RAZBOROV'S METHOD OF APPROXIMATIONS

W. T. GOWERS

1. Introduction

1.1. Open exposition problems

This document contains an exposition of Razborov's celebrated proof [3] that the monotone circuit complexity of the clique function is superpolynomial. It is closely related to some lectures I gave about this proof, but it is not a set of lecture notes in the usual sense, because my aim has very much not been to present the result in the concise form that the phrase "lecture notes" would suggest. When I first came across Razborov's proof, or rather a stronger version proved by Alon and Boppana [1], I was astonished that anybody could have thought of such an argument. As a result of lecturing it three times over the last fifteen years or so, I still find it breathtakingly clever, but I have come to understand better where the ideas came from. This does not in any way lessen my respect and admiration for the result (since I firmly believe that ideas *never* spring from nowhere), but I now think that I can explain how it could have come into existence without having to use the hypothesis that Razborov has supernatural powers.

To use a phrase of Timothy Chow [2], I am attempting to give a convincing solution to an *exposition problem*. As he puts it: "All mathematicians are familiar with the concept of an *open research problem*. I propose the less familiar concept of an *open exposition problem*. Solving an open exposition problem means explaining a mathematical subject in a way that renders it totally perspicuous. Every step should be motivated and clear; ideally, students should feel that they could have arrived at the results themselves." He also quotes Donald Newman (who used the term "natural proof", amusingly given that I am discussing a famous result of Razborov), who described such a proof as "not having any ad hoc constructions or *brilliancies*." Chow, incidentally, was attempting to solve the problem, "How did anyone think of forcing?" and his article, A Beginner's Guide to Forcing, is highly recommended.

A solution to an open exposition problem could be a straightforward historical account of how a proof was actually discovered, but what makes the concept interesting is that

an exposition problem can also be solved if one gives a plausible account of how a proof *could* have been discovered, even if was in fact not discovered that way. And indeed, in this case there are good reasons to suppose that the account I give here is *not* a faithful account of how the result was discovered. I do not know for sure, but my guess is that the true historical picture is like this: Razborov began by discovering an argument that gave a roughly quadratic lower bound for deciding whether a graph contains a triangle; he then generalized this argument to every fixed k and observed that it worked up to $k \approx \log n$; finally, Alon and Boppana modified Razborov's argument to obtain the proof discussed here that yields a stronger bound. I, however, shall present a way of coming up with Alon and Boppana's argument that does not involve thinking about small cases first.

Since Razborov is still young and alive, and since his proof has been read and digested by many people, I should perhaps add the qualification that the exposition problem I hope to have presented a solution to here was not truly open. However, I am not aware of a fully worked out solution to it in the literature.

1.2. **Statement of Razborov's theorem**

So as not to assume any background in the theory of computational complexity (though such a background is essential if you want to understand why this result is so interesting), let me begin with some notation and then an equivalent statement of Razborov's theorem.

Let $\Gamma(m,r)$ be the set of all graphs with vertex set $\{1,2,\dots,m\}$ that contain a clique of size r. If we let $n = \binom{m}{2}$ then we can identify $\Gamma(m,r)$ with a subset of the cube $\{0,1\}$ as follows. First, we fix a sensible bijection between subsets of $\{1,2,\dots,m\}$ of size 2 and $\{1,2,\dots,n\}$. For the sake of definiteness, we could write the subsets of $\{1,2,\dots,m\}$ in the lexicographical order $12,13,14,\dots,1m,23,24,\dots,2m,\dots,(m-1)m$ and map them to $1,2,\dots,n$, respectively. Let us call this bijection β. Next, for each graph G with vertex set $\{1,2,\dots,m\}$ we define a Boolean sequence $x \in \{0,1\}^n$ by setting x_i to be 1 if $\beta^{-1}(i)$ is an edge of G and 0 otherwise. Let us call this Boolean sequence $\beta(G)$. Finally, we identify $\Gamma(m,r)$ with $\beta(\Gamma(m,r))$. That is, we shall be interested in the set of all Boolean sequences $\beta(G)$ where G is a graph with vertex set $\{1,2,\dots,m\}$ that contains a clique of size r, and we shall call this $\Gamma(m,r)$ as well.

We shall be concerned with the following question. For $1 \leq i \leq n$ let E_i be the set of all Boolean sequences $x \in \{0,1\}^n$ such that $x_i = 1$. Thus, geometrically speaking, E_i is a half space (in the vector space $\mathbb{F}_2^n$). What Razborov showed is that it is impossible to build the

set $\Gamma(m,r)$ out of these half spaces by taking intersections and unions only a small number of times. Here is the theorem stated more precisely.

Theorem 1.1. *Let $E_1, E_2, \dots, E_M$ be a sequence of sets such that E_i is the half space $\{x : x_i = 1\}$ when $i \le n$, every other E_i can be written as $E_j \cap E_k$ or $E_j \cup E_k$ for some $j, k < i$, and $E_M = \Gamma(m,r)$. Then for suitably chosen r (depending on n), M is bounded below by a function that grows faster than any polynomial in n.*

The proof we present is a modification of Razborov's original argument due to Alon and Boppana. They show that if one takes r to be around $(1/4)(m/\log m)^{2/3}$, then M must be at least $\exp(C(m/\log m)^{1/3})$ for some constant $C > 0$, which is exponentially large in a power of m (and therefore also in a power of n). Razborov's original argument took r to be fixed and obtained a bound of m^r, up to logarithmic factors. Then optimizing r (to be around $\log m$) he obtained a bound of $m^{C \log m}$.

Very briefly, this result is important because a similar result where one was allowed to take complements as well as unions and intersections would prove that P$\neq$NP. Before Razborov's argument, the best known lower bound for M was $4n$, so this was a major breakthrough. (The best known lower bound when complements are allowed is still, unfortunately, only $3n$, and even a lower bound of $n \log\log n$ would be considered a huge advance.)

1.3. **A few ideas that don't work**

Before we dive in, let us think briefly about how one might try to prove a result of this kind. The obvious general plan of attack would be to define some notion of "complexity" with the following properties.

- The sets $E_1, \dots, E_n$ have very low complexity.
- If A and B have low complexity then so do $A \cap B$ and $A \cup B$ (and A^c in the case when complements are allowed).
- The set $\Gamma(m,r)$ has very high complexity.

If we could do that, and if we had good enough statements for the second and third items above, then we might hope to prove by induction an upper bound for E_M in terms of M, and to deduce from this that E_M has lower complexity than $\Gamma(m,r)$ unless M is large.

Just to show the kinds of difficulties that arise, let us try a simple-minded definition of complexity in order to see why it doesn't work. An obvious way in which the sets

$E_1, \dots, E_n$ are "simple" is that membership of E_i depends just on the one coordinate x_i. What about membership of $E_i \cap E_j$ or $E_i \cup E_j$? These depend just on the coordinates x_i and x_j. How about pursuing this idea and defining the complexity $\kappa(X)$ of a set X to be the number of coordinates you need to specify in order to demonstrate that x belongs or fails to belong to X?

It is not hard to see that $\kappa(A \cup B)$ and $\kappa(A \cap B)$ are both at most $\kappa(A) + \kappa(B)$. From this we can deduce by an easy inductive argument that $\kappa(E_{n+k})$ is at most 2^k if the sets E_i satisfy the hypotheses of Razborov's theorem. Since no set has complexity greater than n, this argument cannot lead to a lower bound that is better than $n + \log_2 n$.

However, it was not a complete waste of time to present it, because it gives a chance to point out a simple way of improving the argument, which is relevant to what Razborov did. Let us stick with the definition of $\kappa(X)$ just given, but let us improve our inductive hypothesis. The hypothesis we used above (without actually stating it) was that $\kappa(E_{n+k}) \leq 2^k$. But a much more sensible inductive hypothesis is the following: for every k there is a set I_k of size at most $2k$ such that membership of any of the sets $E_{n+1}, \dots, E_{n+k}$ just depends on the values of x_i with $i \in I_k$.

Now let us see what happens if we take the set E_{n+k+1}. It is the union or intersection of two earlier sets E_s and E_t. If $s \leq n$, then membership of E_s depends only on x_s, and otherwise we can determine it if we know those x_i with $i \in I_k$, and similarly for t. Therefore, membership of E_{n+k+1} depends only on those coordinates x_i for which $i \in I_k \cup (\{s,t\} \cap \{1,2,\dots,n\})$, which is a set of size at most $|I_k| + 2 \leq 2(k+1)$, so we are done.

From the above result it follows that $\kappa(E_{n+k}) \leq 2k$. The point to note about the proof was that in order to improve the first bound we replaced a *numerical* parameter by a *set-valued* parameter, which allowed us to exploit the fact that a union of two sets A and B has size significantly smaller than $|A| + |B|$ if A and B have a substantial overlap.

One can be more careful still about the inductive hypothesis and obtain an upper bound of $k + 1$. Given a collection of subsets $A_1, \dots, A_r \subset \{1, 2, \dots, n\}$, define a graph $H(A_1, \dots, A_r)$ to be the union of all the cliques $K(A_i)$ and let us define $\mu(A_1, \dots, A_r)$ to be the number of edges in the smallest graph that has the same components as $H(A_1, \dots, A_r)$. Let $J(A_1, \dots, A_r)$ be such a graph (which will be a spanning forest of $H(A_1, \dots, A_r)$).

Now let A_s be the set of all i such that membership of E_s has some dependence on the coordinate of x_i. Then A_s is equal either to $A_p \cup A_q$ for some $p, q < s$, or to some pair $\{i, j\}$, or to $\{i\} \cup A_p$ for some $p < s$. In each case, we have merged at most two components of

$H(A_1, \dots, A_r)$ to obtain $H(A_1, \dots, A_{r+1})$ (counting isolated vertices as singleton components), and it is easy to see that we can merge the same two components of $J(A_1, \dots, A_r)$ by adding a single edge.

It is well known that a connected graph with r vertices has at least $r - 1$ edges, so the above argument proves that A_r cannot have size more than $r + 1$.

Nevertheless, the point remains that this notion of complexity is useless for the purposes of proving Razborov's theorem. The same applies to all other simple-minded definitions. (If complements are allowed, then something much stronger can be said: to oversimplify a bit, it has been shown that no proof like this can work unless the notion of complexity is extremely strange.)

2. Approximating by sets in a lattice

Razborov's idea was as follows. You begin by defining a collection $\mathcal{L}$ of subsets of $\{0,1\}^n$ that you will later think of as "simple". This collection should contain the empty set, $\{0,1\}^n$ and all the half spaces $E_1, \dots, E_n$. Along with $\mathcal{L}$ you define two operations, $\sqcap$ and $\sqcup$, that approximate the operations $\cap$ and $\cup$. The resulting structure is known in the literature as a *legitimate lattice*.

You try to define your legitimate lattice $\mathcal{L}$ in such a way that all its sets have some special property from which you can deduce that none of them is "close" to the set $\Gamma(m, r)$. But that aim is in tension with the other aim, which is to make the operations $\sqcap$ and $\sqcup$ as close as possible to the usual operations $\cap$ and $\cup$. Why do these aims work against each other? Well, if those operations actually *were* $\cap$ and $\cup$ then you would be able to generate *all* monotone subsets of $\{0,1\}^n$, so the closer they are to $\cap$ and $\cup$, the more sets you are likely to be able to generate and the harder it will be to avoid getting close to $\Gamma(m, r)$.

There are two stages to the proof. The first is a fairly easy abstract result that says that *if* a suitable lattice $\mathcal{L}$ can be found, then Razborov's theorem follows. The second, which is the truly remarkable step (though as I shall point out, there is a certain non-obviousness to the first step as well), is actually to provide such a lattice and prove that it works. In this section we shall concentrate on the first step.

Suppose, then, that $\mathcal{L}$ is a legitimate lattice. Suppose that we build up a sequence of sets $E_1, \dots, E_n, E_{n+1}, \dots, E_M$ as above, and at the same time we build a parallel sequence of sets $F_1, \dots, F_n, F_{n+1}, \dots, F_M$ belonging to $\mathcal{L}$ by replacing the operations $\cap$ and $\cup$ by their approximations $\sqcap$ and $\sqcup$. What can we say about the symmetric difference between E_M and F_M?

To answer this question, let us try to express this symmetric difference in terms of earlier symmetric differences. Suppose first that $E_M = E_r \cap E_s$. Then $F_M = F_r \sqcap F_s$. Therefore

$$E_M \triangle F_M = (E_r \cap E_s) \triangle (F_r \sqcap F_s)$$

After our experiences with the naive notion of complexity, we want to give a "set-theoretic bound" for this, in the sense that we would like to prove that it is contained in a set that is not too large (in some sense that may or may not be straightforward cardinality). To that end, let us prove a sort of "set-theoretic triangle inequality".

Lemma 2.1. *Let A, B and C be three sets. Then $A \triangle C \subset (A \triangle B) \cup (B \triangle C)$.*

Proof. If $x \in A \setminus C$, then either $x \in A \setminus B$ or $x \in B \setminus C$, so $x \in (A \triangle B) \cup (B \triangle C)$. By symmetry, the same is true if $x \in C \setminus A$. □

Returning to our previous calculation, we can deduce that

$$(E_r \cap E_s) \triangle (F_r \sqcap F_s) \subset \Big((E_r \cap E_s) \triangle (F_r \cap F_s)\Big) \cup \Big((F_r \cap F_s) \triangle (F_r \sqcap F_s)\Big)$$

To deal with this, let us prove another very easy lemma, which we can think of as asserting the "set-theoretic continuity" of the operations $\cap$ and $\cup$.

Lemma 2.2. *Let A, B, C and D be sets. Then*

$$(A \cap B) \triangle (C \cap D) \subset (A \triangle C) \cup (B \triangle D)$$

and

$$(A \cup B) \triangle (C \cup D) \subset (A \triangle C) \cup (B \triangle D)$$

Proof. If $x \in (A \cap B) \setminus (C \cap D)$, then either $x \notin C$, in which case $x \in A \setminus C$, or $x \notin D$, in which case $x \in B \setminus D$. The first statement then follows by symmetry. The proof of the second statement is similar (or can be deduced from the first by looking at complements). □

Remark. In the back of one's mind here should be a dictionary that takes A, B, C, D to real numbers w, x, y, z, replaces symmetric differences such as $A \triangle B$ by differences such as $|x - y|$, replaces unions of symmetric differences by $+$, other unions and intersections by either $+$ or $-$, and $\subset$ by $\leq$. So for instance, the first statement in Lemma 2.2 could be regarded as similar to the inequality $|(w + x) - (y + z)| \leq |w - y| + |x - z|$. Similarly, Lemma 2.1 corresponds to the usual triangle inequality in $\mathbb{R}$.

Applying this to our calculation, we see that

$$(E_r \cap E_s) \triangle (F_r \cap F_s) \subset (E_r \triangle F_r) \cup (E_s \triangle F_s),$$

and putting everything together we deduce that

$$E_M \triangle F_M \subset (E_r \triangle F_r) \cup (E_s \triangle F_s) \cup \Big((F_r \cap F_s) \triangle (F_r \sqcap F_s)\Big).$$

A very similar argument shows that if $E_M = E_r \cup E_s$, then

$$E_M \triangle F_M \subset (E_r \triangle F_r) \cup (E_s \triangle F_s) \cup \Big((F_r \cup F_s) \triangle (F_r \sqcup F_s)\Big).$$

Therefore, we can adopt a set-theoretic inductive hypothesis that is in a similar spirit to the second inductive hypothesis of the previous section. We have more or less proved the inductive step, so let us state the inductive hypothesis as a lemma.

Lemma 2.3. *Let the sets E_i have the properties discussed above. Then for each $M > n$ there is a union U_M of at most $M - n$ "error sets", each of which is either of the form $(A \cap B) \triangle (A \sqcap B)$ or of the form $(A \cup B) \triangle (A \sqcup B)$, for two sets $A, B \in \mathcal{L}$, such that the sets $E_r \triangle F_r$ are all subsets of U when $r \leq M$.*

Proof. The result is trivial when $M \leq n$. Suppose we have proved the result up to $M - 1$. Then the above calculations show that there exist $r, s < M$ such that either $E_M \triangle F_M \subset U_{M-1} \cup \Big((F_r \cap F_s) \triangle (F_r \sqcap F_s)\Big)$ or $E_M \triangle F_M \subset U_{M-1} \cup \Big((F_r \cup F_s) \triangle (F_r \sqcup F_s)\Big)$. Either way, we have established the statement for M. □

3. How should we choose a legitimate lattice?

The lattice chosen by Alon and Boppana (which modifies the lattice chosen by Razborov) looks completely extraordinary if one just gives the definition, as Alon and Boppana do in their paper. And then everything magically works out. Here we shall approach the definition very slowly with the aim of removing as much as possible of the magic.

Let us concentrate first on the fact that we want the set $\Gamma(m, r)$ not to be easily approximable by a set in the lattice $\mathcal{L}$. We can say precisely what we mean by "approximable" here: given a set $A \subset \{0,1\}^n$, we define the *distance* $\rho(A, \mathcal{L})$ from A to $\mathcal{L}$ to be the smallest t such that we can write A as $L \cup U_t$, where $L \in \mathcal{L}$ and U_t is a union of t error sets of the type described in Lemma 2.3.

One of Razborov's ideas for how to achieve that was to choose $\mathcal{L}$ in such a way that no set in $\mathcal{L}$ would contain more than half the cliques of size r, except for the set $\{0,1\}^n$ itself (which is obliged to be in $\mathcal{L}$). If we do this, then the argument will naturally split into two parts. We want to show that if $\Gamma(m, r) = L \cup U_t$, then t must be very large, and it is natural to look at the two cases where L either is or is not the set $\{0,1\}^n$ (which we identify with the set of all graphs).

If L is *not* the set of all graphs, then L contains at most half the cliques of size r. In that case, we will be done if we can prove that the error sets all contain few cliques of size r, since the set $\Gamma(m,r)$ contains all cliques of size r. Notice that we are finding ourselves drawn to a notion of smallness that is not simply cardinality: we regard an error set as small if it contains few cliques of size r. (Once again, recall that a point in $\{0,1\}^n$ is identified with a graph with vertex set $\{1,2,\dots,m\}$.)

How about if L *is* the set of all graphs? In that case, it contains lots of graphs that are not in $\Gamma(m,r)$, so we somehow have to get rid of them by removing error sets. An obvious class of graphs that belong to L but not to $\Gamma(m,r)$ is complete $(r-1)$-partite graphs. So one fairly natural way to proceed is to try to show that the error sets also do not contain many of these.

3.1. **A more refined approximation lemma**

At this point, another idea enters the picture, which goes some way towards explaining why monotone circuits are easier to handle than general circuits. Suppose we define our operations $\sqcap$ and $\sqcup$ in such a way that for any $A,B\in\mathcal{L}$ we have $A\sqcap B\subset A\cap B$ and $A\sqcup B\supset A\cup B$. Let us write $\delta_\sqcap(A,B)$ for the set $(A\cap B)\setminus(A\sqcap B)$ and $\delta_\sqcup(A,B)$ for the set $(A\sqcup B)\setminus(A\cup B)$. Then we can modify the proof of the earlier approximation result and obtain the following lemma.

Lemma 3.1. *Let $\mathcal{L}$ be a legitimate lattice such that $A\sqcap B\subset A\cap B$ and $A\sqcup B\supset A\cup B$ for every $A,B\in\mathcal{L}$. Let $E_1,\dots,E_M$ be a sequence of sets such that E_i is the half space $\{x:x_i=1\}$ when $i\leq n$ and each subsequent E_i is either the intersection or the union of two earlier sets in the sequence. Let $F_1,\dots,F_M$ be the corresponding sequence in $\mathcal{L}$: that is, $F_i=E_i$ when $i\leq n$, $F_r=F_s\sqcap F_t$ if $E_r=E_s\cap E_t$, and $F_r=F_s\sqcup F_t$ if $E_r=E_s\cup E_t$. Then for each i between $n+1$ and M we can find a pair (A_i,B_i) of sets in $\mathcal{L}$ such that $E_r\setminus F_r\subset\bigcup_{i=n+1}^r\delta_\sqcap(A_i,B_i)$ and $F_r\setminus E_r\subset\bigcup_{i=n+1}^r\delta_\sqcup(A_i,B_i)$ for every $r\leq M$.*

Proof. We prove this result by induction on M. If $M\leq n$ then there is nothing to prove. Now suppose we know the result for $M-1$ and let (A_i,B_i) $(n+1\leq i\leq M-1)$ be the pairs of sets we have chosen. There are four cases to consider, with all four proofs being very similar. For completeness we shall give them all.

If $E_M = E_r \cap E_s$, then

$$\begin{aligned} E_M \setminus F_M &= (E_r \cap E_s) \setminus (F_r \sqcap F_s) \\ &\subset \Big((E_r \cap E_s) \setminus (F_r \cap F_s)\Big) \cup \Big((F_r \cap F_s) \setminus (F_r \sqcap F_s)\Big) \\ &\subset (E_r \setminus F_r) \cup (E_s \setminus F_s) \cup \delta_\sqcap(F_r, F_s). \end{aligned}$$

Let us therefore choose (A_M, B_M) to be (F_r, F_s) in this case, and we then know that $E_M \setminus F_M \subset \bigcup_{i=n+1}^{M} \delta_\sqcap(A_i, B_i)$, as required.

We also have that

$$\begin{aligned} F_M \setminus E_M &= (F_r \sqcap F_s) \setminus (E_r \cap E_s) \\ &\subset \Big((F_r \sqcap F_s) \setminus (F_r \cap F_s)\Big) \cup \Big((F_r \cap F_s) \setminus (E_r \cap E_s)\Big) \\ &\subset (F_r \setminus E_r) \cup (F_s \setminus E_s), \end{aligned}$$

where the proof of the last inclusion uses our assumption that $A \sqcap B \subset A \cap B$ for every $A, B \in \mathcal{L}$, so in this case we can deduce from the inductive hypothesis that $F_M \setminus E_M \subset \bigcup_{i=n+1}^{M-1} \delta_\sqcup(A_i, B_i)$, which is stronger than we need.

If $E_M = E_r \cup E_s$, then

$$\begin{aligned} E_M \setminus F_M &= (E_r \cup E_s) \setminus (F_r \sqcup F_s) \\ &\subset \Big((E_r \cup E_s) \setminus (F_r \cup F_s)\Big) \cup \Big((F_r \cup F_s) \setminus (F_r \sqcup F_s)\Big) \\ &\subset (E_r \setminus F_r) \cup (E_s \setminus F_s), \end{aligned}$$

which is good enough, by the inductive hypothesis, however we choose A_M and B_M.

Finally,

$$\begin{aligned} F_M \setminus E_M &= (F_r \sqcup F_s) \setminus (E_r \cup E_s) \\ &\subset \Big((F_r \sqcup F_s) \setminus (F_r \cup F_s)\Big) \cup \Big((F_r \cup F_s) \setminus (E_r \cup E_s)\Big) \\ &\subset (F_r \setminus E_r) \cup (F_s \setminus E_s) \cup \delta_\sqcup(F_r, F_s)., \end{aligned}$$

In this case we set $A_M = F_r$, $B_M = F_s$, and the inductive hypothesis implies that $F_M \setminus E_M \subset \bigcup_{i=n+1}^{M} \delta_\sqcup(A_i, B_i)$. □

3.2. **Why the refinement helps**

The great advantage of this result over Lemma 2.3 is that it allows us to choose the operation $\sqcap$ in such a way that the error sets $\delta_\sqcap(A,B)$ take care of cliques of size r, and the error sets $\delta_\sqcup(A,B)$ take care of the $(r-1)$-partite graphs. To see what this means, let us return to our hoped-for two-case argument. If L is not the set of all graphs, and if we define $\mathcal{L}$ in such a way that this implies that L contains at most $\frac{1}{2}\binom{m}{r}$ cliques of size r, then $\Gamma(m,r)\setminus L$ contains at least $\frac{1}{2}\binom{m}{r}$ cliques. So if the error sets $\delta_\sqcap(A,B)$ all contain very few cliques, then it is not possible to express $\Gamma(m,r)\setminus L$ as a union of few error sets of this type, so Lemma 3.1 implies that the number of operations used to create $\Gamma(m,r)$ must be large. If on the other hand, L is the set of all graphs, then $L\setminus\Gamma(m,r)$ contains all $(r-1)$-partite graphs, so Lemma 2.3 tells us that we are done if we can prove that each error set of the type $\delta_\sqcup(A,B)$ contains few $(r-1)$-partite graphs.

If we allowed complements, then we would not be able to separate out the two tasks in this way, which is why a proof of this type is possible for monotone circuit lower bounds and not for general circuit lower bounds.

3.3. **Sets whose minimal elements are cliques**

Let us now focus on the two requirements that every non-trivial $L\in\mathcal{L}$ contains at most $\frac{1}{2}\binom{m}{r}$ cliques of size r, and that the error sets $\delta_\sqcap(A,B)$ contain very few cliques.

We can clarify our task somewhat by observing that a monotone subset of $\{0,1\}^n$ is determined by its minimal elements. What are our requirements if we rephrase them in terms of the minimal elements of our sets?

Let $L\subset\{0,1\}^n$ and let A be the set of minimal elements of L. Thinking about L as a set of graphs, we are saying that A is some collection of graphs, no one of which contains another, and L is the set of all graphs that contain at least one graph in A. A clique $K(X)$ belongs to L if and only if there is some graph in A all of whose vertices of non-zero degree all belong to X. This implies that if we replace A by the set B of all cliques that contain a graph in A and then let C be the set of all minimal graphs in B, then the set of graphs containing an element of C will be the same as the set of graphs containing an element of A. So at least from the point of view of the first requirement, we could restrict attention to sets defined as follows. Take $\mathcal{A}$ to be a collection of subsets of $\{1,2,\dots,m\}$ and let $\lceil\mathcal{A}\rceil$ be the set of graphs that contain a clique $K(X)$ for some $X\in\mathcal{A}$.

Notice that $\lceil\mathcal{A}\rceil \cup \lceil\mathcal{B}\rceil = \lceil\mathcal{A}\cup\mathcal{B}\rceil$. It follows that, from the point of view of the complete $(r-1)$-partite side of the story, we don't really mind restricting attention to sets of this special form, since they are closed under unions. (Later we shall of course have to restrict further what sets we take and this will no longer be the case.)

But what about the error sets $\delta_\sqcap(A,B)$? Indeed, how should we define $\lceil\mathcal{A}\rceil \sqcap \lceil\mathcal{B}\rceil$? If we are restricting our attention to sets of the form $\lceil\mathcal{C}\rceil$ and trying to make our error set as small as possible, then we want $\mathcal{C}$ to be as large as possible subject to the condition that $\lceil\mathcal{C}\rceil$ is a subset both of $\lceil\mathcal{A}\rceil$ and $\lceil\mathcal{B}\rceil$. This condition tells us that every $Z \in \mathcal{C}$ must contain $X \cup Y$ for some $X \in \mathcal{A}$ and $Y \in \mathcal{B}$. So a natural choice for $\mathcal{C}$ is precisely that: we take all sets of the form $X \cup Y$ with $X \in \mathcal{A}$ and $Y \in \mathcal{B}$. Of course, we shall need to regard this as a temporary definition until we know exactly what we are taking as $\mathcal{L}$.

Before we do that, let us briefly think about cliques in the error set $(\lceil\mathcal{A}\rceil \cap \lceil\mathcal{B}\rceil) \setminus (\lceil\mathcal{A}\rceil \sqcap \lceil\mathcal{B}\rceil)$ if we adopt this definition. For a clique $K(W)$ to belong to this set, W must contain some $X \in \mathcal{A}$, and also some $Y \in \mathcal{B}$, but it must not contain any $X \cup Y$ with $X \in \mathcal{A}$ and $Y \in \mathcal{B}$.

Can *any* graph belong to the error set? The answer is going to have to be yes, or we would have an inductive proof that every set that could be computed by a monotone circuit was of the form $\lceil\mathcal{A}\rceil$, which is definitely not true. So let's give a very simple example. Let $\mathcal{A}$ be the graph that contains just the single edge 12 and let $\mathcal{B}$ be the graph that contains just the single edge 23. Then for a graph G to belong to $\lceil\mathcal{A}\rceil \cap \lceil\mathcal{B}\rceil$ we need G to contain the two edges 12 and 23. But for it to belong to $\lceil\mathcal{A}\rceil \sqcap \lceil\mathcal{B}\rceil$ we need it to contain the clique $K(\{1,2,3\})$. So any graph that contains 12 and 23 but not 13 will belong to the error set. However, *no clique* will belong to the error set, and this is what we care about.

Thus, so far we have a collection of sets and two operations that give rise to error sets that are empty of the graphs we mind about. This suggests two things. First, this collection of sets is obviously not going to do the job for us, because if it did then the results we could use it to prove would be far too strong. And indeed, it doesn't do the job for us because the set of all graphs that contain a clique of size r is of the form $\lceil\mathcal{A}\rceil$. But more positively, it suggests that restricting attention to sets of this form is a sensible thing to do.

3.4. **A closure operation and where it comes from**

Our plan now is to define $\mathcal{L}$ to be the collection of all sets $\lceil\mathcal{A}\rceil$ such that $\mathcal{A}$ satisfies some condition. But a major question that we have not even begun to think about is what condition we might take. In order to get us started, let us imagine that we have a class $\mathbb{A}$

of subsets $\mathcal{A}$ of the power set of $\{1,2,\dots,m\}$, and see what properties it would be good for $\mathbb{A}$ to have.

At this stage, I must apologize for how many levels of the set-theoretic hierarchy I am using. A property of $\mathbb{A}$ is a set of possible $\mathbb{A}$s. Each $\mathbb{A}$ is a set of possible $\mathcal{A}$s. Each $\mathcal{A}$ is a set of possible Xs. And each X is a set of elements of $\{1,2,\dots,m\}$. So I am implicitly thinking about sets of sets of sets of sets. The usual way to get one's head round a situation like this is to forget some of the bottom layers and to think of the top layers as properties rather than sets. So for example we could say that each set $\mathcal{A}$ is a *basis* for some monotone set (where this is a non-standard name for the set of minimal elements of the monotone set). We are trying to find some property of bases, and in order to do so we shall think about conditions that this property should satisfy.

To make this easier to think about, let us call $\mathcal{A}$ *simple* if it has the property in question, even though we do not know what the property is. So now we are asking what conditions the class of simple set systems should satisfy.

One of them is that if $\mathcal{A}$ is simple, then either $\lceil\mathcal{A}\rceil$ contains all graphs or it contains at most $\frac{1}{2}\binom{m}{r}$ cliques of size r. What could cause $\lceil\mathcal{A}\rceil$ to contain few cliques of size r? This is a question entirely about set systems: we want $\mathcal{A}$ to be such that at most half the sets of size r have a subset in $\mathcal{A}$.

A second condition is that if $\mathcal{A}$ and $\mathcal{B}$ are simple, then there should be a simple set $\mathcal{C}$ such that almost every set of size r that has a subset in $\mathcal{C}$ has subsets in $\mathcal{A}$ and $\mathcal{B}$ as well.

A third condition concerns the error sets in the other direction. If $\mathcal{A}$ and $\mathcal{B}$ are simple, then there should be some simple set $\mathcal{C}$ that contains $\mathcal{A}\cup\mathcal{B}$, and almost every complete $(r-1)$-partite graph that contains $K(Y)$ for some $Y\in\mathcal{C}$ should also contain $K(X)$ for some $X\in\mathcal{A}\cup\mathcal{B}$.

It would be nice to arrive at a definition without making any wild guesses. The first two conditions seem a bit too general to suggest any particular notion of simplicity, but we can get some mileage out of the third, because it is not just a necessary condition but also a sufficient one for the error sets $\delta_{\sqcup}(A,B)$ to be small.

What it tells us is that every time we take a union such as $\lceil\mathcal{A}\rceil\cup\lceil\mathcal{B}\rceil$, we can afford to throw in extra sets, provided that as we do so we increase only very slightly the number of complete $(r-1)$-partite graphs that contain $K(X)$ for one of the sets in our collection.

Let us write $[\mathcal{A}]$ for the set of complete $(r-1)$-partite graphs that contain $K(X)$ for some $X\in\mathcal{A}$. Under what circumstances can we add sets to $\mathcal{A}\cup\mathcal{B}$ and increase by only a very small amount the size of $[\mathcal{A}\cup\mathcal{B}]$? To get some purchase on this question, let us

consider what happens if we add just *one* set. Let us write $\mathcal{C}$ for $\mathcal{A} \cup \mathcal{B}$, and let Z be the set we add.

A useful way of regarding a complete $(r-1)$-partite graph is as a colouring of the set $\{1, 2, \dots, m\}$ with $r-1$ colours (not necessarily all used), or equivalently as a function from $\{1, 2, \dots, m\}$ to $\{1, 2, \dots, r-1\}$. A colouring belongs to $[\mathcal{C}]$ if and only if there exists $X \in \mathcal{C}$ such that every element of X gets a different colour. If every element of X gets a different colour, let us say that X is *properly coloured*. So a colouring belongs to $[\mathcal{C} \cup \{Z\}] \setminus [\mathcal{C}]$ if Z is properly coloured but no set in $\mathcal{C}$ is properly coloured.

The statement that adding Z to $\mathcal{C}$ does not introduce too many new complete $(r-1)$-partite graphs into $[\mathcal{C}]$ can be reformulated as follows: if you choose a random $(r-1)$-colouring of $\{1, 2, \dots, m\}$, then the probability that Z is properly coloured and no set in $\mathcal{C}$ is properly coloured is small.

It would be nice to know what sets Z are likely to have this property. It seems that the more Z intersects with the existing sets in $\mathcal{C}$, the harder it will be for Z to be properly coloured and all the sets in $\mathcal{C}$ to fail to be properly coloured.

But how can we estimate this probability? Well, one way of doing it is to calculate exactly the probability that Z is properly coloured (since this is an easy calculation) and then try to estimate the probabilty that, given this information, no set in $\mathcal{C}$ is properly coloured.

However, this conditional probability still seems hard to estimate, because it depends on the precise way that the various sets in $\mathcal{C}$ intersect. So let us go a stage further and enumerate the elements of $\mathcal{C}$ as $X_1, \dots, X_N$, and then think about the conditional probabilities $\mathbb{P}[X_1 \text{ not PC} | Z \text{ PC}]$, $\mathbb{P}[X_2 \text{ not PC} | X_1 \text{ not PC}, Z \text{ PC}]$, and so on.

It is not hard to work out the first of these probabilities in terms of the sizes of the sets $X_1 \setminus Z$ and $X_1 \cap Z$. However, the second one becomes more complicated because it depends on the sizes of $X_2 \setminus Z$, $X_2 \cap Z$ and $(X_1 \cap X_2) \setminus Z$. And the situation gets rapidly worse as more sets are added. In general, the more the sets $X_i \setminus Z$ overlap, the less small the main probability that we are trying to estimate will be.

This difficulty suggests a possible definition. Obviously our lives would be much easier if the events $[X_i \text{ not PC} | Z \text{ PC}]$ were independent. And this will be the case if $X_i \cap X_j \subset Z$ for every $i \neq j$. So we know that we can throw in Z if we can find a reasonably large collection of sets $X_1, \dots, X_s \in \mathcal{C}$ such that $X_i \cap X_j \subset Z$ for every $i \neq j$ and the individual probabilities $\mathbb{P}[X_i \text{ not PC} | Z \text{ PC}]$ are not too close to 1.

This suggests a way of defining a notion of simplicity. We say that a set-system $\mathcal{C}$ *generates* Z if there exist $X_1, \dots, X_s \in \mathcal{C}$ such that $X_i \cap X_j \subset Z$ whenever $i \neq j$. If the numbers work out well enough, this will imply that the probability that Z is properly coloured and no set in $\mathcal{C}$ is properly coloured is tiny. This will allow us to keep adding sets that are generated by the set-system so far, and potentially allows us to define a simple set as one that is *closed*: that is, one that already contains all the sets that it generates. We can also define the *closure* $\mathcal{C}^*$ of a set $\mathcal{C}$ to be the smallest closed set that contains $\mathcal{C}$, or equivalently the set that you obtain by adding sets that you can generate until you end up with a closed set.

For this to work, it will be necessary that we do not add *too* many sets to $\mathcal{C}$, since a huge number of tiny probabilities can add up to something that is no longer tiny. So now we have a problem: in the worst case, how many sets does one need to add to $\mathcal{A} \cup \mathcal{B}$ to make it closed? We shall return to this question later.

3.5. **A further restriction on the sets in $\mathcal{L}$**

Let me briefly clear up a potential confusion. On the one hand, I have suggested that the set-systems I am looking at are *antichains*, that is, set-systems $\mathcal{C}$ such that if $X \in \mathcal{C}$ and X is a proper subset of Y, then $Y \notin \mathcal{C}$. But the closure operation does not take antichains to antichains, so what is going on?

A simple way round this apparent difficulty is to talk not about antichains but about monotone set-systems. This is equivalent because an antichain $\mathcal{C}$ can be replaced by the collection of all Y that contain some $X \in \mathcal{C}$, and a monotone system $\mathcal{D}$ can be replaced by the set of all its minimal elements. It is easy to check that these two operations are inverse to each other. It is also trivial that the closure of a monotone set-system is itself a monotone set-system.

Now let us make an observation that is quite encouraging. One might wonder whether choosing a generous closure operation could cause problems on the cliques side. That is, could it lead us to accept non-trivial sets that contain more than $\frac{1}{2}\binom{m}{r}$ cliques of size r, and could it lead to error sets of the form $\delta_{\sqcap}(A, B)$ that are too large?

The answer to the first question is that the more sets we allow ourselves to generate in a closure operation, the fewer set-systems are closed. So we cannot do this kind of harm on the cliques side. And the answer to the second question is that if we think of $\mathcal{A}$ and $\mathcal{B}$ as monotone sets (rather than antichains), then $\lceil \mathcal{A} \rceil \sqcap \lceil \mathcal{B} \rceil$ turns out to be $\lceil \mathcal{A} \cap \mathcal{B} \rceil$, if we use our earlier definition. Furthermore, if $\mathcal{A}$ and $\mathcal{B}$ are closed, then so is $\mathcal{A} \cap \mathcal{B}$. So there

really is no reason not to make a closure operation as generous as we can, subject to the condition that the error sets $\delta_\sqcup(A,B)$ are small.

However, we still have work to do, because as things stand, the error sets $\delta_\sqcap(\lceil\mathcal{A}\rceil,\lceil\mathcal{B}\rceil)$ contain no cliques, and it seems rather unlikely that we can get away with that. It might be interesting to try to find a closed set, under the current definition, that is a reasonably good approximation to the set of all graphs containing a clique of size r. (At the time of writing I would expect such an example to exist, but I don't know for certain that it does – it could be that the problem is merely in proving that it it does not.) But let us leave that problem aside and try to think whether we could afford to throw away some sets from $\lceil\mathcal{A}\cap\mathcal{B}\rceil$ when we come to define $\lceil\mathcal{A}\rceil\sqcap\lceil\mathcal{B}\rceil$, in order to exploit the fact that we do not need it to contain *no* cliques of size r, but just to contain *few* cliques of size r.

We find ourselves asking the following general question: given a set-system $\mathcal{C}$, when can we say that a particular set $X\in\mathcal{C}$ is making a small contribution to the number of cliques in the set $\lceil\mathcal{C}\rceil$? Obviously if X has a proper subset Y that also belongs to $\mathcal{C}$, then X makes no further contribution, so to think about this question it is more convenient to formulate it in terms of antichains rather than monotone set-systems.

It is not very easy to think of any generalizations of this condition: one could try to come up with circumstances where very few graphs that contain a clique $K(X)$ do not also contain $K(Y)$ for some other $Y\in\mathcal{C}$, but no simple set of such circumstances suggests itself.

Or rather, nothing suggests itself except perhaps the ridiculously simple idea that we could think about dropping X if there are very few cliques of size r that contain $K(X)$. Of course, this just means that we should drop all large X.

The symmetry of the problem suggests that if we do that then the only sensible way of doing it will be to choose some cutoff l and drop X if it has size greater than l. But is there any point in doing this?

At first sight, it might seem as though this line of thought is going nowhere. After all, if we fix some $l<r$, then every clique of size r contains a clique of size $l+1$, so even if each individual $K(X)$ with $|X|=l$ makes a small contribution to the number of r-cliques in $\lceil\mathcal{C}\rceil$, the total contribution from all of them is not small. However, this argument is not quite right, since we are not looking at arbitrary set-systems $\mathcal{C}$ but at *closed* set-systems, and it is not clear that every X of size $l+1$ could be a minimal element of $\mathcal{C}$.

We are now ready to start proving lemmas, all of them fully motivated in advance.

4. Minimal elements of closed sets

The remarks at the end of the previous section make it very clear that we should think about the following question: what can we say about the set of minimal elements of a closed set-system $\mathcal{C}$? More precisely, since the number of cliques containing $K(X)$ depends only on the size of X, we should ask ourselves how many minimal elements of size k a closed set-system can contain. If the answer turns out to be that it cannot contain very many, then we will be able to get rid of minimal sets of large cardinality.

So far, I have given only a rather weak reason for wanting to do that, which is that it seems too much to expect that we could get away with not doing so. But let me give a much stronger reason as well. Earlier, we were a bit worried that we didn't have an upper bound on the number of times we might need to add sets in order to get from $\mathcal{A}\cup\mathcal{B}$ to $(\mathcal{A}\cup\mathcal{B})^*$. But if we could throw away all sets of size greater than l, then we would have a *trivial* upper bound of $\binom{m}{l}$, which might well be good enough, given that we are free to choose s. (At this stage, there is absolutely no guarantee that any of these ideas will work, but trying things to see whether they work is part of the natural process of mathematical discovery.)

If $\mathcal{D}$ is the set of minimal elements of a closed set-system $\mathcal{C}$, then $\mathcal{D}$ is an antichain, and also it is not possible to find s sets $X_1,\dots,X_s\in\mathcal{D}$, a further set $X\in\mathcal{D}$ (not necessarily distinct from all the X_i), and a proper subset $Y\subset X$, such that $X_i\cap X_j\subset Y$ whenever $i\neq j$. Let us call this property $P(s)$.

The next lemma tells us the maximum cardinality of a collection of sets of size k that has property $P(s)$. It may seem rather miraculous that the closure operation, which we did not design with the following lemma in mind, gives rise to an exact bound with a simple proof, and indeed, this is a phenomenon that I cannot explain except by saying that we have been lucky. However, if such luck makes you feel uneasy, you may be reassured to know that we do not depend on it too heavily: what really matters is that we obtain a bound that is independent of the size of the ground set (which in our case is $\{1,2,\dots,m\}$), and a little thought makes it highly plausible that such a bound should exist: after all, the condition $P(s)$ implies that we cannot find $s+1$ sets such that any two of them have the same intersection. A collection of such sets is called a *sunflower* with $s+1$ petals, and a result of Erdős and Rado implies that a system of sets of size k with no sunflower with $s+1$ petals has size at most $k!s^k$. We could afford to use this bound, but it just happens that we can prove the next lemma fairly easily, so we might as well do so.

Before we give the lemma, here is an example that shows that it is best possible. Let $W_1, \dots, W_k$ be disjoint sets of size $s-1$ and let $\mathcal{D}$ be the collection of all sets X such that $|X \cap W_i| = 1$ for every i. Then if $X_1, \dots, X_s$ and X belong to $\mathcal{D}$ and Y is a proper subset of X, there must exist i such that $Y \cap W_i = \emptyset$ and by the pigeonhole principle there must be some p and q such that $X_p \cap X_q \cap W_i \neq \emptyset$. Therefore, $\mathcal{D}$ has property $P(s)$ and has cardinality $(s-1)^k$.

Lemma 4.1. *Let $\mathcal{D}$ be a collection of sets of size k that satisfies property $P(s)$. Then $|\mathcal{D}| \leq (s-1)^k$.*

Proof. We prove the result by induction on s. First of all, if $s = 1$, then if X is a set of size k that belongs to $\mathcal{D}$, then $\emptyset$ is a proper subset of X and $\mathcal{D}$ fails property $P(s)$ for trivial reasons (because one cannot find two distinct terms in a sequence of length 1). So $\mathcal{D}$ has to be empty and the result holds in this case. (If $k = 0$ then $\mathcal{D}$ could be $\{\emptyset\}$, so we shall interpret $(s-1)^k$ to be 1 when $s = 1$ and $k = 0$.)

Now let us suppose that we know the result for $s-1$, and let $\mathcal{D}$ be a set-system with property $P(s)$. Let Z be an arbitrary element of $\mathcal{D}$. For each $W \subset Z$, let $\mathcal{D}_W$ be the collection of all sets $X \setminus Z$ such that $X \cap Z = W$. Note that these sets are non-empty, since $\mathcal{D}$ is an antichain.

Let us prove that $\mathcal{D}_W$ satisfies property $P(s-1)$. If $U_1, \dots, U_{s-1}$ and U are elements of $\mathcal{D}_W$, V is a proper subset of U, and $U_i \cap U_j \subset V$ for every $i \neq j$, then all the sets $U_i \cup W$ and $U \cup W$ belong to $\mathcal{D}$, $U \cup V$ is a proper subset of $U \cup W$, and $(U_i \cup W) \cap (U_j \cup W) \subset V \cup W$ for every $i \neq j$. Also, $Z \cap (U_i \cup W) = W \subset V \cup W$ as well, which contradicts the assumption that $\mathcal{D}$ has property $P(s)$.

Since every element of $\mathcal{D}_W$ has size $k - |W|$, our inductive hypothesis implies that $|\mathcal{D}_W| \leq (s-2)^{k-|W|}$. It follows that

$$|\mathcal{D}| \leq \sum_{W \subset Z} (s-2)^{k-|W|} = \sum_{j=0}^{k} \binom{k}{j} (s-2)^{k-j} = (s-1)^k$$

by the binomial theorem. The result is proved. □

Let us see what this tells us about the contribution to the number of cliques of size r made by minimal elements of $\mathcal{C}$ of size greater than l.

Corollary 4.2. *Let $\mathcal{C}$ be a closed set-system, and suppose that $2r(s-1) \leq m$. Then the number of sets of size r that contain an element of $\mathcal{C}$ but that do not contain a element of $\mathcal{C}$ of size at most l is at most $2^{-l}\binom{m}{r}$.*

Proof. Let X be a subset of $\{1,2,\dots,m\}$ of size j. The number of subsets of $\{1,2,\dots,m\}$ of size r that contain X is $\binom{m-j}{r-j}$, which is at most $(r/m)^j\binom{m}{r}$. Therefore, the number of sets of size r that contain a minimal element of $\mathcal{C}$ of size j is, by the previous lemma, at most $(r(s-1)/m)^j\binom{m}{r} \leq 2^{-j}\binom{m}{r}$, by our assumption (which was of course made with a view to this number being small). Since every set that contains an element of $\mathcal{C}$ contains a minimal element of $\mathcal{C}$, it follows that the number of sets of size r that contain an element of $\mathcal{C}$ but not a set of size at most l in $\mathcal{C}$ is at most $\sum_{j>l} 2^{-j}\binom{m}{r} = 2^{-l}\binom{m}{r}$, as claimed. □

This comes very close to proving a fact that we were hoping for: that if $\mathcal{C}$ is a closed set, then either $\lceil\mathcal{C}\rceil$ consists of all graphs or it contains at most $\frac{1}{2}\binom{m}{r}$ cliques of size r. From the corollary we can read off that the number of sets of size r that contain an element of $\mathcal{C}$ but do not contain an element of size at most 1 is at most $\frac{1}{2}\binom{m}{r}$. But what about the ones that *do* contain a singleton?

The simplest way to deal with this is to make the not unreasonable convention that if X is a singleton, then every graph contains the clique $K(X)$. After all, $K(X)$ contains no edges, so if we think of it as a set of edges, then it is the empty set. So we shall indeed adopt this convention. Therefore, if $\mathcal{C}$ is a closed set, then either $\lceil\mathcal{C}\rceil$ is the set of all graphs, or it contains at most $\frac{1}{2}\binom{m}{r}$ cliques.

But our main reason for proving Corollary 4.2 was to prove that we could afford to throw away sets of size greater than l. More precisely, it tells us that we can afford to define the operation $\sqcap$ as follows. If $\mathcal{A}$ and $\mathcal{B}$ are two closed sets, then let $\mathcal{C}$ be the set of all minimal elements of $\mathcal{A}\cap\mathcal{B}$ (which, since we assume that $\mathcal{A}$ and $\mathcal{B}$ are monotone, is the collection of all minimal sets of the form $X\cup Y$ with $X\in\mathcal{A}$ and $Y\in\mathcal{B}$). Let $\mathcal{D}$ be the set of all sets in $\mathcal{C}$ of size at most l, and let $\lceil\mathcal{A}\rceil\sqcap\lceil\mathcal{B}\rceil = \lceil\mathcal{D}\rceil$.

Unfortunately, there doesn't seem to be any reason for the collection of all sets that have a subset in $\mathcal{D}$ to be closed. We can get round this problem by restricting our attention entirely to sets of size at most l. Write $[m]^{(\leq l)}$ for the set of all subsets of $\{1,2,\dots,m\}$ of size at most l. Instead of looking at monotone sets, let us look at monotone subsets of $[m]^{(\leq l)}$, meaning set-systems $\mathcal{A}$ such that every set in $\mathcal{A}$ has size at most l, and if $X\in\mathcal{A}$, $X\subset Y$, and $|Y|\leq l$, then $Y\in\mathcal{A}$. Now we can define $\lceil\mathcal{A}\rceil\sqcap\lceil\mathcal{B}\rceil$ to be $\lceil\mathcal{A}\cap\mathcal{B}\rceil$.

To make this work, we also redefine the closure of $\mathcal{A}$ to be the collection of all sets *of size at most l* that can be generated from $\mathcal{A}$ and previously generated sets.

We have arrived at our definition of $\mathcal{L}$. It consists of all sets of the form $\lceil\mathcal{A}\rceil$, where $\mathcal{A}$ is a closed subset of $[m]^{(\leq l)}$. This definition depends on the parameters l and s (which comes in via the definition of the closure operation). We shall choose these parameters later to

optimize the bound we obtain. We stress that the restriction to sets of size at most l is not arbitrary and unmotivated: we wanted to restrict the possible minimal elements that might occur, we realized that we could afford to throw away sets of size greater than l, we had a problem about sets ceasing to be closed, and we fixed the problem by restricting attention to subsets of $[m]^{(\leq l)}$.

Just to be completely explicit about it, the two operations are defined as follows. Given closed subsets $\mathcal{A}$ and $\mathcal{B}$ of $[m]^{(\leq l)}$, we let $\lceil\mathcal{A}\rceil \sqcap \lceil\mathcal{B}\rceil$ be $\lceil\mathcal{A}\cap\mathcal{B}\rceil$ and we let $\lceil\mathcal{A}\rceil \sqcup \lceil\mathcal{B}\rceil$ be $\lceil(\mathcal{A}\cup\mathcal{B})^*\rceil$, where the closure is now defined in the new way that restricts attention to sets of size at most l.

After all this justification, let us quickly prove that the error sets $\delta_\sqcap(A,B)$ are small.

Corollary 4.3. *Let $\mathcal{A}$ and $\mathcal{B}$ be two closed subsets of $[m]^{(\leq l)}$. Then $(\lceil\mathcal{A}\rceil \cap \lceil\mathcal{B}\rceil) \setminus \lceil\mathcal{A}\cap\mathcal{B}\rceil$ contains at most $4.2^{-l/2}\binom{m}{r}$ cliques of size r.*

Proof. Every clique in $(\lceil\mathcal{A}\rceil \cap \lceil\mathcal{B}\rceil) \setminus \lceil\mathcal{A}\cap\mathcal{B}\rceil$ has a vertex set Z that contains a minimal element X of $\mathcal{A}$ and a minimal element Y of $\mathcal{B}$ but no element of $\mathcal{A}\cap\mathcal{B}$. Since $\mathcal{A}$ and $\mathcal{B}$ are monotone sets, $X\cup Y$ will belong to both $\mathcal{A}\cap\mathcal{B}$ unless it has size greater than l. Therefore, at least one of X and Y has size greater than $l/2$. It follows that every clique in $(\lceil\mathcal{A}\rceil \cap \lceil\mathcal{B}\rceil) \setminus \lceil\mathcal{A}\cap\mathcal{B}\rceil$ contains either a minimal element of $\mathcal{A}$ of size greater than $l/2$ or a minimal element of $\mathcal{B}$ of size greater than $l/2$. As in the proof of Corollary 4.2, this implies that the number of such cliques of size r is at most $2\sum_{j>l/2} 2^{-j}\binom{m}{r}$, which is at most $4.2^{-l/2}\binom{m}{r}$, as claimed. □

5. The error sets $\delta_\sqcup(A,B)$

We designed the closure operation in such a way that adding a set X to a set-system $\mathcal{C}$ that generates X increases by only a very small amount the number of complete $(r-1)$-partite graphs contained in $\lceil\mathcal{C}\rceil$. Indeed, we even knew how we were intending to go about proving this. The only thing we have not done is carry out the argument and see whether it is good enough, but we have good reason to suppose that it will be if we choose s large enough. So let us go ahead and prove what we need.

Recall that we reformulated our task as follows: we would like to prove that if $\mathcal{C}$ is a set-system and κ is a random $(r-1)$-colouring of $\{1,2,\dots,m\}$, then the probability that some set in $\mathcal{C}^*$ is properly coloured but no set in $\mathcal{C}$ is properly coloured is tiny.

Lemma 5.1. *Let $\mathcal{C}$ be a subset of $[m]^{(\leq l)}$ and let κ be a random $(r-1)$-colouring of $\{1,2,\dots,m\}$. Suppose that $l^2 \leq r/4$. Then the probability that some set in $\mathcal{C}^*$ is properly coloured but no set in $\mathcal{C}$ is properly coloured is at most $2^{-s}m^l$.*

Proof. Let us produce $\mathcal{C}^*$ from $\mathcal{C}$ by adding a sequence of sets $Y_1,\dots,Y_M$, each one generated by the sets in $\mathcal{C}$ or sets that have appeared earlier in the sequence. Since $\sum_{j=0}^{l}\binom{m}{j} \leq m^l$, we will be done if we can prove that for each set we add, the probability that it is properly coloured and no earlier set is properly coloured is at most 2^{-s}.

If we generate a set X, then there must be sets $X_1,\dots,X_s$ already generated, or in $\mathcal{C}$, such that $X_i \cap X_j \subset X$ for every $i \neq j$. Then

$$\mathbb{P}[X \text{ is PC and no } X_i \text{ is PC}] = \mathbb{P}[X \text{ is PC}]\,\mathbb{P}[\text{No } X_i \text{ is PC}|X \text{ is PC}].$$

The probability that X is properly coloured is easy to calculate, but it turns out not to have much effect on the final bound, so we shall not bother. As for the probability that no X_i is properly coloured, given that X is properly coloured, we know from the fact that the sets $X_i \setminus X$ are disjoint that the events $[X_i \text{ is not PC}|X \text{ is PC}]$ are independent, so we shall just multiply their probabilities together.

What, then, is the probability that X_i is not properly coloured, given that X is properly coloured? Let $|X_i \cap X| = p_i$ and $|X_i \setminus X| = q_i$. Then the colours of the elements of $X_i \cap X$ are guaranteed to be distinct, so for X_i to be properly coloured we have to colour the remaining q_i points with distinct colours chosen from the $r-1-p_i$ colours still allowed. The probability that we do this is at least

$$\frac{r-1-p_i}{r-1}\frac{r-1-p_i-1}{r-1}\cdots\frac{r-1-p_i-q_i+1}{r-1}.$$

Each bracket is at least $\frac{r-1-l}{r-1} = 1-\frac{l}{r-1}$, and there are at most l brackets, so the probability is at least $(1-\frac{l}{r-1})^l$, which is at least $1/2$ if $l^2 \leq r/4$. Therefore, the probability that no X_i is properly coloured given that X is properly coloured is at most 2^{-s}. The result follows. □

What we have shown if we put all these calculations together is that the monotone circuit complexity of the function that is 1 if a graph on m vertices contains a clique of size r and 0 otherwise is at least the minimum of $m^{-l}2^s$ and $2^{l/2}/8$, provided that $2r(s-1) \leq m$ and $l^2 \leq r/4$. A brief calculation shows that we can maximize this minimum by taking r to be about $m^{2/3}(\log n)^{-2/3}$, s to be about $m^{1/3}(\log n)^{2/3}$ and l to be about $m^{1/3}(\log n)^{-1/3}$. The resulting lower bound is then $\exp(c(m/\log m)^{1/3})$ for some absolute constant $c > 0$.

In terms of n, this gives us a lower bound of $\exp(cn^{1/6}/(\log n)^{1/3})$ for the monotone circuit complexity of a function in NP, so the problem is solved.

6. References

[1] N. Alon and R. Boppana, *The monotone circuit complexity of Boolean functions*, Combinatorica **7** (1987), 1-23.

[2] T. Chow, *A beginner's guide to forcing*, in Communicating Mathematics, Contemp. Math, **479** (2009), 25-40.

[3] A. A. Razborov, *Some lower bounds for the monotone complexity of some Boolean functions*, Soviet Math. Dokl. **31** (1985), 354-357.