Proof Complexity as a Computational Lens 2025/26 November 6, 2025

Lecture 3

Lecturer: Jakob Nordström *Scribe: Marc Vinyals, Additional-Scribe-Here*

1 Recap of Last Lecture and Plans for Today

In the last lecture we proved an exponential lower bound

$$L_{\mathcal{R}}(PHP_n^{n+1} \vdash \bot) = \exp(\Omega(n)) \tag{1.1}$$

on resolution refutations on pigeonhole principle formulas PHP_n^{n+1}. Observe that PHP_n^{n+1} has size $N = \Theta(n^3)$, so the lower bound is, in fact, $\exp\big(\Omega\big(\sqrt[3]{N}\big)\big)$ in terms of formula size, and not truly exponential (by which we mean $\exp(\Omega(N))$ with the coefficient in the exponent scaling linearly with N).

In order to prove this lower bound we used the Prosecutor–Defendant game, where Prosecutor asks whether pigeon i flies into hole j, and Defendant replies in a way that delays an explicit contradiction for as long as possible. Good Defendant strategies for this game imply resolution lower bounds, so we constructed such a strategy: Defendant picks a random matching of $n/4$ pigeons to $n/4$ pigeonholes, and then answers according to this matching when asked about pigeons in this matching, and otherwise for other pigeons says that they do not go into holes for as long as possible. It is easy to see that there are exponentially many different choices for a matching. It takes some more work to show that, before Prosecutor wins, there has be be a record where a noticeable fraction of (information about) this random matching is written down. Once this is shown, however, it follows that Prosecutor needs exponentially many records, which immediately gives an exponential resolution lower bound.

Today we look at formulas encoding (a contradiction of) the handshaking lemma: *"The sum of vertex degrees in an undirected graph $G = (V, E)$ is even"*, or, in math symbols, that $\sum_{v \in V} \deg(v) \equiv 0 \pmod 2$. Our goal is to obtain lower bounds that are *truly* exponential in the size of the formula.

2 A General Version of the Prosecutor–Defendant Game

In the last lecture we only defined the Prosecutor–Defendant game by Pudlák [Pud00] for pigeonhole principle formulas, but it is not hard to see that it generalizes to any formula. There is nothing special about asking whether pigeon i flies to hole j compared to simply asking whether a variable in the formula is assigned value 0 or 1.

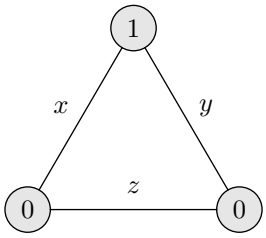
Let us describe how the game is played on any unsatisfiable CNF formula F (which is fixed and known to both players). Prosecutor maintains a record R, which is just a partial truth value assignment to $Vars(F)$. Every record R has an associated instruction of type Ask or Forget.

Ask Ask the value of a variable x. Defendant answers $b \in \{0, 1\}$; Prosecutor adds $\{x \mapsto b\}$ to R.

Forget Forget some values, i.e., shrink the assignment to $R' \subsetneq R$.

A *winning position* R for Prosecutor is an assignment falsifying some clause $C \in F$ (these are the "explicit contradictions" we had for PHP formulas in the last lecture). A *strategy* $\mathcal{S}$ for Prosecutor is a (finite) set of records such that there is an instruction/record for every possible partial assignment that can arise during play. A Prosecutor strategy is *complete* if in addition Prosecutor can always win (in a finite number of steps) regardless of how Defendant plays. The size of a strategy $\mathcal{S}$ is the number of records in it.

The reason we are studying the Prosecutor–Defendant game is that it is essentially equivalent to resolution: Complete Prosecutor strategies for a formula F are just a different way of describing resolution refutations of F and vice versa.



(a) Triangle graph with odd labelling.

$$
\begin{aligned}
&(x \vee y)\\
\wedge\ &(\overline{x} \vee \overline{y})\\
\wedge\ &(x \vee \overline{z})\\
\wedge\ &(\overline{x} \vee z)\\
\wedge\ &(y \vee \overline{z})\\
\wedge\ &(\overline{y} \vee z)
\end{aligned}
$$

(b) Corresponding Tseitin formula.

Figure 1: Example Tseitin formula.

Lemma 2.1 ([Pud00]). *F has a resolution refutation in length L if and only if Prosecutor has a complete strategy of size $\Theta(L)$.*

We saw the proof of the $\Rightarrow$ direction during the last lecture, and this is the only direction that we will need. The $\Leftarrow$ direction is left as an exercise (not very hard, but not trivial either).

3 Tseitin Formulas

We now proceed to define the formulas that we use to encode the handshaking lemma discussed above. When we describe these formulas, it will be convenient to use a notation x^b for literals with the polarity of the literal in the exponent, where we write $x^1 = x$ and $x^0 = \overline{x}$, so that x^b is the literal satisfied by setting $x \mapsto b$. Let $G = (V, E)$ be an undirected graph. Throughout this lecture we denote the order of the graph by $n = |V|$. We have one variable x_e for every edge $e \in E$, and we often identify variables and edges. We say that a function $\chi \colon V \to \{0, 1\}$ is a *charge* function.

Let us write $PARITY_{v,\chi}$ to denote the set of clauses encoding

$$\sum_{e \ni v} x_e \equiv \chi(v) \pmod 2\,, \tag{3.1}$$

i.e., that the number of true edges incident to v is equal to $\chi(v) \pmod 2$ (or, as we will often say in what follows, that $\chi(v)$ is the *parity* of the sum). If we expand the above constraint into CNF, we get the set of clauses

$$PARITY_{v,\chi} = \left\{ \bigvee_{e \ni v} x_e^{1-b_e} \;\middle|\; \sum_{e \ni v} b_e \not\equiv \chi(v) \pmod 2 \right\}. \tag{3.2}$$

In other words, we have a set of constraints that forbids every assignment to $\{x_e \mid e \ni v\}$ that gets the parity wrong. Indeed, if $\{x_e \mapsto b_e\}_{e \ni v}$ is a partial assignment such that $\sum b_e \not\equiv \chi(v) \pmod 2$, then we have a clause $\bigvee_{e \ni v} x_e^{1-b_e}$ that forces at least one literal $x_e^{1-b_e}$ to be satisfied, i.e., it must be the case that the assignment maps $x_e \mapsto 1 - b_e$, showing that we do not have this assignment.

Every vertex imposes a constraint on the parity of the edges incident to it, and the collection of all these constraints (or, rather, the union of the clauses encoding them), is the Tseitin formula. That is, the *Tseitin formula over G with respect to χ* is defined as

$$Ts(G, \chi) = \bigwedge_{v \in V} PARITY_{v,\chi}\,. \tag{3.3}$$

See Figure 1 for an example.

If G has bounded degree d, i.e., if all vertices have at most d incident edges, then it is not hard to see that $Ts(G, \chi)$ is a d-CNF formula with at most $2^{d-1}|V|$ clauses and $|E| \leq d|V|/2$ variables. Since we want the initial formula to be of size linear in the graph, when talking about Tseitin formulas in what follows we always assume that G has constant bounded degree unless stated otherwise.

We say that the *charge* of a function $\chi\colon V \to \{0,1\}$ is $\sum_{v\in V}\chi(v)$. In particular, an *odd-charge* function has charge $\sum_{v\in V}\chi(v) \equiv 1 \pmod 2$ and an *even-charge* function satisfies $\sum_{v\in V}\chi(v) \equiv 0 \pmod 2$.

Proposition 3.1 ([Urq87]). *If G is connected, then $Ts(G,\chi)$ is unsatisfiable if and only if χ is an odd-charge function.*

The proof of Proposition 3.1 is left as an exercise.

Since we want the initial formula to be unsatisfiable, in what follows we always assume that the graph G underlying the Tseitin formula is connected and that the charge function χ is of odd charge unless stated otherwise.

Today we want to prove that if G is a *well-connected* graph, then resolution refutations of $Ts(G,\chi)$ require truly exponential length $\exp(\Omega(N))$ measured in the size N of the formula. Let us make two quick comments:

- For an even-charge χ this is obviously true (satisfiable formulas are very hard to refute, wink wink, nudge nudge).
- For an odd-charge χ we only really care about the charge being odd. All odd-charge functions on V are equivalent from the point of view of resolution, since if we have a resolution refutation π of $Ts(G,\chi)$ for some odd χ and connected G and χ' is another odd-charge function, then just by flipping signs of literals in π we can obtain a refutation π' of $Ts(G,\chi')$. We will not need this observation, and so will not prove it—it is just a side remark.

4 Expander Graphs

For a Tseitin formula to be hard we would like the underlying formula to be very well connected, since this will prevent a proof to reason about a specific part of the graph in isolation. But we also want the degree to be bounded, so that the original formula is not too large. Is it possible to satisfy both constraints? Fortunately, graphs of bounded degree that are nevertheless very well connected are well-studied objects in theoretical computer science, and are known as *expander graphs*. There are several different ways of measuring expansion—below we outline three possible ways:

Vertex expansion A graph is a vertex expander if every small-to-medium vertex set $U \subseteq V(G)$ has many neighbours in $N(U) \setminus U$.

Edge expansion A graph is an edge expander if every small-to-medium-large $U \subseteq V$ has many outgoing edges to $V \setminus U$.

Algebraic expansion A graph is an algebraic (or spectral) expander if the gap between the two largest eigenvalues in the (normalized) adjacency matrix is large.

It turns out that these notions are all tightly connected, but proving this is out of scope for this course. In fact, expander graphs on their own could easily be made into the topic of one (or several) advanced courses, but here we will limit ourselves to mentioning that an excellent survey paper on expanders is [HLW06]. In today's lecture, the concept we will need is *edge expansion*.

In order to define formally what an edge expander is, we need a piece of notation. For $G = (V,E)$ and $U \subseteq V$, let ∂U denote the set of outgoing edges from U,[1] i.e.,

$$\partial U = \{(u,v) \in E \mid u \in U, v \in V \setminus U\}\,. \tag{4.1}$$

Now we can formally define edge expansion.

[1]The reader is warned that the notation ∂U is sometimes used to denote the related concepts of *unique neighbours* in a unique-neighbour (or boundary) expander graph, but we will not need this concept in the current lecture.

Definition 4.1 (Edge expansion). An undirected graph $G = (V, E)$ is a (d, δ)-*edge expander* if G has bounded degree d and for every vertex set $U \subseteq V$ of size $|U| \leq |V|/2$ it holds that $|\partial U| \geq \delta |U|$.

That is, a constant fraction of edges incident to U are exiting U.

We are ready to state the goal of today's lecture more precisely. Recall that we want to prove exponential lower bounds. What does this mean? This is an asymptotic statement, so it means that as formulas get larger and larger, refutation length scales exponentially. Tseitin formulas are defined in terms of graphs, and have size linear in the size of the graphs (assuming bounded degree). So we we want to prove that there is a family of graphs of growing size such that the resolution refutations of the Tseitin formulas over these graphs have to have exponentially growing length.

Theorem 4.2 ([Urq87]). *Fix $d \in \mathbb{N}^+$, $\delta > 0$, and suppose that $\{G_n\}_{n=3}^{\infty}$ is a family of n-vertex, (d, δ)-edge expanders. Then for any family of odd-charge functions $\chi_n \colon V(G_n) \to \{0, 1\}$ it holds that the CNF formula family $\{Ts(G_n, \chi_n)\}_{n=3}^{\infty}$ requires resolution refutations of length $\exp(\Omega(n))$.*

Let us make a few remarks. The formula size is $\Theta(n)$, so the lower bound is truly exponential. The concrete constants will depend on d and δ. We will not reproduce the original proof by Urquhart, but will do our own Prosecutor–Defendant style proof (cooked up together with Massimo Lauria and Per Austrin, but any errors are the responsibility of the lecturer, of course). We will actually prove a slightly weaker result, where we require the stronger assumption that the graphs are expanders with expansion factor δ not just positive, but with $\delta > 1$.

But first a natural question: Is Theorem 4.2 a non-vacuous theorem, i.e., are there such edge expander graphs as in the assumption in the theorem? Yes, there are such graphs, and in fact just picking random d-regular graphs will do.

Theorem 4.3 ([Bol88]). *Fix $d \in \mathbb{N}^+$, $d \geq 3$. Then there exists a universal constant $\delta > 0$ such that asymptotically almost surely a random d-regular graph is a (d, δ)-edge expander*

We say that a family of events $\{\mathcal{E}_n\}_{n=1}^{\infty}$ happens *asymptotically almost surely* (sometimes abbreviated *a.a.s.*) if $\lim_{n\to\infty} \Pr[\mathcal{E}_n] = 1$. Sometimes the same notion is also refered to as "with high probability" (w.h.p.), but we will try to stick to a.a.s. in this course.

In fact, it is also possible to give explicit constructions of families of expander graphs, but even if the constructions are not too convoluted, proving that they are expanders is highly nontrivial.

Let us make a quick detour to get better acquainted with edge expansion. The maximal edge expansion of a graph is known as the *isoperimetric number* or *Cheeger constant* $h(G)$, i.e.,

$$h(G) = \min_{U \subseteq V, |U| \leq |V|/2} \frac{|\partial U|}{|U|} \,. \tag{4.2}$$

How good expanders could we expect to be able to find? For a random graph and any subset $U \subseteq V(G)$ of size $|U| \leq |V(G)|/2$, we would expect roughly half of the edges in U to go to $V \setminus U$. So a random d-regular graph might have edge expansion something like $d/2$ if we are lucky. This is indeed the case for d large enough—in fact, we get this from a more refined version of Theorem 4.3.

Theorem 4.3' ([Bol88]). *For every $\epsilon > 0$ there is a $d \in \mathbb{N}^+$ such that a random d-regular n-vertex graph has edge expansion at least $d/2 - \epsilon$ asymptotically almost surely as $n \to \infty$.*

Bollobás actually calculates a more precise expression for the expansion, from which it follows that random 6-regular graphs have edge expansion strictly greater than 1 a.a.s., and random 4-regular graphs have expansion 0.4 a.a.s.

In the proof of Theorem 4.2 it will also be convenient to use another (fourth) notion of expansion.

Definition 4.4 (Connectivity expansion). An undirected graph $G = (V, E)$ is a (d, c)-*connectivity expander* if it has bounded degree d and for every edge set $E' \subseteq E$, $|E'| \leq cn$, it holds that the graph $G' = (V, E \setminus E')$ has a connected component of size strictly greater than $|V|/2$.

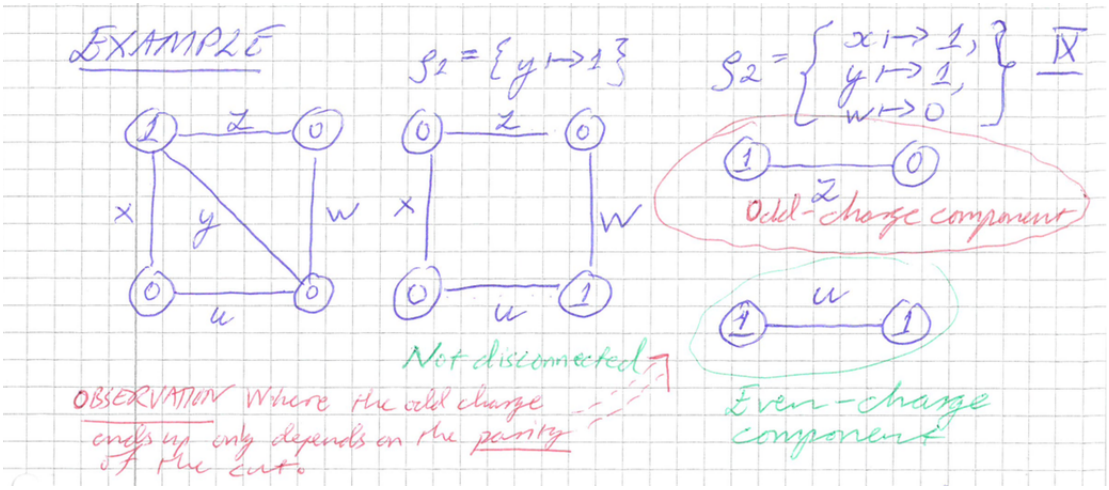


Figure 2: Charge update

We want the graph to be connected when you just remove a linear-sized subset of edges. This is obviously impossible since the graph has bounded degree d—by removing just d edges one can disconnect a vertex. But a graph is a connectivity expander if this is (essentially) the only thing that can happen, i.e., if removing a not too large linear-sized subset of edges can only disconnect fairly small subsets of vertices, and in particular leaves a large, connected component consisting of more than half of the graph. Since most of the graph is actually connected, we can even say that G is "morally connected."

Every edge expander is also a connectivity expander, so we only need to ask for the first property.

Proposition 4.5. *Every (d, δ)-edge expander is a (d, c)-connectivity expander for $c = \delta/4$.*

The proof of this proposition is left as an exercise.

5 Defendant Strategy for Tseitin Formulas

Let us go back to proving the lower bound on Tseitin formulas in Theorem 4.2. We will do so by coming up with a good strategy for Defendant, similarly to what we did for the pigeonhole principle formulas: pick a random subset of variables, fix in advance the answers on that subset, and for variables outside of this subset try to give "safe answers" to delay the inevitable defeat for as long as possible.

To save typing in what follows, let us say that any edge set E' of size $|E'| \leq cn$ in a (d, c)-connectivity expander is of *moderate size* and that the (unique) component of size greater than $|V|/2$ is *the large component*.

Observe that assigning values to some variables is equivalent to removing some edges from the graph (and updating the vertex charges accordingly). Making sure that we are left with a large component after removing some edges is not enough to guarantee that Defendant will not be caught red-handed—it is crucial that the small components are not of odd charge. Otherwise, if during play some small component of odd charge arises, Prosecutor can cross-examine Defendant about this particular component and quickly and cheaply prove a contradiction.

For any moderate-size edge-set E', let us say that an assignment $\rho\colon E' \to \{0, 1\}$ is *charge-preserving* if in the graph $G' = (V, E \setminus E')$ the large component has odd charge and any small disconnected component has even charge (meaning that the corresponding subformula is satisfiable by Proposition 3.1). Note that we update the charge function χ for G' by plugging in edge values from ρ. If $e \ni v$, then we update $\chi'(v) = \chi(v) + \rho(e)$. See Figure 2 for an example. Observe that the component where the odd charge ends up only depends on the parity of the cut.

Here is the Defendant strategy at a high level for $G = (V, E)$ a graph that is both a (d, δ)-edge expander and a (d, c)-connectivity expander for $c = \delta/4$.

First pick uniformly at random a set $E' \subset E$ of size $cn/2$. Sample uniformly at random a charge-preserving assignment to E'. Note that different edges are *not* set independently. In Figure 2, if $x, u \in E'$, then once we have randomly set x we must set $u = x$. But in turns out that many edges in E' will in fact be set independently at random (although we will need to argue this crucial fact carefully).

Call this assignment ρ_{init}. Defendant will maintain a partial assignment ρ such that

- $\rho \supseteq \rho_{\text{init}}$, and
- ρ is consistent with Prosecutor record R.

In fact, we will have $\rho = \rho_{\text{init}} \cup R$.

When Prosecutor asks about the value of x_e (edge e), Defendant does the following:

- If $e \in \text{dom}(\rho)$, answer according to $\rho(x_e)$. In this case $e \in \text{dom}(\rho_{\text{init}})$, otherwise Prosecutor knows the answer and would not ask.
- If $e \notin \text{dom}(\rho)$, answer with a value $b \in \{0, 1\}$ so that the odd charge in $G' = (V, E \setminus (\text{dom}(\rho) \cup \{e\}))$ stays in the large component. If this is not possible (since the large component disappeared), give up, or answer arbitrarily.

When Prosecutor forgets to get $R' \subsetneq R$, Defendant simply updates ρ to $\rho = \rho_{\text{init}} \cup R'$.

6 Proof of the Lower Bound

Now we want to implement the same lower bound strategy as for the pigeonhole principle. This means that we want to prove the following:

1. Before Prosecutor wins, there needs to be an informative record containing lots of edges.
2. Such records contain lots of information about Defendant's initial random choice.
3. Hence, any given informative record is exponentially unlikely to be consistent with a particular random choice.
4. So the strategy contains exponentially many (informative) records.

Let us deal with the first item on our to-do list right away.

Observation 6.1. *Before Prosecutor wins there must be a record with more than $cn/2$ edges.*

Proof. A winning position for Prosecutor is a falsified vertex constant, i.e., an odd-charge disconnected component of size 1.

Let E' be an initial random choice by Defendant. Let E'' be the edges in Prosecutor's record. As long as $|E' \cup E''| \leq cn$ Defendant is making sure the odd charge is in the large component. Hence, before Prosecutor wins we must have $|E' \cup E''| = |E''| + |E' \setminus E''| \geq cn$ or $|E''| \geq cn - |E' \setminus E''| \geq cn/2$. □

We call a record with $\geq cn/2$ edges *informative*. We want to prove that a fixed informative record R has exponentially small probability of being consistent with Defendant's initial random choice E'. For any edge $e \in R$ we have that

$$\Pr_{\rho_{\text{init}}}[e \in E'] = \frac{cn/2}{|E(G)|} \geq \frac{cn/2}{dn/2} = \frac{c}{d}\,. \tag{6.1}$$

By linearity of expectation it holds that

$$\text{E}_{\rho_{\text{init}}}[|\text{dom}(R) \cap E'|] \geq \frac{cn}{2} \cdot \frac{c}{d} = \frac{c^2}{2d} n\,. \tag{6.2}$$

By concentration of measure (the same kind of calculations as for the pigeonhole principle) we can now conclude that the random variable $|\text{dom}(R) \cap E'|$ is at least a half of the expected value except with

exponentially small probability; more formally, it holds that there exists an $\epsilon' > 0$ such that, for n large enough,

$$\Pr_{\rho_{\text{init}}}\left[|\text{dom}(R) \cap E'| \leq \frac{c^2}{4d}n\right] \leq 2^{-\epsilon' n}. \tag{6.3}$$

Fix $E_1 = \text{dom}(R) \cap E'$ and assume for now that $|E_1| \geq c^2 n/4d$. By construction we also have that $|E_1| \leq |E'| = cn/2$.

Lemma 6.2 (Key technical lemma). *Suppose that G is a (d, δ)-edge expander for $\delta > 1$ and let $E_1 \subseteq E(G)$ be any moderate-size edge set. Then there is a subset $E_2 \subseteq E_1$ of size at least $\gamma|E_1|$ for some $\gamma > 0$ such that if ρ is a uniformly randomly sampled charge-preserving assignment to E_1, it holds that all edges in E_2 are assigned uniformly and independently at random.*

Recall that by Proposition 4.5 G is also a (d, c)-connectivity expander for $c = \delta/4 = 1/4$, so E_1 is an edge set of moderate size if $|E_1| \leq cn$. Taking Lemma 6.2 on faith for now, we can prove Theorem 4.2 (for edge expansion $\delta > 1$).

Proof of Theorem 4.2. Let $\mathcal{S}$ be a complete strategy for Prosecutor for $Ts(G_n, \chi_n)$ for n large enough. Any game goes through an informative record R with probability 1. This record R is consistent with ρ_{init}. If we can prove for any fixed informative R that

$$\Pr_{\rho_{\text{init}}}[R \text{ and } \rho_{\text{init}} \text{ consistent}] \leq 2^{-\epsilon n}, \tag{6.4}$$

then it follows that the size of $\mathcal{S}$, which is at least the number of informative records in $\mathcal{S}$, is at least $2^{\epsilon n}$. We can estimate the probability in (6.4) by

$$\Pr_{\rho_{\text{init}}}[R \text{ and } \rho_{\text{init}} \text{ consistent}] \leq \Pr_{\rho_{\text{init}}}\left[|\text{dom}(R) \cap E'| \leq \frac{c^2}{4d}n\right] + \Pr_{\rho_{\text{init}}}\left[R \text{ and } \rho_{\text{init}} \text{ consistent} \middle| |\text{dom}(R) \cap E'| > \frac{c^2}{4d}n\right] \tag{6.5}$$

By (6.3) above we have that the first term on the right is upper-bounded by $2^{-\epsilon' n}$. To bound the second term we can apply Lemma 6.2 to deduce that at least $\gamma|E_1| \geq \gamma c^2 n/4d$ edges are set uniformly and independently at random. These edges agree with R with probability at most $(1/2)^{\gamma c^2 n/4d} = 2^{-\epsilon'' n}$ for some $\epsilon'' > 0$. Combining both observations, we get that

$$\Pr[R \text{ and } \rho_{\text{init}} \text{ consistent}] \leq 2^{-\epsilon' n} + 2^{-\epsilon'' n} \leq 2^{-\epsilon n} \tag{6.6}$$

for some $\epsilon > 0$ chosen small enough, which establishes Theorem 4.2. □

Lemma 6.2 follows from the following two lemmas. Recall that we say that a set of edges E' does not disconnect a graph $G(E, V)$ if the graph $G' = (V, E \setminus E')$ remaining after this edge set has been removed is still a connected graph.

Lemma 6.3. *For $G = (V, E)$ a (d, δ)-expander with $\delta > 0$, suppose that $E_1 \subseteq E(G)$ is a moderate-size set and that $E_2 \subseteq E_1$ does not disconnect G. Then uniformly random sampling of a charge-preserving assignment to E_1 gives a uniformly random sample of $\{0, 1\}^{|E_2|}$ for the edges in E_2.*

In what follows, let us use the shorthand $\{0, 1\}^{E_2}$ for the set of all possible assignments to edges $e \in E_2$ (which, as before, we identify with the variable set $\{x_e \mid e \in E_2\}$).

Lemma 6.4. *Let $G = (V, E)$ be a (d, δ)-expander with $\delta > 1$ and let $E_1 \subseteq E$ be any moderate-size set. Then there is a subset of edges $E_2 \subseteq E_1$ of size $|E_2| = \Omega(|E_1|)$ such that E_2 does not disconnect G.*

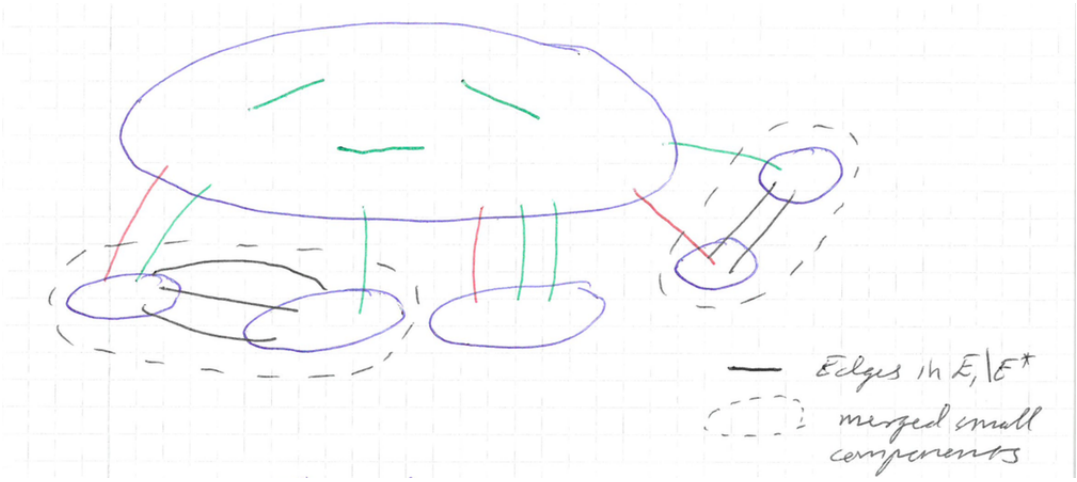


Figure 3: Obtaining E'

Strictly speaking, we might feel uncomfortable using the $\Omega(\cdot)$ asymptotic notation in the statement above, since we refer to a single graph. This is standard (ab)use of notation however, and above and in similar contexts it is simply taken to mean that that there exists some global constant $\gamma > 0$ such that the inequality $|E_2| \geq \gamma |E_1|$ holds.

Given these two lemmas, Lemma 6.2 follows immediately. Let us write down the details for completeness.

Proof of Lemma 6.2. Consider the set E_1 in Lemma 6.2, which is of moderate size. Lemma 6.4 guarantees that there exists a set E_2 of size $|E_2| \geq \gamma |E_1|$ for some $\gamma > 0$ such that $G' = (V, E \setminus E_2)$ is connected. Now Lemma 6.3 says that when we randomly sample a charge-preserving assignment to E_1 we get uniform and independent random bits in E_2. □

Lemma 6.3 is mostly some linear algebra juggling and is left as an exercise. Let us instead show Lemma 6.4.

Proof of Lemma 6.4. Let $E_1 \subseteq E(G)$ be any moderate-size set, i.e., such that $|E_1| \leq cn$ for $c = \delta/4$. Look at all the small connected components in $G' = (V, E \setminus E_1)$. Let the sum of their sizes be s. We get two cases depending on how large s is. Note that by assumption $s < n/2$.

Case 1: $s \leq |E_1|/2d$**.** The sum of sizes of small components is rather small. Then the total number of edges not completely inside the large component is at most $sd \leq |E_1|/2$ so if we pick E_2 to be the edges in E_1 between vertices in the large component we have $|E_2| \geq |E_1|/2$.

Case 2: $s > |E_1|/2d$**.** The sum of the sizes of the small components is larger, but not too large (by connectivity expansion). Remove from E_1 any edges not incident to the large component (black edges in Figure 3) to get $E' \subseteq E_1$. This might merge some small components, but they are still smallish and their total size s stays the same.

Observe that after our edge removal observation every edge in E' either is fully contained inside the large component or goes from a small component to the large component. There are no edges in E' between small components, because if so these small components would have been merged into a larger (though still small) component. Note that we have lost potentially quite a few edges in E_1, which seems dangerous, but we will prove that there are enough edges left anyway. The reason for this is that by edge expansion it holds that

$$|E'| \geq \delta s > \delta |E_1|/2d\,. \tag{6.7}$$

Fix one edge per (merged) small component going to the large component (red edges in Figure 3). Let $E'' = E' \setminus \{\text{red edges}\}$. Since $\delta > 1$, every small component has at least two edges to the large

components, so the red edges are at most half of E'. The remaining edges in E'' are the green edges in Figure 3. By construction, $G'' = (V, E \setminus E'')$ is connected, and using (6.7) we conclude that

$$|E''| \geq |E'|/2 > \frac{\delta |E_1|}{4d} = \Omega(|E_1|)\,. \tag{6.8}$$

The lemma follows. □

It remains to prove Lemma 6.3: if E' does not disconnect G, then sampling random charge-preserving assignments to a superset $E'' \supseteq E'$ yields uniform random bits on E'. We leave this as a linear algebra exercise, but we provide some useful background facts in an appendix to these notes.

A Linear Algebra Appendix

Suppose that V is a vector space over some field $\mathbb{F}$, such as the real numbers $\mathbb{R}$ or the complex numbers $\mathbb{C}$. Here we will have $\mathbb{F} = \mathbb{F}_2$, i.e., the field with two elements $0, 1$ such that

$$0 + 0 = 0 \qquad 0 + 1 = 1 \qquad 1 + 1 = 0 \tag{A.1}$$

$$0 \cdot f = 0 \qquad 1 \cdot f = f \tag{A.2}$$

(This field is also known under the name $\mathrm{GF}(2)$, the *Galois field* with two elements, but in TCS the notation $\mathbb{F}_2$ tends to be more common.) Most of the basic facts you have learned about vector spaces still hold in this setting.

An *affine subspace* A of V is a set $A = \{\mathbf{a} + \mathbf{u} \mid \mathbf{u} \in U\}$ for some fixed $\mathbf{a} \in V$ and some fixed (linear) subspace $U \subseteq V$. That is, an affine space is a linear space shifted by a constant vector. If $\mathbf{u}, \mathbf{v} \in A$, then $\mathbf{u} - \mathbf{v} \in U$, but in general it might be that $\mathbf{u} + \mathbf{v} \notin U$. Also, $\mathbf{u} + \mathbf{v}$ and $\mathbf{u} - \mathbf{v}$ might not be in A. If it is always the case that $\mathbf{u}, \mathbf{v} \in A$ implies that $\mathbf{u} + \mathbf{v} \in A$, then A is linear, i.e., we have an offset $\mathbf{a} = \mathbf{0}$.

For simplicity, in what follows let us focus on affine subspaces of $\{0,1\}^m = \mathbb{F}_2^m$. Since $+$ and $-$ is the same operation in $\mathbb{F}_2$, we have $\mathbf{u} + \mathbf{v} \in U$ in this setting if $\mathbf{u}, \mathbf{v} \in A$. For the same reason we have that if $\mathbf{u}, \mathbf{v}, \mathbf{w} \in A$, then $\mathbf{u} + \mathbf{v} + \mathbf{w} \in A$ for affine subspaces of $\mathbb{F}_2^m$.

Fact A.1. *Any affine subspace $A \subseteq \{0,1\}^m$ of dimension $n \leq m$ is generated by $M\mathbf{x} + \mathbf{a}$, for some fixed (but not unique) $m \times n$ matrix M of (full) rank n and some fixed size-m column vector $\mathbf{a}$, when we let $\mathbf{x}$ range over (all column vectors in) $\{0,1\}^n$. Uniform random sampling from A can be performed by choosing a uniformly random $\mathbf{x} \in \{0,1\}^n$.*

We will just accept this as true—you can take it as a definition if you like.

Proposition A.2. *Let $A \subseteq \{0,1\}^m$ be an affine subspace. Suppose for a subset of coordinates $S \subseteq [m]$ that all bitstrings in $\{0,1\}^s$ are supported by A (i.e., there are vectors $\mathbf{u} \in A$ that agree with any bit pattern in $\{0,1\}^s$). Then a uniformly random sample from A yields uniformly random and independent bits when restricted to $\{0,1\}^s$.*

The proof is left as an exercise, but let us hint at two possible solutions.

Approach 1 Argue that (by the way we have defined things) in any affine subspace $A \subseteq \{0,1\}^m$ any bit that is not fixed to 0 or 1 is 0 in exactly half of the vectors and 1 in exactly half of the vectors (why?). Repeat this on bit after bit in S, using that fixing a bit yields another affine subspace $A' \subseteq A$ (why?). Argue that every fixed bit pattern in $\{0,1\}^s$ must appear in a fraction $1/2^{|S|}$ of the vectors in A.

Approach 2 The rows $\{R_i \mid i \in S\}$ in M must be linearly independent (why?). This means that the submatrix consisting of these rows has rank $|S|$, and hence there exists a set of T columns, $|T| = |S|$, such that the submatrix on rows S and columns T is invertible. Argue that for any choice of the values of $\mathbf{x}$ outside of the coordinates in T the values of $M\mathbf{x} + \mathbf{a}$ over $\{0,1\}^T$, restricted to the coordinates in S, is one-to-one and hence uniform over random $\mathbf{x}$.

Phrased differently, Proposition A.2 says that if an affine subspace is supported on the uniform distribution of some set of coordinates S, then sampling from A uniformly at random and restricting to the coordinates in S yields the uniform distribution over $\{0,1\}^S$.

Observation A.3. *Let $G = (V, E)$ be any connected graph; let $\chi\colon V \to \{0,1\}$ be any odd-charge function; and let $E' \subseteq E$ be a set of edges such that $G' = (V, E \setminus E')$ has a (unique) connected component of size larger than $|V|/2$. Then the set of charge-preserving assignments $A \subseteq \{0,1\}^{E'}$ form an affine subspace.*

To prove this we need another observation.

Observation A.4. *Let $G = (V, E)$ be a connected graph with a charge function χ and let $E' \subseteq E$ be a minimal set of edges that disconnects G into two connected subgraphs G_1 and G_2. Then the total charges of the subgraphs G_1 and G_2 resulting from any assignment $\rho\colon E' \to \{0,1\}$ depend only on χ and on the parity of $\sum_{e\in E'} \rho(e)$.*

The proof of Observation A.4 is left as an exercise. Using this latter observation, however, it is straightforward, if a bit tedious, to prove Observation A.3.

Proof sketch for Observation A.3. Consider $G' = (V, E \setminus E')$. Let G_0 be the unique large component and G_i, $i = 1, \ldots, s$, the small connected components. Let E_{ij} be the edges between G_i and G_j. Look first at $G^0 = \left(V, E \setminus \bigcup_{j=1}^{s} E_{0j}\right)$. This yields one affine constraint per small connected component in G^0 requiring that the odd charge component is pushed into the large component. Now consider $E_{ij} \neq \emptyset$ in order for $1 \leq i < j \leq s$. If adding E_{ij} to the set of previously considered edges adds new connected components, i.e., splits one small component into two smaller ones, the constraint that both new small components should get even charge is an affine constraint on the edges E_{ij} and the previously assigned edges. All of these constraints can be collected in matrix form $M'\mathbf{y} = \mathbf{b}$, and the set of solutions can be written as $M\mathbf{x} + \mathbf{b}$ for some $\mathbf{x}$ of suitable dimension. □

Proposition A.5. *Suppose $G = (V, E)$ is a connected graph with an odd-charge function χ and let $E_1 \subseteq E$ be such that $G_1 = (V, E \setminus E_1)$ has a unique connected component of size strictly larger than $|V|/2$. Then for any fixed $E_2 \subseteq E_1$ the set of charge-preserving assignments to E_1 has full support on $\{0,1\}^{E_2}$ if and only if $G_2 = (V, E \setminus E_2)$ is connected.*

Proof. Let us do the "only if" direction first and then the "if" direction.

($\Leftarrow$) If G_2 is connected, then clearly we can assign all edges $e \in E_2$ arbitrarily, since there is only one component and its charge stays the same. Assignments to edges in $E_1 \setminus E_2$ will take care of pushing charges in the right direction if and when the graph is disconnected. Hence we have full support on $\{0,1\}^{E_2}$.

($\Rightarrow$) Pick a minimal set $E' \subseteq E_2$ that disconnects G. The parity of this set of edges must be such that the odd charge stays in the large components. This means that once we have assigned all edges in E' except one, this last edge has its value determined by the other edges (otherwise the odd charge will go into some small component). Hence, we can only have half of the assignments to $\{0,1\}^{E'}$ in any charge-preserving assignment and, in particular, do not have full support on $\{0,1\}^{E_2}$. □

Given the above observations and propositions, it is straightforward to combine them into a proof of Lemma 6.3. We leave the details to the reader as an exercise.

References

[Bol88] Béla Bollobás. The isoperimetric number of random regular graphs. *European Journal of Combinatorics*, 9(3):241–244, May 1988.

[HLW06] Shlomo Hoory, Nathan Linial, and Avi Wigderson. Expander graphs and their applications. *Bulletin of the American Mathematical Society*, 43(4):439–561, October 2006.

[Pud00] Pavel Pudlák. Proofs as games. *American Mathematical Monthly*, pages 541–550, 2000.

[Urq87] Alasdair Urquhart. Hard examples for resolution. *Journal of the ACM*, 34(1):209–219, January 1987.