Combinatorial anti-concentration inequalities, with applications

Jacob Fox* Matthew Kwan † Lisa Sauermann‡

Abstract

We prove several different anti-concentration inequalities for functions of independent Bernoulli-distributed random variables. First, motivated by a conjecture of Alon, Hefetz, Krivelevich and Tyomkyn, we prove some "Poisson-type" anti-concentration theorems that give bounds of the form $1/e + o(1)$ for the point probabilities of certain polynomials. Second, we prove an anti-concentration inequality for polynomials with nonnegative coefficients which extends the classical Erdős–Littlewood–Offord theorem and improves a theorem of Meka, Nguyen and Vu for polynomials of this type. As an application, we prove some new anti-concentration bounds for subgraph counts in random graphs.

1 Introduction

In probabilistic combinatorics (and probability in general), many arguments are heavily dependent on *concentration* inequalities, which show that certain random variables are likely to lie in a small interval around their mean. For example, if X takes the binomial distribution $\mathrm{Bin}(n,p)$, which has mean $\mu = np$ and variance $\sigma^2 = p(1-p)n$, then typically $X = \mu \pm O(\sigma)$. In the other direction, *anti-concentration* inequalities give *upper* bounds on the probability that a random variable falls into a small interval or is equal to a particular value. The *Lévy concentration function* Q_X of a random variable X is defined by

$$Q_X(t) := \sup_{x\in\mathbb{R}} \Pr(x \le X \le x+t).$$

Returning to the example $X \in \mathrm{Bin}(n,p)$, we can compute $\Pr(X = x) = O(1/\sigma)$ for all $x \in \mathbb{N}$, which implies that $Q_X(t) = O((t+1)/\sigma)$. Bounds of this type can be proved for a variety of different kinds of random variables. See for example [32, 39] for surveys on anti-concentration.

In the above example $X \in \mathrm{Bin}(n,p)$, in the case where p is fixed and n is large, the above concentration and anti-concentration phenomena can both be explained by comparison to a Gaussian distribution. More generally, as an important example generalising the binomial distribution, let $a_1, \dots, a_n$ be a fixed sequence of nonzero real numbers, let $\xi_1, \dots, \xi_n$ be a sequence of i.i.d. p-Bernoulli-distributed random variables (meaning that $\Pr(\xi_i = 1) = p$, $\Pr(\xi_i = 0) = 1-p$) and let $X := a_1\xi_1 + \cdots + a_n\xi_n$. If $1 \le |a_i| = O(1)$ for each i, then one can apply a quantitative central limit theorem to compare X to a Gaussian distribution and show that $\Pr(|X - x| < 1) = O(1/\sqrt{n})$ for any $x \in \mathbb{R}$ (and therefore $Q_X(t) = O((t+1)/\sqrt{n})$). Remarkably, the same result holds even when we require no upper bound on the $|a_i|$, meaning that X may be far from Gaussian and may not even be particularly well-concentrated. This is the content of the Erdős–Littlewood–Offord theorem[1] [13]. A precursor to the Erdős–Littlewood–Offord theorem was first used by Littlewood and Offord [27] in their study of random polynomials more than 50 years ago, and since then, the theorem and its variants have played an important role in probability, especially in random matrix theory (see for example [37, 38]).

Observe that $a_1\xi_1 + \cdots + a_n\xi_n$ is a linear polynomial in the ξ_i, so a natural variation on the Littlewood–Offord problem is to consider polynomials of higher degree. This problem seems to have been first

*Department of Mathematics, Stanford University, Stanford, CA 94305. Email: jacobfox@stanford.edu. Research supported by a Packard Fellowship and by NSF Career Award DMS-1352121.

†Department of Mathematics, Stanford University, Stanford, CA 94305. Email: mattkwan@stanford.edu. Research supported in part by SNSF project 178493.

‡Department of Mathematics, Stanford University, Stanford, CA 94305. Email: lsauerma@stanford.edu.

[1] The Erdős–Littlewood–Offord theorem was most famously stated for the case where $p = 1/2$. This case is somewhat simpler because we can assume that all the a_i are positive: changing the sign of a_i only changes the distribution of X by a translation. However, with modern techniques it is not difficult to deduce a similar estimate for any fixed $p \in (0,1)$; see for example [9, Lemma A.1].

studied by Rosiński and Samorodnitsky [34] in connection with Lévy chaos, but it was later popularised by Costello, Tao and Vu [10] when they used a quadratic variant of the Littlewood–Offord inequality in their proof of Weiss' conjecture that a random symmetric ± 1 matrix typically has full rank. Anti-concentration inequalities for higher-degree polynomials have since found several applications in the theory of Boolean functions (see for example [31, 35]). The current most general result is due to Meka, Nguyen, and Vu [31], and gives a bound in terms of the *rank* of a polynomial, as follows. For a real multilinear degree-d polynomial f in n variables, consider the d-uniform hypergraph on the vertex set $\{1, \dots, n\}$ with a hyperedge $\{i_1, \dots, i_d\}$ if the coefficient of $x_{i_1} \dots x_{i_d}$ in f has absolute value at least 1. Then the rank of f is defined to be the largest matching in this hypergraph. For example, if all $\binom{n}{d}$ degree-d coefficients of f have absolute value at least 1, then f has rank $\lfloor n/d \rfloor = \Omega(n)$. Meka, Nguyen and Vu proved that for fixed $d \in \mathbb{N}$ and $p \in (0, 1)$, any multilinear degree-d rank-r polynomial f in n variables, any $x \in \mathbb{R}$, and $\boldsymbol{\xi} = (\xi_1, \dots, \xi_n) \in \mathrm{Ber}(p)^n$, we have

$$\Pr(|f(\boldsymbol{\xi}) - x| < 1) \le \frac{(\log r)^{O(1)}}{\sqrt{r}}.$$

Up to the polylogarithmic factor, this result is best-possible in multiple regimes. Consider for example the polynomial $x_1 \dots x_d + x_{d+1} \dots x_{2d} + \dots + x_{(r-1)d+1} \dots x_{rd}$, with only linearly many nonzero coefficients, or the polynomial $(x_1 + \dots + x_n)^d$ with $\Theta(n^d)$ nonzero coefficients.

Our first result is that if the coefficients of f are nonnegative, then we can remove the polylogarithmic factor in the Meka–Nguyen–Vu theorem, even with a slightly looser notion of rank. For a multilinear polynomial f in n variables, let $r(f)$ be the largest matching in the (non-uniform) hypergraph on the vertex set $\{1, \dots, n\}$ with a hyperedge $\{i_1, \dots, i_k\}$ if the coefficient of $x_{i_1} \dots x_{i_k}$ in f has absolute value at least 1.

Theorem 1.1. *Fix $d \in \mathbb{N}$ and $p \in (0, 1)$, let f be a degree-d multilinear polynomial in n variables with nonnegative coefficients, and let $\boldsymbol{\xi} \in \mathrm{Ber}(p)^n$. Then for any $x \in \mathbb{R}$, with $r(f)$ as defined above, we have*

$$\Pr(|f(\boldsymbol{\xi}) - x| < 1) \le O\Big(1/\sqrt{r(f)}\Big).$$

We remark that polynomials of Bernoulli random variables with nonnegative coefficients arise naturally in probabilistic combinatorics. An important example is the number of copies of a fixed graph H in a random graph $\mathbb{G}(n, p)$ (we will say more about this in Section 1.3). Actually there is also a rich theory of concentration inequalities for these kinds of polynomials, due primarily to Kim and Vu (see [24] for a survey).

Actually, it seems that for many polynomials that arise in combinatorics, their polynomial structure is less important than the fact that they are strongly monotone: changing some ξ_i from 0 to 1 tends to cause a large increase in the value of $f(\boldsymbol{\xi})$. Our next result extends Theorem 1.1 in this setting.

Theorem 1.2. *Fix $p \in (0, 1)$. Consider a function $f : \{0, 1\}^n \to \mathbb{R}$, let $\boldsymbol{\xi} \in \mathrm{Ber}(p)^n$, and define the random variables*

$$\Delta_i(\boldsymbol{\xi}) := f(\xi_1, \dots, \xi_{i-1}, 1, \xi_{i+1}, \dots, \xi_n) - f(\xi_1, \dots, \xi_{i-1}, 0, \xi_{i+1}, \dots, \xi_n).$$

Suppose for some positive s (which may be a function of n) that $\Pr(\Delta_i(\boldsymbol{\xi}) \le 2s) \le n^{-\omega(1)}$ for all $i \in \{1, \dots, n\}$. Then, for any $x \in \mathbb{R}$,

$$\Pr(|f(\boldsymbol{\xi}) - x| < s) \le \max_t \binom{n}{t} p^t (1 - p)^{n-t} + o\big(1/\sqrt{n}\big) = O\big(1/\sqrt{n}\big).$$

We will prove Theorems 1.1 and 1.2 in Section 2. Both proofs are quite similar (and quite short), and proceed along similar lines to Erdős' original proof of the Erdős–Littlewood–Offord theorem: the events in question are "almost" antichains in the n-dimensional Boolean lattice. We remark that the main term of Theorem 1.2 is best-possible: consider the case $f(\boldsymbol{\xi}) = \xi_1 + \dots + \xi_n$, with any $s < 1/2$.

The above discussion concerns the regime where p is fixed and n is large, in which case we expect anti-concentration behaviour to be "Gaussian-like". However, if p is allowed to be a decaying function of n, then we cannot hope for bounds as strong as $O(1/\sqrt{n})$. Indeed, consider the case $X \in \mathrm{Bin}(n, p)$ with $p = 1/n$. The Poisson limit theorem (see for example [23, p. 64]) shows that X is asymptotically Poisson, which implies that $\Pr(X = x) \le 1/e + o(1)$ for each $x \in \mathbb{N}$. To our knowledge there is not yet a theory of anti-concentration that generalises this fact, though a recent conjecture of Alon, Hefetz, Krivelevich and Tyomkyn [2] hints at the existence and utility of such a theory. We discuss this in the next subsection, and in Section 1.2 we will present some general "Poisson-type" inequalities for certain polynomials.

1.1 Edge-statistics in graphs

For an n-vertex graph G and some $0 \le k \le n$, consider a uniformly random set of k vertices $A \subseteq V(G)$ and define the random variable $X_{G,k} := e(G[A])$ to be the number of edges induced by the random k-set A. Motivated by connections to graph inducibility[2], Alon, Hefetz, Krivelevich and Tyomkyn [2] recently initiated the study of the anti-concentration of $X_{G,k}$, and made the following three conjectures.

Conjecture 1.3 ([2, Conjecture 6.2])**.** *Suppose $k \to \infty$ and $n/k \to \infty$, and consider ℓ satisfying $\ell = \Omega(k^2)$ and $\binom{k}{2} - \ell = \Omega(k^2)$. Then $\Pr(X_{G,k} = \ell) = O\Big(1/\sqrt{k}\Big)$.*

Conjecture 1.4 ([2, Conjecture 6.1])**.** *Suppose $k \to \infty$ and $n/k \to \infty$, and consider ℓ satisfying $\ell = \omega(k)$ and $\binom{k}{2} - \ell = \omega(k)$. Then $\Pr(X_{G,k} = \ell) = o(1)$.*

Conjecture 1.5 ([2, Conjecture 1.1])**.** *Suppose $k \to \infty$ and n grows sufficiently rapidly in terms of k. Then for all $0 < \ell < \binom{k}{2}$ we have $\Pr(X_{G,k} = \ell) \le 1/e + o(1)$.*

There has already been a lot of progress on these conjectures. Kwan, Sudakov and Tran [26] proved Conjecture 1.4 and proved that in the setting of Conjecture 1.3, we have $\Pr(X_{G,k} = \ell) = (\log k)^{O(1)}/\sqrt{k}$. Combining the results of [26] with several new ideas, Conjecture 1.5 was then proved, independently by Fox and Sauermann [19] and by Martinsson, Mousset, Noever and Trujić [29].

Actually, Kwan, Sudakov and Tran's work on Conjectures 1.3 and 1.4 involved an application of the Meka–Nguyen–Vu polynomial anti-concentration inequality mentioned earlier. To illustrate the connection between polynomial anti-concentration and this problem, instead of the random size-k subset $A \subseteq V(G)$, consider the closely related random subset $A^{\mathrm{Ber}} \subseteq V(G)$, where each of the n vertices is included with probability k/n independently. Then, $X_{G,k}^{\mathrm{Ber}} := e\big(G\big[A^{\mathrm{Ber}}\big]\big)$ can be interpreted as a quadratic polynomial of a $\mathrm{Ber}(k/n)^n$-distributed random vector, whose coefficients correspond to edges in the graph.

As our first application of our new anti-concentration theorems, we observe that Theorem 1.2 can be used to prove a stronger "Bernoulli version" of Conjecture 1.3, in the more general setting of hypergraphs.

Proposition 1.6. *Fix $r \in \mathbb{N}$, suppose $k \to \infty$ and $n \ge 2k$, and consider any ℓ satisfying $\ell = \Omega(k^r)$. Then for any r-uniform hypergraph G, we have $\Pr\Big(\Big|X_{G,k}^{\mathrm{Ber}} - \ell\Big| \le k^{r-1}\Big) = O\Big(1/\sqrt{k}\Big)$.*

Note that Proposition 1.6 is best-possible, as can be seen by considering the case where G is a clique. We defer the proof of Proposition 1.6 to Section 4.

While the Littlewood–Offord point of view has been very useful for attacking Conjecture 1.3, the proofs of Conjecture 1.5 in [19, 29] proceeded along rather different lines. Our original motivation for developing Poisson-type analogues to the Littlewood–Offord problem (where p may go to zero) was that they may give a simpler proof of Conjecture 1.5 and faciliate generalisation to hypergraphs (a problem that was also suggested by Alon, Hefetz, Krivelevich and Tyomkyn). While we did not manage to achieve this original goal, we were able to prove several Poisson-type anti-concentration inequalities (stated in the next subsection), one of which (Theorem 1.10) implies the following "Bernoulli version" of Conjecture 1.5. The short deduction can be found in Section 4.

Proposition 1.7. *Suppose $n/k \to \infty$. Then for any $\ell \ne 0$ and any r-uniform hypergraph G, we have $\Pr\Big(X_{G,k}^{\mathrm{Ber}} = \ell\Big) \le 1/e + o(1)$.*

1.2 Poisson-type anti-concentration inequalities for polynomials

Consider first the Littlewood–Offord case where $X = a_1\xi_1 + \cdots + a_n\xi_n$, for some fixed sequence $(a_1, \dots, a_n) \in \mathbb{R}^n$ and a random vector $(\xi_1, \dots, \xi_n) \in \mathrm{Ber}(p)^n$. We want to prove an anti-concentration theorem for the case where p is small. Of course, if p is *extremely* small, then we are very likely to see $\xi_1 = \cdots = \xi_n = 0$, meaning that $\Pr(X = 0) \approx 1$. Discounting this trivial case, we are able to prove the following theorem (in Section 3).

[2]Roughly speaking, the *inducibility* of a graph H measures the maximum number of induced copies of H a large graph can have. This notion was introduced in 1975 by Pippenger and Golumbic [33], and has enjoyed a recent surge of interest; see for example [3, 21, 41, 25].

Theorem 1.8. *Consider a sequence* $(a_1, \dots, a_n) \in \mathbb{R}^n$, *let* $\boldsymbol{\xi} \in \mathrm{Ber}(p)^n$ *and let* $X := a_1\xi_1 + \dots + a_n\xi_n$. *Then for any* $x \neq 0$,

$$\Pr(X = x) \leq \frac{1}{e} + o_{p\to 0}(1).$$

(The notation $o_{p\to 0}(1)$ refers to a function $g(p)$, not depending on n, such that $g(p) \to 0$ as $p \to 0$). We remind the reader that in the case where p does not tend to zero, the Littlewood–Offord theorem gives a bound of $O(1/\sqrt{n})$ on the point probabilities of X.

One might hope to prove that the same conclusion holds whenever X is a polynomial of bounded degree with zero constant coefficient. Unfortunately, this is not true in general: for example, if $X = \sum_{i=1}^n \xi_i - \sum_{i=1}^n \sum_{j=i+1}^n \xi_i\xi_j$, and $p = 1/n$, then $\Pr(X = 1) = 3/(2e) + o(1)$. Nevertheless, we are able to prove that $\Pr(X = x)$ is bounded away from 1 for any $x \neq 0$, as follows.

Proposition 1.9. *Consider an* n*-variable polynomial* f *with degree at most* d, *and let* $\boldsymbol{\xi} \in \mathrm{Ber}(p)^n$ *for some* $p \leq 1/2$. *Then for any* x *not equal to the constant coefficient of* f,

$$\Pr(f(\boldsymbol{\xi}) = x) \leq 1 - 2^{-d}.$$

We prove Proposition 1.9 in Section 3, with the combinatorial Nullstellensatz (see [1]). Next, one way to recover the "$1/e$" behaviour is to consider only polynomials with nonnegative coefficients, as in Theorem 1.1. We also prove the following theorem in Section 3.

Theorem 1.10. *Consider an* n*-variable polynomial* f *with nonnegative coefficients, and consider a random vector* $\boldsymbol{\xi} \in \mathrm{Ber}(p)^n$. *Then for any* x *not equal to the constant coefficient of* f,

$$\Pr(f(\boldsymbol{\xi}) = x) \leq \frac{1}{e} + o_{p\to 0}(1).$$

We emphasise that Theorem 1.10 makes no assumption on the degree of the polynomial f.

1.3 Subgraph counts in random graphs

Fix $p \in (0, 1)$ and let $G \in \mathbb{G}(n, p)$ be a random labelled graph on the vertex set $\{1, \dots, n\}$ where every pair of vertices is included as an edge with probability p independently. This is called the binomial or Erdős–Renyi model of random graphs. For a fixed graph H, let X_H be the number of copies of H in G. The study of X_H and its distribution is a fundamental topic in the theory of random graphs (see for example [7, 22]). It is well-known that for any H with no isolated vertices, X_H satisfies a central limit theorem, but the anti-concentration behaviour of X_H is not as well-understood. In this setting where p is fixed, one can deduce[3] from a quantitative central limit theorem by Barbour, Karoński and Ruciński [4] that $\Pr(X_H = x) \leq O(1/\sqrt{n})$ for all x. As an application of their Littlewood–Offord-type polynomial anti-concentration inequality mentioned earlier in this paper, Meka, Nguyen and Vu proved the stronger bound that $\Pr(X_H = x) \leq n^{o(1)-1}$. This was a consequence of a more general result concerning random graphs of the form G_p, obtained by starting with a fixed graph G and including each edge of G with probability p independently. Specifically, Meka, Nguyen and Vu observed that if G has r edge-disjoint copies of H, and $X_H(G_p)$ is the number of copies of H in G_p, then $X_H(G_p)$ can be interpreted as a rank-r polynomial of independent p-Bernoulli random variables, so $\Pr(X_H(G_p) = x) \leq r^{o(1)-1/2}$ for all $x \in \mathbb{N}$. Since the polynomial corresponding to $X_H(G_p)$ has nonnegative coefficients, one can use Theorem 1.1 in place of the Meka–Nguyen–Vu anti-concentration inequality to improve this as follows.

Corollary 1.11. *Fix* $p \in (0, 1)$ *and let* G *be a graph with* r *edge-disjoint copies of* H. *Then for any* $x \in \mathbb{N}$ *we have*

$$\Pr(X_H(G_p) = x) = O\big(1/\sqrt{r}\big).$$

In particular, in $\mathbb{G}(n, p)$ *we have*

$$\Pr(X_H = x) \leq O(1/n).$$

We believe that in $\mathbb{G}(n, p)$, the above bound is far from optimal.

[3]The central limit theorem of Barbour, Karoński and Ruciński is not stated with a metric that allows one to directly read off an estimate for the distribution function of X_H. But, it is possible to deduce such an estimate with the method of [36, Proposition 1.2.2].

Conjecture 1.12. *Fix $p \in (0,1)$ and fix a graph H with h non-isolated vertices. Let $G \in \mathbb{G}(n,p)$. Then for any $x \in \mathbb{N}$,*

$$\Pr(X_H = x) = O\Big(1/\sqrt{\mathrm{Var}(X_H)}\Big) = O\big(1/n^{h-1}\big).$$

Conjecture 1.12 would imply that $Q_{X_H}(t) = O\big((t+1)n^{1-h}\big)$. If true, this is best-possible; anything stronger would contradict the central limit theorem known to hold for X_H. Although it is not obvious how to prove Conjecture 1.12, we can use Theorem 1.2 to obtain the optimal bound $Q_{X_H}\left(n^{h-2}\right) = O(1/n)$ for anti-concentration at a "coarse" scale.

Theorem 1.13. *Fix $p \in (0,1)$ and fix a graph H with h vertices and at least one edge. Let $G \in \mathbb{G}(n,p)$. Then for any $x \in \mathbb{N}$,*

$$\Pr\big(|X_H - x| \le n^{h-2}\big) = O(1/n).$$

The short deduction of Theorem 1.13 appears in Section 5.

With a bit more effort, one can combine Theorem 1.2 with some inductive arguments to prove an almost-optimal bound in the case where H is a clique.

Theorem 1.14. *Fix $p \in (0,1)$ and $h \in \mathbb{N}$. Then $\Pr(X_{K_h} = x) \le n^{o(1)+1-h}$ for all $x \in \mathbb{N}$.*

The proof of Theorem 1.14 appears in Section 5. We note that, after we had proved Theorem 1.14 and were working on writing this paper, Berkowitz [6] released a preprint proving a *local limit theorem* that gives an asymptotic estimate for the point probabilities of X_{K_h} in terms of the density of a normal distribution (see also [20, 5]). This local limit theorem directly implies Theorem 1.14 and a strengthening of Conjecture 1.12 in the case where H is a clique. However, we still feel that it is worthwhile to include the proof of Theorem 1.14 in this paper: our proof is simpler and more combinatorial, and with some more work the ideas can be generalised to give a comparable bound for a larger class of subgraphs H. In a separate paper [17] we will introduce some additional ideas to generalise Theorem 1.14 to all connected H.

We remark that the number of cliques of each size is determined by the Tutte polynomial of a graph (see for example [11, Theorem 2.4]), so Theorem 1.14 has the following corollary.

Corollary 1.15. *The probability that two independently chosen random graphs from $\mathbb{G}(n,1/2)$ have the same Tutte polynomial is $n^{-\omega(1)}$.*

Corollary 1.15 improves on a bound of $O(1/\log n)$ for this probability due to Loebl, Matoušek and Pangrác [28, Corollary 1.3] (the study of this question was motivated by a conjecture of Bollobás, Pebody and Riordan [8] that almost all graphs are determined by their Tutte polynomial).

1.4 Structure of the paper and outline of the proofs

The rest of the paper is organised as follows. First, in Section 2 we prove Theorems 1.1 and 1.2. The rough idea for both proofs is the same, and is motivated by Erdős' proof of the Littlewood–Offord theorem. We consider a process that flips the bits ξ_i from zero to one in a random order, where we start with $\boldsymbol{\xi}$ being the all-zero vector, and end with $\boldsymbol{\xi}$ being the all-one vector. We show that our random variable $f(\boldsymbol{\xi})$ tends to increase fairly substantially on each flip (for Theorem 1.1, this is where we use the assumption that the coefficients are nonnegative). We deduce that during our process, $f(\boldsymbol{\xi})$ does not tend to spend very long in the vicinity of any given value. This can then be translated into an anti-concentration result.

Next, in Section 3 we prove Theorem 1.8, Proposition 1.9 and Theorem 1.10. First, Proposition 1.9 has a fairly routine proof, using the combinatorial Nullstellensatz. Second, Theorems 1.8 and 1.10 are proved in a unified way, via a careful induction on n.

In the next two sections we give some applications: in Section 4 we prove Propositions 1.6 and 1.7, and in Section 5 we prove Theorems 1.13 and 1.14. These are all quite direct deductions from the theorems proved in Sections 2 and 3, with the exception of Theorem 1.14. Roughly speaking, the idea for the proof of Theorem 1.14 is to fix a vertex v and then decompose the random variable $X_{K_h}(G)$ (counting copies of K_h in $G \in \mathbb{G}(n,p)$) as $X_{K_h} = X_{K_h}(G - v) + X_{K_{h-1}}(G[N_G(v)])$. That is to say, every copy of K_h in G either does not use the vertex v, or it is comprised of the vertex v and a copy of K_{h-1} inside the neighbourhood $N_G(v)$ of v. We then apply Theorem 1.2 to $X_{K_h}(G-v)$, and deal with $X_{K_{h-1}}(G[N_G(v)])$ by induction on h. The main challenge for this approach is that the random variables $X_{K_h}(G-v)$ and $X_{K_{h-1}}(G[N_G(v)])$ are not independent.

Finally, Section 6 contains some concluding remarks, including some open questions and some further miscellaneous results.

1.5 Notation

We use standard asymptotic notation throughout, and all asymptotics are as $n \to \infty$ unless stated otherwise. For functions $f = f(n)$ and $g = g(n)$ we write $f = O(g)$ to mean there is a constant C such that $|f| \le C|g|$, we write $f = \Omega(g)$ to mean there is a constant $c > 0$ such that $f \ge c|g|$ for sufficiently large n, we write $f = \Theta(g)$ to mean that $f = O(g)$ and $f = \Omega(g)$, and we write $f = o(g)$ or $g = \omega(f)$ to mean that $f/g \to 0$ as $n \to \infty$.

We also use standard graph-theoretic notation: $V(G)$ and $E(G)$ are the sets of vertices and (hyper)edges of a (hyper)graph G, and $v(G)$ and $e(G)$ are the cardinalities of these sets. The subgraph of G induced by a vertex subset U is denoted $G[U]$, the neighbourhood of a vertex v in a graph G is denoted $N_G(v)$, and the degree of v is denoted $\deg_G(v) = |N_G(v)|$.

For a zero-one vector $\boldsymbol{x} \in \{0,1\}^n$, we write $|\boldsymbol{x}|$ for the number of entries that are ones. For a real number x, the floor and ceiling functions are denoted $\lfloor x \rfloor = \max\{i \in \mathbb{Z} : i \le x\}$ and $\lceil x \rceil = \min\{i \in \mathbb{Z} : i \ge x\}$. Finally, all logarithms are in base e.

1.6 Concentration inequalities

For the convenience of the reader, in this section we collect some standard concentration inequalities that will be used throughout the paper (since these inequalities are standard, we will refer to them by name and not by their theorem number). First, we will frequently need to use Chernoff bounds for the binomial and hypergeometric distributions. The following bounds can be found in [22, Corollary 2.3 and Theorem 2.10].

Lemma 1.16 (Chernoff bound)**.** *Suppose X has a binomial or hypergeometric distribution, and consider $0 < \varepsilon \le 3/2$. Then*

$$\Pr(|X - \mathbb{E}X| \ge \varepsilon \mathbb{E}X) \le 2\exp\left(-\frac{\varepsilon^2}{3}\mathbb{E}X\right).$$

Second, we will need (a simple consequence of) the Azuma–Hoeffding inequality, as follows. See for example [22, Corollary 2.27].

Lemma 1.17. *Let $X_1, \dots, X_n$ be indepenent random variables, and let $X = f(X_1, \dots, X_n)$ be some function of these random variables. Suppose that if we change the value of some X_i, then the value of X changes by at most c. Then for every $t > 0$, we have*

$$\Pr(|X - \mathbb{E}X| \ge t) \le 2\exp\left(-\frac{t^2}{2nc^2}\right).$$

2 Generalising Littlewood–Offord to nonnegative polynomials

In this section we prove Theorems 1.1 and 1.2.

Proof of Theorem 1.1. Let H be the hypergraph in the definition of $r(f)$, with a hyperedge for each coefficient of f with size at least 1. Let M be a matching of size $r = r(f)$ in this hypergraph, and condition on any outcome of the variables whose indices do not appear in M. We may assume the remaining variables (corresponding to the vertices of M) are $\xi_1, \dots, \xi_N$, with $r \le N \le rd$. Then, $f(\boldsymbol{\xi})$ is a polynomial in $\xi_1, \dots, \xi_N$. Abusing notation, we write $\boldsymbol{\xi} = (\xi_1, \dots, \xi_N)$; we will not need to worry about the specific values of any of the ξ_i we have conditioned on.

Also, it will be more convenient to estimate probabilities of the form $\Pr(|f(\boldsymbol{\xi})-x|<1/2)$ than of the form $\Pr(|f(\boldsymbol{\xi})-x|<1)$. It suffices to show that $\Pr(|f(\boldsymbol{\xi})-x|<1/2)=O(1/\sqrt{r})$, because we can cover the length-2 interval $\{y:|f(\boldsymbol{\xi})-x|<1\}$ with three (open) length-1 intervals. For the rest of the proof we fix some $x\in\mathbb{R}$.

Choose $N_1=pN-o(r)$ and $N_2=pN+o(r)$ so that $\Pr(N_1\le|\boldsymbol{\xi}|\le N_2)=1-o(1/\sqrt{r})$ (such N_1,N_2 exist by the Chernoff bound). Let $\sigma:\{1,\dots,N\}\to\{1,\dots,N\}$ be a uniformly random permutation, and let $\boldsymbol{\xi}^t$ be the length-N zero-one vector with a 1 in positions $\sigma(1),\dots,\sigma(t)$. Let Y be the number of t satisfying $N_1\le t\le N_2$ and $|f(\boldsymbol{\xi}^t)-x|<1/2$.

Recall that $\max_t\binom{N}{t}p^t(1-p)^{N-t}=\Theta(1/\sqrt{N})$ (one can prove this with Stirling's approximation). We can use linearity of expectation to estimate $\mathbb{E}Y$, as follows (recalling that $r=\Theta(n)$).

$$\begin{aligned}\mathbb{E}Y&=\sum_{t=N_1}^{N_2}\frac{|\{\boldsymbol{x}\in\{0,1\}^N:|\boldsymbol{x}|=t,\ |f(\boldsymbol{x})-x|<1/2\}|}{\binom{N}{t}}\\&=\sum_{t=N_1}^{N_2}\frac{\Pr(|f(\boldsymbol{\xi})-x|<1/2\text{ and }|\boldsymbol{\xi}|=t)}{\binom{N}{t}p^t(1-p)^{N-t}}\\&\ge\frac{1}{\max_t\binom{N}{t}p^t(1-p)^{N-t}}\sum_{t=N_1}^{N_2}\Pr(|f(\boldsymbol{\xi})-x|<1/2\text{ and }|\boldsymbol{\xi}|=t)\\&=\Theta\big(\sqrt{N}\big)\Big(\Pr(|f(\boldsymbol{\xi})-x|<1/2)-\Pr(|\boldsymbol{\xi}|<N_1)-\Pr(|\boldsymbol{\xi}|>N_2)\Big)\\&=\Theta\big(\sqrt{r}\big)\Pr(|f(\boldsymbol{\xi})-x|<1/2)-o(1).\end{aligned}\tag{1}$$

We can also estimate $\mathbb{E}Y$ a different way, using the relationship between the $\boldsymbol{\xi}^t$. Let X_t be the number of $e\in M$ such that $e\cap\sigma(\{1,\dots,t\})=|e|-1$ (that is, all but one of the elements of e have been "activated" by time t). If $t=pN+o(r)=(1+o(1))pN$ then the probability that any particular $e\in M$ contributes to X_t is $(1+o(1))|e|p^{|e|-1}(1-p)=\Theta(1)$, so $\mathbb{E}X_t=\Theta(N)$. Also, changing σ by a transposition changes X_t by at most 2, as M is a matching. So, by a McDiarmid-type concentration inequality for random permutations (see for example [30, Section 3.2]), for each $N_1\le t\le N_2$ we have

$$\Pr(X_t<\mathbb{E}X_t/2)=\exp\left(-\Omega\left(\frac{(\mathbb{E}X_t/2)^2}{N\cdot 2^2}\right)\right)=e^{-\Omega(N)}.$$

Now, observe that $f(\boldsymbol{\xi}^t)$ is increasing in t, because f has nonnegative coefficients. For $N_1\le t\le N_2$, let $\mathcal{E}_t$ be the event that $|f(\boldsymbol{\xi}^t)-x|<1/2$, but $f(\boldsymbol{\xi}^s)-x\le-1/2$ for $N_1\le s<t$ (that is, t is the first time that $f(\boldsymbol{\xi}^t)$ enters the desired range). Note that $Y=0$ unless some $\mathcal{E}_t$ occurs.

For any t, condition on a specific outcome of $(\sigma(1),\dots,\sigma(t))$ such that $\mathcal{E}_t$ holds and such that $X_t\ge\mathbb{E}X_t/2=\Theta(N)$. Let U be the set of $i\notin\sigma(\{1,\dots,t\})$ such that there is $e\in M$ with $i\in e$ and $e\setminus\{i\}\subseteq\sigma(\{1,\dots,t\})$. By definition we have $|U|=X_t$. Let $\tau:=\min\{s\in\{t+1,\dots,N\}:\sigma(\tau)\in U\}$ be the first time that we have $\sigma(\tau)\in U$. By the definition of U, some edge $e\in M$ will be "activated" at time τ, so $f(\boldsymbol{\xi}^\tau)\ge f(\boldsymbol{\xi}^t)+1$ and in particular $|f(\boldsymbol{\xi}^\tau)-x|\ge1/2$. Under our conditioning, $\tau-t+1$ is stochastically dominated by the geometric distribution $\mathrm{Geom}(X_t/(N-t))$, which has expected value $(N-t)/X_t=O(1)$. We have proved that $\mathbb{E}[Y\,|\,\mathcal{E}_t\cap\{X_t\ge\mathbb{E}X_t/2\}]=O(1)$.

Recall that we can have $Y>0$ only if some $\mathcal{E}_t$ occurs, and observe that the $\mathcal{E}_t$ are disjoint and that $Y\le N$ with probability 1. So,

$$\begin{aligned}\mathbb{E}Y&\le\sum_{t=N_1}^{N_2}\Pr(\mathcal{E}_t)\mathbb{E}[Y\,|\,\mathcal{E}_t\cap\{X_t\ge\mathbb{E}X_t/2\}]+N\Pr(X_t<\mathbb{E}X_t/2\text{ for some }t)\\&=O(1)\Pr(\mathcal{E}_{N_1}\cup\dots\cup\mathcal{E}_{N_2})+N^2e^{-\Omega(N)}=O(1).\end{aligned}$$

Combining this with (1), the desired result follows. □

Proof of Theorem 1.2. We proceed in almost the same way as in the proof of Theorem 1.1.

Let $\alpha:=\max_t\binom{n}{t}p^t(1-p)^{n-t}$, and observe that $\alpha=\Theta(1/\sqrt{n})$ (this can be proved with Stirling's approximation; see for example [12, Proposition 1]). Let $|\boldsymbol{\xi}|$ be the number of ones in $\boldsymbol{\xi}$, which has a binomial distribution. By the Chernoff bound, we can choose $n_1 = pn - O\big(\sqrt{n\log n}\big)$ and $n_2 = pn + O\big(\sqrt{n\log n}\big)$ such that $\Pr(n_1 \le |\boldsymbol{\xi}| \le n_2) \ge 1 - o(1/\sqrt{n})$. Observe that $\Pr(|\boldsymbol{\xi}| = t) \ge n^{-O(1)}$ for $n_1 \le t \le n_2$ (this can be proved by comparison to the modal probability α or by direct computation using Stirling's inequality; see for example [12, Proposition 1]). Let $\sigma : \{1,\dots,n\} \to \{1,\dots,n\}$ be a uniformly random permutation, and for each $0 \le t \le n$ let $\boldsymbol{\xi}^t$ be the length-n zero-one vector with a one in positions $\sigma(1),\dots,\sigma(t)$, and zeros in the other positions. Let Y be the number of t satisfying $n_1 \le t \le n_2$ and $\big|f\big(\boldsymbol{\xi}^t\big) - x\big| < s$.

The same calculation as in the proof of Theorem 1.1 gives

$$\mathbb{E}Y \ge \alpha^{-1}\Pr(|f(\boldsymbol{\xi}) - x| < s) - o(1), \tag{2}$$

but on the other hand, by the choice of n_1, n_2, for all i we have

$$\begin{aligned}
\Pr\big(\Delta_i\big(\boldsymbol{\xi}^t\big) < 2s \text{ for some } t \in \{n_1,\dots,n_2\}\big) &\le \sum_{t=n_1}^{n_2} \Pr\Big(\Delta_i(\boldsymbol{\xi}) < 2s \,\Big|\, |\boldsymbol{\xi}| = t\Big) \\
&\le \Pr(\Delta_i(\boldsymbol{\xi}) < 2s) \sum_{t=n_1}^{n_2} \frac{1}{\Pr(|\boldsymbol{\xi}| = t)} \\
&= n^{-\omega(1)} O\Big(\sqrt{n\log n}\Big) n^{O(1)} = n^{-\omega(1)}.
\end{aligned}$$

Let $\mathcal{E}$ be the event that $\Delta_i(\boldsymbol{\xi}^r) \ge 2s$ for all i and all $n_1 \le r \le n_2$, so that $\Pr(\overline{\mathcal{E}}) = n \cdot n^{-\omega(1)} = n^{-\omega(1)}$. Note that $\Delta_i(\boldsymbol{\xi}^t) = f(\boldsymbol{\xi}^{t+1}) - f(\boldsymbol{\xi}^t)$ for $i = \sigma(t+1)$. Therefore, if $\mathcal{E}$ holds and $\big|f\big(\boldsymbol{\xi}^t\big) - x\big| < s$ for some t, then $f(\boldsymbol{\xi}^r) - x \ge s$ for all r satisfying $t < r \le n_2$. That is to say, if $\mathcal{E}$ holds then $Y \le 1$. Since Y can never be greater than $n_2 - n_1 + 1 \le n$, it follows that

$$\mathbb{E}Y \le 1 \cdot \Pr(\mathcal{E}) + n\Pr(\overline{\mathcal{E}}) \le 1 + n \cdot n^{-\omega(1)} \le 1 + o(1).$$

Combining this with (2), we obtain $\Pr(|f(\boldsymbol{\xi}) - x| < s) \le (1 + o(1))\alpha = \alpha + o(1/\sqrt{n})$, as desired. □

3 Poisson-type anti-concentration

In this section we prove Theorem 1.8, Proposition 1.9 and Theorem 1.10. First, Proposition 1.9 will be a corollary of the following non-asymptotic bound for anti-concentration of polynomials of unbiased coin flips.

Lemma 3.1. *Consider a multilinear n-variable polynomial f with degree $d \ge 1$, and let $\boldsymbol{\xi} \in \mathrm{Ber}(1/2)^n$. Then for any $\ell \in \mathbb{R}$,*

$$\Pr(f(\boldsymbol{\xi}) = \ell) \le 1 - 2^{-d}.$$

We prove Lemma 3.1 with the combinatorial Nullstellensatz, whose statement is as follows (see [1, Theorem 1.2])

Theorem 3.2. *Let f be an n-variable polynomial over an arbitrary field $\mathbb{F}$, with degree $\sum_{i=1}^n t_i$ (where each t_i is a nonnegative integer). Suppose that the coefficient of $x_1^{t_1}\dots x_n^{t_n}$ is nonzero. If $S_1,\dots,S_n$ are subsets of $\mathbb{F}$ with $|S_i| > t_i$, then there is $\boldsymbol{s} \in S_1 \times \dots \times S_n$ with $f(\boldsymbol{s}) \ne 0$.*

Proof of Lemma 3.1. Suppose without loss of generality that the coefficient of $\xi_1\dots\xi_d$ is nonzero, and condition on any outcomes for $\xi_{d+1},\dots,\xi_n$. Then, $f(\boldsymbol{\xi}) - \ell$ becomes a degree-d polynomial of $\xi_1,\dots,\xi_d$, and the coefficient of $\xi_1\dots\xi_d$ is nonzero. By Theorem 3.2 with $S_i = \{0,1\}$, at least one of the 2^d equally likely outcomes of $(\xi_1,\dots,\xi_d)$ gives $f(\boldsymbol{\xi}) - \ell \ne 0$. We have proved that

$$\Pr(f(\boldsymbol{\xi}) = \ell \,|\, \xi_{d+1},\dots,\xi_n) \le 1 - 2^{-d},$$

and the desired result then follows from the law of total probability. □

Now we prove Proposition 1.9.

Proof of Proposition 1.9. First note that we can assume f is multilinear, because $\xi_i^2 = \xi_i$ for each i. Consider any x not equal to the constant coefficient of f. Let $\boldsymbol{\xi}' \in \mathrm{Ber}(2p)^n$ and $\boldsymbol{\gamma} \in \mathrm{Ber}(1/2)^n$ be independent random vectors, so that $(\xi_1'\gamma_1, \dots, \xi_n'\gamma_n)$ has the same distribution as $\boldsymbol{\xi}$. Note that if we condition on any outcome of $\boldsymbol{\xi}'$ then $f(\xi_1'\gamma_1, \dots, \xi_n'\gamma_n)$ becomes a multilinear polynomial of $\boldsymbol{\gamma}$ whose constant coefficient is the same as the constant coefficient of f. If this polynomial is constant then $\Pr(f(\boldsymbol{\xi}) = x) = 0$, and otherwise Lemma 3.1 gives $\Pr(f(\boldsymbol{\xi}) = x) \le 1 - 2^{-d}$. □

Next, we give a unified proof of Theorems 1.8 and 1.10. For $0 < p < 1$, define

$$\tau(p) := \sup_{n \in \mathbb{N}} \Pr(X_{n,p} = 1) = \sup_{n \in \mathbb{N}} np(1-p)^{n-1},$$

where $X_{n,p} \in \mathrm{Bin}(n,p)$. It is a straightforward computation to determine the limiting behaviour of $\tau(p)$ as $p \to 0$, as follows.

Lemma 3.3. *We have $\tau(p) \le 1/e + o_{p \to 0}(1)$.*

Proof. For $0 < p < 1$, define $\eta_p : [0,\infty) \to \mathbb{R}$ by $\eta_p(n) := np(1-p)^{n-1}$. We compute

$$\eta_p'(n) = p(1-p)^{n-1}(1 + n\log(1-p)),$$

so $\eta_p'(n) = 0$ only when $n = -1/\log(1-p)$. Since $\eta_p(0) = 0$ and $\eta_p(n) \to 0$ as $n \to \infty$, we have

$$\tau(p) \le \sup_{n \in [0,\infty)} \eta_p(n) = \eta_p(-1/\log(1-p)) = \frac{-p}{e(1-p)\log(1-p)}.$$

This converges to $1/e$ as $p \to 0$, by L'Hôpital's rule. □

The following lemma then implies Theorems 1.8 and 1.10.

Lemma 3.4. *Let f be an n-variable polynomial, with zero constant coefficient, which is either of degree 1 or has all coefficients nonnegative. Consider any $p \in (0,1)$ and $\ell \ne 0$. Then*

$$\Pr(f(\boldsymbol{\xi}) = \ell) \le \tau(p),$$

where $\boldsymbol{\xi} \in \mathrm{Ber}(p)^n$.

Proof. We prove this by induction on n (it is trivially true for $n = 0$). So, consider some $n > 0$, and assume that the statement is true for the case $n-1$. As in the proof of Lemma 3.1 we can assume that f is multilinear.

If $a_1 = \dots = a_n = \ell$ then the only way we can have $f(\boldsymbol{\xi}) = \ell$ is if exactly one of the ξ_i is equal to one. So, in this case $\Pr(f(\boldsymbol{\xi}) = \ell) = \Pr(X_{n,p} = 1) \le \tau(p)$.

Otherwise, there must be some $a_j \ne \ell$. Suppose without loss of generality that $a_n \ne \ell$, so $\ell - a_n \ne 0$. Define $(n-1)$-variable polynomials g and h by $f(x_1, \dots, x_n) = g(x_1, \dots, x_{n-1}) + a_n x_n + h(x_1, \dots, x_{n-1})x_n$, and observe that g and h both have zero constant coefficient. Let $\boldsymbol{\xi}' = (\xi_1, \dots, \xi_{n-1})$. Applying the induction hypothesis to g and $g + h$ gives

$$\begin{aligned}\Pr(f(\boldsymbol{\xi}) = \ell) &= \Pr(\xi_n = 0)\Pr(f(\boldsymbol{\xi}) = \ell \,|\, \xi_n = 0) + \Pr(\xi_n = 1)\Pr(f(\boldsymbol{\xi}) = \ell \,|\, \xi_n = 1)\\ &= (1-p)\Pr\big(g(\boldsymbol{\xi}') = \ell\big) + p\Pr\big(g(\boldsymbol{\xi}') + h(\boldsymbol{\xi}') = \ell - a_n\big)\\ &\le (1-p)\tau(p) + p\tau(p)\\ &\le \tau(p).\end{aligned}$$

□

4 Anti-concentration of the edge-statistic

In this section we give the short proofs of Propositions 1.6 and 1.7. First, note that Proposition 1.7 is an immediate consequence of Theorem 1.10.

Proof of Proposition 1.7. Let $\boldsymbol{\xi} \in \mathrm{Ber}(k/n)^n$, and let $f(\boldsymbol{\xi}) := \sum_{e\in E(G)} \prod_{i\in e} \xi_i$. This polynomial has nonnegative coefficients. Then, $X^{\mathrm{Ber}}_{G,k}$ has the same distribution as $f(\boldsymbol{\xi})$, and since we are assuming $n/k \to \infty$ we can apply Theorem 1.10. □

Next, it is almost as simple to deduce Proposition 1.6 from Theorem 1.2. We need the following well-known lemma, which may be proved by induction, iteratively deleting any vertices with degree less than d/r (see for example [18, Lemma 2.5]).

Lemma 4.1. *Let G be an r-uniform hypergraph with average degree d. Then G has an induced subgraph with minimum degree at least d/r.*

Proof of Proposition 1.6. We can obtain our random subset A^{Ber} by first sampling a random subset A', where each element is present with probability $2k/n$, then deleting each element from A' with probability $1/2$. By the Chernoff bound, with probability $1 - e^{-\Omega(k)}$ we have $(3/2)k \le |A'| \le 3k$; consider such an outcome of A', and let $G' := G[A']$. Let X be a random variable having the same distribution as X conditioned on this outcome of A'. It now suffices to show that $\Pr\big(|X-\ell| \le k^{r-1}\big) = O\Big(1/\sqrt{k}\Big)$.

Now, note that $\mathbb{E}X = e(G')/4$, and observe that deleting an element of A^{Ber} can change X by at most $|A'|^{r-1}$. If $e(G') \le \ell$ then $|\ell - \mathbb{E}X| \ge 3\ell/4 \ge \ell/2 + k^{r-1}$ and therefore by the Azuma–Hoeffding inequality we have

$$\Pr\big(|X-\ell| \le k^{r-1}\big) \le \Pr(|X-\mathbb{E}X| \ge \ell/2) \le \exp\left(-\Omega\left(\frac{(\ell/2)^2}{|A'|^{2(r-1)}\cdot |A'|}\right)\right) = e^{-\Omega(k)}.$$

So, we may assume that $e(G') \ge \ell = \Omega(k^r)$. By Lemma 4.1, this implies that G' has an induced subgraph $G'[B]$ with minimum degree $\Omega\big(k^{r-1}\big)$. Condition on any outcome of $A^{\mathrm{Ber}} \setminus B$, and for $i \in B$ let $\xi_i := \mathbb{1}_{i\in A^{\mathrm{Ber}}}$, so that $X_{G,k}$ can be viewed as a function of $(\xi_i)_{i\in B}$. Recalling the definition of Δ_i from Theorem 1.2 we have $\Delta_i = \big|N_G(i) \cap A^{\mathrm{Ber}}\big| \ge \big|N_G(i) \cap A^{\mathrm{Ber}} \cap B\big|$. By our minimum degree assumption, $\big|N_G(i) \cap A^{\mathrm{Ber}} \cap B\big|$ has a binomial distribution with mean $\Omega(k^{r-1})$, so by the Chernoff bound, with probability $1 - e^{-\Omega(k^{r-1})}$ each $\Delta_i = \Omega\big(k^{r-1}\big)$, which allows us to apply Theorem 1.2. (This gives us a bound for the probability that X falls in an interval of length $\Omega(k^{r-1})$, which suffices because we can cover any interval of length $2k^{r-1}$ with $O(1)$ such intervals). □

5 Anti-concentration for subgraph counts in random graphs

First we give the simple deduction of Theorem 1.13 from Theorem 1.2.

Proof of Theorem 1.13. For this proof it is convenient to redefine X_H to count *labelled* copies of H (this changes the anti-concentration behaviour by a constant factor depending on the number of automorphisms of H). For a pair of distinct vertices $x, y \in V(G)$, define $\Delta_{x,y}$ to be the difference

$$X_H(G + \{x,y\}) - X_H(G - \{x,y\})$$

(that is, the number of copies of H that would be created or destroyed by flipping the status of $\{x,y\}$). Observe that

$$\mathbb{E}\Delta_{x,y} = 2e(H)p^{e(H)-1}n(n-1)\dots(n-h+1) = \Omega\big(n^{h-2}\big), \tag{3}$$

Also, observe that for any vertex other than x or y, changing the set of edges adjacent to that vertex can affect $\Delta_{x,y}$ by at most $O(n^{h-3})$. So, by the Azuma–Hoeffding inequality (with the vertex exposure martingale) it follows that

$$\Pr(\Delta_{x,y} \le \mathbb{E}\Delta_{x,y}/2) = \exp\left(-\Omega\left(\frac{n^{2(h-2)}}{n\cdot n^{2(h-3)}}\right)\right) = n^{-\omega(1)}. \tag{4}$$

Let D be the common value of the $\mathbb{E}\Delta_{x,y}$. We now apply Theorem 1.2 with $\binom{n}{2} = \Theta(n^2)$ variables and with $s = D/4 = \Omega(n^{h-2})$, and observe that the interval $\{r \in \mathbb{R} : |r - x| \le n^{h-2}\}$ (having length $O(n^{h-2})$) can be covered by $O(1)$ intervals of length $2s$. □

Next, we turn to Theorem 1.14. First we illustrate the high-level strategy of the proof, which is by induction on h. Recall that $X_{K_{h-1}} = X_{K_{h-1}}(G)$ is the number of copies of K_h in $G \in \mathbb{G}(n,p)$. Now, let X be the number of copies of K_h in $G \in \mathbb{G}(n,p)$ which contain some fixed vertex v (this is equal to the number of copies of K_{h-1} in $G[N_G(v)]$). Then, we have the decomposition $X_{K_h} = X_{K_h}(G-v) + X$, where $X_{K_h}(G-v)$ is typically much larger than X. One may then hope to establish anti-concentration of X_{K_h} by first showing that $X_{K_h}(G-v)$ is anti-concentrated at a "coarse" scale as in Theorem 1.13, then establishing anti-concentration of X on a finer scale. For this second step, we may first observe that the approximate value of X is primarily driven by $|N_G(v)|$ (which has a binomial distribution and is thus easy to study), and if we condition on $N_G(v)$ then X is the number of copies of K_{h-1} in a fixed vertex subset of $G-v$, which we may study with the induction hypothesis.

The main complication with this approach is that it does not suffice to analyse $X_{K_h}(G-v)$ and X separately, because in principle they could correlate with each other in a way that increases the concentration probabilities. So, we must analyse the concentration behaviour of X conditioned on an outcome of $G-v$. Our approach is to show that $G-v$ is very likely to have certain properties that ensure that, conditioned on this outcome of $G-v$, X has approximately the concentration behaviour we would expect unconditionally. For this, we will need something a bit stronger than Theorem 1.14 as our induction hypothesis, as follows.

Definition 5.1. For real numbers $c \in (0,1/2)$ and $q \in (0,1)$, we say an n-vertex graph G is (c,q,h)-*dispersed* if for all $cn \le k \le (1-c)n$ and all ℓ, the number of induced subgraphs of G with k vertices and exactly ℓ copies of K_h is at most $\binom{n}{k}q$.

Theorem 5.2. *For any constants $c \in (0,1/2)$, $p \in (0,1)$ and $h \in \mathbb{N}$, there are functions $\alpha = \alpha_{h,p,c}$ and $\phi = \phi_{h,p,c}$, with $\lim_{n\to\infty} \alpha(n) = 0$ and $\lim_{n\to\infty} \phi(n) = \infty$, such that the random graph $G \in \mathbb{G}(n,p)$ is $\big(c, n^{\alpha+1-h}, h\big)$-dispersed with probability at least $1 - n^{-\phi}$.*

To see that Theorem 5.2 implies Theorem 1.14, observe that we can obtain a random graph $G \in \mathbb{G}(n,p)$ by first taking a random graph $G' \in \mathbb{G}(2n,p)$, and then taking a random subset of n vertices of G'. Theorem 5.2 tells us that G' is very likely to be $\big(1/3, n^{o(1)+1-h}, h)\big)$-dispersed, and if it is, then the definition of being dispersed gives the required bound on the point probabilities of X_{K_h}.

Proof of Theorem 5.2. As in the proof of Theorem 1.13, we count labelled cliques, which affects X_H by a constant factor $h!$. The proof is by induction on h (the case $h = 1$ is trivial). Fix $\varepsilon > 0$ and $t \in \mathbb{N}$. We will prove that with probability at least $1 - n^{o(1)+1+h-\varepsilon t}$, $G \in \mathbb{G}(n,p)$ is $\big(c, n^{\varepsilon+1-h}, h\big)$-dispersed (asymptotics are allowed to depend on t and ε, which we view as fixed constants for most of the proof). After we have proved this, we can then let $\varepsilon \to 0$ and $t\varepsilon \to \infty$.

For $cn \le r \le (1-c)n$ and $0 \le \ell \le \binom{n}{h}$, let $Z_{r,\ell}$ be the number of sets of r vertices in G that induce exactly ℓ copies of K_h. We need to show that with probability $1 - n^{o(1)+1+h-\varepsilon t}$ we have $Z_{r,\ell} \le n^{\varepsilon+1-h}\binom{n}{r}$ for all $cn \le r \le (1-c)n$ and all $0 \le \ell \le \binom{n}{h}$. We upper-bound $\mathbb{E}Z_{r,\ell}^t$. Note that if we randomly choose a sequence $S_1, \dots, S_t$ of r-vertex sets (with replacement), then with probability $1 - n^{-\omega(1)}$ we have

$$\begin{aligned} |S_1 \cup \cdots \cup S_{i-1}| &\le 1 - (c/2)^{i-1}n, \\ |S_i \setminus (S_1 \cup \cdots \cup S_{i-1})| &\ge (c/2)^i n = \Omega(n) \end{aligned} \tag{5}$$

for each $i \in \{1, \dots, t\}$. This can be proved by repeatedly applying a Chernoff bound for the hypergeometric distribution.

Let $\mathcal{S}$ be the collection of sequences $(S_1, \dots, S_t)$ which satisfy (5) for each $i \in \{1, \dots, t\}$. For any $(S_1, \dots, S_t) \in \mathcal{S}$, with $G_i := G[S_i]$, we wish to prove that

$$\Pr(X_{K_h}(G_i) = \ell \text{ for each } i) \le n^{o(1)+(1-h)t}. \tag{6}$$

It will follow from (6) that

$$\begin{aligned} \mathbb{E}Z_{r,\ell}^t &= \sum_{(S_1,\dots,S_t)\in\mathcal{S}} \Pr(X_{K_h}(G[S_i]) = \ell \text{ for each } i) + \sum_{(S_1,\dots,S_t)\notin\mathcal{S}} \Pr(X_{K_h}(G[S_i]) = \ell \text{ for each } i) \\ &\le \binom{n}{r}^t n^{o(1)+t(1-h)} + \binom{n}{r}^t n^{-\omega(1)} \end{aligned}$$

$$= n^{o(1)+t(1-h)}\binom{n}{r}^t,$$

so

$$\Pr\left(Z_{r,\ell} \ge n^{\varepsilon+1-h}\binom{n}{r}\right) = \Pr\left(Z_{r,\ell}^t \ge n^{t(\varepsilon+1-h)}\binom{n}{r}^t\right) \le \frac{\mathbb{E}Z_{r,\ell}^t}{n^{t(\varepsilon+1-h)}\binom{n}{r}^t} = n^{o(1)-t\varepsilon}.$$

We can then take a union bound over all the (at most $n\binom{n}{h} \le n^{h+1}$) possibilities for r, ℓ.

So, it suffices to prove (6). For the rest of the proof we fix a sequence $(S_1,\dots,S_t) \in \mathcal{S}$. The t events $X_{K_h}(G_i) = \ell$ are not independent, but by the choice of $S_1,\dots,S_t$, for each i there is still a lot of randomness in G_i after exposing outcomes of $G_1,\dots,G_{i-1}$. The plan is to show that for each i, if we condition on an outcome of $G_i^\cap := G[S_i \cap (S_1 \cup \dots \cup S_{i-1})]$, then unless $G_i^\cap$ has some atypical properties, there is still enough randomness to guarantee $\Pr(X_{K_h}(G_i) = \ell) \le n^{o(1)+(1-h)}$.

For each i fix some $v_i \in S_i \setminus (S_1 \cup \dots \cup S_{i-1})$ (which is possible by (5)), let $N_i = N_G(v_i) \cap S_i$, and define

$$X_i = X_{K_h}(G_i) - X_{K_h}(G_i - v_i) = X_{K_{h-1}}(G[N_i])$$

to be the number of copies of K_h in G_i which contain v_i. Also, let $n' = r - 1 = \Omega(n)$ be the common size of the sets $S_i \setminus \{v_i\}$, let $c' = \min\{p, 1-p\}/2$, let $I = \{k \in \mathbb{N} : c'n' \le k \le (1-c')n'\}$ and let $E_k = \mathbb{E}\big[X_i \,\big|\, |N_i| = k\big] = p^{\binom{h-1}{2}}\binom{k}{h-1}$. Let $\beta = \alpha_{h-1,p,c'}(n') = o(1)$ and $\psi = (\log n)^{1/2} = \omega(1)$, recalling the notation in the statement of Theorem 5.2. We say that an outcome of $G_i - v_i$ is *good* if

1. it is $\big(c', (n')^{\beta+2-h}, h-1\big)$-dispersed;
2. for each $k \in I$, at most $\binom{n'}{k}n^{-\psi}$ size-k subsets $S \subseteq S_i \setminus \{v_i\}$ fail to satisfy

$$\big|X_{K_{h-1}}(G_i[S]) - E_k\big| \le n^{h-2}\log n.$$

Then, for $\{x,y\} \subseteq S_i \setminus \{v_i\}$, define $\Delta_{x,y}^{(i)}$ to be the difference

$$X_{K_h}((G_i - v_i) + \{x,y\}) - X_{K_h}((G_i - v_i) - \{x,y\})$$

(that is, the number of copies of K_h in $G_i - v_i$ that would be created or destroyed by flipping the status of $\{x,y\}$). Note that each $\mathbb{E}\Delta_{x,y}^{(i)}$ is equal to some common value $D = \Theta((n')^{h-2}) = \Theta(n^{h-2})$, with essentially the same calculation as in (3). Fix some $\chi = \omega(1)$ that grows sufficiently slowly to satisfy certain inequalities we will encounter later in the proof. Say that an outcome G^* of $G_i^\cap$ is *good-inducing* if

$$\Pr\Big(G_i - v_i \text{ is good, } \Delta_{x,y}^{(i)} \ge D/2 \text{ for all } x,y \in S_i \setminus \{v_i\} \,\Big|\, G_i^\cap = G^*\Big) \ge 1 - n^{-\chi}.$$

We now break the remainder of the proof into a sequence of claims. First, we need to show that it is very likely that each $G_i^\cap$ is good-inducing (here we specify χ).

Claim 5.3. *There is $\chi = \omega(1)$ such that $G_i^\cap$ is good-inducing for each $i \in \{1,\dots,t\}$, with probability $1 - n^{-\omega(1)}$.*

We defer the proof of Claim 5.3 until later. It will be a fairly straightforward consequence of the induction hypothesis and a concentration inequality. Next, recalling (5), note that after exposing $G_i^\cap$ there are still $\Omega(n^2)$ edges of G_i left unexposed. So, we can apply Theorem 1.2 with $s = D/4$ (and $O(m)$ different values of x) to prove the following claim, establishing anti-concentration of X_{K_h} at a "coarse" scale.

Claim 5.4. *For any $x \in \mathbb{R}$, any real $m \ge 1$ and any good-inducing outcome G^* of $G_i^\cap$, we have*

$$\Pr\big(|X_{K_h}(G_i - v_i) - x| < mn^{h-2} \,\big|\, G_i^\cap = G^*\big) = O\Big(\frac{m}{n}\Big).$$

Then, note that if $G_i - v_i$ is good, typically $X_i = X_{K_{h-1}}(G[N_i])$ is approximately equal to $E_{|N_i|}$ (specifically, this follows from the second property of being good). So, the following claim establishes anti-concentration of X_i.

Claim 5.5. *For any $x \in \mathbb{R}$, we have*

$$\Pr\big(\big|E_{|N_i|} - x\big| \le n^{h-2}\log n\big) = O\bigg(\frac{\log n}{\sqrt{n}}\bigg).$$

Further, we have

$$\Pr\Big(\big|E_{|N_i|} - \mathbb{E}X_i\big| > n^{h-3/2}\log n\Big) = n^{-\omega(1)}.$$

We defer the proof of Claim 5.5 until later. The proof is fairly simple, since $|N_i|$ is binomially distributed, and we have an explicit formula for E_k. Next, recalling that X_i is the number of copies of K_{h-1} in $G[N_i]$, the following claim is a direct consequence of the first property of being good (that $G_i - v_i$ is $\big(c', (n')^{\beta+2-h}, h-1\big)$-dispersed). Indeed, after conditioning on the event that $|N_i| = k$, note that N_i is a uniformly random k-vertex subset of $G_i - v_i$.

Claim 5.6. *For any $x \in \mathbb{R}$, any $k \in I$, and any good outcome G' of $G_i - v_i$, we have*

$$\Pr(X_i = x \,|\, G_i - v_i = G',\ |N_i| = k) \le n^{\beta+2-h} = n^{o(1)+2-h}.$$

Finally, the following claim follows directly from the Chernoff bound, since $|N_i|$ has a binomial distribution with parameters $n' = \Omega(n)$ and p.

Claim 5.7. *For each i,*

$$\Pr(|N_i| \notin I) = n^{-\omega(1)}$$

Before proving Claims 5.3 and 5.5, we show how the above claims can be used to deduce (6). Let T_x be the set of all $k \in \mathbb{N}$ such that $|E_k - x| \le n^{h-2}\log n$. Then, $\Pr(|N_i| \in T_x) = O(\log n/\sqrt{n})$ by the first part of Claim 5.5. Next, consider any good outcome G' of $G_i - v_i$. If $|N_i| \notin T_x$ then in order to have $X_i = x$ we must have $|X_i - E_{|N_i|}| > n^{h-2}\log n$. So, by the second property of being good, we have

$$\Pr(X_i = x \,|\, G_i - v_i = G',\ |N_i| \in I \setminus T_x) \le n^{-\psi} = n^{-\omega(1)}.$$

By Claims 5.6 and 5.7, it follows that for any $x \in \mathbb{R}$ we have

$$\begin{aligned}
&\Pr(X_i = x \,|\, G_i - v_i = G') \\
&\quad\le \sum_{k \in I \cap T_x} \Pr(X_i = x \,|\, G_i - v_i = G',\ |N_i| = k)\cdot \Pr(|N_i| = k) \\
&\qquad\quad + \Pr(X_i = x \,|\, G_i - v_i = G',\ |N_i| \in I \setminus T_x) + \Pr(|N_i| \notin I) \\
&\quad\le \Pr(|N_i| \in T_x) n^{o(1)+2-h} + n^{-\omega(1)} + n^{-\omega(1)} = n^{o(1)+3/2-h}.
\end{aligned} \tag{7}$$

For $i \in \{1, \dots, t\}$, let $\mathcal{F}_i$ be the event that $|X_{K_h}(G_i - v_i) + \mathbb{E}X_i - \ell| \le 2n^{h-3/2}\log n$ (note that $\mathbb{E}X_i$ is an unconditional expectation and $\mathcal{F}_i$ only depends on $G_i - v_i$). By Claim 5.4 with $x = \ell - \mathbb{E}X_i$ and $m = 2\sqrt{n}\log n$, for any good-inducing outcome G^* of $G_i^\cap$ we have $\Pr(\mathcal{F}_i \,|\, G_i^\cap = G^*) = O(\log n/\sqrt{n})$. Also, by the second part of Claim 5.5 and the second property of being good, for any good outcome G' of $G_i - v_i$ (not satisfying $\mathcal{F}_i$) we have

$$\Pr\big(X_i = \ell - X_{K_h}(G_i - v_i) \,\big|\, \overline{\mathcal{F}_i},\ G_i - v_i = G'\big) = n^{-\omega(1)}.$$

Using (7), for any good-inducing outcome G^* of $G_i^\cap$ we then have

$$\begin{aligned}
&\Pr(X_{K_h}(G_i) = \ell \,|\, G_i^\cap = G^*) \\
&\quad= \Pr(X_{K_h}(G_i - v_i) + X_i = \ell \,|\, G_i^\cap = G^*) \\
&\quad\le \Pr(\mathcal{F}_i \,|\, G_i^\cap = G^*)\Pr(X_i = \ell - X_{K_h}(G_i - v_i) \,|\, \mathcal{F}_i,\ G_i - v_i \text{ is good},\ G_i^\cap = G^*) \\
&\qquad\quad + \Pr\big(X_i = \ell - X_{K_h}(G_i - v_i) \,\big|\, \overline{\mathcal{F}_i},\ G_i - v_i \text{ is good},\ G_i^\cap = G^*\big) \\
&\qquad\quad + \Pr(G_i - v_i \text{ is not good} \,|\, G_i^\cap = G^*) \\
&\quad\le O\bigg(\frac{\log n}{\sqrt{n}}\bigg) n^{o(1)+3/2-h} + n^{-\omega(1)} + n^{-\omega(1)} = n^{o(1)+1-h}.
\end{aligned} \tag{8}$$

Now, let $\mathcal{H}_i$ be the event that $X_{K_h}(G_i) = \ell$ and that $G_{i+1}^{\cap}$ is good-inducing (if $i = t$ this is just the event that $X_{K_h}(G_t) = \ell$). Observe that $G_1^{\cap} = \emptyset$ is not actually random, so Claim 5.3 implies that it is always good-inducing. Applying (8) we have

$$\Pr(\mathcal{H}_i \mid \mathcal{H}_1, \dots, \mathcal{H}_{i-1}) \le \Pr(X_{K_h}(G_i) = \ell \mid \mathcal{H}_1, \dots, \mathcal{H}_{i-1}) \le n^{o(1)+1-h},$$

so

$$\Pr(\mathcal{H}_1 \cap \cdots \cap \mathcal{H}_t) = \prod_{i=1}^{t} \Pr(\mathcal{H}_i \mid \mathcal{H}_1, \dots, \mathcal{H}_{i-1}) \le n^{o(1)+(1-h)t}.$$

Finally, by Claim 5.3, we have

$$\begin{aligned}\Pr(X_{K_h}(G_i) = \ell \text{ for each } i) &\le \Pr(\mathcal{H}_1 \cap \cdots \cap \mathcal{H}_t) + \Pr(\text{some } G_i^{\cap} \text{ is not good-inducing}) \\ &\le n^{o(1)+(1-h)t} + n^{-\omega(1)} = n^{o(1)+(1-h)t},\end{aligned}$$

concluding the proof of (6). Pending the proofs of Claims 5.3 and 5.5, which are given below, this completes the proof of Theorem 5.2. □

Claim 5.3 will be a consequence of the law of total expectation and the following claim.

Claim 5.8. *Let $\mathcal{A}_i$ be the event that $G_i - v_i$ fails to satisfy the first property of being good, let $\mathcal{B}_i$ be the event that it fails to satisfy the second property of being good, and let $\mathcal{C}_i$ be the event that $\Delta_{x,y}^{(i)} < D/2 = \mathbb{E}\Delta_{x,y}^{(i)}/2$ for some $x, y \in S_i \setminus \{v_i\}$. Then, for each $i \in \{1, \dots, t\}$, we have $\Pr(\mathcal{A}_i \cup \mathcal{B}_i \cup \mathcal{C}_i) = n^{-\omega(1)}$.*

Proof. First, we have $\Pr(\mathcal{A}_i) \le (n')^{-\phi_{h-1,p,c'}(n')} = n^{-\omega(1)}$, by Theorem 5.2 for $h-1$ (which we are assuming as our induction hypothesis).

Second, we have $\Pr(\mathcal{C}_i) = e^{-\Omega(n')} = n^{-\omega(1)}$ with exactly the same argument as in (4) in the proof of Theorem 1.13 (using the Azuma–Hoeffding inequality) and the union bound.

Third, we consider $\mathcal{B}_i$. For each $k \in I$ and each subset $S \subseteq S_i - v_i$ of size k, consider the random variable $X_{K_{h-1}}(G_i[S])$. This random variable has mean $E_k = \Omega(n^{h-1})$ and flipping the status of an edge causes a change of at most $O(n^{h-3})$. So, by the Azuma–Hoeffding inequality we have

$$\Pr\big(\big|X_{K_{h-1}}(G_i[S]) - E_k\big| > n^{h-2}\log n\big) = \exp\left(-\Omega\left(\frac{\big(n^{h-2}\log n\big)^2}{n^2 \cdot n^{2(h-3)}}\right)\right) = e^{-\Omega\left((\log n)^2\right)}.$$

Hence, the expected number of subsets S for which $\big|X_{K_{h-1}}(G_i[S]) - E_k\big| > n^{h-2}\log n$ is $\binom{n'}{k}e^{-\Omega\left((\log n)^2\right)}$, and by Markov's inequality, the probability that this occurs for more than $\binom{n'}{k}e^{-(\log n)^{3/2}} = \binom{n'}{k}n^{-\psi}$ subsets is at most $e^{-\Omega\left((\log n)^2\right)} = n^{-\omega(1)}$. We can then take the union bound over all $k \in I \subseteq \{1, \dots, n\}$ to obtain $\Pr(\mathcal{B}_i) \le n^{-\omega(1)}$. □

Now we prove Claim 5.3.

Proof of Claim 5.3. Fix some i; we will show that $G_i^{\cap}$ is good-inducing with probability $n^{-\omega(1)}$. We can then take the union bound over all i.

Let W be the random variable $\Pr(\mathcal{A}_i \cup \mathcal{B}_i \cup \mathcal{C}_i \mid G_i^{\cap})$ (which depends on $G_i^{\cap}$). By the law of total expectation and Claim 5.8, we have $\mathbb{E}W = \Pr(\mathcal{A}_i \cup \mathcal{B}_i \cup \mathcal{C}_i) \le f$ for some $f = n^{-\omega(1)}$, so $\Pr\big(W \ge \sqrt{f}\big) \le \sqrt{f}$ by Markov's inequality. Since $\sqrt{f}$ is still of the form $n^{-\omega(1)}$, letting $\chi = -\log\sqrt{f}/\log n$, the desired result follows. □

Proof of Claim 5.5. For any $k \le n'$, note that

$$E_k - E_{k-1} = p^{\binom{h-1}{2}}\binom{k}{h-1} - p^{\binom{h-1}{2}}\binom{k-1}{h-1} = O\big(n^{h-2}\big).$$

Now, the second inequality then follows from the Azuma–Hoeffding inequality: we have just observed that adding or removing a vertex from N_i changes $E_{|N_i|}$ by $O\big(n^{h-2}\big)$, so

$$\Pr\Big(\big|E_{|N_i|}-\mathbb{E}X_i\big|>n^{h-3/2}\log n\Big)=\exp\left(-\Omega\left(\frac{\big(n^{h-3/2}\log n\big)^2}{n\cdot n^{2(h-2)}}\right)\right)=e^{-\Omega((\log n)^2)}=n^{-\omega(1)}.$$

For the first inequality, we can now assume that $x\ge\mathbb{E}X_i-n^{h-3/2}\log n-n^{h-2}\log n=\Omega(n^{h-1})$. Therefore, we can only have $|E_k-x|\le n^{h-2}\log n$ if $k=\Omega(n)$. If $k=\Omega(n)$ then we can compute $E_k-E_{k-1}=\Omega(n^{h-2})$, so there are only $O(\log n)$ values of k which yield $|E_k-x|\le n^{h-2}\log n$. Since $|N_i|$ has the binomial distribution $\mathrm{Bin}(n',p)$, the probability that $|N_i|$ takes one of these values is $O(\log n/\sqrt{n})$, recalling that $n'=\Omega(n)$. □

6 Concluding remarks

In this paper we have proved several new anti-concentration inequalities and given some applications. There are many interesting directions of future research.

First, we still do not have a complete understanding of anti-concentration for bounded-degree polynomials in the "Gaussian" regime where p is fixed. Most obviously, it would be very interesting to remove the polylogarithmic factor from the Meka–Nguyen–Vu inequality, for polynomials which have both positive and negative coefficients. As noted in [26], this would imply Conjecture 1.3.

Also, while the Meka–Nguyen–Vu inequality gives an almost-optimal bound on $Q_{f(\boldsymbol{\xi})}(1)$ (for a bounded-degree polynomial f and $\boldsymbol{\xi}\in\mathrm{Ber}(p)^n$), our understanding of the whole concentration function $Q_{f(\boldsymbol{\xi})}$ is still quite limited, even for "dense" polynomials with many large coefficients. For example, if f has degree $d=O(1)$ and $\Omega(n^d)$ coefficients with absolute value at least 1, then the Meka–Nguyen–Vu inequality gives $Q_{f(\boldsymbol{\xi})}(r)=O((\log n)^{O(1)}(r+1)/\sqrt{n})$, whereas it seems likely that the correct bound should be $Q_{f(\boldsymbol{\xi})}(r)=O((r^{1/d}+1)/\sqrt{n})$ (attained by the polynomial $(x_1+\dots+x_n-pn)^d$).

Second, regarding the "Poisson" regime where p may be a vanishing function of n, we were not able to find a polynomial anti-concentration inequality that implies Conjecture 1.5, which was our initial motivation for this study. There are invariance principles (see [15, 16]) which allow us to compare $X_{G,k}$ to polynomials of Bernoulli random variables, but as we observed in Section 1.2, in general there are bounded-degree polynomials yielding point probabilities much larger than $1/e$. The invariance principles in [15, 16] yield polynomials with special structure (*harmonic* polynomials), and perhaps it would be feasible to prove an analogue of Theorem 1.10 for such polynomials, which might yield a new proof of Conjecture 1.5 and a generalisation for hypergraphs.

So far, in order to avoid trivialities, when we allow p to decrease with n we have been considering probabilities of the form $\Pr(X=x)$ only when $x\neq 0$. A different way to avoid trivialities would be to impose that the a_i are nonzero and explicitly specify the dependence of p on n; of particular interest may be the Poisson regime where $p=\lambda/n$ for some constant λ. However, in this setting there do not seem to be theorems that are quite as elegant as Theorem 1.8. In the case where all the a_i are positive, one can imitate Erdős' proof of the Erdős–Littlewood–Offord theorem to prove a bound of the form

$$\Pr(a_1\xi_1+\dots+a_n\xi_n=x)\le\max_{x\in\mathbb{Z}}\Pr(Z=x)+o(1),$$

where Z has the Poisson distribution $\mathrm{Po}(\lambda)$. If we do not require that the a_i are all positive, we get some more complicated behaviour. We can resolve the linear case with Fourier analysis, as follows.

Proposition 6.1. *Fix $\lambda>0$ and consider a linear polynomial $X=\sum_{i=1}^n a_i\xi_i$, where $\boldsymbol{\xi}\in\mathrm{Ber}(\lambda/n)^n$. Then for any $x\in\mathbb{R}$,*

$$\Pr(X=x)\le\frac{I_0(\lambda)}{e^\lambda}+o(1),$$

where

$$I_0(\lambda)=\sum_{i=0}^\infty\frac{(\lambda/2)^{2i}}{(i!)^2}$$

is an evaluation of a modified Bessel function of the first kind.

This bound is best-possible, as can be proved by considering the case where $a_1, \dots, a_{\lfloor n/2 \rfloor} = 1$ and $a_{\lfloor n/2 \rfloor + 1}, \dots, a_n = -1$.

Proof of Proposition 6.1. First, with a standard reduction we may assume all the a_i are integers. Indeed, we can view $\mathbb{R}$ as a vector space over the rational numbers $\mathbb{Q}$, and choose a projection map $P : \mathbb{R} \to \mathbb{Q}$ such that $P(a_i) \neq 0$ for each i. Clearing denominators by multiplying by some integer d, we obtain nonzero integers $a_i' = dP(a_i)$ such that whenever we have $a_1\xi_1 + \dots + a_n\xi_n = x$, we have $a_1'\xi_1 + \dots + a_n'\xi_n = dP(x)$. That is, any anti-concentration bound for the random variable $a_1'\xi_1 + \dots + a_n'\xi_n$ implies the same bound for $a_1\xi_1 + \dots + a_n\xi_n$.

So, we assume each a_i is an integer (therefore we may also assume x is an integer). We do Fourier analysis over $\mathbb{Z}/N\mathbb{Z}$, for some prime N very large relative to n, x and the a_i. For all $a \in \mathbb{Z}/N\mathbb{Z}$, let $f_a = (1 - \lambda/n)\delta_0 + (\lambda/n)\delta_a$, so that

$$\hat{f}_a(k) = \frac{\lambda}{n} e^{-2\pi i a k/N} + \left(1 - \frac{\lambda}{n}\right).$$

Also, we have

$$\begin{aligned} \Pr(X = x) &= f_{a_1} * \dots * f_{a_n}(x) \\ &= \frac{1}{N} \sum_{k=0}^{N-1} e^{2\pi i x k/N} \prod_{j=1}^{n} \hat{f}_{a_j}(k) \\ &\leq \frac{1}{N} \sum_{k=0}^{N-1} \prod_{j=1}^{n} \left|\hat{f}_{a_j}(k)\right| \\ &\leq \prod_{j=1}^{n} \left(\frac{1}{N} \sum_{k=0}^{N-1} \left|\hat{f}_{a_j}(k)\right|^n \right)^{1/n} \\ &= \frac{1}{N} \sum_{k=0}^{N-1} \left|\hat{f}_1(k)\right|^n, \end{aligned}$$

where the second equality is by the Fourier inversion formula, and the second inequality is by Hölder's inequality. Now, taking $N \to \infty$ gives

$$\Pr(X = x) \leq \int_0^1 \left| \frac{\lambda}{n} e^{-2\pi i x} + \left(1 - \frac{\lambda}{n}\right) \right|^n \mathrm{d}\, x.$$

As $n \to \infty$ we can compute

$$\left| \frac{\lambda}{n} e^{-2\pi i x} + \left(1 - \frac{\lambda}{n}\right) \right|^n \to \left| e^{-\lambda(1 - \cos(2\pi x) + i \sin(2\pi x))} \right| = e^{-\lambda + \lambda \cos(2\pi x)},$$

and it is known (see for example [40, Eq. (3), p. 181]) that

$$\int_0^1 e^{\lambda \cos(2\pi x)} \,\mathrm{d}\, x = I_0(\lambda),$$

so by the dominated convergence theorem, $\Pr(X = x) \leq I_0(\lambda)e^{-\lambda} + o(1)$. □

We also think it might be interesting to study the situation for general p (in particular, the intermediate regime between $p = \lambda/n$ "Poisson" behaviour and $p = 1/2$ "Gaussian" behaviour). The linear case would be a good start, as follows.

Question 6.2. *Let $\boldsymbol{a} = (a_1, \dots, a_n) \in (\mathbb{R} \setminus \{0\})^n$, $\boldsymbol{\xi} \in \mathrm{Ber}(p)^n$ for some $0 < p \leq 1/2$ and $X = a_1\xi_1 + \dots + a_n\xi_n$. What upper bounds (in terms of n and p) can we give on the maximum point probability $Q_X(0) = \max_{x \in \mathbb{R}} \Pr(X = x)$?*

Also, we remark that the constant $1/e$ in Theorems 1.8 and 1.10 appears in several other combinatorial and probabilistic problems, such as in a well-known conjecture of Feige [14].

Finally, on the subject of subgraph counts in random graphs, it may also be interesting to study anti-concentration of the number of *induced* copies X'_H of a subgraph H in a random graph $\mathbb{G}(n,p)$. (This question was also raised by Meka, Nguyen and Vu [31]). Using Theorem 1.2 in the same way as the proof of Theorem 1.13, one can prove that $\Pr\big(|X'_H - x| \le n^{h-2}\big) = O\big(\frac{1}{n}\big)$, provided p is different to the edge-density of H. The natural analogue of Conjecture 1.12 is that for a fixed graph H and fixed $p \in (0,1)$, we have

$$\max_{x\in\mathbb{N}} \Pr(X'_H = x) = O\left(1/\sqrt{\operatorname{Var}(X'_H)}\right).$$

We remark that the behaviour of $\sqrt{\operatorname{Var}(X'_H)}$ is not entirely trivial: for most values of p it has order $\Theta(n^{h-1})$, but when p is exactly equal to the edge-density of H it may have order $\Theta(n^{h-3/2})$ or $\Theta(n^{h-2})$ (see [22, Theorem 6.42]).

Acknowledgements. We would like to thank Van Vu for some clarifications regarding his work with Meka and Nguyen [31]. Also, after we completed a draft of this paper, Anders Martinsson and Frank Mousset told us that one can give an alternative proof of Theorem 1.10 using the ideas in [29, Section 2]. Finally, in a previous version of this paper we stated a specific conjecture in the setting of Question 6.2, which was observed by Mihir Singhal to be incorrect.

References

[1] N. Alon, *Combinatorial Nullstellensatz*, Combin. Probab. Comput. **8** (1999), no. 1-2, 7–29, Recent trends in combinatorics (Mátraháza, 1995).

[2] N. Alon, D. Hefetz, M. Krivelevich, and M. Tyomkyn, *Edge-statistics on large graphs*, Combin. Probab. Comput. **29** (2020), no. 2, 163–189.

[3] J. Balogh, P. Hu, B. Lidický, and F. Pfender, *Maximum density of induced 5-cycle is achieved by an iterated blow-up of 5-cycle*, European J. Combin. **52** (2016), part A, 47–58.

[4] A. D. Barbour, M. Karoński, and A. Ruciński, *A central limit theorem for decomposable random variables with applications to random graphs*, J. Combin. Theory Ser. B **47** (1989), no. 2, 125–145.

[5] R. Berkowitz, *A quantitative local limit theorem for triangles in random graphs*, arXiv preprint arXiv:1610.01281 (2016).

[6] R. Berkowitz, *A local limit theorem for cliques in $G(n,p)$*, arXiv preprint arXiv:1811.03527 (2018).

[7] B. Bollobás, *Random graphs*, second ed., Cambridge Studies in Advanced Mathematics, vol. 73, Cambridge University Press, Cambridge, 2001.

[8] B. Bollobás, L. Pebody, and O. Riordan, *Contraction-deletion invariants for graphs*, J. Combin. Theory Ser. B **80** (2000), no. 2, 320–345.

[9] J. Bourgain, V. H. Vu, and P. M. Wood, *On the singularity probability of discrete random matrices*, J. Funct. Anal. **258** (2010), no. 2, 559–603.

[10] K. P. Costello, T. Tao, and V. Vu, *Random symmetric matrices are almost surely nonsingular*, Duke Math. J. **135** (2006), no. 2, 395–413.

[11] A. de Mier and M. Noy, *On graphs determined by their Tutte polynomials*, Graphs Combin. **20** (2004), no. 1, 105–119.

[12] S. R. Dunbar, *Topics in probability theory and stochastic processes: The moderate deviations result*, 2012, URL: `https://www.math.unl.edu/~sdunbar1/ProbabilityTheory/Lessons/BernoulliTrials/ModerateDeviations/moderatedeviations.pdf`. Last visited on 2018/12/03.

[13] P. Erdős, *On a lemma of Littlewood and Offord*, Bull. Amer. Math. Soc. **51** (1945), 898–902.

[14] U. Feige, *On sums of independent random variables with unbounded variance and estimating the average degree in a graph*, SIAM J. Comput. **35** (2006), no. 4, 964–984.

[15] Y. Filmus, G. Kindler, E. Mossel, and K. Wimmer, *Invariance principle on the slice*, 31st Conference on Computational Complexity, LIPIcs. Leibniz Int. Proc. Inform., vol. 50, Schloss Dagstuhl. Leibniz-Zent. Inform., Wadern, 2016, Art. No. 15, 10 pages.

[16] Y. Filmus and E. Mossel, *Harmonicity and invariance on slices of the Boolean cube*, 31st Conference on Computational Complexity, LIPIcs. Leibniz Int. Proc. Inform., vol. 50, Schloss Dagstuhl. Leibniz-Zent. Inform., Wadern, 2016, Art. No. 16, 13 pages.

[17] J. Fox, M. Kwan, and L. Sauermann, *Anticoncentration for subgraph counts in random graphs*, arXiv preprint arXiv:1905.12749 (2019).

[18] J. Fox, M. Kwan, and B. Sudakov, *Acyclic subgraphs of tournaments with high chromatic number*, arXiv preprint arXiv:1912.07722 (2019).

[19] J. Fox and L. Sauermann, *A completion of the proof of the edge-statistics conjecture*, Advances in Combinatorics 2020:4.

[20] J. Gilmer and S. Kopparty, *A local central limit theorem for triangles in a random graph*, Random Structures Algorithms **48** (2016), no. 4, 732–750.

[21] D. Hefetz and M. Tyomkyn, *On the inducibility of cycles*, J. Combin. Theory Ser. B **133** (2018), 243–258.

[22] S. Janson, T. Łuczak, and A. Ruciński, *Random graphs*, Cambridge University Press, 2000.

[23] O. Kallenberg, *Foundations of modern probability*, Probability and its Applications (New York), Springer-Verlag, New York, 1997.

[24] J. H. Kim and V. H. Vu, *Concentration of multivariate polynomials and its applications*, Combinatorica **20** (2000), no. 3, 417–434.

[25] D. Král', S. Norin, and J. Volec, *On the exact maximum induced density of almost all graphs and their inducibility*, J. Combin. Theory Ser. A **161** (2019), 359–363.

[26] M. Kwan, B. Sudakov, and T. Tran, *Anticoncentration for subgraph statistics*, J. London Math. Soc. **99**, no. 3 (2019), 757–777.

[27] J. E. Littlewood and A. C. Offord, *On the number of real roots of a random algebraic equation. III*, Rec. Math. [Mat. Sbornik] N.S. **12(54)** (1943), 277–286.

[28] M. Loebl, J. Matoušek, and O. Pangrác, *Triangles in random graphs*, Discrete Math. **289** (2004), no. 1-3, 181–185.

[29] A. Martinsson, F. Mousset, A. Noever, and M. Trujić, *The edge-statistics conjecture for $\ell \ll k^{6/5}$*, Israel J. Math. **234** (2019), no. 2, 677–690.

[30] C. McDiarmid, *Concentration*, Probabilistic methods for algorithmic discrete mathematics, Algorithms Combin., vol. 16, Springer, Berlin, 1998, pp. 195–248.

[31] R. Meka, O. Nguyen, and V. Vu, *Anti-concentration for polynomials of independent random variables*, Theory Comput. **12** (2016), Paper No. 11, 16 pages.

[32] H. H. Nguyen and V. H. Vu, *Small ball probability, inverse theorems, and applications*, Erdős centennial, Bolyai Soc. Math. Stud., vol. 25, János Bolyai Math. Soc., Budapest, 2013, pp. 409–463.

[33] N. Pippenger and M. C. Golumbic, *The inducibility of graphs*, J. Combin. Theory Ser. B **19** (1975), no. 3, 189–203.

[34] J. Rosiński and G. Samorodnitsky, *Symmetrization and concentration inequalities for multilinear forms with applications to zero-one laws for Lévy chaos*, Ann. Probab. **24** (1996), no. 1, 422–437.

[35] A. Razborov and E. Viola, *Real advantage*, ACM Trans. Comput. Theory **5** (2013), no. 4, Art. 17, 8 pages.

[36] N. Ross, *Fundamentals of Stein's method*, Probab. Surv. **8** (2011), 210–293.

[37] T. Tao and V. Vu, *From the Littlewood-Offord problem to the circular law: universality of the spectral distribution of random matrices*, Bull. Amer. Math. Soc. (N.S.) **46** (2009), no. 3, 377–396.

[38] T. Tao and V. H. Vu, *Inverse Littlewood-Offord theorems and the condition number of random discrete matrices*, Ann. of Math. (2) **169** (2009), no. 2, 595–632.

[39] V. Vu, *Anti-concentration inequalities for polynomials*, A journey through discrete mathematics, Springer, Cham, 2017, pp. 801–810.

[40] G. N. Watson, *A treatise on the theory of Bessel functions*, Cambridge Mathematical Library, Cambridge University Press, Cambridge, 1995, Reprint of the second (1944) edition.

[41] R. Yuster, *On the exact maximum induced density of almost all graphs and their inducibility*, J. Combin. Theory Ser. B **136** (2019), 81–109.