

# Testing thresholds for high-dimensional sparse random geometric graphs

Siqi Liu\*   Sidhanth Mohanty†   Tselil Schramm‡   Elizabeth Yang§

November 23, 2021

## Abstract

The random geometric graph model  $\text{Geo}_d(n, p)$  is a distribution over graphs in which the edges capture a latent geometry. To sample  $G \sim \text{Geo}_d(n, p)$ , we identify each of our  $n$  vertices with an independently and uniformly sampled vector from the  $d$ -dimensional unit sphere  $\mathbb{S}^{d-1}$ , and we connect pairs of vertices whose vectors are “sufficiently close,” such that the marginal probability of an edge is  $p$ . Because of the underlying geometry, this model is natural for applications in data science and beyond.

We investigate the problem of testing for this latent geometry, or in other words, distinguishing an Erdős-Rényi graph  $G(n, p)$  from a random geometric graph  $\text{Geo}_d(n, p)$ . It is not too difficult to show that if  $d \rightarrow \infty$  while  $n$  is held fixed, the two distributions become indistinguishable; we wish to understand how fast  $d$  must grow as a function of  $n$  for indistinguishability to occur.

When  $p = \frac{\alpha}{n}$  for constant  $\alpha$ , we prove that if  $d \geq \text{polylog}(n)$ , the total variation distance between the two distributions is close to 0; this improves upon the best previous bound of Brennan, Bresler, and Nagaraj (2020), which required  $d \gg n^{3/2}$ , and further our result is nearly tight, resolving a conjecture of Bubeck, Ding, Eldan, & Rácz (2016) up to logarithmic factors. We also obtain improved upper bounds on the statistical indistinguishability thresholds in  $d$  for the full range of  $p$  satisfying  $\frac{1}{n} \leq p \leq \frac{1}{2}$ , improving upon the previous bounds by polynomial factors.

Our analysis uses the Belief Propagation algorithm to characterize the distributions of (subsets of) the random vectors *conditioned on producing a particular graph*. In this sense, our analysis is connected to the “cavity method” from statistical physics. To analyze this process, we rely on novel sharp estimates for the area of the intersection of a random sphere cap with an arbitrary subset of  $\mathbb{S}^{d-1}$ , which we prove using optimal transport maps and entropy-transport inequalities on the unit sphere. We believe these techniques may be of independent interest.

---

\*UC Berkeley. sliu18@berkeley.edu. Supported in part by the Berkeley Haas Blockchain Initiative and a donation from the Ethereum Foundation.

†UC Berkeley. sidhanthm@cs.berkeley.edu. Supported by a Google PhD Fellowship.

‡Stanford University. tselil@stanford.edu.

§UC Berkeley. elizabeth\_yang@berkeley.edu. Supported by the NSF GRFP under Grant No. DGE 1752814.

# Contents

<b>1</b>	<b>Introduction</b>	<b>1</b>
1.1	Our results . . . . .	2
1.2	Prior and related work . . . . .	3
<b>2</b>	<b>Technical overview</b>	<b>6</b>
2.1	Relative entropy tensorization . . . . .	6
2.2	A geometric interpretation of neighborhood probability . . . . .	7
2.3	Neighborhood containment as a constraint satisfaction problem. . . . .	8
2.4	Further connections between Erdős-Rényi and geometric random graphs. . . . .	8
<b>3</b>	<b>Preliminaries</b>	<b>9</b>
3.1	Divergences between probability distributions . . . . .	9
3.2	Sphere caps and dot products of unit vectors . . . . .	10
3.3	Subgaussian random variables . . . . .	10
3.4	Random graphs . . . . .	11
3.5	Conditional vector distributions . . . . .	11
3.6	Constraint satisfaction problems and Belief Propagation . . . . .	12
<b>4</b>	<b>Concentration via optimal transport</b>	<b>12</b>
4.1	Optimal transport and the Wasserstein metric . . . . .	12
4.2	The concentration of sphere cap measure . . . . .	13
4.3	Different parameterizations of the sphere cap concentration . . . . .	16
<b>5</b>	<b>Concentration for intersections of caps and anti-caps</b>	<b>17</b>
<b>6</b>	<b>Stochastic domination by Erdős-Rényi graphs</b>	<b>21</b>
<b>7</b>	<b>Distribution of a neighborhood</b>	<b>23</b>
7.1	Setup and outline . . . . .	24
7.2	A relaxation of containment probabilities . . . . .	26
7.3	Belief propagation for relaxed distribution . . . . .	27
7.4	Relating relaxed and true distributions . . . . .	38
7.5	From neighborhood containment to neighborhood equality probabilities . . . . .	40
<b>8</b>	<b>Total variation bound</b>	<b>41</b>
8.1	Bounding neighborhood relative entropy . . . . .	42
<b>A</b>	<b>Extending the analysis of the signed triangle statistic to all densities</b>	<b>51</b>
<b>B</b>	<b>Deferred proofs from the preliminaries</b>	<b>53</b>

# 1 Introduction

The study of random graphs has been incredibly influential, not only in modeling applications, but also in the development of algorithms and in the study of mathematics. For example, consider the simple Erdős-Rényi random graph model  $G(n, p)$ , in which  $n$  nodes are each connected independently with probability  $p$ . This model was introduced by Erdős and Rényi in 1959 [ER59], and since then its study has blossomed. To list only a few of the many fruits of this line of research:  $G(n, p)$  graphs have been used to demonstrate the existence of many combinatorial objects via the probabilistic method (e.g. [AS04]); they have been used in algorithm design both as a benchmark and as a starting point for analysis in worst-case or semirandom settings (e.g. [FK01, BCC<sup>+</sup>10]); they have been useful for proving conditional and unconditional lower bounds in complexity theory (e.g. [Jer92, Fei02, Ros08, Gri01, Sch08]); and finally  $G(n, p)$  graphs are a common model in statistical physics [Nis81, Pan13] and network science [HLL83, CK16]. Though deep questions remain, decades of intensive study of the  $G(n, p)$  distribution have been rewarded with a rich understanding and many unforeseen insights.

Our primary object of study is the *random geometric graph* model on the sphere,  $\text{Geo}_d(n, p)$ . In this model, a graph is sampled by choosing  $n$  vectors  $\mathbf{v}_1, \dots, \mathbf{v}_n$  uniformly at random from the  $d$ -dimensional sphere  $\mathbb{S}^{d-1}$ , identifying  $\mathbf{v}_i$  with vertex  $i$ , then connecting each pair  $i, j$  whose vectors  $\mathbf{v}_i, \mathbf{v}_j$  have inner product exceeds a threshold  $\tau(p)$ , chosen so that  $\Pr[\langle \mathbf{v}_i, \mathbf{v}_j \rangle > \tau(p)] = p$ . A compelling aspect of this model is that the graph structure is derived from a geometric representation of the vertices; this makes it suitable for modeling applications in, say, data science, where we think of network nodes as representable by a feature vector in some  $d$ -dimensional space, with neighboring nodes sharing similar features.

Though random geometric graphs on the sphere (and also on other domains, such as  $[0, 1]^d$ ) have been studied extensively (see the monograph [Pen03]), the focus has been on the low-dimensional setting, where we think of  $d$  is fixed as  $n \rightarrow \infty$ . However, the high-dimensional setting, in which  $d \rightarrow \infty$  as a function of  $n$ , is still poorly understood. In low dimensions, one can understand  $\text{Geo}_d(n, p)$  as a fine random discretization of  $\mathbb{S}^{d-1}$ , but we no longer expect  $\text{Geo}_d(n, p)$  to behave in this way in the high-dimensional setting.

This opens an intriguing possibility: what are the properties of  $\text{Geo}_d(n, p)$  in the high-dimensional setting? And, what are the undiscovered applications of these random objects in algorithm design and in mathematics? One reason to study this question is that the high-dimensional setting is a more faithful model for graphs arising from modern high-dimensional data (e.g. high-dimensional feature vectors). Thus, high-dimensional geometric random graphs could constitute a more useful benchmark for algorithmic methods, or a more useful starting point for designing algorithms for the semirandom setting. Another reason to study this distribution is that it may yield graphs with properties that we have not yet realized are possible; this has historical precedent, for example, Erdős-Rényi graphs witnessing the existence of certain Ramsey graphs [AS04].

Devroye, György, Lugosi, and Udina [DGLU11] made the first exploration of  $\text{Geo}_d(n, p)$  in the high-dimensional regime. They study the classic property of the chromatic number in random geometric graphs, and also raise the following fundamental question: for which  $d = d(n, p)$  is it possible to test for the presence of an underlying geometry? When can one distinguish  $\text{Geo}_d(n, p)$  from  $G(n, p)$ ?

For intuition, one can see that if  $d \rightarrow \infty$  for  $n$  fixed, the  $\text{Geo}_d(n, p)$  distribution approaches  $G(n, p)$ : the vectors  $\mathbf{v}_1, \dots, \mathbf{v}_n$  are effectively mutually orthogonal,  $\tau(p)$  is very small, and conditioning on the

presence of the edge  $(i, j)$  (or equivalently on  $\langle \mathbf{v}_i, \mathbf{v}_j \rangle > \tau(p)$ ) has little impact on the probability that the edges  $(i, k)$  and/or  $(j, k)$  are present. How fast must  $d$  grow as a function of  $n$  to realize this limiting behavior? In [DGLU11] the authors observe that if  $d \gg \exp(n^2)$ , a central limit theorem implies that  $\text{Geo}_d(n, p)$  and  $G(n, p)$  are indistinguishable; they also note that this bound is likely far from tight.

Determining the asymptotic threshold in  $d$  when geometric graphs become indistinguishable from Erdős-Rényi graphs is a most basic question that we must address if we wish to make a serious study of high-dimensional random geometric graphs. Bubeck, Ding, Eldan, and Rácz [BDER16] were the first to tackle this question, showing that in dense graphs (when  $p = \Theta(1)$ ), the threshold occurs at  $d \asymp n^3$ ; when  $d \gg n^3$  the total variation distance goes to 0, and when  $d \ll n^3$  the “signed triangle count” provides a good test statistic. But in the arguably more interesting sparse case  $p = \Theta(\frac{1}{n})$ , [BDER16] are only able to establish that signed triangle counts distinguish when  $d = O(\log^3 n)$ . They conjecture that when  $d = \Omega(\log^3 n)$ , the distributions are indistinguishable.

The current best bound, due to Brennan, Bresler, and Nagaraj [BBN20], asserts that in the regime  $p = \Theta(\frac{1}{n})$ ,  $d_{\text{TV}}(\text{Geo}_d(n, p), G(n, p)) \rightarrow 0$  so long as  $d \gg n^{3/2}$ . In essence, their bound relies on the fact that independent random vectors  $\mathbf{v}_i, \mathbf{v}_j \sim \mathbb{S}^{d-1}$  have  $|\langle \mathbf{v}_i, \mathbf{v}_j \rangle| = \tilde{O}(\frac{1}{\sqrt{d}})$  with high probability; when  $d \gg n^{3/2}$ , these inner products are small enough relative to  $n$  that  $\mathbf{v}_i$  has negligible projection (of order  $\approx \sqrt{\sum_{j \neq i} \langle \mathbf{v}_i, \mathbf{v}_j \rangle^2} = O(\sqrt{n/d})$ ) into  $\text{span}\{\mathbf{v}_j\}_{j \neq i}$ , which is enough to guarantee approximate independence of edges. This argument is carried out with technical sophistication in [BBN20], but clearly, this technique cannot be extended to  $d < n$ , much less to  $d = \text{polylog } n$ , as the vectors in  $\{\mathbf{v}_j\}_{j \neq i}$  then span all of  $\mathbb{R}^d$ .

In this work, we come close to closing this gap in the sparse regime, confirming the conjecture of [BDER16] up to polylogarithmic factors: we show that if  $p = \Theta(\frac{1}{n})$ , the total variation distance goes to zero when  $d \gg \log^{36} n$ . We also give a separate improved bound on the total variation distance for the entire parameter regime  $p \ll 1$ , showing that the total variation distance goes to 0 when  $d \gg n^3 p^2$ , improving upon previous bounds by polynomial factors. Our proof relies on a number of novel technical contributions. Crucially, we must understand, for “typical”  $\mathbf{G} \sim \text{Geo}_d(n, p)$ , the distribution of (subsets of the) vectors  $\mathbf{v}_1, \dots, \mathbf{v}_n$  sampled from  $\mathbb{S}^{d-1}$  conditioned on giving a vector embedding for  $\mathbf{G}$ . In order to understand this complicated conditional distribution, we use the Belief Propagation (BP) algorithm, where our variables are the vertices  $[n]$  and their “labels” are vectors in  $\mathbb{S}^{d-1}$ . To analyze BP, we rely on a (to our knowledge) novel concentration inequality, which we prove using optimal transport, for the area of the intersection of a random spherical cap with any subset  $L \subseteq \mathbb{S}^{d-1}$ . We also demonstrate a coupling of  $\mathbf{G} \sim \text{Geo}_d(n, p)$  and  $\mathbf{G}_+ \sim G(n, p + o(p))$  which produces  $\mathbf{G} \subseteq \mathbf{G}_+$  (meaning  $\mathbf{G}$  is a subgraph of  $\mathbf{G}_+$ ) with high probability. We feel that these techniques may find other applications, and be of independent interest.

## 1.1 Our results

Our main result is an indistinguishability result for sparse Random Geometric Graphs and Erdős-Rényi graphs when the dimension  $d$  exceeds  $\text{polylog } n$ .

**Theorem 1.1.** *For any fixed constant  $\alpha \geq 1$ , if  $d = \Omega(\log^{36} n)$ , then*

$$\lim_{n \rightarrow \infty} d_{\text{TV}}\left(\text{Geo}_d\left(n, \frac{\alpha}{n}\right), G\left(n, \frac{\alpha}{n}\right)\right) = 0.$$

Our result settles the conjecture of [BDER16] up to logarithmic factors, an exponential improvement over the previous bound of [BBN20], which required  $d \gg n^{3/2}$ . We remark that we have not made an effort to optimize the logarithmic factors; it is possible that our current proofs in combination with chaining-style arguments will yield  $\log^3 n$ , matching their conjecture. We also obtain an improved result for general  $p = \Omega(\frac{1}{n})$ :

**Theorem 1.2.** *For any fixed constant  $\alpha > 0$ , if  $\frac{\alpha}{n} < p < \frac{1}{2}$  and  $d = \tilde{\Omega}(p^2 n^3)$ ,*

$$\lim_{n \rightarrow \infty} d_{\text{TV}}(\text{Geo}_d(n, p), G(n, p)) = 0.$$

This improves by polynomial factors (in  $p$  and  $n$ ) on the previous bound of [BBN20], which required  $d \gg \min\{pn^3 \log \frac{1}{p}, p^2 n^{7/2} \text{polylog} n\}$  and  $d \gg n \text{polylog} n$ . However, this result is not tight (at least for small  $p$ ) since in particular it does not recover Theorem 1.1. Given that we have come close to establishing the conjecture of [BDER16] in the sparse case, it is tempting to interpolate between the upper and lower bounds of [BDER16] in the  $p = \Theta(1)$  regime and their conjecture for the  $p = \Theta(\frac{1}{n})$  regime and speculate that for all  $p \leq \frac{1}{2}$ , the testing threshold occurs at  $d \asymp (nH(p))^3 = O(n^3 p^3 \log^3 \frac{1}{p})$ , for  $H(p)$  the binary entropy function. In Appendix A, we show that the “signed triangle count” test statistic analyzed by [BDER16] in the  $p = \Theta(1)$  regime can in fact distinguish whenever  $d \ll (nH(p))^3$  for all  $p = \Omega(\frac{1}{n})$ , establishing that the testing threshold occurs at some  $d = \Omega((nH(p))^3)$ . If  $d \asymp (nH(p))^3$  is indeed the threshold, our Theorem 1.2 is tight up to a factor of  $\tilde{O}(p)$ .

**Relaxed stochastic dominance by Erdős-Rényi graphs.** En route to proving Theorem 1.1 and Theorem 1.2, we establish a result which may be of independent interest. We show that whenever  $d = \tilde{\Omega}(p^2 n^2)$ , a random geometric graph  $\mathbf{G} \sim \text{Geo}_d(n, p)$  can be coupled with an Erdős-Rényi graph  $\mathbf{G}_+ \sim G(n, p + o(p))$  (and with  $\mathbf{G}_- \sim G(n, p - o(p))$ ) in such a way that the Erdős-Rényi graph “stochastically dominates”<sup>1</sup> the geometric graph with high probability over the coupling, in the sense that every edge in  $\mathbf{G}$  is also contained in  $\mathbf{G}_+$  (respectively, every edge in  $\mathbf{G}_-$  is also contained in  $\mathbf{G}$ ).

**Proposition 1.3.** *For any constant  $\alpha > 0$  there exist constants  $C_1, C_2 > 0$  such that if  $\frac{\alpha}{n} \leq p \leq \frac{1}{2}$  and  $d \geq C_1 \cdot (n^2 p^2 + \log^4 n) \log^4 n$ , for any  $\varepsilon \geq C_2 \sqrt{\frac{1}{d}(np + \log n) \log^4 n}$ , one can simultaneously sample  $\mathbf{G}_- \sim G(n, (1 - \varepsilon)p)$ ,  $\mathbf{G} \sim \text{Geo}_d(n, p)$ , and  $\mathbf{G}_+ \sim G(n, (1 + \varepsilon)p)$  in a correlated manner so that with probability at least  $1 - n^{-\Omega(\log n)}$ ,  $\mathbf{G}_- \subseteq \mathbf{G} \subseteq \mathbf{G}_+$ .*

The notation  $\mathbf{G} \subseteq \mathbf{G}_+$  means that  $\mathbf{G}$  is a subgraph of  $\mathbf{G}_+$ . We find it intriguing that this coupling succeeds whenever  $d = \tilde{\Omega}(p^2 n^2)$ , which is well below the speculative “interpolated” threshold  $d = \tilde{\Omega}(p^3 n^3)$  for large  $p$ .

## 1.2 Prior and related work

**Prior Works.** Random geometric graphs in fixed dimension are a well-studied model, with connections to Poisson processes and continuum percolation. The tools used to study fixed-dimensional random graphs are of a very different flavor from ours; for example, in low dimensions it is often useful to compare  $\text{Geo}_d(n, p)$  to a random process on an appropriate low-dimensional infinite lattice which discretizes the space. We refer the reader to the survey [Wal11] and the monograph [Pen03].

---

<sup>1</sup>Strictly speaking we only have stochastic dominance conditioned on the success of the coupling; that is why we say that it is “relaxed” stochastic dominance.

The study of random geometric graphs in high dimension was initiated by [DGLU11]. The main result of [DGLU11] is a bound on the clique number of  $\text{Geo}_d(n, p)$ ; for example in the dense case  $p = \Theta(1)$ , they show that the clique numbers of  $\text{Geo}_d(n, p)$  and  $G(n, p)$  become indistinguishable when  $d = \Omega(\log^3 n)$ . The authors of [DGLU11] also note that if  $d \rightarrow \infty$  fast enough as a function of  $n$  (they require  $d \gg \exp(n^2)$ ),  $\text{Geo}_d(n, p)$  and  $G(n, p)$  are indistinguishable. This naturally raises the question: for which  $d$  is it possible to test for the underlying geometry?

The authors of [BDER16] are the first to directly study the testing phase transition in  $d$ . They show that at any density, if  $d \gg n^3$ , then the total variation distance between  $\text{Geo}_d(n, p)$  and  $G(n, p)$  goes to zero with  $n$ , and conversely, if  $d \ll n^3$ , then in the dense regime  $p = \Theta(1)$ , the signed triangle count statistic furnishes a hypothesis test between the distributions. However in the sparse regime  $p = \Theta(\frac{1}{n})$ , their results are not as conclusive; they are able to show that triangle counts furnish a test if and only if  $d \ll \log^3 n$ , and they conjecture that if  $d \gg \log^3 n$  the total variation distance goes to zero. Briefly, their bound on the TV distance is via a reduction to the indistinguishability of Wishart matrices and matrices from the Gaussian Orthogonal Ensemble (GOE); the idea is that one can obtain a sample from  $G(n, p)$  by thresholding the off-diagonal entries of a GOE matrix  $\mathbf{B}$  with independent entries sampled from  $\mathcal{N}(0, \frac{1}{d})$  at threshold  $\approx \tau(p)$ , and similarly obtain a sample from  $\text{Geo}_d(n, p)$  by thresholding the off-diagonal entries of a Wishart matrix  $\mathbf{A}\mathbf{A}^\top$  at  $\approx \tau(p)$ , where  $\mathbf{A} \in \mathbb{R}^{n \times d}$  with rows sampled independently from  $\mathcal{N}(0, \frac{1}{d} \mathbb{1}_d)$ . Hence, if  $\mathbf{B}$  and  $\mathbf{A}\mathbf{A}^\top$  are indistinguishable from their off-diagonal entries, one can conclude that  $G(n, p)$  and  $\text{Geo}_d(n, p)$  are indistinguishable as well. It makes sense, then, that the result may not be tight in the sparse regime, as thresholding at a higher value  $\tau(p)$  (corresponding to a sparser graph) intuitively reveals less information about the original GOE or Wishart matrix. The proof in [BDER16] directly compares the Wishart and GOE densities to obtain the TV bound. Independently, [JL15] obtain the same bounds on the TV distance of Wishart and GOE matrices.

Following the work of [BDER16], [BBN20] study the question of detecting underlying geometry in greater generality. The authors show that for any  $\frac{\log n}{n^2} < p \leq \frac{1}{2}$ , if  $d \gg \min\left\{pn^3 \log \frac{1}{p}, p^2 n^{7/2} \text{polylog} n\right\}$  and  $d \gg n \log^4 n$ , then  $d_{\text{TV}}(\text{Geo}_d(n, p), G(n, p)) \rightarrow_n 0$ . In particular, they match the bound appearing in [BDER16] for dense graphs, and in sparse graphs with  $p = \Theta(\frac{1}{n})$  they improve the bound to  $d \gg n^{3/2} \text{polylog} n$ . At a high level, their proof applies information-theoretic inequalities to reduce the question to bounding the Chi-square divergence of the marginal of a single edge in  $\text{Geo}_d(n, p)$  vs.  $G(n, p)$ ; they then bound this Chi-squared divergence by showing that if one conditions  $\mathbf{v}_1, \dots, \mathbf{v}_n \sim (\mathbb{S}^{d-1})^{\otimes n}$  on the event that they produce a “typical” random geometric graph  $G$  excluding edge  $(n-1, n)$ , then the vectors  $\mathbf{v}_n, \mathbf{v}_{n-1}$  remain sufficiently independent that  $\Pr[\langle \mathbf{v}_{n-1}, \mathbf{v}_n \rangle > \tau(p)] \approx p$ . They achieve this by showing that  $\mathbf{v}_{n-1}$  and  $\mathbf{v}_n$  each have a small projection onto  $\text{span}\{\mathbf{v}_1, \dots, \mathbf{v}_{n-2}\}$ , even after conditioning on their adjacency into  $[n-2]$ . Their argument is essentially *oblivious to the particular choice of  $G$* , and merely uses properties that hold with high probability over independently sampled vectors in  $\mathbb{S}^{d-1}$ . This argument relies on  $\mathbf{v}_1, \dots, \mathbf{v}_{n-2}$  not spanning the entirety of  $\mathbb{R}^d$ , and the authors of [BBN20] explicitly state that improving their bound to any  $d < n$  requires a different approach.

**Techniques.** We describe our techniques in detail in Section 2, but here we discuss some connections in the literature. To improve upon the bounds of [BBN20], we draw upon tools from several areas. At the heart of the proofs of both Theorem 1.1 and Theorem 1.2 are new sharp concentration of measure results for intersections of random spherical caps with arbitrary subsets of  $\mathbb{S}^{d-1}$ ; to

prove these bounds, we make use of optimal transport inequalities in the Wasserstein metric (see e.g. [Vil08]). To our knowledge this is the first application of optimal transport in this context, and may be of independent interest. These tail bounds (in combination with Pinsker’s inequality and some simple arguments) are enough to yield [Theorem 1.2](#), proving that the total variation distance goes to zero if  $d \gg n^3 p^2$ .

Then, to break the  $d \asymp n$  barrier for sparse graphs and prove [Theorem 1.1](#), we must answer the following question: if  $\mathbf{v}_1, \dots, \mathbf{v}_n$  are sampled uniformly conditioned on producing a vector embedding of a fixed graph  $G$ , how strong are the correlations in the marginal distributions on  $\{\mathbf{v}_i\}_{i \in S}$  for small subsets  $S \subset [n]$ ? This is similar to establishing decay of correlations, as in the analysis of Gibbs sampling (c.f. [DSVW04, Wei06]). To achieve this, we carry out a rigorous analysis in the style of the “cavity method” from statistical physics (c.f. [MP03]): we compute the marginal over the depth- $\frac{\log n}{\log \log n}$ -neighborhood of vertices in  $S$  via the Belief Propagation (BP) algorithm, under arbitrary boundary conditions on the rest of the graph. The cavity method has been previously used to compute solution geometry phase transitions for a number of prominent discrete spin systems, such as coloring, Ising and Potts models, and more (c.f. [DMSS14, COKPZ18, Pan09, Tal03]); cavity-style arguments have also been employed in a similar way to establish decay of correlation properties in the analysis of Glauber dynamics, as in [GKM15]. In our instantiation of Belief Propagation, the variables are the vertices in the graph, and their “labels” or “assignments” are vectors in  $\mathbb{S}^{d-1}$ , the constraints are that vectors corresponding to edges have inner product at least  $\tau$ , and hence the “messages” passed from variable to variable are convolutions of marginal distributions with spherical caps; the concentration of measure for intersections of sets with random spherical caps is again useful in the analysis of this BP.

**Applications of high-dimensional geometric graphs in Theoretical CS.** A line of work including [FS02, FLS04] has utilized a distribution similar to  $\text{Geo}_d(n, p)$ , to obtain integrality gaps for semidefinite relaxations of max-cut or graph coloring. In their setting, a graph  $G$  is sampled by first placing  $n$  vertices in a “regular” configuration on a  $d$ -dimensional sphere, after which  $n' = \exp(\Theta(d))$  vertices are randomly subsampled independently with some probability  $q$ ,<sup>2</sup> and the sample is a graph induced on these  $n'$  vertices, in which vertices are connected if their inner product lies in some range of distances  $(\tau_1, \tau_2)$  (as opposed to the  $\text{Geo}_d(n, p)$  case in which the connectivity criteria is inner product at least  $\tau(p)$ ). The resulting  $G$  comes with a natural embedding into  $\mathbb{S}^{d-1}$ , which is utilized in constructing the semidefinite programming certificate. Similar constructions are also used in [KTW14] to give the optimal approximation ratio of constraint satisfaction problems assuming the unique games conjecture. Though these graphs are not sampled from  $\text{Geo}_d(n, p)$ , they are sampled from a distribution over a high-dimensional sphere which is qualitatively similar. We feel that this points to the promise of high-dimensional random geometric graphs for applications in theoretical computer science.

**Phase transitions between Wishart and GOE matrices.** The works mentioned above are those most closely related to our results. We mention as well additional works concerning the phase transition between Wishart and GOE matrices: the work of [RR19] studies the phase transition between Wishart and GOE at a higher resolution in the dense regime, deriving the expression for the total variation distance as a function of  $c = d/n^3$  when  $\lim_{n \rightarrow \infty} d/n$  is finite. The result [BG18] generalizes

---

<sup>2</sup>For small enough  $q$  these are equivalent to  $n'$  independent samples from the uniform measure over  $\mathbb{S}^{d-1}$ .

the  $d \gg n^3$  bound to Wishart matrices in vectors drawn from log-concave measures, and in [EM20] the authors study the phase transition for Wishart matrices in vectors drawn from non-isotropic Gaussian measures, drawing conclusions for hypothesis testing dense geometric graphs derived from these ensembles as well. Other similar questions that have been studied are hypothesis testing in *noisy* Wishart matrices [LR21], in which the entries are with some probability independently re-sampled from the Gaussian distribution, and *masked* hypothesis testing between Wishart and GOE matrices, in which only a subset of the matrix’s entries are revealed [BBH21]. In both of these cases, it has been demonstrated the presence of noise or masking can shift the threshold in interesting ways.

## Organization of the paper

In Section 2, we explain our techniques and the proofs of Theorem 1.1 and Theorem 1.2 at a high level. Section 3 contains preliminaries and definitions. In Section 4 we use optimal transport to derive tail bounds for the measure of the intersection of a random spherical cap with an arbitrary subset (or distribution) on  $\mathbb{S}^{d-1}$ , and in Section 5 we utilize these bounds to get tight concentration for a sequence of random caps and anti-caps. In Section 6 we prove our coupling Proposition 1.3. Then, in Section 7, we use the Belief Propagation algorithm to analyze the marginal distributions on vectors conditioned on producing a specific graph  $G$ ; we then put these ingredients together in Section 8 to prove Theorem 1.2 and Theorem 1.1. Finally, in Appendix A we show that the signed triangle count hypothesis test of [BDER16] can be extended to work so long as  $d \ll (np \log \frac{1}{p})^3$  for the full range of  $p$ , and in Appendix B we give some deferred proofs from Section 3.

## 2 Technical overview

### 2.1 Relative entropy tensorization

Our goal is to determine  $d$  at which the total variation distance  $d_{\text{TV}}(G(n, p), \text{Geo}_d(n, p))$  goes to 0 as  $n \rightarrow \infty$ . Like the authors of [BBN20], we relate the TV distance between these two distributions to their relative entropy (Definition 3.1)  $D(\text{Geo}_d(n, p) \| G(n, p))$  via Pinsker’s inequality (Theorem 3.3), and then apply the tensorization of the relative entropy (Claim 8.1). Roughly, the tensorization says that given a decomposition of  $G(n, p)$  as a product distribution, we can reduce the problem of bounding  $D(\text{Geo}_d(n, p) \| G(n, p))$  to bounding the relative entropy over (potentially simpler) distributions with smaller support.

$G(n, p)$  is conveniently a product distribution over edges. However, unlike [BBN20], we do not use this straightforward decomposition of  $G(n, p)$  by edge. Instead, let  $\mu_t$  be the distribution of vertex  $t$ ’s edges to  $[t - 1]$ . Similarly, let  $\nu_t$  be the marginal distribution of vertex  $t$ ’s edges to  $[t - 1]$  over the graph being sampled from  $\text{Geo}_d(n, p)$ . Our bound via tensorization now becomes

$$D(\text{Geo}_d(n, p) \| G(n, p)) = \sum_{t=1}^n \mathbf{E}_{\mathbf{G}_{t-1} \sim \text{Geo}(t-1, p)} [D(\nu_t(\cdot | \mathbf{G}_{t-1}) \| \mu_t)] \leq n \cdot \mathbf{E}_{\mathbf{G}_{n-1} \sim \text{Geo}(n-1, p)} [D(\nu_n(\cdot | \mathbf{G}_{n-1}) \| \mu_n)]$$

where the final inequality follows after applying the chain rule for relative entropy (Claim 8.2).

**The coupling view.** The tensorization inequality reduces bounding the TV distance to comparing the probability distribution of the neighborhood of the “final” vertex in  $G(n, p)$  and  $\text{Geo}(n, p, d)$ .

Specifically, we study  $\mathbf{E}_{\mathbf{G}_{n-1} \sim \text{Geo}(n-1, p)} [\mathbf{D}(v_n(\cdot) \mid \mathbf{G}_{n-1}) \parallel \mu_n]$  by considering the following scenario: we already have a graph  $\mathbf{G}_{n-1}$  sampled on  $n-1$  vertices, and we want to incorporate vertex  $n$  into our graph. By the definition of the Erdős-Rényi distribution,  $\mu_n$  will sample the neighbor set  $S \subseteq [n-1]$  with probability  $p^{|S|}(1-p)^{n-1-|S|}$ . For a random geometric graph, we can sample a vector  $\mathbf{v}_n \sim \rho$ , and take its dot products to vectors  $\mathbf{v}_1, \dots, \mathbf{v}_{n-1}$  sampled uniformly from  $\mathbb{S}^{d-1}$  conditioned on producing  $\mathbf{G}_{n-1}$ , to determine the neighbors of  $n$  in  $\mathbf{G}$  (which we denote by  $N_{\mathbf{G}}(n)$ ). Our goal now is to compare  $\Pr_{\mathbf{G} \sim \text{G}(n, p)}[N_{\mathbf{G}}(n) = S]$  and  $\Pr_{\mathbf{G} \sim \text{Geo}_d(n, p)}[N_{\mathbf{G}}(n) = S]$  for  $S \subseteq [n-1]$ .

## 2.2 A geometric interpretation of neighborhood probability

For  $\mathbf{G} \sim \text{Geo}_d(n, p)$ , if vertex  $i$  is associated to a (random) vector  $\mathbf{v}_i$ , and  $(i, j)$  is an edge, we consequently know that  $\langle \mathbf{v}_i, \mathbf{v}_j \rangle \geq \tau$ . On the sphere  $\mathbb{S}^{d-1}$ , the locus of points where  $\mathbf{v}_j$  can be, conditioned on  $(i, j)$  being an edge, is a sphere cap centered at  $\mathbf{v}_i$  with a  $p$  fraction of the sphere's surface area, which we denote by  $\text{cap}(\mathbf{v}_i)$ . Similarly, if we know that  $i$  and  $j$  are *not* adjacent, the locus of points where  $\mathbf{v}_j$  can fall is the complement of a sphere cap, which we call an “anti-cap,” with measure  $1-p$ .

Equipped with this geometric picture, we can view the probability that vertex  $n$ 's neighborhood is exactly equal to  $S \subseteq [n-1]$  as the measure  $\rho(\mathbf{L}_S)$ , where  $\mathbf{L}_S \subseteq \mathbb{S}^{d-1}$  is a random set defined as

$$\mathbf{L}_S := \left( \bigcap_{i \in S} \text{cap}(\mathbf{v}_i) \right) \cap \left( \bigcap_{i \notin S} \overline{\text{cap}(\mathbf{v}_i)} \right)$$

To show that the distance between  $\text{Geo}_d(n, p)$  and  $\text{G}(n, p)$  is small, we must show that  $\rho(\mathbf{L}_S)$  concentrates around  $p^{|S|}(1-p)^{n-1-|S|}$ , which is the probability that  $n$ 's neighborhood is equal to  $S$  under the Erdős-Rényi model.

**Optimal transport.** The backbone of our result is a (to our knowledge) novel application of optimal transport. In [Section 4](#), we prove for a generic distribution  $\mathcal{L}$  supported on  $\mathbb{S}^{d-1}$ , and  $\mathbf{z}$  sampled uniformly at random over  $\mathbb{S}^{d-1}$ , that

$$\Pr_{\mathbf{x} \sim \mathcal{L}} [\langle \mathbf{x}, \mathbf{z} \rangle \geq \tau] \in (1 \pm \varepsilon) \cdot p \text{ for } \varepsilon \leq \tilde{O} \left( \sqrt{\frac{\log \|\mathcal{L}\|_\infty}{d}} \right)$$

with high probability over  $\mathbf{z}$ . In other words, how tightly the random variable  $X_{\mathcal{L}}(\mathbf{z}) = \Pr_{\mathbf{x} \sim \mathcal{L}}[\langle \mathbf{x}, \mathbf{z} \rangle \geq \tau]$  concentrates is directly related to the maximum value of its relative density,  $\|\mathcal{L}\|_\infty$ .

To give some intuition for this result, first consider the case when  $\mathcal{L} = \rho$ , the uniform distribution over  $\mathbb{S}^{d-1}$ . Then, the variable  $X_\rho(\mathbf{z}) = \Pr_{\mathbf{x} \sim \rho}[\langle \mathbf{x}, \mathbf{z} \rangle \geq \tau] = p$  deterministically. Now, when  $\mathcal{L} \neq \rho$ , we can work with a transport map  $\mathcal{D}$  between  $\rho$  and  $\mathcal{L}$ , and we can couple  $\mathbf{y} \sim \mathcal{L}$  and  $\mathbf{x} \sim \rho$  according to  $\mathcal{D}$ , so that

$$X_{\mathcal{L}}(\mathbf{z}) = \Pr_{\mathbf{y} \sim \mathcal{L}} [\langle \mathbf{y}, \mathbf{z} \rangle \geq \tau] = \Pr_{\substack{(\mathbf{x}, \mathbf{y}) \sim \mathcal{D}(\rho, \mathcal{L}) \\ \mathbf{e} = \mathbf{y} - \mathbf{x}}} [\langle \mathbf{x}, \mathbf{z} \rangle \geq \tau - \langle \mathbf{e}, \mathbf{z} \rangle].$$

The smaller  $\|\mathcal{L}\|_\infty$  is, the smaller the average of the transport distance  $\|\mathbf{e}\| = \|\mathbf{x} - \mathbf{y}\|$ ; further when  $\mathbf{z} \sim \rho$  the quantity  $\langle \mathbf{e}, \mathbf{z} \rangle$  concentrates tightly around  $\frac{1}{\sqrt{d}}\|\mathbf{e}\|$ . In this way, we translate the concentration of transport distance into tail bounds on  $|X_\rho(\mathbf{z}) - X_{\mathcal{L}}(\mathbf{z})|$ .

To analyze  $\rho(\mathbf{L})$  for  $\mathbf{L}$  the intersection of caps and anti-caps defined above, we will apply the above in sequence inside a martingale concentration argument, building up  $\mathbf{L}$  one cap at a time ([Lemma 5.1](#), [Corollary 6.1](#)). Using this approach, our transport result alone is enough to conclude [Theorem 1.2](#). (The proof is assembled in [Section 8.1.1](#).)

**The need to resample vectors.** In the general  $p$  setting, we can think of our analysis of  $v_n(\cdot|\mathbf{G}_{n-1})$  as considering a fixed vector embedding  $\mathbf{v}_1, \dots, \mathbf{v}_{n-1}$  of  $\mathbf{G}_{n-1}$ , and then analyzing the probability that  $n$  connects to some  $S \subseteq [n-1]$ . When  $p = \frac{\alpha}{n}$ , this does not yield tight results; moreover, one can show that this is not due to loose tail bounds on  $\rho(\mathbf{L})$ , as our concentration results have matching anti-concentration results.

Hence, in order to prove [Theorem 1.1](#), we must additionally consider the concentration of  $\rho(\mathbf{L}_S)$  on average over vector embeddings of  $\mathbf{G}_{n-1}$  as well. We will first sample  $\mathbf{G}_{n-1}$ , and then for each set  $S$ , we bound the deviation in the random variable  $\rho(\mathbf{L}_S) = \rho(\cap_{i \in S} \text{cap}(\mathbf{u}_i) \cap \cap_{j \notin S} \overline{\text{cap}}(\mathbf{u}_j))$  conditioned on  $\mathbf{u}_1, \dots, \mathbf{u}_{n-1}$  producing  $\mathbf{G}_{n-1}$ . To do this, we will use a ‘‘cavity-method’’ style argument: we will view all vectors at distance  $> \ell = \frac{\log n}{\log \log n}$  from  $S$  as fixed and arbitrary, and then exactly compute the marginal distributions over  $\mathbf{u}_i$  for  $i$  at distance  $\leq \ell$  from  $S$ , conditional on forming  $\mathbf{G}_{n-1}$ .

### 2.3 Neighborhood containment as a constraint satisfaction problem.

We first reduce the need for high-probability estimates for  $\Pr_{\text{Geo}, \mathbf{G}_{n-1}}[N(k) = S]$  to obtaining estimates for  $\Pr_{\text{Geo}, \mathbf{G}_{n-1}}[N(k) \supseteq S]$  instead. This simplification is possible because the measure of anti-cap intersections concentrates dramatically better than the measure of cap intersections. With this step, we eliminate the need to study anti-correlations between  $\mathbf{v}_i, \mathbf{v}_j$  that do not have an edge between them.

Given  $S$  and  $\mathbf{G}_{n-1}$  (and its corresponding vectors), we fix all vectors except those corresponding to the depth- $\frac{\log n}{\log \log n}$  neighborhood of  $S$  in  $\mathbf{G}_{n-1}$ , which is with high probability a union of trees. To formally analyze the distribution of the unfixed vectors upon resampling them, we set up a 2-CSP (constraint satisfaction problem) instance over a continuous alphabet that encodes the edges of  $\mathbf{G}_{n-1}$  within the trees around  $S$ : each node has a vector-valued variable in  $\mathbb{S}^{d-1}$ , and the constraints are that nodes joined by an edge must have vectors with inner product at least  $\tau$ .

**Belief propagation.** Since our 2-CSPs are over trees, the *belief propagation (BP) algorithm* exactly computes the marginal distribution of each variable vector (see [Section 3.6](#) for the definition of BP). Using our results on the concentration of  $\Pr_{\mathbf{x} \sim \mathcal{L}}[\langle \mathbf{x}, \mathbf{z} \rangle \geq \tau]$  over  $\mathbf{z}$  uniform on  $\mathbb{S}^{d-1}$ , we can quantify the TV distance between the marginal distributions of our resampled vectors and the uniform distribution over  $\mathbb{S}^{d-1}$ . At a high level, the farther some  $\mathbf{v}_i$  is from a fixed vector in our 2-CSP, the closer its distribution is to uniform. The key insight is that the message from  $i$  to its neighbor  $j$  in our belief propagation algorithm correspond to a convolution of the marginal distribution of  $\mathbf{v}_i$  with a spherical cap. We can then use our concentration of measure for spherical caps from [Section 4](#) to show that convolutions of spherical caps mix to uniform rapidly, causing the correlations between far away vertices to decay. This can be seen as a form of the ‘‘decay of correlations’’ phenomenon. This analysis gives us the finer-grained control over  $\Pr[N_{\mathbf{G}}(n) = S]$  needed to conclude [Theorem 1.1](#).<sup>3</sup>

### 2.4 Further connections between Erdős-Rényi and geometric random graphs.

In the course of proving [Theorem 1.1](#), we prove [Proposition 1.3](#), which demonstrates a coupling of  $\mathbf{G}_- \sim \mathbf{G}(n, p - o(p))$ ,  $\mathbf{G} \sim \text{Geo}_d(n, p)$ , and  $\mathbf{G}_+ \sim \mathbf{G}(n, p + o(p))$  that satisfies  $\mathbf{G}_- \subseteq \mathbf{G} \subseteq \mathbf{G}_+$  with high

<sup>3</sup>Unfortunately, this analysis only works for  $p = \frac{\alpha}{n}$ ; otherwise the neighborhoods around vertices in  $S$  are only trees at a depth which is too shallow for the correlations to decay sufficiently, so the resampled vectors’ distributions are not close enough to uniform.

probability. We use this coupling in the proof of [Theorem 1.1](#) to leverage known structural results on Erdős–Rényi graphs to reason about the structure of random geometric graphs. For instance, to upper bound the probability that the depth- $\ell$  neighborhood of some  $i \in [n]$  forms a tree under  $\text{Geo}_d(n, p)$ , we rely on known bounds on the probability of this event under  $G(n, p(1 \pm o(1)))$ . We remark that this coupling may be of independent interest.

### 3 Preliminaries

**Basic notation.** Throughout this paper, we use **boldface** for random variables. We use  $\log x$  to denote the natural base logarithm, and for  $x \in [0, 1]$ ,  $H(x) = x \log \frac{1}{x} + (1-x) \log \frac{1}{1-x}$  denotes the binary entropy function (with the understanding that  $H(0) = H(1) = 0$ ). We use standard big- $O$  notation, and we use  $\tilde{O}$  and  $\tilde{\Omega}$  to hide polylog  $n$  factors. The notation  $f(x) \gg g(x)$  to denote that  $\lim_{x \rightarrow \infty} \frac{g(x)}{f(x)} = 0$ ; the argument  $x$  will be clear from context.

We use  $\rho$  to denote the uniform distribution on  $\mathbb{S}^{d-1}$ . Given a function  $f : \mathbb{S}^{d-1} \rightarrow \mathbb{R}$ , we define its  $p$ -th norm for  $p < \infty$  as:

$$\|f\|_p := \left( \mathbf{E}_{\mathbf{z} \sim \rho} |f(\mathbf{z})|^p \right)^{1/p}$$

and for  $p = \infty$  as:

$$\|f\|_\infty := \sup_z |f(z)|.$$

Given a distribution  $\nu$  on  $\mathbb{S}^{d-1}$ , we overload notation and use  $\|\nu\|_p$  to denote  $\left\| \frac{d\nu}{d\rho} \right\|_p$ . We will also frequently use the symbol  $\nu$  itself to denote its relative density to  $\rho$ . For a set  $A$ , we overload notation and use  $A$  to denote the uniform distribution on  $A$  when it is clear from context. We'll use  $\mathbf{X} | \mathcal{E}$  to denote the random variable  $\mathbf{X}$  sampled from the conditional distribution of  $\mathbf{X}$  conditioned on the event  $\mathcal{E}$ .

Given a graph  $G$ , and a subset of its vertices  $S$ , we use  $G[S]$  to denote the induced subgraph of  $G$  on  $S$ . We use  $B_G(v, \ell)$  to denote the ball of radius- $\ell$  around a vertex  $v$  in graph  $G$ , and similarly use  $S_G(v, \ell)$  to denote the corresponding sphere of radius- $\ell$ . We use  $N_G(i)$  to denote the neighbors of vertex  $i$  in graph  $G$ .

Given a collection of vectors  $V = (v_1, \dots, v_n) \in (\mathbb{S}^{d-1})^n$  the associated *geometric graph* denoted  $\text{gg}(V, p) = ([n], \mathcal{E})$  is given by choosing the edge set as all  $\{i, j\}$  where  $\langle v_i, v_j \rangle \geq \tau(p)$ . In particular, when  $V \sim \rho^{\otimes n}$ ,  $\text{gg}(V, p)$  is distributed as  $\text{Geo}_d(n, p)$ . For an  $n$ -vertex graph  $G$ , we use  $\rho^G$  to denote the conditional distribution  $\rho^G = \rho^{\otimes n} | \text{gg}(V, p) = G$ .

#### 3.1 Divergences between probability distributions

We use the relative entropy to compare probability distributions.

**Definition 3.1** (Relative Entropy). Given two probability distributions  $\mu$  and  $\nu$  over  $\Omega$  where  $\nu$  is absolutely continuous with respect to  $\mu$ , the *relative entropy* between  $\mu$  to  $\nu$  is:

$$D(\nu \| \mu) := \int_{\Omega} \log \left( \frac{d\nu}{d\mu}(x) \right) d\nu(x).$$

If  $\nu$  is not absolutely continuous with respect to  $\mu$ , then we define the relative entropy to be  $\infty$ .

A simple but useful observation is the following.

**Observation 3.2.** If for all  $x \in \Omega$ ,  $\frac{d\nu}{d\mu}(x) \leq C$ : (1)  $D(\nu\|\mu) \leq \log C$ , and (2) for any event  $\mathcal{E}$ :  $\nu(\mathcal{E}) \leq C \cdot \mu(\mathcal{E})$ .

We will use Pinsker's inequality to bound the total variation distance between two probability distributions in terms of the relative entropy.

**Theorem 3.3** (Pinsker's inequality). *For distributions  $\mu, \nu$  over the same domain,  $d_{\text{TV}}(\nu, \mu)^2 \leq \frac{1}{2}D(\nu\|\mu)$ .*

See e.g. [Mas07, Theorem 2.16] for a proof.

### 3.2 Sphere caps and dot products of unit vectors

Here, we introduce some useful bounds on the measure of sphere caps, and a concentration bound on the dot products of two unit vectors. Proofs for some of these lemmas are provided in the appendix, along with some additional facts about the geometry of the unit sphere that we use to prove them.

**Definition 3.4** ( $p$ -cap). For a vector  $v \in \mathbb{S}^{d-1}$ , its  $p$ -cap is  $\text{cap}_p(v) := \{x \in \mathbb{S}^{d-1} : \langle v, x \rangle \geq \tau(p)\}$ . Similarly, we define its  $p$ -anticap as  $\overline{\text{cap}}_p(v) := \{x \in \mathbb{S}^{d-1} : \langle v, x \rangle < \tau(p)\}$ . We drop the  $p$  in the subscript when its value is clear from context.

Recall Lévy's theorem for concentration of measure on the unit sphere [Mat13, Theorem 14.1.1]. We use it to upper bound the measure of a sphere cap with threshold  $\tau$ .

**Lemma 3.5.** *Let  $y \in \mathbb{S}^{d-1}$  be any vector. Then:  $\Pr_{\mathbf{w} \sim \mathbb{S}^{d-1}}[|\langle \mathbf{w}, y \rangle| \geq \tau] \leq 4 \exp(-\tau^2 d/2)$ .*

We now present a convenient upper bound on the dot product threshold  $\tau(p)$  of a  $p$ -cap. Its proof is provided in [Appendix B](#).

**Lemma 3.6.** *For any  $p \leq \frac{1}{2}$ , we have  $\tau(p) \leq \sqrt{\frac{2 \log(1/p)}{d}}$ .*

The next lemma helps us understand the deviations in cap volume  $p$  when we make small adjustments to its dot product threshold  $\tau(p)$ . Its proof is also in [Appendix B](#).

**Lemma 3.7.** *Fix  $x \in \mathbb{S}^{d-1}$ . Let  $p := \Pr_{z \sim \rho}[\langle z, x \rangle \geq \tau]$ . For any  $\varepsilon \geq 0$ , there is a universal constant  $C_{3.7}$  such that:*

$$\Pr_{z \sim \rho}[\tau - \varepsilon \leq \langle z, x \rangle \leq \tau + \varepsilon] \leq p \cdot \left( C_{3.7} \varepsilon \exp(2d\tau\varepsilon) \sqrt{d \log(1/p)} \right)$$

### 3.3 Subgaussian random variables

We will need concentration inequalities for martingales that arise as a sum of subgaussian random variables.

**Definition 3.8** (Subgaussian norm). The *subgaussian norm* of a real-valued random variable  $\mathbf{X}$  is defined as:

$$\|\mathbf{X}\|_{\psi_2} := \inf\{K > 0 : \mathbf{E} \exp(\mathbf{X}^2/K^2) \leq 2\}.$$

We say  $\mathbf{X}$  is a *subgaussian random variable* if  $\|\mathbf{X}\|_{\psi_2} < \infty$ .

We will need a version of Azuma's inequality for martingales with centered subgaussian increments. A proof may be found in [VH14], Lemma 3.7.

**Lemma 3.9** (Subgaussian martingale concentration inequality). *There exists a constant  $C_{3.9} > 0$  such that if  $\mathbf{X}_0, \mathbf{X}_1, \dots, \mathbf{X}_m$  is a martingale sequence with respect to a filtration  $(\mathcal{V}_t)_{t \in [m]}$  and  $K_i := \sup_{\mathcal{V}_{i-1}} \|\mathbf{X}_i - \mathbf{X}_{i-1}\|_{\psi_2} < \infty$  for all  $i \in [m]$ , then:*

$$\Pr[|\mathbf{X}_m - \mathbf{X}_0| \geq t] \leq 2 \exp\left(-C_{3.9} \frac{t^2}{\sum_{i=1}^m K_i^2}\right).$$

Also, we will make use of the following statement to bound the subgaussian norm of a random variable in terms of its tail probabilities.

**Lemma 3.10** ([Ver18], Proposition 2.5.2). *There exists a constant  $C_{3.10} > 0$  such that if  $\mathbf{X}$  is a random variable satisfying  $\Pr[|\mathbf{X}| > t] \leq 2 \exp\left(-\frac{t^2}{K^2}\right)$  for all  $t \geq 0$ , then  $\|\mathbf{X}\|_{\psi_2} \leq C_{3.10} \cdot K$ .*

### 3.4 Random graphs

In this section, we include some facts about Erdős–Rényi random graphs. First, we state a high-probability upper bound of  $O\left(\frac{\log n}{\log \log n}\right)$  on the maximum degree in both random graph models when  $p = \frac{\alpha}{n}$  for constant  $\alpha$ . This allows us to ignore graphs where the max degree is too high when we upper bound  $d_{\text{TV}}(\mathbf{G}(n, p), \text{Geo}_d(n, p))$ . Since the degree of each vertex is distributed like  $\text{Binom}(n, p)$ , by applying the standard tail bound for a Binomial random variable and taking a union bound over all vertices we get the following.

**Lemma 3.11.** *Let  $\Delta(G)$  denote the maximum degree of a graph  $G$ . If  $p = \frac{\alpha}{n}$ , then for both  $\mathbf{G} \sim \mathbf{G}(n, p)$  and  $\mathbf{G} \sim \text{Geo}_d(n, p)$  and for all  $d$ :*

$$\Pr[\Delta(\mathbf{G}) \geq k] \leq n \cdot \left(\frac{k}{e\alpha}\right)^{-k}.$$

More generally, we know the following bound on the number of vertices in an Erdős–Rényi graph at distance  $\leq \ell$  from any given vertex, as well as a high probability statement about their structure.

**Lemma 3.12** ([BLM15, Lemma 29]). *Let  $\mathbf{G} \sim \mathbf{G}(n, p)$  for  $p = \frac{\alpha}{n}$  for constant  $\alpha$ . Define  $\mathbf{h}_t(v)$  as the number of vertices with distance exactly  $t$  from  $v$  in  $\mathbf{G}$ . Then for any vertex  $v$ , there are constants  $c, C$  such that:*

$$\Pr[\exists t \geq 0 : \mathbf{h}_t(v) > s\alpha^t] \leq C \exp(-cs).$$

**Lemma 3.13** ([BLM15, Lemma 30]). *Let  $\mathbf{G} \sim \mathbf{G}(n, p)$  for  $p = \frac{\alpha}{n}$  for constant  $\alpha$ , and let  $B_{\mathbf{G}}(v, t)$  be the set of all vertices with distance  $\leq t$  from vertex  $v$  in  $\mathbf{G}$ . Then, for any vertex  $v$ , there is a constant  $c'$  such that:*

$$\Pr[B_{\mathbf{G}}(v, t) \text{ is not a tree}] \leq \frac{c'\alpha^t}{n}$$

### 3.5 Conditional vector distributions

Given an  $n$ -vertex graph  $G$ , we are interested in  $\mathbf{V} \sim \rho^{\otimes n} | \text{gg}(\mathbf{V}, p) = G$ , whose distribution we shorten to  $\rho^G$ . A simple but crucial observation for us is the following.

**Observation 3.14.** Let  $f : (\mathbb{S}^{d-1})^n \rightarrow \{0, 1\}$  be a Boolean-valued function. If  $\mathbf{E}_{\mathbf{V} \sim \rho^{\otimes n}} f(\mathbf{V}) \leq \delta$ , then:

$$\Pr_{\mathbf{G} \sim \text{Geo}_d(n, p)} \left[ \mathbf{E}_{\mathbf{V} \sim \rho^G} f(\mathbf{V}) \geq \sqrt{\delta} \right] \leq \sqrt{\delta}.$$

*Proof.* We can write:  $\mathbf{E}_{\mathbf{V} \sim \rho^{\otimes n}} f(\mathbf{V}) = \mathbf{E}_{\mathbf{G} \sim \text{Geo}_d(n, p)} \mathbf{E}_{\mathbf{V} \sim \rho^G} f(\mathbf{V}) \leq \delta$ . The statement then follows from Markov's inequality.  $\square$

### 3.6 Constraint satisfaction problems and Belief Propagation

**Definition 3.15.** A *constraint satisfaction problem instance* (CSP instance)  $\mathcal{I}$  consists of a *variable set*  $V$  and a *constraint set*  $E$ . The variables  $v \in V$  each belong to an alphabet  $\Sigma$ , while the constraints  $f \in E$  consist of a  $k$ -tuple of variables  $\partial f$  and a function  $\psi_f : \Sigma^k \rightarrow \{0, 1\}$ . An assignment to variables  $c : V \rightarrow \Sigma$  is *satisfying* if for each constraint  $f \in E$ ,  $\psi_f(c(\partial f)) = 1$ , i.e. it is satisfied.

**Remark 3.16.** In our setting, we choose  $\Sigma = \mathbb{S}^{d-1}$ .

**Definition 3.17.** We can represent any CSP instance  $\mathcal{I}$  as a bipartite graph  $F$ , which we call a *factor graph*. The two sides of the bipartition are  $V$  and  $E$ , and we place an edge between  $v \in V$  and  $f \in E$  if  $v$  participates in  $f$ . We use  $\partial v$  and  $\partial f$  to denote the neighborhoods of variables  $v$  and clauses  $f$ , respectively, in this graph.

When the factor graph  $F$  does not contain cycles, the marginal on a variable  $v$  can be computed exactly from the fixed point of the belief propagation algorithm. We state this precisely below.

**Definition 3.18** (Belief propagation fixed point). A *belief propagation fixed point* for a factor graph  $F$  is a collection of *messages*

$$\left\{ m^{v \rightarrow f}, m^{f \rightarrow v} \right\}_{v \in V, f \in E}$$

for all pairs  $v, f$  such that  $f$  is a neighbor of  $v$ , where each message is a probability distribution on  $\Sigma$ , such that

$$m^{f \rightarrow v}(x) \propto \int_{c|_{\partial f}: c(v)=x} \psi_f(c|_{\partial f}) \prod_{v' \in \partial f \setminus v} m^{v' \rightarrow f}(c(v')) \quad (1)$$

$$m^{v \rightarrow f}(x) \propto \prod_{f' \in \partial v \setminus f} m^{f' \rightarrow v}(x) \quad (2)$$

**Theorem 3.19** (e.g. [MM09], Theorem 14.1). *Suppose  $F$  is a forest factor graph corresponding to a CSP instance  $\mathcal{I}$  where every vertex is attached to a unary constraint. Then there is a unique belief propagation fixed point and the marginal distribution  $\nu$  on variable  $v$  over the uniform distribution over satisfying assignments to  $\mathcal{I}$  is given via the following formula.*

$$\nu \propto \prod_{f \in \partial v} m^{f \rightarrow v}.$$

## 4 Concentration via optimal transport

In this section we establish that for a probability distribution  $\nu$  over  $\mathbb{S}^{d-1}$ , a random  $p$ -cap on the sphere contains a  $p$ -fraction of  $\nu$ 's measure with high probability, where the strength of the concentration depends on  $\|\nu\|_\infty$ . We do so by analyzing the optimal transport mapping  $\mathcal{D}$  between  $\nu$  and the uniform measure  $\rho$ .

### 4.1 Optimal transport and the Wasserstein metric

The Wasserstein metric quantifies the ‘‘physical distance’’ between a pair of probability distributions. We say  $\pi(x, y)$  is a transport coupling between distributions  $\mu$  and  $\nu$  if  $\pi(x, \cdot) = \mu(x)$  and  $\pi(\cdot, y) = \nu(y)$ .

**Definition 4.1** (Wasserstein Distance). Let  $\mu$  and  $\nu$  be two probability distributions over  $\Omega$ , and  $\Pi$  be the set of all transport couplings  $\pi(x, y)$  between  $\mu$  and  $\nu$ . Then, their *Wasserstein-2 distance* is

$$W_2(\mu, \nu) := \sqrt{\inf_{\pi \in \Pi} \int_{\Omega \times \Omega} \|x - y\|_2^2 d\pi(x, y)}$$

It is straightforward to verify that  $W_2(\cdot, \cdot)$  is in fact a metric.

In other words, the square of the Wasserstein distance is the average squared Euclidean distance  $\|\mathbf{x} - \mathbf{y}\|^2$  between  $(\mathbf{x}, \mathbf{y}) \sim \pi$  for the most efficient “transport coupling”  $\pi$ . For intuition, we may think of  $\mu$ ’s density of a pile of sand over  $\Omega$ , and of  $\pi$  as a map that shifts grains of sand from  $\mu$  to form the shape of  $\nu$  in such a way that minimizes the average distance traveled.

**Fact 4.2** ([BGL13, Proposition 9.1.2]). *Given two probability distributions  $\mu$  and  $\nu$  over  $\Omega$ , there exists a coupling  $\pi$  such that:*

$$\int_{\Omega \times \Omega} \|x - y\|_2^2 d\pi(x, y) = W_2(\mu, \nu)^2.$$

We call any such  $\pi$  an *optimal coupling*.

We specifically will need bounds on the Wasserstein-2 distance between an arbitrary distribution  $\nu$  on the unit sphere and the uniform distribution on the unit sphere  $\rho$ . We obtain a handle on the distances we need via an *entropy-transport inequality*, which bounds the Wasserstein-2 distance in terms of the relative entropy. The following is a direct corollary of [Vil08, Theorem 22.17, part (i)] and [DEKL14, Corollary 2].

**Lemma 4.3** (Talagrand’s  $T_2$  inequality on the sphere). *For any distribution  $\nu$  on  $\mathbb{S}^{d-1}$ , and  $\rho$  the uniform measure on  $\mathbb{S}^{d-1}$ ,*

$$W_2(\nu, \rho) \leq \sqrt{\frac{2}{d-1} \cdot D(\nu \| \rho)}.$$

For a reader interested in the proof of [Lemma 4.3](#), we recommend the proof of the analogous statement in Gaussian space by [Tal96] due to its relative simplicity. In fact, it is possible to derive a slightly weaker bound of  $\sqrt{\frac{2}{d} \cdot D(\nu \| \rho) + \frac{2}{\sqrt{d}}}$  from the Gaussian case via an elementary proof. We also find it worthwhile to point to [BGL13, Theorem 9.2.1] for a comprehensive exposition.

## 4.2 The concentration of sphere cap measure

Let  $\rho$  be the uniform measure over  $\mathbb{S}^{d-1}$ , let  $\nu$  be a probability measure over  $\mathbb{S}^{d-1}$ , and let  $\mathcal{D}$  be a coupling between them. For  $(\mathbf{x}, \mathbf{y}) \sim \mathcal{D}$  we use the convention that  $\mathbf{x}$  is distributed according to  $\nu$  and  $\mathbf{y}$  is distributed according to  $\rho$ . We first prove quantitative bounds on the concentration of  $\|\mathbf{x} - \mathbf{y}\|$  when  $(\mathbf{x}, \mathbf{y}) \sim \mathcal{D}$ .

**Lemma 4.4.** *Let  $\nu$  be a distribution over  $\mathbb{S}^{d-1}$  and let  $\mathcal{D}$  be the optimal transport coupling of  $\nu$  and  $\rho$ . Then for all  $t > 0$ ,*

$$\Pr_{(\mathbf{x}, \mathbf{y}) \sim \mathcal{D}} \left[ \|\mathbf{x} - \mathbf{y}\| \geq t + \sqrt{\frac{2}{d-1} \log \|\nu\|_\infty} \right] \leq \exp \left( -\frac{d-1}{8} t^2 \right).$$

*Proof.* Let  $(\mathbf{x}, \mathbf{y}) \sim \mathcal{D}$  with  $\mathbf{x} \sim \nu$  and  $\mathbf{y} \sim \rho$ . Let  $\mathcal{E}_s$  be the event that  $\|\mathbf{x} - \mathbf{y}\| \geq s$ , and let  $p_s := \mathbf{Pr}_{\mathcal{D}}[\mathcal{E}_s]$ . Let  $\mathcal{D}^s = \mathcal{D}|\mathcal{E}_s$ , and let  $\mathcal{D}_\nu^s, \mathcal{D}_\rho^s$  be the marginal distributions of  $\mathcal{D}^s$  on  $\mathbf{x}$  and  $\mathbf{y}$  respectively. We claim

$$W_2(\mathcal{D}_\nu^s, \mathcal{D}_\rho^s) \geq s.$$

Indeed, suppose not; then, one could obtain a coupling which further decreases the transport distance between  $\nu$  and  $\rho$ , contradicting the optimality of  $\mathcal{D}$ . Now, since  $W_2$  is a metric:

$$s \leq W_2(\mathcal{D}_\nu^s, \mathcal{D}_\rho^s) \leq W_2(\mathcal{D}_\rho^s, \rho) + W_2(\mathcal{D}_\nu^s, \rho) \leq \sqrt{\frac{2}{d-1} \mathbf{D}(\mathcal{D}_\rho^s \| \rho)} + \sqrt{\frac{2}{d-1} \mathbf{D}(\mathcal{D}_\nu^s \| \rho)}, \quad (3)$$

where we have used the triangle inequality in conjunction with Talagrand's  $T_2$  inequality (Lemma 4.3). Finally, since  $\mathcal{D}_\rho^s(x) = \rho(x|\mathcal{E}_s) \leq \frac{1}{p_s} \rho(x)$ ,

$$\mathbf{D}(\mathcal{D}_\rho^s \| \rho) = \int_{\mathbb{S}^{d-1}} \mathcal{D}_\rho^s(x) \cdot \log \frac{\mathcal{D}_\rho^s(x)}{\rho(x)} dx \leq \int_{\mathbb{S}^{d-1}} \mathcal{D}_\rho^s(x) \cdot \log \frac{1}{p_s} dx = \log \frac{1}{p_s},$$

and similarly, because  $\mathcal{D}_\nu^s(x) = \nu(x|\mathcal{E}_s) \leq \frac{\|\nu\|_\infty}{p_s} \nu(x) \leq \frac{\|\nu\|_\infty}{p_s} \rho(x)$ ,

$$\mathbf{D}(\mathcal{D}_\nu^s \| \rho) = \int_{\mathbb{S}^{d-1}} \mathcal{D}_\nu^s(x) \cdot \log \frac{\mathcal{D}_\nu^s(x)}{\rho(x)} dx \leq \int_{\mathbb{S}^{d-1}} \mathcal{D}_\nu^s(x) \cdot \log \frac{\|\nu\|_\infty}{p_s} dx = \log \frac{1}{p_s} + \log \|\nu\|_\infty.$$

Putting these together with (3) and using  $\sqrt{a+b} \leq \sqrt{a} + \sqrt{b}$ , we have

$$s \leq 2 \sqrt{\frac{2}{d-1} \log \frac{1}{p_s}} + \sqrt{\frac{2}{d-1} \log \|\nu\|_\infty},$$

and then re-arranging we have

$$p_s \leq \exp \left( -\frac{d-1}{8} \left( s - \sqrt{\frac{2}{d-1} \log \|\nu\|_\infty} \right)^2 \right),$$

when  $s \geq \sqrt{\frac{2}{d-1} \log \|\nu\|_\infty}$ . Applying a change of variables completes the proof.  $\square$

Having established in Lemma 4.4 that the optimal transport map  $\pi(\nu, \rho)$  between  $\mathbf{x} \sim \nu$  and  $\mathbf{y} \sim \rho$  has bounded length with high probability, we can translate this into a tail bound for the inner product  $\langle \mathbf{z}, \mathbf{x} - \mathbf{y} \rangle$  for a random vector  $\mathbf{z} \sim \rho$ .

**Lemma 4.5.** *Let  $\nu$  be a distribution on  $\mathbb{S}^{d-1}$ , and let  $\mathcal{D}$  be the optimal transport coupling of  $\nu$  and  $\rho$ . For  $z \in \mathbb{S}^{d-1}, t \in \mathbb{R}^+$  and any  $\kappa > 0$ , define  $X(z, t)$  as:*

$$X(z, t) := \mathbf{Pr}_{(\mathbf{x}, \mathbf{y}) \sim \mathcal{D}} \left[ |\langle \mathbf{z}, \mathbf{x} - \mathbf{y} \rangle| \geq \left( \sqrt{\frac{2}{d-1} \log \|\nu\|_\infty} + \sqrt{\frac{8\kappa}{d-1}} \right) \cdot t \right].$$

Then:

$$\mathbf{Pr}_{\mathbf{z} \sim \rho} [X(\mathbf{z}, t) \geq 2 \exp(-dt^2/4) + \exp(-\kappa)] \leq 2 \exp(-dt^2/4).$$

**Remark 4.6.** One should think of  $X(z, t)$  as a measure of how often a randomly chosen transport vector  $\mathbf{x} - \mathbf{y}$ , with  $(\mathbf{x}, \mathbf{y}) \sim \mathcal{D}$ , has a large projection in the  $z$  direction.

*Proof of Lemma 4.5.* For any  $z \in \mathbb{S}^{d-1}$  and  $(\mathbf{x}, \mathbf{y}) \sim \mathcal{D}$ , suppose  $|\langle z, \mathbf{x} - \mathbf{y} \rangle| \geq \left( \sqrt{\frac{2}{d-1} \log \|v\|_\infty} + \sqrt{\frac{8\kappa}{d-1}} \right) \cdot t$ , then  $|\langle z, \mathbf{x} - \mathbf{y} \rangle| \geq \|\mathbf{x} - \mathbf{y}\| \cdot t$  or  $\|\mathbf{x} - \mathbf{y}\| \geq \sqrt{\frac{2}{d-1} \log \|v\|_\infty} + \sqrt{\frac{8\kappa}{d-1}}$ . Defining

$$Y(z, t) := \Pr_{(\mathbf{x}, \mathbf{y}) \sim \mathcal{D}} [|\langle z, \mathbf{x} - \mathbf{y} \rangle| \geq \|\mathbf{x} - \mathbf{y}\| \cdot t],$$

we can write

$$X(z, t) \leq Y(z, t) + \Pr_{(\mathbf{x}, \mathbf{y}) \sim \mathcal{D}} \left[ \|\mathbf{x} - \mathbf{y}\| \geq \sqrt{\frac{8\kappa}{d-1}} + \sqrt{\frac{2}{d-1} \log \|v\|_\infty} \right] \leq Y(z, t) + \exp(-\kappa) \quad (4)$$

where to obtain the upper bound we have applied Lemma 4.4. Next, we prove that

$$\Pr_{z \sim \rho} [Y(z, t) \geq 2 \exp(-dt^2/4)] \leq 2 \exp(-dt^2/4), \quad (5)$$

by showing  $\mathbf{E}[Y(z, t)] \leq 4 \exp(-dt^2/2)$ , which implies (5) via Markov's inequality.

$$\begin{aligned} \mathbf{E}_{z \sim \rho} [Y(z, t)] &= \Pr_{\substack{z \sim \rho \\ (\mathbf{x}, \mathbf{y}) \sim \mathcal{D}}} [|\langle z, \mathbf{x} - \mathbf{y} \rangle| \geq \|\mathbf{x} - \mathbf{y}\| \cdot t] = \Pr_{\substack{z \sim \rho \\ (\mathbf{x}, \mathbf{y}) \sim \mathcal{D}}} \left[ \left| \langle z, \frac{\mathbf{x} - \mathbf{y}}{\|\mathbf{x} - \mathbf{y}\|} \rangle \right| \geq t \right] \\ &\leq 4 \exp(-dt^2/2). \end{aligned} \quad (\text{by Lemma 3.5})$$

We can then complete the proof starting at (5) as follows:

$$\begin{aligned} \Pr_{z \sim \rho} [Y(z, t) + \exp(-\kappa) \geq 2 \exp(-dt^2/4) + \exp(-\kappa)] &\leq 2 \exp(-dt^2/4) \\ \Pr_{z \sim \rho} [X(z, t) \geq 2 \exp(-dt^2/4) + \exp(-\kappa)] &\leq 2 \exp(-dt^2/4). \end{aligned} \quad (\text{by (4)})$$

This yields the desired conclusion.  $\square$

Since a  $p$ -cap around  $x$  is given by the set of vectors  $z$  with inner product  $\langle x, z \rangle \geq \tau(p)$ , and  $\langle x, z \rangle = \langle y, z \rangle \pm |\langle x - y, z \rangle|$ , we can finally use Lemma 4.5 to relate  $X(z) = \Pr_{\mathbf{x} \sim v} [\langle z, \mathbf{x} \rangle > \tau(p)]$ , the measure of  $v$  that falls into the  $p$ -cap of  $z$ , to  $p = \Pr_{\mathbf{y} \sim \rho} [\langle z, \mathbf{y} \rangle > \tau(p)]$ . That is, we can now show that  $X(z)$  for  $z \sim \rho$  concentrates tightly around  $p$ , so that most vectors in  $\mathbb{S}^{d-1}$  contain very close to a  $p$ -fraction of  $v$ 's mass in their  $p$ -caps.

**Lemma 4.7.** *Let  $v$  be a distribution on  $\mathbb{S}^{d-1}$ . For  $z \in \mathbb{S}^{d-1}$ , let  $X(z) := \Pr_{\mathbf{x} \sim v} [\langle \mathbf{x}, z \rangle > \tau(p)]$ , and for any  $\kappa > 0$ , let  $u(t) := \left( \sqrt{\frac{8}{d-1} \log \|v\|_\infty} + \sqrt{\frac{8\kappa}{d-1}} \right) \cdot t$ . Then for any  $t \geq 0$ :*

$$\Pr_{z \sim \rho} [ |X(z) - p| > p \cdot \varepsilon(t) ] \leq 2 \exp(-dt^2/4)$$

where  $\varepsilon(t) := C_{3.7} \cdot u(t) \cdot \exp(2d\tau(p)u(t)) \cdot \sqrt{d \log \frac{1}{p} + \frac{2 \exp(-dt^2/4) + \exp(-\kappa)}{p}}$ .

*Proof.* Let  $\mathcal{D}$  be the optimal transport coupling between  $v$  and  $\rho$ . For any  $z \in \mathbb{S}^{d-1}$  and  $t \geq 0$ :

$$\begin{aligned} X(z) &= \Pr_{(\mathbf{x}, \mathbf{y}) \sim \mathcal{D}} [\langle \mathbf{y}, z \rangle > \tau(p) - \langle z, \mathbf{x} - \mathbf{y} \rangle] \\ &\leq \Pr_{(\mathbf{x}, \mathbf{y}) \sim \mathcal{D}} [\langle \mathbf{y}, z \rangle > \tau(p) - \max\{\langle z, \mathbf{x} - \mathbf{y} \rangle, u(t)\}] \\ &\leq \Pr_{\mathbf{y} \sim \rho} [\langle \mathbf{y}, z \rangle > \tau(p) - u(t)] + \Pr_{(\mathbf{x}, \mathbf{y}) \sim \mathcal{D}} [|\langle z, \mathbf{x} - \mathbf{y} \rangle| > u(t)] \end{aligned}$$

$$\leq p \left( 1 + C_{3.7} \cdot u(t) \cdot \exp(2d\tau(p)u(t)) \cdot \sqrt{d \log \frac{1}{p}} \right) + \Pr_{(\mathbf{z}, \mathbf{y}) \sim \mathcal{D}} [|\langle \mathbf{z}, \mathbf{x} - \mathbf{y} \rangle| > u(t)] \quad (6)$$

where the last step of the chain of inequalities follows from [Lemma 3.7](#). Identically,

$$X(\mathbf{z}) \geq p \left( 1 - C_{3.7} \cdot u(t) \cdot \exp(2d\tau(p)u(t)) \cdot \sqrt{d \log \frac{1}{p}} \right) - \Pr_{(\mathbf{z}, \mathbf{y}) \sim \mathcal{D}} [|\langle \mathbf{z}, \mathbf{x} - \mathbf{y} \rangle| > u(t)] \quad (7)$$

Then by [Lemma 4.5](#), when  $\mathbf{z} \sim \rho$ , we can obtain an upper bound of  $2\exp(-dt^2/4) + \exp(-\kappa)$  on the second term in the RHS of (6) and (7) that holds except with probability  $2\exp(-dt^2/4)$ , which implies

$$\Pr_{\mathbf{z} \sim \rho} \left[ |X(\mathbf{z}) - p| > p \cdot C_{3.7} \cdot u(t) \cdot \exp(2d\tau(p)u(t)) \cdot \sqrt{d \log \frac{1}{p}} + 2\exp(-dt^2/4) + \exp(-\kappa) \right] \leq 2\exp(-dt^2/4).$$

thus completing the proof.  $\square$

### 4.3 Different parameterizations of the sphere cap concentration

We'll now derive a few useful corollaries of [Lemma 4.7](#). First, for intuition, consider the following immediate consequence regarding the intersection of a set in  $\mathbb{S}^{d-1}$  with a random sphere cap.

**Corollary 4.8.** *Let  $d \geq \log^{10} n$  and let  $Q \subseteq \mathbb{S}^{d-1}$  be a set such that  $\rho(Q) \geq \frac{1}{n^{\log^3 n}}$ . Then for  $\mathbf{z} \sim \rho$ :*

$$\frac{\rho(\text{cap}(\mathbf{z}) \cap Q)}{\rho(\text{cap}(\mathbf{z})) \cdot \rho(Q)} \notin \left( 1 \pm \frac{\log^5 n}{\sqrt{d}} \pm n^{-\log^2 n} \right)$$

with probability at most  $n^{-\Omega(\log^3 n)}$ .

We will make use of the following convenient specialization of [Lemma 4.7](#).

**Corollary 4.9.** *Let  $\nu$  be a distribution on  $\mathbb{S}^{d-1}$  and for  $z \in \mathbb{S}^{d-1}$ , let  $X(z) := \Pr_{\mathbf{x} \sim \nu} [\langle \mathbf{x}, z \rangle \geq \tau(p)]$ . Then for  $s \leq 1$  and for some constant  $C_{4.9}$ ,*

$$\Pr_{\mathbf{z} \sim \rho} [|X(\mathbf{z}) - p| > ps] \leq 2 \exp \left( - \frac{ds^2}{C_{4.9} \left( \sqrt{\log \|\nu\|_\infty} + \sqrt{\log \frac{d}{p}} \right)^2 \cdot \log \frac{1}{p} \cdot \log \frac{d}{p}} \right)$$

*Proof.* The idea is to apply [Lemma 4.7](#) by setting  $\kappa = 4 \log \frac{d}{p}$  and using the parameterization

$$t = \frac{s}{2 \left( \sqrt{8 \log \|\nu\|_\infty} + \sqrt{32 \log \frac{1}{p}} \right) \sqrt{\log \frac{1}{p}}}.$$

Clearly, the statement is true when

$$s \leq \sqrt{\frac{C_{4.9}}{2d}} \cdot \left( \sqrt{\log \|\nu\|_\infty} + \sqrt{\log \frac{d}{p}} \right) \cdot \sqrt{\log \frac{1}{p} \cdot \log \frac{d}{p}}$$

since  $\Pr [|X(z) - p| > ps] \leq 1 < 2 \exp(-1/2)$ . Hence, we restrict our attention to when

$$s \in H := \left[ \sqrt{\frac{C_{4.9}}{2d}} \cdot \left( \sqrt{\log \|\nu\|_\infty} + \sqrt{\log \frac{d}{p}} \right) \cdot \sqrt{\log \frac{1}{p} \cdot \log \frac{d}{p}}, 1 \right].$$

Let  $\varepsilon(t)$  and  $u(t)$  be as in the statement of [Lemma 4.7](#). Using  $s \leq 1$  and [Lemma 3.6](#), we know  $\exp(2d\tau(p)u(t)) \leq O(1)$ . This tells us that for some constant  $C$ ,

$$\varepsilon(t) \leq C \left( \sqrt{\frac{8}{d-1} \log \|v\|_\infty} + \sqrt{\frac{32 \log \frac{1}{p}}{d-1}} \right) \sqrt{d \log \frac{1}{p} t} + \frac{2 \exp(-dt^2/4) + \frac{p^4}{d^4}}{p}.$$

We can choose our constant  $C_{4.9}$  to be large enough so that when  $s \in H$ , then  $t \geq 4\sqrt{\frac{\log \frac{d}{p}}{d}}$ . Observe that once  $t \geq 4\sqrt{\frac{\log \frac{d}{p}}{d}}$ , for large enough  $d$ :

$$\begin{aligned} \varepsilon(t) &\leq 2C \left( \sqrt{\frac{8}{d-1} \log \|v\|_\infty} + \sqrt{\frac{32 \log \frac{1}{p}}{d-1}} \right) \sqrt{d \log \frac{1}{p} t} \\ &= C' \cdot 2 \left( \sqrt{8 \log \|v\|_\infty} + \sqrt{32 \log \frac{1}{p}} \right) \cdot \sqrt{\log \frac{1}{p} \cdot t} \\ &= C' s \end{aligned}$$

for some other constant  $C'$ . Then by [Lemma 4.7](#), when  $s \in H$ :

$$\begin{aligned} \Pr[|X(z) - p| > ps] &\leq 2 \exp \left( -\frac{ds^2}{4C'^2 \left( \sqrt{8 \log \|v\|_\infty} + \sqrt{32 \log \frac{1}{p}} \right)^2 \log \frac{1}{p}} \right) \\ &\leq 2 \exp \left( -\frac{ds^2}{C_{4.9} \left( \sqrt{\log \|v\|_\infty} + \sqrt{\log \frac{d}{p}} \right)^2 \cdot \log \frac{1}{p} \cdot \log \frac{d}{p}} \right) \end{aligned}$$

where the second inequality arises from choosing  $C_{4.9}$  to be large enough, which completes the proof.  $\square$

[Corollary 4.9](#) can be extended to the case when  $\mathbf{z} \sim \mu$  with a worse quantitative upper bound depending on  $\mu$  from [Observation 3.2](#).

**Corollary 4.10.** *Let  $\nu$  and  $\mu$  be distributions on  $\mathbb{S}^{d-1}$  and for  $z \in \mathbb{S}^{d-1}$ , let  $X(z) := \Pr_{\mathbf{x} \sim \nu}[\langle \mathbf{x}, z \rangle > \tau(p)]$ . Then for  $s \leq 1$  we have:*

$$\Pr_{\mathbf{z} \sim \mu}[|X(\mathbf{z}) - p| > ps] \leq 2 \exp \left( -\frac{ds^2}{C_{4.9} \left( \sqrt{\log \|v\|_\infty} + \sqrt{\log \frac{d}{p}} \right)^2 \cdot \log \frac{1}{p} \cdot \log \frac{d}{p}} \right) \cdot \|\mu\|_\infty$$

## 5 Concentration for intersections of caps and anti-caps

In this section, we prove concentration of measure for the intersection of random  $p$ -caps and  $p$ -anticaps with any fixed set  $L \subseteq \mathbb{S}^{d-1}$ .

**Lemma 5.1** (Concentration for the intersection of  $j$  caps and  $k - j$  anti-caps). *Let  $\rho$  be the uniform measure over  $\mathbb{S}^{d-1}$ , and let  $L \subset \mathbb{S}^{d-1}$ , and let  $k > 0$  be an integer. For  $(\mathbf{v}_1, \dots, \mathbf{v}_k) \sim \rho^{\otimes k}$ , let*

$$\mathbf{S}_i := \begin{cases} \text{cap}(\mathbf{v}_i) & \text{if } i \leq j \\ \overline{\text{cap}}(\mathbf{v}_i) & \text{if } i > j, \end{cases}$$

and let  $\mathbf{L}_t = L \cap \bigcap_{i=1}^t \mathbf{S}_i$ . Then the ratio  $\mathbf{R} := \frac{\rho(\mathbf{L}_t)}{\rho(L)p^j(1-p)^{k-j}}$  is concentrated as follows:

$$\Pr[|\mathbf{R} - 1| > s] \leq \varepsilon_1(s) + k\varepsilon_1(.5) + \varepsilon_2$$

where

$$\varepsilon_1(s) := 2 \exp\left(-\frac{ds^2}{C'(j + (k-j)p^2)F(j)}\right) \\ \varepsilon_2 := \frac{4k}{p^2} \exp\left(-\frac{d}{CF(j)}\right)$$

$F(j) := \left(\sqrt{\log \frac{1}{\rho(L)} + j \log \frac{1}{p} + (k-j) \log \frac{1}{1-p}} + \sqrt{\log \frac{d}{p}}\right)^2 \log \frac{1}{p} \log \frac{d}{p}$ , and  $C, C' > 0$  are universal constants.

A caricature of the proof of the above lemma is the following. The main takeaway from [Section 4](#) is that the intersection of a random  $p$ -cap with a set  $L$  has measure  $(1 \pm \varepsilon)p\rho(L)$  where the fluctuation  $\varepsilon$  is centered and has typical magnitude  $\tilde{O}\left(\sqrt{\log \frac{1}{\rho(L)}/d}\right)$ . As an upshot the intersection of a random  $p$ -anti-cap with  $L$  has measure  $\rho(L) - (1 \pm \varepsilon)p\rho(L) = \left(1 \pm \frac{p\varepsilon}{1-p}\right)(1-p)\rho(L)$ . Thus, when intersecting  $j$  random caps and  $(k-j)$  random anticaps, we expect a multiplicative fluctuation that resembles  $\left(1 \pm \left(\sqrt{j}\varepsilon + \sqrt{k-j} \cdot \frac{p}{1-p} \cdot \varepsilon\right)\right)$ . Since the fluctuation term per anti-cap is roughly  $p\varepsilon$  instead of merely  $\varepsilon$ , we get extremely strong concentration for intersections of anti-caps, as articulated by the below corollary. For example, in the  $p = \frac{a}{n}$  regime, an intersection of  $m = \Theta(n)$  anti-caps will have measure  $(1-p)^m(1 \pm \varepsilon)$  with typical deviations  $\varepsilon = o(\sqrt{1/n})$  if  $d > \text{poly} \log n$  for a sufficiently large power of  $\log n$ . (See also [Corollary 6.1](#) for another application of [Lemma 5.1](#)).

**Corollary 5.2** (Intersection of anti-caps). *Suppose  $m, n, d \in \mathbb{Z}_+$ ,  $a \geq 1$ , and  $p \in \mathbb{R}_+$  satisfy  $m \leq n$ ,  $d \leq n^{100}$ ,  $\frac{1}{n^2} \ll p \leq \frac{1}{2}$ , and  $mp \leq \log^a n$  for  $n$  sufficiently large. Let  $L \subset \mathbb{S}^{d-1}$  with  $\rho(L) \geq e^{-\log^a n}$ . Let  $\mathbf{v}_1, \dots, \mathbf{v}_m \sim \mathbb{S}^{d-1}$  uniformly at random, and let  $\mathbf{A} = \{w \in \mathbb{S}^{d-1} \mid \langle w, \mathbf{v}_i \rangle \leq \tau(p) \forall i \in [m]\}$  be the intersection of anti-caps of the  $\mathbf{v}_i$ 's. Then there exist universal constants  $C, C'$  such that for all  $n$  sufficiently large, for all  $t \geq 0$ ,*

$$\Pr_{\mathbf{A}} \left[ \left| \frac{\rho(\mathbf{A} \cap L)}{(1-p)^m \rho(L)} - 1 \right| > t \cdot \sqrt{mp^2} \right] \leq \exp \left( -C \min \left\{ \frac{t^2 \cdot d}{(\log n)^{a+2}}, \frac{\sqrt{d} - C'(\log n)^{a+3}}{(\log n)^{a+2}} \right\} \right).$$

*Proof.* Using that  $\log(1+x) \leq x$  for  $x > 0$ ,

$$\frac{1}{\log \frac{1}{p} \log \frac{d}{p}} \sqrt{F(j)} = \sqrt{\log \frac{1}{\rho(L)} + m \log(1 + \frac{p}{1-p})} + \sqrt{\log \frac{d}{p}} \leq \sqrt{\log \frac{1}{\rho(L)}} + \sqrt{m \frac{p}{1-p}} + \sqrt{\log \frac{d}{p}} \leq c \cdot \log^a n,$$

for  $c > 0$  a universal constant, where in the final inequality we have applied the assumption  $p \leq \frac{1}{2}$  and the assumption  $p \gg n^{-2}$ . Re-arranging, we have  $F(j) \leq c'(\log n)^{a+2}$  for some constant  $c'$ . Hence, [Lemma 5.1](#) implies that there exist constants  $c'', c'''$  so that for  $n$  sufficiently large,

$$\Pr_{\mathbf{A}} \left[ \left| \frac{\rho(\mathbf{A} \cap L)}{(1-p)^m \rho(L)} - 1 \right| > s \right] \leq 2 \exp \left( -c'' \frac{ds^2}{mp^2(\log n)^{a+2}} \right) + \left( 2k + \frac{4m}{p^2} \right) \cdot \exp \left( -c''' \frac{d}{(\log n)^{a+2}} \right),$$

where we used our bounds on  $p, k, m$  to combine the  $\varepsilon_2$  and  $k\varepsilon_1(\frac{1}{2})$  terms from [Lemma 5.1](#) into our second term. The conclusion now follows by substituting  $s = t\sqrt{mp^2}$  and applying asymptotic simplifications.  $\square$

Our proof proceeds by a martingale argument. First, we notice that the (rescaled) area of the intersection of sets is a martingale.

**Observation 5.3** (Scaled intersection is a Martingale). Define  $\mathbf{R}_t := \frac{\rho(\mathbf{L}_t)}{\rho(L)\prod_{i=1}^t \rho(\mathbf{S}_i)}$ , and  $\mathbf{L}_t = L \cap \bigcap_{i=1}^t \mathbf{S}_i$ , as introduced in [Lemma 5.1](#). Then  $(\mathbf{R}_t)_{t \in [k]}$  is a martingale with respect to the filtration  $(\mathcal{V}_t)_{t \in [k]}$  induced by  $\mathbf{v}_1, \mathbf{v}_2, \dots$ , with  $\mathbf{E}[\mathbf{R}_t | \mathcal{V}_{t-1}] = \mathbf{R}_{t-1}$  and  $|\mathbf{R}_t| \leq \frac{1}{\rho(L)p^j(1-p)^{k-j}}$ .

*Proof.* The quantities  $\rho(\mathbf{S}_i)$  are fixed for all  $i$ . To see that  $\mathbf{R}_t \geq 0$  is bounded, note that  $\mathbf{R}_t \leq \frac{1}{p^j(1-p)^{k-j}}$ , as  $\rho(\mathbf{L}_t) \leq \rho(L)$ . Now, note that by definition  $\mathbf{R}_t = \frac{\rho(\mathbf{S}_t \cap \mathbf{L}_{t-1})}{\rho(L)\prod_{i=1}^t \rho(\mathbf{S}_i)}$ . Since  $\mathbf{v}_t \sim \mathbb{S}^{d-1}$  independently of  $\mathcal{V}_{t-1}$ ,  $\mathbf{E}_{\mathbf{v}_t \sim \mathbb{S}^{d-1}}[\rho(\mathbf{S}_t \cap \mathbf{L}_{t-1})] = \rho(\mathbf{S}_t) \cdot \rho(\mathbf{L}_{t-1})$ . The conclusion now follows, since  $\mathbf{E}_{\mathbf{v}_t \sim \mathbb{S}^{d-1}}[\mathbf{R}_t] = \mathbf{R}_{t-1}$ .  $\square$

Now, we will prove concentration for the martingale introduced in [Observation 5.3](#), using the concentration inequality for martingales with subgaussian increments, [Lemma 3.9](#). The  $(\mathbf{R}_t)_{t \in [k]}$  do not quite have subgaussian increments as described. Thus we must “tame” them by making some minor technical modifications.

*Proof of Lemma 5.1.* Recall the martingale sequence  $\mathbf{R}_t$  defined in [Observation 5.3](#). We are then interested in deviation bounds for  $|\mathbf{R}_k - 1|$ . Our proof strategy is to couple  $\mathbf{R}_t$  with a more well-behaved martingale sequence and use the subgaussian martingale concentration inequality on the more well-behaved sequence. Let  $\mathbf{T}$  be the first time at which  $\mathbf{R}_T > 2$  or  $\mathbf{R}_T < \frac{1}{2}$ . Define the process  $(\mathbf{Q}_t)_{t \geq 1}$  so that

$$\mathbf{Q}_i = \begin{cases} \mathbf{R}_i & i \leq \mathbf{T} \\ \mathbf{R}_T & \text{otherwise.} \end{cases}$$

Note that the sequence  $(\mathbf{Q}_t)_{t \geq 1}$  is a martingale. This is because when  $t \leq \mathbf{T}$ ,  $\mathbf{Q}_t - \mathbf{Q}_{t-1} | \mathcal{V}_{t-1}$  has the same distribution as  $\mathbf{R}_t - \mathbf{R}_{t-1} | \mathcal{V}_{t-1}$  and hence  $\mathbf{E}[\mathbf{Q}_t - \mathbf{Q}_{t-1} | \mathcal{V}_{t-1}] = 0$ , and when  $t \geq \mathbf{T}$ ,  $\mathbf{Q}_t - \mathbf{Q}_{t-1} | \mathcal{V}_{t-1}$  is identically zero. We will need to make an additional modification on top of  $\mathbf{Q}_i$  to make the random variable well-behaved, since [Corollary 4.9](#) only guarantees that  $\Pr[|\mathbf{Q}_i - \mathbf{Q}_{i-1}| > s]$  remains subgaussian up to  $s = 1$ . We will truncate  $\mathbf{Q}_i$  as described by the following definition:

**Definition 5.4** ( $(\alpha, \beta)$ -truncation). Given a centered random variable  $\mathbf{X}$ , we define a random variable  $\mathbf{X}_{\alpha, \beta}$  for  $\alpha, \beta \in \mathbb{R}_{\geq 0}$ . First, let  $\theta = \min\left(\frac{\Pr[|\mathbf{X}| > \alpha]}{\beta}, 1\right)$ . Now, define

$$\mathbf{X}_{\alpha, \beta} := \begin{cases} \mathbf{X} & \text{with probability } (1 - \theta)\Pr[|\mathbf{X}| \leq \alpha] \\ \mathbf{E}[\mathbf{X} | |\mathbf{X}| > \alpha] \cdot \beta & \text{with probability } (1 - \theta) \cdot \theta \\ 0 & \text{otherwise.} \end{cases}$$

$\mathbf{X}_{\alpha, \beta}$  is well-defined; in the case when  $\theta = 1$  it takes value 0 deterministically, and otherwise if  $\theta < 1$  the probabilities sum to 1.

We make two useful observations about  $(\alpha, \beta)$ -truncations.

**Observation 5.5.** For a random variable  $\mathbf{X}$  and parameters  $\alpha, \beta$ ,  $\mathbf{E}\mathbf{X}_{\alpha, \beta} = (1 - \theta)\mathbf{E}\mathbf{X}$ .

**Observation 5.6.** For a random variable  $\mathbf{X}$  and parameters  $\alpha, \beta$  satisfying  $\beta \leq 1$ ,  $d_{\text{TV}}(\mathbf{X}, \mathbf{X}_{\alpha, \beta}) \leq 2\theta$ .

We prove these observations at the end of the section; for now, we proceed with the proof of [Lemma 5.1](#).

Define  $\Delta_i := \mathbf{Q}_i - \mathbf{Q}_{i-1}$ , and define  $\tilde{\Delta}_i | \mathcal{V}_{i-1} := (\Delta_i | \mathcal{V}_{i-1})_{\alpha, \beta}$  where

$$\alpha = \begin{cases} 1 & \text{if } i \leq j \\ \frac{p}{1-p} & \text{if } i > j. \end{cases}$$

and  $\beta = p^2\alpha$ . We have chosen  $\alpha$  such that [Corollary 4.9](#) guarantees that we can control the tail probability  $\Pr[|\Delta_i| > t]$  for  $t < \alpha$ . Since  $|\Delta_i| \leq \frac{1}{p}$ ,  $|\mathbf{E}[\Delta_i | |\Delta_i| > \alpha]| \cdot \beta \leq p\alpha$ , and consequently  $|\tilde{\Delta}_i| \leq \alpha$ . Let  $(\tilde{\mathbf{Q}}_t)_{t \geq 1}$  be the random process obtained by setting  $\tilde{\mathbf{Q}}_1 = \mathbf{Q}_1$  and  $\tilde{\mathbf{Q}}_t = \tilde{\mathbf{Q}}_{t-1} + \tilde{\Delta}_t$  for  $t \geq 2$ . By [Observation 5.5](#),  $(\tilde{\mathbf{Q}}_t)_{t \geq 1}$  is a martingale.

Next, we obtain bounds on the subgaussian norms of  $\tilde{\Delta}_i | \mathcal{V}_{i-1}$ . For  $\mathcal{V}_{i-1}$  such that  $i > \mathbf{T}$ ,  $\|\tilde{\Delta}_i | \mathcal{V}_{i-1}\|_{\psi_2} = 0$ . For any  $\mathcal{V}_{i-1}$  where  $i \leq \mathbf{T}$  and  $s \geq p\alpha$ :

$$\begin{aligned} \Pr[|\tilde{\Delta}_i| > s | \mathcal{V}_{i-1}] &\leq \Pr[|\Delta_i| > s | \mathcal{V}_{i-1}] \mathbf{1}[s \leq \alpha] \\ &= \Pr[|\mathbf{R}_i - \mathbf{R}_{i-1}| > s | \mathcal{V}_{i-1}] \mathbf{1}[s \leq \alpha] \\ &= \Pr\left[\mathbf{R}_{i-1} \cdot \left| \frac{\rho(\mathbf{S}_i \cap \mathbf{L}_{i-1})}{\rho(\mathbf{S}_i) \cdot \rho(\mathbf{L}_{i-1})} - 1 \right| > s | \mathcal{V}_{i-1}\right] \mathbf{1}[s \leq \alpha]. \end{aligned} \quad (8)$$

Since  $i \leq \mathbf{T}$ , we know  $\mathbf{R}_{i-1} \leq 2$  and  $\rho(\mathbf{L}_{i-1}) \geq \frac{\rho(L) \prod_{t=1}^i \rho(\mathbf{S}_t)}{2} \geq \frac{\rho(L) p^j (1-p)^{k-j}}{2}$ . When  $i \leq j$ :

$$\Pr[|\tilde{\Delta}_i| > s | \mathcal{V}_{i-1}] \leq 2 \exp\left(-\frac{ds^2}{C \cdot \left(\sqrt{\log \frac{1}{\rho(L)}} + j \log \frac{1}{p} + (k-j) \log \frac{1}{1-p} + \sqrt{\log \frac{d}{p}}\right)^2 \log \frac{1}{p} \log \frac{d}{p}}\right)$$

and when  $i > j$ :

$$\Pr[|\tilde{\Delta}_i| > s | \mathcal{V}_{i-1}] \leq 2 \exp\left(-\frac{ds^2}{C \cdot p^2 \cdot \left(\sqrt{\log \frac{1}{\rho(L)}} + j \log \frac{1}{p} + (k-j) \log \frac{1}{1-p} + \sqrt{\log \frac{d}{p}}\right)^2 \log \frac{1}{p} \log \frac{d}{p}}\right)$$

for all  $s > 0$  and for some constant  $C > 0$ , where the case of  $s \leq p\alpha$  holds vacuously since the right hand side exceeds 1, and the case of  $s > p\alpha$  holds by (8) and [Corollary 4.9](#). Using  $F(j)$  to denote

$$\left(\sqrt{\log \frac{1}{\rho(L)}} + j \log \frac{1}{p} + (k-j) \log \frac{1}{1-p} + \sqrt{\log \frac{d}{p}}\right)^2 \log \frac{1}{p} \log \frac{d}{p},$$

we can more legibly write the above as:

$$\Pr[|\tilde{\Delta}_i| > s | \mathcal{V}_{i-1}] \leq 2 \exp\left(-\frac{ds^2}{C p^{2 \cdot \mathbf{1}[i > j]} \cdot F(j)}\right)$$

Consequently,  $\|\tilde{\Delta}_i | \mathcal{V}_{i-1}\|_{\psi_2} \leq C p^{\mathbf{1}[i > j]} \cdot \sqrt{\frac{F(j)}{d}}$  for all  $\mathcal{V}_{i-1}$  by [Lemma 3.10](#).

By the subgaussian martingale concentration inequality ([Lemma 3.9](#)), for some constant  $C' > 0$ :

$$\Pr[|\tilde{\mathbf{Q}}_k - 1| \geq s] \leq \varepsilon_1(s) := 2 \exp\left(-\frac{ds^2}{C'(j + (k-j)p^2)F(j)}\right).$$

By a union bound we also know:

$$\Pr[\exists i : |\tilde{\mathbf{Q}}_i - 1| > s] \leq k\varepsilon_1(s).$$

It remains to relate  $\tilde{\mathbf{Q}}_k$  back to  $\mathbf{R}_k$ . By [Observation 5.6](#), and a union bound, there is a coupling between the sequence of  $\tilde{\Delta}_i$  and the sequence of  $\Delta_i$  such that  $\tilde{\Delta}_i = \Delta_i$  for all  $i$  except with probability at most  $2k \max_{i \in [k]} \frac{\Pr[\Delta_i > \alpha]}{p\alpha}$ , which with [Corollary 4.9](#) we can bound by

$$\varepsilon_2 := \frac{4k}{p^2} \exp\left(-\frac{d}{CF(j)}\right)$$

The upshot of the above is:

$$\Pr[\exists i \in [k] : |\mathbf{Q}_i - 1| > s \text{ or } \mathbf{Q}_i \neq \tilde{\mathbf{Q}}_i] = \Pr[\exists i \in [k] : |\tilde{\mathbf{Q}}_i - 1| > s \text{ or } \mathbf{Q}_i \neq \tilde{\mathbf{Q}}_i] \leq k\varepsilon_1(s) + \varepsilon_2.$$

Suppose  $|\mathbf{Q}_i - 1| < .5$  for  $i = 1, \dots, k$ , then  $\mathbf{Q}_i = \mathbf{R}_i$  for  $i = 1, \dots, k$ . Thus:

$$\begin{aligned} \Pr[\exists i \in [k] : \tilde{\mathbf{Q}}_i \neq \mathbf{R}_i] &\leq \Pr[\exists i \in [k] : \tilde{\mathbf{Q}}_i \neq \mathbf{R}_i \text{ or } \mathbf{Q}_i \neq \tilde{\mathbf{Q}}_i] \\ &= \Pr[\exists i \in [k] : \mathbf{Q}_i \neq \mathbf{R}_i \text{ or } \mathbf{Q}_i \neq \tilde{\mathbf{Q}}_i] \\ &\leq k\varepsilon_1(.5) + \varepsilon_2. \end{aligned}$$

Consequently,

$$\begin{aligned} \Pr[|\mathbf{R}_k - 1| > s] &\leq \Pr[|\tilde{\mathbf{Q}}_k - 1| > s] + \Pr[\tilde{\mathbf{Q}}_k \neq \mathbf{R}_k] \\ &\leq \varepsilon_1(s) + k\varepsilon_1(.5) + \varepsilon_2. \end{aligned} \quad \square$$

We now give the deferred proofs of our observations regarding truncated variables.

*Proof of [Observation 5.5](#).* If  $\theta = 1$ ,  $\mathbf{X}_{\alpha, \beta} = 0$  and the statement holds. Otherwise (and actually in any case), we can write:

$$\begin{aligned} \mathbf{E} \mathbf{X}_{\alpha, \beta} &= (1 - \theta) \Pr[|\mathbf{X}| \leq \alpha] \mathbf{E}[\mathbf{X} \mid |\mathbf{X}| \leq \alpha] + (1 - \theta) \cdot \frac{\Pr[|\mathbf{X}| > \alpha]}{\beta} \cdot \mathbf{E}[\mathbf{X} \mid |\mathbf{X}| > \alpha] \cdot \beta \\ &= (1 - \theta) \mathbf{E}[\mathbf{X} \cdot \mathbf{1}[|\mathbf{X}| \leq \alpha]] + (1 - \theta) \mathbf{E}[\mathbf{X} \cdot \mathbf{1}[|\mathbf{X}| > \alpha]] = (1 - \theta) \mathbf{E} \mathbf{X}. \end{aligned} \quad \square$$

*Proof of [Observation 5.6](#).* If  $\theta = 1$ , the bound holds trivially, so assume that  $\theta < 1$ . We can couple  $\mathbf{X}$  and  $\mathbf{X}_{\alpha, \beta}$  by sampling  $\mathbf{X}_{\alpha, \beta}$  in the following way: sample  $\mathbf{c} \sim \text{Ber}(1 - \theta)$  and  $\mathbf{X}$ . If  $\mathbf{c} = 1$  and  $|\mathbf{X}| \leq \alpha$ , let  $\mathbf{X}_{\alpha, \beta} = \mathbf{X}$ ; otherwise let  $\mathbf{X}_{\alpha, \beta} = \mathbf{E}[\mathbf{X} \mid |\mathbf{X}| > \alpha] \cdot \beta$  with probability  $(1 - \theta)\theta$  and 0 with the remaining probability. If  $\beta \leq 1$ , then by definition  $\theta \geq \beta\theta = \Pr[|\mathbf{X}| > \alpha]$ . Hence under this coupling,  $\mathbf{X} \neq \mathbf{X}_{\alpha, \beta}$  with probability at most  $\theta + \beta\theta \leq 2\theta$ , yielding the total variation bound.  $\square$

## 6 Stochastic domination by Erdős-Rényi graphs

Here, we will prove [Proposition 1.3](#): we'll show that so long as  $d = \tilde{\Omega}((np)^2)$ , there is a three-way ‘‘coupling’’ between  $\mathbf{G} \sim \text{Geo}_d(n, p)$ ,  $\mathbf{G}_- \sim \text{G}(n, (1 - \tilde{O}(\sqrt{\frac{np}{d}}))p)$ , and  $\mathbf{G}_+ \sim \text{G}(n, (1 + \tilde{O}(\sqrt{\frac{np}{d}}))p)$  such that with high probability  $\mathbf{G}$  is sandwiched between  $\mathbf{G}_-$  and  $\mathbf{G}_+$ . This will be crucial for the proof of [Theorem 1.1](#) in reasoning about the structure of random geometric graphs.

We will need the following corollary of [Lemma 5.1](#), which lower bounds the area of intersections of  $j$  random  $p$ -caps and  $k - j$  random anti- $p$ -caps.

**Corollary 6.1** (Area of  $\approx pk$  caps and  $\approx (1-p)k$  anti-caps). *Let  $n, j, k, d \in \mathbb{Z}_+$ , and let  $p \in \mathbb{R}_+$ , satisfying  $\frac{1}{n^2} \ll p \leq \frac{1}{2}$ ,  $1 \leq d \leq n^{100}$ ,  $j \leq k \leq n$ , with  $j = pk + \Delta$ . Draw  $\mathbf{v}_1, \dots, \mathbf{v}_k$  uniformly from  $\mathbb{S}^{d-1}$  and let  $\mathbf{L} = \left(\bigcap_{i=1}^j \text{cap}(\mathbf{v}_i)\right) \cap \left(\bigcap_{i=j+1}^k \overline{\text{cap}}(\mathbf{v}_i)\right)$ . Then there exist constants  $C_{6.1}, C'_{6.1} > 0$  such that for all  $n$  sufficiently large,*

$$\Pr \left[ \left| \frac{\rho(\mathbf{L})}{e^{-kH(p)} \left(\frac{p}{1-p}\right)^\Delta} - 1 \right| \geq t \right] \leq \exp \left( -C_{6.1} \cdot \min \left( \frac{dt^2}{M(k, p, \Delta)^2 \log n}, \frac{d}{M(k, p, \Delta)^2 \log n} - C'_{6.1} \log n \right) \right),$$

where  $M(k, p, \Delta) = \max(kH(p), |\Delta| \log \frac{1}{p}, \log n)$ .

*Proof.* We will apply [Lemma 5.1](#). Applying asymptotic simplifications, in our case,

$$\frac{\sqrt{F(j)}}{\log \frac{1}{p} \log \frac{d}{p}} = \sqrt{kH(p) + \Delta \log \frac{1-p}{p}} + \sqrt{\log \frac{d}{p}} \leq c \cdot \left( \sqrt{kH(p) + |\Delta| \log \frac{1}{p}} + \sqrt{\log n} \right),$$

for  $c > 0$  a universal constant. Re-arranging and applying the inequality  $a^2 + b^2 \geq \frac{1}{2}ab$ ,

$$F(j) \leq c' \cdot \log \frac{1}{p} \log n \left( kH(p) + |\Delta| \log \frac{1}{p} + \log n \right).$$

Now applying [Lemma 5.1](#) and letting  $B = kH(p) + |\Delta| \log \frac{1}{p}$  (and using  $j + (k-j)p^2 = kp + (1-p^2)\Delta + k(1-p)p^2 \leq 2kp + |\Delta|$ , and hence  $F(j)(j + (k-j)p^2) \leq c''(B + \log n)B \log n$  for some constant  $c''$ ),

$$\begin{aligned} \Pr \left[ \left| \frac{\rho(\mathbf{L})}{e^{-kH(p)} \left(\frac{p}{1-p}\right)^\Delta} - 1 \right| \geq t \right] &\leq 2 \exp \left( -c_1 \frac{ds^2}{(B + \log n)B \log n} \right) + 2k \exp \left( -c_2 \frac{d}{(B + \log n)B \log n} \right) \\ &\quad + \frac{4k}{p^2} \exp \left( -c_3 \frac{d}{(B + \log n) \log \frac{1}{p} \log n} \right) \\ &\leq 2 \exp \left( -c_1 \frac{ds^2}{M^2 \log n} \right) + \left( 2k + \frac{4k}{p^2} \right) \exp \left( -c_3 \frac{d}{M^2 \log n} \right), \end{aligned}$$

for universal constants  $c_1, c_2, c_3 > 0$  and  $M = \max(kH(p), |\Delta| \log \frac{1}{p}, \log n)$ . Applying asymptotic simplifications yields our desired result.  $\square$

**Proposition** (Restatement of [Proposition 1.3](#)). *For any constant  $\alpha > 0$  there exist constants  $C_1, C_2 > 0$  such that if  $\frac{\alpha}{n} \leq p \leq \frac{1}{2}$  and  $d \geq C_1 \cdot (n^2 p^2 + \log^4 n) \log^4 n$ , for any  $\varepsilon \geq C_2 \sqrt{\frac{1}{d}(np + \log n) \log^4 n}$ , one can simultaneously sample  $\mathbf{G}_- \sim \mathbf{G}(n, (1-\varepsilon)p)$ ,  $\mathbf{G} \sim \text{Geo}_d(n, p)$ , and  $\mathbf{G}_+ \sim \mathbf{G}(n, (1+\varepsilon)p)$  in a correlated manner so that with probability at least  $1 - n^{-\Omega(\log n)}$ ,  $\mathbf{G}_- \subseteq \mathbf{G} \subseteq \mathbf{G}_+$ .*

*Proof of Proposition 1.3.* We describe how to sample  $\mathbf{G}_-, \mathbf{G}, \mathbf{G}_+$  jointly so that the above holds. We consider vertices arriving one at a time, and we'll determine the adjacency of the arriving vertex to all previous vertices in a way that correlates  $\mathbf{G}_-, \mathbf{G}, \mathbf{G}_+$ . In  $\mathbf{G}$ , each vertex is identified with a vector; an arriving vertex's vector starts with the entirety of  $\mathbb{S}^{d-1}$ , then iteratively refines the set of candidates by intersecting with the spherical caps (or anti-caps) of the previous vectors.

1. Choose  $\mathbf{v}_1 \sim \mathbb{S}^{d-1}$  uniformly at random.

2. For each  $i \in \{2, \dots, n\}$ :
  - (a) Initialize the set of candidates for vector  $\mathbf{v}_i$  as  $\mathbf{A}_{i,0} = \mathbb{S}^{d-1}$ .
  - (b) For each  $k \in \{1, \dots, i-1\}$ :
    - i. Choose a threshold value  $\theta \sim \text{Unif}([0, 1])$ .
    - ii. If  $\theta < (1 - \varepsilon)p$ , add the edge  $(i, k)$  to  $\mathbf{G}_-$ .
    - iii. Let  $\mathbf{S}_k = \{u \in \mathbb{S}^{d-1} \mid \langle u, \mathbf{v}_k \rangle \geq \tau(p)\}$ . If  $\theta < \frac{\rho(\mathbf{A}_{i,k-1} \cap \mathbf{S}_k)}{\rho(\mathbf{A}_{i,k-1})}$ , let  $\mathbf{A}_{i,k} = \mathbf{A}_{i,k-1} \cap \mathbf{S}_k$ . Otherwise let  $\mathbf{A}_{i,k} = \mathbf{A}_{i,k-1} \cap \overline{\mathbf{S}_k}$ .
    - iv. If  $\theta < (1 + \varepsilon)p$ , add the edge  $(i, k)$  to  $\mathbf{G}_+$ .
  - (c) Choose  $\mathbf{v}_i$  uniformly at random from  $\mathbf{A}_{i,i-1}$ .
3. Set  $\mathbf{G} = \text{gg}_{\tau(p)}(\mathbf{v}_1, \dots, \mathbf{v}_n)$ .

The marginal distributions over  $\mathbf{G}_-$  and  $\mathbf{G}_+$  are clearly  $\text{G}(n, (1 - \varepsilon)p)$  and  $\text{G}(n, (1 + \varepsilon)p)$  respectively. The marginal distribution over  $\mathbf{G}$  is  $\text{Geo}_d(n, p)$ : since we choose  $\mathbf{A}_{i,k} = \mathbf{A}_{i,k-1} \cap \mathbf{S}_k$  independently with probability proportional to the measure  $\rho(\mathbf{A}_{i,k-1} \cap \mathbf{S}_k)$ , each  $\mathbf{v}_i \sim \mathbb{S}^{d-1}$  independently of  $\mathbf{v}_1, \dots, \mathbf{v}_{i-1}$ .

Note that for  $i > k$ , the edge  $(i, k)$  will be present in  $\mathbf{G}$  if and only if  $\mathbf{A}_{i,k} = \mathbf{A}_{i,k-1} \cap \mathbf{S}_k$ . Hence, the sampled graphs will satisfy  $\mathbf{G}_- \subseteq \mathbf{G} \subseteq \mathbf{G}_+$  if and only if for every  $i > k$  it is the case that

$$(1 - \varepsilon)p \leq \frac{\rho(\mathbf{A}_{i,k-1} \cap \mathbf{S}_k)}{\rho(\mathbf{A}_{i,k-1})} \leq (1 + \varepsilon)p. \quad (9)$$

We now argue that this occurs with high probability. Each  $\mathbf{A}_{i,k-1}$  is an intersection of  $\mathbf{j} = \text{Binom}(p, k-1)$  caps and  $k-1-\mathbf{j}$  anti-caps. Using Bernstein's inequality, with probability  $1 - n^{-\log n}$ ,  $\mathbf{j} = kp + \Delta$  for  $|\Delta| \leq C \cdot \log n \max(\sqrt{kp}, \log n)$  for  $C$  a universal constant. Hence, applying [Corollary 6.1](#) with  $t = \frac{1}{2}$  and a union bound, we have that so long as  $d \gg \max(kH(p), \log^3 n)^2 \log^2 n = O((np + \log^2 n)^2 \log^4 n)$ , with probability at least  $1 - \binom{n}{2} \cdot n^{-\Omega(\log n)}$ ,  $\rho(\mathbf{A}_{i,k}) \geq \frac{1}{2} \exp^{-C' \cdot kH(p) - \log^2 n}$  for all  $1 \leq k < i \leq n$ , and hence  $\log(\|\rho_{\mathbf{A}_{i,j}}\|_\infty) \leq C' \cdot (n \cdot H(p) + \log^2 n)$ .

Using this, we may apply [Lemma 4.7](#) with  $\kappa = 4 \log \frac{d}{p}$  and  $t = 4 \log n \sqrt{\frac{1}{d}}$ , in conjunction with a union bound over all  $1 \leq j < i \leq n$  to conclude that the probability that (9) holds for all such  $i > j$  is at least  $1 - n^{-\Omega(\log n)}$  for any  $\varepsilon \geq C'' \sqrt{\frac{1}{d}(np + \log n) \log^4 n}$ , where  $C''$  is a universal constant. This completes the proof.  $\square$

## 7 Distribution of a neighborhood

In this section, we analyze the probability distribution of the neighborhood of a vertex in a random geometric graph conditioned on the remaining graph. Throughout this section we will assume that  $\log^{36} n \leq d \leq n^{100}$  and  $p = \frac{\alpha}{n}$  for constant  $\alpha \geq 1$ .<sup>4</sup>

Concretely, let  $\mathbf{G} \sim \text{Geo}_d(n, p)$  and let  $\mathbf{G}_{n-1}$  be the induced subgraph on  $[n-1]$ . We analyze the distribution of  $N_{\mathbf{G}}(n) | \mathbf{G}_{n-1}$ , and in particular prove that it closely tracks the neighborhood distribution in an Erdős-Rényi graph.

<sup>4</sup>The upper bound of  $n^{100}$  on  $d$  isn't actually necessary for our proof techniques, but this assumption slightly simplifies calculations. Since indistinguishability results are already known when  $d \geq n^{100}$  from prior work, we prioritize having somewhat less symbol-dense calculations.

**Lemma 7.1.** *With probability  $1 - n^{-\Omega(\log n)}$  over the randomness of  $\mathbf{G}_{n-1}$ , for every  $S \subseteq [n-1]$  such that  $|S| \leq \log^2 n$ , the following is true. Let  $\ell < \frac{\log n}{\log \log n}$  be such that the balls  $\{B_{\mathbf{G}_{n-1}}(x, \ell) : x \in S\}$  are pairwise disjoint trees, and let  $d > \log^{20} n$ . Then:*

$$\Pr[N_{\mathbf{G}}(n) = S | \mathbf{G}_{n-1}] \in (1 \pm 2\eta(\ell)) p^{|S|} (1-p)^{n-1-|S|}$$

where

$$\eta(\ell) := \min \left\{ \frac{\log^8 n}{\sqrt{d}}, \max \left\{ 2 \left( \sqrt{\frac{\log^{28} n}{d}} \right)^\ell, 8 \cdot \sqrt{\frac{\log^{11} n}{nd}} \right\} \right\}.$$

We prove [Lemma 7.1](#) in [Section 7.5](#), after establishing the necessary ingredients.

## 7.1 Setup and outline

The distribution of  $N_{\mathbf{G}}(n)$  is given by sampling  $\mathbf{w}_n \sim \rho$ , then sampling  $(\mathbf{w}_1, \dots, \mathbf{w}_{n-1}) \sim \rho^{\mathbf{G}_{n-1}}$ , and choosing  $N_{\mathbf{G}}(n) = \{x : \langle \mathbf{w}_x, \mathbf{w}_n \rangle \geq \tau(p)\}$ . For every  $S \subseteq [n-1]$  satisfying  $|S| \leq \log^2 n$ , we are interested in  $\Pr[N_{\mathbf{G}}(n) = S | \mathbf{G}_{n-1}]$ , which is the expected measure of the intersection of some collection of caps and anticaps. In particular:

$$\Pr[N_{\mathbf{G}}(n) = S | \mathbf{G}_{n-1}] = \mathbf{E}_{(\mathbf{w}_i)_{i \in [n]} \sim \rho^{\mathbf{G}_{n-1}}} \rho \left( \bigcap_{j \in S} \text{cap}(\mathbf{w}_j) \cap \bigcap_{j \in [n-1] \setminus S} \overline{\text{cap}}(\mathbf{w}_j) \right).$$

The random variable  $\rho \left( \bigcap_{j \in S} \text{cap}(\mathbf{w}_j) \cap \bigcap_{j \in [n-1] \setminus S} \overline{\text{cap}}(\mathbf{w}_j) \right)$  may appear daunting at first due to the complicated correlation structure of  $(\mathbf{w}_i)_{i \in [n-1]}$ . Our situation is greatly simplified by tight concentration for the measure of the intersection of random anticaps with sets of lower bounded measure. In particular, we show in [Section 7.5](#) that with high probability:

$$\rho \left( \bigcap_{j \in S} \text{cap}(\mathbf{w}_j) \cap \bigcap_{j \in [n-1] \setminus S} \overline{\text{cap}}(\mathbf{w}_j) \right) \in \left( 1 \pm \tilde{O} \left( \sqrt{\frac{1}{nd}} \right) \right) \cdot \rho \left( \bigcap_{j \in S} \text{cap}(\mathbf{w}_j) \right) (1-p)^{n-1-|S|}.$$

This lets us write  $\Pr[N_{\mathbf{G}}(n) = S | \mathbf{G}_{n-1}]$  as

$$\Pr[N_{\mathbf{G}}(n) = S | \mathbf{G}_{n-1}] \in \left( 1 \pm \tilde{O} \left( \sqrt{\frac{1}{nd}} \right) \right) (1-p)^{n-1-|S|} \cdot \mathbf{E} \rho \left( \bigcap_{j \in S} \text{cap}(\mathbf{w}_j) \right).$$

Now,

$$\mathbf{E} \rho \left( \bigcap_{j \in S} \text{cap}(\mathbf{w}_j) \right) = \Pr[N_{\mathbf{G}}(n) \supseteq S | \mathbf{G}_{n-1}]$$

and so to study the probability that  $N_{\mathbf{G}}(n) = S$ , it suffices to study the probability that  $N_{\mathbf{G}}(n) \supseteq S$ , which can equivalently be written as:

$$\Pr[N_{\mathbf{G}}(n) \supseteq S | \mathbf{G}_{n-1}] = \Pr[\forall x \in S : \langle \mathbf{w}_x, \mathbf{w}_n \rangle \geq \tau(p)].$$

We next notice that we are spared from working with the potentially complicated full ensemble  $(\mathbf{w}_i)_{i \in [n-1]}$ , since the above depends only on its marginal distribution on  $\mathbf{w}_i$  for  $i \in S$ . Indeed, we prove that for most sets  $S$ , the vectors  $(\mathbf{w}_i)_{i \in S}$  roughly behave like independent and uniform vectors on the unit sphere, a statement which is made concrete below.

**Lemma 7.2.** *With probability  $1 - n^{-\Omega(\log n)}$ , for every  $S \subseteq [n - 1]$  such that  $|S| \leq \log^2 n$ , the following is true. Suppose  $\ell < \frac{\log n}{\log \log n}$  is such that the balls  $\{B_{\mathbf{G}_{n-1}}(x, \ell) : x \in S\}$  are all trees and are pairwise disjoint, then:*

$$\Pr[N_{\mathbf{G}}(n) \supseteq S | \mathbf{G}_{n-1}] \in (1 \pm \eta(\ell)) p^{|S|}$$

where

$$\eta(\ell) := \min \left\{ \frac{\log^8 n}{\sqrt{d}}, \max \left\{ 2 \sqrt{\frac{\log^{28} n}{d}}, 8 \sqrt{\frac{\log^{11} n}{d}} \right\} \right\}.$$

We now give a brief proof overview of [Lemma 7.2](#). The way we think about the quantity we wish to obtain a handle on is articulated by the following equality.

$$\Pr[N_{\mathbf{G}}(n) \supseteq S | \mathbf{G}_{n-1}] = \mathbf{E}_{\substack{\mathbf{w}_n \sim \rho \\ (\mathbf{w}_i)_{i \in [n-1]} \sim \rho^{\mathbf{G}_{n-1}}}} \prod_{x \in S} \mathbf{1}[\langle \mathbf{w}_x, \mathbf{w}_n \rangle \geq \tau(p)].$$

If  $(\mathbf{w}_x)_{x \in S}$  was a collection of **independent** random vectors distributed **uniformly** on the sphere then the above quantity is exactly equal to  $p^{|S|}$ . Our proof of [Lemma 7.2](#) is by establishing that both of these properties are approximately true. The following gives a rough outline of the proof along with how it's organized in this section.

- Our intuition is that any “large” correlations between vertices  $x$  and  $y$  are explained by the existence of paths in between them. Thus, we study the distribution of  $(\mathbf{w}_x)_{x \in S} | (\mathbf{w}_x)_{x \in B(S, \ell)}$  for some  $\ell$  where  $(\mathbf{w}_x)_{x \in B(S, \ell)}$  is sampled according to  $\rho^{\mathbf{G}_{n-1}}$  and attempt to prove approximate independence for this ensemble.
- In [Section 7.2](#), we define a constraint satisfaction problem instance on a forest factor graph with variable set indexed by the vertices of  $B(S, \ell)$ , and in [Section 7.4](#) show that the marginals of the uniform distribution  $\mathcal{F}'$  on satisfying assignments to this CSP instance on the variables indexed by  $S$  is very close to the distribution of  $(\mathbf{w}_x)_{x \in S}$ .
- $(\mathbf{w}_x)_{x \in S}$  is an independent ensemble when sampled from  $\mathcal{F}'$  for the simple reason that they are all in different connected components of the factor graph. The bulk of the technical work in this section is in proving that each  $\mathbf{w}_x$  “behaves like” a uniform vector, which is performed in [Section 7.3](#).
- The key relevant property of the uniform distribution on the sphere  $\rho$  for us is that for every  $w$ ,  $\Pr_{v \sim \rho}[\langle v, w \rangle \geq \tau(p)] = p$ . Thus, for each vertex  $x \in S$ , we show the following about the marginal of  $\mathcal{F}'$  on  $\mathbf{w}_x$ .

For most  $w$ ,  $\Pr[\langle \mathbf{w}_x, w \rangle \geq \tau(p)] = p(1 \pm \varepsilon)$ .

- The way we establish this “pseudo-uniformity” property of  $\mathbf{w}_x$  is by using the fact that its marginal can be computed exactly from belief propagation fixed point messages by virtue of arising from a forest factor graph. Our analysis of the belief propagation fixed point messages is based on diffusion properties of the random walk on the sphere where each transition moves from a point  $v$  to a random point in the  $p$ -cap around  $v$ .

## 7.2 A relaxation of containment probabilities

We now turn our attention towards proving [Lemma 7.2](#).

We use  $\mathbf{W}$  to refer to the collection  $(\mathbf{w}_i)_{i \in [n-1]} \sim \rho^{\mathbf{G}_{n-1}}$ , and for any subset  $A \subseteq [n]$ , we use the notation  $\mathbf{W}_A$  to refer to the subcollection  $(\mathbf{w}_i)_{i \in A}$ . Let  $\ell$  be an integer chosen so that  $\{B_{\mathbf{G}_{n-1}}(x, \ell) : x \in S\}$  are pairwise disjoint trees, and  $\ell < \frac{\log n}{\log \log n}$ . Define  $K := \bigcup_{x \in S} V(B_{\mathbf{G}_{n-1}}(x, \ell - 1))$  and denote the remaining vertices as  $R := [n - 1] \setminus K$ . Note that it is possible for  $\ell = 0$ , in which case  $K$  is empty. We can write

$$\Pr[N_{\mathbf{G}}(n) \supseteq S | \mathbf{G}_{n-1}] = \mathbf{E}_{\mathbf{w}_n \sim \rho} \left[ \mathbf{E}_{\mathbf{W}_R} \left[ \mathbf{Pr}_{\mathbf{W}_K | \mathbf{W}_R} [\forall x \in S : \langle \mathbf{w}_x, \mathbf{w}_n \rangle \geq \tau(p)] \right] \right]. \quad (10)$$

The distribution of  $\mathbf{W}_K | \mathbf{W}_R$  is uniform on the space of vectors producing  $\mathbf{G}_{n-1}$ , conditioned on the choice of  $\mathbf{W}_R$ :

$$\mathcal{F}(\mathbf{W}_R) := \{(w_i)_{i \in K} : \text{gg}(w_1, \dots, w_{n-1}) = \mathbf{G}_{n-1}, \text{ and } w_r = \mathbf{w}_r \text{ for } r \in R\}.$$

We “relax”  $\mathcal{F}(\mathbf{W}_R)$  to a slightly larger set  $\mathcal{F}'(\mathbf{W}_R) \supseteq \mathcal{F}(\mathbf{W}_R)$ , which admits a description in terms of the solution space of a constraint satisfaction problem on a forest; in particular, the induced subgraph  $T(\mathbf{G}_{n-1}) := \mathbf{G}_{n-1}[K]$ . The reason for doing so is we can precisely calculate marginals of the uniform distribution on  $\mathcal{F}'(\mathbf{W}_R)$  using the belief propagation algorithm. To help define  $\mathcal{F}'(\mathbf{W})$ , we let  $L(\mathbf{W}_R) := \bigcap_{j \in R} \overline{\text{cap}}(\mathbf{w}_j)$ , and similarly let  $L_i(\mathbf{W}_R) := \bigcap_{j \in R \setminus N_R(i)} \overline{\text{cap}}(\mathbf{w}_j) \cap \bigcap_{j \in N_R(i)} \text{cap}(\mathbf{w}_j)$  where  $N_R(i)$  denotes the set of neighbors of  $i$  within  $R$ . Observe that for  $i \in V$  with distance less than  $\ell - 1$  to some  $j \in S$ , the set  $L_i = L(\mathbf{W}_R)$ .

**Definition 7.3.**  $\mathcal{F}'(\mathbf{W}_R)$  is the collection of satisfying assignments to the following CSP instance  $\mathcal{I}(\mathbf{W}_R)$  on variables indexed by  $V(K)$ , comprised of unary constraints and binary constraints.

- **Unary constraints.** For each  $i \in V(K)$ ,  $v_i \in L_i(\mathbf{W}_R)$ . In words, the set of neighbors of the vector assignment  $v_i$  within  $R$  must be equal to  $N_{\mathbf{G}_{n-1}}(i) \cap R$ .
- **Binary constraints.** For every  $\{i, j\} \in T(\mathbf{G}_{n-1})$ :  $\langle v_i, v_j \rangle \geq \tau(p)$ . In words, for every edge  $\{i, j\}$ , the vector assignments  $v_i$  and  $v_j$  must also have an edge in between them.

We use  $F(\mathbf{W}_R)$  to denote the factor graph representation of  $\mathcal{I}(K)$ .

Notice that [Definition 7.3](#) enforces that  $\text{gg}(\mathbf{W}_K \cup \mathbf{W}_R)$  contains  $\mathbf{G}_{n-1}$  as an edge-induced subgraph. In [Section 7.4](#), we will show that with high probability,  $\text{gg}(\mathbf{W}_k \cup \mathbf{W}_R)$  is also precisely equal to  $\mathbf{G}_{n-1}$ .

Our main result about  $\mathcal{F}'$  is the following.

**Lemma 7.4.** *Suppose  $\ell \geq 1$ . Then for every  $S$  such that  $|S| \leq \log^2 n$ :*

$$\mathbf{Pr}_{\substack{\mathbf{w}_n \sim \mathbb{S}^{d-1} \\ \mathbf{W}_K \sim \mathcal{F}'(\mathbf{W}_R)}} [\forall x \in S : \langle \mathbf{w}_x, \mathbf{w}_n \rangle \geq \tau(p)] \in (1 \pm \eta(\ell)) p^{|S|}$$

except with probability at most  $n^{-\Omega(\log n)}$  over the randomness of  $\mathbf{G}_{n-1}$  where

$$\eta(\ell) := \max \left\{ \left( C \sqrt{\frac{\log^{27} n}{d}} \right)^\ell, 4 \sqrt{\frac{\log^{11} n}{nd}} \right\}$$

for some absolute constant  $C$ .<sup>5</sup>

The proof of [Lemma 7.4](#) is carried out in [Section 7.3](#).

<sup>5</sup>Recall that [Lemma 7.4](#) is a stepping stone towards proving [Lemma 7.2](#), and the case of  $\ell = 0$  is handled directly within the proof of [Lemma 7.2](#).

### 7.3 Belief propagation for relaxed distribution

**BP update rule.** Recall that given a CSP instance  $\mathcal{I}$  with variable set  $V$  and constraint set  $E$ , the marginals on specific variables can be accurately computed from the belief propagation fixed point messages when the factor graph is a forest using [Theorem 3.19](#).

In our setting, the BP equations are the following:

$$\text{Variable-to-constraint messages. } m^{v \rightarrow f} = \frac{\prod_{e \in \partial v \setminus f} m^{e \rightarrow v}}{\int \prod_{e \in \partial v \setminus f} m^{e \rightarrow v}(x) d\rho(x)} \quad (11)$$

$$\text{Constraint-to-variable messages (unary case). } m^{f \rightarrow v} = \frac{f}{\int f(x) d\rho(x)} \quad (12)$$

**Constraint-to-variable messages (binary case).** Let  $\partial f = \{v, w\}$ , then:

$$m^{f \rightarrow v}(x_v) = \frac{\int f(x_v, x_w) \cdot m^{w \rightarrow f}(x_w) d\rho(x_w)}{\int \int f(x_v, x_w) \cdot m^{w \rightarrow f}(x_w) d\rho(x_w) d\rho(x_v)}. \quad (13)$$

#### 7.3.1 Interpreting binary constraint-to-vertex messages as convolutions

The denominator in the binary case of constraint-to-vertex messages further simplifies as follows:

$$\begin{aligned} \int \int f(x_v, x_w) \cdot m_t^{w \rightarrow f}(x_w) d\rho(x_v) d\rho(x_w) &= \int m_t^{w \rightarrow f}(x_w) \int \mathbf{1}[\langle x_v, x_w \rangle \geq \tau(p)] d\rho(x_v) d\rho(x_w) \\ &= p \int m_t^{w \rightarrow f}(x_w) d\rho(x_w) \\ &= p. \end{aligned}$$

The final equality comes from the fact that each message  $m_t^{w \rightarrow f}$  is a distribution. Thus, for binary constraints  $f$  such that  $\partial f = \{v, w\}$  we can rewrite the constraint-to-vertex messages as:

$$m_{t+1}^{f \rightarrow v}(x_v) = \int \frac{f(x_v, x_w)}{p} m_t^{w \rightarrow f}(x_w) d\rho(x_w),$$

which, in particular, can be written as  $Pm^{w \rightarrow f}$  for a linear operator  $P$ , defined as follows:

**Definition 7.5.** Let  $P$  be the linear operator defined so that for any function  $h : \mathbb{S}^{d-1} \rightarrow \mathbb{R}$ ,

$$Ph(x) = \frac{1}{p} \int_{\text{cap}_p(x)} h(y) d\rho(y),$$

which we alternately denote  $\frac{1}{p}h(\text{cap}_p(x))$ .

In words, the operator  $P$  convolves its input with the uniform distribution over a spherical cap. Since  $P$  is a convolution operator, it preserves the  $\ell_1$ -norm of nonnegative functions.

**Observation 7.6.** Suppose  $v$  is a nonnegative function, then  $\|Pv\|_1 = \|v\|_1$ . Additionally, for an arbitrary function  $v$ ,  $\|Pv\|_1 \leq \|v\|_1$ .

When  $v$  is a distribution in particular,  $P$  can be construed as the transition operator of the Markov chain on  $\mathbb{S}^{d-1}$  where a single step entails walking from  $v$  to a uniformly random point in  $\text{cap}(v)$ .

The following useful observations are immediate from how we define the operator  $P$ .

**Observation 7.7.** For any function  $v$ ,  $\|Pv\|_\infty \leq \frac{\|v\|_1}{p}$ . This is because  $v(\text{cap}(z)) \leq \|v\|_1$  for any  $z \in \mathbb{S}^{d-1}$ .

### 7.3.2 A potential function for analyzing BP.

Here, we introduce the notion of the *spread profile*, which we will use as potential function for tracking how close the density defined by a message is to uniform over  $L(\mathbf{W}_R)$ . We'll show that with each successive application of the operator  $P$  which occurs when messages pass from the leaves to the root of the tree, the spread of the function improves, until ultimately the spread of the root's function guarantees that it is close to uniform.

**Definition 7.8.** Given a function  $v$  on  $\mathbb{S}^{d-1}$ , its *deviation profile* is the function  $\text{Dev}_v : \mathbb{R}_+ \rightarrow [0, 1]$ :

$$\text{Dev}_v(\varepsilon) := \Pr_{\mathbf{z} \sim \rho} \left[ v(\mathbf{z}) \notin \mathbf{E}_{\mathbf{y} \sim \rho} [v(\mathbf{y})] \pm \varepsilon \cdot \|v\|_1 \right].$$

The *spread profile* of  $v$ , denoted  $\text{Spr}_v(\delta)$ , is

$$\text{Spr}_v(\delta) := \inf\{\varepsilon \in \mathbb{R}_+ : \text{Dev}_v(\varepsilon) \leq \delta\}.$$

In the special case where  $v$  is a relative density of a distribution with respect to  $\rho$ ,

$$\text{Dev}_v(\varepsilon) = \Pr_{\mathbf{z} \sim \rho} [ |v(\mathbf{z}) - 1| > \varepsilon ].$$

We comment that we can think of the spread profile as an “inverse” to the deviation profile, since it takes in a tail probability and returns the corresponding deviation  $\varepsilon$ .

It now follows from averaging arguments that the deviation profile and spread profile of a distribution  $v$  give us useful upper and lower bounds on  $d_{\text{TV}}(v, \rho)$ .

**Observation 7.9.** For any  $\varepsilon > 0$  and distribution  $v$ ,

$$\text{Dev}_v(\varepsilon) \cdot \varepsilon \leq d_{\text{TV}}(v, \rho) \leq \varepsilon + \text{Dev}_v(\varepsilon) \cdot \|v\|_\infty$$

Similarly, for any  $\delta > 0$  and distribution  $v$ ,

$$\delta \cdot \text{Spr}_v(\delta) \leq d_{\text{TV}}(v, \rho) \leq \text{Spr}_v(\delta) + \delta \cdot \|v\|_\infty$$

We list some properties of the deviation and spread profiles that will be useful in our analysis.

**Observation 7.10.** Both the deviation profile and spread profile are non-increasing functions.

**Observation 7.11.** For  $v = \rho$ , we have  $\text{Dev}_v(\varepsilon) = 0$  for all  $\varepsilon > 0$  and  $\text{Spr}_v(\delta) = 0$  for all  $\delta > 0$ .

**Observation 7.12.** Both the deviation and spread profiles are invariant under a constant factor multiplication to  $v$ , i.e. for  $\alpha \neq 0$ :

$$\text{Dev}_{\alpha v}(\varepsilon) = \text{Dev}_v(\varepsilon) \text{ and } \text{Spr}_{\alpha v}(\delta) = \text{Spr}_v(\delta)$$

**Observation 7.13.** For  $v = v_1 + v_2$ , and  $\varepsilon > 0$ , the deviation profile satisfies:

$$\text{Dev}_v \left( \varepsilon \cdot \frac{\|v_1\|_1 + \|v_2\|_1}{\|v\|_1} \right) \leq \text{Dev}_{v_1}(\varepsilon) + \text{Dev}_{v_2}(\varepsilon)$$

The “triangle inequality” follows from the following containment of events:

$$\{v(\mathbf{z}) \notin \mathbf{E}_{\mathbf{y} \sim \rho} [v(\mathbf{y})] \pm \varepsilon \cdot (\|v_1\|_1 + \|v_2\|_1)\} \subseteq \{v_1(\mathbf{z}) \notin \mathbf{E}_{\mathbf{y} \sim \rho} [v_1(\mathbf{y})] \pm \varepsilon \cdot \|v_1\|_1\} \cup \{v_2(\mathbf{z}) \notin \mathbf{E}_{\mathbf{y} \sim \rho} [v_2(\mathbf{y})] \pm \varepsilon \cdot \|v_2\|_1\}.$$

Similarly, we also have a “triangle inequality” for the spread profile. For  $v = v_1 + v_2$  and  $\delta_1, \delta_2 > 0$ :

$$\|v\|_1 \cdot \text{Spr}_v(\delta_1 + \delta_2) \leq \|v_1\|_1 \cdot \text{Spr}_{v_1}(\delta_1) + \|v_2\|_1 \cdot \text{Spr}_{v_2}(\delta_2)$$

We can think of the left hand side as the  $(1 - \delta_1 - \delta_2)$ -confidence interval of  $v$ , and the right hand terms as the  $(1 - \delta_1)$ - and  $(1 - \delta_2)$ -confidence intervals of  $v_1$  and  $v_2$ , respectively.

### 7.3.3 Diffusion of distributions under convolutions with caps.

We will show that the cap-convolution operator  $P$  “flattens” functions over  $\mathbb{S}^{d-1}$ . Concretely:

**Lemma 7.14.** *Let  $v$  be any function on  $\mathbb{S}^{d-1}$  with  $\mathbf{E}_{\mathbf{y} \sim \rho} v(\mathbf{y}) = 0$ , and suppose  $d \geq \log^{10} n \cdot \left(1 + \log^3 \frac{\|v\|_\infty}{\|v\|_1}\right)$ . Then:*

$$\text{Dev}_{Pv} \left( \varepsilon \cdot \frac{\|v\|_1}{\|Pv\|_1} \right) \leq n^{-\log^4 n},$$

where  $\varepsilon := \sqrt{\frac{1}{d}} \cdot C \log^{5.5} n \cdot \left(2 \log \frac{\|v\|_\infty}{\|v\|_1} + 8 \log \frac{d}{p}\right)$  for  $C$  an absolute constant. Thus,  $\text{Spr}_{Pv}(n^{-\log^4 n}) \leq \varepsilon \cdot \frac{\|v\|_1}{\|Pv\|_1}$ .

A special case of Lemma 7.14 is where  $v = \mu - 1$  for  $\mu$  a relative density. In this scenario we have

$$n^{-\log^4 n} \geq \text{Dev}_{Pv} \left( \varepsilon \frac{\|v\|_1}{\|Pv\|_1} \right) = \mathbf{Pr}_{\mathbf{z} \sim \rho} \left[ |Pv(\mathbf{z})| > \varepsilon \frac{\|v\|_1}{\|Pv\|_1} \cdot \|Pv\|_1 \right] = \mathbf{Pr}_{\mathbf{z} \sim \rho} [ |P\mu(\mathbf{z}) - 1| > \varepsilon \cdot d_{\text{TV}}(\mu, \rho) ],$$

and hence by an averaging argument  $d_{\text{TV}}(P\mu, \rho) \leq \varepsilon \cdot d_{\text{TV}}(\mu, \rho) + n^{-\log^4 n}$ . In such a way, Lemma 7.14 characterizes the diffusion of distributions under  $P$ .

This is our main ingredient in the proof of Lemma 7.1; the geometric decay of  $\left(\sqrt{\frac{\log^{28} n}{d}}\right)^\ell$  in the definition of  $\eta(\ell)$  is due to a factor of  $\varepsilon^\ell$  arising from the  $\ell$  applications of  $P$  occur as the messages travel from from the leaves of  $K$  to the roots (i.e., the vertices of  $S$ ).

The key property we use to establish Lemma 7.14 is that the value of  $v(\text{cap}(\mathbf{z}))$  concentrates when  $\mathbf{z}$  is chosen uniformly at random from  $\mathbb{S}^{d-1}$  and  $\|v\|_\infty$  is reasonably bounded.

*Proof of Lemma 7.14.* First, let us write  $v = v^+ - v^-$  where  $v^+ = \max(v, 0)$  and  $v^- = -\min(v, 0)$ . Here, both  $v^+$  and  $v^-$  are nonnegative functions, and  $\|v\| = 2\|v^+\| = 2\|v^-\| = \|v^+\| + \|v^-\|$  for  $\|\cdot\|$  any  $\ell_p$  norm. Since  $Pv = Pv^+ - Pv^-$ , by Observation 7.13 and Observation 7.12:

$$\text{Dev}_{Pv} \left( \varepsilon \cdot \frac{\|Pv^+\|_1 + \|Pv^-\|_1}{\|Pv\|_1} \right) \leq \text{Dev}_{Pv^+}(\varepsilon) + \text{Dev}_{Pv^-}(\varepsilon) = \text{Dev}_{P \frac{v^+}{\|v^+\|}}(\varepsilon) + \text{Dev}_{P \frac{v^-}{\|v^-\|}}(\varepsilon).$$

Note that  $\frac{1}{\|v^+\|_1} v^+$  and  $\frac{1}{\|v^-\|_1} v^-$  are both probability measures. By Definition 7.5, for any  $x \in \mathbb{S}^{d-1}$ :

$$P \frac{v^+(x)}{\|v^+\|_1} = \frac{1}{p} \cdot \frac{v^+(\text{cap}(x))}{\|v^+\|_1} \quad \text{and} \quad P \frac{v^-(x)}{\|v^-\|_1} = \frac{1}{p} \cdot \frac{v^-(\text{cap}(x))}{\|v^-\|_1}$$

By Corollary 4.9 with  $s$  set as  $\frac{1}{d} \cdot C \log^5 n \cdot \left(2 \log \frac{\|v^+\|_\infty}{\|v^+\|_1} + 8 \log \frac{d}{p}\right)$  for some constant  $C > 0$ :

$$\mathbf{Pr}_{\mathbf{x} \sim \rho} \left[ \left| \frac{1}{p} \cdot \frac{v^+(\text{cap}(\mathbf{x}))}{\|v^+\|_1} - 1 \right| > C \cdot C_{4.9} \sqrt{\frac{1}{d} \cdot \log\left(\frac{1}{p}\right)} \cdot \log^5(n) \cdot \left(2 \log \frac{\|v^+\|_\infty}{\|v^+\|_1} + 8 \log \frac{d}{p}\right) \right] \leq \frac{1}{2} \cdot n^{-\log^4 n}.$$

Using  $p \geq \frac{1}{n}$ ,  $\|v^+\|_\infty \leq \|v\|_\infty$  and  $\|v^+\|_1 = \frac{1}{2}\|v\|$  we can conclude:

$$\text{Dev}_{P \frac{v^+}{\|v^+\|}}(\varepsilon) \leq \frac{1}{2} n^{-\log^4 n}.$$

Identically,  $\text{Dev}_{P \frac{v^-}{\|v^-\|}}(\varepsilon) \leq \frac{1}{2} n^{-\log^4 n}$ , and thus:

$$\text{Dev}_{Pv} \left( \varepsilon \cdot \frac{\|Pv^+\|_1 + \|Pv^-\|_1}{\|Pv\|_1} \right) \leq n^{-\log^4 n}.$$

Since  $v^+$  and  $v^-$  are nonnegative functions,  $\|Pv^+\| = \|v^+\|$  and  $\|Pv^-\| = \|v^-\|$ , and the desired statement then immediately follows.  $\square$

**Diffusion of intersection of anti-caps under  $P$ .** Since the BP constraints require that the vectors be contained in  $L(\mathbf{W}_R)$ , we would also like to understand the diffusion of  $L(\mathbf{W}_R)$  under  $P$ . Since  $L(\mathbf{W}_R)$  is an intersection of anticaps, we can greatly strengthen [Lemma 7.14](#) for such sets.

**Lemma 7.15.** *Define*

$$\begin{aligned} g(\mathbf{G}_{n-1}) &:= \Pr_{\mathbf{w}_1, \dots, \mathbf{w}_{n-1} \sim \rho^{\mathbf{G}_{n-1}}} \left[ \text{Dev}_{PL(\mathbf{W}_R)}(\varepsilon) > n^{-\log^4 n} \right] \\ &= \Pr_{\mathbf{w}_1, \dots, \mathbf{w}_{n-1} \sim \rho^{\mathbf{G}_{n-1}}} \left[ \text{Spr}_{PL(\mathbf{W}_R)}(n^{-\log^4 n}) > \varepsilon \right] \end{aligned}$$

for  $\varepsilon = \sqrt{\frac{\log^{11} n}{nd}}$ . Then,  $g(\mathbf{G}_{n-1})$  is at most  $O(n^{-\log^2 n})$  except with probability  $O(n^{-\log^2 n})$ .

*Proof.* While we ultimately want control on  $\text{Dev}_{PL(\mathbf{W}_R)}$ , our starting point is to control only  $\text{Dev}_{PL(\mathbf{W})}$ . We take advantage of concentration over  $\mathbf{W}$  to deduce concentration over  $\mathbf{W}_T$ , for any  $T \subseteq [n-1]$  of sufficiently large size. We will show below that the following hold:

(I) Concentration of anticap intersection: with probability  $1 - O(n^{-\log^2 n})$  over randomness of  $\mathbf{G}_{n-1}$ ,

$$\Pr_{\mathbf{w}_1, \dots, \mathbf{w}_{n-1} \sim \rho^{\mathbf{G}_{n-1}}} \left[ \text{Dev}_{PL(\mathbf{W})}(\varepsilon/2) > n^{-\log^5 n} \right] \leq O(n^{-\log^2 n}).$$

(II) Bounded pairwise cap intersections: for any  $x \in \mathbb{S}^{d-1}$ :

$$\Pr_{\mathbf{w}_1, \dots, \mathbf{w}_{n-1} \sim \rho^{\mathbf{G}_{n-1}}} \left[ \exists i \in [n-1] \text{ s.t. } \rho(\text{cap}(x) \cap \text{cap}(\mathbf{w}_i)) \geq 2p^2 \right] \leq O(n^{-\log^3 n/2})$$

except with probability  $O(n^{-\log^2 n})$  over the randomness of  $\mathbf{G}_{n-1}$ .

Assume first that [Property \(I\)](#) and [Property \(II\)](#) hold. Define

$$h(\mathbf{w}_1, \dots, \mathbf{w}_{n-1}) := \Pr_{\mathbf{x} \sim \rho} \left[ \exists i \in [n-1] \text{ s.t. } \rho(\text{cap}(\mathbf{x}) \cap \text{cap}(\mathbf{w}_i)) \geq 2p^2 \right]$$

[Property \(II\)](#) implies:

$$\Pr_{\mathbf{w}_1, \dots, \mathbf{w}_{n-1} \sim \rho^{\mathbf{G}_{n-1}}} \left[ h(\mathbf{w}_1, \dots, \mathbf{w}_{n-1}) \leq O(n^{-\log^3 n/4}) \right] \leq O(n^{-\log^3 n/4}) \quad (14)$$

except with probability  $O(n^{-\log^3 n/4})$ . By Markov's inequality, whenever [\(14\)](#) holds:

$$\Pr_{\mathbf{w}_1, \dots, \mathbf{w}_{n-1} \sim \rho^{\mathbf{G}_{n-1}}} \left[ h(\mathbf{w}_1, \dots, \mathbf{w}_{n-1}) \leq O(n^{-\log^3 n/8}) \right] \leq O(n^{-\log^3 n/8}).$$

Suppose  $\mathbf{w}_1, \dots, \mathbf{w}_{n-1}$  are such that  $\text{Dev}_{PL(\mathbf{W})}(\varepsilon/2) \leq n^{-\log^4 n}$  and  $h(\mathbf{w}_1, \dots, \mathbf{w}_{n-1}) \leq O(n^{-\log^3 n/8})$ . By our assumptions, these conditions hold with probability at least  $1 - O(n^{-\log^2 n})$  when  $\mathbf{w}_1, \dots, \mathbf{w}_{n-1} \sim \rho^{\mathbf{G}_{n-1}}$ , with probability at least  $1 - O(n^{-\log^2 n})$  over the randomness of  $\mathbf{G}_{n-1}$ . For  $\mathbf{x} \sim \mathbb{S}^{d-1}$ , with probability at least  $1 - O(n^{-\log^2 n})$ :

$$\begin{aligned} \rho \left( \bigcap_{i=1}^{n-1} \overline{\text{cap}}(\mathbf{w}_i) \cap \text{cap}(\mathbf{x}) \right) &\in (1 \pm \varepsilon/2)(1-p)^{n-1} p \\ \rho(\text{cap}(\mathbf{x}) \cap \text{cap}(\mathbf{w}_i)) &\leq 2p^2 \quad \forall i \in [n-1]. \end{aligned}$$

As a consequence, with probability at least  $1 - O\left(n^{-\log^2 n}\right)$  as well:

$$\begin{aligned} \rho\left(\bigcap_{i \in [n-1]} \overline{\text{cap}}(\mathbf{w}_i) \cap \text{cap}(\mathbf{x})\right) &= \rho\left(\left(\bigcap_{i \in R} (\overline{\text{cap}}(\mathbf{w}_i) \cap \text{cap}(\mathbf{x}))\right) \setminus \bigcup_{j \in [n-1] \setminus R} (\text{cap}(\mathbf{w}_j) \cap \text{cap}(\mathbf{x}))\right) \\ &\geq \rho\left(\bigcap_{i \in R} \overline{\text{cap}}(\mathbf{w}_i) \cap \text{cap}(\mathbf{x})\right) - 2(n-1-|R|)p^2. \end{aligned}$$

The inequality follows from assuming the sets  $\text{cap}(\mathbf{w}_j)$  for  $j \in [n-1] \setminus R$  are disjoint in the worst case. On the other hand, as  $R \subseteq [n-1]$ ,

$$\rho\left(\bigcap_{i \in [n-1]} \overline{\text{cap}}(\mathbf{w}_i) \cap \text{cap}(\mathbf{x})\right) \leq \rho\left(\bigcap_{i \in R} \overline{\text{cap}}(\mathbf{w}_i) \cap \text{cap}(\mathbf{x})\right).$$

The above can be rearranged as:

$$\rho\left(\bigcap_{i \in [n-1]} \overline{\text{cap}}(\mathbf{w}_i) \cap \text{cap}(\mathbf{x})\right) \leq \rho\left(\bigcap_{i \in R} \overline{\text{cap}}(\mathbf{w}_i) \cap \text{cap}(\mathbf{x})\right) \leq \rho\left(\bigcap_{i \in [n-1]} \overline{\text{cap}}(\mathbf{w}_i) \cap \text{cap}(\mathbf{x})\right) + 2(n-1-|R|)p^2,$$

and thus

$$\rho\left(\bigcap_{i \in R} \overline{\text{cap}}(\mathbf{w}_i) \cap \text{cap}(\mathbf{x})\right) \in p \cdot (1-p)^{|R|} \cdot \left((1-\varepsilon) \cdot (1-p)^{n-1-|R|}, (1+\varepsilon) \cdot (1+C(n-1-|R|)p)\right)$$

for constant  $C$ . By applying a union bound to [Lemma 3.12](#) along with [Proposition 1.3](#) and our bounds on  $\ell$  and  $|S|$ :

$$n-1-|R| = |K| \leq n^{\log(2\alpha)/\log \log n} \cdot \log^2 n$$

except with probability at most  $n^{-\Omega(\log n)}$ . Thus, the above simplifies to:

$$\rho\left(\bigcap_{i \in R} \overline{\text{cap}}(\mathbf{w}_i) \cap \text{cap}(\mathbf{x})\right) \in p \cdot (1-p)^{|R|} \cdot (1-\varepsilon, 1+\varepsilon).$$

This establishes that  $\text{Dev}_{PL}(\mathbf{W}_R)(\varepsilon) \leq O\left(n^{-\log^5 n}\right) \leq n^{-\log^4 n}$ .

It remains to prove [Property \(I\)](#) and [Property \(II\)](#). [Property \(I\)](#) follows from a combination of [Observation 3.14](#) and [Corollary 5.2](#) with parameter setting  $t = C\sqrt{\frac{\log^{11} n}{d}}$  for some constant  $C$  and  $m = n$ . To prove [Property \(II\)](#), first observe that applying a union bound to [Corollary 4.8](#) implies that when  $\mathbf{w}_1, \dots, \mathbf{w}_{n-1} \sim \rho^{\otimes n-1}$ ,  $\rho(\text{cap}(x) \cap \text{cap}(\mathbf{w}_i))$  is at most  $2p^2$  for all  $i \in [n-1]$  except with probability at most  $n^{-\log^3 n/2}$ . [Property \(II\)](#) then follows from [Observation 3.14](#), hence completing our proof.  $\square$

**Applying  $P$  to distributions close to  $L(\mathbf{W}_R)$ .** We now describe our main result about the diffusion properties of  $P$  when applied to distributions close to the uniform distribution on  $L(\mathbf{W}_R)$ .

**Lemma 7.16.** *Let  $\nu$  be a distribution over  $\mathbb{S}^{d-1}$  such that  $\|\nu\|_\infty \leq n^{\log^5 n}$ , and let  $\delta = d_{\text{TV}}(\nu, L(\mathbf{W}_R))$ . Then,*

$$\text{Spr}_{P\nu}\left(2n^{-\log^4 n}\right) \leq \max\left\{2\sqrt{\frac{\log^{11} n}{nd}}, C\delta \cdot \sqrt{\frac{\log^{23} n}{d}}\right\}$$

for a universal constant  $C$ .

*Proof.* We can express  $\nu = L(\mathbf{W}_R) + \Delta$ . Then, the triangle inequality for  $\text{Spr}_{P\nu}$  as articulated in [Observation 7.13](#) implies:

$$\text{Spr}_{P\nu}(2n^{-\log^4 n}) \leq \text{Spr}_{PL(\mathbf{W}_R)}(n^{-\log^4 n}) + \|P\Delta\|_1 \cdot \text{Spr}_{P\Delta}(n^{-\log^4 n}). \quad (15)$$

By [Lemma 7.15](#),  $\text{Spr}_{PL(\mathbf{W}_R)}(n^{-\log^4 n}) \leq \sqrt{\frac{\log^{11} n}{nd}}$ .

We now turn our attention to  $P\Delta$ . We would like to understand  $P\Delta$  through [Lemma 7.14](#), but the lower bound on  $d$  of  $\log^{10} n \cdot \left(1 + \log^3 \frac{\|v\|_\infty}{\|v\|_1}\right)$  in the hypothesis of its statement prevents us from applying it when  $\delta$  is too small, and hence we case on the value of  $\delta$ . Suppose  $\delta \leq n^{-2\log^4 n}$ , then by [Observation 7.6](#),  $\|P\Delta\|_1 \leq n^{-2\log^4 n}$  and thus by Markov's inequality  $\text{Spr}_{P\Delta}(n^{-\log^4 n}) \leq n^{\log^4 n}$ . Plugging this into (15) gives:

$$\text{Spr}_{P\nu}(2n^{-\log^4 n}) \leq \sqrt{\frac{\log^{11} n}{nd}} + n^{-\log^4 n} \leq 2\sqrt{\frac{\log^{11} n}{nd}}.$$

When  $\delta > 2n^{-\log^4 n}$ , by [Lemma 7.14](#), we have for some constant  $C$ :

$$\begin{aligned} \text{Spr}_{P\Delta}(n^{-\log^4 n}) &\leq \sqrt{\frac{1}{d}} \cdot C \log^{5.5} n \cdot \left(2 \log \frac{n^{\log^5 n}}{\|\Delta\|_1} + 8 \log \frac{d}{p}\right) \cdot \frac{\|\Delta\|_1}{\|P\Delta\|_1} \\ &\leq \sqrt{\frac{1}{d}} \cdot C \log^{5.5} n \cdot \left(4 \log^6 n + 8 \log \frac{d}{p}\right) \cdot \frac{\|\Delta\|_1}{\|P\Delta\|_1} \end{aligned}$$

where the second inequality uses  $\|\Delta\|_1 = \delta > n^{-2\log^4 n}$ . Consequently, plugging into (15),

$$\text{Spr}_{P\nu}(n^{-\log^4 n}) \leq \sqrt{\frac{\log^{11} n}{nd}} + C \sqrt{\frac{\log^{23} n}{d}} \cdot \delta$$

which completes the proof.  $\square$

### 7.3.4 Analyzing vertex-to-constraint messages.

Recall that our vertex-to-constraint messages are given as a product of the vertex's incoming messages. Each vertex's incoming messages are themselves measures over  $\mathbb{S}^{d-1}$  with bounded spread; we show that as long as a vertex's degree is bounded, the spread of their product is not too large.

**Lemma 7.17.** *Let  $\nu_1, \dots, \nu_j$  be distributions over  $\mathbb{S}^{d-1}$  with intersecting support. For each  $i \in [j]$ , suppose  $\text{Spr}_{\nu_i}(\varepsilon) \leq \eta$  and  $\|\nu_i\|_\infty \leq \frac{1}{p}$ . Further, assume  $\eta(j+1) < \frac{1}{2}$  and  $\frac{j\varepsilon}{p^j} \leq \eta$ . Let  $\nu$  be the distribution whose density is  $\nu(x) \propto \prod_{i=1}^j \nu_i(x)$ . Then,*

$$\text{Spr}_\nu(j \cdot \varepsilon) \leq 8j\eta \quad \text{and} \quad \|\nu\|_\infty \leq \frac{2}{p^j}.$$

As a consequence of [Observation 7.9](#):

$$d_{\text{TV}}(\nu, \rho) \leq 10j\eta.$$

*Proof.* Let  $B$  be the “bad” domain defined  $B := \{x \in \mathbb{S}^{d-1} : \exists i \text{ s.t. } \nu_i \notin [1-\eta, 1+\eta]\}$ . We know  $\rho(B) \leq j\varepsilon$  by our assumption on the spread profiles of  $\nu_1, \dots, \nu_j$ . Thus, for  $x \in \mathbb{S}^{d-1} \setminus B$ :

$$\frac{(1-\eta)^j}{Z} \leq \nu(x) \leq \frac{(1+\eta)^j}{Z}$$

where

$$Z = \int \prod_{i=1}^j v_i(x) d\rho(x).$$

To obtain upper and lower bounds on  $Z$ , we write

$$\begin{aligned} Z &= \int \prod_{i=1}^j v_i(x) \mathbf{1}[x \in \mathbb{S}^{d-1} \setminus B] d\rho(x) + \int \prod_{i=1}^j v_i(x) \mathbf{1}[x \in B] d\rho(x) \\ &\begin{cases} \geq (1-\eta)^j (1-j\epsilon) \geq (1-\eta)^{j+1} \geq 1-(j+1)\eta \\ \leq (1+\eta)^j + \frac{j\epsilon}{p^j} \leq 1 + \frac{j\eta}{1-j\eta} + \eta \leq 1 + \frac{(j+1)\eta}{1-j\eta}. \end{cases} \end{aligned}$$

We can obtain lower and upper bounds on  $v(x)$  for  $x \in \mathbb{S}^{d-1} \setminus B$  as follows.

$$\begin{aligned} v(x) &\geq \frac{(1-\eta)^j}{1 + \frac{(j+1)\eta}{1-j\eta}} \geq \frac{1-j\eta}{1 + \frac{(j+1)\eta}{1-j\eta}} = 1 - \frac{j\eta + \frac{(j+1)\eta}{1-j\eta}}{1 + \frac{(j+1)\eta}{1-j\eta}} \geq 1 - 5j\eta \\ v(x) &\leq \frac{(1+\eta)^j}{1-(j+1)\eta} \leq \frac{1 + \frac{j\eta}{1-j\eta}}{1-(j+1)\eta} = 1 + \frac{(j+1)\eta + \frac{j\eta}{1-j\eta}}{1-(j+1)\eta} \leq 1 + 8j\eta. \end{aligned}$$

Since  $Z \geq 1-(j+1)\eta \geq \frac{1}{2}$ , and each  $\|v_i\|_\infty \leq \frac{1}{p}$ ,  $\|v\|_\infty \leq \frac{2}{p^j}$ . □

### 7.3.5 Tracking the spread in BP fixed point messages

We now study the spread properties of the BP fixed point messages.

In this section let  $T$  be a tree with a distinguished root vertex  $r$ . Let  $F$  be a factor graph associated to the tree where the variable vertices are the vertices of  $T$ , with a binary constraint  $\mathbf{1}[\langle v_i, v_j \rangle \geq \tau(p)]$  for every edge  $\{i, j\} \in T$ , and a unary constraint  $\mathbf{1}[v_i \in L_i]$  for every vertex  $i \in V(T)$ .

**Assumption 7.18.** We assume that the degree of every vertex in  $T$  is bounded by  $\log^2 n$ , and that for every  $i \in V(T)$ ,  $\rho(L_i) \geq n^{-\log^3 n}$ .

For a variable vertex  $v \in F$  for  $v \neq r$ , we use  $\text{Par}(v)$  to denote the constraint vertex corresponding to the edge  $\{v, w\}$  where  $w$  is the parent of  $v$  in  $T$ . We use  $\text{Ch}(v)$  to denote the set of constraint vertices corresponding to edges  $\{v, w\}$  for  $w$  that are children of  $v$  in  $T$ . Finally, we use  $\text{Un}(v)$  to denote the unary constraint vertex incident to  $v$ .

**Remark 7.19.** Observe that trees that arise from rooting connected components of  $T(\mathbf{G}_{n-1})$  at vertices in  $S$  always have all leaves equidistant to the root and also satisfy the inequality

$$\text{Spr}_{PL_i}(n^{-\log^4 n}) \leq \sqrt{\frac{\log^{11} n}{nd}}$$

for every internal vertex  $i$ , which are properties that trees meeting our above assumptions need not satisfy. This may suggest adding these properties as assumptions too. However, we stick to the more general setting as it is crucial in our proof relating the relaxed and true distributions.

We place our vertices in to “tiers”, which will let us quantify the spread of messages sent out by them.

**Definition 7.20.** We say  $i \in V(T)$  is a *free vertex* if  $L_i = L(\mathbf{W}_R)$  and  $\text{Spr}_{PL_i}(n^{-\log^4 n}) \leq \sqrt{\frac{\log^{11} n}{nd}}$ . We say  $i \in T$  is a *anchored vertex* if it is not a free vertex. We use  $\text{Anch}(T)$  to denote the set of anchored vertices in  $T$ .

**Definition 7.21.** Given a vertex  $i \in V(T)$  we say its *tier* is the distance to its closest descendant that is an anchored vertex. In particular,

$$\text{Tier}(v) = \min\{\beta : v \text{ has a distance-}\beta \text{ descendant in } \text{Anch}(T)\}.$$

We use the convention that  $v$  is a descendant of itself and hence  $\text{Tier}(v) = 0$  when  $v$  is an anchored vertex.

For a variable vertex  $i$ , the closeness of the distribution obtained by combining messages from its children to  $\rho$  is governed by its tier.

**Lemma 7.22.** For a variable vertex  $i$  in  $F$ , let  $v_i$  be the distribution

$$v_i \propto \prod_{a \in \text{Ch}(i)} m^{a \rightarrow i}.$$

Then:

$$\text{Spr}_{v_i}(n^{-\log^4 n/2}) \leq \min \left\{ \varepsilon, \max \left\{ \varepsilon^{\text{Tier}(i)}, C \sqrt{\frac{\log^{15} n}{nd}} \right\} \right\} \quad \text{and} \quad \|v_i\|_\infty \leq \frac{2}{p^{\log^2 n}}.$$

where  $\varepsilon = C \sqrt{\frac{\log^{27} n}{d}}$  for some universal constant  $C$ .

*Proof.* In this proof, we will use  $j_1, \dots, j_s$  to denote the children of  $i$  in  $T$ . Then  $\text{Ch}(i) = \{a_1, \dots, a_s\}$  where  $a_t$  connects  $i$  and  $j_t$  in the factor graph.

We first show the weaker statement that  $\text{Spr}_{v_i}(n^{-\log^4 n/2}) \leq \varepsilon$  and  $\|v_i\|_\infty \leq \frac{2}{p^{\log^2 n}}$  for every  $i \in V(T)$  via induction on the depth of the subtree rooted at  $i$ . When this depth is equal to 0, the vertex  $i$  has no children, and then the result is clearly true since  $v_i = \rho$ .

Now, suppose the weaker statement is true for every vertex with a depth- $q$  subtree. Let  $i$  be a vertex with a depth- $(q+1)$  subtree. Note that  $m^{a_t \rightarrow i} = P m^{j_t \rightarrow a_t}$ . We know that  $m^{j_t \rightarrow a_t} \propto v_{j_t} \cdot \mathbf{1}[L_{j_t}]$ . On one hand

$$\|v_{j_t} \cdot \mathbf{1}[L_{j_t}]\|_\infty \leq \frac{2}{p^{\log^2 n}}$$

and on the other hand by the induction hypothesis and our assumption on  $L_{j_t}$ ,

$$\|v_{j_t} \cdot \mathbf{1}[L_{j_t}]\|_1 \geq (1 - \varepsilon) \cdot (n^{-\log^3 n} - n^{-\log^4 n/2}) \geq \frac{1}{2} n^{-\log^3 n},$$

and thus

$$\|m^{j_t \rightarrow a_t}\|_\infty \leq n^{\log^5 n}.$$

By [Lemma 7.16](#), we know:

$$\text{Spr}_{m^{a_t \rightarrow i}}(n^{-\log^4 n}) \leq C \sqrt{\frac{\log^{23} n}{d}}$$

for some absolute constant  $C$ . We also know  $\|m^{a_t \rightarrow i}\|_\infty \leq \frac{1}{p}$  by [Observation 7.7](#), and that  $s \leq \log^2 n$  by assumption. Therefore, by [Lemma 7.17](#),

$$\text{Spr}_{v_i}(n^{-\log^4 n/2}) \leq C \sqrt{\frac{\log^{27} n}{d}}$$

which establishes the weaker statement.

Next, we prove  $\text{Spr}_{v_i}(n^{-\log^4 n/2}) \leq \max\left\{\varepsilon^{\text{Tier}(i)}, \sqrt{\frac{\log^{15} n}{nd}}\right\}$ . We prove this statement by induction on  $\text{Tier}(i)$ . The base case of  $0 \leq \text{Tier}(i) \leq 1$  follows from the weaker statement proved above.

Now, suppose  $\text{Tier}(i) \geq 2$ . Note that  $\|v_{j_t}\|_\infty \leq \frac{2}{p \log^2 n}$  and  $m^{j_t \rightarrow a_t} \propto v_{j_t} \cdot \mathbf{1}[L_{j_t}]$ . Since  $\text{Tier}(j_t) \geq \text{Tier}(i) - 1$ , we know  $\text{Tier}(j_t) \geq 1$ , which in particular means that  $j_t$  is a free vertex and consequently

$$\text{Spr}_{PL_{j_t}}(n^{-\log^4 n}) \leq \sqrt{\frac{\log^{11} n}{nd}}.$$

By our induction hypothesis, the value of  $v_{j_t} \cdot \mathbf{1}[L_{j_t}]$  is in  $1 \pm \max\left\{\varepsilon^{\text{Tier}(j_t)}, \sqrt{\frac{\log^{15} n}{nd}}\right\}$  on a subregion of  $L_{j_t}$  of measure at least:

$$\rho(L_{j_t}) - n^{-\log^4 n} \geq (1 - n^{-\log^4 n/2}) \cdot \rho(L_{j_t})$$

by our assumption that  $\rho(L_{j_t}) \geq n^{-\log^3 n}$ . Combined with our bound on  $\|v_{j_t}\|_\infty$ , we know that:

$$\|v_{j_t} \cdot \mathbf{1}[L_{j_t}]\|_1 \in \rho(L_{j_t}) \cdot \left(1 \pm \max\left\{\varepsilon^{\text{Tier}(j_t)}, \sqrt{\frac{\log^{15} n}{nd}}\right\} \pm n^{-\log^4 n/2}\right).$$

We can then write  $m^{j_t \rightarrow a_t}$  as  $L_{j_t} + \Delta$  where  $\|\Delta\|_1 \leq \max\left\{\varepsilon^{\text{Tier}(j_t)}, \sqrt{\frac{\log^{15} n}{nd}}\right\}$  and  $\|\Delta\|_\infty \leq n^{\log^5 n}$ . By [Lemma 7.16](#),  $m^{a_t \rightarrow i} = P m^{j_t \rightarrow a_t}$  satisfies the following for some constant  $C$ :

$$\begin{aligned} \text{Spr}_{m^{a_t \rightarrow i}}(n^{-\log^4 n}) &\leq \max\left\{\sqrt{\frac{\log^{11} n}{nd}}, C \sqrt{\frac{\log^{23} n}{d}} \max\left\{\varepsilon^{\text{Tier}(j_t)}, \sqrt{\frac{\log^{15} n}{nd}}\right\}\right\} \\ &= \max\left\{\sqrt{\frac{\log^{11} n}{nd}}, C \sqrt{\frac{\log^{23} n}{d}} \cdot \varepsilon^{\text{Tier}(j_t)}\right\} \end{aligned}$$

where the equality is due to our lower bound on  $d$ . Additionally, since  $\|m^{a_t \rightarrow i}\|_\infty \leq \frac{1}{p}$  and  $s \leq \log^2 n$ , we can use [Lemma 7.17](#) to conclude:

$$\begin{aligned} \text{Spr}_{v_i}(n^{-\log^4 n/2}) &\leq C \log^2 n \max\left\{\sqrt{\frac{\log^{11} n}{nd}}, \varepsilon^{\text{Tier}(i)-1} \sqrt{\frac{\log^{23} n}{d}}\right\} \\ &\leq \max\left\{\varepsilon^{\text{Tier}(i)}, C \sqrt{\frac{\log^{15} n}{nd}}\right\}, \end{aligned}$$

which completes the proof.  $\square$

### 7.3.6 Neighborhood containment for relaxed distribution

As final preparation to prove [Lemma 7.4](#), we list some properties that hold with high probability over the randomness of  $\mathbf{G}_{n-1}$  and  $\mathbf{W}_R$  for the forest  $T(\mathbf{G}_{n-1})$ , and proceed by conditioning on these high probability events for the rest of this section.

**Lemma 7.23.** *The following events simultaneously occur with probability at least  $1 - n^{-\Omega(\log n)}$  over the randomness of  $\mathbf{G}_{n-1}$  and  $\mathbf{W}_R$ , and all choices of  $T(\mathbf{G}_{n-1})$ ,*

- (1) *The maximum degree in  $\mathbf{G}_{n-1}$ , and hence the maximum degree in  $T(\mathbf{G}_{n-1})$  is bounded by  $\log^2 n$ .*
- (2) *For every  $x \in T(\mathbf{G}_{n-1})$ ,  $\rho(L_x(\mathbf{W}_R)) \geq n^{-\log^3 n}$ .*
- (3)  *$\text{Spr}_{PL(\mathbf{W}_R)}(n^{-\log^4 n}) \leq \sqrt{\frac{\log^{11} n}{nd}}$ .*
- (4)  *$|K| \leq n^{\max\{\log(2\alpha), 1\}/\log \log n}$ .*

*Proof.* [Property \(1\)](#) follows from [Lemma 3.11](#), [Property \(3\)](#) follows from [Lemma 7.15](#), and [Property \(4\)](#) follows from [Proposition 1.3](#) and [Lemma 3.12](#). It remains to prove [Property \(2\)](#). To obtain a lower bound on  $\rho(L_x(\mathbf{W}_R))$ , we use:

$$\rho(L_x(\mathbf{W}_R)) \geq \rho(L_x(\mathbf{W})).$$

We will first prove that with probability  $1 - n^{-\Omega(\log n)}$ , for all  $Y \subseteq [n-1]$  such that  $|Y| \leq \log^2 n$ ,

$$\rho\left(\bigcap_{i \in Y} \text{cap}(\mathbf{w}_i) \cap \bigcap_{i \in [n-1] \setminus Y} \overline{\text{cap}}(\mathbf{w}_i)\right) \geq n^{-\log^3 n},$$

which would then imply the desired statement. Suppose  $\mathbf{w}_1, \dots, \mathbf{w}_{n-1} \sim \rho^{\otimes n-1}$ , then by [Lemma 5.1](#) under setting the  $s$  parameter to, say,  $1/2$ , and a union bound over all  $Y$  such that  $|Y| \leq \log^2 n$ :

$$\rho\left(\bigcap_{i \in Y} \text{cap}(\mathbf{w}_i) \cap \bigcap_{i \in [n-1] \setminus Y} \overline{\text{cap}}(\mathbf{w}_i)\right) \geq n^{-\log^3 n} \quad \forall Y \subseteq [n-1], |Y| \leq \log^2 n \quad (16)$$

except with probability at most  $n^{-\log n}$ . By [Observation 3.14](#), (16) is true with probability  $1 - n^{-\Omega(\log n)}$  when  $\mathbf{w}_1, \dots, \mathbf{w}_{n-1} \sim \rho^{\mathbf{G}_{n-1}}$  with probability  $1 - n^{-\Omega(\log n)}$  over the randomness of  $\mathbf{G}_{n-1}$ , which establishes [Property \(2\)](#).  $\square$

Now we proceed with the proof of [Lemma 7.4](#).

**Lemma** (Restatement of [Lemma 7.4](#)). *Suppose  $\ell \geq 1$ . Then for every  $S$  such that  $|S| \leq \log^2 n$ :*

$$\Pr_{\substack{\mathbf{w}_n \sim \mathbb{S}^{d-1} \\ \mathbf{W}_K \sim \mathcal{F}'(\mathbf{W}_R)}} [\forall x \in S : \langle \mathbf{w}_x, \mathbf{w}_n \rangle \geq \tau(p)] \in (1 \pm \eta(\ell)) p^{|S|}$$

*except with probability at most  $n^{-\Omega(\log n)}$  over the randomness of  $\mathbf{G}_{n-1}$  where*

$$\eta(\ell) := \max \left\{ \left( C \sqrt{\frac{\log^{27} n}{d}} \right)^\ell, 4 \sqrt{\frac{\log^{11} n}{nd}} \right\}$$

*for some absolute constant  $C$ .*

*Proof of Lemma 7.4.* Let  $\mathbf{W}_K \sim \mathcal{F}'(\mathbf{W}_R)$ , and without loss of generality let  $S = \{1, \dots, s\}$ . First observe that  $\mathbf{w}_1, \dots, \mathbf{w}_s$  are independent conditioned on  $\mathbf{W}_R$ . Now, we use  $C_i$  to denote the connected component of  $T(\mathbf{G}_{n-1})$  where  $i$  belongs. With probability  $1 - n^{-\Omega(\log n)}$  over the randomness of  $\mathbf{G}_{n-1}$  and  $\mathbf{W}_R$ , the events in the statement of Lemma 7.23 occur. For the rest of the proof we assume that these events occur. Then every  $C_i$  satisfies Assumption 7.18 and  $\text{Anch}(C) \subseteq S_{\mathbf{G}_{n-1}}(v, \ell - 1)$ . Thus,  $\text{Tier}(i) = \ell - 1$ , and consequently by Lemma 7.22 the distribution

$$v_i \propto \prod_{a \in \text{Ch}(i)} m^{a-i}$$

satisfies the following for some absolute constant  $C$ :

$$\text{Spr}_{v_i}(n^{-\log^4 n}) \leq \max \left\{ \sqrt{\frac{\log^{27} n}{d}}^{\ell-1}, C \sqrt{\frac{\log^{15} n}{nd}} \right\} \quad \text{and} \quad \|v_i\|_\infty \leq \frac{2}{p^{\log^2 n}}.$$

The marginal on  $\mathbf{w}_i$  is

$$\tilde{v}_i \propto v_i \cdot \mathbf{1}[L_x(\mathbf{W}_R)].$$

$L_x(\mathbf{W}_R) = L(\mathbf{W}_R)$  when  $\ell \geq 2$ , and hence for  $\ell \geq 1$ , this satisfies

$$\tilde{v}_i = L(\mathbf{W}_R) + \Delta$$

with  $\|\Delta\|_1 \leq 2 \max \left\{ \left( C \sqrt{\frac{\log^{27} n}{d}} \right)^{\ell-1}, C \sqrt{\frac{\log^{15} n}{nd}} \right\}$  and  $\|\tilde{v}_i\|_\infty \leq n^{\log^5 n}$ .

Then by the definition of  $P$ :

$$\Pr_{\substack{\mathbf{w}_n \sim \mathbb{S}^{d-1} \\ \mathbf{W}_K \sim \mathcal{F}'(\mathbf{W}_R)}} [\forall x \in S : \langle \mathbf{w}_x, \mathbf{w}_n \rangle \geq \tau(p)] = p^s \mathbf{E}_{\mathbf{w}_n \sim \mathbb{S}^{d-1}} \prod_{i=1}^s P \tilde{v}_i(\mathbf{w}_n)$$

From Lemma 7.16, we know:

$$\begin{aligned} \text{Spr}_{P \tilde{v}_i}(2n^{-\log^4 n}) &\leq \max \left\{ 2 \sqrt{\frac{\log^{11} n}{nd}}, 2C \cdot \left( C \sqrt{\frac{\log^{27} n}{d}} \right)^{\ell-1} \cdot \sqrt{\frac{\log^{23} n}{d}} \right\} \\ &\leq \max \left\{ 2 \sqrt{\frac{\log^{11} n}{nd}}, \left( C \sqrt{\frac{\log^{27} n}{d}} \right)^\ell \right\}. \end{aligned}$$

Using the above along with the fact that  $\|P \tilde{v}_i\|_\infty \leq \frac{1}{p}$  for all  $i$  implies:

$$\Pr_{\substack{\mathbf{w}_n \sim \mathbb{S}^{d-1} \\ \mathbf{W}_K \sim \mathcal{F}'(\mathbf{W}_R)}} [\forall x \in S : \langle \mathbf{w}_x, \mathbf{w}_n \rangle \geq \tau(p)] \in \left( 1 \pm \max \left\{ 4 \sqrt{\frac{\log^{11} n}{nd}}, \left( 2C \sqrt{\frac{\log^{27} n}{d}} \right)^\ell \right\} \right) p^s$$

which completes the proof.  $\square$

## 7.4 Relating relaxed and true distributions

In this section, we relate the uniform distribution on  $\mathcal{F}(\mathbf{W}_R)$  to the uniform distribution on  $\mathcal{F}'(\mathbf{W}_R)$ . We first prove that a random sample from  $\mathcal{F}'(\mathbf{W}_R)$  falls in  $\mathcal{F}(\mathbf{W}_R)$  with high probability.

**Lemma 7.24.** *With probability at least  $1 - n^{-\Omega(\log n)}$  over the randomness of  $\mathbf{G}_{n-1}$  and  $\mathbf{W}_R$ ,*

$$\Pr_{(\mathbf{w}_i)_{i \in K} \sim \mathcal{F}'(\mathbf{W}_R)} \left[ (\mathbf{w}_i)_{i \in K} \notin \mathcal{F}(\mathbf{W}_R) \right] \leq pn^{O(1/\log \log n)}.$$

*Proof.* It suffices to prove that for every pair of vertices  $x, y \in K$  such that  $d_{T(\mathbf{G}_{n-1})}(x, y) \geq 2$ , the probability that  $\langle \mathbf{w}_x, \mathbf{w}_y \rangle \geq \tau(p)$  is at most  $2p$ . We can then apply the union bound over all  $x, y$  pairs, of which there are  $n^{O(1/\log \log n)}$  by [Property \(4\) of Lemma 7.23](#). Let's assume that the events in [Lemma 7.23](#) hold, which happens with probability at least  $1 - n^{-\Omega(\log n)}$ . Consequently,  $T(\mathbf{G}_{n-1})$  satisfies [Assumption 7.18](#). The marginal on  $\mathbf{w}_x$  is given by

$$\tilde{v}_x \propto \prod_{a \in \text{Ch}(x)} m^{a \rightarrow x} \cdot \mathbf{1}[L_x].$$

By [Lemma 7.22](#) and the fact that  $\rho(L_x) \geq n^{-\log^3 n}$ , we can write  $\tilde{v}_x$  as  $L_x + \Delta$  where  $\|\Delta\|_1 \leq C \sqrt{\frac{\log^{27} n}{d}}$  for some constant  $C$  and  $\|\tilde{v}_x\|_\infty \leq n^{\log^5 n}$ . By [Corollary 4.10](#), for  $\mathbf{w}_x \sim \tilde{v}_x$ , (1)  $\rho(\text{cap}(\mathbf{w}_x) \cap L_y) \leq \frac{3p}{2} \cdot \rho(L_y)$ , and (2)  $\rho(\text{cap}(\mathbf{w}_x) \cap L_z) \geq \frac{p}{2} \cdot \rho(L_z)$  for all  $z \in V(T(\mathbf{G}_{n-1}))$  simultaneously except with probability at most  $n^{-\log^2 n}$ .

The distribution of  $\mathbf{w}_y | \mathbf{w}_x$  is the marginal on variable indexed by  $y$  of the uniform distribution over satisfying assignments to the following CSP instance. Define  $T'$  as the tree obtained by deleting the subtree rooted at  $x$  from  $T(\mathbf{G}_{n-1})$ . Let  $u$  denote the parent of  $x$  in  $T$ . Now, let  $F'$  be the factor graph of the CSP instance  $\mathcal{F}'$  with variables indexed by  $V(T')$  and constraints:

- **Unary constraints.** For each  $i \in V(T')$ , if  $i \neq u$ , then  $v_i \in L_i(\mathbf{W}_R)$ , and if  $i = u$ , then  $v_u \in L_u(\mathbf{W}_R) \cap \text{cap}(\mathbf{w}_x)$ .
- **Binary constraints.** For every  $\{i, j\} \in T'$ :  $\langle v_i, v_j \rangle \geq \tau(p)$ .

The distribution of  $\mathbf{w}_y | \mathbf{w}_x$  can be computed from the belief propagation fixed point on  $F'$ .

Suppose  $\mathbf{w}_x$  is such that (1) and (2) hold. Since (2) holds,  $F'$  satisfies [Assumption 7.18](#). By [Lemma 7.22](#) along with  $\rho(L_y) \geq n^{-\log^3 n}$  and the assumption that (1) holds, we know:

$$\Pr_{\mathbf{w}_y | \mathbf{w}_x} \left[ \mathbf{w}_y \in \text{cap}(\mathbf{w}_x) \cap L_y \right] \leq 1.6p.$$

Since (1) and (2) hold simultaneously except with probability at most  $n^{-\log^2 n}$ :

$$\Pr_{\mathbf{w}_x, \mathbf{w}_y} \left[ \langle \mathbf{w}_x, \mathbf{w}_y \rangle \geq \tau(p) \right] = \Pr_{\mathbf{w}_x, \mathbf{w}_y} \left[ \mathbf{w}_y \in \text{cap}(\mathbf{w}_x) \cap L_y \right] \leq 2p. \quad \square$$

Our next ingredient is to prove that individual instantiations of vectors from the uniform distribution over  $\mathcal{F}'(\mathbf{W}_R)$  produce an intersection of sphere caps whose measure concentrates reasonably well.

**Lemma 7.25.** *With probability at least  $1 - n^{-\Omega(\log n)}$  over the randomness of  $\mathbf{W}_R$  and  $\mathbf{G}_{n-1}$ ,*

$$\Pr_{(\mathbf{w}_i)_{i \in K} \sim \mathcal{F}'(\mathbf{W}_R)} \left[ \rho \left( \bigcap_{j \in S} \text{cap}(\mathbf{w}_j) \right) \notin \left( 1 \pm O \left( \frac{\log^{15} n}{\sqrt{d}} \right) \right) p^{|S|} \right] \leq n^{-\log^5 n}.$$

*Proof.* Each  $j \in S$  is in a separate connected component of the factor graph  $F(\mathbf{W}_R)$  and so the collection  $(\mathbf{w}_j)_{j \in S}$  is a collection of independent random vectors. Denoting the distribution of  $\mathbf{w}_j$  with  $\tilde{v}_j$ , we know by [Lemma 7.23](#) that with probability at least  $1 - n^{-\Omega(\log n)}$  over the randomness of  $\mathbf{G}_{n-1}$  and  $\mathbf{W}_R$  that every connected component of  $T(\mathbf{G}_{n-1})$  satisfies [Assumption 7.18](#). [Lemma 7.22](#) along with [Property \(2\)](#) from [Lemma 7.23](#) then implies that  $\|\tilde{v}_j\|_\infty \leq n^{\log^5 n}$ . Since  $|S| \leq \log^2 n$ , the statement follows from an iterated application of [Corollary 4.10](#) and a union bound over all applications.  $\square$

We are now finally prepared to prove [Lemma 7.2](#).

**Lemma** (Restatement of [Lemma 7.2](#)). *With probability  $1 - n^{-\Omega(\log n)}$ , for every  $S \subseteq [n-1]$  such that  $|S| \leq \log^2 n$ , the following is true. Suppose  $\ell < \frac{\log n}{\log \log n}$  is such that the balls  $\{\mathbf{B}_{\mathbf{G}_{n-1}}(x, \ell) : x \in S\}$  are all trees and are pairwise disjoint, then:*

$$\Pr[N_{\mathbf{G}}(n) \supseteq S | \mathbf{G}_{n-1}] \in (1 \pm \eta(\ell)) p^{|S|}$$

where

$$\eta(\ell) := \min \left\{ \frac{\log^8 n}{\sqrt{d}}, \max \left\{ 2\sqrt{\frac{\log^{28} n}{d}}, 8\sqrt{\frac{\log^{11} n}{d}} \right\} \right\}.$$

*Proof of Lemma 7.2.* We first establish that  $\eta(\ell) \leq \frac{\log^8 n}{\sqrt{d}}$ . Recall that:

$$\Pr[N_{\mathbf{G}}(n) \supseteq S | \mathbf{G}_{n-1}] = \mathbf{E}_{\mathbf{W} \sim \rho^{\mathbf{G}_{n-1}}} \rho \left( \bigcap_{i \in S} \text{cap}(\mathbf{w}_i) \right).$$

Suppose  $\mathbf{W} \sim \rho^{\otimes n-1}$ , then by an iterated application of [Corollary 4.8](#) and a union bound, with probability at least  $1 - n^{-\Omega(\log^3 n)}$ ,

$$\rho \left( \bigcap_{i \in S} \text{cap}(\mathbf{w}_i) \right) \in \left( 1 \pm \frac{\log^8 n}{\sqrt{d}} \right). \quad (17)$$

By [Observation 3.14](#), the same probability bound for (17) holding is true even when  $\mathbf{W} \sim \rho^{\mathbf{G}_{n-1}}$  with probability at least  $1 - n^{-\Omega(\log^2 n)}$  over the randomness of  $\mathbf{G}_{n-1}$ .

The fact that  $\rho(\bigcap_{i \in S} \text{cap}(\mathbf{w}_i))$  is always in  $[0, 1]$  and the upper bound on  $d$  together imply that

$$\Pr[N_{\mathbf{G}}(n) \supseteq S | \mathbf{G}_{n-1}] \in \left( 1 \pm O\left(\frac{\log^8 n}{\sqrt{d}}\right) \right) p^{|S|}$$

with probability at least  $1 - n^{-\Omega(\log^3 n)}$ . This establishes the desired statement when  $\ell = 0$ .

Henceforth, we can assume  $\ell \geq 1$ , which in particular means  $K$  is nonempty. Now we turn our attention to proving that:

$$\Pr[N_{\mathbf{G}}(n) \supseteq S | \mathbf{G}_{n-1}] \in (1 \pm 2\eta(\ell)) \cdot p^{|S|}$$

where  $\eta(\ell) := \max \left\{ \sqrt{\frac{\log^{28} n}{d}}, 4\sqrt{\frac{\log^{11} n}{nd}} \right\}$ . Let  $\mathbf{G}_{n-1}$  and  $\mathbf{W}_R$  be such that the conclusions of [Lemma 7.4](#), [Lemma 7.24](#), and [Lemma 7.25](#) hold. By [Lemma 7.4](#),

$$\mathbf{E}_{\mathbf{W}_K \sim \mathcal{F}'(\mathbf{W}_R)} \rho \left( \bigcap_{i \in S} \text{cap}(\mathbf{w}_i) \right) \in (1 \pm \eta(\ell)) p^{|S|}.$$

Denoting  $\Pr_{\mathbf{W}_K \sim \mathcal{F}'(\mathbf{W}_R)}[\mathbf{W}_K \in \mathcal{F}(\mathbf{W}_R)]$  as  $q$ , we can also write:

$$\mathbf{E}_{\mathbf{W}_K \sim \mathcal{F}'(\mathbf{W}_R)} \rho \left( \bigcap_{i \in S} \text{cap}(\mathbf{w}_i) \right) = q \cdot \mathbf{E}_{\mathbf{W}_K \sim \mathcal{F}'(\mathbf{W}_R) | \mathbf{W}_K \in \mathcal{F}(\mathbf{W}_R)} \rho \left( \bigcap_{i \in S} \text{cap}(\mathbf{w}_i) \right) + (1-q) \cdot \mathbf{E}_{\mathbf{W}_K \sim \mathcal{F}'(\mathbf{W}_R) | \mathbf{W}_K \notin \mathcal{F}(\mathbf{W}_R)} \rho \left( \bigcap_{i \in S} \text{cap}(\mathbf{w}_i) \right).$$

The first term on the right hand side of the above can be written as:

$$q \cdot \mathbf{E}_{\mathbf{W}_K \sim \mathcal{F}'(\mathbf{W}_R)} \rho \left( \bigcap_{i \in S} \text{cap}(\mathbf{w}_i) \right).$$

By [Lemma 7.25](#) and [Observation 3.2](#) along with the fact that  $\rho(\bigcap_{i \in S} \text{cap}(\mathbf{w}_i))$  is in  $[0, 1]$ , the second term in the right hand side is in

$$(1-q) \cdot \left( \left( 1 \pm \frac{\log^{15} n}{\sqrt{d}} \right) \cdot p^{|S|} \pm \frac{n^{-\log^5 n}}{1-q} \right) \subseteq (1-q) \cdot \left( 1 \pm \frac{\log^{15} n}{\sqrt{d}} \right) \cdot p^{|S|} \pm n^{-\log^5 n}$$

Rearranging the above in the expression

$$q \cdot \mathbf{E}_{\mathbf{W}_K \sim \mathcal{F}'(\mathbf{W}_R) | \mathbf{W}_K \in \mathcal{F}(\mathbf{W}_R)} \rho \left( \bigcap_{i \in S} \text{cap}(\mathbf{w}_i) \right) + (1-q) \cdot \mathbf{E}_{\mathbf{W}_K \sim \mathcal{F}'(\mathbf{W}_R) | \mathbf{W}_K \notin \mathcal{F}(\mathbf{W}_R)} \rho \left( \bigcap_{i \in S} \text{cap}(\mathbf{w}_i) \right) \in (1 \pm \eta(\ell)) p^{|S|}$$

gives

$$\mathbf{E}_{\mathbf{W}_K \sim \mathcal{F}'(\mathbf{W}_R)} \rho \left( \bigcap_{i \in S} \text{cap}(\mathbf{w}_i) \right) \in \left( 1 \pm \frac{1-q}{q} \cdot \frac{\log^{15} n}{\sqrt{d}} \pm \frac{\eta(\ell)}{q} \right) p^{|S|}$$

By [Lemma 7.24](#),  $q \geq 1 - pn^{O(1/\log \log n)}$ , which implies that:

$$\Pr[N_{\mathbf{G}}(n) \supseteq S | \mathbf{G}_{n-1}] = \mathbf{E}_{\mathbf{W}_K \sim \mathcal{F}'(\mathbf{W}_R)} \rho \left( \bigcap_{i \in S} \text{cap}(\mathbf{w}_i) \right) \in (1 \pm 2\eta(\ell)) \cdot p^{|S|}$$

The above holds with probability at least  $1 - n^{-\Omega(\log n)}$  over the randomness of  $\mathbf{G}_{n-1}$  and  $\mathbf{W}_R$  since we only assumed the conclusions of [Lemma 7.4](#), [Lemma 7.24](#) and [Lemma 7.25](#) hold, which completes the proof of [Lemma 7.2](#).  $\square$

## 7.5 From neighborhood containment to neighborhood equality probabilities

In this section we prove [Lemma 7.1](#) using [Lemma 7.2](#).

**Lemma** (Restatement of [Lemma 7.1](#)). *With probability  $1 - n^{-\Omega(\log n)}$  over the randomness of  $\mathbf{G}_{n-1}$ , for every  $S \subseteq [n-1]$  such that  $|S| \leq \log^2 n$ , the following is true. Let  $\ell < \frac{\log n}{\log \log n}$  be such that the balls  $\{B_{\mathbf{G}_{n-1}}(x, \ell) : x \in S\}$  are pairwise disjoint trees, and let  $d > \log^{20} n$ . Then:*

$$\Pr[N_{\mathbf{G}}(n) = S | \mathbf{G}_{n-1}] \in (1 \pm 2\eta(\ell)) p^{|S|} (1-p)^{n-1-|S|}$$

where

$$\eta(\ell) := \min \left\{ \frac{\log^8 n}{\sqrt{d}}, \max \left\{ 2 \left( \sqrt{\frac{\log^{28} n}{d}} \right)^\ell, 8 \cdot \sqrt{\frac{\log^{11} n}{nd}} \right\} \right\}.$$

*Proof of [Lemma 7.1](#).* Let  $\mathbf{C} = \bigcap_{j \in S} \text{cap}(\mathbf{w}_j)$ , and let  $\mathbf{A} = \bigcap_{j \in [n-1] \setminus S} \overline{\text{cap}}(\mathbf{w}_j)$ . We note that the area of  $\mathbf{C}$  is unlikely to be too small:

**Claim 7.26.** With probability at least  $1 - n^{-\Omega(\log n)}$ ,  $\rho(\mathbf{C}) \geq \frac{1}{2}p^{|\mathcal{S}|} \geq \Omega(n^{-\log n})$ .

We defer the proof of the claim to the end. We can thus use [Claim 7.26](#) to apply [Corollary 5.2](#) and conclude that with probability at least  $1 - n^{-\Omega(\log n)}$ ,

$$\rho(\mathbf{A} \cap \mathbf{C}) \in \left(1 \pm \sqrt{\frac{\log^{11} n}{nd}}\right) \cdot \rho(\mathbf{C}) \cdot (1-p)^{n-1-|\mathcal{S}|}. \quad (18)$$

Now from [Lemma 7.2](#), with probability  $1 - n^{-\Omega(\log n)}$ ,

$$\mathbf{E}_{\rho^{\mathbf{G}_{n-1}}} [\rho(\mathbf{C})] = \Pr[N_{\mathbf{G}}(n)S \mid \mathbf{G}_{n-1}] \in (1 \pm \eta(\ell))p^{|\mathcal{S}|}. \quad (19)$$

Let  $\mathcal{E}$  be the event that [Claim 7.26](#), (18) and (19) hold. We then have that

$$\begin{aligned} \Pr[N_{\mathbf{G}}(n) = S \mid \mathbf{G}_{n-1}, \mathcal{E}] &= \mathbf{E}[\rho(\mathbf{A} \cap \mathbf{C}) \mid \mathcal{E}] = \left(1 \pm \sqrt{\frac{\log^{11} n}{nd}}\right) \cdot \mathbf{E}[\rho(\mathbf{C}) \mid \mathcal{E}] \cdot (1-p)^{n-1-|\mathcal{S}|} \\ &= \left(1 \pm \sqrt{\frac{\log^{11} n}{nd}}\right) \cdot (1 \pm \eta(\ell))p^{|\mathcal{S}|} (1-p)^{n-1-|\mathcal{S}|}, \end{aligned}$$

and since  $\sqrt{\frac{\log^{11} n}{nd}} \leq \eta(\ell)/2 \ll 1$ , we have our conclusion.

Now we prove [Claim 7.26](#). We'll apply [Lemma 5.1](#) to conclude that any  $m = |\mathcal{S}|$  vectors from  $\rho^{\otimes n}$  have cap intersection area  $\geq e^{-\text{poly} \log n}$  with overwhelmingly large probability, then apply [Observation 3.14](#) to conclude that the same is true for vectors sampled from  $\rho^{\mathbf{G}_{n-1}}$ .

If  $\mathbf{v}_1, \dots, \mathbf{v}_n \sim \rho^{\otimes n}$ , then from [Lemma 5.1](#) combined with a union bound over subsets of  $[n]$  of size  $|\mathcal{S}|$ , so long as  $d \geq \log^{20} n$ ,

$$\Pr_{\rho^{\otimes n}} \left[ \rho \left( \bigcap_{j \in \mathcal{S}} \text{cap}(\mathbf{v}_j) \right) < \frac{1}{2}p^{|\mathcal{S}|} \right] \leq \binom{n}{|\mathcal{S}|} \cdot \Pr_{\rho^{\otimes |\mathcal{S}|}} \left[ \rho \left( \bigcap_{j=1}^{|\mathcal{S}|} \text{cap}(\mathbf{v}_j) \right) < \frac{1}{2}p^{|\mathcal{S}|} \right] \leq \binom{n}{|\mathcal{S}|} \cdot n^{-\log^3 n} \leq n^{-.5 \log^3 n},$$

for  $n$  sufficiently large. Hence by [Observation 3.14](#),

$$\Pr_{\mathbf{w}_1, \dots, \mathbf{w}_{n-1} \sim \rho^{\mathbf{G}_{n-1}}} \left[ \rho \left( \bigcap_{j \in \mathcal{S}} \text{cap}(\mathbf{w}_j) \right) < \frac{1}{2}p^{|\mathcal{S}|} \right] \leq n^{-.25 \log^3 n},$$

with probability at least  $1 - n^{-.25 \log^3 n}$  over the randomness of  $\mathbf{G}_{n-1}$ , completing the proof.  $\square$

## 8 Total variation bound

In this section, we prove [Theorem 1.2](#) and [Theorem 1.1](#). As was done in the prior work by Brennan, Bresler, and Nagaraj [[BBN20](#)], we use an analogue of the tensorization of the relative entropy for non-product measures:

**Claim 8.1** (Relative entropy tensorization, similar to Lemma 2.1 of [[BBN20](#)]). Suppose  $\mu = \mu_1 \otimes \dots \otimes \mu_n$  is a product measure and  $\nu$  is a measure over the same domain. Let  $\nu_t$  denote the marginal of  $\nu$  on the  $t$ -th coordinate  $x_t$ , and let  $x_{a:b}$  denote coordinates  $a$  through  $b$  of  $x$ . Then

$$D(\nu \parallel \mu) = \sum_{t=1}^n \mathbf{E}_{x_{1:t-1} \sim \nu} [D(\nu_t(x_t \mid x_{1:t-1}) \parallel \mu_t)].$$

*Proof.* By the chain rule for relative entropy,

$$\mathbf{E}_{\mathbf{x} \sim \nu} \log \frac{\nu(\mathbf{x})}{\mu(\mathbf{x})} = \sum_{t=1}^n \mathbf{E}_{\mathbf{x} \sim \nu} \log \frac{\nu_t(\mathbf{x}_t | \mathbf{x}_{1:t-1})}{\mu_t(\mathbf{x}_t)} = \sum_{t=1}^n \mathbf{E}_{\mathbf{x}_{1:t-1} \sim \nu} \left( \mathbf{E}_{\mathbf{x}_t \sim \nu_t | \mathbf{x}_{1:t-1}} \log \frac{\nu_t(\mathbf{x}_t | \mathbf{x}_{1:t-1})}{\mu_t(\mathbf{x}_t)} \right),$$

by linearity of expectation and by definition of the marginal distribution. The expression on the right simplifies using the definition of the relative entropy, completing the proof.  $\square$

In combination with Pinsker's inequality, this lemma reduces bounding the TV distance between a product measure  $\mu$  and a general measure  $\nu$ , to bounding the relative entropy  $D(\nu_t(x_t | x_{1:t-1}) \| \mu_t)$ .

**Claim 8.2.** Let  $\mu_t$  be the distribution of the neighborhood of  $t$  to vertices  $[t-1]$  under  $G(n, p)$ , and let  $\nu_t(\cdot | G_{t-1})$  be the distribution of the neighborhood of  $t$  under  $\text{Geo}_d(n, p)$ , conditioned on the subgraph  $G_{t-1} \sim \text{Geo}_d(t-1, p)$  on the vertices  $[t-1]$ . Then,

$$2d_{\text{TV}}(\text{Geo}_d(n, p), G(n, p))^2 \leq n \cdot \mathbf{E}_{G_{n-1} \sim \text{Geo}(n-1, p)} [D(\nu_n(\cdot | G_{n-1}) \| \mu_t)]$$

*Proof.* Applying Pinsker's inequality ([Theorem 3.3](#)),  $2 \cdot d_{\text{TV}}(\nu, \mu)^2 \leq D(\nu \| \mu)$ . We then apply [Claim 8.1](#) with  $\mu = G(n, p)$  and  $\nu = \text{Geo}_d(n, p)$ , and  $\mu_t, \nu_t$  as defined in the claim:

$$2d_{\text{TV}}(\text{Geo}_d(n, p), G(n, p))^2 \leq \sum_{t=1}^n \mathbf{E}_{G_{t-1} \sim \text{Geo}(t-1, p)} [D(\nu_t(\cdot | G_{t-1}) \| \mu_t)]$$

Let  $G_S$  denote a graph over vertices  $S$ , and let  $\nu_t^S, \mu_t^S$  refer to the distribution of vertex  $t$ 's neighbors in  $S$  under  $\text{Geo}_d(n, p)$  and  $G(n, p)$  respectively. By symmetry,  $\mathbf{E}_{x_{[n] \setminus t} \sim \nu} [D(\nu_t^{[n] \setminus t}(\cdot | x_{[n] \setminus t}) \| \mu_n)]$  is the same for all  $t \in [n]$ . Via the chain rule for relative entropy, and the non-negativity of relative entropy,

$$D(\nu_t^{[n] \setminus t}(\cdot | G_{[n] \setminus t}) \| \mu_n) \geq D(\nu_t^{[t-1]}(\cdot | G_{[n] \setminus t}) \| \mu_t^{[t-1]}) = D(\nu_t^{[t-1]}(\cdot | G_{t-1}) \| \mu_t^{[t-1]})$$

The final equality comes from the fact that  $t$ 's neighbors in  $[t-1]$  only depends on  $G_{t-1}$ .

Upper bounding each  $D(\nu_t(\cdot | G_{t-1}) \| \mu_t)$  by  $D(\nu_n(\cdot | G_{n-1}) \| \mu_n)$  completes the proof.  $\square$

## 8.1 Bounding neighborhood relative entropy

Via [Claim 8.2](#), our goal now is to upper bound:

$$\mathbf{E}_{G_{n-1} \sim \text{Geo}(n-1, p)} [D(\nu_n(\cdot | G_{n-1}) \| \mu_t)] = \mathbf{E}_{G_{n-1} \sim \nu_{[n-1]}} \mathbf{E}_{S \sim \nu_n(\cdot | G_{n-1})} \log \frac{\nu_n(S | G_{n-1})}{\mu_n(S)} \leq o\left(\frac{1}{n}\right).$$

For most events under  $\nu_{[n-1]}$  and  $\nu_n(\cdot | G_{n-1})$ , we will upper bound the relative entropy via a Chi-square-like quantity: we use the Chi-squared distance, but we allow the omission of a low-probability event  $\mathcal{E}$  (to allow the removal of events which cause the Chi-square distance to blow up). We then specialize the resulting Chi-square-like bound ([Lemma 8.3](#)) by removing different low-probability events for the general  $p$  case ([Section 8.1.1](#)) and the sparse case ([Section 8.1.2](#)), and separately conclude [Theorem 1.2](#) and [Theorem 1.1](#). We now formally state the Chi-square bound:

**Lemma 8.3.** Let  $\mathcal{E}$  be an event satisfying both

$$\Pr_{G_{n-1}, \text{Geo}_d(n-1, p), S \sim \nu_n(\cdot | G_{n-1})} [\mathcal{E}] \leq o\left(\frac{1}{n^2 \log n}\right) \text{ and } \Pr_{G_{n-1} \sim \text{Geo}_d(n-1, p), S \sim \mu_n} [\mathcal{E}] \leq o\left(\frac{1}{n^2 \log n}\right)$$

and for  $S \subseteq [n-1]$ , define  $\Delta_{G_{n-1}}(S) = \frac{\nu_n(S | G_{n-1})}{\mu_n(S)} - 1$ . Then,

$$\mathbf{E}_{G_{n-1} \sim \nu_{[n-1]}} \mathbf{E}_{S \sim \nu_n(\cdot | G_{n-1})} \log \frac{\nu_n(S | G_{n-1})}{\mu_n(S)} \leq \mathbf{E}_{G_{n-1} \sim \nu_{[n-1]}} \mathbf{E}_{S \sim \mu_n} \Delta_{G_{n-1}}(S)^2 \cdot \mathbf{1}(\overline{\mathcal{E}}) + o\left(\frac{1}{n}\right)$$

*Proof.* Before we introduce the Chi-square bound, we first perform some conditioning on  $\bar{\mathcal{E}}$ .

$$\begin{aligned} \mathbf{E}_{\mathbf{G}_{n-1} \sim \nu_{[n-1]}} \mathbf{E}_{\mathbf{S} \sim \nu_n(\cdot | \mathbf{G}_{n-1})} \log \frac{v_n(\mathbf{S} | \mathbf{G}_{n-1})}{\mu_n(\mathbf{S})} &= \mathbf{E}_{\mathbf{G}_{n-1} \sim \nu_{[n-1]}} \mathbf{E}_{\mathbf{S} \sim \nu_n(\cdot | \mathbf{G}_{n-1})} \log (\Delta_{G_{n-1}}(\mathbf{S}) + 1) \\ &= \mathbf{E}_{\mathbf{G}_{n-1} \sim \nu_{[n-1]}} \mathbf{E}_{\mathbf{S} \sim \nu_n(\cdot | \mathbf{G}_{n-1})} \log (\Delta_{G_{n-1}}(\mathbf{S}) + 1) \cdot \mathbf{1}(\bar{\mathcal{E}}) + \max_S \left( \log \frac{v_n(\mathbf{S} | \mathbf{G}_{n-1})}{\mu_n(\mathbf{S})} \right) \cdot \mathbf{Pr}_{\text{Geo}_d(n,p)}[\mathcal{E}] \end{aligned}$$

The maximum of  $\log \frac{v_n(\mathbf{S} | \mathbf{G}_{n-1})}{\mu_n(\mathbf{S})}$  is upper bounded by  $n \log \frac{1}{p}$ , which comes from bounding  $v_n(\mathbf{S} | \mathbf{G}_{n-1})$ , a probability, above by 1, and achieving the smallest possible  $\mu_n(\mathbf{S})$  when  $|\mathbf{S}| = n - 1$ . Applying our assumption on  $\mathbf{Pr}[\mathcal{E}]$ , the term  $\max_S \left( \log \frac{v_n(\mathbf{S} | \mathbf{G}_{n-1})}{\mu_n(\mathbf{S})} \right) \cdot \mathbf{Pr}[\mathcal{E}]$  is at most  $o\left(\frac{1}{n}\right)$ .

We now turn our attention to the remaining expectation term:

$$\begin{aligned} \mathbf{E}_{\mathbf{G}_{n-1} \sim \nu_{[n-1]}} \mathbf{E}_{\mathbf{S} \sim \nu_n(\cdot | \mathbf{G}_{n-1})} \log (\Delta_{G_{n-1}}(\mathbf{S}) + 1) \cdot \mathbf{1}(\bar{\mathcal{E}}) &\leq \mathbf{E}_{\mathbf{G}_{n-1} \sim \nu_{[n-1]}} \mathbf{E}_{\mathbf{S} \sim \nu_n(\cdot | \mathbf{G}_{n-1})} \Delta_{G_{n-1}}(\mathbf{S}) \cdot \mathbf{1}(\bar{\mathcal{E}}) \\ &= \mathbf{E}_{\mathbf{G}_{n-1} \sim \nu_{[n-1]}} \mathbf{E}_{\mathbf{S} \sim \mu_n} (1 + \Delta_{G_{n-1}}(\mathbf{S})) \Delta_{G_{n-1}}(\mathbf{S}) \cdot \mathbf{1}(\bar{\mathcal{E}}) \\ &= \mathbf{E}_{\mathbf{G}_{n-1} \sim \nu_{[n-1]}} \mathbf{E}_{\mathbf{S} \sim \mu_n} \Delta_{G_{n-1}}(\mathbf{S}) \cdot \mathbf{1}(\bar{\mathcal{E}}) + \mathbf{E}_{\mathbf{G}_{n-1} \sim \nu_{[n-1]}} \mathbf{E}_{\mathbf{S} \sim \mu_n} \Delta_{G_{n-1}}(\mathbf{S})^2 \cdot \mathbf{1}(\bar{\mathcal{E}}) \end{aligned}$$

The inequality follows because  $1 + x \leq e^x$  for all  $x$ , and the second equality follows from a change in the randomness of  $\mathbf{S}$ , and because  $(1 + \Delta_{G_{n-1}}(\mathbf{S})) = \frac{v_n(\mathbf{S} | \mathbf{G}_{n-1})}{\mu_n(\mathbf{S})}$ .

By the definition of  $\Delta_{G_{n-1}}$ , we can further simplify the expectation of  $\Delta_{G_{n-1}}(\mathbf{S}) \cdot \mathbf{1}(\bar{\mathcal{E}})$ :

$$\begin{aligned} \left| \mathbf{E}_{\mathbf{G}_{n-1} \sim \nu_{[n-1]}} \mathbf{E}_{\mathbf{S} \sim \mu_n} \Delta_{G_{n-1}}(\mathbf{S}) \cdot \mathbf{1}(\bar{\mathcal{E}}) \right| &= \left| \mathbf{Pr}_{\mathbf{G}_{n-1} \sim \text{Geo}_d(n-1,p), \mathbf{S} \sim \mu_n}[\bar{\mathcal{E}}] - \mathbf{Pr}_{\mathbf{G}_{n-1} \sim \text{Geo}_d(n-1,p), \mathbf{S} \sim \nu_n(\cdot | \mathbf{G}_{n-1})}[\mathcal{E}] \right| \\ &\leq \mathbf{Pr}_{\mathbf{G}_{n-1} \sim \text{Geo}_d(n-1,p), \mathbf{S} \sim \mu_n}[\mathcal{E}] + \mathbf{Pr}_{\mathbf{G}_{n-1} \sim \text{Geo}_d(n-1,p), \mathbf{S} \sim \nu_n(\cdot | \mathbf{G}_{n-1})}[\mathcal{E}] \leq o\left(\frac{1}{n^2 \log n}\right) \end{aligned}$$

which completes the proof.  $\square$

**Remark 8.4.** We will ultimately choose the event  $\mathcal{E}$  based on when the Chi-square estimate is too loose of an upper bound on the relative entropy. If the conditions of [Lemma 8.3](#) hold, the overall TV bound we want would follow from

$$\mathbf{E}_{\mathbf{G}_{n-1} \sim \text{Geo}_d(n-1,p)} \mathbf{E}_{\mathbf{S} \sim \text{Binom}(n-1,p)} \left[ \left( \frac{v_n(\mathbf{S} | \mathbf{G}_{n-1})}{p^{|\mathbf{S}|} (1-p)^{n-1-|\mathbf{S}|}} - 1 \right)^2 \cdot \mathbf{1}(\bar{\mathcal{E}}) \right] = o\left(\frac{1}{n}\right). \quad (20)$$

### 8.1.1 The general $p$ case

The goal of this section is to prove [Theorem 1.2](#).

**Theorem** (Restatement of [Theorem 1.2](#)). *For any fixed constant  $\alpha > 0$ , if  $\frac{\alpha}{n} < p < \frac{1}{2}$  and  $d = \tilde{\Omega}(p^2 n^3)$ ,*

$$\lim_{n \rightarrow \infty} d_{\text{TV}}(\text{Geo}_d(n, p), \text{G}(n, p)) = 0.$$

*Proof.* We apply [Lemma 8.3](#), when  $\mathcal{E}$  is the failure of degree concentration:  $|\mathbf{S}| \geq pn + \Delta$ , with  $\Delta = 10 \max(\log n, pn)$ . Via [Lemma 3.11](#),  $\mathcal{E}$  has probability  $O(n^{-3})$ , regardless of whether  $\mathbf{S} \sim \mu_n$  or  $\mathbf{S} \sim \nu_n(\cdot | \mathbf{G}_{n-1})$  and  $\mathbf{G}_{n-1} \sim \text{Geo}_d(n-1, p)$ . (The latter distribution is equivalent to  $\mathbf{G}_n \sim \text{Geo}$ .) We thus satisfy:

$$\mathbf{Pr}_{\mathbf{G}_{n-1} \sim \text{Geo}_d(n-1,p), \mathbf{S} \sim \mu_n}[\mathcal{E}], \mathbf{Pr}_{\mathbf{G}_{n-1} \sim \text{Geo}_d(n-1,p), \mathbf{S} \sim \nu_n(\cdot | \mathbf{G}_{n-1})}[\mathcal{E}] \leq o\left(\frac{1}{n^2 \log n}\right)$$

After the reduction to (20) in Remark 8.4, we complete the proof by bounding:

$$\begin{aligned} & \mathbf{E}_{\mathbf{G}_{n-1} \sim \nu_{[n-1]}} \mathbf{E}_{\mathbf{S} \sim \text{Binom}(n-1, p)} \left[ \left( \frac{\nu_n(\mathbf{S} | \mathbf{G}_{n-1})}{p^{|\mathbf{S}|}(1-p)^{n-1-|\mathbf{S}|}} - 1 \right)^2 \cdot \mathbf{1}(\overline{\mathcal{E}}) \right] \\ & \leq \int_0^\infty \mathbf{Pr}_{\mathbf{G}_{n-1} \sim \nu_{[n-1]}, \mathbf{S} \sim \text{Binom}(n-1, p)} \left[ \left( \frac{\nu_n(\mathbf{S} | \mathbf{G}_{n-1})}{p^{|\mathbf{S}|}(1-p)^{n-1-|\mathbf{S}|}} - 1 \right)^2 > t \mid \overline{\mathcal{E}} \right] dt \end{aligned}$$

We now apply Corollary 6.1 to control these tail probabilities. By conditioning on  $\overline{\mathcal{E}}$ , we may assume  $\Delta = 10 \max(pn, \log n)$  in the tail bound, as this choice of  $\Delta$  leads to the worst case tail probability for all  $|\mathbf{S}| \leq pn + \Delta$ . Recall that  $M(n, p, \Delta) = \max(nH(p), |\Delta| \log \frac{1}{p}, \log n)$  in Corollary 6.1, and note that  $p^{|\mathbf{S}|}(1-p)^{n-1-|\mathbf{S}|} = e^{-nH(p)} \cdot \left(\frac{p}{1-p}\right)^\Delta$ .

$$\begin{aligned} \int_0^\infty \mathbf{Pr} \left[ \left( \frac{\nu_n(\mathbf{S} | \mathbf{G}_{n-1})}{p^{|\mathbf{S}|}(1-p)^{n-1-|\mathbf{S}|}} - 1 \right)^2 > t \mid \overline{\mathcal{E}} \right] dt & \leq \int_0^{\frac{1}{n \log n}} 1 \cdot dt + \int_{\frac{1}{n \log n}}^1 \exp\left(\frac{-C_{6.1}d \cdot t}{M(n, p, \Delta)^2 \log n}\right) dt \\ & \quad + \int_1^{e^{2nH(p)} \cdot \left(\frac{p}{1-p}\right)^\Delta} \exp\left(\frac{-C_{6.1}d}{M(n, p, \Delta)^2 \log n} + C'_{6.1} \log n\right) dt \\ & \leq \frac{1}{n \log n} - \frac{M(n, p, \Delta)^2 \log n}{C_{6.1}d} \exp\left(\frac{-C_{6.1}d \cdot t}{M(n, p, \Delta)^2 \log n}\right) \Big|_{\frac{1}{n \log n}} \\ & \quad + e^{2nH(p)} \cdot \left(\frac{p}{1-p}\right)^\Delta \cdot \exp\left(\frac{-C_{6.1}d}{M(n, p, \Delta)^2 \log n} + C'_{6.1} \log n\right) \end{aligned}$$

In the first line, we split the integral based on the appropriate tail bound expression in Corollary 6.1, and also remark that it suffices to consider  $t \leq e^{2nH(p)} \cdot \left(\frac{p}{1-p}\right)^\Delta$ , as  $\frac{\nu_n(\mathbf{S} | \mathbf{G}_{n-1})}{p^{|\mathbf{S}|}(1-p)^{n-1-|\mathbf{S}|}}$  is maximized at that value when  $|\mathbf{S}| \leq np + \Delta$ . If we choose  $d \geq n \cdot M(n, p, \Delta)^2 \cdot \log^3 n$ , we can bound each summation term individually to obtain the desired bound on  $\mathbf{E}_{\mathbf{G}_{n-1} \sim \nu_{[n-1]}} \mathbf{E}_{\mathbf{S} \sim \text{Binom}(n-1, p)} \left[ \left( \frac{\nu_n(\mathbf{S} | \mathbf{G}_{n-1})}{p^{|\mathbf{S}|}(1-p)^{n-1-|\mathbf{S}|}} - 1 \right)^2 \cdot \mathbf{1}(\overline{\mathcal{E}}) \right]$ .

$$\int_0^\infty \mathbf{Pr} \left[ \left( \frac{\nu_n(\mathbf{S} | \mathbf{G}_{n-1})}{p^{|\mathbf{S}|}(1-p)^{n-1-|\mathbf{S}|}} - 1 \right)^2 > t \mid \overline{\mathcal{E}} \right] dt \leq \frac{1}{n \log n} + \frac{1}{n} \cdot e^{-\log n} + e^{2nH(p)} \cdot e^{-Cn \log^2 n} \leq o\left(\frac{1}{n}\right)$$

We note that  $M(n, p, \Delta) \leq \max(np \log n, \log^2 n)$  for  $\frac{1}{n} \leq p \leq \frac{1}{2}$ , so we can restate our requirement on  $d$  as  $d = \tilde{\Omega}(n^3 p^2)$ .  $\square$

### 8.1.2 The sparse case

For the rest of this section, we assume  $p = \frac{\alpha}{n}$ , where  $\alpha$  is a constant. The goal of this section is to prove Theorem 1.1, reproduced below:

**Theorem** (Restatement of Theorem 1.1). *For any fixed constant  $\alpha \geq 1$ , if  $d = \Omega(\log^{36} n)$ , then*

$$\lim_{n \rightarrow \infty} d_{\text{TV}}(\text{Geo}_d(n, \frac{\alpha}{n}), G(n, \frac{\alpha}{n})) = 0.$$

*Proof.* As with the general case, we apply Lemma 8.3, but let  $\mathcal{E}$  be the union of the following events occurring for  $\mathbf{G} \sim \text{Geo}_d(n-1, p)$ :

- Failure of degree concentration:  $|\mathbf{S}| \geq 10 \log^2 n$ .

- Failure of [Proposition 1.3](#): If we sample  $\mathbf{G}_- \sim G(n-1, (1-\varepsilon)p)$ , and  $\mathbf{G}_+ \sim G(n-1, (1+\varepsilon)p)$ , corresponding to  $\mathbf{G}_{n-1}$  via the three-way coupling promised by [Proposition 1.3](#) with  $\varepsilon > \sqrt{\frac{n \cdot H(p)}{d}}$ , the graphs do not satisfy  $\mathbf{G}_- \subseteq \mathbf{G}_{n-1} \subseteq \mathbf{G}_+$ .

- Failure of [Lemma 7.1](#): There exists an  $S$  such that  $|S| \leq \log^2 n$  for which the neighborhood probability

$$\frac{v_n(S | \mathbf{G}_{n-1})}{p^{|S|}(1-p)^{n-1-|S|}} \notin (1 \pm \eta(\ell_{\mathbf{G}_{n-1}, S})).$$

Here, we can define  $\eta(\ell) := \min \left\{ \frac{\log^8 n}{\sqrt{d}}, \max \left\{ \left( \frac{\log^{29} n}{\sqrt{d}} \right)^\ell, \frac{\log^{12} n}{\sqrt{nd}} \right\} \right\}$ , and

$$\ell_{G,S} := \min \left\{ \frac{\log n}{\log \log n}, \arg \max_{\ell} \left\{ \bigcup_{i \in S} B_G(i, \ell) \text{ form } |S| \text{ disjoint trees in } G \right\} \right\}$$

Via [Lemma 3.11](#), the first event has probability  $n^{-\Omega(\log n)}$ , regardless of whether we are working over  $\mathbf{G}_{n-1} \sim \text{Geo}_d(n-1, p), S \sim \mu_n$  or  $\mathbf{G}_{n-1} \sim \text{Geo}_d(n-1, p), S \sim v_n(\cdot | \mathbf{G}_{n-1})$ .

The second and third events also exclusively deal with  $\mathbf{G}_{n-1}$ . The probability of the second event is  $n^{-\Omega(\log n)}$  by [Proposition 1.3](#). The third event also has probability at most  $n^{-\Omega(\log n)}$  over  $\mathbf{G}_{n-1} \sim \text{Geo}_d(n-1, p)$  by [Lemma 7.1](#). We thus satisfy the conditions of [Lemma 8.3](#):

$$\Pr_{\mathbf{G}_{n-1} \sim \text{Geo}_d(n-1, p), S \sim \mu_n} [\mathcal{E}], \Pr_{\mathbf{G}_{n-1} \sim \text{Geo}_d(n-1, p), S \sim v_n(\cdot | \mathbf{G}_{n-1})} [\mathcal{E}] \leq o \left( \frac{1}{n^2 \log n} \right)$$

It suffices to show that the following quantity is at most  $o\left(\frac{1}{n}\right)$ :

$$\mathbf{E}_{\mathbf{G}_{n-1} \sim \text{Geo}_d(n-1, p), S \sim \text{Binom}(n-1, p)} \left[ \left( \frac{v_n(\mathbf{S} | \mathbf{G}_{n-1})}{p^{|\mathbf{S}|}(1-p)^{n-1-|\mathbf{S}|}} - 1 \right)^2 \cdot \mathbf{1}[\overline{\mathcal{E}}] \right].$$

Due to the  $\mathbf{1}[\overline{\mathcal{E}}]$  term, we can simplify the expression inside the expectation while assuming the outcome of [Lemma 7.1](#) holds, so this expression is bounded by

$$\mathbf{E}_{\mathbf{G}_{n-1} \sim \text{Geo}_d(n-1, p), S \sim \text{Binom}(n-1, p)} \left[ \eta(\ell_{\mathbf{G}_{n-1}, \mathbf{S}})^2 \cdot \mathbf{1}[\overline{\mathcal{E}}] \right]$$

We can think of  $\ell_{\mathbf{G}_{n-1}, \mathbf{S}}$  itself as a random variable over  $\{0, \dots, \lceil \frac{\log n}{\log \log n} \rceil\}$ , whose randomness is governed by  $\mathbf{G}_{n-1}$ . Wishfully, if  $\ell_{\mathbf{G}_{n-1}, \mathbf{S}} \geq \Omega\left(\frac{\log n}{\log \log n}\right)$  always held true, then  $\eta(\ell_{\mathbf{G}_{n-1}, \mathbf{S}})^2 \leq O\left(\frac{\log^{24} n}{nd}\right)$ . Then, as desired,  $\mathbf{E}_{\mathbf{G}_{n-1} \sim \text{Geo}_d(n-1, p)} \mathbf{E}_{S \sim \text{Binom}(n-1, p)} \left[ \eta(\ell_{\mathbf{G}_{n-1}, \mathbf{S}})^2 \cdot \mathbf{1}[\overline{\mathcal{E}}] \right] \leq o\left(\frac{1}{n}\right)$  when  $d \geq \Omega(\log^{30} n)$ .

The reality is: with some (small) probability,  $\ell_{\mathbf{G}_{n-1}, \mathbf{S}}$  is small, so  $\eta(\ell_{\mathbf{G}_{n-1}, \mathbf{S}})^2$  could be as large as  $\frac{\log^{16} n}{d}$ . We need to upper bound the probability that  $\ell_{\mathbf{G}_{n-1}, \mathbf{S}}$  is small, so that the contribution of this event to  $\mathbf{E}_{\mathbf{G}_{n-1} \sim \text{Geo}_d(n-1, p)} \mathbf{E}_{S \sim \text{Binom}(n-1, p)} \left[ \eta(\ell_{\mathbf{G}_{n-1}, \mathbf{S}})^2 \cdot \mathbf{1}[\overline{\mathcal{E}}] \right]$  remains at most  $o\left(\frac{1}{n}\right)$ .

But in a sparse Erdős-Rényi graph  $\mathbf{G}_+ \sim G(n, (1+\varepsilon)p)$ , it is well-known that  $\ell_{\mathbf{G}_+, \mathbf{S}} \geq \Omega\left(\frac{\log n}{\log \log n}\right)$  with high probability when  $\mathbf{S} \sim \text{Binom}(n, p)$ , and further there exist known bounds on the probability that  $\ell_{\mathbf{G}_+, \mathbf{S}}$  is small. We can now use the success of the coupling between  $\mathbf{G}_{n-1}$  and  $\mathbf{G}_+$ , guaranteed by  $\overline{\mathcal{E}}$ , to argue that the same bounds hold over  $\mathbf{S}$  and  $\mathbf{G}_{n-1}$  with high probability. After we make this relationship between  $\ell_{\mathbf{G}_{n-1}, \mathbf{S}}$  and  $\ell_{\mathbf{G}_+, \mathbf{S}}$  explicit below, the remainder of the proof is accounting for the events when  $\ell_{\mathbf{G}_+, \mathbf{S}}$  is small.

Before diving into the distribution of  $\ell_{\mathbf{G}_+, \mathbf{S}}$ , we can first simplify

$$\begin{aligned}
& \mathbf{E}_{\mathbf{G}_{n-1} \sim \text{Geo}_d(n-1, p)} \mathbf{E}_{\mathbf{S} \sim \text{Binom}(n-1, p)} \left[ \eta(\ell_{\mathbf{G}_{n-1}, \mathbf{S}})^2 \cdot \mathbf{1}[\overline{\mathcal{E}}] \right] \\
& \leq \mathbf{E}_{\mathbf{G}_{n-1} \sim \text{Geo}_d(n-1, p)} \mathbf{E}_{\mathbf{S} \sim \text{Binom}(n-1, p)} \left[ \min \left( \frac{\log^{16} n}{d}, \max \left( \left( \frac{\log^{29} n}{d} \right)^{\ell_{\mathbf{G}_{n-1}, \mathbf{S}}}, \frac{\log^{24} n}{nd} \right) \right) \cdot \mathbf{1}[\overline{\mathcal{E}}] \right] \\
& \leq \mathbf{E}_{\mathbf{G}_+ \sim \text{G}(n-1, (1+\varepsilon)p)} \mathbf{E}_{\mathbf{S} \sim \text{Binom}(n-1, p)} \left[ \min \left( \frac{\log^{16} n}{d}, \max \left( \left( \frac{\log^{29} n}{d} \right)^{\ell_{\mathbf{G}_+, \mathbf{S}}}, \frac{\log^{24} n}{nd} \right) \right) \cdot \mathbf{1}[\overline{\mathcal{E}}] \right] \\
& \leq \mathbf{E}_{\mathbf{S} \sim \text{Binom}(n-1, p)} \mathbf{E}_{\mathbf{G}_+ \sim \text{G}(n-1, (1+\varepsilon)p)} \left[ \min \left( \frac{\log^{16} n}{d}, \max \left( \left( \frac{\log^{29} n}{d} \right)^{\ell_{\mathbf{G}_+, \mathbf{S}}}, \frac{\log^{24} n}{nd} \right) \right) \cdot \mathbf{1}[\overline{\mathcal{E}}] \right]
\end{aligned}$$

The second-to-last line relies on the success of the coupling from [Proposition 1.3](#), with  $\varepsilon > \sqrt{\frac{n \cdot H(p)}{d}}$ , and noting that when  $\mathbf{G}_{n-1} \subseteq \mathbf{G}_+$ , we have  $\ell_{\mathbf{G}_+, \mathbf{S}} \leq \ell_{\mathbf{G}_{n-1}, \mathbf{S}}$ . The last line involves switching the order of the expectations.

Again, as the bound need only apply when  $\overline{\mathcal{E}}$  occurs, we may assume  $|\mathbf{S}| \leq \log^2 n$ . The theorem now follows directly from the claim below.

**Claim.** For all  $|\mathbf{S}| \leq \log^2 n$ ,  $\mathbf{E}_{\mathbf{G}_+ \sim \text{G}(n-1, (1+\varepsilon)p)} \left[ \min \left( \frac{\log^{16} n}{d}, \max \left( \left( \frac{\log^{29} n}{d} \right)^{\ell_{\mathbf{G}_+, \mathbf{S}}}, \frac{\log^{24} n}{nd} \right) \right) \right] \leq o\left(\frac{1}{n}\right)$ .

*Proof of Claim.* To understand the magnitude of  $d^{\ell_{\mathbf{G}_{n-1}, \mathbf{S}}}$ , we need to case on different depths  $\ell$ .

$$\begin{aligned}
& \mathbf{E}_{\mathbf{G}_+ \sim \text{G}(n-1, (1+\varepsilon)p)} \left[ \min \left( \frac{\log^{16} n}{d}, \max \left( \left( \frac{\log^{29} n}{d} \right)^{\ell_{\mathbf{G}_+, \mathbf{S}}}, \frac{\log^{24} n}{nd} \right) \right) \right] = \mathbf{Pr}_{\mathbf{G}_+ \sim \text{G}(n-1, (1+\varepsilon)p)} [\ell_{\mathbf{G}_+, \mathbf{S}} > \log_d(nd)] \cdot \frac{\log^{16} n}{nd} \\
& \quad + \mathbf{Pr}_{\mathbf{G}_+ \sim \text{G}(n-1, (1+\varepsilon)p)} [\ell_{\mathbf{G}_+, \mathbf{S}} = 0] \cdot \frac{\log^{24} n}{d} + \sum_{\ell=1}^{\log_d(nd)} \mathbf{Pr}_{\mathbf{G}_+ \sim \text{G}(n-1, (1+\varepsilon)p)} [\ell_{\mathbf{G}_+, \mathbf{S}} = \ell] \cdot \left( \frac{\log^{29} n}{d} \right)^\ell \\
& \leq \frac{\log^{16} n}{nd} + \mathbf{Pr}_{\mathbf{G}_+ \sim \text{G}(n-1, (1+\varepsilon)p)} [\ell_{\mathbf{G}_+, \mathbf{S}} = 0] \cdot \frac{\log^{24} n}{d} + \sum_{\ell=1}^{\log_d(nd)} \mathbf{Pr}_{\mathbf{G}_+ \sim \text{G}(n-1, (1+\varepsilon)p)} [\ell_{\mathbf{G}_+, \mathbf{S}} \leq \ell] \cdot \left( \frac{\log^{29} n}{d} \right)^\ell
\end{aligned}$$

The  $\frac{\log^{16} n}{nd}$  term is at most  $o\left(\frac{1}{n}\right)$  when  $d \geq \Omega(\log^{35} n)$ . To understand the event  $\ell_{\mathbf{G}_{n-1}, \mathbf{S}} \leq \ell$ , we compute the probability of  $\bigcup_{i \in \mathbf{S}} B_{\mathbf{G}_{n-1}}(i, \ell+1)$  does not form  $|\mathbf{S}|$  disjoint trees. This event is contained in the union of the following events:

$\mathcal{I}(\ell)$ : There exist  $i, j \in \mathbf{S}$  that are distance  $\leq 2\ell+2$  apart, so  $B_{\mathbf{G}_{n-1}}(i, \ell+1)$  and  $B_{\mathbf{G}_{n-1}}(j, \ell+1)$  **intersect**.

$\mathcal{C}(\ell)$ : For some  $i \in \mathbf{S}$ , the ball  $B_{\mathbf{G}_{n-1}}(i, \ell+1)$  contains a **cycle**.

Applying the union bound to events  $\mathcal{I}$  and  $\mathcal{C}$ , we can bound:

$$\begin{aligned}
& \mathbf{Pr}_{\mathbf{G}_+ \sim \text{G}(n-1, (1+\varepsilon)p)} [\ell_{\mathbf{G}_+, \mathbf{S}} = 0] \cdot \frac{\log^{24} n}{d} \leq \mathbf{Pr}_{\mathbf{G}_+ \sim \text{G}(n-1, (1+\varepsilon)p)} [\mathcal{I}(0)] \cdot \frac{\log^{24} n}{d} + \mathbf{Pr}_{\mathbf{G}_+ \sim \text{G}(n-1, (1+\varepsilon)p)} [\mathcal{C}(0)] \cdot \frac{\log^{24} n}{d} \\
& \sum_{\ell=1}^{\log_d(nd)} \mathbf{Pr}_{\mathbf{G}_+ \sim \text{G}(n-1, (1+\varepsilon)p)} [\ell_{\mathbf{G}_+, \mathbf{S}} = \ell] \cdot \left( \frac{\log^{29} n}{d} \right)^\ell \leq \sum_{\ell=1}^{\log_d(nd)} \mathbf{Pr}_{\mathbf{G}_+ \sim \text{G}(n-1, (1+\varepsilon)p)} [\mathcal{I}(\ell)] \cdot \left( \frac{\log^{29} n}{d} \right)^\ell \\
& \quad + \sum_{\ell=1}^{\log_d(nd)} \mathbf{Pr}_{\mathbf{G}_+ \sim \text{G}(n-1, (1+\varepsilon)p)} [\mathcal{C}(\ell)] \cdot \left( \frac{\log^{29} n}{d} \right)^\ell
\end{aligned}$$

**Contribution of event  $\mathcal{I}$ :** To analyze event  $\mathcal{I}$ , we consider its complement  $\overline{\mathcal{I}}$ . We lower bound the probability of  $\overline{\mathcal{I}}$  by counting how many ways we can choose the vertices  $[i]$  via this greedy process: we first choose  $v \in V$ , and then set  $V := V \setminus B_{\mathcal{G}_{n-1}}(v, 2\ell + 2)$  before choosing the next vertex. Using [Lemma 3.12](#) and a union bound over all  $|S|$  vertices, with probability  $\geq 1 - \frac{1}{n^2}$ , the size of each  $B_{\mathcal{G}_{n-1}}(v, 2\ell + 2)$  is at most  $c \log^3 n (pn)^{2\ell+2}$ , for a universal constant  $c$ . This greedy process ensures that all vertices chosen for  $S$  are distance  $\geq 2\ell + 2$  apart.

$$\begin{aligned} \Pr_{\mathcal{G}_+ \sim \mathcal{G}(n-1, (1+\varepsilon)p)} [\overline{\mathcal{I}(\ell)}] &\geq \frac{n \cdot (n - c \log^3 n (pn)^{2\ell+2}) \cdots (n - (|S| - 1)c \log^3 n (pn)^{2\ell+2})}{n^{|S|}} \cdot \left(1 - \frac{1}{n^2}\right) \\ &\geq \left[1 - \frac{c \log^3 n |S| (pn)^{2\ell}}{n}\right]^{|S|} - \frac{1}{n^2} \geq 1 - \frac{c \log^3 n |S|^2 (pn)^{2\ell+2}}{n} - \frac{1}{n^2} \end{aligned}$$

First, we consider what happens when  $\ell_{\mathcal{G}_+, S} = 0$ . (Recall  $p = \frac{\alpha}{n}$ .)

$$\Pr_{\mathcal{G}_+ \sim \mathcal{G}(n-1, (1+\varepsilon)p)} [\mathcal{I}(0)] \cdot \frac{\log^{24} n}{d} \leq \left(\frac{c \log^3 n |S|^2 (pn)^2}{n} + \frac{1}{n^2}\right) \cdot \frac{\log^{24} n}{d} \leq \frac{5c\alpha^2 \log^{29} n}{nd}.$$

When  $d \geq \Omega(\log^{36} n)$ , this quantity is at most  $o(\frac{1}{n})$ . Next, we consider what happens when  $\ell_{\mathcal{G}_+, S} > 0$ .

$$\begin{aligned} \sum_{\ell=1}^{\log_d(nd)} \Pr_{\mathcal{G}_+ \sim \mathcal{G}(n-1, (1+\varepsilon)p)} [\mathcal{I}(\ell)] \cdot \left(\frac{\log^{29} n}{d}\right)^\ell &\leq \sum_{\ell=1}^{\log_d(nd)} \left(\frac{c \log^3 n |S|^2 (pn)^{2\ell+2}}{n} + \frac{1}{n^2}\right) \cdot \left(\frac{\log^{29} n}{d}\right)^\ell \\ &\leq 2 \left(\sum_{\ell=1}^{\log_d(nd)} \frac{c(\log^{29\ell+3} n) |S|^2 (pn)^{2\ell+2}}{n \cdot d^\ell}\right) \end{aligned}$$

The summand when  $\ell = 1$ ,  $\frac{c(\log^{32} n) |S|^2 \alpha^4}{nd}$  has the largest magnitude, if  $\frac{\alpha^2 \log^{29} n}{d} \leq 1$ . Thus, we can upper bound the summation by  $4c\alpha^4 \cdot \frac{\log^{33} n \log_d(nd)}{n \log^{36} n} \leq o(\frac{1}{n})$ , as desired for the claim.

**Contribution of event  $\mathcal{C}$ :** Let the event  $\mathcal{C}(\ell)_i$  denote the event that  $B_{\mathcal{G}_{n-1}}(i, \ell + 1)$  contains a cycle for  $i \in S$ . Using a union bound, we have  $\Pr_{\mathcal{G}_+ \sim \mathcal{G}(n-1, (1+\varepsilon)p)} [\mathcal{C}(\ell)] \leq |S| \cdot \Pr_{\mathcal{G}_+ \sim \mathcal{G}(n-1, (1+\varepsilon)p)} [\mathcal{C}(\ell)_i]$ . Now, using [Lemma 3.13](#), there exists a universal constant  $c'$  such that

$$|S| \cdot \Pr_{\mathcal{G}_{n-1} \sim \mathcal{G}(n-1, (1+\varepsilon)p)} [\mathcal{C}_j \text{ for } \ell] \leq |S| \cdot \frac{c'(pn)^{\ell+1}}{n}$$

First, we consider the contribution of the event  $\ell_{\mathcal{G}_+, S} = 0$ . (Recall  $p = \frac{\alpha}{n}$ .)

$$\Pr_{\mathcal{G}_+ \sim \mathcal{G}(n-1, (1+\varepsilon)p)} [\mathcal{C}(\ell)] \cdot \frac{\log^{24} n}{d} \leq \frac{c'(\log^{24} n) |S| (pn)}{nd} \leq \frac{2c'\alpha \log^{25} n}{nd} \leq o\left(\frac{1}{n}\right)$$

When  $d \geq \Omega(\log^{35} n)$ , this quantity is at most  $o(\frac{1}{n})$ . We lastly consider the events when  $\ell_{\mathcal{G}_+, S} > 0$ .

$$\sum_{\ell=1}^{\log_d(nd)} \Pr_{\mathcal{G}_+ \sim \mathcal{G}(n-1, (1+\varepsilon)p)} [\mathcal{C}(\ell)] \cdot \left(\frac{\log^{29} n}{d}\right)^\ell \leq \sum_{\ell=1}^{\log_d(nd)} \frac{c' \log^{29\ell} |S| (pn)^{\ell+1}}{nd^\ell}$$

Again, the term for  $\ell = 1$ ,  $\frac{c'(\log^{29} n) |S| (pn)^2}{d}$ , is the largest term of the summation, if  $\frac{\alpha \log^{29} n}{d} \leq 1$ . The summation is thus upper bounded by  $2c'\alpha^2 \cdot \frac{\log^{31} n \log_d(nd)}{n \log^{36} n} \leq o(\frac{1}{n})$  as well.  $\square$

## Acknowledgments

Much of this work was conducted when we were participating in the Simons Institute Fall 2020 program on “Probability, Geometry, and Computation in High Dimensions”. We thank the Simons Institute for their support, and the program for inspiration. S.M. would like to thank Abhishek Shetty for inspiring conversations on geometric concentration.

## References

- [AS04] Noga Alon and Joel H Spencer. *The probabilistic method*. John Wiley & Sons, 2004. [1](#)
- [BBH21] Matthew Brennan, Guy Bresler, and Brice Huang. De Finetti-style results for Wishart matrices: Combinatorial structure and phase transitions. *arXiv preprint arXiv:2103.14011*, 2021. [6](#)
- [BBN20] Matthew Brennan, Guy Bresler, and Dheeraj Nagaraj. Phase transitions for detecting latent geometry in random graphs. *Probability Theory and Related Fields*, 178(3):1215–1289, 2020. [2](#), [3](#), [4](#), [6](#), [41](#), [54](#)
- [BCC<sup>+</sup>10] Aditya Bhaskara, Moses Charikar, Eden Chlamtac, Uriel Feige, and Aravindan Vijayaraghavan. Detecting high log-densities: an  $o(n^{1/4})$  approximation for densest  $k$ -subgraph. In *Proceedings of the forty-second ACM symposium on Theory of computing*, pages 201–210, 2010. [1](#)
- [BDER16] Sébastien Bubeck, Jian Ding, Ronen Eldan, and Miklós Z Rácz. Testing for high-dimensional geometry in random graphs. *Random Structures & Algorithms*, 49(3):503–532, 2016. [2](#), [3](#), [4](#), [6](#), [51](#), [52](#)
- [BG18] Sébastien Bubeck and Shirshendu Ganguly. Entropic CLT and phase transition in high-dimensional Wishart matrices. *International Mathematics Research Notices*, 2018(2):588–606, 2018. [5](#)
- [BGK<sup>+</sup>01] Andreas Brieden, Peter Gritzmann, Ravindran Kannan, Victor Klee, László Lovász, and Miklós Simonovits. Deterministic and randomized polynomial-time approximation of radii. *Mathematika*, 48(1-2):63–105, 2001. [53](#)
- [BGL13] Dominique Bakry, Ivan Gentil, and Michel Ledoux. *Analysis and geometry of Markov diffusion operators*, volume 348. Springer Science & Business Media, 2013. [13](#)
- [BLM15] Charles Bordenave, Marc Lelarge, and Laurent Massoulié. Non-backtracking spectrum of random graphs: community detection and non-regular ramanujan graphs. In *2015 IEEE 56th Annual Symposium on Foundations of Computer Science*, pages 1347–1357. IEEE, 2015. [11](#)
- [CK16] Daniel Cullina and Negar Kiyavash. Improved achievability and converse bounds for erdos-rényi graph matching. *ACM SIGMETRICS Performance Evaluation Review*, 44(1):63–72, 2016. [1](#)

- [COKPZ18] Amin Coja-Oghlan, Florent Krzakala, Will Perkins, and Lenka Zdeborová. Information-theoretic thresholds from the cavity method. *Advances in Mathematics*, 333:694–795, 2018. [5](#)
- [DEKL14] Jean Dolbeault, Maria J Esteban, Michal Kowalczyk, and Michael Loss. Sharp interpolation inequalities on the sphere: new methods and consequences. In *Partial Differential Equations: Theory, Control and Approximation*, pages 225–242. Springer, 2014. [13](#)
- [DGLU11] Luc Devroye, András György, Gábor Lugosi, and Frederic Udina. High-dimensional random geometric graphs and their clique number. *Electronic Journal of Probability*, 16:2481–2508, 2011. [1](#), [2](#), [4](#)
- [DMSS14] Amir Dembo, Andrea Montanari, Allan Sly, and Nike Sun. The replica symmetric solution for potts models on d-regular graphs. *Communications in Mathematical Physics*, 327(2):551–575, 2014. [5](#)
- [DSVW04] Martin Dyer, Alistair Sinclair, Eric Vigoda, and Dror Weitz. Mixing in time and space for lattice spin systems: A combinatorial view. *Random Structures & Algorithms*, 24(4):461–479, 2004. [5](#)
- [EM20] Ronen Eldan and Dan Mikulincer. Information and dimensionality of anisotropic random geometric graphs. In *Geometric Aspects of Functional Analysis*, pages 273–324. Springer, 2020. [6](#)
- [ER59] Paul Erdős and Alfréd Rényi. On random graphs. I. *Publicationes Mathematicae*, 1959. [1](#)
- [Fei02] Uriel Feige. Relations between average case complexity and approximation complexity. In *Proceedings of the thirty-fourth annual ACM symposium on Theory of computing*, pages 534–543, 2002. [1](#)
- [FK01] Uriel Feige and Joe Kilian. Heuristics for semirandom graph problems. *Journal of Computer and System Sciences*, 63(4):639–671, 2001. [1](#)
- [FLS04] Uriel Feige, Michael Langberg, and Gideon Schechtman. Graphs with tiny vector chromatic numbers and huge chromatic numbers. *SIAM Journal on Computing*, 33(6):1338–1368, 2004. [5](#)
- [FS02] Uriel Feige and Gideon Schechtman. On the optimality of the random hyperplane rounding technique for MAX CUT. *Random Structures & Algorithms*, 20(3):403–440, 2002. [5](#)
- [GKM15] David Gamarnik, Dmitriy Katz, and Sidhant Misra. Strong spatial mixing of list coloring of graphs. *Random Structures & Algorithms*, 46(4):599–613, 2015. [5](#)
- [Gri01] Dima Grigoriev. Linear lower bound on degrees of positivstellensatz calculus proofs for the parity. *Theoretical Computer Science*, 259(1-2):613–622, 2001. [1](#)
- [HLL83] Paul W Holland, Kathryn Blackmond Laskey, and Samuel Leinhardt. Stochastic block-models: First steps. *Social networks*, 5(2):109–137, 1983. [1](#)

- [Jer92] Mark Jerrum. Large cliques elude the Metropolis process. *Random Structures & Algorithms*, 3(4):347–359, 1992. [1](#)
- [JL15] Tiefeng Jiang and Danning Li. Approximation of rectangular beta-Laguerre ensembles and large deviations. *Journal of Theoretical Probability*, 28(3):804–847, 2015. [4](#)
- [KTW14] Subhash Khot, Madhur Tulsiani, and Pratik Worah. A characterization of strong approximation resistance. In *Proceedings of the forty-sixth annual ACM symposium on Theory of computing*, pages 634–643, 2014. [5](#)
- [LR21] Suqi Liu and Miklós Z Rácz. Phase transition in noisy high-dimensional random geometric graphs. *arXiv preprint arXiv:2103.15249*, 2021. [6](#)
- [Mas07] Pascal Massart. Concentration inequalities and model selection. 2007. [10](#)
- [Mat13] Jiri Matousek. *Lectures on discrete geometry*, volume 212. Springer Science & Business Media, 2013. [10](#)
- [MM09] Marc Mézard and Andrea Montanari. *Information, physics, and computation*. Oxford University Press, 2009. [12](#)
- [MP03] Marc Mézard and Giorgio Parisi. The cavity method at zero temperature. *Journal of Statistical Physics*, 111(1):1–34, 2003. [5](#)
- [Nis81] Hidetoshi Nishimori. Internal energy, specific heat and correlation function of the bond-random Ising model. *Progress of Theoretical Physics*, 66(4):1169–1181, 1981. [1](#)
- [Pan09] Dmitry Panchenko. Cavity method in the spherical sk model. In *Annales de l’IHP Probabilités et statistiques*, volume 45, pages 1020–1047, 2009. [5](#)
- [Pan13] Dmitry Panchenko. *The Sherrington-Kirkpatrick model*. Springer Science & Business Media, 2013. [1](#)
- [Pen03] Mathew Penrose. *Random geometric graphs*, volume 5. Oxford university press, 2003. [1](#), [3](#)
- [Ros08] Benjamin Rossman. On the constant-depth complexity of k-clique. In *Proceedings of the fortieth annual ACM symposium on Theory of computing*, pages 721–730, 2008. [1](#)
- [RR19] Miklós Z Rácz and Jacob Richey. A smooth transition from Wishart to GOE. *Journal of Theoretical Probability*, 32(2):898–906, 2019. [5](#)
- [Sch08] Grant Schoenebeck. Linear level Lasserre lower bounds for certain k-CSPs. In *2008 49th Annual IEEE Symposium on Foundations of Computer Science*, pages 593–602. IEEE, 2008. [1](#)
- [Tal96] Michel Talagrand. Transportation cost for Gaussian and other product measures. *Geometric & Functional Analysis GFA*, 6(3):587–600, 1996. [13](#)
- [Tal03] Michel Talagrand. *Spin glasses: a challenge for mathematicians: cavity and mean field models*, volume 46. Springer Science & Business Media, 2003. [5](#)

- [Ver18] Roman Vershynin. *High-dimensional probability: An introduction with applications in data science*, volume 47. Cambridge university press, 2018. 11
- [VH14] Ramon Van Handel. Probability in high dimension. Technical report, Princeton University, 2014. 10
- [Vil08] Cédric Villani. *Optimal transport: old and new*, volume 338. Springer Science & Business Media, 2008. 5, 13
- [Wal11] Mark Walters. Random geometric graphs. *Surveys in combinatorics*, 392:365–402, 2011. 3
- [Wei06] Dror Weitz. Counting independent sets up to the tree threshold. In *Proceedings of the thirty-eighth annual ACM symposium on Theory of computing*, pages 140–149, 2006. 5

## A Extending the analysis of the signed triangle statistic to all densities

Associate the graph  $G$  with the string  $G \in \{0, 1\}^{\binom{[n]}{2}}$  with  $G_{ij} = \mathbf{1}[i \sim j \text{ in } G]$ , and define the signed triangle count statistic

$$T(\mathbf{G}) = \sum_{i < j < k \in [n]} (G_{ij} - p)(G_{jk} - p)(G_{ik} - p).$$

In [BDER16], Theorem 2, the authors prove that for any fixed  $p \in (0, 1)$ ,  $T(\mathbf{G})$  distinguishes between  $\mathbf{G} \sim \text{Geo}_d(n, p)$  and  $\mathbf{G} \sim G(n, p)$  whenever  $d \ll n^3 = \Theta((nH(p))^3)$ . They also show in Theorem 3 that the *unsigned* triangle count is a distinguishing statistic for  $p = \Theta(\frac{1}{n})$  whenever  $d \ll \log^3 n = \Theta((nH(p))^3)$ .

Here we show that the analysis of  $T(\mathbf{G})$  from [BDER16] can be extended to show that the signed count  $T(\mathbf{G})$  distinguishes between  $G(n, p)$  and  $\text{Geo}_d(n, p)$  for most  $p$ , so long as  $d \ll (nH(p))^3$ .

**Lemma A.1.** *If  $\frac{1}{n^2} \ll p \leq 1 - \delta$  for any fixed constant  $\delta > 0$ , then so long as  $d \ll (nH(p))^3$ ,*

$$\left| \mathbf{E}_{\text{Geo}_d(n, p)} T(\mathbf{G}) - \mathbf{E}_{G(n, p)} T(\mathbf{G}) \right| \gg \max \left( \sqrt{\mathbf{Var}_{\text{Geo}_d(n, p)} T(\mathbf{G})}, \sqrt{\mathbf{Var}_{G(n, p)} T(\mathbf{G})} \right),$$

where  $H(p) = p \log \frac{1}{p} + (1 - p) \log \frac{1}{1-p}$  is the binary entropy function.

Utilizing the independence of edges in  $G(n, p)$ , it is easy to show that

$$\mathbf{E}_{G(n, p)} T(\mathbf{G}) = 0 \quad \text{and} \quad \mathbf{Var}_{G(n, p)} T(\mathbf{G}) = \binom{n}{3} p^3 (1-p)^3.$$

In [BDER16], in the course of proving their Lemma 1, the authors prove that for any  $p < 0.49$  there exists a universal constant  $C$  such that

$$\mathbf{E}_{\text{Geo}_d(n, p)} [T(\mathbf{G})] \geq C \binom{n}{3} p^3 \sqrt{\frac{(\log \frac{1}{p})^3}{d}}.$$

Hence, if  $d \ll (nH(p))^3$ , one can already see that the expectation of  $T(\mathbf{G})$  under  $\text{Geo}_d(n, p)$  overwhelms the standard deviation of  $T(\mathbf{G})$  under  $G(n, p)$ . All that remains is to bound the variance

under  $\text{Geo}_d(n, p)$ . For this, [BDER16] do not obtain sharp estimates for all  $p$ , obtaining simpler bounds for the special cases  $p = \Theta(1)$  and  $p = \Theta(\frac{1}{n})$ . En route, the authors prove the following claim:

**Claim A.2** (From the proof of Lemma 1 in [BDER16]). Let  $Q = \Pr_{\text{Geo}_d(n, p)}[\mathbf{G}_{12}\mathbf{G}_{13}\mathbf{G}_{23} = 1 \mid \mathbf{G}_{23} = 1]$ . There exists a universal constant  $c > 0$  such that for all  $p < 0.49$ ,  $Q = p^2(1 + \varepsilon)$  for  $\varepsilon \geq c\sqrt{\frac{\log^3 \frac{1}{p}}{d}}$ .

Here we will use this claim to extend the analysis of [BDER16] to all  $p \in (\frac{1}{n^2}, 0.49)$ .

*Proof of Lemma A.1.* What remains is to bound the variance of  $T(\mathbf{G})$  under  $\mathbf{G} \sim \text{Geo}_d(n, p)$ , so hereafter all expectations are taken with respect to  $\mathbf{G} \sim \text{Geo}_d(n, p)$ . We may also assume  $p < 0.49$ , as for constant  $p \in (0.49, 1)$  we may invoke Theorem 2 of [BDER16]. Let  $T_{ijk} = (G_{ij} - p)(G_{ik} - p)(G_{jk} - p)$  and  $\bar{T}_{ijk} = T_{ijk} - \mathbf{E}[T_{ijk}]$  for notational simplicity. We have that

$$\begin{aligned} \mathbf{Var} \left[ \sum_{i < j < k \in [n]} T_{ijk} \right] &= \mathbf{E} \left[ \sum_{i < j < k} \bar{T}_{ijk}^2 + 2 \sum_{i < j < k < \ell \in [n]} \bar{T}_{ijk} \bar{T}_{jkl} + \bar{T}_{ijl} \bar{T}_{jlk} + \bar{T}_{ikl} \bar{T}_{jkl} \right] \\ &= \binom{n}{3} \cdot \mathbf{E}[\bar{T}_{123}^2] + 6 \cdot \binom{n}{4} \cdot \mathbf{E}[\bar{T}_{123} \bar{T}_{234}], \end{aligned} \quad (21)$$

where the first line follows because if triangles  $i, j, k$  and  $a, b, c$  do not share an edge,  $T_{ijk}$  and  $T_{abc}$  are independent, and the second line follows from symmetry.

Because  $G_{ij}^2 = G_{ij}$ , we have that  $(G_{ij} - p)^2 = (1 - 2p)G_{ij} + p^2$ , and so

$$T_{123}^2 = ((1 - 2p)G_{12} + p^2) \cdot ((1 - 2p)G_{13} + p^2) \cdot ((1 - 2p)G_{23} + p^2), \quad (22)$$

$$T_{123}T_{234} = ((1 - 2p)G_{23} + p^2) \cdot (G_{12} - p) \cdot (G_{13} - p) \cdot (G_{24} - p) \cdot (G_{34} - p). \quad (23)$$

Now we will bound each of  $\mathbf{E}[\bar{T}_{123}^2]$  and  $\mathbf{E}[\bar{T}_{123} \bar{T}_{234}]$  separately. First, we bound  $\mathbf{E}[\bar{T}_{123}^2]$  by expanding (22) and utilizing the symmetry of  $\mathbf{G}_{12}, \mathbf{G}_{23}, \mathbf{G}_{13}$ ,

$$\begin{aligned} \mathbf{E}[\bar{T}_{123}^2] &= \mathbf{E}[T_{123}^2] - \mathbf{E}[T_{123}]^2 \\ &= (1 - 2p)^3 \mathbf{E}[\mathbf{G}_{12}\mathbf{G}_{13}\mathbf{G}_{23}] + 3(1 - 2p)^2 p^2 \mathbf{E}[\mathbf{G}_{12}\mathbf{G}_{13}] + 3(1 - 2p)p^4 \mathbf{E}[\mathbf{G}_{12}] + p^6 \\ &\quad - (\mathbf{E}[\mathbf{G}_{12}\mathbf{G}_{13}\mathbf{G}_{23}] - 3p \mathbf{E}[\mathbf{G}_{12}\mathbf{G}_{13}] + 3p^2 \mathbf{E}[\mathbf{G}_{12}] - p^3)^2 \\ &= (1 - 2p)^3 \mathbf{E}[\mathbf{G}_{12}\mathbf{G}_{13}\mathbf{G}_{23}] + 3(1 - 2p)^2 p^4 + 3(1 - 2p)p^5 + p^6 \\ &\quad - (\mathbf{E}[\mathbf{G}_{12}\mathbf{G}_{13}\mathbf{G}_{23}] - p^3)^2, \end{aligned}$$

where we have used that for any  $S \subseteq \binom{[n]}{2}$  which corresponds to a tree,  $\mathbf{E}[\prod_{(i,j) \in S} \mathbf{G}_{ij}] = p^{|S|}$ . Recalling that we have defined  $Q = p^2(1 + \varepsilon) = \mathbf{E}[\mathbf{G}_{12}\mathbf{G}_{13}\mathbf{G}_{23} \mid \mathbf{G}_{23} = 1]$ ,

$$\begin{aligned} \mathbf{E}[\bar{T}_{123}^2] &= (1 - 2p)^3 \cdot p \cdot Q + 3(1 - 2p)^2 p^4 + 3(1 - 2p)p^5 + p^6 - (pQ - p^3)^2 \\ &= p^3(1 - p)^3 + (1 - 2p)^3 p^3 \varepsilon - p^6 \varepsilon^2 \\ &= O((1 + \varepsilon)p^3). \end{aligned} \quad (24)$$

Dealing now with the second term, starting with (23),

$$\begin{aligned} \mathbf{E}[\bar{T}_{123} \bar{T}_{234}] &= \mathbf{E}[T_{123}T_{234}] - \mathbf{E}[T_{123}]^2 \\ &= \mathbf{E}[T_{123}T_{234}] - \varepsilon^2 p^6 \end{aligned}$$

$$\begin{aligned}
&= \mathbf{E}[(1-2p)\mathbf{G}_{23} + p^2)(\mathbf{G}_{12} - p)(\mathbf{G}_{13} - p)(\mathbf{G}_{24} - p)(\mathbf{G}_{34} - p)] - \varepsilon^2 p^6 \\
&= (1-2p) \cdot p \cdot \mathbf{E}[(\mathbf{G}_{12} - p)(\mathbf{G}_{13} - p) \mid \mathbf{G}_{23} = 1]^2 + p^2 \mathbf{E}[(\mathbf{G}_{12} - p)(\mathbf{G}_{13} - p)(\mathbf{G}_{24} - p)(\mathbf{G}_{34} - p)] - \varepsilon^2 p^6
\end{aligned}$$

Where we have used that  $(\mathbf{G}_{12} - p)(\mathbf{G}_{13} - p)$  and  $(\mathbf{G}_{24} - p)(\mathbf{G}_{34} - p)$  are independent conditioned on  $\mathbf{G}_{23}$ . Now again applying the simplification that tree-shaped products are independent,

$$\begin{aligned}
&= (1-2p)p(Q - p^2)^2 + p^2(\mathbf{E}[\mathbf{G}_{12}\mathbf{G}_{13}\mathbf{G}_{24}\mathbf{G}_{34}] - p^4) - \varepsilon^2 p^6 \\
&= (1-2p)p^5\varepsilon^2 + p^2(\mathbf{E}[\mathbf{G}_{12}\mathbf{G}_{13}\mathbf{G}_{24}\mathbf{G}_{34}] - p^4) - \varepsilon^2 p^6.
\end{aligned}$$

Now, we use that  $\mathbf{E}[\mathbf{G}_{12}\mathbf{G}_{13}\mathbf{G}_{24}\mathbf{G}_{34}] \leq \mathbf{E}[\mathbf{G}_{12}\mathbf{G}_{13}\mathbf{G}_{24}\mathbf{G}_{34} \mid \mathbf{G}_{23} = 1]$ , and again applying independence conditioned on  $\mathbf{G}_{23}$ ,

$$\begin{aligned}
&\leq (1-2p)p^5\varepsilon^2 + p^2(\mathbf{E}[\mathbf{G}_{12}\mathbf{G}_{13} \mid \mathbf{G}_{23}]^2 - p^4) - \varepsilon^2 p^6 \\
&= (1-2p)p^5\varepsilon^2 + p^2(p^4(1+\varepsilon)^2 - p^4) - \varepsilon^2 p^6 \\
&= (1-2p)p^5\varepsilon^2 + 2p^6\varepsilon.
\end{aligned}$$

Hence we have

$$\mathbf{E}[\overline{\mathbf{T}}_{123}\overline{\mathbf{T}}_{234}] = O(p^5\varepsilon^2 + 2p^6\varepsilon). \quad (25)$$

Putting together (21), (24), and (25), we have that there exists a constant  $C$  such that

$$\frac{\sqrt{\mathbf{Var}[T(\mathbf{G})]}}{\mathbf{E}[T(\mathbf{G})]} \leq C \cdot \frac{\sqrt{(1+\varepsilon)p^3n^3 + p^5\varepsilon^2n^4 + p^6\varepsilon n^4}}{n^3p^3\varepsilon} \leq C \cdot \left( \sqrt{\frac{1+\varepsilon}{p^3n^3\varepsilon^2}} + \frac{1}{n\sqrt{p}} + \frac{1}{n\sqrt{\varepsilon}} \right).$$

Since we have assumed that  $p \gg \frac{1}{n^2}$ ,  $\frac{1}{n\sqrt{p}} \rightarrow_n 0$ . Also, by Claim A.2,  $\frac{1}{n\sqrt{\varepsilon}} \leq \frac{d^{1/4}}{nc^{1/2}\log^{3/4}\frac{1}{p}} \rightarrow_n 0$  so long as  $d \ll n^4 \log^3 \frac{1}{p}$ , which is implied by our assumption that  $d \ll (nH(p))^3$ . Finally, if  $\varepsilon > 1$ , then applying Claim A.2,

$$\frac{1+\varepsilon}{p^3n^3\varepsilon^2} \leq \frac{2\varepsilon}{p^3n^3\varepsilon^2} \leq \frac{2}{p^3n^3\varepsilon} \leq \frac{2d^{1/2}}{cp^3n^3\log^{3/2}\frac{1}{p}} \rightarrow_n 0,$$

whenever  $d \ll (nH(p))^3$ . Alternatively, if  $\varepsilon < 1$ , applying Claim A.2 again,

$$\frac{1+\varepsilon}{p^3n^3\varepsilon^2} \leq \frac{2}{p^3n^3\varepsilon^2} \leq \frac{2d}{c^2p^3n^3\log^3\frac{1}{p}} \rightarrow_n 0,$$

since  $d \ll (nH(p))^3$ . This concludes the proof.  $\square$

## B Deferred proofs from the preliminaries

We first provide a proof of Lemma 3.6, which provides a convenient upper bound on the dot product threshold  $\tau(p)$  corresponding to a  $p$ -cap.

**Lemma** (Restatement of Lemma 3.6). *For any  $p \leq \frac{1}{2}$ , we have  $\tau(p) \leq \sqrt{\frac{2\log(1/p)}{d}}$ .*

We apply the following constant factor approximation for the measure of a sphere cap in terms of the threshold  $\tau(p)$ , when  $\tau(p) \geq \sqrt{\frac{2}{d}}$  given by [BGK<sup>+</sup>01, Lemma 2.1(b)]:

**Theorem B.1.** Consider a  $p$ -cap where  $\tau(p) \geq \sqrt{\frac{2}{d}}$ . Then:

$$\frac{1}{6\tau(p)\sqrt{d}} (1 - \tau(p)^2)^{(d-1)/2} \leq p \leq \frac{1}{2\tau(p)\sqrt{d}} (1 - \tau(p)^2)^{(d-1)/2}$$

*Proof of Lemma 3.6.* We case on whether  $\tau(p)$  is smaller or larger than  $\sqrt{\frac{2}{d}}$ . In the first case observe that  $\log(1/p) \geq 1$  by our bound on  $p$  and so  $\tau(p) \leq \sqrt{\frac{2}{d}} \leq \sqrt{\frac{2\log(1/p)}{d}}$ .

When  $\tau(p) \geq \sqrt{\frac{2}{d}}$ , we use the upper bound in [Theorem B.1](#). Let  $\tau' = \sqrt{\frac{2\log(1/p)}{d}}$ , and let  $p'$  denote its corresponding tail probability  $\Pr_{x,y \in \mathbb{S}^{d-1}}[\langle x, y \rangle \geq \tau']$ . If we can show  $p' \leq p$ , then we know  $\tau(p) \leq \tau'$ .

$$\begin{aligned} p' &\leq \frac{1}{2\tau'\sqrt{d}} \cdot (1 - t^2)^{(d-1)/2} = \frac{1}{2\sqrt{\log(1/p)}} \cdot \left(1 - \frac{2\log(1/p)}{d}\right)^{(d-1)/2} \\ &\leq \frac{1}{2\sqrt{\log(1/p)}} \cdot p \leq p \end{aligned}$$

Since  $p' \leq p$ , this tells us  $\tau \leq \tau'$  as well, giving us the desired inequality.  $\square$

We next prove [Lemma 3.7](#). Understanding the joint distribution of two independently chosen unit vectors in  $\mathbb{S}^{d-1}$  helps complete this proof. We first present a theorem from [\[BBN20\]](#) that present some useful bounds on this distribution  $\psi$ .

**Lemma B.2** ([\[BBN20, Lemma 5.1\]](#)). Let  $\tau(p)$  be the value of  $t$  at which  $\Pr_{x,y \in \mathbb{S}^{d-1}}[\langle x, y \rangle \geq t] = p$ , and let  $\psi_d$  be the distribution of the inner products of two random unit vectors in  $\mathbb{R}^2$ .

1. For  $0 \leq \tau \leq \frac{1}{2}$  and  $\delta > 0$ , we have:

$$\frac{\psi_d(\tau - \delta)}{\psi_d(\tau)} \leq \exp(2\tau d \delta).$$

2.  $\psi_d(\tau) \leq C_{B.2} p \cdot \max\{\sqrt{d}, d\tau\}$  for a universal constant  $C_{B.2}$ .

$$3. \psi(\tau(p)) = \frac{\Gamma\left(\frac{d}{2}\right)}{\Gamma\left(\frac{d-1}{2}\right)\sqrt{\pi}} \cdot (1 - \tau(p)^2)^{(d-3)/2}.$$

*Proof of Lemma 3.7.* When  $t \geq 0$ , the density  $\psi_d(t)$  is a decreasing function in  $t$ . Thus:

$$\begin{aligned} \Pr_{z \sim \rho}[\tau - \varepsilon \leq \langle z, x \rangle \leq \tau + \varepsilon] &= \int_{\tau - \varepsilon}^{\tau + \varepsilon} \psi_d(t) dt \\ &\leq (2\varepsilon) \cdot [\psi_d(\tau) \cdot \exp(2d\tau\varepsilon)] \end{aligned}$$

In the last line, we used [Part 1](#) of [Lemma B.2](#), and noted that this is an upper bound even when  $\varepsilon > \tau$ . Using [Part 2](#) of [Lemma B.2](#), and [Lemma 3.6](#), we can upper bound  $\psi_d(\tau)$ :

$$\psi_d(\tau) \leq C_{B.2} p \cdot \max\{\sqrt{d}, d\tau\} \leq p \cdot \left(C_{B.2} \sqrt{2} \cdot \sqrt{d} \cdot \sqrt{\log(1/p)}\right)$$

from which the desired bound follows.  $\square$