GAUSSIAN RANDOM GRAPHS AND RAMSEY NUMBERS

ZACH HUNTER, ALEKSA MILOJEVIĆ, AND BENNY SUDAKOV

ABSTRACT. We give a simple proof of the recent remarkable exponential improvement for Ramsey lower bounds, obtained by Ma, Shen and Xie. Our key ingredient is an alternative construction based on Gaussian random graphs, which allows us to simplify their analysis significantly. As a consequence of this simpler analysis, we also obtain better quantitative bounds.

1. INTRODUCTION

For positive integers $k, \ell \geq 1$, the *Ramsey number* $R(\ell, k)$ is the smallest integer n such that any edge-coloring of the complete graph K_n using the colors red and blue either contains a copy of K_ℓ all of whose edges are red or a copy of K_k all of whose edges are blue. These numbers play a central role in *Ramsey theory*, which is itself an essential branch of combinatorics with a long and rich history, with hundreds of papers published each year. Arguably the most famous problem in Ramsey theory is to determine the growth of Ramsey numbers $R(\ell, k)$ when $k = \ell$, also known as the *diagonal Ramsey numbers*, and more generally when $k = C\ell$. Without attempting to comprehensively survey all of the recent results, let us mention that the last couple of years have brought a series of exciting developments on many long-standing questions related to Ramsey numbers of cliques, see e.g. [1, 4, 5, 7, 13, 15, 16, 17] and the references therein.

The first quantitative upper bound on $R(\ell, k)$ comes from a classical paper of Erdős and Szekeres [10] from 1935, and states that $R(\ell, k) \leq \binom{\ell+k-2}{\ell-1}$, thus showing that $R(\ell, \ell) \leq 4^\ell$. After several important improvements by Thomason [22], Conlon [6] and Sah [20], in a recent breakthrough paper Campos, Griffiths, Morris and Sahasrabudhe [4] showed that $R(\ell, \ell) \leq 4^{(1-\varepsilon)\ell}$ for a positive constant $\varepsilon > 0$ (see also [12] and [1] for further developments).

The study of the lower bounds for the Ramsey numbers had at least equal, if not greater, influence on combinatorics. In the paper which introduced the probabilistic method to combinatorics, Erdős [9] in 1947 gave an argument using random graphs which shows $R(\ell, \ell) \geq \frac{\ell}{e\sqrt{2}} 2^{\ell/2}$. Beyond a modest improvement [21] by a factor of 2 using the Lovász Local Lemma, the progress on this question was extremely slow and Erdős' bound stays essentially unchanged to this day. Hence it was a great surprise when Ma, Shen and Xie [15] used a different random graph model to give an exponential improvement to Erdős' lower bound for $R(\ell, k)$ when $k = C\ell$ for a constant $C > 1$. To obtain this improvement, they used a geometrically defined random graph, whose vertices are random points on a high-dimensional sphere and the colors of the edges are determined by the distances between the points. Because of the complicated geometry of high-dimensional spheres, the analysis of this random graph model is long and technical, spanning many pages.

In this paper, we propose an alternative approach to this problem. Although we draw our inspiration from the work of Ma, Shen and Xie, we use random Gaussian vectors to define our graphs. Of course, in the regime of parameters considered here, the length of the Gaussian random vectors we choose is highly concentrated, so the resulting graph is very similar to the one obtained by placing random points on the sphere. However, the key advantage is that the Gaussian measure is a product measure, that is to say its coordinates are independent. This, along with a couple of further ideas, allows us to significantly simplify and shorten the analysis of such graphs, and to obtain tighter bounds.

In order to state our main theorem, let us briefly recall the classical lower bound on $R(\ell, k)$ coming from Erdős' argument. If we color every edge of the graph red with probability p or blue with probability $1 - p$, independently of all other edges, then the probability that ℓ vertices form a red clique is exactly $p^{\binom{\ell}{2}}$, and the probability that k vertices form a blue clique is $(1-p)^{\binom{k}{2}}$. Thus, the expected number of red ℓ-cliques is $\binom{n}{\ell} p^{\binom{\ell}{2}}$, and the expected number of blue k-cliques is $\binom{n}{k} (1-p)^{\binom{k}{2}}$. If for some positive integer n we have that $\binom{n}{\ell} p^{\binom{\ell}{2}} + \binom{n}{k}(1-p)^{\binom{k}{2}} < 1$, then there exists a coloring in which both the number of red ℓ-cliques and blue k-cliques is zero, and therefore we have $R(\ell, k) \geq n$. When $k = C\ell$, the probability p which gives the best lower bound is denoted by p_C, and it can be found as the unique solution of the equation $C = \frac{\log p_C}{\log(1-p_C)}$. In this case, the corresponding lower bound is $R(\ell, k) \gtrsim p_C^{-\ell/2}$. In this paper we improve this bound as follows.

Theorem 1.1. *For any $C > 1$, there exists a constant ε such that, for all sufficiently large ℓ, we have*

$$R(\ell, C\ell) \geq (p_C^{-1/2} + \varepsilon)^\ell,$$

where $p_C \in (0, 1/2)$ is the unique solution to the equation $C = \frac{\log p_C}{\log(1-p_C)}$.

Beyond its simplicity, another advantage of this analysis is that it allows us to obtain quantitative improvements over the work of Ma, Shen and Xie. Namely, as $C \to \infty$ and $p_C \to 0$, Ma, Shen and Xie obtain a gain of ε which goes to 0 with $C \to \infty$. On the other hand, the argument we present here can be extended to show one can take $\varepsilon = (e^{1/24} - 1)p_C^{-1/2}$, thus showing $R(\ell, C\ell) \geq (e^{1/24} p_C^{-1/2})^\ell$ for all sufficiently large C. Let us also remark that when $C \to 1$, our proof gives the same quantitative behavior $\varepsilon = \Omega((C-1)^2)$ as [15]. To focus on the simplest possible presentation, we present this proof only in a forthcoming appendix to this paper.

1.1. **Gaussian random geometric graph** $G(n, d, p)$. Let us now formally define the model of the Gaussian random geometric graph, before saying a couple of words about how to analyze it. Let $n, d \geq 1$ be positive integers, and let $p \in (0, 1/2)$ be a probability. Also, let $c_p > 0$ be the unique real number for which $\mathbb{P}[Z \leq -c_p] = p$, where $Z \sim \mathcal{N}(0, 1)$ is the standard one-dimensional Gaussian.

Definition 1.2. The vertices of the Gaussian random geometric graph $G(n, d, p)$ correspond to vectors $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n$, independently sampled from the d-dimensional normal distribution $\mathcal{N}(0, \frac{1}{d}I_d)$. Further, the vertices $i, j \in [n]$ are connected by an edge if $\langle \boldsymbol{x}_i, \boldsymbol{x}_j \rangle \geq -\frac{c_p}{\sqrt{d}}$.

Note that given a graph $G \sim G(n, d, p)$, we can define a red/blue edge-coloring of the complete graph K_n, by declaring the edges of G to be blue, and all other edges to be red.

Our notion of random geometric graphs differs slightly from the one commonly used in the literature. There are two main reasons. Firstly, it would be common to define the threshold c_p such that $\langle \boldsymbol{x}_i, \boldsymbol{x}_j \rangle \geq -\frac{c_p}{\sqrt{d}}$ occurs exactly with probability $1 - p$, when $\boldsymbol{x}_i, \boldsymbol{x}_j \sim \mathcal{N}(0, \frac{1}{d}I_d)$, since the graph $G(n, d, p)$ would then have expected density *exactly* $1 - p$.

However, we opt for a simpler definition of c_p, and observe that the graph we define has expected density *very close* to $1 - p$ (in fact, the expected density converges to $1 - p$ as d increases). The reason is that the length of vectors $\boldsymbol{x}_i$ is well concentrated around 1, and so $\langle \boldsymbol{x}_i, \boldsymbol{x}_j \rangle$ is essentially the length of the projection of $\boldsymbol{x}_j$ onto $\boldsymbol{x}_i$, which is a Gaussian random variable of variance $\frac{1}{d}$, and as such exceeds the threshold $-c_p/\sqrt{d}$ with probability $1 - p$.

While this first difference is mostly cosmetic, the second one is more substantial. It seems that random geometric graphs of density larger than $1/2$ (i.e. when the threshold is taken to be negative) have not been thoroughly studied in the literature, with most works focusing on the regime of sparse random geometric graphs. However, there is no obvious symmetry relating the two settings, such as taking the complements. Still, as [15] and this work demonstrate, it is worthwhile to study the dense random geometric graphs, especially since some of the techniques from the random geometric graph literature do not work in this setting anymore.

Let us also mention that the random geometric graphs have been thoroughly investigated in their own right. For the early study of random geometric graphs, mostly in low dimension, see Penrose's book *Random Geometric Graphs* [18]. Some 15 years ago, there was a shift and the random geometric graphs in high dimension also started receiving considerable attention. In one of the first such works, Devroye, György, Lugosi and Udina [8] studied the clique numbers of random geometric graphs, showing that, for example, when $d \gg (\log n)^3$ then $\omega(G(n, d, p)) \leq O(\log n)$ with high probability. They also studied the regime when $d \approx (\log n)^2$ (which is the regime relevant for us, since $\ell \approx \log n$ and $d \approx \ell^2$, although they only provided the upper bound $\omega(G(n, d, p)) \leq O((\log n)^3)$ in this case. Thus, their results are not sufficient for the purposes of this paper, and we require a much more precise analysis of the clique numbers when d is as low as $(\log n)^2$.

More generally, one can ask for which d is the distribution of $G(n, d, p)$ close to that of $G(n, p)$ (in total variation distance, say), and this question received a lot of attention over the years (see e.g. [3]). However, this question is not the focus of our article and we will not discuss it at length.

1.2. **Applying Gaussian random graphs to Ramsey number lower bounds.** Let us now discuss how the model $G(n, d, p)$ is used to prove Theorem 1.1. We define the edge-coloring of K_n based on a sample $G \sim G(n, d, p)$ by coloring all edges present in G blue and all missing edges red. We will choose n to be the intended lower bound for the Ramsey number, i.e. $n = (p_C^{-1/2} + \varepsilon)^\ell$ and $d = D^2 \ell^2$, where D is a large constant depending only on C and $\ell \gg D, C$. Finally, the density p of the random geometric graph will be slightly larger than p_C, for the reasons we will explain below.

As usual, our goal will be to show that the expected number of red cliques of size ℓ and blue cliques of size $C\ell$ in this coloring is less than 1. By linearity of expectation, this reduces to bounding the probability that a red clique of size ℓ or a blue clique of size $C\ell$ arises in the above coloring. Equivalently, the goal is to bound $\mathbb{P}[G(C\ell, d, p) \text{ is a clique}]$ and $\mathbb{P}[G(\ell, d, p) \text{ is an independent set}]$. Since these probabilities will play the central role in the proofs, let us fix the notation $P_{\text{red},r} = \mathbb{P}[G(r, d, p) \text{ is an independent set}]$ and $P_{\text{blue},r} = \mathbb{P}[G(r, d, p) \text{ is a clique}]$, where $r \geq 1$ is a positive integer. The bounds on $P_{\text{red},\ell}$ and $P_{\text{blue},r}$ which we obtain are summarized in the following proposition.

Proposition 1.3. *Let $G \sim G(n, d, p)$ be a Gaussian random geometric graph, and let $a = \frac{e^{-c_p^2/2}}{\sqrt{2\pi}}$. If $d \geq D^2 \ell^2$ where D is sufficiently large compared to C, then*

$$(1) \qquad P_{\text{red},\ell} \leq p^{\binom{\ell}{2}} \exp\left(-\frac{a^3}{p^3\sqrt{d}} \binom{\ell}{3} \left(1 + O_p\left(\frac{1}{D}\right)\right)\right),$$

$$(2) \qquad P_{\text{blue},C\ell} \leq (1-p)^{\binom{C\ell}{2}} \exp\left(\frac{a^3}{(1-p)^3\sqrt{d}} \binom{C\ell}{3} \left(1 + O_{p,C}\left(\frac{1}{D}\right)\right)\right).$$

In the above proposition, the notation $O_p(1/D)$ and $O_{p,C}(1/D)$ hides constants depending on p and C, but not on ℓ or d. This will hold for each $O(\cdot)$ notation throughout the paper, and we will often go without writing it, in order to avoid cluttering the formulas. At this point, it is good to say a couple of words about the intuition behind the shape of these bounds. In the first order, the terms $p^{\binom{\ell}{2}}$ and $(1-p)^{\binom{C\ell}{2}}$ correspond to the behavior expected in an Erdős-Rényi random graph which we already discussed above.

The corrections in formulas (1), (2) account for the correlations between the edges of G. Intuitively, if ij, jk are not edges of G, then the vectors $\boldsymbol{x}_i, \boldsymbol{x}_k$ have very negative inner products with $\boldsymbol{x}_j$. In other words, the vectors $\boldsymbol{x}_i, \boldsymbol{x}_k$ are very far from $\boldsymbol{x}_j$, and as an exaggeration, one can imagine that they point in an almost opposite direction from $\boldsymbol{x}_j$. Then, $\boldsymbol{x}_i, \boldsymbol{x}_k$ are most likely close together, making it unlikely that ik is not an edge. Thus, having independent sets in Gaussian random geometric graphs is much less likely than in binomial random graph.

On the other hand, cliques are slightly more likely in Gaussian random graphs, because of the triangle inequality. If pairs ij and jk are adjacent, meaning that the pairs of vectors $\boldsymbol{x}_i, \boldsymbol{x}_j$ and $\boldsymbol{x}_j, \boldsymbol{x}_k$ are somewhat close, then $\boldsymbol{x}_i$ and $\boldsymbol{x}_k$ are more likely to be close as well, thus making the edge ik slightly

more likely. Hence, this positive correlation between neighboring edges is manifested in the fact that the probability of $G(C\ell, d, p)$ being a clique is higher than in the Erdős-Rényi random graph.

At this point, one might reasonably complain - the independent sets are less likely in the graph $G(n, d, p)$ compared to the binomial random graph, but this comes at a price of cliques being more likely, and our goal is to avoid *both* cliques and independent sets. The final observation is that one might increase p slightly (and consequently decrease the density of the graph), thus lowering the probability of cliques and raising the probability of independent sets. If one gets the balance just right, the gain coming from (1) outweighs the losses in (2), and one obtains the graph in which both cliques and independent sets are less likely compared to $G(n, 1 - p_C)$. Let us formalize the last paragraph in the form of the following lemma.

Lemma 1.4. *Let $C > 1$ be a constant and let $p_C \in (0, 1/2)$ be the unique solution to $C = \frac{\log p_C}{\log 1 - p_C}$. If D is sufficiently larger than C, there exist $\varepsilon_1 = \varepsilon_1(C, D) > 0$ and $p \in (0, 1/2)$ such that the clique probabilities in $G(n, d, p)$ with $d = D^2 \ell^2$ are bounded by*

$$P_{\mathrm{red}, \ell} \le \left(p_C - \varepsilon_1 \right)^{\binom{\ell}{2}} \text{ and } P_{\mathrm{blue}, C\ell} \le \left(1 - p_C - \varepsilon_1 \right)^{\binom{C\ell}{2}}.$$

Proof of Lemma 1.4 assuming Proposition 1.3. As we have indicated in the discussion preceding the lemma, in bounds (1), (2) we have gained more in the probability $P_{\mathrm{red}, \ell}$ than we have lost in the bound on $P_{\mathrm{blue}, C\ell}$. This is justified in the following claim.

Claim 1.5.
$$\frac{a^3}{3p_C^2} > \frac{a^3 C}{3(1 - p_C)^2}.$$

Proof. By canceling out $a^3/3$, we reduce the inequality to $1/p_C^2 > C/(1 - p_C)^2$, i.e. to showing that $(1 - p_C)^2 > C p_C^2$. If we recall that $C = \frac{\log p_C}{\log(1 - p_C)} = \frac{\log_2 1/p_C}{\log_2 1/(1 - p_C)}$ and further rearrange the inequality, it reduces to

$$(1 - p_C)^2 \log_2 \frac{1}{1 - p_C} > p_C^2 \log_2 \frac{1}{p_C}.$$

Thus, the goal is to show that the function $f(t) = (1 - t)^2 \log_2 \frac{1}{1 - t} - t^2 \log_2 \frac{1}{t}$ is positive for all $t \in (0, 1/2)$. It is not hard to see that $\lim_{t \to 0^+} f(t) = \lim_{t \to 1/2^-} f(t) = 0$. Moreover, a simple calculation shows that the derivative of $f(t)$ is $f'(t) = \frac{1}{\ln 2} - 2H(t)$, where $H(t) = -t \log_2 t - (1 - t) \log_2(1 - t)$ is the binary entropy function. Thus, $f''(t) = -2H'(t)$, and since H is a strictly increasing function on $t \in (0, 1/2)$, we have $f''(t) < 0$. Hence, f is strictly concave and tends to 0 at the endpoints of the interval $(0, 1/2)$, so it must be positive inside the interval $(0, 1/2)$.   ⊡

By continuity and since $p_C < 1/2$, there exists an open neighborhood $(p_C - \eta, p_C + \eta) \subset (0, 1/2)$ of p_C and a constant ε_0 (where η and ε_0 depend only on p_C) such that

$$\underbrace{\inf_{|p - p_C| < \eta} \frac{a^3}{3p^2}}_{\alpha} > \varepsilon_0 + \underbrace{\sup_{|p - p_C| < \eta} \frac{a^3 C}{3(1 - p)^2}}_{\beta}.$$

If we denote the above infimum and supremum by α and β, we note that both α and β depend only on p_C (and thus only on C). We also have that $\alpha > \varepsilon_0 + \beta$. We can use the Proposition 1.3 to write

$$P_{\mathrm{red}, \ell}^{1/\binom{\ell}{2}} \le p \exp\left(- \left(1 + O(D^{-1})\right) \frac{\alpha \ell}{p\sqrt{d}} \right) \le p \exp\left(- \left(1 + O(D^{-1})\right) \frac{\alpha}{pD} \right), \text{ and}$$

$$P_{\mathrm{blue}, C\ell}^{1/\binom{C\ell}{2}} \le (1 - p) \exp\left(\left(1 + O(D^{-1})\right) \frac{\beta \ell}{(1 - p)\sqrt{d}} \right) \le (1 - p) \exp\left(\left(1 + O(D^{-1})\right) \frac{\beta}{(1 - p)D} \right)$$

Since D is very large compared to C, we choose $|p - p_C| < \eta$, and so by using $e^t = 1 + t + O(t^2)$ for $|t| \leq 1$ we can write

$$P_{\text{red},\ell}^{1/\binom{\ell}{2}} \leq p\left(1 - \frac{\alpha}{pD} + O(D^{-2})\right) = p - \frac{\alpha}{D} + O(D^{-2}), \text{ and}$$

$$P_{\text{blue},\ell}^{1/\binom{C\ell}{2}} \leq (1-p)\left(1 + \frac{\beta}{(1-p)D} + O(D^{-2})\right) = 1 - p + \frac{\beta}{D} + O(D^{-2}).$$

Hence, if we set $p = p_C + \frac{\alpha+\beta}{2D}$ (recall that α, β depend only on p_C) and let D be sufficiently larger than C, we obtain

$$P_{\text{red},\ell} \leq \left(p_C - \frac{\alpha - \beta}{2D} + O(D^{-2})\right)^{\binom{\ell}{2}} \text{ and } P_{\text{blue},C\ell} \leq \left(1 - p_C - \frac{\alpha - \beta}{2D} + O(D^{-2})\right)^{\binom{C\ell}{2}}.$$

By setting $\varepsilon_1 < \varepsilon_0/2D$ and letting D be sufficiently large, we obtain precisely what we wanted. $\qquad\square$

Proof of Theorem 1.1 assuming Proposition 1.3. By Lemma 1.4, there exists some $\varepsilon_1 > 0$ and $p \in (0, 1/2)$ for which the clique probabilities in $G(n, d, p)$ (where $d = D^2\ell^2$) satisfy

$$P_{\text{red},\ell} \leq \left(p_C - \varepsilon_1\right)^{\binom{\ell}{2}} \text{ and } P_{\text{blue},C\ell} \leq \left(1 - p_C - \varepsilon_1\right)^{\binom{C\ell}{2}}.$$

Also, set $n = (p_C^{-1/2} + \varepsilon)^\ell$ for some $\varepsilon \ll \varepsilon_1$, and let us show that the red/blue coloring of K_n defined using the Gaussian random geometric graph has no red ℓ-clique and no blue $C\ell$-clique with high probability.

By the union bound the probability that this coloring has a red ℓ-clique is at most

$$\binom{n}{\ell} \cdot P_{\text{red},\ell} \leq \frac{n^\ell}{\ell!}(p_C - \varepsilon_1)^{\frac{\ell(\ell-1)}{2}} \leq \frac{1}{\ell!}\left(n(p_C - \varepsilon_1)^{\frac{\ell-1}{2}}\right)^\ell = \frac{1}{\ell!}\left((p_C^{-1/2} + \varepsilon)^\ell(p_C - \varepsilon_1)^{\frac{\ell-1}{2}}\right)^\ell.$$

The crucial observation is that, inside the parentheses, we have an exponential term in ℓ with the base of the exponent being $(p_C^{-1/2} + \varepsilon)(p_C - \varepsilon_1)^{1/2} < 1$. Also, note that the additive constant $-1/2$ in the exponent $\frac{\ell-1}{2}$ does not affect the calculation, since ℓ is very large compared to p_C. Hence, the probability that there exists a red ℓ-clique is at most $\frac{1}{\ell!}$.

Similarly, the probability that the coloring contains a blue $C\ell$-clique is at most

$$\binom{n}{C\ell} \cdot P_{\text{blue},C\ell} \leq \frac{n^{C\ell}}{(C\ell)!}(1 - p_C - \varepsilon_1)^{\frac{C\ell(C\ell-1)}{2}} \leq \frac{1}{(C\ell)!}\left(n(1 - p_C - \varepsilon_1)^{\frac{C\ell-1}{2}}\right)^{C\ell}.$$

Again, the base of the exponential term inside the parentheses is $(p_C^{-1/2} + \varepsilon)(1 - p_C - \varepsilon_1)^{C/2} < 1$ (which follows from $(1 - p_C)^C = p_C$). Hence, the probability that there exists a blue $C\ell$-clique is at most $\frac{1}{(C\ell)!}$. This shows that there is an outcome without red ℓ-cliques or blue $C\ell$-cliques, completing the proof. $\quad\square$

Organization. In Section 2, we collect a number of well-known results about the Gaussian distribution, all of whose proofs are rather standard. Although we provide the complete proofs of these results in the appendix, we suggest the reader to skim this section by only reading the statements, since the heart of the paper lies in Sections 3 and 4. In Section 3, we reduce bounding the probabilities $P_{\text{red},r}, P_{\text{blue},r}$ in Proposition 1.3 to an easier problem of bounding the probability that random Gaussian vectors $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_r$ both form a monochromatic clique and have the inner products of expected size (i.e. that they form the so-called perfect sequence). Finally, in Section 4 we present the most important part of our argument, in which we give the remaining ingredients needed to prove Proposition 1.3.

2. Preliminaries

In this section, we will recall a collection of rather standard facts related to Gaussian random variables, which we mostly provide for reference. We will begin by proving two concentration inequalities. Then, we will discuss truncated Gaussian random variables, and some of their properties. Finally, we conclude by bounding exponential moments of a certain quadratic function of truncated Gaussians, which will

play an important role in our proofs later on. The first result we present is the concentration inequality for norms of high-dimensional Gaussian vectors.

Lemma 2.1. *If $\boldsymbol{x} \sim \mathcal{N}(0, \frac{1}{d}I_d)$ and $\delta \in (0,1)$, then $\mathbb{P}\big[\|\boldsymbol{x}\|_2 \in (1-\delta, 1+\delta)\big] \geq 1 - 2\exp(-\delta^2 d/10)$.*

Observe that $d\|\boldsymbol{x}\|_2^2$ is a sum of squares of d independent standard one-dimensional Gaussians, and therefore $d\|\boldsymbol{x}\|_2^2$ follows the so-called χ_d^2 distribution with d degrees of freedom. Hence, Lemma 2.1 follows as a simple application of the famous Laurent-Massart concentration inequality for χ_d^2.

Theorem 2.2 (Lemma 1 in [14]). *Let $Y \sim \chi_d^2$ and let $t \geq 0$. Then*

$$\mathbb{P}\big[Y - d \geq 2\sqrt{dt} + 2t\big] \leq e^{-t} \quad and \quad \mathbb{P}\big[d - Y \geq 2\sqrt{dt}\big] \leq e^{-t}.$$

Next, we give a one-sided tail bound for the length of a projection of a Gaussian vector to a lower-dimensional subspace. For a vector $\boldsymbol{x} \in \mathbb{R}^d$ and a subspace $W \subseteq \mathbb{R}^d$, let $\pi_W(\boldsymbol{x})$ be the projection of $\boldsymbol{x}$ to W.

Lemma 2.3. *Let $1 \leq s \leq C\ell$ and let W be any s-dimensional subspace of $\mathbb{R}^d$. If $\boldsymbol{x} \sim \mathcal{N}(0, \frac{1}{d}I_d)$ is a random Gaussian vector and $\alpha = 100C\log(10/p)$, then*

$$\mathbb{P}\Big[\|\pi_W(\boldsymbol{x})\|_2 \geq \frac{\alpha\sqrt{\ell}}{\sqrt{d}}\Big] \leq \Big(\frac{p}{10}\Big)^{10C\ell}.$$

Since Gaussian random variables are the crucial players in this paper, let us denote by ϕ, Φ the probability density function and the cumulative distribution function of the standard normal $\mathcal{N}(0,1)$. Among their many useful properties, we will often use the following bound on the Mills ratio: for every negative $t < 0$ we have

$$(3) \qquad\qquad\qquad |t| \leq \frac{\phi(t)}{\Phi(t)} \leq |t| + \frac{1}{|t|}.$$

This follows from a simple integration by parts, see e.g. [11]. Another standard fact related to the Gaussian distribution is that the function $\Phi(t)$ is log-concave (i.e. $\log\Phi(t)$ is a concave function). This will be quite useful, since it allows us to derive that $\log\Phi(t+\varepsilon) \leq \log\Phi(t) + \varepsilon(\log\Phi(t))'$, which can also be read as

$$(4) \qquad\qquad\qquad \Phi(t+\varepsilon) \leq \Phi(t)e^{\varepsilon\phi(t)/\Phi(t)}.$$

Truncated Gaussian random variables will also play an important role in the paper. If $Z \sim \mathcal{N}(0, \sigma^2)$ is a one-dimensional Gaussian and $(a,b) \subseteq \mathbb{R}$ an interval, the *truncated Gaussian* is the distribution of Z conditioned on $Z \in (a,b)$. Our distribution will be truncated only on one side, i.e. we will only consider cases where $a = -\infty$ or $b = +\infty$. If $a \neq -\infty$, we have a lower truncated variable, and if $b \neq +\infty$ we have an upper truncated variable. We begin our discussion of truncated Gaussians by computing their expectation.

Lemma 2.4. *Let b, ε be real numbers and let $X \sim \mathcal{N}(0, \frac{1}{d})$. Then, we have the following four formulas*

$$\mathbb{E}\Big[X\Big|X \leq \frac{b}{\sqrt{d}}\Big] = -\frac{\phi(b)}{\Phi(b)\sqrt{d}} \qquad and \quad \mathbb{E}\Big[X\Big|X \leq \frac{b+\varepsilon}{\sqrt{d}}\Big] = -\frac{\phi(b)}{\Phi(b)\sqrt{d}} + O\Big(\frac{\varepsilon}{\sqrt{d}}\Big),$$

$$\mathbb{E}\Big[X\Big|X \geq \frac{b}{\sqrt{d}}\Big] = \frac{\phi(b)}{(1-\Phi(b))\sqrt{d}} \quad and \quad \mathbb{E}\Big[X\Big|X \geq \frac{b+\varepsilon}{\sqrt{d}}\Big] = \frac{\phi(b)}{(1-\Phi(b))\sqrt{d}} + O\Big(\frac{\varepsilon}{\sqrt{d}}\Big).$$

Let us close this section by showing how to bound the exponential moments of a certain quadratic sum of truncated Gaussian random variables. Let us begin by recalling the notion of subgaussian random variables. They will be important to us since they provide a good framework for bounding exponential moments, and since truncated Gaussians are subgaussian.

A random variable X is called *subgaussian* if there exists a positive number σ^2 such that

$$\mathbb{E}\big[\exp\big(\lambda(X - \mathbb{E}X)\big)\big] \leq \exp\big(\sigma^2\lambda^2/2\big) \text{ for all } \lambda \in \mathbb{R}.$$

The smallest such constant σ^2 is called the *variance proxy*. The paper [2] shows that the optimal variance proxy of a truncated Gaussian is at most the variance of the corresponding Gaussian.

Subgaussian random variables possess many useful properties, and here we will survey only those which we will need in the paper. For a more thorough treatment, we suggest to consult Vershynin's book [23]. A basic observation, which follows directly from the definition, is that if $X_1, \ldots, X_k$ are independent subgaussians with variance proxies $\sigma_1^2, \ldots, \sigma_k^2$, then $X_1 + \cdots + X_k$ is also subgaussian with variance proxy $\sigma_1^2 + \cdots + \sigma_k^2$. The most important property of subgaussians will be the following bound on the exponential moments of their squares.

Lemma 2.5. *Let X be a subgaussian of variance proxy σ^2 and with $\mathbb{E}[X] = 0$, and let $0 \leq \lambda < \frac{1}{2\sigma^2}$. Then*

$$\mathbb{E}[e^{\lambda X^2}] \leq 1 + \frac{4\lambda\sigma^2}{1 - 2\lambda\sigma^2}.$$

Finally, the reason we introduced subgaussian random variables is to prove the following lemma, which says that the exponential moments of a certain quadratic sum of subgaussian random variables is in essence controlled by the expectation of this sum.

Lemma 2.6. *Let $X_1, \ldots, X_k$ be subgaussian random variables with variance proxy $1/d$, and let $\lambda \in (-\infty, \infty)$ such that $d \geq 4|\lambda|k$. If $S = \sum_{1 \leq i < j \leq k} X_i X_j$, then*

$$\mathbb{E}[e^{\lambda S}] \leq \exp\left(\lambda \mathbb{E}[S] + \frac{\lambda^2 k^2}{d} \sum_{j=1}^{k} (\mathbb{E}X_j)^2 + \frac{4|\lambda|k}{d}\right).$$

3. Perfect sequences

In this section, we will reduce the problem of bounding $P_{\text{red},\ell}, P_{\text{blue},C\ell}$ to a simpler problem. Recall that in general, for $1 \leq r \leq C\ell$, $P_{\text{red},r}$ denotes the probability that r independent Gaussian vectors $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_r$ form a red clique. Observe that such random vectors are typically almost orthogonal to each other, because of Lemma 2.3.

It turns out that it is much simpler to analyze the behavior of the sequence $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_r$ if indeed *all* inner products $\langle \boldsymbol{x}_i, \boldsymbol{x}_j \rangle$ are small in absolute value. In other words, it is much easier to analyze the probability that a *typical* sequence $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_r$ forms a red clique than a general one, and in both cases we essentially get the same answer.

Following Ma, Shen and Xie [15], who used a very similar notion, we say that $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_r$ is a *perfect sequence* if for all $1 \leq i \leq r$ we have:

$$\|\boldsymbol{x}_i\|_2 \in (1 - \delta, 1 + \delta) \quad \text{and} \quad \|\pi_{\text{span}\{\boldsymbol{x}_1,\ldots,\boldsymbol{x}_{i-1}\}}(\boldsymbol{x}_i)\|_2 \leq \frac{\alpha\sqrt{\ell}}{\sqrt{d}},$$

where $\alpha = 100C\log(10/p)$ and $\delta = \alpha d^{-1/4}$.

Let $P_{\text{red},r}^*$ be the probability that $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_r$ form a red clique and a perfect sequence, and let $P_{\text{blue},r}^*$ be defined analogously. In light of this definition, we now have two tasks - we need to bound $P_{\text{red},r}^*$ and we need to show that $P_{\text{red},r}$ and $P_{\text{red},r}^*$ are essentially the same. The first task will be performed via the following proposition.

Proposition 3.1. *Let $C > 1, p \in (0, 1)$ be real numbers and $\ell, d = D^2\ell^2$ be large positive integers, where D is sufficiently larger than C. For every integer $1 \leq r \leq C\ell$, we have*

$$(5) \qquad P_{\text{red},r}^* \leq p^{\binom{r}{2}} \exp\left(-\left(1 + O_p(D^{-1})\right) \frac{a^3}{p^3\sqrt{d}} \binom{r}{3}\right)$$

$$(6) \qquad P_{\text{blue},r}^* \leq (1-p)^{\binom{r}{2}} \exp\left(\left(1 + O_p(D^{-1})\right) \frac{a^3}{(1-p)^3\sqrt{d}} \binom{r}{3}\right),$$

where $a = \phi(c_p) = \frac{e^{-c_p^2/2}}{\sqrt{2\pi}}$.

The proof of Proposition 3.1 is the heart of our argument, and it will be presented in the next section. In this section, however, we will show how to get the bounds on $P_{\mathrm{red},r}, P_{\mathrm{blue},r}$ from (5), (6).

Proof of Proposition 1.3 based on Proposition 3.1. We will present a bound only for $P_{\mathrm{red},r}$, since the bound on $P_{\mathrm{blue},r}$ can be derived in a completely analogous fashion. Recall, we aim to bound $P_{\mathrm{red},r}$, the probability that $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_r$ form a red clique.

Let us begin the proof by describing a procedure which extracts a perfect sequence $(\boldsymbol{y}_1, \ldots, \boldsymbol{y}_t)$ from $(\boldsymbol{x}_1, \ldots, \boldsymbol{x}_r)$. To do this, we go through the sequence $(\boldsymbol{x}_1, \ldots, \boldsymbol{x}_r)$ in order. In the i-th step, we consider the element $\boldsymbol{x}_i$, and assuming that we have defined $\boldsymbol{y}_1, \ldots, \boldsymbol{y}_j$ so far, we check whether $\|\boldsymbol{x}_i\|_2 \in (1-\delta, 1+\delta)$ and $\|\pi_{\mathrm{span}\{\boldsymbol{y}_1,\ldots,\boldsymbol{y}_j\}}(\boldsymbol{x}_i)\|_2 \le \frac{\alpha\sqrt{\ell}}{\sqrt{d}}$. If both inequalities hold, we set $\boldsymbol{y}_{j+1} = \boldsymbol{x}_i$, and otherwise we simply throw $\boldsymbol{x}_i$ away.

Let I be the set of indices i for which $\boldsymbol{x}_i$ was included in the sequence $(\boldsymbol{y}_1, \ldots, \boldsymbol{y}_t)$. By definition, $(\boldsymbol{y}_1, \ldots, \boldsymbol{y}_t)$ is a perfect sequence. Also, observe that $(\boldsymbol{x}_1, \ldots, \boldsymbol{x}_r)$ is perfect if and only if $I = [r]$.

To bound the probability that $(\boldsymbol{x}_1, \ldots, \boldsymbol{x}_r)$ forms a red ℓ-clique, we use a union bound over all possible sets I produced by the above procedure. If we let $p_I = \mathbb{P}\big[\boldsymbol{x}_1, \ldots, \boldsymbol{x}_r$ form a red clique and the above procedure gives the set $I\big]$, then

$$(7) \qquad P_{\mathrm{red},r} = \mathbb{P}[\boldsymbol{x}_1, \ldots, \boldsymbol{x}_r \text{ form a red clique}] = \sum_{I \subseteq [r]} p_I.$$

As suggested above, the main term of the above sum comes from $I = [r]$. In this case, we are evaluating the probability that the sequence $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_r$ forms a red clique and is perfect, which equals $P^*_{\mathrm{red},r}$ by definition. Let us now bound the contribution of terms coming from other sets I.

If $i \notin I$, then either $\|\boldsymbol{x}_i\|_2 \notin (1-\delta, 1+\delta)$ or $\big\|\pi_{\mathrm{span}\{\boldsymbol{x}_1,\ldots,\boldsymbol{x}_{i-1}\}}\boldsymbol{x}_i\big\|_2 \ge \big\|\pi_{\mathrm{span}\{\boldsymbol{y}_1,\ldots,\boldsymbol{y}_j\}}(\boldsymbol{x}_i)\big\|_2 \ge \frac{\alpha\sqrt{\ell}}{\sqrt{d}}$, where we used that enlarging the subspace only increases the length of the projection. The first event happens with probability at most $2e^{-\delta^2 d/10} \le e^{-\Omega(\sqrt{d})} \ll (p/10)^{10C\ell}$ (due to Lemma 2.1). The second of these events also happens with probability at most $(p/10)^{10C\ell}$, due to Lemma 2.3. Finally, note that both of these statements are true even conditionally on $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_{i-1}$. Thus, we can write the following

$$p_I \le \mathbb{P}\big[(\boldsymbol{x}_i)_{i \in I} \text{ forms a red clique and is perfect}\big]$$

$$\cdot \mathbb{P}\big[\|\boldsymbol{x}_i\|_2 \notin (1-\delta, 1+\delta) \text{ or } \big\|\pi_{\mathrm{span}\{\boldsymbol{x}_1,\ldots,\boldsymbol{x}_{i-1}\}}(\boldsymbol{x}_i)\big\|_2 \ge \frac{\alpha\sqrt{\ell}}{\sqrt{d}} \text{ for all } i \notin I\big].$$

The first probability is $P^*_{\mathrm{red},|I|}$, while the second one is at most $(p/10)^{10C\ell \cdot (r-|I|)}$, due to the independence of $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_r$. Hence, we have $p_I \le P^*_{\mathrm{red},|I|}(p/10)^{10C\ell \cdot (r-|I|)}$. Let $r - |I| = u$. We now use the bounds on $P^*_{\mathrm{red},r-u}$ coming from Proposition 3.1

$$P^*_{\mathrm{red},r-u} \le p^{\binom{r-u}{2}} \exp\left(-(1 + O(D^{-1}))\frac{a^3}{p^3\sqrt{d}}\binom{r-u}{3}\right)$$

Observe now that

$$\left(\frac{p}{10}\right)^{10C\ell \cdot u} \le p^{\binom{r}{2} - \binom{r-u}{2}} \cdot 2^{-C\ell u} \exp\left(-(1 + O(D^{-1}))\frac{a^3}{p^3\sqrt{d}}\left(\binom{r}{3} - \binom{r-u}{3}\right)\right),$$

since $\binom{r}{2} - \binom{r-u}{2} \le ru \le C\ell u$ and $\frac{a^3}{p^3\sqrt{d}}\left(\binom{r}{3} - \binom{r-u}{3}\right) \le \frac{a^3}{p^3}\frac{ur^2}{\sqrt{d}} \le \frac{a^3}{p^3}\frac{u(C\ell)^2}{D\ell} \le C\ell u$ (recall that D is sufficiently large as a function of p, C). So,

$$p_I \le P^*_{\mathrm{red},r-u} \cdot \left(\frac{p}{10}\right)^{10C\ell \cdot u} \le 2^{-2C\ell u}p^{\binom{r}{2}} \exp\left(-(1 + O(D^{-1}))\frac{a^3}{p^3\sqrt{d}}\binom{r}{3}\right).$$

Let us now split the sum in (7) depending on the size of I. Recalling that $u = r - |I|$, we have

$$P_{\mathrm{red},r} \le p^{\binom{r}{2}} \exp\left(-\left(1 + O(D^{-1})\right)\frac{a^3}{p^3\sqrt{d}}\binom{r}{3}\right) \sum_{u=0}^{r} \binom{r}{u} 2^{-2C\ell u}.$$

Since $r \le C\ell$, we easily see that $\sum_{u=0}^{r} \binom{r}{u} 2^{-2C\ell u} = (1 + 2^{-C\ell})^r \le \exp\left(r/2^{C\ell}\right) \le 2$, and so

$$P_{\mathrm{red},r} \le 2p^{\binom{r}{2}} \exp\left(-\left(1 + O(D^{-1})\right)\frac{a^3}{p^3\sqrt{d}}\binom{r}{3}\right).$$

This suffices to complete the proof, by absorbing the constant factor 2 into the $O_p(\frac{r^3}{D\sqrt{d}})$ error term in the exponent. $\qquad\square$

4. Changing the perspective

Our goal in this section is to bound the probability that $G \sim G(r,d,p)$ is a clique or an independent set and that the Gaussian vectors $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_r$ corresponding to its vertices form a perfect sequence, i.e. to bound $P^*_{\mathrm{red},r}, P^*_{\mathrm{blue},r}$. In other words, we will prove Proposition 3.1. The first step in doing this will be to change coordinates to a more convenient basis, as follows.

For a sequence of r points $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_r \sim \mathcal{N}(0, \frac{1}{d}I_d)$, let us reveal the subspaces $\mathrm{span}\{\boldsymbol{x}_1\}, \mathrm{span}\{\boldsymbol{x}_1, \boldsymbol{x}_2\},$ $\ldots, \mathrm{span}\{\boldsymbol{x}_1, \ldots, \boldsymbol{x}_r\}$. We may rotate the setup so that $\mathrm{span}\{\boldsymbol{x}_1\}$ is aligned with the first basis vector $\boldsymbol{e}_1$, $\mathrm{span}\{\boldsymbol{x}_1, \boldsymbol{x}_2\}$ is the span of first two basis vectors $\boldsymbol{e}_1, \boldsymbol{e}_2$, and so on, until $\mathrm{span}\{\boldsymbol{x}_1, \ldots, \boldsymbol{x}_r\} = \mathrm{span}\{\boldsymbol{e}_1, \ldots, \boldsymbol{e}_r\}$. After the rotation, the vector $\boldsymbol{x}_i$ will have the first $i - 1$ coordinates distributed as independent $\mathcal{N}(0, \frac{1}{d})$ Gaussians, while the last coordinate will follow the distribution $\sqrt{\frac{1}{d}\chi^2_{d-i+1}}$. Recall that the χ^2_k distribution is defined as the sum of squares of k one-dimensional standard normal random variables. The mentioned rotation can be obtained by applying the Gram-Schmidt process to the sequence $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_r$, and the set of vectors obtained after the rotation is sometimes known as the Bartlett decomposition.

Definition 4.1. Let r be a positive integer with $2 \le r \le d$. For $1 \le i \le r$, let us define a random vector $\boldsymbol{y}_i \in \mathbb{R}^d$ as follows: the first $i - 1$ coordinates of $\boldsymbol{y}_i$ are independent Gaussian $\mathcal{N}(0, \frac{1}{d})$, the i-th coordinate follows the distribution $\sqrt{\frac{1}{d}\chi^2_{d-i+1}}$, and all further coordinates are zero.

Lemma 4.2. Let $d \ge r \ge 1$ be positive integers, let $\boldsymbol{y}_1, \ldots, \boldsymbol{y}_r$ be sampled as above, and let $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_r \sim \mathcal{N}(0, \frac{1}{d}I_d)$ be independent. Then, the collection of inner products $\{\langle \boldsymbol{x}_i, \boldsymbol{x}_j \rangle\}_{1 \le i,j \le r}$ has the same joint distribution as $\{\langle \boldsymbol{y}_i, \boldsymbol{y}_j \rangle\}_{1 \le i,j \le r}$.

In particular, if G' is a random graph on the vertex set $[r]$ where the vertices ij are adjacent if $\langle \boldsymbol{y}_i, \boldsymbol{y}_j \rangle \ge -\frac{c_p}{\sqrt{d}}$, then G' follows the same distribution as the random graph $G(r,d,p)$.

Proof. Let us denote the standard basis of $\mathbb{R}^d$ by $\boldsymbol{e}_1, \ldots, \boldsymbol{e}_d$. Let us sample $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_r \sim \mathcal{N}(0, \frac{1}{d}I_d)$ independently and consider the following rotation. Let O be a $d \times d$ orthogonal matrix such that $\mathrm{span}\{O\boldsymbol{x}_1\} = \mathrm{span}\{\boldsymbol{e}_1\}, \mathrm{span}\{O\boldsymbol{x}_1, O\boldsymbol{x}_2\} = \mathrm{span}\{\boldsymbol{e}_1, \boldsymbol{e}_2\}, \cdots, \mathrm{span}\{O\boldsymbol{x}_1, \ldots, O\boldsymbol{x}_r\} = \mathrm{span}\{\boldsymbol{e}_1, \ldots, \boldsymbol{e}_r\}$. Note that such a matrix exists whenever $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_r$ are linearly independent (which happens with probability 1), and can be explicitly constructed using the Gram-Schmidt algorithm.

Let $\boldsymbol{z}_i = O\boldsymbol{x}_i$, and we claim that $(\boldsymbol{z}_1, \ldots, \boldsymbol{z}_r)$ follows the same distribution as $(\boldsymbol{y}_1, \ldots, \boldsymbol{y}_r)$, which is sufficient to conclude the proof (since O does not change the inner products).

Observe first that $\boldsymbol{z}_i(j) = 0$ when $j > i$, since $\boldsymbol{z}_i$ lies in the span of $\boldsymbol{e}_1, \ldots, \boldsymbol{e}_i$. Further, note that $\pi_{\mathrm{span}\{\boldsymbol{e}_1, \ldots, \boldsymbol{e}_{i-1}\}}(\boldsymbol{z}_i)$ is the same as the projection of $\boldsymbol{x}_i$ to $\mathrm{span}\{\boldsymbol{x}_1, \ldots, \boldsymbol{x}_{i-1}\}$, which is simply a $(i - 1)$-dimensional Gaussian distribution with covariance matrix $\frac{1}{d}I_{i-1}$. Thus, every coordinate of $\pi_{\mathrm{span}\{\boldsymbol{e}_1, \ldots, \boldsymbol{e}_{i-1}\}}(\boldsymbol{z}_i)$ has a Gaussian distribution with variance $\frac{1}{d}$, as needed. Finally, $\boldsymbol{z}_i(i)$ is the residual

length of $\boldsymbol{x}_i$ which does not lie in the subspace $\mathrm{span}\{\boldsymbol{x}_1, \ldots, \boldsymbol{x}_{i-1}\}$. However, since the Gaussian distribution is rotationally invariant, this is simply the length of a $(d-i+1)$-dimensional Gaussian, which is distributed as $\sqrt{\frac{1}{d}\chi^2_{d-i+1}}$. This completes the proof. $\qquad\square$

Although we have initially defined vectors $\boldsymbol{y}_1, \ldots, \boldsymbol{y}_r$ as elements of $\mathbb{R}^d$, they are always zero on coordinates $r+1, \ldots, d$, and so we might as well write $\boldsymbol{y}_1, \ldots, \boldsymbol{y}_r \in \mathbb{R}^r$. The main advantage of working in this new basis is that it simplifies the calculation of the inner products. More precisely, for $i < j$, we have

$$\langle \boldsymbol{y}_i, \boldsymbol{y}_j \rangle = \boldsymbol{y}_j(i)\boldsymbol{y}_i(i) + \sum_{k=1}^{i-1} \boldsymbol{y}_i(k)\boldsymbol{y}_j(k).$$

Recall that $\boldsymbol{y}_i(k)$ denotes the k-th coordinate of $\boldsymbol{y}_i$. Since each coordinate $\boldsymbol{y}_i(k), \boldsymbol{y}_j(k)$ is expected to have size roughly $1/\sqrt{d}$ and $\boldsymbol{y}_i(i) \approx 1$, we see that the leading term of the above expression is $\boldsymbol{y}_j(i)$. In other words, whether the edge ij is present or not is essentially determined by $\boldsymbol{y}_j(i)$, with small corrections coming from the other coordinates $\boldsymbol{y}_i(k), \boldsymbol{y}_j(k)$. Throughout the proof, we find it useful to arrange the vectors $\boldsymbol{y}_1, \ldots, \boldsymbol{y}_r$ into a lower-triangular matrix M whose rows are $\boldsymbol{y}_1, \ldots, \boldsymbol{y}_r$.

$\boldsymbol{y}_1(1)$			
$\boldsymbol{y}_2(1)$	$\boldsymbol{y}_2(2)$		
$\boldsymbol{y}_3(1)$	$\boldsymbol{y}_3(2)$	$\boldsymbol{y}_3(3)$	
$\boldsymbol{y}_4(1)$	$\boldsymbol{y}_4(2)$	$\boldsymbol{y}_4(3)$	$\boldsymbol{y}_4(4)$

Since $\boldsymbol{y}_1, \ldots, \boldsymbol{y}_r$ is obtained from $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_r$ by a rotation, the notion of a perfect sequence remains unchanged. In particular, this means that each vector $\boldsymbol{y}_i$ has length approximately 1, and that the most of this length comes from the diagonal entry $\boldsymbol{y}_i(i)$. More precisely, we have the following deterministic statement.

Claim 4.3. *If $\boldsymbol{y}_1, \ldots, \boldsymbol{y}_r \in \mathbb{R}^r$ is a perfect sequence, then $M_{i,i} = \boldsymbol{y}_i(i) \in (1 - 2\delta, 1 + \delta)$ for all $1 \leq i \leq r$.*

Proof. We have $M_{i,i}^2 = \boldsymbol{y}_i(i)^2 = \|\boldsymbol{y}_i\|_2^2 - \|\pi_{i-1}(\boldsymbol{y}_i)\|_2^2 \geq (1-\delta)^2 - \frac{\alpha^2\ell}{d}$, where we have used that $\boldsymbol{y}_1, \ldots, \boldsymbol{y}_r$ is perfect in the last inequality. Since $\delta = \alpha d^{-1/4}$, we have $\alpha^2\ell/d \leq \alpha^2/\sqrt{d} = \delta^2$, and so $M_{i,i}^2 \geq (1-\delta)^2 - \delta^2 = 1 - 2\delta$, implying that $M_{i,i} \geq 1 - 2\delta$. On the other hand, we have $M_{i,i} \leq \|\boldsymbol{y}_i\| \leq 1 + \delta$. $\quad\square$

The entries on the diagonal will not play an important role in our proof of Proposition 3.1. Thus, we will condition on the specific outcome of $M_{1,1}, \ldots, M_{r,r}$ and assume each diagonal entry is of size $1 \pm 2\delta$. Furthermore, for the sake of shortening the formulas, we will not specify the conditioning on the diagonal entries explicitly.

The main idea in the proof of Proposition 3.1 is to bound the probability that $\boldsymbol{y}_{s+1}, \ldots, \boldsymbol{y}_r$ is a clique or an independent set, for every $0 \leq s \leq r-1$, conditionally on the first s columns of M (note that if we condition on these columns, the connections between vectors $\boldsymbol{y}_1, \ldots, \boldsymbol{y}_s$ and all other vectors are already determined). The bound will then be proved by induction on s, starting from $s = r-1$ and going all the way down to $s = 0$. However, before we do that, let us introduce some terminology which will be used in the proof.

For $1 \leq i < j \leq r$, let E_{ij} be the event that $\langle \boldsymbol{y}_i, \boldsymbol{y}_j \rangle \geq -c_p/\sqrt{d}$, and let $\overline{E}_{ij}$ be its complement, i.e. the event that $\langle \boldsymbol{y}_i, \boldsymbol{y}_j \rangle < -c_p/\sqrt{d}$. Let $C_s = \bigwedge_{s < i < j \leq r} E_{ij}$ be the event that the vertices $s+1, \ldots, r$ induce a clique in $G(r, d, p)$ and let $I_s = \bigwedge_{s < i < j \leq r} \overline{E}_{ij}$ be the event they induce an independent set. Finally, B_r denotes the event that $\boldsymbol{y}_1, \ldots, \boldsymbol{y}_r$ is a perfect sequence, and that this event implies that $\|\pi_i(\boldsymbol{y}_{i+1})\|_2 \leq \frac{\alpha\sqrt{\ell}}{\sqrt{d}}$ for all $1 \leq i \leq r-1$.

Also, for $1 \leq s \leq i \leq r$, we write $\pi_s(\boldsymbol{y}_i)$ for the projection of $\boldsymbol{y}_i$ to the first s coordinates (which is a shorthand for $\pi_{\mathrm{span}\{\boldsymbol{y}_1, \ldots, \boldsymbol{y}_s\}}(\boldsymbol{y}_i)$). Further, let M_i denote the set of entries in the i-th column of the matrix M strictly below the diagonal.

Proposition 4.4. *Let $C > 1, p \in (0,1)$ be real numbers and $\ell, d = D^2\ell^2$ be large positive integers, where D is sufficiently larger than C. Also, let s, r be nonnegative integers such that $1 \leq s + 1 \leq r \leq C\ell$. Suppose that the first s columns of M are fixed, and denote them by $M[s] = (M_1, \ldots, M_s)$. Then, the probabilities over the random choice of the remaining columns $M_{s+1}, \ldots, M_r$ of the events $C_s \wedge B_r, I_s \wedge B_r$ are at most*

$$(8) \quad \mathbb{P}\Big[I_s \wedge B_r \Big| M[s]\Big] \leq p^{\binom{r-s}{2}} \exp\left(-\frac{a\sqrt{d}}{p} \sum_{s < i < j \leq r} \langle \pi_s(\boldsymbol{y}_i), \pi_s(\boldsymbol{y}_j)\rangle - \frac{a^3}{p^3\sqrt{d}}\binom{r-s}{3} + O\Big(\frac{(r-s)^4}{d} + (r-s)\Big)\right)$$

$$(9) \quad \mathbb{P}\Big[C_s \wedge B_r \Big| M[s]\Big] \leq (1-p)^{\binom{r-s}{2}} \exp\left(\frac{a\sqrt{d}}{1-p} \sum_{s < i < j \leq r} \langle \pi_s(\boldsymbol{y}_i), \pi_s(\boldsymbol{y}_j)\rangle + \frac{a^3}{(1-p)^3\sqrt{d}}\binom{r-s}{3}\right.$$
$$\left. + O\Big(\frac{(r-s)^4}{d} + (r-s)\Big)\right).$$

Proof. We argue by reverse induction on s. Note that for $s = r - 1$, the statement is vacuous, since there the whole matrix is revealed and the product on the right hand side of Equation (8,9) is empty. Thus, we may use this as a basis of the induction. Let us first focus on bounding the probability of $C_s \wedge B_r$.

Suppose now that the statement holds for s and let us prove it for $s - 1$, where only the columns $M[s-1] = (M_1, \ldots, M_{s-1})$ are fixed. We will reduce the inductive step to two auxiliary claims, which we will then prove separately. To do this, we sample the column M_s by setting all of its entries below the diagonal to be an independent Gaussian $\mathcal{N}(0, \frac{1}{d})$. This is sufficient to determine whether or not the vertex s is connected to the vertices $s + 1, \ldots, r$. Let us begin by estimating this probability in our first auxiliary claim.

Claim 4.5. *Given $M_1, \ldots, M_{s-1}$, the probability (over the random choice of M_s) that s is connected to all of $s + 1, \ldots, r$ is at most*

$$(10) \quad \mathbb{P}\left[\bigwedge_{s < i \leq r} E_{s,i} \Big| M[s-1]\right] \leq (1-p)^{r-s} \exp\left(\frac{a\sqrt{d}}{1-p} \sum_{i=s+1}^{r} \langle \pi_{s-1}(\boldsymbol{y}_i), \pi_{s-1}(\boldsymbol{y}_s)\rangle + O\big((r-s)\delta\big)\right).$$

Also, the probability that s is not connected to any of $s + 1, \ldots, r$ is at most

$$(11) \quad \mathbb{P}\left[\bigwedge_{s < i \leq r} \overline{E}_{s,i} \Big| M[s-1]\right] \leq p^{r-s} \exp\left(-\frac{a\sqrt{d}}{p} \sum_{i=s+1}^{r} \langle \pi_{s-1}(\boldsymbol{y}_i), \pi_{s-1}(\boldsymbol{y}_s)\rangle + O\big((r-s)\delta\big)\right).$$

Although the proof of Claim 4.5 is quite simple, we postpone it for later. For each outcome of M_s, the induction hypothesis provides a bound on the probability of $C_s \wedge B_r$ over the random choice of $M_{s+1}, \ldots, M_r$. Thus, in order to perform an induction step, we will integrate this bound over all outcomes of M_s satisfying the event $C_s \wedge B_r$. At this point, it is important to observe that $C_{s-1} = C_s \wedge \bigwedge_{s < i \leq r} E_{si}$, i.e. that vertices $s, \ldots, r$ form a clique precisely when vertices $s+1, \ldots, r$ form a clique and s is connected to all vertices among $s + 1, \ldots, r$. Hence, by the law of total probability, we can write

$$(12) \quad \mathbb{P}\Big[C_{s-1} \wedge B_r \Big| M[s-1]\Big] = \mathbb{E}_{M_s}\left[\mathbb{P}\Big[C_s \wedge B_r \Big| M_s\Big] \Big| \bigwedge_{s < i \leq r} E_{s,i}, M[s-1]\right] \mathbb{P}\left[\bigwedge_{s < i \leq r} E_{s,i} \Big| M[s-1]\right].$$

Since we have already estimated the latter term, we focus on the first one. We use the induction hypothesis to upper bound $\mathbb{P}\Big[C_s \wedge B_r \Big| M_s\Big]$ as follows

$$(13) \quad \mathbb{P}\Big[C_s \wedge B_r \Big| M_s\Big] \leq (1-p)^{\binom{r-s}{2}} \exp\left(\frac{a\sqrt{d}}{1-p} \sum_{s < i < j \leq r} \langle \pi_s(\boldsymbol{y}_i), \pi_s(\boldsymbol{y}_j)\rangle + \frac{a^3}{(1-p)^3\sqrt{d}}\binom{r-s}{3} + O\Big(\frac{(r-s)^4}{d}\Big)\right).$$

Note that most terms on the right-hand side are independent of the entries of M_s. In fact, the only terms which depend on it are $\exp\left(\frac{a\sqrt{d}}{1-p} \sum_{s < i < j \leq r} \boldsymbol{y}_i(s)\boldsymbol{y}_j(s)\right)$. Thus, we focus on bounding their expectation.

Claim 4.6. *If the entries of M_s are sampled as Gaussians $\mathcal{N}(0, \frac{1}{d})$, then we have*

(14)
$$
\mathbb{E}_{M_s}\left[\exp\left(\frac{a\sqrt{d}}{1-p}\sum_{s<i<j\leq r}\boldsymbol{y}_i(s)\boldsymbol{y}_j(s)\right)\Big|\bigwedge_{s<i\leq r}E_{s,i}, M[s-1]\right] \leq \exp\left(\frac{a^3}{(1-p)^3\sqrt{d}}\binom{r-s}{2}+O\Big(\frac{(r-s)^3}{d}+1\Big)\right)
$$

(15)
$$
\mathbb{E}_{M_s}\left[\exp\left(-\frac{a\sqrt{d}}{p}\sum_{s<i<j\leq r}\boldsymbol{y}_i(s)\boldsymbol{y}_j(s)\right)\Big|\bigwedge_{s<i\leq r}\overline{E}_{s,i}, M[s-1]\right] \leq \exp\left(-\frac{a^3}{p^3\sqrt{d}}\binom{r-s}{2}+O\Big(\frac{(r-s)^3}{d}+1\Big)\right)
$$

To complete the induction step using the two auxiliary claims, we perform some simple but tedious bookkeeping. By plugging in the bound (14) into (13), we obtain

$$
\mathbb{E}_{M_s}\left[\mathbb{P}\left[C_s \wedge B_r\big|M_s\right]\Big|\bigwedge_{s<i\leq r}E_{s,i}, M[s-1]\right] \leq
$$

$$
\leq (1-p)^{\binom{r-s}{2}}\exp\left(\frac{a\sqrt{d}}{1-p}\sum_{s<i<j\leq r}\langle\pi_{s-1}(\boldsymbol{y}_i), \pi_{s-1}(\boldsymbol{y}_j)\rangle+\frac{a^3}{(1-p)^3\sqrt{d}}\binom{r-s}{3}+O\Big(\frac{(r-s)^4}{d}+(r-s)\Big)\right)
$$

$$
\cdot \exp\left(\frac{a^3}{(1-p)^3\sqrt{d}}\binom{r-s}{2}+O\Big(\frac{(r-s)^3}{d}+1\Big)\right)
$$

$$
\leq (1-p)^{\binom{r-s}{2}}\exp\left(\frac{a\sqrt{d}}{1-p}\sum_{s<i<j\leq r}\langle\pi_{s-1}(\boldsymbol{y}_i), \pi_{s-1}(\boldsymbol{y}_j)\rangle+\frac{a^3}{(1-p)^3\sqrt{d}}\binom{r-s+1}{3}+O\Big(\frac{(r-s+1)^4}{d}+(r-s+1)\Big)\right).
$$

By plugging into the equation (12) the bound on $\mathbb{P}\left[\bigwedge_{s<i\leq r}E_{s,i}\big|M[s-1]\right]$ from (10) and the bound of (13), we obtain

$$
\mathbb{P}\left[C_{s-1}\wedge B_r\big|M[s-1]\right] \leq (1-p)^{\binom{r-s}{2}}\exp\left(\frac{a\sqrt{d}}{1-p}\sum_{s<i<j\leq r}\langle\pi_{s-1}(\boldsymbol{y}_i), \pi_{s-1}(\boldsymbol{y}_j)\rangle+\frac{a^3}{(1-p)^3\sqrt{d}}\binom{r-s+1}{3}\right.
$$

$$
\left.+O\Big(\frac{(r-s+1)^4}{d}\Big)\right)
$$

$$
\cdot (1-p)^{r-s}\exp\left(\frac{a\sqrt{d}}{1-p}\sum_{i=s+1}^{r}\langle\pi_{s-1}(\boldsymbol{y}_i), \pi_{s-1}(\boldsymbol{y}_s)\rangle+O\big((r-s)\delta\big)\right)
$$

$$
\leq (1-p)^{\binom{r-s+1}{2}}\exp\left(\frac{a\sqrt{d}}{1-p}\sum_{s\leq i<j\leq r}\langle\pi_{s-1}(\boldsymbol{y}_i), \pi_{s-1}(\boldsymbol{y}_j)\rangle+\frac{a^3}{(1-p)^3\sqrt{d}}\binom{r-s+1}{3}+O\Big(\frac{(r-s+1)^4}{d}\Big)\right).
$$

A quick inspection shows that this is precisely the bound we need on $\mathbb{P}\left[C_{s-1}\wedge B_r\big|M[s-1]\right]$, and thus our induction step is complete. Let us conclude by noting that nothing in this bookkeeping relied on the fact that we are estimating the probability of $C_s \wedge B_r$. In fact, an analogous proof would apply unchanged to estimate the probability of $I_s \wedge B_r$, and therefore we choose not to spell it out explicitly. Let us now finish the discussion of Proposition 4.4 by proving the auxiliary claims.

Proof of Claim 4.5. Since the proofs of both bounds are analogous, we only show the first one. Recall that $E_{s,i}$ holds whenever $\boldsymbol{y}_i(s)\boldsymbol{y}_s(s) \geq -c_p/\sqrt{d} - \langle\pi_{s-1}(\boldsymbol{y}_i), \pi_{s-1}(\boldsymbol{y}_s)\rangle$. Note that the only random variable in this equation is $\boldsymbol{y}_i(s)$, since $\boldsymbol{y}_s(s)$ was fixed and all of the other ones are determined by $M[s-1]$. Since the variables $\boldsymbol{y}_{s+1}(s), \ldots, \boldsymbol{y}_r(s)$ are independent, so are the events $E_{s,s+1}, \ldots, E_{s,r}$.

Thus, it is sufficient to compute the probabilities of the individual events $E_{s,i}$. Thus,

$$
\mathbb{P}\left[E_{s,i}|M[s-1]\right] = \mathbb{P}\left[\boldsymbol{y}_i(s) \geq -\frac{1}{\boldsymbol{y}_s(s)}\Big(c_p/\sqrt{d}+\langle\pi_{s-1}(\boldsymbol{y}_i), \pi_{s-1}(\boldsymbol{y}_s)\rangle\Big)\Big|M[s-1]\right].
$$

Recall that $\boldsymbol{y}_s(s) \in (1-2\delta, 1+\delta)$, and note that by Cauchy-Schwarz inequality $|\langle \pi_{s-1}(\boldsymbol{y}_i), \pi_{s-1}(\boldsymbol{y}_s) \rangle| \leq \|\pi_{s-1}(\boldsymbol{y}_i)\| \cdot \|\pi_{s-1}(\boldsymbol{y}_s)\| \leq \frac{\alpha^2 \ell}{d} = O(\frac{1}{D\sqrt{d}})$. Therefore the cutoff for $E_{s,i}|M[s-1]$ to hold equals

$$b_i = -c_p/\sqrt{d} - \langle \pi_{s-1}(\boldsymbol{y}_i), \pi_{s-1}(\boldsymbol{y}_s) \rangle + O(\delta/\sqrt{d}).$$

Since $\boldsymbol{y}_i(s) \sim \mathcal{N}(0, \frac{1}{d})$, log-concavity of Φ (see (4)) implies that

$$\mathbb{P}[\boldsymbol{y}_i(s) \geq b_i] = \mathbb{P}[\mathcal{N}(0,1) \leq -\sqrt{d}b_i] = \mathbb{P}[\mathcal{N}(0,1) \leq c_p] \exp\left(\frac{\phi(c_p)}{\Phi(c_p)}(-\sqrt{d}b_i - c_p)\right).$$

In our setup, we have $\mathbb{P}[\mathcal{N}(0,1) \leq c_p] = \Phi(c_p) = 1 - p$ and $\phi(c_p) = a$. Thus, we have

$$\mathbb{P}[E_{s,i}] = \mathbb{P}[\boldsymbol{y}_i(s) \geq b_i] \leq (1-p) \exp\left(\frac{a\sqrt{d}}{1-p} \langle \pi_{s-1}(\boldsymbol{y}_i), \pi_{s-1}(\boldsymbol{y}_s) \rangle + O(\delta)\right),$$

where we have used that $-\sqrt{d}b_i - c_p = \pi_{s-1}(\boldsymbol{y}_i), \pi_{s-1}(\boldsymbol{y}_s)\rangle + O(\delta)$. Multiplying over all $s < i \leq r$, we obtain

$$\prod_{s<i\leq r} \mathbb{P}\left[E_{s,i}\Big|M[s-1]\right] \leq (1-p)^{r-s} \exp\left(\frac{a\sqrt{d}}{1-p} \sum_{i=s+1}^{r} \langle \pi_{s-1}(\boldsymbol{y}_i), \pi_{s-1}(\boldsymbol{y}_s) \rangle + O((r-s)\delta)\right),$$

which is sufficient to complete the proof due to the independence of the events $E_{s,i}$. $\square$

Proof of Claim 4.6. Again, we will focus on proving (14), since the proof of (15) is analogous. Recall that $E_{s,i}$ holds whenever $\boldsymbol{y}_i(s)\boldsymbol{y}_s(s) \geq -c_p/\sqrt{d} - \langle \pi_{s-1}(\boldsymbol{y}_i), \pi_{s-1}(\boldsymbol{y}_s) \rangle$. Since the diagonal entries $\boldsymbol{y}_s(s)$ are fixed, the events $E_{s,i}$ are independent and depend only on $\boldsymbol{y}_i(s)$. So, conditioning on their intersection leaves the variables $\boldsymbol{y}_{s+1}(s), \ldots, \boldsymbol{y}_r(s)$ independent. Moreover, conditioning on the event $E_{s,i}$, each $\boldsymbol{y}_i(s)$ follows the distribution of a lower truncated Gaussian with the cutoff $b_i = (-c_p/\sqrt{d} - \langle \pi_{s-1}(\boldsymbol{y}_i), \pi_{s-1}(\boldsymbol{y}_s) \rangle)/\boldsymbol{y}_s(s)$.

As we have observed already, truncated Gaussian random variables have variance proxy at most $1/d$, and so we can apply Lemma 2.6 with $\lambda = \frac{a\sqrt{d}}{1-p}$, $k = r - s$, random variables $\boldsymbol{y}_{s+1}(s), \ldots, \boldsymbol{y}_r(s)$ and $S = \sum_{s<i<j\leq r} \boldsymbol{y}_i(s)\boldsymbol{y}_j(s)$ to obtain

$$\mathbb{E}[e^{\lambda S}] \leq \exp\left(\lambda\mathbb{E}[S] + \frac{\lambda^2 k^2}{d} \sum_{j=s+1}^{r} (\mathbb{E}\boldsymbol{y}_j(s))^2 + \frac{4\lambda k}{d}\right).$$

Here, it is important to verify that $d \geq 4\lambda k$, which follows since $d = D^2\ell^2$, $\lambda = aD\ell/(1-p)$ and $k = C\ell$, reducing the inequality to $D\ell \geq aC\ell/(1-p)$, which holds since D is sufficiently larger than C. Hence, in order to complete the proof, we only need to verify that

$$\lambda\mathbb{E}[S] + \frac{\lambda^2 k^2}{d} \sum_{j=s+1}^{r} (\mathbb{E}\boldsymbol{y}_j(s))^2 + \frac{4\lambda k}{d} \leq \frac{a}{(1-p)^3\sqrt{d}}\binom{r-s}{2} + O\left(\frac{(r-s)^2\delta}{\sqrt{d}} + \frac{(r-s)^3}{d} + 1\right).$$

The last term on the left hand side is clearly $O(1)$, since $4\lambda k/d \leq 1$.

Using the independence of $\boldsymbol{y}_{s+1}(s), \ldots, \boldsymbol{y}_r(s)$, we can upper bound $\mathbb{E}[S]$ as follows

$$\mathbb{E}[S] = \sum_{s<i<j\leq r} \mathbb{E}[\boldsymbol{y}_i(s)]\mathbb{E}[\boldsymbol{y}_j(s)] \leq \frac{1}{2}\left(\sum_{s<i\leq r} \mathbb{E}[\boldsymbol{y}_i(s)]\right)^2 \leq \frac{r-s}{2} \sum_{s<i\leq r} \mathbb{E}[\boldsymbol{y}_i(s)]^2,$$

where we have used the inequality between the arithmetic and the quadratic mean in the last step. Thus, we have

$$\lambda\mathbb{E}[S] + \frac{\lambda^2 k^2}{d} \sum_{j=s+1}^{r} (\mathbb{E}\boldsymbol{y}_j(s))^2 \leq \lambda\frac{r-s}{2}\left(1 + \frac{2\lambda k}{d}\right) \sum_{i=s+1}^{r} \mathbb{E}[\boldsymbol{y}_i(s)]^2.$$

We now need to calculate the expectations of $\boldsymbol{y}_i(s)$, which we do using Lemma 2.4. Let us recall that $\boldsymbol{y}_i(s)$ is a lower truncated Gaussian with the cutoff $b_i = -(c_p/\sqrt{d} + \langle \pi_{s-1}(\boldsymbol{y}_i), \pi_{s-1}(\boldsymbol{y}_s) \rangle)/\boldsymbol{y}_s(s)$. If we write $b_i = -c_p/\sqrt{d} + \varepsilon_i$, then $|\varepsilon_i| = |b_i + c_p/\sqrt{d}| = O(\frac{1}{D\sqrt{d}} + \frac{\delta}{\sqrt{d}}) = O(\frac{1}{D\sqrt{d}})$. The last inequality hold

since $|\langle \pi_{s-1}(\boldsymbol{y}_i), \pi(\boldsymbol{y}_j)\rangle| \leq \frac{\alpha^2 \ell}{d} \leq O(\frac{1}{D\sqrt{d}})$ and since $\boldsymbol{y}_s(s)$ is fixed and in $(1-2\delta, 1+\delta)$. Then, Lemma 2.4 gives

$$\mathbb{E}[\boldsymbol{y}_i(s)|E_{s,i}] = \frac{\phi(c_p)}{\Phi(c_p)\sqrt{d}} + O(|\varepsilon_i|) = \frac{a}{(1-p)\sqrt{d}} + O\left(\frac{1}{D\sqrt{d}}\right).$$

Recalling that $\lambda = \frac{a\sqrt{d}}{1-p}$ and $\frac{r-s}{\sqrt{d}} \leq \frac{C\ell}{D\ell} \leq O(\frac{1}{D})$, we get

$$\lambda \frac{r-s}{2}\left(1 + \frac{2\lambda(r-s)}{d}\right) \sum_{i=s+1}^{r} \mathbb{E}[\boldsymbol{y}_i(s)]^2 \leq \frac{a\sqrt{d}(r-s)}{2(1-p)}\left(1 + O\left(\frac{r-s}{\sqrt{d}}\right)\right) \sum_{i=s+1}^{r} \left[\frac{a^2}{(1-p)^2 d} + O\left(\frac{1}{Dd}\right)\right]$$

$$\leq \frac{a^3}{(1-p)^3\sqrt{d}}\binom{r-s}{2}\left(1 + O(\frac{1}{D})\right),$$

which is what we needed to show. $\qquad \square$

$\qquad \square$

The only remaining step is to prove Proposition 3.1 from Proposition 4.4.

Proof of Proposition 3.1. By plugging in $s = 0$ in Proposition 4.4, we find that

$$\mathbb{P}[\boldsymbol{x}_1, \ldots, \boldsymbol{x}_r \text{ form a red clique and are perfect}] \leq p^{\binom{r}{2}} \exp\left(-\frac{a^3}{p^3\sqrt{d}}\binom{r}{3} + O\left(\frac{r^4}{d} + r\right)\right)$$

$$= p^{\binom{r}{2}} \exp\left(-\frac{a^3}{p^3\sqrt{d}}\binom{r}{3}\left(1 + O\left(\frac{1}{D}\right)\right)\right),$$

since $\frac{r}{\sqrt{d}} \leq O(1/D)$. Similarly,

$$\mathbb{P}[\boldsymbol{x}_1, \ldots, \boldsymbol{x}_r \text{ form a blue clique and are perfect}] \leq (1-p)^{\binom{r}{2}} \exp\left(\frac{a^3}{(1-p)^3\sqrt{d}}\binom{r}{3} + O\left(\frac{r^4}{d} + r\right)\right)$$

$$= (1-p)^{\binom{r}{2}} \exp\left(\frac{a^3}{(1-p)^3\sqrt{d}}\binom{r}{3}\left(1 + O\left(\frac{1}{D}\right)\right)\right).$$

This completes the proof. $\qquad \square$

REFERENCES

[1] P. Balister, B. Bollobás, M. Campos, S. Griffiths, E. Hurley, R. Morris, J. Sahasrabudhe and M. Tiba. *Upper bounds for multicolour Ramsey numbers.* arXiv:2410.17197.

[2] M. Barreto, O. Marchal, and J. Arbel. *Optimal sub-Gaussian variance proxy for truncated Gaussian and exponential random variables* arXiv:2403.08628.

[3] S. Bubeck, J. Ding, R. Eldan and M.Z. Rácz. *Testing for high-dimensional geometry in random graphs.* Random Structures Algorithms, 49 (2016), 503–532.

[4] M. Campos, S. Griffiths, R. Morris and J. Sahasrabudhe. *An exponential improvement for diagonal Ramsey.* to appear in Ann. of Math.

[5] M. Campos, M. Jenssen, M. Michelen and J. Sahasrabudhe. *A new lower bound for the Ramsey numbers $R(3, k)$.* arXiv:2505.13371.

[6] D. Conlon. *A new upper bound for diagonal Ramsey numbers.* Ann. of Math., 170 (2009), 941–960.

[7] D. Conlon and A. Ferber. *Lower bounds for multicolour Ramsey numbers.* Adv. Math., 378 (2021), Paper No. 107528, 5 pp.

[8] L. Devroye, A. György, G. Lugosi and F. Udina. *High-dimensional random geometric graphs and their clique number.* Electron. J. Probab., 16 (2011), Paper no. 90, 2481–2508.

[9] P. Erdős. *Some remarks on the theory of graphs.* Bull. Amer. Math. Soc., 53 (1947), 292–294.

[10] P. Erdős and G. Szekeres. *A combinatorial problem in geometry.* Compos. Math., 2 (1935), 463–470.

[11] R.D. Gordon. *Values of Mills' ratio of area to bounding ordinate and of the normal probability integral for large values of the argument.* Ann. Math. Statist., 12 (1941), 364–366.
[12] P. Gupta, N. Ndiaye, S. Norin and L. Wei. *Optimizing the CGMS upper bound on Ramsey numbers.* arXiv:2407.19026.
[13] Z. Hefty, P. Horn, D. King and F. Pfender. *Improving $R(3,k)$ in just two bites.* arXiv:2510.19718.
[14] B. Laurent and P. Massart. *Adaptive estimation of a quadratic functional by model selection* Ann. Stat., 28 (2000): 1302–1338.
[15] J. Ma, W. Shen and S. Xie. *An exponential improvement for Ramsey lower bounds.* arXiv:2507.12926.
[16] S. Mattheus and J. Verstraëte. *The asymptotics of $r(4,t)$.* Ann. of Math., 199 (2024), 919–941.
[17] D. Mubayi and J. Verstraëte. *A note on pseudorandom Ramsey graphs.* J. Eur. Math. Soc., 26 (2024) 153–161.
[18] M. Penrose. *Random Geometric Graphs.* Oxford Studies in Probability, 5, Oxford University Press, 2003.
[19] F.P. Ramsey. *On a Problem of Formal Logic.* Proc. London Math. Soc., 30 (1930), 264–286.
[20] A. Sah. *Diagonal Ramsey via effective quasirandomness.* Duke Math. J., 172 (2023), 545–567.
[21] J. Spencer. *Ramsey's theorem—A new lower bound.* J. Combin. Theory Ser. A, 18 (1975), 108–115.
[22] A. Thomason. *An upper bound for some Ramsey numbers.* J. Graph Theory, 12 (1988), 509–517.
[23] R. Vershynin. *High-Dimensional Probability: An Introduction with Applications in Data Science.* Cambridge Series in Statistical and Probabilistic Mathematics, 47, Cambridge University Press, 2018.

Appendix A. Proofs of Preliminaries

Proof of Lemma 2.1. If $\|\boldsymbol{x}\|_2 \notin (1-\delta, 1+\delta)$, then we also have $\|\boldsymbol{x}\|_2^2 \notin (1-\delta, 1+\delta)$. As we have already observed, if we define $Y = d\|\boldsymbol{x}\|_2^2$ we have $Y \sim \chi_d^2$. Thus, we can write

$$\mathbb{P}\big[\|\boldsymbol{x}\|_2 \notin (1-\delta, 1+\delta)\big] \leq \mathbb{P}\big[\|\boldsymbol{x}\|_2^2 \notin (1-\delta, 1+\delta)\big] = \mathbb{P}\Big[Y \notin \big((1-\delta)d, (1+\delta)d\big)\Big].$$

By Laurent-Massart inequality we have $\mathbb{P}[|Y - d| \geq 2\sqrt{dt} + 2t] \leq 2e^{-t}$, and plugging in $t = \delta^2 d/10$, we obtain

$$\mathbb{P}\big[\|\boldsymbol{x}\|_2 \notin (1-\delta, 1+\delta)\big] = \mathbb{P}\Big[|Y - d| \geq \delta d\Big] \leq \mathbb{P}\Big[|Y - d| \geq 2\sqrt{dt} + 2t\Big] \leq 2e^{-t} = 2\exp(-\delta^2 d/10),$$

where we have used that $2\sqrt{dt} + 2t = 2\frac{\delta d}{\sqrt{10}} + 2\frac{\delta^2 d}{10} \leq \delta d$ since $\delta < 1$. $\qquad\square$

Proof of Lemma 2.3. Since the normal distribution $\mathcal{N}(0, \frac{1}{d}I_d)$ is invariant under the action of the orthogonal group, i.e. under rotations, we may assume that W is simply the subspace spanned by the first s coordinate vectors $\boldsymbol{e}_1, \ldots, \boldsymbol{e}_s$.

It is then easy to see that $\pi_W(\boldsymbol{x})$ follows the distribution $\mathcal{N}(0, \frac{1}{d}I_s)$ inside W. Hence, $\sqrt{d}\pi_W(\boldsymbol{x}) \sim \mathcal{N}(0, I_s)$ and we can apply Theorem 2.2 to obtain a tail bound on $\|\sqrt{d}\pi_W(\boldsymbol{x})\|_2^2 \sim \chi_s^2$. In particular, we have

$$\mathbb{P}\Big[\|\pi_W(\boldsymbol{x})\|_2 \geq \frac{\alpha\sqrt{\ell}}{\sqrt{d}}\Big] = \mathbb{P}\Big[\big\|\sqrt{d}\pi_W(\boldsymbol{x})\big\|_2 \geq \alpha\sqrt{\ell}\Big] \leq \mathbb{P}\Big[\chi_s^2 \geq \alpha\ell\Big] \leq \mathbb{P}\Big[\chi_s^2 > s + \frac{\alpha\ell}{2}\Big].$$

Picking $t = \frac{\alpha\ell}{8}$, we get $2\sqrt{st} + 2t = 2\sqrt{\frac{\alpha\ell s}{8}} + 2\frac{\alpha\ell}{8} \leq \frac{\alpha\ell}{2}$, where the inequality follows from the choice of $\alpha > 8C$ and $s \leq C\ell$. Hence, by the Laurent-Massart inequality, we get

$$\mathbb{P}\Big[\|\pi_W(\boldsymbol{x})\|_2 \geq \frac{\alpha\sqrt{\ell}}{\sqrt{d}}\Big] \leq \mathbb{P}\Big[\chi_s^2 > s + 2\sqrt{st} + 2t\Big] \leq \exp(-\alpha\ell/8) \leq \Big(\frac{p}{10}\Big)^{10C\ell},$$

due to the definition of α. $\qquad\square$

Proof of Lemma 2.4. We will start by proving the exact formulas, when the cutoff is exactly b. Let $Z = \sqrt{d}X$ be the standard Gaussian. Observe that the function $\phi(z)$ satisfies $\phi'(z) = -ze^{-z^2/2}/\sqrt{2\pi} = -z\phi(z)$. Hence, for any $t \in \mathbb{R}$, we can write $\int_{-\infty}^{t} z\phi(z)\,dz = -\phi(t)$. We can now compute the expectations of the truncated Gaussians:

$$\mathbb{E}[X|X \leq b/\sqrt{d}] = \frac{1}{\sqrt{d}}\mathbb{E}\Big[Z \,\Big|\, Z \leq b\Big] = \frac{1}{\sqrt{d}} \cdot \frac{\int_{-\infty}^{b} z\phi(z)\,dz}{\mathbb{P}[Z \leq b]} = \frac{1}{\sqrt{d}} \cdot \frac{-\phi(b)}{\Phi(b)}.$$

To compute the other expectation, we can symmetrically write $\int_t^\infty z\phi(z)\,dz = \phi(t)$. Then,

$$\mathbb{E}[X|X \geq b/\sqrt{d}] = \frac{1}{\sqrt{d}}\mathbb{E}\left[Z \mid Z \geq b\right] = \frac{1}{\sqrt{d}} \cdot \frac{\int_b^\infty z\phi(z)\,dz}{\mathbb{P}[Z \geq b]} = \frac{1}{\sqrt{d}} \cdot \frac{\phi(b)}{1-\Phi(b)}.$$

To prove the second part of the statement, we define the function $f(t) = \mathbb{E}[X|X \leq t/\sqrt{d}]$. We have already shown that $f(t) = -\frac{\phi(t)}{\Phi(t)\sqrt{d}}$. Moreover, note that $f'(t) = \frac{t\phi(t)\Phi(t)+\phi(t)^2}{\Phi(t)^2\sqrt{d}}$ is a function bounded by $O(1/\sqrt{d})$ on the real line. This is because $\sqrt{d}f'(t) = \frac{\phi(t)}{\Phi(t)}\left(t + \frac{\phi(t)}{\Phi(t)}\right)$ is a smooth function on the real line with $\lim_{t\to\infty}\sqrt{d}f'(t) = 0$ and

$$2 \geq \left(|t| + \frac{1}{|t|}\right)\frac{1}{|t|} \geq \frac{\phi(t)}{\Phi(t)}\left(t + \frac{\phi(t)}{\Phi(t)}\right) \geq 0,$$

for all $t \leq -1$ due to (3). Finally, by the Mean value theorem, we conclude that $f(b+\varepsilon) = f(b) + f'(c) \cdot \varepsilon$ for some $c \in (b, b+\varepsilon)$. From this, we have

$$\mathbb{E}\left[X|X \leq \frac{b+\varepsilon}{\sqrt{d}}\right] = \mathbb{E}\left[X|X \leq \frac{b}{\sqrt{d}}\right] + f'(c)\varepsilon = -\frac{\phi(b)}{\Phi(b)\sqrt{d}} + O\left(\frac{\varepsilon}{\sqrt{d}}\right),$$

which is what we needed. The computation of $\mathbb{E}[X|X \geq b+\varepsilon]$ is analogous. $\qquad\square$

Proof of Lemma 2.5. Before we start with the proof, observe that X satisfies Gaussian-like tail bounds. Namely, for any $u \in (0,\infty)$, we can set $\theta = u/\sigma^2$ and obtain the following

$$\mathbb{P}[X \geq u] = \mathbb{P}[e^{\theta X} \geq e^{\theta u}] \leq \frac{\mathbb{E}[e^{\theta X}]}{e^{\theta u}} \leq e^{\theta^2\sigma^2/2-\theta u} = e^{-u^2/2\sigma^2}.$$

Note that the first inequality here is due to Markov's inequality and the second one to the subgaussian property of X. Thus, we also have $\mathbb{P}[|X| \geq u] \leq 2e^{-u^2/2\sigma^2}$

Now, we can compute the needed expectation as follows. Since $e^{\lambda X^2} \geq 1$, we have

$$\mathbb{E}[e^{\lambda X^2}] = 1 + \int_1^\infty \mathbb{P}[e^{\lambda X^2} \geq t]dt = 1 + \int_1^\infty \mathbb{P}\left[|X| \geq \sqrt{\frac{\log t}{\lambda}}\right]dt \leq 1 + 2\int_1^\infty e^{-\log t/2\lambda\sigma^2}dt,$$

where the last inequality comes from the tail bound on X. Finally, the integral in question is easily computable when $2\lambda\sigma^2 < 1$ since then we have

$$\mathbb{E}[e^{\lambda X^2}] \leq 1 + 2\int_1^\infty t^{-1/2\lambda\sigma^2}dt = 1 + 2\left(\frac{t^{1-1/2\lambda\sigma^2}}{1-1/2\lambda\sigma^2}\bigg|_1^\infty\right) = 1 - 2\frac{1}{1-1/2\lambda\sigma^2} = 1 + \frac{4\lambda\sigma^2}{1-2\lambda\sigma^2}. \quad \square$$

Proof of Lemma 2.6. If A is the all-ones matrix with zeros on the diagonal, then $S = \frac{1}{2}X^\top AX$, where $X = (X_1, \ldots, X_k)^\top$ is a subgaussian vector in $\mathbb{R}^k$. Then, we can write

$$X^\top AX = \mathbb{E}[X]^\top A\mathbb{E}[X] + 2(X - \mathbb{E}[X])^\top A\mathbb{E}[X] + (X - \mathbb{E}X)^\top A(X - \mathbb{E}X).$$

Introducing $Y = X - \mathbb{E}X$ and using that A is zero on the diagonal, this reduces to $X^\top AX = \mathbb{E}[X^\top AX] + 2Y^\top A\mathbb{E}[X] + Y^\top AY$. If we let $\Sigma_0 = \frac{1}{2}\mathbb{E}[X^\top AX] = \mathbb{E}[S], \Sigma_1 = Y^\top A\mathbb{E}[X]$ and $\Sigma_2 = \frac{1}{2}Y^\top AY$, then we can write $S = \Sigma_0 + \Sigma_1 + \Sigma_2$, and thus $\mathbb{E}[e^{\lambda S}] = e^{\lambda\Sigma_0}\mathbb{E}[e^{\lambda\Sigma_1}e^{\lambda\Sigma_2}]$. The Cauchy-Schwarz inequality then gives

$$(16) \qquad\qquad \mathbb{E}[e^{\lambda S}] \leq e^{\lambda\Sigma_0}\sqrt{\mathbb{E}[e^{2\lambda\Sigma_1}]}\sqrt{\mathbb{E}[e^{2\lambda\Sigma_2}]}.$$

Since the entries of Y are independent, we have

$$\mathbb{E}[e^{2\lambda\Sigma_1}] \leq \prod_{i=1}^k \mathbb{E}[e^{2\lambda(A\mathbb{E}[X])_i \cdot Y_i}] \leq \prod_{i=1}^k \exp\left(\frac{2\lambda^2(A\mathbb{E}[X])_i^2}{d}\right)$$

$$(17) \qquad\qquad \leq \exp\left(\sum_{i=1}^k \frac{2\lambda^2}{d}k\sum_{j=1}^k(\mathbb{E}X_j)^2\right) = \exp\left(\frac{2\lambda^2 k^2}{d}\sum_{j=1}^k(\mathbb{E}X_j)^2\right),$$

where the second inequality follows since Y_i is subgaussian with mean zero and variance proxy at most $1/d$ and the third inequality by Cauchy-Schwarz.

Thus, in order to complete the proof, we only need to show that $\mathbb{E}[e^{2\lambda\Sigma_2}] \leq \exp(8|\lambda|k/d)$. To do this, we consider two cases - when $\lambda \geq 0$ and when $\lambda < 0$. Let us start with $\lambda \geq 0$. Since $2\Sigma_2 = \sum_{i\neq j} Y_i Y_j = \left(\sum_{i=1}^k Y_i\right)^2 - \sum_{i=1}^k Y_i^2 \leq \left(\sum_{i=1}^k Y_i\right)^2$, we have

$$\mathbb{E}\big[e^{2\lambda\Sigma_2}\big] \leq \mathbb{E}\bigg[\exp\bigg(\lambda\bigg(\sum_{i=1}^k Y_i\bigg)^2\bigg)\bigg].$$

Note that the random variable $Z = \sum_{i=1}^k Y_i$ is a subgaussian random variable with variance proxy at most k/d, as a sum of k subgaussians of variance proxy $1/d$. Therefore, using Lemma 2.5 we can bound the expectation of $\exp(\lambda Z^2)$ as

$$\mathbb{E}\big[e^{2\lambda\Sigma_2}\big] \leq \mathbb{E}\big[e^{\lambda Z^2}\big] \leq 1 + \frac{4\lambda k/d}{1 - 2\lambda k/d} \leq 1 + \frac{8\lambda k}{d} \leq \exp\bigg(\frac{8\lambda k}{d}\bigg).$$

When $\lambda < 0$, we use $2\Sigma_2 = \left(\sum_{i=1}^k Y_i\right)^2 - \sum_{i=1}^k Y_i^2 \geq -\sum_{i=1}^k Y_i^2$, leading to

$$\mathbb{E}\big[e^{2\lambda\Sigma_2}\big] \leq \mathbb{E}\bigg[\exp\bigg(-\lambda\sum_{i=1}^k Y_i^2\bigg)\bigg] = \prod_{i=1}^k \mathbb{E}[e^{-\lambda Y_i^2}],$$

where the last inequality is due to the independence of the variables Y_i. Since Y_i has variance proxy at most $1/d$, by Lemma 2.5, we find that

$$\mathbb{E}[e^{2\lambda\Sigma_2}] \leq \prod_{i=1}^k \mathbb{E}[e^{-\lambda Y_i^2}] \leq \prod_{i=1}^k \bigg(1 + \frac{4|\lambda|/d}{1 - 2|\lambda|/d}\bigg) \leq \prod_{i=1}^k \exp\bigg(\frac{8|\lambda|}{d}\bigg) = \exp\bigg(\frac{8|\lambda|k}{d}\bigg).$$

By plugging in the bounds (17) and $\mathbb{E}[e^{2\lambda\Sigma_2}] \leq \exp(8|\lambda|k/d)$ into (16), we find that

$$\mathbb{E}\big[e^{\lambda S}\big] \leq e^{\lambda\mathbb{E}[S]}\sqrt{\exp\bigg(\frac{2\lambda^2 k^2}{d}\sum_{j=1}^k (\mathbb{E}X_j)^2\bigg)}\sqrt{\exp\bigg(\frac{8|\lambda|k}{d}\bigg)} \leq \exp\bigg(\lambda\mathbb{E}[S] + \frac{\lambda^2 k^2}{d}\sum_{j=1}^k (\mathbb{E}X_j)^2 + \frac{4|\lambda|k}{d}\bigg),$$

which is what we needed to prove. □