first,

$$M = \begin{pmatrix} \text{all } +1 \text{ off the diagonal} & \text{random signs} \\ \text{random signs} & \text{random signs} \end{pmatrix},$$

where the top-left block is $k \times k$. Such a block should force a large eigenvalue.

7.2 A k-clique forces a large eigenvalue

Recall that for a symmetric matrix M,

$$\|M\|_{\mathrm{op}} = \max_{\|x\|_2 = 1} |x^T M x| = \max_{x \neq 0} \frac{|x^T M x|}{\|x\|_2^2}.$$

In particular, to lower-bound the operator norm it is enough to exhibit a single vector.

Lemma 20. *If G contains a k-clique S, then $\|M\|_{\mathrm{op}} \geq k - 1$.*

Proof. Let $\mathbf{1}_S$ be the indicator vector of S. Every ordered pair of distinct vertices in S contributes $+1$ to $\mathbf{1}_S^T M \mathbf{1}_S$, and the diagonal contributes 0, so

$$\mathbf{1}_S^T M \mathbf{1}_S = k(k-1), \qquad \|\mathbf{1}_S\|_2^2 = k.$$

Therefore $\|M\|_{\mathrm{op}} \geq \dfrac{\mathbf{1}_S^T M \mathbf{1}_S}{\|\mathbf{1}_S\|_2^2} = k - 1.$ □

It remains to show that under the null distribution the norm is much smaller.

7.3 The random matrix theorem and the spectral test

Theorem 21 (Random matrix theorem)**.** *Let M be symmetric with $M_{ii} = 0$, and let the entries $M_{ij} = M_{ji}$, $i < j$, be independent uniform signs. Then*

$$\|M\|_{\mathrm{op}} \leq C\sqrt{n}$$

with probability at least $1 - e^{-cn}$, for universal constants $C, c > 0$.

The entries have size 1, so a priori the norm could be of order n; cancellations keep the largest eigenvalue at scale $\sqrt{n}$. We prove the theorem in Section 9 using an ε-net.

Spectral test.

1. Construct the signed adjacency matrix M.
2. Compute $\|M\|_{\mathrm{op}}$.
3. Output H_1 if $\|M\|_{\mathrm{op}} > C\sqrt{n}$.
4. Otherwise output H_0.

Under the null, $\|M\|_{\mathrm{op}} \leq C\sqrt{n}$ with high probability by Theorem 21. Under the planted distribution, $\|M\|_{\mathrm{op}} \geq k - 1$ *deterministically* by Lemma 20. Thus the algorithm succeeds whenever $k - 1 > C\sqrt{n}$, in particular for $k \geq C'\sqrt{n}$ with a sufficiently large constant C'. The operator norm of a symmetric matrix can be computed in polynomial time, so we obtain the following (the result is due to Alon, Krivelevich and Sudakov).

Corollary 22. *A polynomial-time spectral algorithm distinguishes $G(n, 1/2)$ from $\mathcal{P}_1(n, k)$ whenever $k \geq C'\sqrt{n}$.*

This improves the degree threshold $C\sqrt{n \log n}$ by removing the factor $\sqrt{\log n}$. No polynomial-time algorithm is known for planted clique of size $o(\sqrt{n})$.

8 Discretizing the unit sphere

To prove Theorem 21 we have to control

$$\|M\|_{\mathrm{op}} = \sup_{x \in \mathbb{S}^{n-1}} |x^T M x|, \qquad \mathbb{S}^{n-1} = \{x \in \mathbb{R}^n : \|x\|_2 = 1\}.$$

For a *fixed* x, the quantity $x^T M x$ is a sum of independent bounded random variables, so it is well concentrated. But a concentration bound for one fixed vector does not directly control this supremum over uncountably many vectors. We need a finite approximation of the sphere. This is an extremely common idea: the *ε-net argument*.

Definition 23 (ε-net)**.** *A set $\mathcal{N} \subseteq \mathbb{S}^{n-1}$ is an* ε-net *if every $x \in \mathbb{S}^{n-1}$ has some $y \in \mathcal{N}$ satisfying*

$$\|x - y\|_2 \le \varepsilon.$$

Definition 24 (Covering number)**.** *For $A \subseteq \mathbb{R}^n$, the* covering number $N(A, \varepsilon)$ *is the minimum size of an ε-net contained in A. Equivalently, it is the smallest number of balls of radius ε, centered in A, that cover A.*

Our goal is an exponential-size net: $N(\mathbb{S}^{n-1}, 1/10) \le 2^{O(n)}$. This may look enormous, but a fixed-vector tail of size $e^{-\Omega(n)}$ can survive a union bound over $2^{O(n)}$ points. To bound covering numbers we use the duality with packings.

Definition 25 (Packing)**.** *A set $\mathcal{P} \subseteq A$ is an* ε-packing *if $\|x - y\|_2 > \varepsilon$ for every distinct $x, y \in \mathcal{P}$. The* packing number $P(A, \varepsilon)$ *is the largest cardinality of an ε-packing in A.*

For an ε-packing, the balls of radius $\varepsilon/2$ around its points are pairwise disjoint: a common point z would give $\|x - y\|_2 \le \|x - z\|_2 + \|z - y\|_2 \le \varepsilon$.

Lemma 26 (Covering-packing lemma)**.** *Every inclusion-maximal ε-packing $\mathcal{P} \subseteq A$ is an ε-net for A. Consequently, $N(A, \varepsilon) \le P(A, \varepsilon)$.*

Proof. Suppose instead that some $x \in A$ satisfies $\|x - y\|_2 > \varepsilon$ for every $y \in \mathcal{P}$. Then $\mathcal{P} \cup \{x\}$ is still an ε-packing, contradicting maximality. A packing of maximum cardinality is in particular inclusion-maximal, which gives the inequality. □

Lemma 27 (Volume bound)**.** *Every ε-packing $\mathcal{P} \subseteq \mathbb{S}^{n-1}$ satisfies $|\mathcal{P}| \le \left(1 + \frac{2}{\varepsilon}\right)^n$.*

Proof. The balls $B(p, \varepsilon/2)$, $p \in \mathcal{P}$, are pairwise disjoint. Since $\|p\|_2 = 1$, each of them lies inside $B(0, 1 + \varepsilon/2)$: for $z \in B(p, \varepsilon/2)$, $\|z\|_2 \le \|z - p\|_2 + \|p\|_2 \le 1 + \varepsilon/2$. Comparing volumes of $V = \bigcup_{p \in \mathcal{P}} B(p, \varepsilon/2)$ and of the large ball,

$$|\mathcal{P}| \ \mathrm{Vol}\left(B(0, \varepsilon/2)\right) = \mathrm{Vol}(V) \le \mathrm{Vol}\left(B(0, 1 + \varepsilon/2)\right).$$

Volume in $\mathbb{R}^n$ scales as the n-th power of the radius, therefore

$$|\mathcal{P}| \le \frac{\mathrm{Vol}(B(0, 1 + \varepsilon/2))}{\mathrm{Vol}(B(0, \varepsilon/2))} = \left(\frac{1 + \varepsilon/2}{\varepsilon/2}\right)^n = \left(1 + \frac{2}{\varepsilon}\right)^n. \qquad \square$$

Combining Lemmas 26 and 27:

Corollary 28. *$N(\mathbb{S}^{n-1}, \varepsilon) \le \left(1 + \frac{2}{\varepsilon}\right)^n$. In particular, for $\varepsilon = 1/10$ there exists a net $\mathcal{N} \subseteq \mathbb{S}^{n-1}$ with $|\mathcal{N}| \le 21^n = 2^{O(n)}$.*

This is the right scale: the sphere is continuous, but constant-accuracy discretization costs only exponentially many points.

9 Proof of the random matrix theorem

Step 1: a fixed quadratic form is a bounded independent sum. Fix $x \in \mathbb{S}^{n-1}$. Since M is symmetric with zero diagonal,

$$x^T M x = 2 \sum_{i<j} M_{ij} x_i x_j.$$

The variables M_{ij}, $i < j$, are independent uniform signs, so each summand $2M_{ij}x_ix_j$ has mean zero and magnitude at most $2|x_ix_j|$. The coefficient squares sum to at most 2:

$$\sum_{i<j} (2x_ix_j)^2 = 4 \sum_{i<j} x_i^2 x_j^2 = 2 \left[\Big(\sum_i x_i^2 \Big)^2 - \sum_i x_i^4 \right] \le 2 \Big(\sum_i x_i^2 \Big)^2 = 2.$$

Hence Hoeffding's inequality with $B^2 \le 2$ gives, for every fixed $x \in \mathbb{S}^{n-1}$,

$$\mathbb{P}\left[|x^T M x| \ge t \right] \le 2e^{-t^2/4}. \tag{3}$$

Note that this tail bound does not depend on the dimension.

Step 2: union bound over the net. Let $\mathcal{N}$ be a 1/10-net with $|\mathcal{N}| \le 21^n$ (Corollary 28). Taking $t = C\sqrt{n}$ in (3), $\mathbb{P}[|x^T M x| > C\sqrt{n}] \le 2e^{-C^2 n/4}$ for each fixed $x \in \mathcal{N}$, and a union bound gives

$$\mathbb{P}\left[\exists x \in \mathcal{N} : |x^T M x| > C\sqrt{n} \right] \le 2 \cdot 21^n e^{-C^2 n/4}.$$

For sufficiently large C (e.g. $C = 4$, since $\ln 21 < 4$) this is $e^{-\Omega(n)}$.

Step 3: a bound on the net controls the whole sphere. Define $T := \max_{x \in \mathcal{N}} |x^T M x|$. Choose $u \in \mathbb{S}^{n-1}$ with $\|M\|_{\mathrm{op}} = |u^T M u|$. By the net property there exists $x \in \mathcal{N}$ with $\|u - x\|_2 \le \varepsilon$, $\varepsilon = 1/10$. Write $h := u - x$, so $\|h\|_2 \le \varepsilon$. Writing $u = x + h$ and expanding,

$$u^T M u - x^T M x = x^T M h + h^T M x + h^T M h,$$

we get, using $\|h\|_2 \le \varepsilon$ and $\|x\|_2, \|u\|_2 \le 1$,

$$\begin{aligned} \|M\|_{\mathrm{op}} = |u^T M u| &\le |x^T M x| + |x^T M h| + |h^T M x| + |h^T M h| \\ &\le T + \varepsilon \|M\|_{\mathrm{op}} + \varepsilon \|M\|_{\mathrm{op}} + \varepsilon^2 \|M\|_{\mathrm{op}}. \end{aligned}$$

The approximation errors are proportional to the unknown norm, and we can absorb them:

$$(1 - 2\varepsilon - \varepsilon^2) \|M\|_{\mathrm{op}} \le T, \qquad \text{i.e.} \qquad \|M\|_{\mathrm{op}} \le \frac{T}{1 - 2\varepsilon - \varepsilon^2} = \frac{100}{79} T \quad \text{for } \varepsilon = \frac{1}{10}.$$

Conclusion. With probability at least $1 - e^{-\Omega(n)}$, Step 2 gives $T \le C\sqrt{n}$, and therefore by Step 3

$$\|M\|_{\mathrm{op}} \le \frac{100}{79} C\sqrt{n} = O(\sqrt{n}).$$

This completes the proof of Theorem 21. □

The proof combines geometry and probability:

$$\text{fixed-vector concentration} \longrightarrow \varepsilon\text{-net of the sphere}$$
$$\longrightarrow \text{union bound over } 2^{O(n)} \text{ points} \longrightarrow \|M\|_{\mathrm{op}} = O(\sqrt{n}).$$

The probability of failure for one direction is exponentially small; geometry reduces the continuous sphere to exponentially many directions; the two exponential scales balance.

10 Summary

1. Average-case problems include a distribution over inputs.
2. Planted clique is statistically detectable for $k \gtrsim 2\log_2 n$, but efficient algorithms need a larger clique.
3. Moment-generating functions turn Markov's inequality into Gaussian tail bounds.
4. Degree concentration yields an algorithm for $k \gtrsim \sqrt{n \log n}$.
5. A clique creates a signed-adjacency eigenvalue of size $k - 1$.
6. Nets and concentration show that null eigenvalues have size $O(\sqrt{n})$.
7. The spectral algorithm succeeds for $k \gtrsim \sqrt{n}$.

Typical inputs can contain structure that worst-case analysis does not expose; concentration makes that structure algorithmically visible.

LLM usage disclosure

An LLM was used to help produce an initial outline of these notes. The text was subsequently rewritten and edited by the author.