Graph Colouring Is Hard on Average for Polynomial Calculus and Nullstellensatz

Jonas Conneryd
Lund University
University of Copenhagen

Susanna F. de Rezende
Lund University

Jakob Nordström
University of Copenhagen
Lund University

Shuo Pang
University of Copenhagen
Lund University

Kilian Risse
EPFL

Abstract

We prove that polynomial calculus (and hence also Nullstellensatz) requires linear degree to refute that sparse random regular graphs as well as sparse Erdős-Rényi random graphs are 3-colourable. Using the known relation between size and degree for polynomial calculus proofs, this implies strongly exponential lower bounds on proof size.

1 Introduction

Determining the *chromatic number* of a graph G, i.e., how many colours are needed for the vertices of G if no two vertices connected by an edge should have the same colour, is one of the classic 21 problems shown NP-complete in the seminal work of Karp [Kar72]. This *graph colouring problem*, as it is also referred to, has been extensively studied since then, but there are still major gaps in our understanding.

The best known approximation algorithm computes a graph colouring within at most a factor $O(n(\log\log n)^2/(\log n)^3)$ of the chromatic number [Hal93], and it is known that approximating the chromatic number to within a factor $n^{1-\varepsilon}$ is NP-hard [Zuc07]. Even under the promise that the graph is 3-colourable, the most parsimonious algorithm with guaranteed polynomial running time needs $O(n^{0.19996})$ colours [KT17]. This is very far from the lower bounds that are known—it is NP-hard to $(2k-1)$-colour a k-colourable graph [BBKO21], but the question of whether colouring a 3-colourable graph with 6 colours in NP-hard remains open [KO22]. It is widely believed that any algorithm for graph colouring has to run in exponential time in the worst case, and the currently fastest algorithm for 3-colouring has time complexity $O(1.3289^n)$ [BE05]. A survey on various algorithms and techniques for so-called exact algorithms is [Hus15].

Graph colouring instances of practical interest might not exhibit such exponential-time behaviour, however, and in such a context it is relevant to study algorithms without worst-case guarantees and examine how they perform in practice. To understand such algorithms from a computational complexity viewpoint, it is natural to investigate bounded models of computation that are strong enough to describe the reasoning performed by the algorithms and prove unconditional lower bounds for these models.

1.1 Previous Work

Focusing on random graphs, McDiarmid [McD84] developed a method for determining k-colourability that captures a range of algorithmic approaches. Beame et al. [BCMM05] showed that this method could in turn be simulated by the resolution proof system, and established average-case exponential lower bounds for resolution proofs of non-colourability for random graph instances sampled so as not to be k-colourable with exceedingly high probability.

Different algebraic approaches for k-colourability have been considered in [AT92, Lov94, Mat74, Mat04]. Bayer [Bay82] seems to have been the first to use Hilbert's Nullstellensatz to attack graph colouring. Informally, the idea is to write the problem as a set of polynomial equations $\{p_i(x_1,\dots,x_n)=0 \mid i\in[m]\}$ in such a way that legal colourings correspond to common roots for these polynomials, and then show that there are other polynomials $q_1,\dots,q_m$ such that $\sum_{i=1}^{m} q_i p_i = 1$. This latter equality is referred to as a *Nullstellensatz certificate* of non-colourability, and the *degree* of this certificate is the largest degree of any polynomial $q_i p_i$ in the sum. Later papers based on Nullstellensatz and Gröbner bases such as [DL95, Mnu01, HW08] culminated in an award-winning sequence of works [DLMM08, DLMO09, DLMM11, DMP+15] with surprisingly good performance.

For quite some time, no strong lower bounds were known for these algebraic methods or the corresponding proof systems *Nullstellensatz* [BIK+94] and *polynomial calculus* [CEI96, ABRW02]. On the contrary, the authors of [DLMO09] reported that essentially all benchmarks they studied turned out to have Nullstellensatz certificates of small constant degree. The degree lower bounds $k+1$ for k colours in [DMP+15] remained the best known until optimal, linear, degree lower bounds for polynomial calculus were established in [LN17] using a reduction from lower bounds for so-called functional pigeonhole principle formulas [MN15]. A more general reduction framework in [AO19] yielded optimal degree lower bounds also for *Sherali-Adams* [SA90] and *sums-of-squares* [Las01, Par00],

as well as weakly exponential size lower bounds for *Frege proofs* [CR79, Rec75] of bounded depth.

The lower bounds discussed in the last paragraph are not quite satisfactory, however, in that it is not clear how much they actually tell us about the graph colouring problem, as opposed to the hardness of the problems being reduced from. It seems both natural and desirable to establish optimal average-case lower bounds for random graphs, just as for resolution in [BCMM05], but this goal has remained elusive for almost two decades, as pointed out, e.g., in [LN17, Lau18, BN21]. The strongest result in this direction seems to be the recent logarithmic degree lower bound in sums-of-squares for Erdős-Rényi graphs with edge probability $1/2$ and $k = n^{1/2+\varepsilon}$ colours [KM21]. Since this result is for a problem encoding using inequalities, it is not clear whether this has any implications for Nullstellensatz or polynomial calculus over the reals, and for other fields nothing has been known for the latter two proof systems—not even logarithmic lower bounds.

1.2 Our Contribution

In this work, we establish optimal linear degree lower bounds and exponential size lower bounds for polynomial calculus proofs of non-colourability of random graphs.

Theorem 1.1 (informal). *For any $d \geq 6$, polynomial calculus (and hence also Nullstellensatz) requires linear degree to refute that random d-regular graphs $\mathbb{G}_{n,d}$, as well as Erdős-Rényi random graphs $\mathbb{G}(n, d/n)$, are 3-colourable. These degree lower bounds hold over any field, and also imply exponential lower bounds on proof size.*

We prove our lower bound for the standard encoding in proof complexity, where variables $x_{v,i}$ indicate whether vertex v is coloured with colour i or not. It should be pointed out, however, that just as the results in [LN17], our degree lower bounds apply to the k-colourability encoding introduced in [Bay82] and used in computational algebra papers such as [DLMM08, DLMO09, DLMM11, DMP+15], where a primitive kth root of unity is adjoined to the field and different colours of a vertex v are encoded by a variable x_v taking different powers of this root of unity.

Our lower bound proofs crucially use a new idea for proving degree lower bounds for colouring graphs with large girth [RT22]. After translating this proof from the root-of-unity encoding to the Boolean indicator variable encoding, and replacing the proof in terms of girth with a strengthened argument using carefully chosen properties of random graphs, we obtain a surprisingly clean and simple solution to the long-standing open problem of showing average-case polynomial calculus degree lower bounds for graph colouring.

1.3 Discussion of Proof Techniques

In most works on algebraic and semialgebraic proof systems such as Nullstellensatz, polynomial calculus, Sherali-Adams and sums-of-squares, the focus has been on proving upper and lower bounds on the degree of proofs. Even when proof size is the measure of interest, almost all size lower bounds have been established via degree lower bounds combined with general results saying that for all of the above proof systems except Nullstellensatz strong enough lower bounds on degree imply lower bounds on size [IPS99, AH19].

At a high level the techniques for proving degree lower bounds for the different proof systems have a similar flavour. For the static proof systems, i.e., Nullstellensatz, Sherali-Adams and sums-of-squares, it is enough to show that the dual is feasible and thus rule out low-degree proofs. In more detail, for Nullstellensatz, one constructs a *design* [Bus98], which is a linear functional mapping low degree monomials to the underlying field, satisfying that it maps 1 to a non-zero element and low degree monomials multiplied by any polynomial p_i in the problem encoding to 0. If such a

functional can be found, it is clear that there cannot exist any low-degree Nullstellensatz certificate $\sum_{i=1}^{m} q_i p_i = 1$ of unsatisfiability, as the design would map the left hand side of the equation to 0 but the right hand side to a non-zero element of the field. For sums-of-squares, the analogous functional furthermore has to map squares of low-degree polynomials to non-negative numbers. Such a *pseudo-expectation* can be viewed as a fake random distribution over satisfying assignments to the problem, which is indistinguishable from a true distribution for an adversary using only low-degree polynomials.

Polynomial calculus is different from these proof systems in that it does not present the certificate of unsatisfiability as a static object, but instead, given a set of polynomials $\mathcal{P}$, dynamically derives new polynomials in the ideal generated by $\mathcal{P}$. The derivation ends when it reaches the polynomial 1, i.e., the multiplicative identity in the field, showing that there is no solution. Since all polynomials derived lie in the ideal of $\mathcal{P}$, reducing them modulo this ideal always yields the polynomial 0. To prove degree lower bounds one designs a *pseudo-reduction operator* or *R-operator* [Raz98], which maps all low-degree polynomials derived from $\mathcal{P}$ to 0 but sends 1 to 1, and which is indistinguishable from a true ideal reduction operator if one is limited to reasoning with low-degree polynomials. This means that for a bounded-degree adversary it seems like the set of input polynomials are consistent.

Following the method in [AR03], a pseudo-reduction operator R can be constructed by defining it on low-degree monomials and extending to low-degree polynomials by linearity. For every monomial m, we identify a set of related input polynomials $S(m)$, let $\langle S(m)\rangle$ be the ideal generated by these polynomials, and define $R(m) = R_{\langle S(m)\rangle}(m)$ to be the reduction of m modulo the ideal $\langle S(m)\rangle$. Intuitively, we think of $S(m)$ as the (satisfiable) subset of polynomials that might possibly have been used in a low-degree derivation of m, but since the constant monomial 1 is not derivable in low degree it gets an empty associated set of polynomials, meaning that $R(1) = R_{\langle S(1)\rangle}(1) = 1$. In order for R to look like a real reduction operator, we need to show that for polynomials p and p' of not too high degree it holds that $R(p + p') = R(p) + R(p')$ and $R(p \cdot R(p')) = R(p \cdot p')$. The first equality is immediate since R is defined to be a linear operator, but the second equality is not so clear. Since the polynomials p and p' will be reduced modulo different ideals—in fact, this will be the case even for different monomials within the same polynomial—a priori there is no reason why R should behave nicely with respect to multiplication.

Proving that an R-operator behaves like a true reduction operator for low-degree polynomials is typically the most challenging technical step in the lower bound proof. Very roughly, the proof method in [AR03] goes as follows. Suppose that m and m' are monomials with associated polynomial sets $S(m)$ and $S(m')$, respectively. Using expansion properties of the constraint-variable incidence graph for the input polynomials, we argue that the true reduction operator will not change if we reduce both monomials modulo the larger ideal $\langle S(m) \cup S(m')\rangle$ generated by the union of their associated sets of polynomials. This implies that we have $R(m') = R_{\langle S(m')\rangle}(m') = R_{\langle S(m)\cup S(m')\rangle}(m')$ and $R(m \cdot m') = R_{\langle S(m)\cup S(m')\rangle}(m \cdot m')$ and if so it is clear that $R(m \cdot R(m')) = R(m \cdot m')$ holds, just like for reduction modulo an actual ideal. To prove this is a delicate balancing act, though, since the ideals will need to be large enough to guarantee non-trivial reduction, but at the same time small enough so that different ideals can be "patched together" with only local adjustments.

All previous attempts to apply this lower bound strategy to the graph colouring problem have failed. For other polynomial calculus lower bounds it has been possible to limit the interaction between different polynomials in the input. For graph colouring, however, applying the reduction operator intuitively corresponds to partial colourings of subsets of vertices, and it has not been known how to avoid that locally assigned colours propagate new colouring constraints through the rest of the graph. In technical language, what is needed is a way to order the vertices in the graph so that there will be no long ordered paths of vertices along which colouring constraints can spread.

It has seemed far from obvious how to construct such an ordering, or even whether it should exist, and due to this technical problem it has not been possible to join local ideal reduction operators into a globally consistent R-operator.

This technical problem was addressed in a recent paper [RT22] by an ingenious, and in hindsight very simple, idea. The main insight is to consider a proper colouring of the graph with χ colours, and then order the vertices in each colour class consecutively. In this way, order-decreasing paths are of length at most χ and one can guarantee some form of locality. Once this order is in place, the final challenge is to ensure that small cycles do not interfere when "patching together" reductions. In [RT22] this was avoided precisely by ensuring that the graph should have high girth, which resulted in a degree lower bound linear in the girth of the graph. In terms of graph size, this cannot give better than logarithmic lower bounds, however, since the girth is at most logarithmic in the number of vertices for any graph of chromatic number larger than 3 [Bol78].

In our work, we use the same ordering as in [RT22], but instead of girth use the fact that random graphs are locally very sparse. Once the necessary technical concepts are in place, the proof becomes quite simple and elegant, which we view as an extra strength of our result.

1.4 Outline of This Paper

The rest of this paper is organized as follows. In Section 2 we present some preliminaries. In Section 3 we introduce our techniques and provide a proof overview. In Section 4 we prove a linear polynomial calculus degree lower bound for 4-colourability on random regular graphs, which serves as a blueprint for our stronger results. In Section 5 and Section 6 we improve the lower bound to hold for 3-colourability on random regular graphs and the Erdős-Rényi graph, respectively. We conclude with some final remarks and open problems in Section 7.

2 Preliminaries

We briefly review the necessary preliminaries from proof complexity, algebra and graph theory. We use standard asymptotic notation, and all logarithms in this paper have base 2.

2.1 Proof Complexity

Polynomial calculus (PC) [CEI96] is a proof system that uses algebraic reasoning to deduce that a set $\mathcal{P}$ of polynomials over a field $\mathbb{F}$ involving the variables $x_1, \ldots, x_n$ is infeasible, i.e., that the polynomials in $\mathcal{P}$ have no common root. To prove that $\mathcal{P}$ is infeasible, polynomial calculus interprets $\mathcal{P}$ as a set of generators of an ideal and then derives new polynomials in this ideal through two derivation rules:

$$\textit{Linear combination: } \frac{p \quad q}{ap + bq}, \ a, b \in \mathbb{F} \tag{2.1a}$$

$$\textit{Multiplication: } \frac{p}{xp}, \ x \text{ any variable} \tag{2.1b}$$

A *polynomial calculus derivation* π of a polynomial p from $\mathcal{P}$ is a sequence $(p_1, \ldots, p_\tau)$ such that $p_\tau = p$ and each polynomial p_i is either in $\mathcal{P}$ or obtained by applying one of the derivation rules (2.1a)-(2.1b) to polynomials in $\mathcal{P}$ or polynomials p_j with $j < i$. A *polynomial calculus refutation of* $\mathcal{P}$ is a derivation of the constant polynomial 1 from $\mathcal{P}$, which is then a proof that $\mathcal{P}$ is infeasible. Often we are interested in systems of polynomial equations with Boolean variables, in which case

we also add the Boolean axioms $\{x_1^2 - x_1, \ldots, x_n^2 - x_n\}$ to $\mathcal{P}$. It is a standard fact that polynomial calculus is sound and complete when the Boolean axioms are present.

The most common complexity measures associated to a polynomial calculus refutation are its *size* and its *degree*. Let p be a polynomial expanded into a linear combination of distinct monomials. The *size* of p is the number of monomials in p, and the *degree* of p is the maximal degree of a monomial in p. The size of a polynomial calculus refutation π is the sum of the sizes of the polynomials in π, and the degree of π is the maximal degree of a polynomial in π. We follow the convention of not counting applications of the Boolean axioms toward degree or size by tacitly working over $\mathbb{F}[x_1, \ldots, x_n]/\langle x_1^2 - x_1, \ldots, x_n^2 - x_n\rangle$, which only strengthens any lower bound on either measure. The size and degree measures in polynomial calculus are tightly related through the *size-degree relation* [IPS99], which states that if $\mathcal{P}$ consists of polynomials with constant degree and D is the minimal degree of any polynomial calculus refutation of $\mathcal{P} \cup \{x_1^2 - x_1, \ldots, x_n^2 - x_n\}$, then any refutation of $\mathcal{P} \cup \{x_1^2 - x_1, \ldots, x_n^2 - x_n\}$ must have size $\exp\left(\Omega\left(D^2/n\right)\right)$.

The size-degree relation also applies to the stronger proof system *polynomial calculus resolution (PCR)* [ABRW02], which is defined analogously but where for each variable x_i appearing in $\mathcal{P}$ we also introduce a formal negation $\bar{x}_i$ enforced by adding the equations $x_i + \bar{x}_i - 1 = 0$ to $\mathcal{P}$. It is not hard to see that polynomial calculus and polynomial calculus resolution are equivalent with respect to degree, so to prove a lower bound on polynomial calculus resolution size it suffices to prove a lower bound on polynomial calculus degree. Finally, we remark that a polynomial calculus degree lower bound also applies to the weaker *Nullstellensatz* proof system mentioned in Section 1.1 and Section 1.2.

2.2 Algebra Background

The following material is standard and can be found in, e.g., [MN15]. Let $\mathbb{F}$ be a field and let $\mathbb{F}[x_1, \ldots, x_n]$ be the polynomial ring over $\mathbb{F}$ in n variables. For all our purposes it will suffice to work in the linear space $\mathbb{F}[x_1, \ldots, x_n]/\langle x_1^2 - x_1, \ldots, x_n^2 - x_n\rangle$ of multilinear polynomials, which we do from now on. A *term* is a monomial multiplied by an element of $\mathbb{F}$. A total ordering $\prec$ on the monomials of $\mathbb{F}[x_1, \ldots, x_n]$ is *admissible* if

1. If $Deg(m_1) < Deg(m_2)$, then $m_1 \prec m_2$.
2. For any monomials m_1, m_2 and m such that $m_1 \prec m_2$ and m does not share any variables with either m_1 or m_2, it holds that $mm_1 \prec mm_2$.

We identify the order of a term with the order of its corresponding monomial. The *leading term* of a polynomial $p = \sum_i t_i$ is the largest term in p according to $\prec$. For an ideal I over $\mathbb{F}[x_1, \ldots, x_n]/\langle x_1^2 - x_1, \ldots, x_n^2 - x_n\rangle$, a term t is *reducible modulo* I if it is the leading term of a polynomial q in I. Otherwise, t is *irreducible modulo* I.

A standard fact which is important for our purposes is that for any ideal I, any polynomial p in $\mathbb{F}[x_1, \ldots, x_n]/\langle x_1^2 - x_1, \ldots, x_n^2 - x_n\rangle$ can be uniquely written as $p = q + r$, where $q \in I$ and r is a linear combination of irreducible terms modulo I. We call r the *reduction of p modulo I*. The *reduction operator* R_I is the operator that takes $p = q + r$ to r. Another standard fact we use is that if I_1 and I_2 are ideals such that $I_1 \subseteq I_2$, then $R_{I_2}(pq) = R_{I_2}(pR_{I_1}(q))$.

2.3 Graph Theory

All graphs $G = (V, E)$ considered in this paper are finite, undirected and simple. Given $U, W \subseteq V$, the set of edges between vertices in U is denoted by $E(U)$, the neighbourhood of U in W is denoted

by $N_W(U)$, the set of edges with one endpoint in U and the other in W is denoted by $E(U,W)$, and the induced subgraph of G on U is denoted by $G[U]$. A *simple path of length* ℓ is a tuple of distinct vertices $(v_1,\dots,v_{\ell+1})$ where $(v_i,v_{i+1}) \in E$ for all i. A *simple cycle* is a simple path, except that $v_1 = v_{\ell+1}$. A *k-colouring* of G is a mapping $\chi : V \to [k]$ such that for each edge (u,v) in E it holds that $\chi(u) \neq \chi(v)$. The *chromatic number* $\chi(G)$ is the smallest k such that a colouring exists for G. We say that G is *k-colourable* if $\chi(G) \leq k$.

We will consider two widely studied random graph models. One is the *Erdős-Rényi random graph* $\mathbb{G}(n,p)$ on n vertices where each edge appears independently with probability p. The other is the *random d-regular graph* $\mathbb{G}_{n,d}$ which is a graph selected uniformly at random from the set of d-regular graphs with n vertices. Here a graph is *d-regular* if all of its vertices have degree d, which is possible if and only if $d < n$ and dn is even. For a fixed random graph model $\mathbb{G} = \{\mathbb{G}_n\}_{n=1}^{\infty}$, we say a graph property P holds *asymptotically almost surely* if $\lim_{n\to\infty} \Pr_{G\sim\mathbb{G}_n}[G \text{ has property } P] = 1$.

The graph property that underpins our lower bounds is *sparsity*: a graph G is *(ℓ,ε)-sparse* if for every set $U \subseteq V$ of size at most ℓ, it holds that $|E(U)| \leq (1+\varepsilon)|U|$. For d-regular graphs, it is not hard to see that sparseness is equivalent to the more familiar notion of *expansion*; G is (ℓ,ε)-sparse if and only if G is an $(\ell, d-2(1+\varepsilon))$-expander, where G is an *(ℓ,γ)-expander* if for every subset $U \subseteq V$ of size at most ℓ it holds that $|E(U, G\setminus U)| \geq \gamma|U|$. The following lemma is folklore, see, e.g., [Raz17, Lemma 4.15]. For the precise version that we state below, we provide a proof in the Appendix.

Lemma 2.1 (Sparsity lemma). *Let $G = (V,E) \sim \mathbb{G}$ where $\mathbb{G} = \mathbb{G}_{n,d}$ or $\mathbb{G}(n,d/n)$. If $3 \leq d \leq o(\sqrt{\ln(n)})$, then for every ε such that $d^2/\ln n < \varepsilon < 0.9d-1$, asymptotically almost surely G is $(d^{-30(1+\varepsilon)/\varepsilon}n, \varepsilon)$-sparse.*

We frequently use that large subsets of sparse graphs are 3-colourable, stated in the next lemma.

Lemma 2.2. *Let $G = (V,E)$ be a graph that is (ℓ,ε)-sparse where $\varepsilon < 1/2$. Then for every $U \subseteq V$ of size at most ℓ, $G[U]$ is 3-colourable.*

Proof. The proof is by induction on $|U|$. For the base case, certainly a graph consisting of one vertex is 3-colourable. For the induction step, suppose the claim holds for sets of size at most $s-1$ and consider a set U of size $s \leq \ell$. The average vertex degree in $G[U]$ is $2|E(U)|/s$, which is at most $2(1+\varepsilon) < 3$ by the sparsity condition. Therefore, since graph degrees are integral there exists a vertex v in U with degree at most 2 in $G[U]$. Now consider the set $U \setminus \{v\}$, which is 3-colourable by the inductive hypothesis. We 3-colour $U\setminus\{v\}$ and add v back in. Since the degree of v in U is at most 2 there is at least one colour available for v, so we can extend the 3-colouring to U. □

Finally, we need a rather loose bound on the chromatic number of $\mathbb{G}(n,d/n)$ and $\mathbb{G}_{n,d}$.

Lemma 2.3 ([KPGW10, AN05]). *For $G \sim \mathbb{G}$ where $\mathbb{G}$ is $\mathbb{G}(n,d/n)$ or $\mathbb{G}_{n,d}$, asymptotically almost surely the following hold: $\chi(G) \leq 2d/\log d$; if $d \geq 10$, then $\chi(G) \geq 5$; and if $d \geq 6$, then $\chi(G) \geq 4$.*

2.4 Graph Colouring and Polynomial Calculus

We study the polynomial calculus degree required to refute the system $\mathrm{Col}(G,k)$ of polynomials

$$\sum_{i=1}^{k} x_{v,i} - 1, \qquad v \in V(G) \qquad \text{[Every vertex is assigned a colour]} \tag{2.2a}$$

$$x_{u,i}x_{v,i}, \qquad (u,v) \in E(G) \qquad \text{[No two adjacent vertices get the same colour]} \tag{2.2b}$$

$$x_{v,i}x_{v,i'}, \qquad v \in V(G),\ i \neq i' \qquad \text{[No vertex gets more than one colour]} \tag{2.2c}$$

$$x_{v,i}^2 - x_{v,i}, \qquad v \in V(G),\ i \in [k] \qquad \text{[Boolean axioms]} \tag{2.2d}$$

that states that a graph G is k-colourable. We refer to (2.2a)-(2.2d) as the *k-colourability axioms on G*. It is known [LN17, Proposition 2.2] that a polynomial calculus degree lower bound for $\mathrm{Col}(G, k)$ also applies to *Bayer's formulation* [Bay82] of k-colourability, also known as the *roots-of-unity encoding*. This encoding has received considerable attention in computational algebra [DLMM08, DLMO09, DLMM11, DMP+15, RT22].

To prove degree lower bounds for $\mathrm{Col}(G, k)$, we use a lemma from [Raz98].

Lemma 2.4 ([Raz98]). *Let $\mathcal{P}$ be a set of polynomials in $\mathbb{F}[x_1, \ldots, x_n]/\langle x_1^2 - x_1, \ldots, x_n^2 - x_n\rangle$ with no common root. Suppose there exists an $\mathbb{F}$-linear operator R such that*

1. *$R(1) = 1$.*
2. *$R(p) = 0$, for each p in $\mathcal{P}$.*
3. *For every term t of degree at most $D - 1$ and every variable x, it holds that $R(xt) = R(xR(t))$.*

Then every polynomial calculus refutation of $\mathcal{P}$ requires degree at least D.

We call a linear operator satisfying items 1-3 in Lemma 2.4 an *R-operator*.

3 Techniques and Proof Overview

In this section we introduce the technical tools needed for our lower bounds and provide a proof overview.

3.1 Closure and Ordering by Colouring

For our R-operator, we need an admissible ordering on monomials, which will come from a linear ordering on the vertices of the underlying graph to be specified later. For now, assume $G = (V, E)$ is a graph with a linear ordering $\prec$ on V. An *increasing (decreasing) path* in G is a path $(v_1, \ldots, v_\tau)$ such that $v_i \prec v_{i+1}$ ($v_i \succ v_{i+1}$) for all i in $[\tau - 1]$. For u, v in V, we say that u is a *descendant* of v if there exists a decreasing path from v to u. A subset U of V is *downward-closed* in G if it holds that for every vertex u in U, all descendants of u are also in U. Let D_U denote the set of vertices v such that v is a descendant of some vertex u in U. The *descendant graph* of U, denoted by $\mathrm{Desc}(U)$, is $G[U \cup D_U]$. (Note that the descendant graph of any vertex set is downward-closed.) A *b-hop with respect to U* is a simple path or simple cycle of length $b \geq 2$ whose endpoints are in U and whose other vertices are in $V \setminus U$.

Remark 3.1. Let us give some examples of the properties of G when there are no b-hops in G with respect to U. To start with, if there are no 2-hops with respect to U, then every vertex in $N_{V\setminus U}(U)$ has a unique neighbour in U, and if there are no 3-hops with respect to U, then $N_{V\setminus U}(U)$ is an independent set. In general, the absence of b-hops with respect to U for more values of b will imply similar properties for sets of vertices with small distance to U. Jumping ahead a bit, our lower bounds rely on such properties implied by the absence of $\{2, 3\}$-hops for 4-colouring and $\{2, 3, 4, 5\}$-hops (as well as the absence of some additional small shapes) for 3-colouring.

The point of the next definition is to construct a set with no $\{2, 3\}$-hops that contains a given subset $U \subseteq V$ and is downward-closed.

Definition 3.2 (Closure [RT22]). Let $G = (V, E)$ be a graph and let $U \subseteq V$. Set $H_0 = \mathrm{Desc}(U)$. While there exists a 2-hop or 3-hop, say Q, with respect to $V(H_i)$, set $H_{i+1} = \mathrm{Desc}(V(H_i) \cup V(Q))$; otherwise, stop and set $H_i = H_{\mathrm{end}}$. The *closure* of U, denoted by $\mathrm{Cl}(U)$, is defined to be H_{end}.

We will need the properties of the closure collected in the next lemma.

Lemma 3.3. *Let $U \subseteq U' \subseteq V$. Then the following properties of the closure hold.*

1. *(Well-definedness) Let $(H_0, \dots, H_{\text{end}})$ and $(H'_0, \dots, H'_{\text{end}})$ be two sequences of the construction in Definition 3.2 starting from U. Then $H_{\text{end}} = H'_{\text{end}}$.*
2. *(Monotonicity) $U \subseteq \text{Cl}(U)$ and $\text{Cl}(U) \subseteq \text{Cl}(U')$.*
3. *(Idempotence) $\text{Cl}(\text{Cl}(U)) = \text{Cl}(U)$.*

Proof. First of all, we remark that if P is a $\{2,3\}$-hop with respect to U, then either $V(P) \subseteq U'$ or P contains at least one shorter hop with respect to U'.

For item 1, we show by induction on i that $V(H_i)$ is contained in $V(H'_{\text{end}})$. The base case $i = 0$ is immediate since $V(H_0) = V(H'_0)$ and $V(H'_0) \subseteq V(H'_{\text{end}})$ by construction. For the induction step, suppose that $V(H_{i-1}) \subseteq V(H'_{\text{end}})$. If there is no 2-hop or 3-hop with respect to $V(H_{i-1})$ we are done, so let $V(Q_i)$ be the hop added to $V(H_{i-1})$ in the ith round. By the remark above, either $V(Q_i) \subseteq V(H'_{\text{end}})$, in which case we are also done, or Q contains another hop with respect to $V(H'_{\text{end}})$, but this contradicts that the sequence stops at H'_{end}. It follows by induction that $V(H_{\text{end}}) \subseteq V(H'_{\text{end}})$. By symmetry it also holds that $V(H'_{\text{end}}) \subseteq V(H_{\text{end}})$, and thus $H_{\text{end}} = H'_{\text{end}}$.

For item 2, $U \subseteq \text{Cl}(U)$ by definition. The fact that $\text{Cl}(U) \subseteq \text{Cl}(U')$ can be proved in the same way as item 1.

Item 3 follows by definition since there are no 2-hops or 3-hops in G with respect to $\text{Cl}(U)$. □

The linear ordering on V that we use is the following, which is crucial both in [RT22] and in our arguments.

Definition 3.4 (χ-ordering). Let $G = (V, E)$ be a graph, and let $\chi : V \to [c]$ be a colouring of G. A *χ-ordering* on V is a linear ordering $\prec$ on V such that $u \prec v$ whenever $\chi(u) < \chi(v)$.

A χ-ordering yields an admissible ordering on monomials in the variables $\{x_{v,i}\}_{v \in V(G), i \in [k]}$ of $\text{Col}(G, k)$ as follows. First, order the variables in any way which satisfies that $x_{u,i} \prec x_{v,j}$ whenever $u \prec v$, for all i and j. Then, order the monomials first by degree and then lexicographically according to this variable ordering.

We conclude with two basic facts that are frequently used in the rest of the paper.

Observation 3.5. *If G is χ-ordered by a colouring $\chi : V \to [c]$, the length of any decreasing or increasing path in G is bounded by $c - 1$. If, moreover, G has maximum degree d, then for any $U \subseteq V$ it holds that $|\text{Desc}(U)| \leq cd^{c-1}|U|$.*

3.2 Proof Overview

Recall from Section 1.2 that we define an R-operator on monomials m as reduction modulo I_m, where I_m is an ideal generated by a set of axioms that are "highly relevant" to m. The goal is to satisfy properties 1-3 in Lemma 2.4, which typically requires two technical lemmas with this approach. We call the first of these the *size lemma*, which says that I_m is not too large (under some semantic measure) compared to the degree of m. We call the other the *reduction lemma*, which says that for every ideal I that contains I_m and is not too large under the same measure, it holds that $R_{I_m}(m) = R_I(m)$. Proving these lemmas is the technical bulk of the lower bound proof.

In [RT22], the notions of closure and χ-ordering are used to define I_m for each monomial m, and both the size lemma and reduction lemma are proved using locally tree-like properties of the

underlying graph. For our lower bound for 4-colouring on random regular graphs (Theorem 4.1), we use the same closure construction as in [RT22] but we can not prove the size lemma and reduction lemma in the same way. Instead, we use sparsity properties of random regular graphs. For the improvement to 3-colouring on random regular graphs (Theorem 5.1), we need to strengthen the closure construction (Definition 5.2) and use a more refined argument based on graph contraction in the proof of the reduction lemma (Lemma 5.4). Finally, to establish the analogous lower bound for Erdős-Rényi random graphs (Theorem 6.1), the presence of vertices with high degree requires us to use a different closure (Definition 6.3). Proving the size lemma on the Erdős-Rényi graph requires additional concentration properties of random graphs (for which in particular we need to use more than the usual Chernoff-Hoeffding bounds) together with a labelling argument (Claim 6.5).

4 Lower Bounds for 4-colourability on Random Regular Graphs

In this section we prove Theorem 4.1. In Section 5 we improve the result to $k \geq 3$, but the proof of Theorem 4.1 already contains the main steps and is more intuitive. Therefore, we present a full proof here and describe the necessary technical changes in Section 5.

Theorem 4.1. *Let G be a graph sampled from $\mathbb{G}_{n,d}$. Then, for any $k \geq 4$, asymptotically almost surely every polynomial calculus refutation of $\mathrm{Col}(G,k)$ requires degree $2^{-\mathrm{O}(d)} \cdot n$.*

For constant d, Theorem 4.1 together with the size-degree relation mentioned in Section 2.1 implies polynomial calculus size lower bounds for $\mathrm{Col}(G,k)$ of the form $\exp \Omega(n)$. The constant term in the exponent in Theorem 4.1 can be taken to be 350.

We prove Theorem 4.1 by a series of lemmas. We write $\mathbb{F}(G,k)$ for $\mathbb{F}[x_{v,i} \mid v \in V(G),\ i \in [k]]$ modulo the Boolean axioms on the variables. In what follows, all monomials are assumed to be in $\mathbb{F}(G,k)$, so in particular they are multilinear. For a subset $U \subseteq V$, we define I_U to be the ideal generated by the polynomials in $\mathrm{Col}(G[U],k)$.

Definition 4.2 (Monomial closure). Let m be a monomial, and let $U_m \subseteq V$ be the set of vertices mentioned by variables in m. The *closure* of m, denoted by $\mathrm{Cl}(m)$, is defined to be $\mathrm{Cl}(U_m)$.

Fix a χ-ordering and an admissible ordering induced by χ on monomials. We define R as $R(m) = R_{I_{\mathrm{Cl}(m)}}(m)$, extended linearly to arbitrary polynomials in $\mathbb{F}(G,k)$. To verify items 1-3 in Lemma 2.4 we need some facts about the closure of a monomial m. The first such fact is that the closure is not too large compared to the degree of m if the closure is taken with respect to a χ-ordering on G when G is sparse.

Lemma 4.3 (Size lemma). *Let $G = (V,E)$ be a graph on n vertices with maximal degree d and chromatic number $c \geq 4$, and suppose G is (ℓ,ε)-sparse with $\varepsilon = 1/(4c)$. Moreover, suppose G is χ-ordered by $\chi : V \to [c]$. Let $U \subseteq V$ have size $D \leq \ell/20c$. Then it holds that $|\mathrm{Cl}(U)| \leq 20cd^cD$.*

Proof. Let $(H_0 = \mathrm{Desc}(U), H_1, \ldots, H_{\mathrm{end}} = \mathrm{Cl}(U))$ be a sequence that defines $\mathrm{Cl}(U)$ through the iterative process in Definition 3.2. For each H_i we will define a set $S_i \subseteq V$ such that $H_i = \mathrm{Desc}(S_i)$. We will then use the fact that G is sparse to argue that the subgraph of G induced by S_i becomes too dense to exist in G after a bounded number of rounds. We use this fact to argue that H_{end} is the descendant graph of a set that is not too large, from where the result follows by Observation 3.5.

We set $S_0 = U$. Let Q_i be the hop added to H_{i-1} at round i, and let u, v denote the endpoints of Q_i (possibly, $u = v$). Define S_i from S_{i-1} as follows: Let P_u and P_v be two shortest decreasing paths from S_{i-1} to u and v respectively, and set $S_i = S_{i-1} \cup V(P_u \cup P_v \cup Q_i)$.

Claim 4.4. S_i is well-defined.

Proof. We need to show that for every vertex v in H_i, there exists a decreasing path from S_i to v. The proof of this fact is by induction, where the base case $i = 1$ follows by the definition of H_0 and S_0. For the induction step, suppose the claim holds for $i-1$. By definition, the vertices in H_i which are not in H_{i-1} are descendants of a vertex in Q_i, and all vertices of Q_i are contained in S_i. □

Claim 4.5. The number of vertices in $S_i \setminus S_{i-1}$ is at most $2c + |V(Q_i)| - 4$, and $|E(S_i)| \geq |E(S_{i-1})| + |S_i \setminus S_{i-1}| + 1$.

Proof. By Observation 3.5, P_u and P_v contain at most c vertices each. Denote the graph $P_u \cup P_v \cup Q_i$ by F. Note that F is connected and $3 \leq |V(F)| \leq 2c + |V(Q_i)| - 2$. Moreover, the endpoints of F are contained in S_{i-1} and all other vertices in F are outside of S_{i-1}. By our choice of P_u and P_v there are two cases, depending on whether the endpoints of F are distinct or not.

Case 1: If $|V(F) \cap S_{i-1}| = 2$, the number of vertices in $S_i \setminus S_{i-i}$ is $|V(F)|-2$, while $|E(F)| \geq |V(F)|-1$ since F is connected.

Case 2: If $|V(F) \cap S_{i-1}| = 1$, F contains a cycle. In this case $|S_i \setminus S_{i-i}| = |V(F)| - 1$ while $|E(F)| \geq |V(F)|$. Moreover, in this case $|V(F)| \leq 2c + |V(Q_i)| - 3$.

In both cases, $|S_i \setminus S_{i-i}| \leq 2c + |V(Q_i)| - 4$ and $|E(S_i)| \geq |E(S_{i-1})| + |S_i \setminus S_{i-1}| + 1$, and the claim follows. □

Now we can prove our upper bound on i. By Claim 4.5 it follows that

$$\frac{|E(S_{4D+1})|}{|S_{4D+1}|} = \frac{|E(U)| + \sum_{i=1}^{4D} |E(S_{i+1}) \setminus E(S_i)|}{|U| + \sum_{i=1}^{4D} |S_{i+1} \setminus S_i|} \geq \frac{|E(U)| + \sum_{i=1}^{4D} (|S_{i+1} \setminus S_i| + 1)}{|U| + \sum_{i=1}^{4D} |S_{i+1} \setminus S_i|} > 1 + \frac{1}{4c}, \tag{4.1}$$

which contradicts that G is $(\ell, 1/(4c))$-sparse since $|S_{4D+1}| < 15cD < \ell$. Therefore, the number of rounds in the construction of $\mathrm{Cl}(U)$ is bounded by $4D$.

At round i in the construction, H_i is the descendant graph of $U \cup \bigcup_{j \leq i} V(Q_i)$. In every round i, the hop Q_i added to $V(H_{i-1})$ contains at most 2 vertices not already in H_i, so a rather loose upper bound on $|U \cup \bigcup_{j \leq i} V(Q_i)|$ in the last round, where $i \leq 4D$ as we showed, is $20D$. (We want this bound to be loose in order for the same estimate to apply for a modified closure which we define in Section 5.) With this estimate in place, it follows from Observation 3.5 that $|\mathrm{Cl}(U)| \leq 20cd^cD$, as desired.

□

The next lemma informally states that there is no difference between reducing a monomial m modulo $I_{\mathrm{Cl}(m)}$ and reducing modulo a slightly larger ideal.

Lemma 4.6 (Reduction lemma). *Let $G = (V, E)$ be a graph on n vertices with chromatic number $c \geq 4$, and suppose G is (ℓ, ε)-sparse with $\varepsilon = 1/(4c)$. Let m be a monomial with closure $\mathrm{Cl}(m)$. Then, for any $U \subseteq V$ such that $|U| \leq \ell$ and $\mathrm{Cl}(m) \subseteq U$, it holds that m is reducible modulo I_U if and only if m is reducible modulo $I_{\mathrm{Cl}(m)}$.*

Proof. One direction is immediate. For the other, we want to prove that if m is the leading term of a polynomial in I_U, it is also the leading term of a polynomial in $I_{\mathrm{Cl}(m)}$. For brevity, we write $\mathrm{Cl}(m)$ for $\mathrm{Cl}(m)$ from now on. The generators of I_U are the k-colourability axioms restricted to U. We partition these axioms into three sets: A_1 is the axioms on J; A_2 is the axioms on $U \setminus (J \cup N_{U \setminus J}(J))$;

and A_3 contains the rest, which is the union of the edge axioms on $E(J \cup N_{U\setminus J}(J), U \setminus J)$ and the vertex axioms on $N_{U\setminus J}(J)$. Consider a polynomial f in I_U with leading term m. Then we can write

$$f = \sum_{a_1 \in A_1} a_1 p_a + \sum_{a_2 \in A_2} a_2 p_{a_2} + \sum_{a_3 \in A_3} a_3 p_{a_3}\,, \tag{4.2}$$

where each $p_{a_1}, p_{a_2}, p_{a_3}$ is some polynomial in $\mathbb{F}(G,k)$. Note that G is $(\ell, 1/(4c))$-sparse and $|U \setminus J| \leq \ell$, so it follows by Lemma 2.2 that $U \setminus J$ is 3-colourable. Take any 3-colouring witnessing this fact and assign it to the variables in $U \setminus (J \cup N_{U\setminus J}(J))$. This results in a 3-colouring of $U \setminus (J \cup N_{U\setminus J}(J))$ for which at most 2 colours are used for the neighbours in U of any vertex in $N_{U\setminus J}(J)$, since we started with a colouring of $U \setminus J$. Write ρ for the corresponding assignment to the variables. We have that

$$f\restriction_\rho = \sum_{a \in A_1} a \cdot p_a\restriction_\rho + \sum_{a'' \in A_3} a''\restriction_\rho \cdot p_{a''}\restriction_\rho\,, \tag{4.3}$$

where the middle term in (4.2) has disappeared since we assigned a 3-colouring to $U \setminus (J \cup N_{U\setminus J}(J))$, and moreover all of the axioms in A_1 are untouched. Note that m is still the leading term of $f\restriction_\rho$ since we did not assign any of its variables and the order of any other terms in f can only decrease after a restriction.

What remains is to transform $f\restriction_\rho$ into a polynomial in I_J while retaining m as its leading term. Note that there are no 3-hops with respect to J by definition, so the vertices in $N_{U\setminus J}(J)$ form an independent set in U. Hence, we can write the axioms a'' in A_3 before applying ρ as

$$\sum_{i=1}^{k} x_{v,i} - 1\,, \qquad v \in N_{U\setminus J}(J) \tag{4.4a}$$

$$x_{u,i}x_{v,i}\,, \qquad i \in [k],\ (u,v) \in E(J, N_{U\setminus J}(J)) \tag{4.4b}$$

$$x_{u,i}x_{v,i}\,, \qquad i \in [k],\ (u,v) \in E\big(U \setminus (J \cup N_{U\setminus J}(J)), N_{U\setminus J}(J)\big) \tag{4.4c}$$

$$x_{v,i}x_{v,i'}\,, \qquad v \in N_{U\setminus J}(J)\,, i \neq i' \tag{4.4d}$$

After applying ρ, at most two colours are made unavailable to each v in $N_{U\setminus J}(J)$ since we started with a 3-colouring to $U \setminus J$. For simplicity, suppose exactly two colours are made unavailable for each v in $N_{U\setminus J}(J)$ (the argument is easily extended to the other cases). Fix one such v, and suppose without loss of generality that the unavailable colours for v are 1 and 2. Then further set $x_{v,1}$ and $x_{v,2}$ to 0, which makes the axioms in (4.4c) that mention v vanish. Since there are no 2-hops with respect to J, there is a unique u in J for each v in $N_{U\setminus J}(J) \setminus J$ in (4.4b). Therefore, after applying the restrictions, the axioms that mention v become

$$\sum_{i=3}^{k} x_{v,i} - 1; \qquad x_{u,i}x_{v,i}\,, i \in [3,k],\ \{u\} = N_J(\{v\}); \qquad x_{v,i}x_{v,i'}\,, i \neq i' \text{ and } i, i' \neq \{1,2\}. \tag{4.5}$$

We now apply the degree-1 substitution

$$x_{v,3} \mapsto x_{u,4}; \qquad x_{v,4} \mapsto \sum_{1 \leq i \leq k, i \neq 4} x_{u,i}; \qquad x_{v,i} \mapsto 0\,, i > 4 \tag{4.6}$$

to f that turns each axiom in (4.5) into a linear combination of the axioms in J. (Note that it is here we use that $k \geq 4$ so that at least 2 colours are still available for v.) To see this, note that the axioms in (4.4a) on v turn into the vertex axioms on u, and the axioms (4.4b) and (4.4c) on v turn into sums

of the corresponding axioms on u. Therefore, applying the analogous procedure in turn for all v in $N_{U\setminus J}(J)$ turns $f\restriction_\rho$ into a polynomial f^* in I_J.

It remains to argue that m is the leading term of f^*. To see this, we note that since J is downward-closed, it must be the case that $u \prec v$ in the ordering on G, and so each monomial in f affected by the substitution decreases in the induced order on $\mathbb{F}(G,k)$. Therefore, since m was the leading term before the substitution and no variable in m was substituted, it follows that m is the leading term of f^*. □

Proof of Theorem 4.1. We assume that G has the following two properties, which hold asymptotically almost surely: $\chi(G) \leq 2d/\log d$ (Lemma 2.3) and G is $(2^{-300d}n, 1/(4\chi(G)))$-sparse (Lemma 2.1). Let $\chi : V \to [c]$ be a colouring of G, where c is the chromatic number of G, and χ-order G according to this colouring. Recall that $R(m) = R_{I_{\mathrm{Cl}(m)}}$. Fix a parameter $D = 2^{-300d}n/(20cd^c)$. We need to verify that items 1-3 in Lemma 2.4 hold for R with this choice of D.

That $R(1) = 1$ is for free since the closure of a constant polynomial is empty by definition. To show that $R(p) = 0$ for each p in the axioms, let m_p be the product of the variables mentioned by p. Note that $Deg(m_p) \leq k$, so by Lemma 4.3 we have that $|\mathrm{Cl}(m_p)| \leq 20c^2d^ck < \ell$. Therefore, for each axiom $p = \sum_j b_j m_j$ we have the sequence of equalities

$$\begin{aligned}
R(p) &= \sum_j b_j R(m_j) && && (4.7a)\\
&= \sum_j b_j R_{I_{\mathrm{Cl}(m_j)}}(m_j) && \text{[By definition]} && (4.7b)\\
&= \sum_j b_j R_{I_{\mathrm{Cl}(m_p)}}(m_j) && \text{[By Lemma 3.3 and Lemma 4.6]} && (4.7c)\\
&= R_{I_{\mathrm{Cl}(m_p)}}\Big(\sum_j b_j m_j\Big) && \text{[By the linearity of } R\text{]} && (4.7d)\\
&= R_{I_{\mathrm{Cl}(m_p)}}(p) && \text{[By definition]} && (4.7e)\\
&= 0\,, &&
\end{aligned}$$

where the last line follows since p is an element of $I_{\mathrm{Cl}(m_p)}$.

Finally, we need to show that for every term t of degree at most $D-1$ and every variable x, it holds that $R(xt) = R(xR(t))$. This follows by a similar sequence of equalities as above, with one subtle step that we record as a separate claim.

Claim 4.7. If t is a term of degree at most $D-1$, it holds that

$$\sum_{t'\in R(t)} R_{I_{\mathrm{Cl}(xt')}}(xt') = \sum_{t'\in R(t)} R_{I_{\mathrm{Cl}(xt)}}(xt') \tag{4.8}$$

Proof. Our choice of D and Lemma 4.3 together imply that $|\mathrm{Cl}(xt)| \leq 20cd^cD < \ell$. Therefore, if we can show that also $\mathrm{Cl}(xt') \subseteq \mathrm{Cl}(xt)$ for each term t' in $R(t)$, we can apply Lemma 4.6, from where the claim follows immediately. To this end, consider a term t' in $R(t)$. Note that $U_{t'} \subseteq \mathrm{Cl}(t)$ and that $U_{xt'} = U_x \cup U_{t'}$. Clearly $U_x \subseteq \mathrm{Cl}(xt)$, and by monotonicity of the closure, also $\mathrm{Cl}(t) \subseteq \mathrm{Cl}(xt)$. Therefore, $U_x \cup U_{t'} \subseteq \mathrm{Cl}(xt)$. Again by monotonicity, $\mathrm{Cl}(xt') = \mathrm{Cl}(U_x \cup U_{t'}) \subseteq \mathrm{Cl}(\mathrm{Cl}(xt))$. Finally, by idempotence of the closure it holds that $\mathrm{Cl}(\mathrm{Cl}(xt)) = \mathrm{Cl}(xt)$. The claim follows. □

The sequence of equalities

$$
\begin{aligned}
R(xR(t)) &= \sum_{t' \in R(t)} R(xt') && && (4.9a)\\
&= \sum_{t' \in R(t)} R_{I_{\mathrm{Cl}(xt')}}(xt') && \text{[By definition]} && (4.9b)\\
&= \sum_{t' \in R(t)} R_{I_{\mathrm{Cl}(xt)}}(xt') && \text{[By Claim 4.7]} && (4.9c)\\
&= R_{I_{\mathrm{Cl}(xt)}}\Big(\sum_{t' \in R(t)} xt' \Big) && \text{[By the linearity of } R\text{]} && (4.9d)\\
&= R_{I_{\mathrm{Cl}(xt)}}(xR_{I_{\mathrm{Cl}(t)}}(t)) && \text{[By the definition of } t'\text{]} && (4.9e)\\
&= R_{I_{\mathrm{Cl}(xt)}}(xt) && \text{[Since } R_{I_2}(xR_{I_1}(t)) = R_{I_2}(xt) \text{ if } I_1 \subseteq I_2\text{]} && (4.9f)\\
&= R(xt), &&
\end{aligned}
$$

together with our parameter choice $D = 2^{-300d}n/(20cd^c) > 2^{-350d}n$ concludes the proof. □

5 Improvement to 3-colourability

In this section we improve our lower bound for 4-colourability (Theorem 4.1) to a lower bound for 3-colourability.

Theorem 5.1. *Let $G = (V, E)$ be a graph sampled from $\mathbb{G}_{n,d}$. Then, for any $k \geq 3$, asymptotically almost surely every polynomial calculus refutation of* $\mathrm{Col}(G, k)$ *requires degree* $2^{-\mathrm{O}(d \log d)} \cdot n$.

As in Section 4, if we set d to be any constant, our results imply polynomial calculus size lower bounds for refuting $\mathrm{Col}(G, k)$ of the form $\exp \Omega(n)$. The constant term in the exponent in Theorem 5.1 can be again be taken to be 350.

Asymptotically almost surely, $\chi(G) \leq 2d/\log d$ (Lemma 2.3) and G is $(2^{-300d \log d}n, 1/(4(d+1)))$-sparse (Lemma 2.1). We prove Theorem 5.1 assuming these properties. We need to apply Lemma 2.2 in a more refined way in Lemma 4.6. This is enabled by a slightly extended notion of closure which in addition to the usual b-hops also involves *b-lassos*, which is a b-hop where the first and last *edges* coincide.

Definition 5.2 (Strong closure). Let $G = (V, E)$ be a graph and let $U \subseteq V$. Let H_1 be the descendant graph of U. While there exists a b-hop for $b = 2, 3, 4, 5$, a 5-lasso, or a 6-lasso, say Q, with respect to H_i, define $H_{i+1} = \mathrm{Desc}(V(H_i) \cup V(Q))$. If there are no b-hops or $\{5, 6\}$-lassos with respect to H_i, stop and set $H_i = H_{\mathrm{end}}$. The *strong closure* of U, denoted by $\mathrm{Cl}^*(U)$, is defined to be H_{end}.

It should be noted that the arguments in this section can be made to work while only including $\{2, 3, 4\}$-hops and 5-lassos in the strong closure. The additional hops and lassos that we include make the arguments cleaner and does not worsen the lower bound in any significant way.

Using the same remark as in the proof of Lemma 3.3, which holds also when including $\{5, 6\}$-lassos, it is not hard to check that the strong closure is well-defined and has the properties in Lemma 3.3. It suffices to prove the analogues of the size lemma and reduction lemma for the strong closure; they are Lemma 5.3 and Lemma 5.4 respectively. The rest of the proof is entirely analogous to the proof of Theorem 4.1.

Lemma 5.3 (Size lemma). *Let $G = (V, E)$ be a graph on n vertices with maximum degree d and chromatic number c, and suppose G is (ℓ, ε')-sparse, where $\varepsilon' = 1/4(d+1)$. Moreover, suppose G is χ-ordered by $\chi : V \to [c]$. Let $U \subseteq V$ have size $D \leq \ell/20d$. Then it holds that $|\mathrm{Cl}^*(U)| \leq 20cd^{c+1}D$.*

Proof sketch. The proof of Lemma 4.3 can be applied almost verbatim; the only differences are that ε is now $1/4(d+1)$ which is at most $1/(4c)$ in any graph (we need this change for (5.2) below) and that the number of vertices added to S_i each round is now at most $2c+2$ instead of $2c$. Our estimates in Lemma 4.3 are loose enough to make the proof go through with the same parameters with these changes. □

Lemma 5.4 (Reduction lemma). *Let $G = (V, E)$ be a graph on n vertices with maximal degree d and chromatic number c, and suppose G is (ℓ, ε')-sparse, where $\varepsilon' = 1/4(d+1)$. Let m be a monomial with strong closure $\mathrm{Cl}^*(m)$. Then, for any $U \subseteq V$ such that $|U| \leq \ell$ and $\mathrm{Cl}^*(m) \subseteq U$, it holds that m is reducible modulo I_U if and only if m is reducible modulo $I_{\mathrm{Cl}^*(m)}$.*

Proof. As in the proof of Lemma 4.6, write J for $\mathrm{Cl}^*(m)$ for brevity. Let $\{v_1, \dots, v_t\}$ denote the vertices in $N_{U\setminus J}(J)$ and write A for $U \setminus (J \cup N_{U\setminus J}(J))$. Our goal is to colour $G[A]$ so that the neighbourhood of each v_i in $N_{U\setminus J}(J)$ is coloured with *one* colour instead of two as in Lemma 4.3, which makes the substitution argument in the proof of Lemma 4.3 go through for $k \geq 3$ instead of $k \geq 4$. To this end, write M_i for $N_{U\setminus J}(v_i) \setminus J$, for i in $[t]$. For i in $[t]$, write B_i for $\{v_i\} \cup M_i$, which is then a star in U with center v_i .

Note the following property of the strong closure: since there are no $\{3,4,5\}$-hops or 5-lassos with respect to J, $B_1, \dots, B_t$ are mutually disjoint and the union of $M_1, \dots, M_t$ is an independent set in G. Moreover, since there are no 6-lassos with respect to J, any two vertices in the same set M_i have no common neighbor in G except v_i. Let G' be the graph obtained from $G[A]$ by contracting each B_i to a single vertex. By the above properties, G' has no self-loops nor multi-edges. To 3-colour $G[A]$ so that the neighbourhood of each v_i has the same colour, it suffices to 3-colour G' since we can then expand G' back to $G[A]$ and use the same colour for all vertices in M_i, again by the above. To see that G' is 3-colourable it suffices to show that G' is $(|G'|, \delta)$-sparse for some $\delta < 1/2$ by Lemma 2.2. To this end, fix any $T' \subseteq V(G')$. We estimate $|E[T']|$ in terms of $|E[T]|$, where T is the preimage of T' in the contraction. Suppose the number of B_i's in T is j. Then $j \leq |T'|$. Write s for $\sum_{i=1}^{j}(|B_i| - 1)$. Then we have that

$$s \leq (d-1)j \leq (d-1)|T'|, \tag{5.1}$$

and moreover, $|T| = |T'| + s$ and $|E(T)| = |E(T')| + s$. The latter is due to the fact that the only edges in T that are contracted are those inside a star B_i in T, and there are no edges between different B_i's. Since G is $(\ell, 1/4(d+1))$-sparse and $|T| \leq \ell$ by assumption, it follows that

$$|E(T')| \leq (1 + \frac{1}{4(d+1)})|T'| + \frac{s}{4(d+1)} \leq (1 + \frac{1}{4(d+1)})|T'| + \frac{(d-1)|T'|}{4(d+1)} < (1 + \frac{1}{4})|T'|, \tag{5.2}$$

so G' is $(|G'|, 1/4)$-sparse. Hence we can 3-colour $G[A]$ in the desired way. Now we note that the absence of $\{2,3\}$-hops allows us to apply the substitution argument in Lemma 4.6, but with $k \geq 3$ instead of $k \geq 4$ using our improved colouring. The lemma follows. □

6 Lower Bounds for 3-colourability on the Erdős-Rényi Graph

In this section we prove Theorem 6.1.

Theorem 6.1. *Let $G = (V, E)$ be a graph sampled from $\mathbb{G}(n, d/n)$. Then, for any $k \geq 3$, asymptotically almost surely every polynomial calculus refutation of $\mathrm{Col}(G, k)$ requires degree $2^{-d^{O(1)}} \cdot n$.*

As in previous sections, Theorem 6.1 implies polynomial calculus size lower bounds for $\mathrm{Col}(G, k)$ of the form $\exp \Omega(n)$ when d is constant. The constant factor in the double exponent can be taken to be 1000.

We modify the argument in Section 5 to work for $\mathbb{G}(n, d/n)$. The steps are the same as before, but now there are vertices with superconstant degree so we cannot use the trivial descendant graph size bound in Observation 3.5. The solution is to contain all vertices with superconstant degree in a single "bad" set, include this set in the closure of every monomial, and reduce modulo this set instead. To this end, given G we fix some additional parameters:

$$p = d/n; \qquad \text{[The edge probability in } \mathbb{G}(n, d/n)\text{]} \tag{6.1a}$$
$$B = d^{801}; \qquad \text{[}(B+1)d \text{ is the degree threshold for the bad set]} \tag{6.1b}$$
$$W = \{v \in V(G) \mid \deg(v) \geq (B+1)d\}; \qquad \text{[The bad set]} \tag{6.1c}$$
$$\varepsilon = 1/(4Bd); \qquad \text{[Sparsity parameter]} \tag{6.1d}$$
$$\alpha = d^{-40/\varepsilon}; \qquad \text{[Sparsity parameter]} \tag{6.1e}$$
$$\beta = B^{-Bd/4}; \qquad \text{[We prove a } \beta n \text{ degree lower bound]} \tag{6.1f}$$

We need one more graph property that holds asymptotically almost surely.

Lemma 6.2. *For $G = (V, E) \sim \mathbb{G}(n, p)$, asymptotically almost surely the following properties hold.*

1. *For every $S \subseteq V$, if $\beta n \leq |S| \leq 100\beta n$ then $|E(S, V \setminus S)| \leq Bd|S|$. In particular, $|N_{V \setminus S}(S)| \leq Bd|S|$.*
2. *$|W| \leq \beta n$.*

Proof. Let $\mathcal{A}$ denote the event in item 1. Note that $|E(S, V \setminus S)|$ is a sum of $|S| \cdot |V \setminus S|$ independently and identically distributed binary random variables, each with expectation $p = d/n$. By the union bound and Bennett's inequality [Ben62] (which improves Chernoff-Hoeffding in our case by a factor of $\log B$ in the exponent) it follows that

$$\Pr[\neg\mathcal{A}] \leq \sum_{i=\lceil \beta n \rceil}^{\lceil 100\beta n \rceil} \binom{n}{i} \exp\left(-\frac{(B \log B) i(n-i)d}{2n}\right) \tag{6.2a}$$
$$< 100\beta n \exp\left(\frac{\beta}{2} n \left(\log \frac{1}{\beta} - (B \log B)(1 - 100\beta)d\right)\right) \tag{6.2b}$$
$$< 100\beta n \exp\left(-\frac{\beta n d B}{100} \log B\right) \tag{6.2c}$$
$$< \exp\left(-\Omega_d(n)\right), \tag{6.2d}$$

which approaches 0 as $n \to \infty$. For item 2, let $\mathcal{B}$ denote the event "G is $(\alpha n, \varepsilon)$-sparse". Then the event $\mathcal{A} \wedge \mathcal{B}$ also holds asymptotically almost surely, so item 2 follows if we prove that it is logically implied by $\mathcal{A} \wedge \mathcal{B}$, or equivalently that $\neg$(item 2) $\wedge\, \mathcal{B}$ implies $\neg\mathcal{A}$. To this end, suppose $|W| > \beta n$. Consider a subset $T \subseteq W$ of size $\lceil \beta n \rceil < \alpha n$. By the definition of T and by $\mathcal{B}$, it holds that

$$|E(T, V \setminus T)| \geq (B+1)d|T| - 2(1+\varepsilon)|T| > Bd|T|, \tag{6.3}$$

which implies $\neg\mathcal{A}$. □

As before we can assume that $\chi(G) \leq 2d/\log d$ and G is $(\alpha n, \varepsilon)$-sparse. From now on, we also assume items 1 and 2 in Lemma 6.2 hold for G and write c for $\chi(G)$ for brevity. Given these properties, fix a set T that contains W and has size $\lceil \beta n \rceil$.

Definition 6.3 (Relative closure). Let $U \subseteq V$ have size at most βn. The *relative closure of U with respect to T* is the strong closure of $U \cup T$, and is denoted by $\mathrm{Cl}^*_T(U)$.

Of course, the properties in Lemma 3.3 hold for the relative closure as well since they hold for the strong closure, so most of the arguments in the previous sections go through for this closure with minor modifications. The point that needs a new argument is the size lemma for the relative closure, where we now have to drop the maximum degree assumption on G and hence cannot use the trivial descendant graph size bound from Observation 3.5.

Lemma 6.4 (Size lemma). *For any $U \subseteq V$ of size at most βn, it holds that $|\mathrm{Cl}^*_T(U)| \leq (2Bd)^{c+1} c \cdot 100\beta n$.*

We remark that with these parameters, for every U such $|U| \leq \beta n$, we have that $|\mathrm{Cl}^*_T(U)| \leq (2Bd)^{c+1} c \cdot 100\beta n < 2^{-(Bd/5)\log B} n < \alpha n$, where the last two inequalities follow from the parameter choice, which gives $2^{-(Bd/5)\log B} < 2^{-160Bd\log d} = \alpha$, and from the assumption that $c \leq 2d/\log d$, respectively.

Proof of Lemma 6.4. Most of the proof of Lemma 4.3 goes through, but there are three differences: First, as in Section 5 we include larger hops in the closure, but as we remarked there the estimates are loose enough to accommodate this change. Second, we have changed the sparseness parameters to $(\alpha n, \varepsilon)$-sparse, but since we assume $c \leq 2d/\log d$ we have that $1/(4c) > \varepsilon$, so the calculation (4.1) still applies. Third, it still holds that if $|U \cup T| \leq 2\beta n < \alpha n$, then the number of rounds in the construction of $\mathrm{Cl}^*_T(U)$ is at most $4|U \cup T|$. Therefore $\mathrm{Cl}^*_T(U)$ is the descendant graph of a set A such that $|A| \leq 20|U \cup T|$. What is needed to finish the proof is an analogue of Observation 3.5, which is the content of the next claim.

Claim 6.5. For every set A such that A contains T and $|A| \leq 100\beta n$, it holds that $\mathrm{Desc}(A) \leq (2Bd)^{c+1} c|A|$.

Proof. Let us label the vertices in $\mathrm{Desc}(A)$ as follows: If $v \in A \cup N_{G\setminus A}(A)$, label v by itself; otherwise v is in $\mathrm{Desc}(A) \setminus (A \cup N_{G\setminus A}(A))$. Then by definition there must be a vertex $u \in N_{G\setminus A}(A)$ and a decreasing path from u to v such that the path does not intersect A (nor T since $T \subseteq A$). In this case, label v by the decreasing path P_{uv}. Note that P_{uv} has length at most $c-1$, and moreover every vertex on $P_{u,v}$ has degree at most $(B+1)d$. The number of such paths is then at most $|N(A)| \cdot c((B+1)d)^{c-1}$. Clearly every vertex in $\mathrm{Desc}(A)$ gets a unique label, and the total number of labels is at most

$$|A| + |N_{G\setminus A}(A)| + |N_{G\setminus A}(A)| \cdot c((B+1)d)^{c-1} < (2Bd)^{c+1} c|A|, \tag{6.4}$$

where we have used that $|N_{G\setminus A}(A)| \leq Bd|A|$ by item 1 of Lemma 6.2. □

Using Claim 6.5 instead of Observation 3.5 in the proof of Lemma 4.3 yields the size lemma for the relative closure . □

Next, we need to establish the reduction lemma as in Section 5.

Lemma 6.6 (Reduction lemma). *Let G be as above. Let m be a monomial with closure $\mathrm{Cl}^*_T(m)$. Then, for any $U \subseteq V$ such that $|U| \leq \beta n$ and $\mathrm{Cl}^*_T(m) \subseteq U$, it holds that m is reducible modulo I_U if and only if m is reducible modulo $I_{\mathrm{Cl}^*_T(m)}$.*

Proof sketch. The proof is essentially analogous to the proof of Lemma 5.4. Our sparseness parameters are now $(\alpha n, \varepsilon)$. We do not have a maximum degree assumption on G, but note that by definition the maximal degree of any vertex in $G \setminus \mathrm{Cl}^*_T(m)$ is $(B+1)d$, so we can bound the star sizes by $(B+1)d$ and carry out the analogue of (5.1). By our parameter choice, $\varepsilon(B+1)d < 1/2$ so the analogue of (5.2) also holds. Once this is done, the rest of the proof is the same as for Lemma 5.4 once we note that $\mathrm{Cl}^*_T(m) \leq \alpha n$ for any m with degree at most βn. □

Proof of Theorem 6.1. As described above, we can run the proof of Theorem 5.1 using Lemma 6.4 and Lemma 6.6 for all monomials of degree at most βn using the R-operator defined by $R(m) = R_{\mathrm{Cl}^*_T(m)}(m)$. With this R-operator it is no longer immediate that $R(1) = 1$, but this follows from the fact that T is 3-colourable by Lemma 2.2, so the 3-colouring axioms on T are satisfiable and in particular 1 is irreducible modulo I_T. Items 2 and 3 in Lemma 2.4 go through unchanged. □

7 Concluding Remarks

In this paper, we show that polynomial calculus requires linear degree to refute that a sparse random regular graph or Erdős-Rényi graph is 3-colourable, which is optimal up to constant factors. This implies exponential size lower bounds for the same problem by the well-known size-degree relation for polynomial calculus [IPS99].

It would be interesting to investigate whether our techniques can be applied to lower bounds for colouring principles in other proof systems, where the most obvious candidates are sums-of-squares and Sherali-Adams. On a similar note, the closure operation defined in [RT22] and extended in this work, are not per se connected to colouring and it is therefore conceivable that it could find applications in a wider context than colouring.

Our degree lower bounds are of the form $n/f(d)$, where d is either degree or average degree depending on the random graph model. In previous works, f can be taken to be polynomial in d [BCMM05, LN17], but in our results f is exponential in d. It is not immediately clear, however, what the correct dependency on d should be. There are randomized algorithms based on semidefinite programming that with high probability can distinguish between 3-colourable graphs and $(\log^2 n)$-regular random graphs [KMS94]. While the precise dependency on d is immaterial in our setting of sparse random graphs, it would be interesting to know if this can be improved.

A Appendix

In this section we prove a quantitative version of the folklore result in Lemma 4.15 in [Raz17], which we previously call the Sparsity lemma. We make no claim of originality.

Theorem A.1 (Sparsity lemma). *Let $G = (V, E) \sim \mathbb{G}$, where $\mathbb{G} = \mathbb{G}_{n,d}$ or $\mathbb{G}(n, d/n)$ and $3 \leq d \leq o(\sqrt{\ln(n)})$. For every ε such that $d^2/\ln n < \varepsilon < 0.9d - 1$, it holds asymptotically almost surely that G is $(d^{-30(1+\varepsilon)/\varepsilon} n, \varepsilon)$-sparse.*

Proof. First, we prove the theorem for $\mathbb{G}_{n,d}$. We sample G from $\mathbb{G}_{n,d}$ in the *configuration model*. In this model, for each vertex v of G there is a cell C_v with d elements $c_1^v, \ldots, c_d^v$. We first sample a G by sampling a perfect matching of $\sqcup_v C_v$ (note that this requires that dn is even) and then collapsing the cells into vertices while retaining matched edges. Next, we repeat this procedure independently until we get a G that is simple (i.e. in the perfect matching, there are no vertices in the same cell matched to each other, and no two cells have more than one edge between them). It's not hard to see that this will give us an element in $\mathbb{G}_{n,d}$ uniformly at random.

The probability that G from the first step is simple is $\sim \exp\left(\frac{1-d^2}{4}\right)$ when $d < \sqrt{2\ln n}$ [Wor99]. So conditioning on G being simple would boost the probability of any 'bad' event about G by at most a factor of $\exp(d^2)$. In the following, we consider the bad event that G is not $(d^{-30(1+\varepsilon)/\varepsilon}n, \varepsilon)$-sparse (defined in the natural sense). We show that this happens with probability at most $n^{-0.1\varepsilon}$, which will prove the conclusion.

To this end, fix $\varepsilon > 0$, and let α be a constant smaller than $1/2d$ (for technical reasons), to be determined later. Let $\mathcal{A}$ denote the event "G is $(\alpha n, \varepsilon)$-sparse". By the union bound,

$$\Pr[\neg\mathcal{A}] \leq \sum_{U\subset V, |U|\leq \alpha n} \Pr[|E(G[U])| \geq (1+\varepsilon)|U|] \tag{A.1}$$

and by symmetry, $\Pr[|E(G[U])| \geq (1+\varepsilon)|U|]$ depends only on $|U|$.

Fix a set U of size $s \leq \alpha n$ and let $S_{ds,q}$ denote a sum of ds independently and identically distributed Bernoulli random variables with head probability $q := s/(n-s)$. We want to upper bound $\Pr[|E(G[U])| \geq (1+\varepsilon)|U|]$ by $\Pr[S_{ds,q} \geq (1+\varepsilon)s]$. For this, we think of the elements in the cells as being matched one at a time in the configuration model. At each step, ds is an upper bound on the number of elements in the cell of a vertex in U that is available to match to, and $d(n-s)$ is a lower bound on the total number of available elements to match to. Since at least one element of a cell in S is matched at each step, there are at most ds steps in total. An edge appears in U if and only if an element in a cell in U is matched to another element in a cell in U, and at each step this happens with probability bounded by $ds/d(n-s) = q$, so indeed

$$\Pr[|E(G[U])| \geq (1+\varepsilon)|U|] \leq \Pr[S_{ds,q} \geq (1+\varepsilon)s]. \tag{A.2}$$

For a, b in $(0,1)$, the *KL divergence* of a and b is $D(a \parallel b) = a\log\frac{a}{b} + (1-a)\log\frac{1-a}{1-b}$. By the Chernoff bound,

$$\Pr[S_{ds,q} \geq (1+\varepsilon)s] \leq 2^{-D\left(\frac{1+\varepsilon}{d} \parallel q\right)ds}. \tag{A.3}$$

Denote $(1+\varepsilon)/d$ by p, and the binary entropy function by H. Then

$$D(p \parallel q) = -H(p) + p\log\frac{1}{q} + (1-p)\log\frac{1}{1-q} \geq -H(p) + p\log\frac{1}{q}. \tag{A.4}$$

We estimate the RHS of (A.4). First, note $H(p) = p\log\frac{1}{p} + (1-p)\log\frac{1}{1-p} < p\log d + \log\frac{1}{1-p}$, and the elementary inequality $\ln(1/(1-y)) \leq y/(1-y)$ gives that $\log\frac{1}{1-p} \leq \frac{1}{\ln 2}\frac{p}{1-p} < 15p$ (recall $\frac{1}{\ln 2} < 1.5$, $p = \frac{(1+\varepsilon)}{d} \leq 0.9$). Hence

$$H(p) < p\left(\log d + 15\right) < 15p\log d. \tag{A.5}$$

Next, for the last term in (A.4), which is $p\log\frac{1}{q} = \frac{p}{q}\left(H(q) - (1-q)\log\frac{1}{1-q}\right)$, the same elementary inequality gives that $\log\frac{1}{1-q} \leq \frac{1}{\ln 2}\frac{q}{1-q} < 2q$, where the last step uses $1/\ln 2 < 1.5$ and $q = s/(n-s) \leq \alpha/(1-\alpha) < 1/(2d-1) \leq 1/5$. Therefore,

$$p\log\frac{1}{q} > \frac{p}{q}H(q) - 2p(1-q). \tag{A.6}$$

Gathering terms in (A.4), (A.5), (A.6), we arrive at the bound

$$D(p \parallel q)\,ds > \left(-15p\log d + \frac{p}{q}H(q) - 2p(1-q)\right)ds > \left(\frac{p}{q}H(q) - 20p\log d\right)ds. \tag{A.7}$$

By plugging in $p = (1+\varepsilon)/d$ and $q = s/(n-s)$, it becomes

$$\begin{aligned} D\left(p \parallel q\right) ds &> (1+\varepsilon)\left((n-s)\, H\left(\frac{s}{n-s}\right) - 20s\log d\right) \\ &> (1+\varepsilon)\left((1-\alpha)\, H(\frac{s}{n}) - \frac{20s}{n}\log d\right) n \end{aligned} \tag{A.8}$$

where the last step is because $s/n \le \alpha$ and $H(s/(n-s)) > H(s/n)$ whenever $1/2 > s/(n-s)$. Now we impose the condition $\alpha \le 0.1\varepsilon(1+\varepsilon)$, which implies that $(1+\varepsilon)(1-\alpha) \ge 1+0.9\varepsilon$. Hence

$$(1+\varepsilon)\left((1-\alpha)\, H\left(\frac{s}{n}\right) - \frac{20s}{n}\log d\right) n \ge \left((1+0.9\varepsilon) H\left(\frac{s}{n}\right) - \frac{20(1+\varepsilon)s}{n}\log d\right) n. \tag{A.9}$$

Altogether, we have that

$$\Pr[\neg\mathcal{A}] \le \sum_{U\subset V, |U|\le \alpha n} \Pr[|E(G[U])| \ge (1+\varepsilon)|U|] \tag{A.10a}$$

$$\le \sum_{s=1}^{\alpha n} \binom{n}{s} 2^{-D\left(\frac{1+\varepsilon}{d} \parallel \frac{s}{n-s}\right) ds} \tag{A.10b}$$

$$\le \sum_{s=1}^{\alpha n} \binom{n}{s} 2^{-\left((1+0.9\varepsilon)H\left(\frac{s}{n}\right) - \frac{20(1+\varepsilon)s}{n}\log d\right) n}, \tag{A.10c}$$

and from the estimate $\binom{n}{s} \le 2^{nH(s/n)}$ it follows that

$$\binom{n}{s} 2^{-\left((1+0.9\varepsilon)H\left(\frac{s}{n}\right) - \frac{20(1+\varepsilon)s}{n} d\log d\right) n} \le 2^{-\left(0.9\varepsilon H\left(\frac{s}{n}\right) - \frac{20(1+\varepsilon)s}{n}\log d\right) n}. \tag{A.11}$$

We want to set α so that $\frac{20(1+\varepsilon)s}{n}\log d < 0.8\varepsilon H\left(\frac{s}{n}\right)$. For this, note that $H\left(\frac{s}{n}\right) > \frac{s}{n}\log\frac{n}{s}$, so

$$\begin{aligned} 0.8\varepsilon H\left(\frac{s}{n}\right) - \frac{20(1+\varepsilon)s}{n}\log d &> 0.8\varepsilon\frac{s}{n}\log\frac{n}{s} - \frac{20(1+\varepsilon)s}{n}\log d \\ &\ge \frac{s}{n}\left(0.8\varepsilon\log\frac{1}{\alpha} - 20(1+\varepsilon)\log d\right) \end{aligned} \tag{A.12}$$

and thus choosing $\alpha = d^{-30(1+\varepsilon)/\varepsilon}$ ensures that (A.12) is positive. Whenever $d \ge 3$, clearly also $d^{-30(1+\varepsilon)/\varepsilon} \le 0.1\varepsilon(1+\varepsilon)$ as required. With this choice of α it follows that

$$\Pr[\neg\mathcal{A}] < \sum_{s=1}^{\alpha n} 2^{-0.1\varepsilon n H\left(\frac{s}{n}\right)}, \tag{A.13}$$

and what is needed to finish the proof is a simple analysis of the growth rate of the terms. The first term is $2^{-0.1\varepsilon n H(1/n)} < 2^{-0.1\varepsilon\log n} = n^{-0.1\varepsilon}$, and $H(s/n + 1/n) > H(s/n) + 1/n$ when say $\frac{s+1}{n} < \frac{1}{3}$, so

$$\Pr[\neg\mathcal{A}] \le \sum_{s=1}^{\alpha n}\left(2^{-0.1\varepsilon}\right)^{nH\left(\frac{s}{n}\right)} < \sum_{s=1}^{\alpha n}\left(2^{-0.1\varepsilon}\right)^{\log n + s - 1} < \frac{n^{-0.1\varepsilon}}{1-2^{-0.1\varepsilon}} = o_n(1), \tag{A.14}$$

which concludes the proof for $\mathbb{G}_{n,d}$.

Next, we turn to the case where $G \sim \mathbb{G}(n, d/n)$. For a subset $U \subseteq V$ of size $s \le \alpha n$, the random variable $E(G[U])$ is a sum of $s(s-1)/2$ Bernoulli random variables with head probability d/n. Therefore, by the Chernoff bound

$$\Pr[|E(G[U])| \ge (1+\varepsilon)|U|] = \Pr[S_{\frac{s(s-1)}{2},\frac{d}{n}} \ge (1+\varepsilon)s] \le 2^{-D\left(\frac{2(1+\varepsilon)}{s-1} \| \frac{d}{n}\right)\frac{s(s-1)}{2}}. \tag{A.15}$$

The KL divergence in the exponent is undefined when $s - 1 \le 2(1+\varepsilon)$, but we can treat it separately: for any subset U' of size s', if $|E(G[U'])| \ge (1+\varepsilon)s'$ then $|E(G[U'])| \ge s'+1$, and

$$\Pr[|E(G[U'])| \ge s'+1] \le \binom{\frac{s'(s'-1)}{2}}{s'+1}\left(\frac{d}{n}\right)^{s'+1} \le \mathrm{O}_d\left(\frac{1}{n^{s'+1}}\right) \tag{A.16}$$

so the sum of the failure probabilities over all U' such that $|U'| - 1 \le 2(1+\varepsilon) < 3d$ is at most $\sum_{s'<3d}\binom{n}{s'}\mathrm{O}_d(1/n^{s'+1}) = \mathrm{O}_d(1/n)$. Henceforth, we assume $s - 1 > 2(1+\varepsilon)$. By definition,

$$D\left(\frac{2(1+\varepsilon)}{s-1} \,\|\, \frac{d}{n}\right) = \frac{2(1+\varepsilon)}{s-1}\log\frac{2n(1+\varepsilon)}{d(s-1)} + \left(1 - \frac{2(1+\varepsilon)}{s-1}\right)\log\frac{1-\frac{d}{n}}{1-\frac{2(1+\varepsilon)}{s-1}} \tag{A.17}$$

where the second term is nonnegative if $d/n \ge 2(1+\varepsilon)/(s-1)$, ensured by letting $\alpha < 2/d$. Now by (A.15), (A.17),

$$\Pr[S_{s(s-1)/2,d/n} \ge (1+\varepsilon)s] \le 2^{-(1+\varepsilon)\left(\log\frac{n}{s-1} + \log\frac{2(1+\varepsilon)}{d}\right)s}. \tag{A.18}$$

Consequently,

$$\Pr[\neg\mathcal{A}] \le \sum_{s=1}^{\alpha n}\binom{n}{s}\Pr[S_{s(s-1)/2,d/n} \ge (1+\varepsilon)s] \tag{A.19a}$$

$$\le \mathrm{O}_d(1/n) + \sum_{s=\lceil 2(1+\varepsilon)+1\rceil}^{\alpha n}\binom{n}{s}2^{-(1+\varepsilon)\left(\log\frac{n}{s} + \log\frac{2(1+\varepsilon)}{d}\right)s}. \tag{A.19b}$$

Now instead of using an entropy estimate, Stirling's approximation $\binom{n}{s} \le (en/s)^s = 2^{s\left(\log(n/s) + \log e\right)}$ suffices for us, whereby

$$\binom{n}{s}2^{-(1+\varepsilon)\left(\log\frac{n}{s} + \log\frac{2(1+\varepsilon)}{d}\right)s} < 2^{-s\left(\varepsilon\log\frac{n}{s} - 2(1+\varepsilon)\log d\right)}. \tag{A.20}$$

We choose α so that for all $s \le \alpha n$, $0.9\varepsilon\log(n/s) > 2(1+\varepsilon)\log d$, and it is clear that $\alpha = d^{-30(1+\varepsilon)/\varepsilon}$ suffices. Then (A.20) is upper bounded by $2^{-0.1\varepsilon s\log(n/s)}$, and we only need to analyze the growth rate of these terms. The first one is bounded by $n^{-0.1\varepsilon}$, and

$$(s+1)\log\frac{n}{s+1} - s\log\frac{n}{s} = \log\left[\frac{n}{s+1}\left(\frac{s}{s+1}\right)^s\right] \ge \log\frac{n}{(s+1)e} \ge 1.$$

Therefore,

$$\Pr[\neg\mathcal{A}] \le \mathrm{O}_d(1/n) + n^{-0.1\varepsilon}\cdot\sum_{s=\lceil 2(1+\varepsilon)+1\rceil}^{\alpha n} 2^{-0.1\varepsilon(s-1)} = \mathrm{o}_n(1) \tag{A.21}$$

and the proof is complete. □

Acknowledgements

The authors would like to thank Gaia Carenini for helpful discussions during the course of this work.

Part of this work was carried out while taking part in the semester programs *Meta-Complexity* and *Satisfiability: Extended Reunion* in the spring of 2023 at the Simons Institute for the Theory of Computing at UC Berkeley.

This work was partially supported by the Wallenberg AI, Autonomous Systems and Software Program (WASP) funded by the Knut and Alice Wallenberg Foundation.

References

[ABRW02] Michael Alekhnovich, Eli Ben-Sasson, Alexander A. Razborov, and Avi Wigderson. Space complexity in propositional calculus. *SIAM Journal on Computing*, 31(4):1184–1211, April 2002. Preliminary version in *STOC '00*.

[AH19] Albert Atserias and Tuomas Hakoniemi. Size-degree trade-offs for Sums-of-Squares and Positivstellensatz proofs. In *Proceedings of the 34th Annual Computational Complexity Conference (CCC '19)*, volume 137 of *Leibniz International Proceedings in Informatics (LIPIcs)*, pages 24:1–24:20, July 2019.

[AN05] Dimitris Achlioptas and Assaf Naor. The two possible values of the chromatic number of a random graph. *Annals of Mathematics*, 162(3):1335–1351, November 2005.

[AO19] Albert Atserias and Joanna Ochremiak. Proof complexity meets algebra. *ACM Transactions on Computational Logic*, 20:1:1–1:46, February 2019. Preliminary version in *ICALP '17*.

[AR03] Michael Alekhnovich and Alexander A. Razborov. Lower bounds for polynomial calculus: Non-binomial case. *Proceedings of the Steklov Institute of Mathematics*, 242:18–35, 2003. Available at http://people.cs.uchicago.edu/~razborov/files/misha.pdf. Preliminary version in *FOCS '01*.

[AT92] Noga Alon and Michael Tarsi. Colorings and orientations of graphs. *Combinatorica*, 12(2):125–134, June 1992.

[Bay82] David Allen Bayer. *The Division Algorithm and the Hilbert Scheme*. PhD thesis, Harvard University, Cambridge, MA, USA, June 1982. Available at https://www.math.columbia.edu/~bayer/papers/Bayer-thesis.pdf.

[BBKO21] Libor Barto, Jakub Bulín, Andrei Krokhin, and Jakub Opršal. Algebraic approach to promise constraint satisfaction. *J. ACM*, 68(4), jul 2021.

[BCMM05] Paul Beame, Joseph C. Culberson, David G. Mitchell, and Cristopher Moore. The resolution complexity of random graph k-colorability. *Discrete Applied Mathematics*, 153(1-3):25–47, December 2005.

[BE05] Richard Beigel and David Eppstein. 3-coloring in time $O(1.3289^n)$. *Journal of Algorithms*, 54(2):168–204, February 2005.

[Ben62] George Bennett. Probability inequalities for the sum of independent random variables. *Journal of the American Statistical Association*, 57(297):33–45, 1962.

[BIK+94] Paul Beame, Russell Impagliazzo, Jan Krajíček, Toniann Pitassi, and Pavel Pudlák. Lower bounds on Hilbert's Nullstellensatz and propositional proofs. In *Proceedings of the 35th Annual IEEE Symposium on Foundations of Computer Science (FOCS '94)*, pages 794–806, November 1994.

[BN21] Samuel R. Buss and Jakob Nordström. Proof complexity and SAT solving. In Armin Biere, Marijn J. H. Heule, Hans van Maaren, and Toby Walsh, editors, *Handbook of Satisfiability*, volume 336 of *Frontiers in Artificial Intelligence and Applications*, chapter 7, pages 233–350. IOS Press, 2nd edition, February 2021.

[Bol78] Béla Bollobás. Chromatic number, girth and maximal degree. *Discrete Mathematics*, 24(3):311–314, 1978.

[Bus98] Samuel R. Buss. Lower bounds on Nullstellensatz proofs via designs. In *Proof Complexity and Feasible Arithmetics*, volume 39 of *DIMACS Series in Discrete Mathematics and Theoretical Computer Science*, pages 59–71. American Mathematical Society, 1998. Available at http://www.math.ucsd.edu/~sbuss/ResearchWeb/designs/.

[CEI96] Matthew Clegg, Jeffery Edmonds, and Russell Impagliazzo. Using the Groebner basis algorithm to find proofs of unsatisfiability. In *Proceedings of the 28th Annual ACM Symposium on Theory of Computing (STOC '96)*, pages 174–183, May 1996.

[CR79] Stephen A. Cook and Robert A. Reckhow. The relative efficiency of propositional proof systems. *Journal of Symbolic Logic*, 44(1):36–50, March 1979. Preliminary version in *STOC '74*.

[DL95] Jesús A. De Loera. Gröbner bases and graph colorings. *Beiträge zur Algebra und Geometrie*, 36(1):89–96, January 1995. Available at https://www.emis.de/journals/BAG/vol.36/no.1/.

[DLMM08] Jesús A. De Loera, Jon Lee, Peter N. Malkin, and Susan Margulies. Hilbert's Nullstellensatz and an algorithm for proving combinatorial infeasibility. In *Proceedings of the 21st International Symposium on Symbolic and Algebraic Computation (ISSAC '08)*, pages 197–206, July 2008.

[DLMM11] Jesús A. De Loera, Jon Lee, Peter N. Malkin, and Susan Margulies. Computing infeasibility certificates for combinatorial problems through Hilbert's Nullstellensatz. *Journal of Symbolic Computation*, 46(11):1260–1283, November 2011.

[DLMO09] Jesús A. De Loera, Jon Lee, Susan Margulies, and Shmuel Onn. Expressing combinatorial problems by systems of polynomial equations and Hilbert's Nullstellensatz. *Combinatorics, Probability and Computing*, 18(4):551–582, July 2009.

[DMP+15] Jesús A. De Loera, Susan Margulies, Michael Pernpeintner, Eric Riedl, David Rolnick, Gwen Spencer, Despina Stasi, and Jon Swenson. Graph-coloring ideals: Nullstellensatz certificates, Gröbner bases for chordal graphs, and hardness of Gröbner bases. In *Proceedings of the 40th International Symposium on Symbolic and Algebraic Computation (ISSAC '15)*, pages 133–140, July 2015.

[Hal93] Magnús M. Halldórsson. A still better performance guarantee for approximate graph coloring. *Information Processing Letters*, 45(1):19–23, January 1993.

[Hus15] Thore Husfeldt. Graph colouring algorithms. In Lowell W. Beineke and Robin J. Wilson, editors, *Topics in Chromatic Graph Theory*, Encyclopedia of Mathematics and its Applications, chapter 13, pages 277–303. Cambridge University Press, May 2015.

[HW08] Christopher J. Hillar and Troels Windfeldt. Algebraic characterization of uniquely vertex colorable graphs. *Journal of Combinatorial Theory, Series B*, 98(2):400–414, March 2008.

[IPS99] Russell Impagliazzo, Pavel Pudlák, and Jiří Sgall. Lower bounds for the polynomial calculus and the Gröbner basis algorithm. *Computational Complexity*, 8(2):127–144, 1999.

[Kar72] Richard M. Karp. Reducibility among combinatorial problems. In *Complexity of Computer Computations*, The IBM Research Symposia Series, pages 85–103. Springer, 1972.

[KM21] Pravesh K. Kothari and Peter Manohar. A stress-free sum-of-squares lower bound for coloring. In *Proceedings of the 36th Annual IEEE Conference on Computational Complexity (CCC '21)*, volume 200 of *Leibniz International Proceedings in Informatics (LIPIcs)*, pages 23:1–23:21, July 2021.

[KMS94] David Karger, Rajeev Motwani, and Madhu Sudan. Approximate graph coloring by semidefinite programming. In *Proceedings 35th Annual IEEE Symposium on Foundations of Computer Science (FOCS '94)*, pages 2–13, November 1994.

[KO22] Andrei Krokhin and Jakub Opršal. An invitation to the promise constraint satisfaction problem. *ACM SIGLOG News*, 9(3):30–59, 2022.

[KPGW10] Graeme Kemkes, Xavier Pérez-Giménez, and Nicholas Wormald. On the chromatic number of random d-regular graphs. *Advances in Mathematics*, 223(1):300–328, January 2010.

[KT17] Ken-Ichi Kawarabayashi and Mikkel Thorup. Coloring 3-colorable graphs with less than $n^{1/5}$ colors. *J. ACM*, 64(1), mar 2017.

[Las01] Jean B. Lasserre. An explicit exact SDP relaxation for nonlinear 0-1 programs. In *Proceedings of the 8th International Conference on Integer Programming and Combinatorial Optimization (IPCO '01)*, volume 2081 of *Lecture Notes in Computer Science*, pages 293–303. Springer, June 2001.

[Lau18] Massimo Lauria. Algorithm analysis through proof complexity. In *Proceedings of the 14th Conference on Computability in Europe (CiE '18), Sailing Routes in the World of Computation*, volume 10936 of *Lecture Notes in Computer Science*, pages 254–263. Springer International Publishing, July 2018.

[LN17] Massimo Lauria and Jakob Nordström. Graph colouring is hard for algorithms based on Hilbert's Nullstellensatz and Gröbner bases. In *Proceedings of the 32nd Annual Computational Complexity Conference (CCC '17)*, volume 79 of *Leibniz International Proceedings in Informatics (LIPIcs)*, pages 2:1–2:20, July 2017.

[Lov94] László Lovász. Stable sets and polynomials. *Discrete Mathematics*, 124(1–3):137–153, January 1994.

[Mat74] Yuri V. Matiyasevich. A criterion for vertex colorability of a graph stated in terms of edge orientations. *Diskretnyi Analiz*, 26:65–71, 1974. English translation of the Russian original. Available at http://logic.pdmi.ras.ru/~yumat/papers/22_paper/.

[Mat04] Yuri V. Matiyasevich. Some algebraic methods for calculating the number of colorings of a graph. *Journal of Mathematical Sciences*, 121(3):2401–2408, May 2004.

[McD84] Colin McDiarmid. Colouring random graphs. *Annals of Operations Research*, 1(3):183–200, October 1984.

[MN15] Mladen Mikša and Jakob Nordström. A generalized method for proving polynomial calculus degree lower bounds. In *Proceedings of the 30th Annual Computational Complexity Conference (CCC '15)*, volume 33 of *Leibniz International Proceedings in Informatics (LIPIcs)*, pages 467–487, June 2015.

[Mnu01] Michal Mnuk. Representing graph properties by polynomial ideals. In *Proceedings of the 4th International Workshop on Computer Algebra in Scientific Computing (CASC '01)*, pages 431–444, September 2001.

[Par00] Pablo A. Parrilo. *Structured Semidefinite Programs and Semialgebraic Geometry Methods in Robustness and Optimization*. PhD thesis, California Institute of Technology, May 2000. Available at http://resolver.caltech.edu/CaltechETD:etd-05062004-055516.

[Raz98] Alexander A. Razborov. Lower bounds for the polynomial calculus. *Computational Complexity*, 7(4):291–324, December 1998.

[Raz17] Alexander A. Razborov. On the width of semialgebraic proofs and algorithms. *Mathematics of Operations Research*, 42(4):1106–1134, May 2017.

[Rec75] Robert A. Reckhow. *On the Lengths of Proofs in the Propositional Calculus*. PhD thesis, University of Toronto, 1975. Available at https://www.cs.toronto.edu/~sacook/homepage/reckhow_thesis.pdf.

[RT22] Julián Ariel Romero Barbosa and Levent Tunçel. Graphs with large girth and chromatic number are hard for Nullstellensatz. Technical Report 2212.05365, arXiv.org, December 2022.

[SA90] Hanif D. Sherali and Warren P. Adams. A hierarchy of relaxations between the continuous and convex hull representations for zero-one programming problems. *SIAM Journal on Discrete Mathematics*, 3:411–430, 1990.

[Wor99] Nicholas Charles Wormald. *Models of random regular graphs*, volume 267 of *London Mathematical Society Lecture Note Series*, pages 239–298. Cambridge University Press, 1999.

[Zuc07] David Zuckerman. Linear degree extractors and the inapproximability of max clique and chromatic number. *Theory of Computing*, 3(6):103–128, August 2007. Preliminary version in *STOC '06*.