Electronic Colloquium on Computational Complexity, Report No. 15 (2012)

Degree Lower Bounds of Tower-Type for Approximating Formulas with Parity Quantifiers

Albert Atserias
Universitat Politècnica de Catalunya,
Barcelona, Spain.

Anuj Dawar
University of Cambridge,
Cambridge, UK.

February 22, 2012

Abstract

Kolaitis and Kopparty have shown that for any first-order formula with parity quantifiers over the language of graphs there is a family of multi-variate polynomials of constant-degree that agree with the formula on all but a $2^{-\Omega(n)}$-fraction of the graphs with n vertices. The proof yields a bound on the degree of the polynomials that is a tower of exponentials of height as large as the nesting depth of parity quantifiers in the formula. We show that this tower-type dependence on the depth of the formula is necessary. We build a family of formulas of depth q whose approximating polynomials must have degree bounded from below by a tower of exponentials of height proportional to q. Our proof has two main parts. First, we adapt and extend known results describing the joint distribution of the parity of the number of copies of small subgraphs on a random graph to the setting of graphs of growing size. Secondly, we analyse a variant of Karp's graph canonical labelling algorithm and exploit its massive parallelism to get a formula of low depth that defines an almost canonical pre-order on a random graph.

1 Introduction

Since the 0-1 law for first-order logic was established [5, 6], there has been much interest in exploring the asymptotic properties of definable classes of graphs. Many extensions of first-order logic have been shown to have a 0-1 law (see for instance [9, 4]) and in many other cases, weaker forms of convergence have been established (see [3]). A recent, remarkable result in this vein is that of Kolaitis and Kopparty [8] who study FO[$\oplus$], the extension of first-order logic with *parity quantifiers*. They show that for every constant edge-probability p and for every FO[$\oplus$] sentence ϕ, there are two explicitly computable rational numbers a_0, a_1 such that for $i \in \{0, 1\}$, as n approaches infinity, the probability that the random graph $G(2n + i; p)$ satisfies ϕ approaches a_i. In other words, ϕ has an asymptotic probability a_0 on the sequence of graphs of even cardinality and a_1 on those of odd cardinality. What is most interesting about this result is that it brings entirely new methods to the analysis of the asymptotic behaviour of logics on graphs, based on discrete analysis and polynomials over finite fields. In particular, it ties this to the study of approximations of circuits by low-degree polynomials, as we explain next.

The 0-1 law for first-order logic, in its general form is a quantifier-elimination result. It states that for any first-order formula ϕ, there is a quantifier-free formula θ such that ϕ is equivalent to θ almost surely on graphs drawn at random from $G(n, 1/2)$. To be precise, ϕ and θ agree on a fraction $1 - 2^{-\epsilon n}$ of the graphs on n vertices. We can say that any first-order formula is well approximated by a quantifier-free formula. This is similar to the phenomenon of depth-reduction for circuits which has a long history in

ISSN 1433-8092

computational complexity theory. For instance, Allender showed that AC^0-circuits have equivalent TC^0-circuits of depth 3 and quasi-polynomial size [1]. The result of Beigel and Tarui that general ACC^0-circuits have equivalent depth-2 circuits of quasi-polynomial size with a symmetric gate at the root [2] has been exploited to remarkable effect recently in the work of Williams [12]. In the context of approximation, one of the best known examples is the Razborov-Smolensky approximation of $AC^0[\oplus]$-circuits by multi-variate polynomials over $\mathbb{Z}_2$ of polylogarithmic degree [10, 11]. The quality of the approximation is still a matter of interesting study: the method yields an approximation that agrees on a fraction $1 - 2^{-(\log n)^c}$ of the inputs, and the question whether this can be improved to something of type $1 - 2^{-n^\epsilon}$ while keeping the degree polylogarithmic is open.

The result by Kolaitis and Kopparty mentioned above is proved by establishing a depth-reduction statement of a similar kind. They prove that every FO[$\oplus$] formula ϕ is well-approximated by a formula of a special form (which we call the KK-normal form in the sequel) which is a Boolean combination of quantifier-free formulas and polynomials of constant degree over $\mathbb{Z}_2$. The polynomials have as variables X_{uv} for every potential edge $\{u, v\}$ over the vertex-set $\{1, \ldots, n\}$. For example, the polynomial that gives the parity of the number of triples that extend the vertex u to a triangle is

$$\sum_{\substack{v \\ v \neq u}} \sum_{\substack{w \\ w \neq u \\ w \neq v}} X_{uv} X_{vw} X_{wu}. \tag{1}$$

At the heart of the construction is the analysis of the bias of certain low degree polynomials of this type on uniformly random inputs. This understanding is then used to carry over a quantifier-elimination argument that eliminates one parity quantifier or one first-order quantifier at a time. Relevant to our work is the fact that, intriguingly, the elimination of each parity quantifier in this argument incurs an exponential loss. The final outcome is that the degree d of the resulting polynomials is bounded from below by a function of tower-type on the number q of parity quantifiers that were eliminated, i.e.

$$d \geq 2^{2^{2^{\cdot^{\cdot^{\cdot}}}}} \tag{2}$$

where the height of the tower is at least q. At first look, the source of this inefficiency in the proof looks technical and it might be tempting to think that some different method could perhaps avoid it altogether.

In this paper we prove that the non-elementary dependence stated in equation (2) cannot be avoided. To be precise, we construct an explicit family of FO[$\oplus$] formulas ϕ_q of quantifier rank q and prove that they cannot be approximated by KK-normal forms whose polynomials have degree bounded by an elementary function of q. Specifically, we prove the following:

Theorem 1. *For every $\epsilon > 0$, there exists $n_0 \geq 0$ such that for every $n \geq n_0$, there exists a formula $\phi(u, v, w)$ such that, for every Boolean combination of FO[$\oplus$]-polynomials of degree bounded by a tower of exponentials of height at most $q/O(1)$, where q is the depth of ϕ, the formulas ϕ and p must disagree on a fraction $1 - \epsilon$ of all graphs with n vertices.*

By an FO[$\oplus$]-polynomial we mean a formula that has a direct translation as a bounded-degree polynomial over $\mathbb{Z}_2$: a sequence of parity quantifiers followed by a conjunction of atomic facts.

The result should be put in contrast with the 0-1 law for first-order logic mentioned above. In that case the approximating formula is quantifier-free, and it can be seen that quantifier free-formulas translate into polynomials of degree at most polynomial in the number of free variables of the formula.

Proof outline and techniques Our proof relies on two technical ingredients. On one hand we analyse a canonical labelling algorithm for graphs going back to Karp [7]. We exploit its massive parallelism to build an FO[$\oplus$]-formula $\psi(u, v)$ of depth $O(\log^* n)$ that works on graphs with n vertices. The formula is designed in such a way that, on almost every graph, it defines a linear pre-order of width at most two on the set of vertices of the graph. The second ingredient is a refined analysis of one of the key tools from the Kolaitis-Kopparty paper. Using and extending their techniques for estimating the frequencies mod 2 of subgraph copies, we show that every FO[$\oplus$]-polynomial $p(u, v, w)$ of degree $\log\log\log n$ must be unable to distinguish some triple of distinct vertices (a, b, c) from any of its permutations, with high probability.

From these two ingredients, the lower bound follows by taking the formula $\psi(u, v) \wedge \psi(v, w)$. On one hand this formula distinguishes at least one permutation of the vertices (a, b, c) from some other because by linearity of the pre-order the classes they lie in must be comparable, but by the width-2 condition on the pre-order not all three vertices can sit in the same class. On the other hand no formula of quantifier rank $\log\log\log n$ is able to distinguish any permutation of some triple (a, b, c) from the others. Since the quantifier rank of ϕ is still $O(\log^* n)$, the tower-type lower bound follows. We provide more details in the body of the paper.

2 Preliminaries

We use $[n]$ to denote the set $\{1, \ldots, n\}$. We write $A = B \pm C$ as convenient notation for $|A - B| \leq C$. We identify the nodes of a complete rooted binary tree with the binary strings that start with the symbol 1: the root is 1, the left child of t is $t0$ and the right child of t is $t1$. The *level-order* of a complete binary tree is $1, 10, 11, 100, 101, 110, 111, 1000, 1001, \ldots$, i.e. ordered first by length, and within each length, in lexicographical order. Note that if the strings are interpreted as numbers written in binary, this is the usual order of the natural numbers. For a natural number $n \geq 1$, we write $\mathrm{bin}_2(n)$ for its unique binary encoding with a leading one.

Let G and H be graphs. A homomorphism from G to H is a mapping $h : V(G) \to V(H)$ that maps edges to edges; i.e. such that if $\{u, v\} \in E(G)$, then $\{h(u), h(v)\} \in E(H)$. Let $\mathrm{Hom}(G, H)$ denote the collection of all homomorphisms from G to H.

The collection of FO[$\oplus$]-formulas over the language of graphs is the smallest class of formulas that contains the atomic formulas $E(x, y)$ and the equalities $x = y$, and is closed under negation, conjunction and disjunction, universal and existential quantification, and parity quantification; i.e. quantification of the form $\oplus x\ \phi(x)$. The meaning of $\oplus x\ \phi$ is that there is an odd number of vertices x that satisfy $\phi(x)$. For a tuple $\mathbf{a} = (a_1, \ldots, a_k)$ and a permutation $\pi \in S_k$, we write $\mathbf{a} \circ \pi$ for the tuple $(a_{\pi(1)}, \ldots, a_{\pi(k)})$. If $p(x_1, \ldots, x_k)$ is a formula with free variables $x_1, \ldots, x_k$, and $y_1, \ldots, y_k$ are variables or constants, we write $p(y_1, \ldots, y_k)$ for the result of replacing each occurrence of x_i by y_i. This applies also to the case where $y_1, \ldots, y_k$ is a permutation of $x_1, \ldots, x_k$.

An atomic type on the variables $x_1, \ldots, x_k$ over the language of graphs is a consistent collection of atomic formulas $E(x_i, x_j)$ or $x_i = x_j$ and negated atomic formulas $\neg E(x_i, x_j)$ or $x_i \neq x_j$ that is maximal with respect to set-inclusion. A positive atomic type is the subset of positive atomic formulas of an atomic type.

3 Fooling polynomials of low degree

In this section we show that every FO[$\oplus$]-formula of a certain general form corresponding to polynomials of low degree is not able to distinguish some triple from any of its permutations, with high probability. Roughly

speaking, the proof strategy is as follows. Fix such a formula $p(x, y, z)$. For every fixed $a, b, c \in [n]$, let $Y(a, b, c)$ be the event that p cannot distinguish any two permutations of a, b, c. Ideally we would like to show that the event $Y(a, b, c)$ has non-negligible probability of happening, and that if $a', b', c' \in [n]$ is a triple disjoint from a, b, c, then the events $Y(a, b, c)$ and $Y(a', b', c')$ are almost independent. If we were able to do this, the result would follow from an application of Chebyshev's inequality. Unfortunately it is not quite true that $Y(a, b, c)$ and $Y(a', b', c')$ are almost independent in general, so we take a detour.

3.1 Formulas and polynomials

In this section, we define the formulas to which our result applies. In short, they are Boolean combinations of FO[$\oplus$]-*polynomials*. An FO[$\oplus$]-*polynomial* is a formula of FO[$\oplus$] consisting of a sequence of parity quantifiers followed by a conjunction of atomic formulas. Thus, in its general form, an FO[$\oplus$]-polynomial p with free variables $u_1, \ldots, u_k$ is a formula of the form

$$\oplus u_{k+1} \cdots \oplus u_m \left(\bigwedge_{i \neq j} u_i \neq u_j \wedge \bigwedge_{\ell=1}^{d} E(u_{i_\ell}, u_{j_\ell}) \right), \tag{3}$$

where i and j range over $\{1, \ldots, m\}$ in the first conjunction, and $i_1, \ldots, i_d$, $j_1, \ldots, j_d$ are indices in $\{1, \ldots, m\}$. The number d of atomic facts in the conjunction is called the *degree* of p.

A conjunction of atomic formulas such as the one in the matrix of the formula (3) corresponds to the graph H on $\{u_1, \ldots, u_m\}$ that has an edge between u_{i_ℓ} and u_{j_ℓ} for each $\ell \in \{1, \ldots, d\}$. Thus, the formula expresses the parity of the number of extensions of $u_1, \ldots, u_k$ to a copy of H. We use the notation $\oplus H(u_1, \ldots, u_k)$ to denote this formula. Note that the degree of $\oplus H(u_1, \ldots, u_k)$ is the number of edges of H.

Example: If H is a triangle containing vertex u, then $\oplus H(u)$ is the formula that expresses the parity of the number of extensions of u to a triangle. Formally, $\oplus H(u)$ is the formula

$$\oplus v \oplus w \, (u \neq v \wedge u \neq w \wedge v \neq w \wedge E(u, v) \wedge E(v, w) \wedge E(w, u)).$$

Note that over undirected graphs, this formula is always false. This is because for every triangle containing u, there are two assignments to the variables v and w which witness H. Thus, the total number of satisfying assignments is twice the number of triangles containing u and is therefore always even. In general, if $H(u_1, \ldots, u_k)$ has an even number of automorphisms that fix $u_1, \ldots, u_k$, then $\oplus H(u_1, \ldots, u_k)$ will always be false, while for graphs with an odd number of automorphisms we get non-trivial formulas.

There is a precise sense in which FO[$\oplus$]-polynomials correspond to polynomials over the Boolean edge-variables X_{uv}. For example, the formula from the previous example corresponds to the family of degree-3 polynomials

$$\sum_{\substack{v \in [n] \\ v \neq u}} \sum_{\substack{w \in [n] \\ w \neq u \\ w \neq v}} X_{uv} X_{vw} X_{wu}.$$

as u ranges over $[n]$.

3.2 Independence and plan of action

The main obstacle to carrying out the argument sketched at the beginning of this section is that it is not true, in general, that the events $p(a, b, c)$ and $p(a', b', c')$ are almost independent, even if a, b, c, a', b', c' are all different. The reason is that the formula $p(x, y, z)$ may include statements about the graph G which do not involve the free variables. These are true or false independently of the choice of a, b, c or a', b', c' and thus create correlations between $p(a, b, c)$ and $p(a', b', c')$.

It is illustrative to give an example how this can happen.

Example: Let $p(x, y)$ be the formula that is the conjunction of the following : (1) $\oplus z\ E(x, z)$ (x has odd degree), (2) $\neg \oplus z\ E(y, z)$ (y has even degree); and (3) $\oplus H$ for some fixed non-trivial rigid graph H (x and y do not appear free in this). This is a Boolean combination of FO[$\oplus$]-polynomials of degree bounded by the number of edges of H.

Note that if $p(a, b)$ holds then $p(b, a)$ must fail. Therefore, the probability that $p(a, b) \not\leftrightarrow p(b, a)$ holds is approximately $2 \cdot \frac{1}{8}$ since each of the three condition in $p(a, b)$ holds with probability approximately $\frac{1}{2}$ almost independently, and similarly for $p(b, a)$. On the other hand, the probability that both $p(a, b) \not\leftrightarrow p(b, a)$ and $p(a', b') \not\leftrightarrow p(b', a')$ hold simultaneously is approximately $4 \cdot \frac{1}{32}$. This is because in each of the four cases in which both hold, condition (3) either holds for both a, b and a', b' or for neither (since x and y do not appear). We are left with five conditions that hold with probability approximately $\frac{1}{2}$ almost independently. Since $4 \cdot \frac{1}{32}$ is not ϵ-close to $(2 \cdot \frac{1}{8})^2$, this shows that $Y(a, b)$ and $Y(a', b')$ are not almost independent.

The example just sketched suggests that we *factor out* the condition that does not depend on neither x nor y from $p(x, y) \leftrightarrow p(y, x)$ since this is the cause for the statistical dependence between $Y(a, b)$ and $Y(a', b')$. However, while such an argument can be made to work in the example above, it is difficult to see how to factor something out when p may involve disjunctions.

The key observation at this point is that the *full type* of (x, y) in terms of its atomic type (the pattern of connections and equalities among x and y) and the truth values of its $\oplus H$'s as H ranges over all small graphs that contain x and y as vertices is enough to determine the truth value of $p(x, y)$. Thus, if we were able to find a full type implying $p(x, y)$ that is symmetric in x and y, we would have reduced the case of general $p(x, y)$ to the case of a $p(x, y)$ that consists of a single term. The argument that we use is a bit more delicate than this, but this is the main idea.

3.3 Normal forms

In this section we introduce some definitions and discuss two different types of *normal forms* for Boolean combinations of FO[$\oplus$]-polynomials.

An *I-labelled graph* is a graph with some vertices labelled by elements of I in such a way that, for every $i \in I$ there is exactly one vertex labelled i, and the set of labelled vertices induces an independent set. The set of labelled vertices of an I-labelled graph H is denoted by $\mathcal{L}(H)$. The vertex labelled by $i \in I$ is denoted by $F(i)$. An I-labelled graph H is *label-connected* if $H \setminus \mathcal{L}(H)$ is connected. Let Conn_I^t be the set of all I-labelled label-connected graphs with at most t unlabelled vertices. We say that H *depends on* label $i \in I$ if $H(i)$ is not an isolated node. We say that H is *label-dependent* if it depends on all its labels. Let $\mathrm{Conn}_I^{*,t}$ be the subset of all labelled graphs in Conn_I^t that are label-dependent.

A *k-labelled graph* is a $[k]$-labelled graph. A *$\leq k$-labelled graph* is an I-labelled graph for some $I \subseteq [k]$. Let H be a $\leq k$-labelled graph with labels $I \subseteq [k]$. A *homomorphism* from H to a pair $(G, \mathbf{a})$, where G is a graph and $\mathbf{a} = (a_1, \ldots, a_k)$ is a tuple in V_G^k, is a homomorphism $\chi \in \mathrm{Hom}(H, G)$ such that $\chi(H(i)) = a_i$ for each $i \in I$. A homomorphism χ from H to $(G, \mathbf{a})$ is *injective* if for any distinct $a, b \in V_H$ such

that $\{a,b\} \not\subseteq \mathcal{L}(H)$ we have $\chi(a) \neq \chi(b)$. Write $\oplus H(G,\mathbf{a})$ for the parity of the number of injective homomorphisms from H to $(G,\mathbf{a})$. We usually omit G and write $\oplus H(\mathbf{a})$. When H is a k-labelled graph (i.e. $I=[k]$), the notation for this in [8] is $[H]_2(G,\mathbf{a})$.

A *KK-normal form* of degree c with free-variables $x_1,\ldots,x_k$ is a Boolean combination of the atomic types on the variables $x_1,\ldots,x_k$ and formulas $\oplus H(x_1,\ldots,x_k)$'s as H ranges over the k-labelled label-connected graphs with labelled vertices $x_1,\ldots,x_k$ and at most $c-\binom{k}{2}$ non-labeled vertices. A *regular normal form* of degree c with free-variables $x_1,\ldots,x_k$ is a Boolean combination of the atomic types on the variables $x_1,\ldots,x_k$ and the $\oplus H(x_1,\ldots,x_k)$'s as H ranges over the $\leq k$-labelled label-connected, label-dependent graphs with labelled vertices within $x_1,\ldots,x_k$ and at most $c-\binom{k}{2}$ non-labelled vertices.

Example: Let $\phi(x,y)$ be the formula

$$\oplus z\ (E(x,z)) \wedge \neg \oplus z\ (E(y,z)),$$

saying that x has odd degree and y has even degree. This is a regular normal form. On the other hand, it is *not* a KK-normal form because the formula $\oplus z\ (E(x,z))$ cannot be put in the form $\oplus H(x,y)$ for any 2-labelled graph H. However, as we will see, it is not hard to transform $\phi(x,y)$ into an equivalent KK-normal form.

Example: Let $p(x,y)$ be the formula

$$(x\neq y \wedge E(x,y) \wedge \neg \oplus\ H_1(x,y) \wedge \oplus H_2(x,y)) \vee (x\neq y \wedge \neg E(x,y) \wedge \oplus H_1(x,y) \wedge \neg \oplus\ H_2(x,y)),$$

where H_1 is the 2-labelled label-connected graph that has three vertices x, y, and z and a single edge between x and z, and H_2 is the 2-labelled label-connected graph that has three vertices x, y, and z and a single edge between y and z. This is a KK-normal form. On the other hand, it is not a regular normal form because H_1 and H_2 are not label-dependent. However, as we will see, it is not hard to transform $\phi(x,y)$ into a regular normal form.

The two examples above are actually logically equivalent and it is a general fact that Boolean combinations of FO[$\oplus$]-polynomials, KK-normal forms, and regular normal forms of the same degree have the same expressive power.

Lemma 1. *Let $k \geq 0$ and $c \geq \binom{k}{2}$ be integers and let $\phi(x_1,\ldots,x_k)$ be a formula of* $\mathrm{FO}[\oplus]$ *that implies $x_i \neq x_j$ when $i \neq j$. The following are equivalent:*

1. *ϕ is equivalent to a Boolean combination of FO[$\oplus$]-polynomials of degree at most c,*
2. *ϕ is equivalent to a KK-normal form of degree at most c,*
3. *ϕ is equivalent to a regular normal form of degree at most c.*

Proof sketch. We only give a proof of (2) $\Rightarrow$ (3) as this is the only part of the above equivalence that we require in the sequel. We need to show how to transform a formula $\oplus H$ where H is a k-labelled label-connected graph H with at most c non-labelled vertices into an equivalent Boolean combination of formulas of the form $\oplus F$ where F is a $\leq k$-labelled label-connected, label-dependent graph with at most c non-labelled vertices. The transformation is done in two steps. In the first step we reduce the number of isolated labelled vertices in H or the number of non-labelled vertices of H at the expense of using possibly label-disconnected graphs. In the second step we get rid of the label-disconnected graphs. Indeed the second step is as in the proof of Lemma 5.6 in [8] so we need only take care of the first step.

If H is already label-dependent, there is nothing to do. Otherwise assume x_k is a labelled vertex that is isolated in H. Then, for every graph G and for every $a_1, \ldots, a_k \in V_G$ all different, $\oplus H(a_1, \ldots, a_k)$ is equivalent to:

$$\sum_{u \in V_H \setminus \mathcal{L}(H)} \tau^{u=x_k}(a_1, \ldots, a_k) \cdot \oplus H^{u=x_k}(a_1, \ldots, a_k) + \oplus H^{-x_k}(a_1, \ldots, a_{k-1}) \mod 2,$$

where:

- $\tau^{u=x_k}$ is the positive atomic type of u over $\mathcal{L}(H) \setminus \{x_k\}$ in H,
- $H^{u=x_k}$ is the k-labelled graph that results from deleting all edges from u to a labelled vertex and identifying u and x_k in H,
- H^{-x_k} is the $\leq k$-labelled graph that results from deleting x_k in H.

To see why this gives $\oplus H(a_1, \ldots, a_k)$ note that the first sum includes all homomorphisms from H that map x_i to a_i for $i \in [k]$ and are injective everywhere except at $\{u, x_k\}$ that are both mapped to a_k, and the last term includes all injective homomorphisms from H^{-x_k} that map x_i to a_i for $i \in [k-1]$. Therefore each injective homomorphism from H^{-x_k} that maps x_i to a_i for $i \in [k-1]$ and uses up a_k is counted exactly twice and cancels. □

3.4 Distribution of frequency vectors

The frequency vector of degree t in a graph G is the $\{0,1\}$-vector indexed by the set of all connected graphs with at most t vertices where the component H is $\oplus H(G)$, i.e. the parity of the number of occurrences of H in G. Kolaitis and Kopparty give an analysis of the distribution of frequency vectors in a random graph $G \sim G(n, 1/2)$, for constant t. Our aim in the present section is to extend this analysis to degrees that grow with n and to $\leq k$-labelled graphs.

Let $\mathrm{Conn}^t_{\leq k}$ be the set of all $\leq k$-labelled label-connected graphs with at most t unlabelled vertices. Let $\mathrm{Conn}^{*,t}_{\leq k}$ be the subset of $\mathrm{Conn}^t_{\leq k}$ containing all graphs that are label-dependent. Let G be a graph, let $\mathbf{a}$ be a tuple in V_G^k, and let $t \geq 0$ be an integer. Let $\mathrm{freq}^{*,t}_{\leq k,G}(\mathbf{a})$ be the $\{0,1\}$-vector indexed by the elements $\mathrm{Conn}^{*,t}_{\leq k}$ that has $\oplus H(\mathbf{a})$ as its component indexed by H. Next we extend the definition of feasible frequency vectors from the KK-article to the setting of $\leq k$-labelled graphs. In defining $\mathrm{FFreq}^*(\tau, \leq k, t)$ we will restrict our attention to atomic types τ that force all k variables distinct. This simplifies matters significantly and is enough for our purposes. If τ is an atomic type on $x_1, \ldots, x_k$ that forces $x_i \neq x_j$ for $i \neq j$, let $\mathrm{FFreq}^*(\tau, \leq k, t)$ denote the set of all *feasible frequency vectors*. Explicitly, these are all the $\{0,1\}$-vectors indexed by $\mathrm{Conn}^{*,t}_{\leq k}$ whose component F belongs to $\mathrm{aut}(F) \cdot \mathbb{Z}/2\mathbb{Z}$. Here $\mathrm{aut}(F)$ denotes the number of automorphisms of F that fix the labels. Let $\mathrm{FFreq}^*_n(\tau, \leq k, t)$ denote the set of $f \in \mathrm{FFreq}^*(\tau, \leq k, t)$ such that $f_{K_1(\emptyset)} = n \mod 2$.

The next lemma describes the distribution of $\mathrm{freq}^{*,t}_{\leq k,G}(\mathbf{a})$ in a random graph. This is analogous to Theorem 2.4 in [8] extended to growing degrees up to $\log\log\log n$, and extended from k-labelled graphs to $\leq k$-labelled graphs.

Lemma 2. *For every $k \geq 0$ there exists $n_0 \geq 0$ such that for every $n \geq n_0$, every atomic type τ on k variables that forces all of them distinct, every $c \leq \log\log\log n$, and every k-tuple $\mathbf{a}$ of distinct elements in $[n]$, the distribution of $\mathrm{freq}^{*,c}_{\leq k,G}(\mathbf{a})$ as $G = G(n, 1/2 \mid \tau(\mathbf{a}))$ is $2^{-\Omega_k(n/\log\log n)}$-close in statistical distance from the uniform distribution over $\mathrm{FFreq}^*_n(\tau, \leq k, c)$.*

Before we go on with the proof, it is worth pointing out the differences between the statement of Lemma 2 and the statement of Theorem 6.12 in the KK-article. Our statement here extends it in two directions, and both directions introduced corresponding points of tension in the statement.

The first difference concerns the extension going from bounded degree in the KK-article to growing degree up to $\log\log\log n$ here. This increased the bound on the statistical distance from $2^{-\Omega_k(n)}$ in the KK-article to $2^{-\Omega_k(n/\log\log n)}$ here. For the purposes of this paper, this difference is minor. This part of the proof consists in adapting the one in the KK-article to the setting of growing degrees. This is achieved by making the ϵ's explicit in the computation of certain Gowers' uniformity norms. It is conceivable that a more careful analysis could push the upper bound on the required degree up to $O(\sqrt{\log n})$ at the cost of increasing the statistical distance some more.

The second difference concerns the extension from k-labelled graphs in the KK-article to $\leq k$-labelled graphs here. This introduced the additional restriction of label-dependency (the * in $\mathrm{freq}^{*,t}$). This restriction is not minor since the result would not be true without it. Luckily, though, Lemma 1 tells us that we can assume label-dependency without loss of generality. In this case discovering the right assumption was an essential step in the proof. Once the concept is up, the proof is again following the original one. One final difference is that we restrict the statement to the case where the atomic type τ forces all variables distinct. This is the case we need anyway and simplifies matters a lot.

Proof sketch of Lemma 2. The issue of handling $\leq k$-labelled graphs instead of k-labelled graphs is not problematic until we realize that the sets of copies of F and F' in $(\mathrm{K}_n, \mathbf{a})$ need not be disjoint even if F and F' are non-isomorphic (i.e.: Proposition 8.1(2) from the KK-article fails in the case of $\leq k$-labelled graphs). This happens, for example, if F and F' are $\leq k$-labelled label-connected graphs that are identical except that F' has one more isolated labelled vertex than F. On the other hand, if F and F' are non-isomorphic and depend on all its labels, then it can be seen that the sets of copies are disjoint. This is enough to carry over the argument in the KK-article.

In order to allow a growing c, it suffices to prove the following lemma that makes the ϵ explicit in Lemma 4.7 from [8]:

Lemma 3. *Let $g : \mathbb{Z}_2^d \to \{-1, 1\}$ be given by $g(\mathbf{y}) = (-1)^{\prod_{i=1}^{d} \mathbf{y}_i}$. Let μ be the uniform distribution on $\mathbb{Z}_2^d$. Then $||g||_{U^d,\mu} < 1 - \epsilon$ where $\epsilon > 1 - \exp(-2^{-d^2-2d+1})$.*

We provide the proof in the appendix. With this bound in hand, the satistical distance becomes $(1-\epsilon)^r$ where $r = \Omega(n/c)$ and $d = O(c^2)$. Plugging in, the distance ends up bounded by

$$\left(\exp\left(-2^{-O(c^4)}\right)\right)^{\Omega(n/c)},$$

which is $2^{-\Omega(n/\log\log n)}$ for $c = \log\log\log n$. □

3.5 The argument itself

Finally we reached the point where we can execute the plan sketched at the beginnng of Section 3. Fix a positive integer k (for the application in Section 5 it suffices to take $k = 3$) and let $p(x_1, \ldots, x_k)$ be a regular normal form of degree $c \leq \log\log\log n$. For every $\mathbf{a} = (a_1, \ldots, a_k) \in [n]^k$, define the following indicator random variables:

$$X(\mathbf{a}) := \mathbb{I}[\ p(\mathbf{a}) \not\leftrightarrow p(\mathbf{a} \circ \pi) \text{ for some } \pi \in S_k\ ],$$
$$Y(\mathbf{a}) := \mathbb{I}[\ p(\mathbf{a}) \leftrightarrow p(\mathbf{a} \circ \pi) \text{ for every } \pi \in S_k\ ].$$

Obviously, $X(\mathbf{a}) = 1 - Y(\mathbf{a})$, and $Y(\mathbf{a})$ is the indicator random variable for the event that p does not distinguish any two permuted versions of $\mathbf{a}$. Our goal is to show that $Y(\mathbf{a})$ holds for some $\mathbf{a}$ with high probability and for this we will follow the plan sketched in section 3.2.

Write $p(\mathbf{x})$ as a DNF on the (Boolean) variables

$$\tau_1(\mathbf{x}), \ldots, \tau_r(\mathbf{x}), \oplus H_1(\mathbf{x}), \ldots, \oplus H_\ell(\mathbf{x}),$$

where $\tau_1, \ldots, \tau_r$ are the atomic types on $x_1, \ldots, x_k$, and $H_1, \ldots, H_\ell$ are the $\leq k$-labelled label-connected, label-dependent graphs with labelled vertices within $x_1, \ldots, x_k$ and c non-labelled vertices. Let us assume that $H_1, \ldots, H_e$ are the ones for which the number of automorphisms that fix its labelled vertices is odd, and that $H_{e+1}, \ldots, H_\ell$ are the rest. Also assume that $H_1, \ldots, H_f$ are the graphs from among $H_1, \ldots, H_e$ that do not have any label at all, and $H_{f+1}, \ldots, H_e$ are the rest. Since exactly one atomic type must hold and each $\oplus H_j(\mathbf{a})$ is false for $j \in \{e+1, \ldots, \ell\}$, we may assume that each term in the DNF formula has the form

$$\tau_i(\mathbf{x}) \cdot \prod_{j \in K} \oplus H_j \cdot \prod_{j \in K'} \overline{\oplus H_j} \cdot \prod_{j \in I} \oplus H_j(\mathbf{x}) \cdot \prod_{j \in I'} \overline{\oplus H_j(\mathbf{x})}. \tag{4}$$

for some $i \in [r]$, some partition (K, K') of $[f]$, and some partition (I, I') of $[e] \setminus [f]$.

Next note that for every $\pi \in S_k$, the sequence of Boolean variables $\tau_1(\mathbf{x} \circ \pi), \ldots, \tau_r(\mathbf{x} \circ \pi)$ is equivalent to a permutation of the sequence of Boolean variables $\tau_1(\mathbf{x}), \ldots, \tau_r(\mathbf{x})$. Similarly, the sequence of Boolean variables $\oplus H_{f+1}(\mathbf{x} \circ \pi), \ldots, \oplus H_e(\mathbf{x} \circ \pi)$ is equivalent to a permutation of the sequence $\oplus H_{f+1}(\mathbf{x}), \ldots, \oplus H_e(\mathbf{x})$. Therefore $p(\mathbf{x})$ and $p(\mathbf{x} \circ \pi)$ are functions of the same Boolean variables and we can write $p(\mathbf{x} \circ \pi)$ also as a DNF formula with terms of the type (4).

From now on, for every $K \subseteq [f]$, let R_K be the term

$$R_K := \prod_{j \in K} \oplus H_j \cdot \prod_{j \in K'} \overline{\oplus H_j},$$

where $K' = [f] \setminus K$. Recall that $H_1, \ldots, H_f$ are all label-free and therefore R_K does not depend on $\mathbf{x}$. Similarly, for every $I \subseteq [e] \setminus [f]$, let $S_I(\mathbf{x})$ be the term

$$S_I(\mathbf{x}) := \prod_{j \in I} \oplus H_j(\mathbf{x}) \cdot \prod_{j \in I'} \overline{\oplus H_j(\mathbf{x})},$$

where $I' = ([e] \setminus [f]) \setminus I$. For every $K \subseteq [f]$, let $p_K(\mathbf{x})$ denote the disjunction of the terms in $p(\mathbf{x})$ that are consistent with R_K. Therefore $p(\mathbf{x})$ is equivalent to the disjunction $\bigvee_{K \subseteq [f]} p_K(\mathbf{x})$.

Let $Z_K(\mathbf{x})$ be the "all-positive-term" defined as follows:

$$Z_K(\mathbf{x}) := \sigma(\mathbf{x}) \cdot R_K \cdot S_{[e] \setminus [f]}(\mathbf{x}),$$

where σ is the atomic type that forces $x_i \neq x_j$ for $i \neq j$, and all possible edges among different x_i, x_j. We show that for every $\mathbf{a} \in [n]^k$, the event $Z_K(\mathbf{a}) = 1$ implies $p(\mathbf{a}) \leftrightarrow p(\mathbf{a} \circ \pi)$ for every $\pi \in S_k$.

Lemma 4. $Z_K(\mathbf{a}) \leq Y(\mathbf{a})$.

Proof. Fix a permutation $\pi \in S_k$. First note that the choice of σ guarantees that $\sigma(\mathbf{a})$ is equivalent to $\sigma(\mathbf{a} \circ \pi)$. Also, the sequence $\oplus H_{f+1}(\mathbf{a} \circ \pi), \ldots, \oplus H_e(\mathbf{a} \circ \pi)$ is equivalent to a permutation of the sequence $\oplus H_{f+1}(\mathbf{a}), \ldots, \oplus H_e(\mathbf{a})$, and all appear positively in $Z_k(\mathbf{a})$. It follows from these term $Z_K(\mathbf{a})$ appears in both DNFs for $p(\mathbf{a})$ and $p(\mathbf{a} \circ \pi)$, or in neither. If it appears in both, then clearly $Z_K(\mathbf{a}) = 1$ implies both $p(\mathbf{a})$ and $p(\mathbf{a} \circ \pi)$. If it does not appear in either, then $Z_K(\mathbf{a}) = 1$ implies $\overline{p(\mathbf{a})}$ and $\overline{p(\mathbf{a} \circ \pi)}$ since $Z_K(\mathbf{a})$ is incompatible with any other term of the DNFs for $p(\mathbf{a})$ and $p(\mathbf{a} \circ \pi)$. In either case, $Z_K(\mathbf{a})$ implies $p(\mathbf{a}) \leftrightarrow p(\mathbf{a} \circ \pi)$. □

At this point it will suffice to show that for every $K \subseteq [f]$, the event $Z_K(\mathbf{a}) = 1$ holds for some $\mathbf{a} \in [n]^k$ with high probability in the probability space conditioned on R_K. From now on, for every event A, write

$$\mathbb{P}_K[\ A\ ] := \mathbb{P}[\ A \mid R_K\ ].$$

Let us start by computing the probability of $Z_K(\mathbf{a})$ for $\mathbf{a} \in [n]^k$ with $a_i \neq a_j$ for $i \neq j$ in this probability space. Let δ be the maximum, over all atomic types $\tau(\mathbf{x})$ that force $x_i \neq x_j$ for $i \neq j$, of the statistical distance between the distribution $\mathrm{freq}^{*,c}_{\leq k,G}(\mathbf{a})$ as $G = G(n, 1/2 \mid \tau(\mathbf{a}))$ and the uniform distribution over $\mathrm{FFreq}^*(\tau, \leq k, c)$. Note that, by symmetry, δ does not depend on $\mathbf{a}$ provided $a_i \neq a_j$ for $i \neq j$.

Lemma 5.

$$\mathbb{P}_K[\ Z_K(\mathbf{a})\ ] = \left(\frac{2^{-e}}{2^{-f}} \pm \delta \cdot \frac{1}{2^{-f} \pm \delta} \mp \delta \cdot \frac{2^{-e}}{2^{-f} \cdot (2^{-f} \pm \delta)}\right) \cdot 2^{-\binom{k}{2}}.$$

Proof. We have

$$\mathbb{P}_K[\ Z_K(\mathbf{a})\ ] = \frac{\mathbb{P}[\ Z_K(\mathbf{a}) \cdot R_K\ ]}{\mathbb{P}[\ R_K\ ]} = \frac{\mathbb{P}[\ S_{[e]\setminus[f]}(\mathbf{a}) \cdot R_K \mid\ \cdot \mathbb{P}[\ \sigma(\mathbf{a})\ ]}{\mathbb{P}[\ R_K\ ]}$$

The denominator is $2^{-f} \pm \delta$ by choice of δ. The numerator is $(2^{-e} \pm \delta) \cdot 2^{-\binom{k}{2}}$ also by choice of δ. The trailing $2^{-\binom{k}{2}}$ factor is $\mathbb{P}[\ \sigma(\mathbf{a})\ ]$. Now:

$$\frac{2^{-e} \pm \delta}{2^{-f} \pm \delta} - \frac{2^{-e}}{2^{-f}} = \frac{2^{-e} \cdot 2^{-f} \pm \delta \cdot 2^{-f} - 2^{-e} \cdot 2^{-f} \mp 2^{-e} \cdot \delta}{2^{-f} \cdot (2^{-f} \pm \delta)}$$

which simplifies to

$$\pm\delta \cdot \frac{1}{2^{-f} \pm \delta} \mp \delta \cdot \frac{2^{-e}}{2^{-f} \cdot (2^{-f} \pm \delta)}.$$

□

Next we compute, for every $\mathbf{a}, \mathbf{a}' \in [n]^k$ with all $a_1, \ldots, a_k, a'_1, \ldots, a'_k$ different, the probability of $Z_K(\mathbf{a}) \cdot Z_K(\mathbf{a}')$ in the probability space conditioned on R_K. Let γ be the maximum, over all atomic types $\tau(\mathbf{x}, \mathbf{x}')$ that force all $x_1, \ldots, x_k, x'_1, \ldots, x'_k$ different, of the statistical distance between the distribution $\mathrm{freq}^{*,c}_{\leq 2k,G}(\mathbf{a}, \mathbf{a}')$ as $G = G(n, 1/2 \mid \tau(\mathbf{a}, \mathbf{a}'))$ and the uniform distribution over $\mathrm{FFreq}^*(\tau, \leq 2k, c)$. Note that, by symmetry, γ does not depend on $\mathbf{a}, \mathbf{a}'$ provided they are all different.

Lemma 6.

$$\mathbb{P}_K[\ Z_K(\mathbf{a}) \cdot Z_K(\mathbf{a}')\ ] = \left(\frac{2^{-2e}}{2^{-2f}} \pm \gamma \cdot \frac{1}{2^{-f} \pm \gamma} \mp \gamma \cdot \frac{2^{-2e+f}}{2^{-f} \cdot (2^{-f} \pm \gamma)}\right) \cdot 2^{-2\binom{k}{2}}.$$

Proof. Let A denote the event that $\sigma(\mathbf{a})$ and $\sigma(\mathbf{a}')$ both hold. We have

$$\begin{aligned}\mathbb{P}_K[\ Z_K(\mathbf{a}) \cdot Z_K(\mathbf{a}')\ ] &= \frac{\mathbb{P}[\ Z_K(\mathbf{a}) \cdot Z_K(\mathbf{a}') \cdot R_K\ ]}{\mathbb{P}[\ R_K\ ]} \\ &= \frac{\mathbb{P}[\ S_{[e]\setminus[f]}(\mathbf{a}) \cdot S_{[e]\setminus[f]}(\mathbf{a}') \cdot R_K \mid A\ ] \cdot \mathbb{P}[\ A\ ]}{\mathbb{P}[\ R_K\ ]}.\end{aligned}$$

The denominator is $2^{-f} \pm \gamma$ by choice of γ. The numerator is $(2^{-2e+f} \pm \gamma) \cdot 2^{-2\binom{k}{2}}$ also by choice of γ. The trailing $2^{-2\binom{k}{2}}$ factor is $\mathbb{P}[\, A \,]$. Now:

$$\frac{2^{-2e+f} \pm \gamma}{2^{-f} \pm \gamma} - \frac{2^{-2e+f}}{2^{-f}} = \frac{2^{-2e+f} \cdot 2^{-f} \pm \gamma \cdot 2^{-f} - 2^{-2e+f} \cdot 2^{-f} \mp 2^{-2e+f} \cdot \gamma}{2^{-f} \cdot (2^{-f} \pm \gamma)}$$

which simplifies to

$$\pm\gamma \cdot \frac{1}{2^{-f} \pm \gamma} \mp \gamma \cdot \frac{2^{-2e+f}}{2^{-f} \cdot (2^{-f} \pm \gamma)}.$$

□

Let us note at this point that the number of $\leq k$-labelled graphs with at most c non-labelled vertices is bounded by $2^{(c+k)^2}$. Therefore, if ℓ is the number of $\leq k$-labelled label-connected graphs with at most c non-labelled vertices, then $\ell \leq \frac{1}{2}\log n$ for sufficiently large n, and in particular

$$2^\ell \leq \sqrt{n}. \tag{5}$$

We use this to prove the main consequence of this analysis up to now:

Lemma 7. *Let $\mathbf{a}, \mathbf{a}' \in [n]^k$ be such that $a_1, \ldots, a_k, a'_1, \ldots, a'_k$ are all different. The following hold:*

1. $\mathbb{P}_K[\, Z_K(\mathbf{a}) \,] \geq n^{-1/2} \cdot 2^{-\binom{k}{2}} - 2^{-\Omega_k(n/\log\log n)}$.
2. $|\mathbb{P}_K[\, Z_K(\mathbf{a}) \cdot Z_K(\mathbf{a}') \,] - \mathbb{P}_K[\, Z_K(\mathbf{a}) \,] \cdot \mathbb{P}_K[\, Z_K(\mathbf{a}') \,]| \leq 2^{-\Omega_k(n/\log\log n)}$.

Proof. By Lemma 2, both δ and γ are $2^{-\Omega(n/\log\log n)}$. On the other hand we have $2^{e-f} \leq 2^e \leq 2^\ell \leq n^{1/2}$ by (5) and also $2^f \leq 2^\ell \leq n^{1/2}$ by (5). Therefore $2^{f-e} \geq 2^{-e} \geq 2^{-\ell} \geq n^{-1/2}$ and $2^{-f} \geq 2^{-\ell} \geq n^{-1/2}$. Now *1* follows from plugging these bounds in Lemma 5 and *2* follows from plugging these bounds in both Lemma 5 and Lemma 6 and recalling that k is a constant. □

Now we conclude by proving the main result of this section:

Lemma 8. *For every $k > 0$ and $\epsilon > 0$, there exists $n_0 \geq 0$ such that for every $n \geq n_0$ and every KK-normal form $p(x_1, \ldots, x_k)$ of degree bounded by $\log\log\log n$, for $G \sim G(n, 1/2)$, the probability that there exists $\mathbf{a} \in [n]^k$ with $a_i \neq a_j$ for $i \neq j$ such that $p(\mathbf{a}) \leftrightarrow p(\mathbf{a} \circ \pi)$ for every $\pi \in S_k$ is at least $1 - \epsilon$.*

Proof. Fix k and ϵ and choose n_0 large. Let $m = \lfloor n/k \rfloor$. Divide $[n]$ into m disjoint k-tuples $(\mathbf{a}_1, \ldots, \mathbf{a}_m)$ arbitrarily but in such a way that $a_{\ell,i} \neq a_{\ell,j}$ for $i \neq j$. Define: $Y = \sum_{\ell \in [m]} Y(\mathbf{a}_\ell)$ and $Z_K = \sum_{\ell \in [m]} Z_K(\mathbf{a}_\ell)$, the second for every $K \subseteq [f]$. Note that by Lemma 4 we have $Z_K \leq Y$. We want to show that $\mathbb{P}_K[\, Z_K = 0 \,] \leq \epsilon$. This will be enough since then

$$\mathbb{P}[\, Y = 0 \,] = \sum_{K \subseteq [e]} \mathbb{P}[\, Y = 0 \mid R_K \,] \cdot \mathbb{P}[\, R_K \,] \leq \sum_{K \subseteq [e]} \mathbb{P}_K[\, Z_K = 0 \,] \cdot \mathbb{P}[\, R_K \,] \leq \epsilon.$$

To show that $\mathbb{P}_K[\, Z_K = 0 \,] \leq \epsilon$ we proceed by the second moment method. To simplify notation, let us fix $K \subseteq [f]$ and abbreviate Z_K by Z, and $Z_K(\mathbf{a}_\ell)$ by Z_ℓ. Similarly, all expectations $\mathbb{E}$, variances $\mathbb{V}$,

and probabilities $\mathbb{P}$ appearing below refer to the probability space $\mathbb{P}_K$. In computing the variance $\mathbb{V}[\, Z \,] = \mathbb{E}[\, Z^2 \,] - \mathbb{E}[\, Z \,]^2$ we have

$$\begin{aligned}
\mathbb{E}[\, Z^2 \,] - \mathbb{E}[\, Z \,]^2 &= \sum_{i=1}^{m}\sum_{j=1}^{m} \mathbb{E}[\, Z_i \cdot Z_j \,] - \sum_{i=1}^{m}\sum_{j=1}^{m} \mathbb{E}[\, Z_i \,] \cdot \mathbb{E}[\, Z_j \,] \\
&\leq \sum_{i=1}^{m} \mathbb{E}[\, Z_i^2 \,] + 2 \cdot \sum_{i \neq j} (\mathbb{E}[\, Z_i \cdot Z_j \,] - \mathbb{E}[\, Z_i \,] \cdot \mathbb{E}[\, Z_j \,]) \\
&\leq \sum_{i=1}^{m} \mathbb{E}[\, Z_i \,] + 2 \cdot \sum_{i \neq j} (\mathbb{P}[\, Z_i \cdot Z_j \,] - \mathbb{P}[\, Z_i \,] \cdot \mathbb{P}[\, Z_j \,]) \\
&= \mathbb{E}[\, Z \,] \pm 2 \cdot \binom{m}{2} \cdot 2^{-\Omega(n/\log\log n)} \\
&= \mathbb{E}[\, Z \,] \pm 2^{-\Omega_k(n/\log\log n)},
\end{aligned}$$

where the first inequality follows from considering the case $i = j$ in the first double sum and ignoring it in the second, the next inequality follows from the fact that Z_i is a 0-1-random variable, the equality after it follows from Lemma 7.2 (recall that $\mathbb{P}$ really stands for $\mathbb{P}_K$ here), and the last equality follows from $m = \lfloor n/k \rfloor$ and the fact that k is a constant.

Now by Lemma 7.1 we have

$$\mathbb{E}[\, Z \,] \geq m \cdot (n^{-1/2} \cdot 2^{-\binom{k}{2}} - 2^{-\Omega_k(n/\log\log n)}) = \Omega_k(n^{1/2}).$$

Applying it to Chebyshev's inequality we obtain

$$\mathbb{P}[\, Z = 0 \,] \leq \frac{\mathbb{V}[\, Z \,]}{\mathbb{E}[\, Z \,]^2} \leq \frac{\mathbb{E}[\, Z \,] \pm 2^{-\Omega_k(n/\log\log n)}}{\mathbb{E}[\, Z \,]^2} \leq \frac{1}{\mathbb{E}[\, Z \,]} + \frac{1}{\mathbb{E}[\, Z \,]^2} \leq \epsilon$$

for sufficiently large n. □

4 Defining a linear pre-order of width two

In this section we construct the formula of very low depth that defines a linear pre-order of width 2 with high probability. The proof strategy is to analyse a variant of an algorithm for graph canonization due to Karp [7], and to exploit its massive implicit parallelism to get formulas of very low depth.

4.1 Plan of action

Informally, the graph canonization algorithm works as follows. For a given graph G, split the vertices into two classes: those of even degree and those of odd degree. Inductively, we split the classes further by dividing the vertices according to the parity of the numbers of neighbours they have in each of the existing classes. We continue this process until no more classes are split.

We will need three facts about this process: (1) that for $G \sim G(n, 1/2)$ the process will reach a state where each class has at most two vertices with high probability, (2) that this will happen in fewer than n "generations" of the splitting process with high probability, and (3) that the process is massively parallel: all the classes created between the $\ell/2$-th generation and the ℓ-th generation are definable in terms of the classes created in the $(\log_2 \ell)$-th generation.

4.2 Splitting procedure

Let $G = (V, E)$ be an undirected graph. For a vertex x and a set B, we write $p(G, x, B)$ for the parity of the number of neighbours that x has in B. We extend this to sets:

$$p(G, A, B) = \sum_{x \in A} p(G, x, B) \mod 2.$$

A *splitting tree* for G is a rooted binary tree T with each node t carrying a label $L_t \subseteq V$ and a sign $M_t \in \{+, -\}$ denoting whether t is *marked* or *unmarked*, and satisfying the following properties:

1. the label of the root is V,
2. no two siblings are marked,
3. if t is an internal node, then [1] $L_{t0} \cup L_{t1} = L_t$ and $L_{t0} \cap L_{t1} = \emptyset$,
4. if s is a leaf, $x, y \in L_s$ and t is marked, then $p(G, x, L_t) = p(G, y, L_t)$.

Given a splitting tree T for G, let $R(T)$ denote the set of unmarked nodes that are either the root or are a left child. Let $R'(T)$ be the subset of nodes in $R(T)$ that are either the root or such that their label and the label of their sibling are both non-empty [2]. One step of the splitting procedure works as follows:

1. let t be the least node in $R(T)$ in level-order [3] and mark it,
2. for every leaf s, let $L_{sa} := \{x \in L_s : p(G, x, L_t) = a\}$ for both $a = 0$ and $a = 1$,
3. make [4] $s0$ and $s1$ the left and right children of s and leave them unmarked.

Let $\mathcal{P}(T)$ be the result of applying one step of the splitting procedure to T. If the node t that is chosen in the first step also belongs to $R'(T)$ we say that the step is *proper*, otherwise *improper*. When $R'(T)$ is empty we say that the procedure *stalls* at T. Note that when it stalls it will never make a proper step again. The procedure starts at the splitting tree T_0 that has only an unmarked root labelled by V.

4.3 Analysis of the splitting procedure

Let T_0 be the tree that has only an unmarked root labelled by V. For $k \geq 1$, let $T_k := \mathcal{P}(T_{k-1})$. Ideally we would like to show that after a modest number of steps, all leaves of the splitting tree are labelled by singletons or empty sets. Unfortunately the splitting procedure is not able to produce a tree with this property in general, not even with high probability on a random graph. The best we will be able to show is that for a randomly generated graph, with high probability all leaves will have at most two vertices.

Let us single out three key desirable properties of T_k, where the third is our goal:

(A_k): T_k has $L_t \neq \emptyset$ for every node t,

(B_k): T_k has been generated through proper steps only,

[1] Karp required also $L_{t0} \neq \emptyset$ and $L_{t1} \neq \emptyset$. For us it will be convenient to not require it and Karp's analysis will still go through with minor modifications that we will point out.

[2] Karp defined $R(T)$ as the set of unmarked nodes t that are either the root or that have a sibling t' such that $|L_{t'}| > |L_t|$, or $|L_{t'}| = |L_t|$ and are a left child. This difference is inessential to the analysis. The only important point is to unambiguously choose one of the two children when both are unmarked and non-empty.

[3] Karp used symmetric order. This difference is not essential for Karp's analysis but is important for us.

[4] Karp's version makes this step only if $L_{sa} \neq \emptyset$ for both $a = 0$ and $a = 1$; this note is related to footnote 1.

(C_k): T_k has $|L_t| \leq 2$ for every leaf t.

In the following we will show:

1. property (A_k) holds with high probability for modest values of k,
2. property (A_k) implies (B_{2^k}) for every graph and every $k \geq 0$,
3. conditioned on (B_{2^k}), property (C_{2^k}) holds with high probability for modest values of k.

Before we analyse the probability of (A_k) we need to introduce some terminology and a lemma from [7]. Let T be a splitting tree for some graph H on the vertices V. A node t of T is called *properly marked* if it is marked and it is either the root or a left child such that its label and the label of its sibling are both non-empty. [5] To every properly marked node t we associate a set $S_t \subseteq V$: the set of all x for which L_t is the unique minimal (with respect to set inclusion) properly marked node containing x. Let $\mathcal{S}(T)$ be the collection of all such sets. For every properly marked node t, let β_t be t together with the set of properly marked nodes $s \neq t$ such that L_s is a maximal subset of L_t. Note that $S_t = \triangle_{s \in \beta_t} L_s$, where $\triangle$ denotes symmetric difference. Define $\ell(x, S_t) := \sum_{s \in \beta_t} p(H, x, L_s) \mod 2$. We will say that another graph G on the vertices V is consistent with T if $p(G, x, L_t) = p(H, x, L_t)$ holds for every $x \in V$ and every properly marked node t.

We state a consequence of Lemmas 4 and 5 in [7] [6]:

Lemma 9. *Let T be the splitting tree of some graph on the vertices V and let H be chosen uniformly at random among the graphs on the vertices V that are consistent with T. If t is a node in $R'(T)$, then the distribution of $\{p(H, x, L_t)\}_{x \in V}$ is uniform over the assignments that satisfy the constraints*

$$p(H, S, L_t) = \ell(L_t, S) \quad \textit{for every } S \in \mathcal{S}(T_k) \setminus \{Y\},$$

where Y is the unique set in $\mathcal{S}(T)$ of which L_t is a proper subset.

In order to be able to make use of this lemma it is important to notice that if G denotes a random graph drawn from $G(n, 1/2)$ and $T_0, T_1, \ldots$ denotes the random sequence of splitting trees produced by this random graph, then the distribution of T_{k+1} conditioned on $T_0, \ldots, T_k$ is equally produced as follows: first choose a graph H uniformly at random among those consistent with T_k, and then run one step of the splitting procedure on T_k with respect to H. This follows from the fact that the marginal of a uniform distribution with respect to a subset of its support is uniformly distributed on that subset.

Now we can analyse the probability of (A_k):

Lemma 10. *Let $n \geq 1$ and $k \geq 1$ be integers such that $4k \leq \log_2 n$, and let $G \sim G(n, 1/2)$. Then, the probability that (A_k) fails is $2^{k+1} \cdot \exp(-n/2^{6k})$.*

Proof. In order to simplify notation, in this proof we let $n_t := |L_t|$. For a node t at depth $\ell \leq k$ in T_k, we say that t is *unbiased* if $|n_t - n \cdot 2^{-\ell}| \leq n \cdot 2^{-(2k-\ell+1)}$ holds, and *biased* otherwise. Note for later use that we allow the error-term $n \cdot 2^{-(2k-\ell+1)}$ to grow with ℓ, but that it always stays below $n \cdot 2^{-\ell}$ because $\ell \leq k$. Let us consider the event defined as follows:

(A'_k): T_k has every node unbiased.

[5] Our properly marked nodes correspond to the marked nodes in Karp's analysis.

[6] It would seem from Lemma 4 in [7] that we also need the constraint $p(G, L_t, L_t) = 0$. However, in our notation this constraint is implicit since $p(G, L_t, L_t)$ counts each edge within L_t exactly twice.

Note that since the error-term for $\ell = k$ is smaller than $n \cdot 2^{-k}$, property (A'_k) implies (A_k). Thus, it suffices to bound the probability that (A'_k) fails.

Since (A'_0) holds, if (A'_k) fails then there is a largest $\ell \in \{0, \ldots, k-1\}$ such that (A'_ℓ) is true and $(A'_{\ell+1})$ is false. Fix $\ell \in \{0, \ldots, k-1\}$, a leaf r of T_ℓ, and $a \in \{0, 1\}$, and we bound the probability that the child ra of r becomes biased in $T_{\ell+1}$ conditioned on T_ℓ satisfying (A'_ℓ). Let t be the node with respect to which the splitting step $\ell + 1$ is made. Since we are assuming that T_ℓ satisfies (A'_ℓ), each label is non-empty and therefore t belongs to $R'(T_\ell)$. Let Y be the unique set in $\mathcal{S}(T_\ell)$ of which L_t is a proper subset. By the discussion after Lemma 9, the tree $T_{\ell+1}$ can be seen as produced by first choosing H uniformly at random among the graphs that are consistent with T_ℓ, and then applying the splitting procedure on T_ℓ with respect to t and H. By Lemma 9, the distribution of $\{p(H, x, L_t)\}_{x \in V}$ is uniform over the assignments that satisfy the constraints

$$p(H, S, L_t) = \ell(L_t, S) \tag{6}$$

for every S in $\mathcal{S}(T_\ell) \setminus \{Y\}$. In particular, since all sets in $\mathcal{S}(T_\ell)$ are pairwise disjoint, if S is the unique minimal set in $\mathcal{S}(T_\ell)$ that contains L_r, then the distribution of $\{p(H, x, L_t)\}_{x \in S}$ is uniform over the assignments that satisfy the constraint (6) for this S only, or no constraint at all if $S = Y$.

Since r is unbiased, the set L_r is non-empty. Fix $x_0 \in L_r \subseteq S$. A different way of generating the distribution $\{p(H, x, L_t)\}_{x \in S}$ without sampling H is by first choosing values for $p(-, x, L_t)$ for $x \in S \setminus \{x_0\}$ uniformly and independently at random, and then setting the value for $p(-, x_0, L_t)$ to the unique value that satisfies the constraint $p(-, S, L_t) = \ell(L_t, S)$, or setting it uniformly and independently at random if $S = Y$. In either case, the number X of elements x in $L_r \setminus \{x_0\}$ for which $p(-, x, L_t) = a$ is a random variable distributed according to the binomial distribution $\mathrm{B}(m, \frac{1}{2})$ with $m = n_r - 1$. Note for later use that $|n_{ra} - X| \leq 1$ because only x_0 could be missed in the count. By Hoeffding's inequality for the binomial distribution, the probability that $\left|X - \frac{1}{2} \cdot m\right| \geq t$ is bounded by $2e^{-2t^2/m}$, which is bounded by

$$2e^{-2t^2/(2n \cdot 2^{-\ell})} \tag{7}$$

because r is unbiased and hence $m = n_r - 1 < n \cdot 2^{-\ell} + n \cdot 2^{-(2k-\ell+1)} \leq 2n \cdot 2^{-\ell}$ because $\ell < k$. Now, if ra were biased we would have

$$\left|n_{ra} - n \cdot 2^{-(\ell+1)}\right| > n \cdot 2^{-(2k-(\ell+1)+1)}.$$

Since $|n_{ra} - X| \leq 1$ and since $\left|\frac{1}{2} \cdot n_r - n \cdot 2^{-(\ell+1)}\right| \leq \frac{1}{2} \cdot n \cdot 2^{-(2k-\ell+1)}$ because r is unbiased, by the triangle inequality this would mean that

$$\left|X - \tfrac{1}{2} \cdot n_r\right| > n \cdot 2^{-(2k-(\ell+1)+1)} - \tfrac{1}{2} \cdot n \cdot 2^{-(2k-\ell+1)} - 1$$

and in particular, using $0 \leq \ell \leq k-1$ and $4k \leq \log_2 n$, that

$$\left|X - \tfrac{1}{2} \cdot (n_r - 1)\right| \geq n \cdot 2^{-3k}.$$

The probability of this happening is bounded by (7) with $t = n \cdot 2^{-3k}$, which is at most $\epsilon := 2\exp(-n \cdot 2^{-6k})$.

The argument is now finished by two union bounds. By the union bound over the 2^ℓ leaves of T_ℓ, the probability that some leaf of T_ℓ generates a biased child is at most $2^\ell \cdot \epsilon$. By the union bound over ℓ, the probability that there exists an $\ell \in \{0, \ldots, k-1\}$ for which (A'_ℓ) holds but $(A'_{\ell+1})$ fails is at most $\sum_{\ell=0}^{k-1} 2^\ell \cdot \epsilon$. Thus, the probability that (A'_k) fails is bounded $2^{k+1} \cdot \exp(-n \cdot 2^{-6k})$. □

Next we observe that (A_k) implies (B_{2^k}).

Lemma 11. *For any graph G and $k \geq 0$, if (A_k) holds, then (B_{2^k}) holds.*

Proof. In a complete binary tree of depth k, the number of *left* children at depth at most k is $\sum_{i=1}^{k} 2^{i-1} = 2^k - 1$. Now, if T_k satisfies (A_k), then the root and every left child at depth at most k has generated a proper step in the process of producing T_{2^k}; these are 2^k proper steps as claimed. □

Finally we note that 3-element sets split with high probability if enough steps are proper. This is similar to Lemma 7 in [7].

Lemma 12. *Let $G \sim G(n, 1/2)$ and let $k \geq 0$. Then, the probability that (B_k) holds and (C_k) fails is at most $\binom{n}{3} \cdot 2^{-2k}$.*

Proof. Fix a 3-element set $A \subseteq V$ and fix $\ell \leq k$. Let S_ℓ denote the event that the set A is not split at step ℓ and P_ℓ denote the event that step ℓ is proper. We aim to show that $\mathbb{P}[\ \bigcap_{\ell=1}^{k} S_\ell \cap \bigcap_{\ell=1}^{k} P_\ell\ ] \leq 2^{-2k}$ and the result then follows by a union bound over all three element subsets.

Now,

$$\mathbb{P}[\ \textstyle\bigcap_{\ell=1}^{k} S_\ell \cap \bigcap_{\ell=1}^{k} P_\ell\ ] = \prod_{\ell=0}^{k-1} \mathbb{P}[\ S_{\ell+1} \cap P_{\ell+1}\ \mid\ \bigcap_{i=1}^{\ell}(S_i \cap P_i)\ ]$$

which is bounded by

$$\prod_{\ell=0}^{k-1} \mathbb{P}[\ S_{\ell+1}\ \mid\ P_{\ell+1} \cap \textstyle\bigcap_{i=1}^{\ell}(S_i \cap P_i)\ ]. \tag{8}$$

So, it suffices to show that each term in (8) is bounded by $\frac{1}{4}$.

Fix $\ell \in \{0, \ldots, k-1\}$ and let $\mathbf{T}$ denote the sequence of splitting trees $T_0, \ldots, T_\ell$. Let $\mathcal{T}$ denote the set of all sequences of splitting trees of length $\ell+1$ and $\mathcal{T}_A$ denote the subset of $\mathcal{T}$ consisting of those sequences $\mathbf{U} = U_0, \ldots, U_\ell$ in which all steps are proper and A does not split at any stage and U_ℓ splits properly, i.e. with respect to a node in $R'(U_\ell)$. In other words, the sequence $\mathbf{U}$ satisfies $P_{\ell+1} \cap \bigcap_{i=1}^{\ell}(S_i \cap P_i)$. We now argue that, for any given $\mathbf{U} \in \mathcal{T}$, we have $\mathbb{P}[\ S_{\ell+1}\ \mid\ \mathbf{T} = \mathbf{U}\ ] \leq \frac{1}{4}$.

Let r be a leaf of T_ℓ such that $A \subseteq L_r$. Let t be the node of T_ℓ with respect to which the splitting step $\ell+1$ is made. We argue that, conditioned on the event that this step is proper, i.e. t belongs to $R'(T_\ell)$, the probability that the elements of A are not split apart in $T_{\ell+1}$ is at most $1/4$. Let Y be the unique set in $\mathcal{S}(T_\ell)$ of which L_t is a proper subset. By the discussion after Lemma 9, the tree $T_{\ell+1}$ can be seen as produced by first choosing H uniformly at random among the graphs that are consistent with T_ℓ, and then applying the splitting procedure on T_ℓ with respect to t and H. By Lemma 9, the distribution of $\{p(H, x, L_t)\}_{x \in V}$ is uniform over the assignments that satisfy the constraints

$$p(G, S, L_t) = \ell(L_t, S). \tag{9}$$

for every $S \in \mathcal{S}(T_\ell) \setminus \{Y\}$. In particular, since all sets in $\mathcal{S}(T_\ell)$ are pairwise disjoint, if S is the unique minimal set in $\mathcal{S}(T_\ell)$ that contains L_r, then the distribution of $\{p(H, x, L_t)\}_{x \in S}$ is uniform over the assignments that satisfy the constraint (9) for this S only, or no constraint at all if $S = Y$. Thus, in case $S \neq Y$ there are $2^{|S \cup L_t|-1}$ choices for $\{p(H, x, L_t)\}_{x \in S}$ and $2^{|S \cup L_t|-3}$ such choices that are constant over A, and in case $S = Y$ there are $2^{|S \cup L_t|}$ choices for $\{p(H, x, L_t)\}_{x \in S}$ and $2^{|S \cup L_t|-2}$ such choices that are constant over A. In both cases this gives probability $1/4$ as claimed.

To complete the argument, let E_ℓ denote the event $P_{\ell+1} \cap \bigcap_{i=1}^{\ell}(S_i \cap P_i)$. We have:

$$\begin{aligned}
\mathbb{P}[\, S_{\ell+1} \mid E_\ell \,] &= \mathbb{P}[\, S_{\ell+1} \cap E_\ell \,] \cdot \mathbb{P}[\, E_\ell \,]^{-1} \\
&= \textstyle\sum_{\mathbf{U}\in\mathcal{T}} \mathbb{P}[\, S_{\ell+1} \cap E_\ell \mid \mathbf{T} = \mathbf{U} \,] \cdot \mathbb{P}[\, \mathbf{T} = \mathbf{U} \,] \cdot \mathbb{P}[\, E_\ell \,]^{-1} \\
&= \textstyle\sum_{\mathbf{U}\in\mathcal{T}_A} \mathbb{P}[\, S_{\ell+1} \mid \mathbf{T} = \mathbf{U} \,] \cdot \mathbb{P}[\, \mathbf{T} = \mathbf{U} \,] \cdot \mathbb{P}[\, E_\ell \,]^{-1} \\
&\leq \tfrac{1}{4} \cdot \textstyle\sum_{\mathbf{U}\in\mathcal{T}_A} \mathbb{P}[\, \mathbf{T} = \mathbf{U} \,] \cdot \mathbb{P}[\, E_\ell \,]^{-1} \\
&= \tfrac{1}{4} \cdot \mathbb{P}[\, E_\ell \,] \cdot \mathbb{P}[\, E_\ell \,]^{-1} \\
&= \tfrac{1}{4}.
\end{aligned}$$

This completes the proof of the lemma. □

We are ready to synthesize what we learned in a single lemma. In its statement, the choice of parameters is made to minimize the probability of failure. Other choices with other goals would work as well.

Lemma 13. *Let $G \sim G(n, 1/2)$. Then, the probability that $T_{\lceil n^{1/7} \rceil}$ does not satisfy $(C_{\lceil n^{1/7} \rceil})$ is at most $2^{-\Omega(n^{1/8})}$.*

Proof. Choose $k = \lceil \frac{1}{7} \log_2 n \rceil$ in Lemma 10 and $k = \lceil n^{1/7} \rceil$ in Lemma 12 and link them through Lemma 11. □

4.4 Defining the splitting steps

In this section we show that sets L_t of the splitting trees T_k are definable by formulas $\psi_t(x)$ of very low quantifier rank. First let us recall that if the splitting step is made with respect to node t, then every leaf s splits into the sets

$$\begin{aligned}
L_{s0} &= \{x \in L_s : p(G, x, L_t) = 0\} \\
L_{s1} &= \{x \in L_s : p(G, x, L_t) = 1\}.
\end{aligned}$$

Note that the nodes at depth ℓ are generated by the ℓ-th splitting step. For every non-root node t in a splitting tree T, let $v_T(t)$ be the node of T that generated t. In the following let $u(1) := 1$ and $u(\ell) := \mathrm{bin}_2(2(\ell-1))$ for every $\ell \geq 2$.

Lemma 14. *Let G be a graph and let $k \geq \ell \geq 1$. Then, for every node t at depth ℓ in T_k, we have $v_{T_k}(t) = u(\ell)$.*

Proof. Let us write $T = T_k$. If t is one of the two nodes at depth 1, then $v_T(t)$ is the root, which agrees with $u(1)$. Assume now that t is a node at depth $\ell \geq 2$. Let num_2 be such that $\mathrm{num}_2(\mathrm{bin}_2(1x)) = 1x$ for every binary string x. We show that $\mathrm{num}_2(v_T(t)) = 2(\ell - 1)$. We proceed by induction on ℓ. For $\ell = 2$ we have it since then $v_T(t)$ is the left child of the root 10, and $\mathrm{num}_2(10) = 2$. Now, if t is a node at depth $\ell \geq 2$ and we assume that $\mathrm{num}_2(v_T(t)) = 2(\ell - 1)$, then for every $a \in \{0, 1\}$ we have

$$\mathrm{num}_2(v_T(ta)) = \mathrm{num}_2(v_T(t)) + 2 = 2(\ell - 1) + 2 = 2((\ell + 1) - 1)$$

where the first follows from the fact that the nodes at level $\ell+1$ are generated by the next left-child following $v_T(t)$ in the level-order, and that the level-order on nodes agrees with the order of the natural numbers when they are read in binary. □

Now, for $a_1, \ldots, a_\ell \in \{0,1\}$, define

$$\psi_{1a_1\cdots a_\ell}(x) := \bigwedge_{\substack{i=1\\a_i=1}}^{\ell} \oplus z\,(\psi_{u(i)}(z) \wedge E(x,z)) \wedge \bigwedge_{\substack{i=1\\a_i=0}}^{\ell} \neg \oplus z\,(\psi_{u(i)}(z) \wedge E(x,z)).$$

Note that $\psi_1(x)$ is true since then the conjunctions are empty. We show that the $\psi_t(x)$ are the formulas we are after.

Lemma 15. *Let G be a graph and let $k \geq \ell \geq 0$. Then, for every node t at depth at most ℓ in T_k, the formula $\psi_t(x)$ defines the set L_t in G.*

Proof. For every non-leaf node t at depth $\ell - 1$ we have $v_T(ta) = u(\ell)$ for both $a = 0$ and $a = 1$ by Lemma 14. Therefore

$$L_{t0} = \{x \in L_t : p(G, x, L_{u(\ell)}) = 0\}$$
$$L_{t1} = \{x \in L_t : p(G, x, L_{u(\ell)}) = 1\},$$

Now, if $t = 1a_1a_2\cdots a_\ell$, then unfolding the recursion we have that L_t is the set of vertices $x \in V$ for which $p(G, x, L_{u(i)}) = a_i$ holds for every $i \in \{1, \ldots, \ell\}$. This is precisely what $\psi_t(x)$ says. □

Note that the quantifier rank of $\psi_t(x)$ depends only on the depth of t. Therefore, let $q(\ell)$ be the quantifier rank of $\psi_t(x)$ for some and hence every t of depth ℓ. Note that $q(\ell)$ is monotone non-decreasing.

Lemma 16. $q(\ell) = O(\log^* \ell)$.

Proof. If t is a node is at depth ℓ, the largest $u(i)$ in the definition of ψ_t is $2(\ell - 1)$. Since q is monotone non-decreasing we have

$$q(\ell) = 1 + q(|\mathrm{bin}_2(2(\ell - 1))|).$$

Since the length of $\mathrm{bin}_2(2(\ell-1))$ is $\log_2(\ell) + O(1)$, this recurrence gives $q(\ell) = O(\log^* \ell)$ as claimed. □

4.5 Defining the linear pre-order

Finally we are ready to prove the main lemma of this section.

Lemma 17. *For every $\epsilon > 0$, there exists $n_0 \geq 0$ such that for all $n \geq n_0$ there is a formula $\psi(x, y)$ of quantifier rank $O(\log^* n)$ such that, for $G \sim G(n, 1/2)$, the probability that φ defines a linear pre-order of width at most 2 is at least $1 - \epsilon$.*

Proof. Fix $\epsilon > 0$ and let n_0 be large enough so that for every $n \geq n_0$ the probability in Lemma 13 is at most ϵ. For fixed $n \geq n_0$, let $k = \lceil n^{1/4} \rceil$, and let $\psi(x, y)$ be the formula:

$$\bigvee_{\substack{s,t\\s\leq t}} \psi_s(x) \wedge \psi_t(y),$$

where s and t range over the leaves of T_k in the disjunction. If T_k has all its leaves labelled by sets of size at most two this defines a linear pre-order of width at most two. By choice of n_0 this happens with probability at least $1 - \epsilon$. Finally, by Lemma 16, the quantifier rank of ψ is $O(\log^* k)$ which is $O(\log^* n)$ since $k \leq n$. □

5 Establishing the lower bound

Here we put it all together to prove Theorem 1.

Theorem 2 (Theorem 1 re-stated)**.** *For every $\delta > 0$, there exists $n_0 \geq 0$ such that for every $n \geq n_0$ there exists a formula $\phi(x, y, z)$ such that, for every Boolean combination of FO[$\oplus$]-polynomials $p(x, y, z)$ of degree bounded by a tower of exponentials of height at most $q/O(1)$, where q is the depth of ϕ, the formulas ϕ and p must disagree on a fraction $1 - \delta$ of all graphs with n vertices.*

Proof. Fix $\delta > 0$ and choose n_0 large enough and $n \geq n_0$. Let $\psi(x, y)$ be the formula from Lemma 17 for $\epsilon = \delta/2$. Let $\phi(x, y, z) := \psi(x, y) \wedge \psi(y, z)$. Let q be the quantifier rank of ϕ, which is the same as ψ. We have $q \leq c \log^* n$ for some constant $c > 0$. We claim that this $\phi(x, y, z)$ witnesses the theorem.

Suppose $p(x, y, z)$ is a Boolean combination of FO[$\oplus$]-polynomials of degree bounded by a tower of exponentials of height $q/c-3$ that agrees with $\phi(x, y, z)$ on more than a δ-fraction of graphs with n vertices. By Lemma 1 we may assume that $p(x, y, z)$ is a regular normal form of the same degree. By the -3 in the choice of the height $q/c-3$ of the tower of exponentials, the degree of $p(x, y, z)$ is bounded by $\log \log \log n$. If n_0 is large enough, with probability at least $1-\delta/2$ there exists a triple a, b, c of distinct vertices for which $Y(a, b, c)$ holds. Also if n_0 is large enough, with probability at least $1 - \delta/2$ the formula $\psi(x, y)$ defines a linear pre-order of width at most 2. By the union bound, with positive probability all three hold. Let G be such that:

1. $\phi(x, y, z)$ and $p(x, y, z)$ agree on G,
2. $\psi(x, y)$ defines a linear pre-order $\preceq$ of width at most 2 in G,
3. there exists a triple of distinct vertices a, b, c of G for which $Y(a, b, c)$ holds.

Now, assume without loss of generality that $a \preceq b \preceq c$: otherwise permute them accordingly. Note that we cannot have $c \preceq a$ as otherwise all three a, b, c would belong to the same class of the pre-order, which is not possible because its width is 2 and a, b, c are distinct. But then $\phi(a, b, c)$ holds and $\phi(c, a, b)$ does not hold, which means that ϕ distinguishes one permutation of (a, b, c) from another. But then p also does; a contradiction to $Y(a, b, c)$. □

Acknowledgment We are grateful to Swastik Kopparty for discussions and comments on a previous version of this paper.

References

[1] E. Allender. A Note on the Power of Threshold Circuits. In *Proceedings of the 30th Symposium on Foundations of Computer Science (FOCS)*, pages 580–584, 1989.

[2] R. Beigel and J. Tarui. On ACC. *Computational Complexity*, 4:350–366, 1994.

[3] K. J. Compton. 0-1 Laws in Logic and Combinatorics. In I. Rival, editor, *NATO Advanced Study Institute on Algorithms and Order*, pages 353–383. Kluwer, 1989.

[4] A. Dawar and E. Grädel. Properties of Almost All Graphs and Generalized Quantifiers. *Fundameta Informaticae*, 98(4):351–372, 2010.

[5] R. Fagin. Probabilities on Finite Models. *Journal of Symbolic Logic*, 41(1):50–58, 1976.

[6] Y. V. Glebskiĭ, D. I. Kogan, M. I. Ligon'kiĭ, and V. A. Talanov. Range and Degree of Realizability of Formulas in the Restricted Predicate Calculus. *Kibernetika,* 2:17–28, 1969.

[7] R. M. Karp. Probabilistic Analysis of a Canonical Numbering Algorithm for Graphs. In *Proceedings of the AMS Symposium in Pure Mathematics,* volume 34, pages 365–378, 1979.

[8] Ph. G. Kolaitis and S. Kopparty. Random Graphs and the Parity Quantifier. In *Proceedings of the 41st ACM Symposium on the Theory of Computing (STOC),* pages 705–714, 2009.

[9] Ph. G. Kolaitis and M. Y. Vardi. Infinitary Logics and 0-1 Laws. *Information and Computation,* 98(2):258–294, 1992.

[10] A. A. Razborov. Lower Bounds on the Size of Bounded Depth Networks over a Complete Basis with Logical Addition. *Mathematical Notes of the Academy of Sciences of the USSR,* 41:333–338, 1987.

[11] R. Smolensky. Algebraic Methods in the Theory of Lower Bounds for Boolean Circuit Complexity. In *Proceedings of the 19th ACM Symposium on the Theory of Computing (STOC),* pages 77–82, 1987.

[12] R. Williams. Non-uniform ACC Circuit Lower Bounds. In *Proceedings of the 26th IEEE Conference on Computational Complexity (CCC),* pages 115–125, 2011.

A Making ϵ explicit

First we show if μ is the uniform distribution over $\mathbb{Z}_2^d$ then the distribution $\mu^{(i)}$ with which the norm $||g||_{U^d,\mu}$ is defined is the uniform distribution over $(\mathbb{Z}_2^d)^{i+1}$.

Lemma 18. *Let μ be the uniform distribution on $\mathbb{Z}_2^d$. Then $\mu^{(i)}$ is the uniform distribution on $(\mathbb{Z}_2^d)^{i+1}$ for every $i \in \mathbb{N}$.*

Proof. By induction on i we show that $\mu^{(i)}(\mathbf{x}, \mathbf{t}_1, \ldots, \mathbf{t}_i) = 2^{-d(i+1)}$ for every $(\mathbf{x}, \mathbf{t}_1, \ldots, \mathbf{t}_i) \in (\mathbb{Z}_2^d)^{i+1}$. Since $\mu^{(0)} = \mu$, the claim is clear for $i = 0$. For $i > 0$ we have

$$\mu^{(i)}(\mathbf{x}, \mathbf{t}_1, \ldots, \mathbf{t}_i) = \frac{\mu^{(i-1)}(\mathbf{x}, \mathbf{t}_1, \ldots, \mathbf{t}_{i-1}) \cdot \mu^{(i-1)}(\mathbf{x} + \mathbf{t}_i, \mathbf{t}_1, \ldots, \mathbf{t}_{i-1})}{\sum_{\mathbf{z}} \mu^{(i-1)}(\mathbf{z}, \mathbf{t}_1, \ldots, \mathbf{t}_{i-1})}.$$

Applying the induction hypothesis this is

$$\frac{2^{-di} \cdot 2^{-di}}{2^d \cdot 2^{-di}} = 2^{-d(i+1)}.$$

□

Lemma 19. *Let $g : \mathbb{Z}_2^d \to \{-1, 1\}$ be given by $g(\mathbf{y}) = (-1)^{\prod_{i=1}^d \mathbf{y}_i}$. Let μ be the uniform distribution on $\mathbb{Z}_2^d$. Then $||g||_{U^d,\mu} < e^{-2^{-d^2-2d+1}}$.*

Proof. Let $p_0 = \mu^{(d)}(\mathbf{0}, \mathbf{t}_0)$ where $\mathbf{t}_0 = (\mathbf{0}, \ldots, \mathbf{0})$ and let $p_1 = \mu^{(d)}(\mathbf{0}, \mathbf{t}_1)$ where $\mathbf{t}_1 = (\mathbf{e}_1, \ldots, \mathbf{e}_d)$ and $\mathbf{e}_j$ is the j-th standard unit vector in $\mathbb{Z}_2^d$. We have

$$(\mathrm{D}_{\mathbf{t}_0} g)(\mathbf{0}) = \prod_{S \subseteq [d]} (-1)^{\prod_{i=1}^d (\mathbf{0}_i + \sum_{j \in S} \mathbf{0}_i)} = \prod_{S \subseteq [d]} (-1)^0 = 1.$$

Also

$$(\mathrm{D}_{\mathbf{t}_1} g)(\mathbf{0}) = \prod_{S \subseteq [d]} (-1)^{\prod_{i=1}^d (\mathbf{0}_i + \sum_{j \in S} \mathbf{e}_{j,i})} = \prod_{\substack{S \subseteq [d] \\ S \neq [d]}} (-1)^{\prod_{i=1}^d (\mathbf{0}_i + \sum_{j \in S} \mathbf{e}_{j,i})} \cdot (-1)^{\prod_{i=1}^d (\mathbf{0}_i + \sum_{j \in [d]} \mathbf{e}_{j,i})}.$$

When S is a proper subset of $[d]$, the factor $\mathbf{0}_i + \sum_{j \in S} \mathbf{e}_{j,i}$ in the exponent vanishes at each $i \in [d] \setminus S$. On the other hand, for $S = [d]$, the factor $\mathbf{0}_i + \sum_{j \in S} \mathbf{e}_{j,i}$ is 1 at each $i \in [d]$. Therefore

$$(\mathrm{D}_{\mathbf{t}_1} g)(\mathbf{0}) = \prod_{\substack{S \subseteq [d] \\ S \neq [d]}} (-1)^0 \cdot (-1)^1 = -1.$$

Thus

$$\begin{aligned}
||g||_{U^d,\mu}^{2^d} = \left| \mathbb{E}_{(x,\mathbf{t}) \sim \mu^{(d)}} [(\mathrm{D}_{\mathbf{t}} g)(x)] \right| &= \left| \sum_{(x,\mathbf{t})} \mu^{(d)}(x, \mathbf{t}) (\mathrm{D}_{\mathbf{t}} g)(x) \right| \\
&\leq |p_0 \cdot (\mathrm{D}_{\mathbf{t}_0} g)(\mathbf{0}) + p_1 \cdot (\mathrm{D}_{\mathbf{t}_1} g)(\mathbf{0})| + \sum_{\substack{(x,\mathbf{t}) \\ (x,\mathbf{t}) \neq (\mathbf{0},\mathbf{t}_0) \\ (x,\mathbf{t}) \neq (\mathbf{0},\mathbf{t}_1)}} \mu^{(d)}(x, \mathbf{t}) \cdot |(\mathrm{D}_{\mathbf{t}} g)(x)| \\
&\leq |p_0 - p_1| + 1 - p_0 - p_1 \\
&= 1 - 2^{-d(d+1)+1},
\end{aligned}$$

where the first inequality follows from the triangle inequality, the second inequality follows from the computations above and the fact that each $(\mathrm{D}_{\mathbf{t}} g)(x)$ has magnitude 1, and the last equality follows from $p_0 = p_1 = 2^{-d(d+1)}$ by the previous lemma. We conclude that

$$||g||_{U^d,\mu} \leq (1 - 2^{-d(d+1)+1})^{1/2^d} \leq e^{-2^{-d(d+1)+1} \cdot 2^{-d}} = e^{-2^{-d^2-2d+1}}.$$

□

ECCC ISSN 1433-8092

http://eccc.hpi-web.de