LOCALIZATION IN RANDOM GEOMETRIC GRAPHS WITH TOO MANY EDGES

BY SOURAV CHATTERJEE[1] AND MATAN HAREL[2]

[1]*Department of Statistics, Stanford University, souravc@stanford.edu*
[2]*Department of Mathematics, Tel Aviv University, matanharel8@tauex.tau.ac.il*

We consider a random geometric graph $G(\chi_n, r_n)$, given by connecting two vertices of a Poisson point process χ_n of intensity n on the d-dimensional unit torus whenever their distance is smaller than the parameter r_n. The model is conditioned on the rare event that the number of edges observed, $|E|$, is greater than $(1 + \delta)\mathbb{E}(|E|)$, for some fixed $\delta > 0$. This article proves that upon conditioning, with high probability there exists a ball of diameter r_n which contains a clique of at least $\sqrt{2\delta\mathbb{E}(|E|)}(1 - \varepsilon)$ vertices, for any given $\varepsilon > 0$. Intuitively, this region contains all the "excess" edges the graph is forced to contain by the conditioning event, up to lower order corrections. As a consequence of this result, we prove a large deviations principle for the upper tail of the edge count of the random geometric graph. The rate function of this large deviation principle turns out to be nonconvex.

1. Introduction. The random geometric graph is a simple stochastic model, first studied in [20] in 1972, for generating a graph: given the parameters n and r, consider a Poisson point process of intensity n on the d-dimensional unit torus, equipped with a translation-invariant metric inherited from a norm $\|\cdot\|$ on Euclidean space (which may not be the Euclidean norm), and declare an edge between any two vertices that are at distance $\leq r$ from each other.

Unlike the well-known Erdős–Rényi random graph, the random geometric graph's definition leads to strong dependence between edges: if three vertices form a "V" shaped graph, they are far more likely to have the third edge of the triangle than if no assumption were made on the other edges, as a consequence of the triangle inequality.

Many properties of this graph model have been studied. The classic monograph of Mathew Penrose [33] studies results pertaining to many graph-theoretical functions of random geometric graphs, including (but not limited to) laws of large numbers and central limit theorems for subgraph counts, independence number and chromatic number, as well as many properties connected to the giant component. Many of the results presented in this monograph have been improved and generalized by Penrose and others in the years since its initial publication. Besides this, there have been investigations into other probabilistic features, such as threshold functions for cover times and mixing times [3] and thresholds for monotone graph functions [18]. This list is far from comprehensive, of course, and the random geometric graph remains an active object of research.

The random geometric graph is also closely related to the random connection continuum percolation model. In that model, the vertex set is given by an (almost surely infinite) Poisson point process of fixed intensity on $\mathbb{R}^d$, and two points are connected with some probability that varies (and usually decreases) with their distance. In particular, the special case in which the radius of connection is deterministically fixed at 1 was the model that initiated the study of this kind of random geometry, in the seminal paper of Gilbert [17]. The properties of interest in this model are the existence of an infinite connected component, as well as the behavior

---

of the subset of $\mathbb{R}^d$ that is at distance at most 1 from one of the vertices of the graph (the so-called "Poisson blob") and its complement (the "vacant set"). Continuum percolation is treated in detail in a book-length monograph by Meester and Roy [30], as well as in the book by Grimmett [19].

Most of the work done on random geometric graphs is concerned with either the behavior of a typical graph—the graph we are likely to see for a given r as n goes to infinity—or typical deviations from that behavior—that is, central limit theorems. In this paper, we are concerned with the behavior of the model conditioned on a rare event. Specifically, we will study the random geometric graph conditioned on having many more edges than is expected (a formal description will follow). The large deviation regime of the upper tail of any subgraph count of the random geometric graph is not well understood, though some bounds are available: Janson [24] established concentration inequalities for U-statistics, a general class of statistics which includes the subgraph counts we are interested in. These upper bounds work in very general settings, but are not tight, even up to constants in the exponent. Large deviation principles have been proven for functionals of random point processes in which the contribution of any particular vertex is uniformly bounded [37], but no such bound is known for functionals with possibly large influence, such as the edge count of the graph.

As motivation for this detailed study, we consider the problem in a more familiar context: the "infamous upper tail" [26] of the triangle count T in the Erdős–Rényi random graph, $G(n, p)$. After many years of development of increasingly strong bounds, the first break-through was made by Kim and Vu [27] and Janson et al. [25] independently, who proved that, for any $\delta > 0$ and whenever $p \gg (\log n)/n$,

$$\exp[-c(\delta)n^2 p^2 \log(1/p)] \leq \mathbb{P}[T > (1 + \delta)\mathbb{E}[T]] \leq \exp[-C(\delta)n^2 p^2],$$

where $c(\cdot)$ and $C(\cdot)$ depend only on δ. Recently, there has been renewed interest in these type of tail estimates. In 2010, Chatterjee [7] and Demarco and Kahn [12] (in independent works) established the correct order of the upper tail of triangles and other small cliques by adding the missing logarithmic term to the upper bound, without providing good control of the leading-order constants. The work of Chatterjee and Dembo [10] on nonlinear large deviations proved that the upper tail probability can be described in terms of a continuous variational problem when p is vanishing sufficiently slowly—namely, when $n^{-1/42} \ll p \ll 1$. Generalizations and expansions of the approach by Eldan [14] established the variational equivalence for $n^{-1/18} \log n \ll p \ll 1$; Cook and Dembo [11] proved the result for $n^{-1/3} \ll p \ll 1$, and Augeri [2] did the same for $n^{-1/2}(\log n)^2 \ll p \ll 1$. Lubetzky and Zhao [29] solved the variational problem for triangles (Bhattacharya, Ganguly, Lubetzky and Zhao [5] did the same for more general subgraphs), which thus calculated both the order and the leading-order constant for the upper tail question in a certain regime of sparse Erdős–Rényi random graphs. The main results and ideas from this body of work are summarized in the survey article [8]. Recently, Harel, Mousset and Samotij [21] used a combinatorial approach to prove that the upper tail probability of the subgraph count of any fixed, regular graph can be expressed in terms of the solution of a discrete variational problem for nearly all values of p where localization is believed to hold. Unfortunately, all the papers described above are only valid for functions of independent Bernoulli random variables, and are therefore not applicable to the problem we are studying here.

In this work, we use the properties the random geometric graph inherits from the geometry of $\mathbb{R}^d$ to evaluate the upper tail large deviation rate function. In addition, we provide a "structure theorem" to describe the graph-theoretical structure of the model conditioned on having too many edges. Specifically, we show that such a conditional model exhibits *localization*. Heuristically, this phenomenon can be described as a scenario in which a small number of vertices will contribute almost all the extra edges that we require the graph to exhibit, while

the edge count of the "bulk" of the graph will remain largely unchanged, in some weak sense. Furthermore, we will show that the geometry of the localized region has the shape of a ball in the given norm (we will make these two statements more precise at the end of the next section). Outside of the aforementioned works of Lubetzky and Zhao [29], Bhattacharya et al. [5] and Harel, Mousset and Samotij [21] in the Erdős–Rényi model, this work is the only (as far as the authors are aware) to establish that the large deviation regime of a subgraph count is (weakly) equivalent to planting a combinatorial structure in the usual, unconditional graph.

The fact that large deviation events may be dominated by configurations with a small number of very large contributions was known relatively early in the history of large deviation theory: a survey by Nagaev [32], summarizing a series of papers written in the Soviet Union in the 1960s and 70s, includes this observation for sums of i.i.d. random variables with stretched-exponential tails. In our context, the natural combinatorial structure for creating many edges with a small number of edges is a "giant clique." The clique number, the (typical) size of the largest clique of the random geometric graph, falls under the general class of scan statistics, and has been shown to focus on two values with high probability for certain values of r (see [31, 34]); however, these works do not explore the large deviation regime. Our work uses techniques from large deviations, concentration inequalities, convex analysis and geometric measure theory. A key component in the proof is a technique for proving localization that has previously appeared in [39] and [9].

2. Definitions and main results. Let χ_n be a Poisson point process of intensity n on the d-dimensional unit torus $\mathbb{T}^d = [0, 1]^d$. For any $S \subset \mathbb{T}^d$, we denote the restriction of χ_n to S by $\chi_n(S)$. Let $N := |\chi_n|$. Recall that N is a Poisson random variable with mean n, and conditional on N, χ_n is just a set of N points, each chosen independently and uniformly at random. Let r_n be a positive sequence that decreases to 0 as $n \to \infty$, and $\|\cdot\|$ be some norm on $\mathbb{R}^d$ that induces a translation-invariant metric on $\mathbb{T}^d$. We define the random geometric graph $G(\chi_n, r_n) := (V, E)$, where $V = \chi_n = \{v_1, \ldots, v_N\}$, enumerated arbitrarily, and E is the set of unordered pairs $\{i, j\}$ such that $\|v_i - v_j\| \leq r_n$. Figure 1 shows a particular instance of $G(\chi_{150}, 0.1)$.
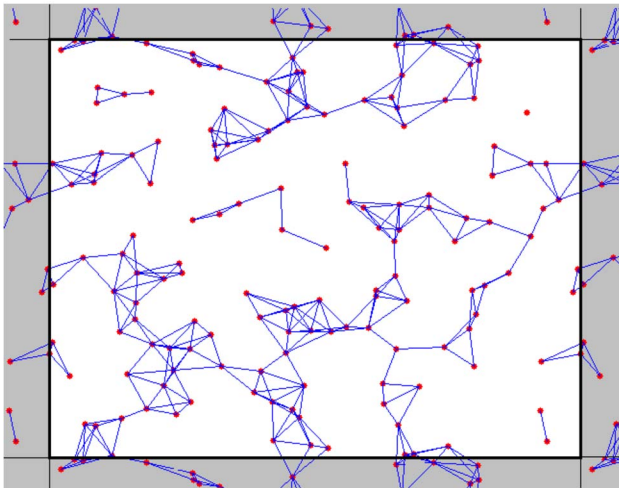

FIG. 1. *An instance of the random geometric graph $G(\chi_{150}, 0.1)$, with respect to the Euclidean norm. The graph has* 148 *vertices and* 343 *edges. The gray area is the white unit square translated, to show periodicity.*

Letting $1_{i,j}$ be the indicator that there is an edge between v_i and v_j, we can calculate the expected value of $|E|$, the number of edges in the graph

$$\mathbb{E}(|E|) = \mathbb{E}\left(\sum_{1 \le i < j \le N} 1_{i,j}\right) = \mathbb{E}\left[\binom{N}{2}\mathbb{E}(1_{1,2} \mid N)\right]$$

$$= \frac{n^2}{2} \cdot \mathbb{P}(\|v_1 - v_2\| \le r_n).$$

Denoting Lebesgue measure on both $\mathbb{R}^d$ and $\mathbb{T}^d$ by $\lambda(\cdot)$, we define

$$v := \lambda[\{x \in \mathbb{R}^d : \|x\| \le 1\}]$$

to be the volume of the unit ball in the norm $\|\cdot\|$. Then, by translation invariance of the metric induced on $\mathbb{T}^d$, $\mathbb{P}(\|v_1 - v_2\| \le r_n)$ is simply $v r_n^d$, as long as r_n is sufficiently small (to ensure the ball on the torus has the same measure as the one in $\mathbb{R}^d$)

$$\mu_n := \mathbb{E}(|E|) = \frac{v \cdot n^2 r_n^d}{2}.$$

We can also compute the variance of $|E|$:

$$\mathrm{Var}(|E|) = \mathbb{E}[\mathrm{Var}(|E| \mid N)] + \mathrm{Var}(\mathbb{E}[|E| \mid N])$$

$$= \mathbb{E}\left[\mathbb{E}\left(\sum_{1 \le i < j \le N, 1 \le i' < j' \le n} (1_{i,j} - v r_n^d)(1_{i',j'} - v r_n^d) \mid N\right)\right]$$

(2.1)
$$+ (v r_n^d)^2 \mathrm{Var}\left[\binom{N}{2}\right]$$

$$= \frac{n^2}{2}(v r_n^d - v^2 r_n^{2d}) + \left(n^3 + \frac{n^2}{2}\right)v^2 r_n^{2d}$$

$$= \mu_n(1 + 2v n r_n^d),$$

where we note that $(i, j) \ne (i', j')$ implies that the indicators $1_{i,j}$ and $1_{i',j'}$ are conditionally independent. This implies that, as long as $\mu_n \to \infty$, $\mathrm{Var}(|E|) \ll \mu_n^2$, and $|E|$ concentrates around its mean by Chebyshev's inequality.

For the rest of the article, we suppose the existence of a fixed constant $\delta^* > 0$ such that, for all sufficiently large n,

(2.2)
$$n^{(\delta^*-2)/d} \le r_n \le n^{-\delta^*/d}.$$

The lower bound ensures that the expected number of edges grows as a positive power of n; the upper bound excludes the possibility of $r_n = n^{-o(1)}$, that is, bounded above and below by n^ε and $n^{-\varepsilon}$, respectively, for any fixed $\varepsilon > 0$ and $n \ge n_0$, for some n_0 depending on ε. We will reuse the notation $n^{o(1)}$ throughout the paper in this sense, and we will allow the (implicit) ε to depend on any fixed parameter other than n. We define the parameter p as

(2.3)
$$p := \lim_{n \to \infty} \frac{\log \mu_n}{\log n},$$

implicitly assuming that the limit exists. This ensures that $\mu_n = f(n)n^p$, where $f(n) = n^{o(1)}$. Notice that

(2.4)
$$\delta^* \le p \le 2 - \delta^*,$$

thanks to (2.2). We will say the random geometric graph is *admissible* if r_n satisfies (2.2) and the limit above exists.

The following theorem is the main result of the paper.

THEOREM 2.1.   *Let $G(\chi_n, r_n)$ be an admissible random geometric graph model on $\mathbb{T}^d$ with respect to some norm $\|\cdot\|$. Define*

$$\tau_n := \nu \cdot (r_n/2)^d,$$

that is, τ_n is the volume of a ball of diameter r_n. Fix $\delta > 0$ and $\varepsilon > 0$, and let $\mathscr{F}_n(\varepsilon)$ be the event that there exists a ball B of diameter r_n such that:

(1) *any convex set $S \subset B$ satisfies*

$$\left| \frac{|\chi_n(S)|}{\sqrt{2\delta\mu_n}} - \frac{\lambda(S)}{\tau_n} \right| < \varepsilon,$$

(2) *for any convex set $S' \subset B^c$ such that $\mathrm{diam}(S') \leq r_n$ and $\lambda(S') > \varepsilon\tau_n$,*

$$\frac{|\chi_n(S')|}{\sqrt{2\delta\mu_n}} < \varepsilon \cdot \frac{\lambda(S')}{\tau_n}.$$

Then

$$\lim_{n\to\infty} \mathbb{P}\big[\mathscr{F}_n(\varepsilon) \mid |E| > (1+\delta)\mu_n\big] = 1.$$

As a consequence of Theorem 2.1, we prove that the upper tail of the edge count of random geometric graphs satisfies a large deviation principle. Recall that a sequence of nonnegative random variables X_n satisfies an *upper tail* large deviation principle with speed $s(n)$ and rate function $I(x)$ if, for any closed set $F \subset (0, \infty)$,

$$\limsup_{n\to\infty} \frac{1}{s(n)} \log \mathbb{P}\left(\frac{X_n - \mathbb{E}[X_n]}{\mathbb{E}[X_n]} \in F \right) \leq - \inf_{x\in F} I(x),$$

and for any open set $G \subset [0, \infty)$,

$$\liminf_{n\to\infty} \frac{1}{s(n)} \log \mathbb{P}\left(\frac{X_n - \mathbb{E}[X_n]}{\mathbb{E}[X_n]} \in G \right) \geq - \inf_{x\in G} I(x).$$

(For more on large deviation principles and their applications see, e.g., [13].) The following theorem gives the upper tail large deviation principle for the number of edges in a random geometric graph.

THEOREM 2.2.   *Let $G(\chi_n, r_n)$ be an admissible random geometric graph model on the d-dimensional torus, with the same assumptions as in Theorem 2.1. Define*

$$I(x) := \left(\frac{2-p}{2} \right) \sqrt{2x},$$

where p is defined as in (2.3). Then $|E|$ satisfies an upper tail large deviation principle with speed $s(n) = \sqrt{\mu_n} \log n$ and rate function $I(x)$.

There are several important features to the two main theorems of this paper: first, both describe models in which the number of edges significantly *exceeds* its mean. The lower tail of the edge count—that is, events of the form $\{|E| < (1-\delta)\mu_n\}$—is likely to satisfy Poisson-like statistics. Its large deviation principle is expected to hold with speed μ_n, and no special combinatorial structure like the "giant clique" of Theorem 2.1 should appear.

Before we go on, let us comment on the precise properties of the giant clique given by our two main theorems. Since the rate function of Theorem 2.2 is strictly increasing, we know that, conditional on $\{|E| > (1+\delta)\mu_n\}$, the event

$$\{(1+\delta')\mu_n > |E| > (1+\delta)\mu_n\}$$

occurs with high probability (i.e., probability at least $1 - \varepsilon$) for any $\delta' > \delta$ and n sufficiently large. Now, if we set $S = B$ in the first stipulation of Theorem 2.1, we see that the ball B of diameter r_n makes up a clique of at least $\sqrt{2\delta\mu_n}(1 - \varepsilon)$ vertices and, therefore, at least $\delta\mu_n(1 - 2\varepsilon)$ edges. Since ε and $\delta' - \delta$ are arbitrarily close to zero, we find that the clique in B has $\delta\mu_n + o(\mu_n)$ edges, whereas the rest of the graph has $\mu_n + o(\mu_n)$ edges. This formalizes our earlier claim that "almost all extra edges in the conditional model are between points in B."

Theorem 2.1 also gives information about the internal geometry of the giant clique. If we pick S to be a proper convex subset of the ball B, we find that $|\chi_n(S)|$ proportional to $\sqrt{2\delta\mu_n}$ times the density of S in B (again, up to lower order corrections). We restricted S to be convex in order to preclude pathological sets, such as sets which are sparse but of large measure (e.g., generalized Cantor sets) or have boundaries that take up a large amount of space. It should be possible to replace convexity with a weaker assumption. That being said, probing the Poisson point process χ_n with convex $S \subset B$ is enough to establish that the *conditional* process, restricted to B, is distributed roughly uniformly, up to errors that vanish in comparison to $\sqrt{\mu_n}$.

Finally, we would like to say that there are no other large cliques in $G(\chi_n, r_n)$ conditioned on $\{|E| > (1+\delta)\mu_n\}$; unfortunately, Theorem 2.1 does not provide this result. Instead, we can only be sure that every other clique outside of the "exceptional" set B has $o(\sqrt{\mu_n})$ vertices, that is, much smaller than the largest clique.

3. The s-graded model. Henceforth in the manuscript, we will suppress the subscript n and write χ, μ, τ and r instead of χ_n, μ_n, τ_n and r_n.

We now present an approximation of the random geometric model which allows us to replace the Poisson point process with a sequence of independent Poisson random variables. To do this, we first discretize space, and then produce a semimetric on the resulting "cells" that approximates the norm $\| \cdot \|$ on the unit torus. We call this the s-graded model.

For a positive integer s, define

$$m := \lfloor s/r \rfloor,$$

so that

$$\frac{s}{r} - 1 \leq m \leq \frac{s}{r}.$$

This definition and (2.3) imply that

$$(3.1) \qquad m^d = n^{2-p+o(1)},$$

where the constant in the $o(1)$ depends on s. Let $T = \{1, 2, \ldots, m\}^d$. Pick $I = (i_1, i_2, \ldots, i_d) \in T$, and define

$$A_I = \left[\frac{i_1 - 1}{m}, \frac{i_1}{m}\right] \times \cdots \times \left[\frac{i_d - 1}{m}, \frac{i_d}{m}\right].$$

The A_I's partition the unit torus into m^d cubes (ignoring sets of measure 0), each of volume $1/m^d$ and, therefore, $X_I = |\chi(A_I)|$ is a Poisson random variable of mean

$$(3.2) \qquad \mathcal{D} := \frac{n}{m^d}.$$

We now define a semimetric on T, induced by the norm on torus:

$$(3.3) \qquad \rho(I, J) = \inf_{x \in A_I^o, y \in A_J^o} \lceil m\|x - y\| \rceil,$$

where the circles indicate the interiors of the sets. Note that the $\rho(\cdot, \cdot)$ is always an integer. Moreover, $\rho(I, J) = z$ if z is the smallest integer such that some point in A_I^o and some point in A_J^o are less than z away, measured in units of $1/m$, the side length of the cubes. We force the points to be in the interior to prevent "trivialities," such as two adjacent cells being distance 0, since they share a boundary. Note that $\rho(\cdot, \cdot)$ does not satisfy the triangle inequality, and hence is only a semimetric. To see this, consider $\mathbb{T}^5$ under the Euclidean norm, and the cells $A_1 = A_{(1,\ldots,1)}$, $A_2 = A_{(2,\ldots,2)}$ and $A_3 = A_{(3,\ldots,3)}$. Since A_1 and A_2 share a corner, $\rho(A_1, A_2) = 1$, and the same holds for $\rho(A_2, A_3)$. However, $\rho(A_1, A_3) = \sqrt{5} > 1 + 1 = \rho(A_1, A_2) + \rho(A_2, A_3)$. But ρ does satisfy a modified triangle inequality of the form $\rho(I, J) \leq \rho(I, K) + \rho(K, J) + C_d$, where C_d depends only on the dimension and choice of norm, though we never make explicit use of this fact.

We are now ready to define the s-graded random geometric graph. Let $G_s(\chi, r) = (V, E_s)$ have the same vertex set as the original graph. For each vertex v, let I_v be the index in T such that $v \in A_{I_v}$; there is ambiguity on the boundary of the A_I's, but that set has Lebesgue measure 0 and, therefore, it has no vertices of χ, almost surely. We say $(v, w) \in E_s$ whenever $\rho(I_v, I_w) \leq s$. Heuristically, the s-graded model allows every point to wander inside a cubical "cage" of side-length $1/m$, and connects any two points that might be connected after we allow this mobility. In this framework, it is clear that E_s becomes smaller as s decreases. In fact, for sufficiently large s, E_s is identical to E; unfortunately, this s will be random. In formulating Theorem 3.1, the main theorem of this section (which is proved in Section 5), we will let s be an arbitrary positive integer, and discuss its asymptotic properties as n goes to infinity. Later, in Sections 6 and 7, we will take s to sufficiently large, and show that the resulting approximation is good enough for our purposes.

Having defined the s-graded model, we now need to compute several quantities related to it, as we did for the random geometric graph in Section 2. We will denote s-graded model variables with tildes, to distinguish them from similar variables defined by the continuous geometry of the $\mathbb{T}^d$. We will say $G_s(\chi, r) = (V, E_s)$ is admissible whenever the random geometric graph $G(\chi, r)$ is admissible.

The major benefit of the s-graded model is that its edge count is very simple to express in terms of X_I, the number of points in each A_I:

$$
\begin{aligned}
|E_s| &= \sum_{I \in T} \left[\binom{X_I}{2} + \frac{1}{2} \sum_{J : 0 < \rho(I, J) \leq s} X_I X_J \right] \\
&= \frac{1}{2} \sum_{I \in T} X_I \left[\left(\sum_{J : \rho(I, J) \leq s} X_J \right) - 1 \right].
\end{aligned}
$$

(3.4)

This random variable is defined in terms of i.i.d. random variables, which eases the analysis greatly. The geometric relations that define the edge count are now completely encoded by ρ. Finally, each X_I only appears in finitely many terms in this expression (i.e., the number of terms involving X_I is uniformly bounded in n). The "finite range" nature of the representation will play a major role in the proof presented.

We quantify this fact as follows: for any $I \in T$, let

$$
\tilde{N}_I := \{J : \rho(I, J) \leq s\}.
$$

Thanks to translation invariance of ρ, the cardinality of this set is independent of the choice of I. Using this parameter, we can compute the expected number of edges in the s-graded

random geometric graph easily:

$$\tilde{\mu}_s := \mathbb{E}(|E_s|)$$

(3.5)
$$= \sum_{I \in T} \mathbb{E}\left[\binom{X_I}{2} + \frac{1}{2} \sum_{J:0<\rho(I,J)\leq s} \mathbb{E}(X_I)\mathbb{E}(X_J)\right]$$

$$= \frac{|\tilde{N}_I|m^d \mathscr{D}^2}{2} = \frac{|\tilde{N}_I|n^2}{2m^d},$$

where we recall that $\mathscr{D}$ is the mean of X_I, and the defining relation (3.2). The variance of $|E_s|$ is also straightforward to calculate from the above representation, though the exact formula is messy. Instead, we produce an upper bound: the variance of $|E_s|$ can be thought of as the sum of $(\mathbb{E}[Q_I \cdot Q_{I'}] - \mathbb{E}[Q_I]^2)$, where Q_I is the summand in (3.4) and $I, I' \in T$. This quantity is maximized when $I = I'$, and is zero if the two terms are independent. Thus, we find that

$$\text{Var}[|E_s|] \leq \sum_{I \in T} |\tilde{S}_I| \cdot \mathbb{E}\left(\left[\binom{X_I}{2} + \frac{1}{2} \sum_{J:0<\rho(I,J)\leq s} X_I X_J - \frac{|\tilde{N}_I|\mathscr{D}^2}{2}\right]^2\right),$$

where

$$\tilde{S}_I := \{J : \tilde{N}_I \cap \tilde{N}_J \neq \varnothing\}.$$

A straightforward (if elaborate) computation can show that this implies that

(3.6)
$$\text{Var}[|E_s|] \leq 16|\tilde{S}_I||\tilde{N}_I|^2 m^d \cdot \max\{\mathscr{D}^3, \mathscr{D}^2\}.$$

In Lemma 5.1 below, we will show that both $|\tilde{N}_I|$ and $|\tilde{S}_I|$ are uniformly bounded in n. Together with (3.5) and (3.2), this implies that, for any s, $\text{Var}[|E_s|] \ll \tilde{\mu}_s^2$, and hence $|E_s|$ concentrates around its mean by Chebyshev's inequality.

As before, we are interested in conditioning the s-graded model on the event $\{|E_s| > (1 + \tilde{\delta})\tilde{\mu}_s\}$. Following Theorem 2.1, we expect that such conditional measures will be concentrated on configurations with many points on sets of diameter s and maximal cardinality. We call a set of indices a *maximal clique set* if it is a subset of T with diameter $\leq s$ that achieves the maximal cardinality of all such sets. We define

(3.7)
$$\tilde{\tau}_s := \max\{|\mathfrak{I}| : \mathfrak{I} \subset T, \text{diam}(\mathfrak{I}) \leq s\},$$

that is, $\tilde{\tau}_s$ is the cardinality of a maximal clique set. Clearly, $\tilde{\tau}_s$ is increasing in s, and

(3.8)
$$\tilde{\tau}_s \geq \tilde{\tau}_1 \geq 2^d,$$

as the diameter of the set $\{I = (\eta_1, \ldots, \eta_n) : \eta_i \in \{1, 2\}\}$ under the semi-metric $\rho(\cdot, \cdot)$ is exactly 1, as all the A_I's share a corner. We will also need an approximate notion of this geometric object: we say a set is a $\tilde{\varepsilon}$-almost maximal clique set if its diameter is bounded above by s, and its cardinality is at least $(1 - \tilde{\varepsilon})\tilde{\tau}_s$.

We can now state the equivalent to Theorem 2.1 for the s-graded model.

THEOREM 3.1. *Let s be a positive integer. Consider $G_s(\chi, r)$, an admissible s-graded random geometric graph. For any $\tilde{\delta} > 0$, define the event $\mathscr{L}_n(\tilde{\delta})$ by*

$$\mathscr{L}_n(\tilde{\delta}) := \{|E_s| > (1 + \tilde{\delta})\tilde{\mu}_s\}.$$

For any $\tilde{\varepsilon} > 0$, let $\mathscr{G}_{n,\tilde{\delta}}(\tilde{\varepsilon})$ be the event there exists a pair of sets $\mathfrak{B}$ and $\mathfrak{C}$ in T such that:

(1) $\mathfrak{B}$ *is a $\tilde{\varepsilon}$-almost maximal clique set,*

(2) *for all $I \in \mathfrak{B}$,*

$$\left| \frac{\tilde{\tau}_s X_I}{(2\tilde{\delta}\tilde{\mu}_s)^{1/2}} - 1 \right| < \tilde{\varepsilon},$$

(3) $\mathfrak{C}$ *satisfies*

$$|\mathfrak{C}| < \tilde{\varepsilon} \cdot \tilde{\tau}_s \quad \text{and} \quad \sum_{I \in \mathfrak{C}} X_I < \tilde{\varepsilon} \cdot (2\tilde{\delta}\tilde{\mu}_s)^{1/2}$$

and

(4) *for all $J \in (\mathfrak{B} \cup \mathfrak{C})^c$,*

$$\frac{\tilde{\tau}_s X_J}{(2\tilde{\delta}\tilde{\mu}_s)^{1/2}} < \tilde{\varepsilon}.$$

There is a universal constant $\tilde{\varepsilon}_0 > 0$ such that the following is true. Take any $\tilde{\varepsilon} \in (0, \tilde{\varepsilon}_0)$, any positive integer s, and any three numbers $0 < \tilde{\delta}_0 \leq \tilde{\delta} \leq \tilde{\Delta}_0$. Then there is an integer n_0 depending only on s, $\tilde{\varepsilon}$, $\tilde{\delta}_0$ and $\tilde{\Delta}_0$, such that whenever $n \geq n_0$,

$$\mathbb{P}[\mathscr{G}_{n,\tilde{\delta}}(\tilde{\varepsilon})^c \cap \mathscr{L}_n(\tilde{\delta})]$$

$$\leq 3\exp\left(-(2\tilde{\delta}\tilde{\mu}_s)^{1/2}\left[\log\left(\frac{(2\tilde{\delta}\tilde{\mu}_s)^{1/2}}{\tilde{\tau}_s \cdot \mathscr{D}}\right) - 1 + (\tilde{\varepsilon}/10)^{10}/2\right]\right).$$

Because of its technical nature, Theorem 3.1 warrants a short explanation. It turns out that it is possible to show that, for some $\eta > 0$,

$$\mathbb{P}[\mathscr{L}_n(\tilde{\delta})] \geq \exp\left(-(2\tilde{\delta}\tilde{\mu}_s)^{1/2}\left[\log\left(\frac{(2\tilde{\delta}\tilde{\mu}_s)^{1/2}}{\tilde{\tau}_s \cdot \mathscr{D}}\right) - 1\right] - Cn^{p/2-\eta}\right).$$

This bound comes from explicitly "planting" a maximal clique set where every cell includes exactly $\lceil (2\tilde{\delta}\tilde{\mu}_s)^{1/2}/\tilde{\tau}_s \rceil$ vertices; we will not prove this fact, but Lemma 6.5 will show a very similar computation for the edge count of the random geometric graph. In Lemma 5.1 below, we will show that $|\tilde{N}_I|$ is uniformly bounded in n for any $s > 0$. Together with the fact that $\tilde{\delta}$ is uniformly bounded above and below in n, this implies that $(2\tilde{\delta}\tilde{\mu}_s)^{1/2} = n^{p/2+o(1)} \gg n^{p/2-\eta}$. Therefore, Theorem 3.1 shows that, with high probability, the event $\mathscr{G}_{n,\tilde{\delta}}(\tilde{\varepsilon})$ occurs in the conditional s-graded model.

The event $\mathscr{G}_{n,\tilde{\delta}}(\tilde{\varepsilon})$ produces a set $\mathfrak{B}$, which is very close to a maximal clique set, in which each X_I is very close to $\sqrt{2\tilde{\delta}\tilde{\mu}_s}/\tilde{\tau}_s$—the value we would expect if we were to spread the $\sqrt{2\tilde{\delta}\tilde{\mu}_s}$ vertices required to make a "giant" clique evenly among the $\tilde{\tau}_s$ elements of a maximal clique set. In addition, we allow for an "exceptional" set $\mathfrak{C}$, where some X_I's may be much larger than this average amount. However, this exceptional set is made up of few indices, and includes few vertices of χ, when compared with $\sqrt{2\tilde{\delta}\tilde{\mu}_s}$. Outside of these two sets, every X_J is at most $\tilde{\varepsilon}\sqrt{2\tilde{\delta}\tilde{\mu}_s}/\tilde{\tau}_s$—a lower order quantity when compared to the bounds on X_I, $I \in \mathfrak{B}$.

When this event fires, the conditional s-graded model has a clique with approximately $\tilde{\delta}\tilde{\mu}_s$ edges. We also know that the vertices are distributed roughly uniformly. Finally, we get a quantitative estimate on the probability that the edge count of the s-graded model exceeds its mean *without* the desired structure occurring. Note that the constants and 10th power of $\tilde{\varepsilon}$ that appears in the quantitative bound are somewhat arbitrary—we made no attempts to optimize them.

Suppose that $\tilde{\varepsilon} < (2\tilde{\tau}_s)^{-1}$. In this case, $\mathscr{G}_{n,\tilde{\delta}}(\tilde{\varepsilon})$ would require $|\mathfrak{B}| \geq \tilde{\tau}_s - 1/2$ and $|\mathfrak{C}| \leq 1/2$, that is, $\mathfrak{C}$ is empty and $\mathfrak{B}$ is a true maximal clique set. Thus, Theorem 3.1 can be used

to show that the s-graded model conditioned on $\mathscr{L}_n(\tilde{\delta})$ will include a maximal clique set housing a clique of at least $\tilde{\delta}\tilde{\mu}_s(1 - o(1))$ edges. Unfortunately, the quantitative estimate on the probability of $\mathscr{G}_{n,\tilde{\delta}}(\tilde{\varepsilon})^c \cap \mathscr{L}_n(\tilde{\delta})$ in this case is not sufficiently good to deduce Theorem 2.1. This is the reason for the introduction of the $\tilde{\varepsilon}$-almost maximal clique sets, which allow us to deduce a stronger upper bound on the probability that $\mathscr{G}_{n,\tilde{\delta}}(\tilde{\varepsilon})$ does not occur—at the price of dealing with more flexible geometric constructions.

4. Outline of the proof.

Before embarking on a proper proof, we sketch the main ideas required. We recall that $\delta^* > 0$ is a fixed positive number and that p is given by $\lim_{n\to\infty} \log \mu / \log n$. We will define

$$a := \frac{\delta^*}{25}.$$

Later, we will also pick two positive real numbers α, β as some quantities depending on p and a. All of these quantities will be fixed throughout the paper. We further note that, for any admissible graph, $r \to 0$ as $n \to \infty$. For the remainder of the paper, we will take the statement "n is sufficiently large" to imply that r is sufficiently small.

We begin by carefully analyzing the s-graded model. We order the indices I by the size of X_I, the point counts of the A_I's. Explicitly, we pick a bijection from T to $\{1, 2, \ldots, m^d\}$ such that

$$X_1 \geq X_2 \geq \cdots \geq X_{m^d}.$$

For notational convenience, we set

$$q = (2\tilde{\delta}\tilde{\mu}_s)^{1/2}, \qquad w = \tilde{\tau}_s \cdot \mathscr{D}.$$

Picking a as above, we set $M = \lceil \mathscr{D} \rceil \cdot n^a$, and let $\mathscr{T}_M$ be the greatest I such that $X_I \geq M$. We define

$$\mathfrak{I} = \{1, 2, \ldots, \mathscr{T}_M\}, \quad \text{ordered by size as above,}$$

to be the set of indices whose associated point counts X_I exceed their mean (corrected for integrality) by a fixed polynomial factor. Furthermore, define

$$Y_I := X_I\big(\log(X_I/\mathscr{D}) - 1\big) + \mathscr{D},$$

and

$$Q(\mathfrak{I}) := \frac{2}{q^2} \sum_{I\in\mathfrak{I}}\left[\binom{X_I}{2} + \frac{1}{2}\sum_{\substack{J\in\tilde{N}_I\cap\mathfrak{I}\\J\neq I}} X_I X_J\right]$$

and

$$V(\mathfrak{I}) := \frac{1}{q}\sum_{I\in\mathfrak{I}} X_I.$$

The first quantity is an appropriately chosen convex function of the X_I's, while the second is a scaled version of the number of edges with both endpoints in the A_I's associated with $\mathfrak{I}$, and the third controls the number of vertices in $\mathfrak{I}$.

Let $\xi > 0$ be a fixed constant independent of n. Consider the event

$$\mathscr{H}_\xi = \left\{Q(\mathfrak{I}) \geq 1 - \frac{\xi}{\log n}\right\} \cap \left\{\frac{1}{q}\sum_{I\in\mathfrak{I}} Y_I \leq \log(q/w) - 1 + \xi\right\}.$$

The main probabilistic analysis of this paper occurs in two sections: the first uses large deviation estimates to control sums of i.i.d. random variables, and the second employs concentration inequalities for more complicated functions. Together, this work allows us to show that, for sufficiently large values of n and small values of ξ,

$$\mathbb{P}\left[\mathscr{H}_\xi^c \cap \mathscr{L}_n(\tilde{\delta})\right] \leq 3\exp\left(-q\left[\log\left(\frac{q}{w}\right) - 1 + \xi/2\right]\right).$$

It turns out that, if we set $\xi \leq (\tilde{\varepsilon}/10)^{10}$, *any* $\mathfrak{I} \subset T$ that satisfies both the quadratic lower bound and the convex upper bound that define $\mathscr{H}_\xi$ (as well as a mild bound on $\mathscr{T}_M$ and $V(\mathfrak{I})$) contains in it a $\tilde{\varepsilon}$-almost maximal clique set $\mathfrak{B}$ and an exceptional set $\mathfrak{C}$ that satisfy the four stipulations of $\mathscr{G}_{n,\tilde{\delta}}(\tilde{\varepsilon})$! This nontrivial statement implies $\mathscr{G}_{n,\tilde{\delta}}(\tilde{\varepsilon})^c \subset \mathscr{H}_\xi^c$, whenever $\xi \leq (\tilde{\varepsilon}/10)^{10}$ – and, in particular, $\{\mathscr{G}_{n,\tilde{\delta}}(\tilde{\varepsilon})^c \cap \mathscr{L}_n(\tilde{\delta})\} \subset \{\mathscr{H}_\xi^c \cap \mathscr{L}_n(\tilde{\delta})\}$. This proves Theorem 3.1.

The proof of the above implication is not straightforward, and we will deduce it in several steps. We emphasize that this is a completely deterministic property of configurations that satisfy a certain set of inequalities. The next two paragraphs sketch the argument used to prove this implication.

Set $\mathscr{T}_V$ to be

$$\mathscr{T}_V := \min\left\{k : V(\{1,\ldots,k\}) > 1 - \frac{2\xi}{\log n}\right\} \quad \text{and} \quad \mathfrak{T} := \{1,\ldots,\mathscr{T}_V\}.$$

Careful use of minimality and Jensen's inequality proves that

$$V(\mathfrak{T}) \leq 1 + \phi(\mathscr{T}_V), \qquad Q(\mathfrak{T}) \geq 1 - \psi(\mathscr{T}_V),$$

where $\phi(\cdot)$ and $\psi(\cdot)$ are explicit functions, bounded above by $1/(\log n)^{1/2}$, that are nonincreasing in their arguments. One of the difficulties we encounter is that we do not have good upper bounds on $\mathscr{T}_V$, and thus must have bounds that improve whenever the parameter grows.

We set $\mathscr{T}_P$ to be the greatest integer I smaller than $\mathscr{T}_V$ that satisfies $X_I > \xi q/\tilde{\tau}_s$. Setting $\mathfrak{P} = \{1, 2, \ldots, \mathscr{T}_P\}$, we now have a set of indices whose associated X_I's are commensurate with q. We proceed to show that the diameter of $\mathfrak{P}$ cannot exceed s without violating either the lower bound on $Q(\mathfrak{T})$ or the upper bound on $V(\mathfrak{T})$. Together with technical estimates that force $\mathscr{T}_P \geq \tilde{\tau}_s(1 - \xi^{1/3})$, we find that $\mathfrak{P}$ is an $\xi^{1/3}$-almost maximal clique set. Moreover, a quantitative version of Jensen's inequality allows us to break $\mathfrak{P}$ into $\mathfrak{B}$ and $\mathfrak{C}$, the required sets. Finally, we can show that $X_{\mathscr{T}_P+1}$ vanishes sufficiently quickly to complete the proof of Theorem 3.1.

We then move on to proving that Theorem 3.1 implies Theorem 2.1. To do so, we first show that we can approximate any convex subset S of a ball of diameter r from both the inside and the outside by a union of A_I's using the tools of geometric measure theory. Next, we use the classical isodiametric inequality to show that the A_I's associated with a $s^{-1/20}$-almost maximal clique set approximate a ball of diameter r, in the sense of the Hausdorff metric.

Next, we fix $\varepsilon > 0$, and show that, for sufficiently large s and $\tilde{\delta} \in [(1 - \varepsilon/16)\delta, \delta]$, the event $\mathscr{G}_{n,\tilde{\delta}}(s^{-1/20})$ will imply $\mathscr{F}_n(\varepsilon)$. We then apply Theorem 3.1 with $\tilde{\delta}$ as above and $\tilde{\varepsilon} = s^{-1/20}$ to get an upper bound on the probability of $\{\mathscr{F}_n(\varepsilon)^c \cap \mathscr{L}_n(\tilde{\delta})\}$. Combining this bound with a good lower bound on the probability of $\{|E| > (1+\delta)\mu\}$ (to be derived directly from the Poisson point process) and a well-known correlation inequality gives Theorem 2.1.

The final section of the paper proves the large deviation principle of Theorem 2.2. We use the first stipulation of Theorem 2.1 and the s-graded model to compute the upper bound.

5. Analysis of the *s*-graded model. In this section, we analyze the *s*-graded model and prove Theorem 3.1. At the very beginning, let us now fix a positive integer *s* and numbers $0 < \tilde{\delta}_0 \leq \tilde{\delta} \leq \tilde{\Delta}_0$. We will figure out the universal constant $\tilde{\varepsilon}_0$ later. Throughout, whenever we say "*n* sufficiently large," we will mean "$n \geq n_0$ for some n_0 that depends only on *s*, $\tilde{\varepsilon}$, $\tilde{\delta}_0$, and $\tilde{\Delta}_0$".

5.1. *Controlling the natural parameters of the s-graded model.* The geometric properties of the *s*-graded model are not quite comparable to those of the random geometric graph; most obviously, the *s*-graded model has a discrete geometry induced by the semi-metric $\rho(\cdot, \cdot)$ on T. We begin with a very useful lemma, which tells us that the parameters of the *s*-graded model are close to their appropriate equivalents on $\mathbb{T}^d$. To do so, we define three operators: first, let $\mathfrak{U}$ send a set of indices to the union of their associated A_I's, that is, for any $\mathfrak{J} \subset T$,

$$(5.1) \qquad \mathfrak{U}(\mathfrak{J}) := \bigcup_{I \in \mathfrak{J}} A_I.$$

In the other direction, we must be more careful. Let $K \subset \mathbb{T}^d$, we define $\mathfrak{R}(K)$ and $\mathfrak{D}(K)$ to be the maximal (resp., minimal) subsets of T such that

$$(5.2) \qquad \mathfrak{U}(\mathfrak{R}(K)) \subset K \quad \text{and} \quad K \setminus K' \subset \mathfrak{U}(\mathfrak{D}(K)),$$

where K' is some subset of K of Lebesgue measure 0; this modification allows us to not deal with certain trivialities. We note that $\mathfrak{R}(K)$ may be empty, and $\mathfrak{D}(K)$ may be T, even when K or $\mathbb{T} \setminus K$ are nonempty. Alternatively, we may define $\mathfrak{D}(K)$ by

$$\mathfrak{D}(K) := \{I \in T : \lambda(K \cap \mathfrak{U}(I)) > 0\}.$$

We recall several definitions: $\mu = \mathbb{E}(|E|)$, $\tilde{\mu}_s = \mathbb{E}(|E_s|)$ and $\tau = \nu(r/2)^d$. We set $|\tilde{N}_I|$ to be the number of indices satisfying $\rho(I, J) \leq s$, and $\tilde{S}_I = \{J : \tilde{N}_I \cap \tilde{N}_J \neq \varnothing\}$. Finally, $\tilde{\tau}_s$ is the cardinality of a maximal clique set (as defined in (3.7)).

LEMMA 5.1. *We have $E \subset E_s$, and there exist constants C, s_0 and n_0 depending only on the dimension and the chosen norm of the torus, such that, if $s \geq s_0$ and $n \geq n_0$, then*

$$\mu \leq \tilde{\mu}_s \leq \mu\left(1 + \frac{C}{s}\right)$$

and

$$m^d \tau \leq \tilde{\tau}_s \leq m^d \tau\left(1 + \frac{C}{s}\right).$$

Furthermore, $|\tilde{N}_I|$, $|\tilde{S}_I|$ and $\tilde{\tau}_s$ are uniformly bounded in n.

In this section, we will only use this lemma to establish that certain quantities are uniform in *n*; in Section 6, we will strongly use the fact that the estimates become tight as *s* grows.

PROOF. Pick an arbitrary I and consider $\mathfrak{U}(\tilde{N}_I)$. By definition of $\rho(\cdot, \cdot)$ and s, this set includes a ball of radius r around *any* point in A_I. Therefore, any pair $(v, w) \in E$ must also be in E_s, giving the first stipulation. Since this inclusion holds for any configuration of the underlying Poisson point process, this also gives $\mu \leq \tilde{\mu}_s$.

Now, define ς to be the diameter of the unit cube under the norm $\| \cdot \|$—that is,

$$(5.3) \qquad \varsigma := \sup_{x, y \in [0,1]^d} \|x - y\|.$$

Fix I, and let x and y be two fixed points in A_I and $\mathfrak{U}(\tilde{N}_I)$, respectively. Letting J be an index for which $y \in A_J$, we pick arbitrary points z and w in A_I and A_J, respectively. Then the triangle inequality for $\|\cdot\|$ implies that

$$\|x - y\| \leq \|x - z\| + \|z - w\| + \|w - z\| \leq \|z - w\| + \frac{2\varsigma}{m},$$

where we bound the first and last terms by ς/m using scaling of the norm. Since z and w are arbitrary, we can take an infimum over all choices of z and w in A_I° and A_J°, respectively, and conclude that

$$\|x - y\| \leq \frac{s + 2\varsigma}{m} \leq \frac{(s + 2\varsigma)r}{s - r},$$

where we use that $m \geq s/r - 1$, by definition. Therefore, $\mathfrak{U}(\tilde{N}_I)$ is contained in a ball of radius $r(1 + 3\varsigma/s)$ around any point in A_I, for sufficiently large value of s and n (recalling that r is vanishing in n). Since each A_I is of measure of m^{-d}, we deduce that

$$|\tilde{N}_I| = m^d \lambda(\mathfrak{U}(\tilde{N}_I)) \leq v m^d r^d \left(1 + \frac{3\varsigma}{s}\right)^d \leq v m^d r^d \left(1 + \frac{6d\varsigma}{s}\right),$$

where the final inequality follows because $(1 + x)^d \leq 1 + (2d)x$ for all sufficiently small x. Substituting this into the definition of $\tilde{\mu}_s$ produces the desired inequality on $\tilde{\mu}_s$. Repeating a similar analysis will show that the set $\mathfrak{U}(\{J : \tilde{N}_I \cap \tilde{N}_J \neq \varnothing\})$ is a subset of some ball of radius $2r(1 + 3\varsigma/s)$, and thus

$$|\tilde{S}_I| \leq v m^d (2r)^d \left(1 + \frac{6d\varsigma}{s}\right).$$

Next, we wish to control $\tilde{\tau}_s$. For the lower bound, let $B \subset \mathbb{T}^d$ be an arbitrary ball (in $\|\cdot\|$) of diameter r. Consider $\mathfrak{O}(B)$. By minimality, $\lambda(A_I \cap B) > 0$ for every $I \in \mathfrak{O}(B)$. Therefore,

$$\max_{I, J \in \mathfrak{O}(B)} \left[\inf_{x \in A_I^\circ, y \in A_J^\circ} \|x - y\| \right] \leq r,$$

which implies, by the definition of $\rho(\cdot, \cdot)$, that the diameter $\mathfrak{O}(B)$ is at most s. Meanwhile, by inclusion, and the fact that $\lambda(A_I) = 1/m^d$ for every I,

$$|\mathfrak{O}(B)| > m^d \lambda(B) = m^d \tau,$$

completing the lower bound.

For the upper bound, pick any $\mathfrak{W} \subset T$ such that

$$\lambda(\mathfrak{U}(\mathfrak{W})) \geq \tau \left(1 + \frac{C}{s}\right).$$

Applying the isodiametric inequality for finite dimensional normed spaces [6], page 93, and choosing C and s_0 sufficiently large gives

$$\text{diam}(\mathfrak{U}(\mathfrak{W})) \geq r \left(1 + \frac{C}{s}\right)^{1/d} \geq r \left(1 + \frac{4\varsigma}{s - r}\right).$$

This implies that the diameter of $\mathfrak{W}$ is at least $s + 1$. Therefore, any set $\mathfrak{W}$ of diameter at most s must satisfy $\lambda(\mathfrak{U}(\mathfrak{W})) < \tau(1 + c/S)$, and

$$|\mathfrak{W}| = m^d \cdot \lambda(\mathfrak{U}(\mathfrak{W})) \leq m^d \tau \left(1 + \frac{C}{s}\right),$$

as required.

The uniform bounds on $|\tilde{N}_I|$, $|\tilde{S}_I|$, and $\tilde{\tau}_s$ follow from $m^d \leq s^d/r^d$ and the above formulae. □

An immediate corollary to this theorem is that, assuming (2.3), $\tilde{\mu}_s = n^{p + o(1)}$.

5.2. *Large deviation estimates.* The probabilistic bounds we need in this work are divided into two parts. The first involves good control on the deviation of sums of i.i.d. random variables. Our main tools here will be Chernoff bounds, as well as exact lower bounds.

Recall from Section 4 that $q = (2\tilde{\delta}\tilde{\mu}_s)^{1/2}$, $w = \tilde{\tau}_s \cdot \mathscr{D}$ and $\mathscr{L}_n(\tilde{\delta}) = \{|E_s| > (1 + \tilde{\delta})\tilde{\mu}_s\}$. By our assumptions on $\tilde{\delta}$, (3.2), (3.5) and Lemma 5.1, we have that $q = n^{p/2+o(1)}$ and $w = n^{p-1+o(1)}$. Since $p/2 > p-1$ for any admissible s-graded model, we can increase n to ensure that $q > 3w$. We will assume this inequality for the rest of the paper.

We begin by recalling some classical bounds on the Poisson distribution (for proof, see [13], page 35, e.g.).

LEMMA 5.2. *Let X_I be a Poisson random variable with mean $\mathscr{D}$. Then, for any $t > \mathscr{D}$,*

$$\mathbb{P}[X_I \geq t] \leq \exp(-t[\log(t/\mathscr{D}) - 1] - \mathscr{D}),$$

and for any $t < \mathscr{D}$,

$$\mathbb{P}[X_I \leq t] \leq \exp(-t[\log(t/\mathscr{D}) - 1] - \mathscr{D}).$$

These bounds, which are given by explicitly computing exponential moment generating functions, are tight up to polynomial factors, and will be very important in the nearly exact computations we do in the proceeding lemmas.

We now define a random ordering of T according to the X_I. Specifically, we pick a bijection from T to $\{1, 2, \ldots, m^d\}$ such that

$$X_1 \geq X_2 \geq \cdots \geq X_{m^d}.$$

This bijection is not unique, as each X_I is integer-valued, and there may be many I's whose associated X_I's are equal. However, all the statements will be true independently of the particular choice of bijection. Next, fix $a = \delta^*/25$, and define M by

(5.4) $$M := \lceil \mathscr{D} \rceil \cdot n^a.$$

The number M is defined so to be a threshold of density for X_I—if $X_I < M$, we say it is in the *bulk* of the graph. We expect that, even conditional on $\mathscr{L}_n(\tilde{\delta})$, most indices I will be in the bulk. To formalize this, we let

$$\mathscr{T}_M := \max\{I : X_I \geq M\}.$$

The next proposition controls the tail of $\mathscr{T}_M$.

PROPOSITION 5.3. *Let*

$$\alpha = \min\{1 - p/2 - a/2, p/2 - a/2\}.$$

Let $\mathscr{A}$ be the event $\{\mathscr{T}_M \geq n^\alpha\}$. Then, for all sufficiently large n,

$$\mathbb{P}[\mathscr{A}] \leq \exp(-n^{p/2+a/3}).$$

The number α will remain fixed to the value above for the remainder of the paper. We note that $\alpha < 2 - p$ for any admissible s-graded model and, therefore, $n^\alpha \ll m^d$ (using (3.1)). Thus, we find that, with very high probability, the complement of the bulk takes up a vanishing proportion of T.

PROOF OF PROPOSITION 5.3. The event $\mathscr{A}$ implies the existence of some $\mathfrak{W} \subset T$ such that, for all $I \in \mathfrak{W}$, $X_I > M$, and $|\mathfrak{W}| > \lceil n^\alpha \rceil$. By the union bound,

$$\mathbb{P}[\mathscr{A}] \leq \binom{m^d}{\lceil n^\alpha \rceil} \cdot \mathbb{P}[X_I > M]^{n^\alpha}.$$

Using the upper tail bound in Lemma 5.2 and the brutal bound $\binom{m}{k} < m^k$, this implies that

$$\mathbb{P}[\mathscr{A}] \leq m^{d(n^\alpha+1)} \cdot \exp(-n^\alpha M[\log(M/\mathscr{D}) - 1])$$

$$\leq \exp(d(n^\alpha + 1)\log m - n^{\alpha+a}\lceil\mathscr{D}\rceil).$$

Since $\log m$ is bounded above by $C\log n$ for some uniform constant C (by Lemma 5.1), we can increase n to ensure that

$$\mathbb{P}[\mathscr{A}] \leq \exp\left(-\frac{n^{\alpha+a} \cdot \lceil\mathscr{D}\rceil}{2}\right).$$

If $\mathscr{D} \leq 1$, then the ceiling function is 1 and $\alpha = p/2 - a/2$. Increasing n until $n^{p/2+a/2}/2 > n^{p/2+a/3}$ completes this case. If $\mathscr{D} > 1$, then we bound $\lceil\mathscr{D}\rceil$ by $\mathscr{D}$ itself. By definition,

$$\mathscr{D}n^{\alpha+a} = n^{\min\{3p/2-1+a/2, p/2+a/2\}+o(1)}.$$

In the case $p \geq 1$, the exponent is always minimized by the second choice. This completes the proof. □

   The second estimate of this section will be used to control the behavior of the elements outside the bulk. Define

$$Y_I = X_I(\log(X_I/\mathscr{D}) - 1) + \mathscr{D},$$

with the convention that $0 \cdot \log 0 = 0$. Note that $Y_I = \mathscr{I}(X_I)$, where $\mathscr{I}$ is the rate function of a Poisson random variable of mean $\mathscr{D}$. This implies that $Y_I \geq 0$ and vanishes only at $\mathscr{D}$. $\mathscr{I}$ is a convex function, and thus we can bound the sum of the Y_I's by a function of the sum of the X_I's, using Jensen's inequality. Furthermore, $\mathbb{P}[Y_I > t]$ should vanish as $\exp(-t)$, by "inverting" the rate function. We formalize this notion in the lemma below.

   LEMMA 5.4.   *For any $\mathscr{D}$, and any positive $\lambda < 1$,*

$$\mathbb{E}[\exp(\lambda Y_I)] \leq \frac{1+\lambda}{1-\lambda}.$$

   PROOF.   The function $\mathscr{I}(x) = x[\log(x/\mathscr{D}) - 1] + \mathscr{D}$ is not invertible, but is piecewise invertible. First, let

$$g_1(x) : [0, \mathscr{D}] \to [0, \mathscr{D}] \text{ be a function such that } (\mathscr{I} \circ g_1)(x) = x.$$

Note that this function is decreasing, with $g_1(0) = \mathscr{D}$ and $g_1(\mathscr{D}) = 0$. For any $x > \mathscr{D}$, we say that $g_1(x) = -\infty$. We define g_2, the second inverse, similarly, except its domain is defined to be $(\mathscr{D}, \infty)$. This inverse is strictly increasing. Thus,

$$\mathbb{P}[Y_I \geq t] = \mathbb{P}[X_I \leq g_1(t)] + \mathbb{P}[X_I \geq g_2(t)].$$

By appealing to the two bounds of Lemma 5.2, we find that both the probabilities above are bounded above by $\exp(-t)$; in fact, if $t > \mathscr{D}$, the first probability is identically zero. Regardless, it will suffice to use the bound $\mathbb{P}[Y_I > t] < 2e^{-t}$. Thus, for any $\lambda < 1$,

$$\mathbb{E}[\exp(\lambda Y_I)] = 1 + \int_1^\infty \mathbb{P}\left[Y_I > \frac{\log t}{\lambda}\right] dt$$

$$\leq 1 + 2\int_1^\infty t^{-1/\lambda} dt$$

$$= \frac{1+\lambda}{1-\lambda},$$

as required. □

We now uses the lemma to control the upper tail of the sum of the Y_I's over *any* sufficiently small subset of T. We will only apply the proposition below on the set $\{1, 2, \ldots, \mathscr{T}_M\}$ (which will be small with good probability from Proposition 5.3), but it is actually more straightforward to consider the existence of a subset with bad properties, in order to avoid conditional probabilities.

PROPOSITION 5.5. *Let Y_I as above, and α as in Proposition 5.3. Define the event*

$$\mathscr{B}_\xi := \left\{ \exists \mathfrak{W} \subset T, |\mathfrak{W}| \leq n^\alpha \text{ such that } \sum_{I \in \mathfrak{W}} Y_I > q\big(\log(q/w) - 1\big) + \xi q \right\}.$$

Then, for all sufficiently large n,

$$\mathbb{P}[\mathscr{B}_\xi] \leq \exp\big(-q[\log(q/w) - 1 + \xi/2]\big).$$

PROOF. Set $t = q(\log(q/w) - 1 + \xi)$. Fix $\mathfrak{W} \subset T$ with cardinality at most n^α. By Chebyshev's inequality,

$$\mathbb{P}\left[\sum_{I \in \mathfrak{W}} Y_I > t\right] \leq \frac{(\exp(\lambda Y_I))^{n^\alpha}}{\exp(\lambda t)}$$

$$\leq \left(\frac{1 + \lambda}{1 - \lambda}\right)^{n^\alpha} \cdot e^{-\lambda t},$$

where the second inequality is Lemma 5.4.

We now set $\lambda = 1 - n^\alpha/t$, noting that, for sufficiently large n, $\lambda > 0$ (since $n^\alpha \ll q$). This turns the above estimate into

$$\mathbb{P}\left[\sum_{I \in \mathfrak{W}} Y_I > t\right] \leq \big(2t/n^\alpha\big)^{n^\alpha} \cdot e^{-t + n^\alpha}$$

$$\leq \exp\big(-t + Cn^\alpha \log n\big).$$

The final step is to apply the union bound:

$$\mathbb{P}[\mathscr{B}] \leq \sum_{k=1}^{\lfloor n^\alpha \rfloor} \binom{m^d}{k} \cdot e^{-t + Cn^\alpha \log n}$$

$$\leq n^\alpha \binom{m^d}{n^\alpha} \cdot e^{-t + Cn^\alpha \log n}$$

$$\leq m^{d(n^\alpha + 1)} \cdot e^{-t + Cn^\alpha \log n}.$$

Recalling Lemma 5.1, we see that the combinatorial term in the final inequality is bounded above by $\exp(Cn^\alpha \log n)$ for some (probably different) C. Since $q \gg n^\alpha \log n$, the entire positive contribution can be bounded above by $\xi q/2$. This completes the proof. $\square$

5.3. *Concentration inequalities.* This section will prove concentration of the edge count of the random geometric graph *restricted to the bulk*. Explicitly, let

$$\hat{X}_I := X_I \cdot 1_{X_I < M}$$

and $|\hat{E}_s|$ be the define analogously with $|E_s|$ by replacing X_I with its truncated version (recall that $M = \lceil \mathscr{D} \rceil \cdot n^a$). In other words, $|\hat{E}_s|$ is the version of the edge count of G_s obtained after deleting all vertices lying in A_I's that satisfy $X_I \geq M$.

For the rest of the paper, fix

$$\beta = p - 2a.$$

Consider the event

$$\mathscr{C} = \{|\hat{E}_s| - \tilde{\mu}_s > n^\beta\}.$$

We control the probability of $\mathscr{C}$ in two regimes. We begin by assuming that $\mathscr{D} < \log n$.

Our strategy for proving an upper bound on $\mathscr{C}$ in this regime relies on Talagrand's convex concentration inequality [38], Theorem 4.1.1. First, let us define the setting: let $\Omega = \prod_{i=1}^N \Omega_i$, where Ω_i are all probability spaces and the measure $\mathbb{P}$ on Ω is the product measure. For a set $A \subset \Omega$, define the set

$$U_A(x) := \{\{s_i\} \in \{0, 1\}^N : \exists y \in A, s_i = 0 \implies x_i = y_i\}.$$

Let $V_A(x)$ be the convex hull of $U_A(x)$, and $d_c(A, x)$ is the ℓ^2 distance of $V_A(x)$ to the origin. For any set A, we denote A_t be the t blowup of A with respect to this metric, that is,

$$A_t := \{x \in \Omega : d_c(A, x) \le t\}.$$

We can now state the inequality.

THEOREM 5.6 (Talagrand's inequality [38]).   *If* Ω, $\mathbb{P}[\cdot]$, *A and* A_t *are as above, then*

$$\mathbb{P}[A](1 - \mathbb{P}[A_t]) \le e^{-t^2/4}.$$

We will not apply this theorem directly; instead, we use a corollary of this theorem frequently used in discrete settings [1], Theorem 7.7.1. To do so, we consider a random variable X defined on the space Ω, and a function f from the natural numbers to the natural numbers. We say that f is a witness function for X if, whenever $X(\omega) \ge t$, there exists $I \subset [n]$ with $|I| \le f(t)$, such that every ω' that agrees with ω in all $i \in I$ has $X(\omega') \ge t$. Furthermore, we assume that $X(\omega)$ is K-Lipschitz with respect to the Hamming distance, that is, $|X(\omega) - X(\omega')| \le K$ whenever ω and ω' differ in at most one coordinate.

THEOREM 5.7 ([1]).   *Let* Ω *be a product space, and* X *a real valued function on* Ω *with Lipschitz constant* K *with respect to the Hamming distance. If* f *is witness function for* X *as above, then, for any* b *and* t,

$$\mathbb{P}[X > b + tK\sqrt{f(b)}]\mathbb{P}[X \le b] \le \exp(-t^2/4).$$

With this preliminary complete, we will prove the following lemma.

LEMMA 5.8.   *Let* $\mathscr{C}$ *be as above, and assume that* $\mathscr{D} < \log n$. *Then, for all sufficiently large* n,

$$\mathbb{P}[\mathscr{C}] \le \exp(-n^{2\beta - p - 6a}).$$

PROOF.   Thanks to our assumption on $\mathscr{D}$, $M < n^a \log n$. We now apply Theorem 5.7 to $X = |\hat{E}_s|$, considered as a function of the X_I's. Since each coordinate is bounded above by $n^a \log n$, X is Lipschitz with $K \le |\tilde{N}_I| n^{2a} \log^2 n$. The function $f(w) = 2w$ is a witness function for $|\hat{E}_s|$; to see this, note that $|\hat{E}_s|$ is the edge count of the s-graded geometric random graph, after we remove any X_I with very high density. As such, we can "witness" the existence of w edges by finding at most $2w$ vertices; the flexibility of the setup allows us to pick these vertices judiciously, avoiding all the isolated ones. Finding $2w$ vertices will

require at most $2w$ distinct coordinates, if each one of them vertices lies in a distinct A_I. Note that this bound may be very loose—whenever $2w > m^d$, we can easily just check *every* A_I to witness $|\hat{E}_s| > w$.

We apply the theorem with $b = \tilde{\mu}_s + n^\beta / \log n$ and

$$t = \frac{n^\beta (1 - 1/\log n)}{|\tilde{N}_I| n^{2a} \log^2 n \cdot [2\tilde{\mu}_s + 2n^\beta / \log n]^{1/2}}.$$

We deduce that

$$\mathbb{P}[\mathscr{C}] \cdot \mathbb{P}[|\hat{E}_s| \le \tilde{\mu}_s + n^\beta / \log n]$$

(5.5)
$$\le \exp\left(-\frac{n^{2\beta}[1 - 1/\log n]^2}{8|\tilde{N}_I|^2 n^{4a} \log^4 n [\tilde{\mu}_s + n^\beta / \log n]}\right)$$

$$\le \exp(-n^{2\beta - p - 5a}),$$

where the final inequality holds for sufficiently large n, using the fact that $\tilde{\mu}_s = n^{p+o(1)} \gg n^\beta / \log n$ and the fact that $|\tilde{N}_I|$ is uniformly bounded in n.

To complete the proof, we must show that $\mathbb{P}[|\hat{E}_s| \le \tilde{\mu}_s + n^\beta / \log n]$ is not too small. Since the mean of $|\hat{E}_s|$ is strictly smaller than the mean of $|E_s|$, it is enough to show that

$$\mathbb{P}[|\hat{E}_s| - \mathbb{E}[|\hat{E}_s|] \ge n^\beta / \log n] < \varepsilon.$$

We will produce a very crude bound on the variance of $|\hat{E}_s|$: let $\hat{Z}_I = \hat{X}_I(\sum_{J \in \tilde{N}_I} \hat{X}_J - 1)$. Then, clearly,

$$\mathrm{Var}[|\hat{E}_s|] = \sum_{I,J} \mathbb{E}[(\hat{Z}_I - \mathbb{E}[\hat{Z}_I])(\hat{Z}_J - \mathbb{E}[\hat{Z}_J])]$$

$$\le \sum_I |\tilde{S}_I| \cdot \mathbb{E}[(\hat{Z}_I - \mathbb{E}[\hat{Z}_I])^2].$$

A straightforward computation will show that, for some constant C independent of n,

$$\mathbb{E}[(\hat{Z}_I - \mathbb{E}[\hat{Z}_I])^2] \le C(\mathscr{D}^2 + \mathscr{D}^4).$$

Since $\mathscr{D} < \log n$ and $|\tilde{S}_I|$ is uniformly bounded in n (from Lemma 5.1), this implies that $\mathrm{Var}[|\hat{E}_s|] < n^{p+o(1)}$, and Chebyshev's inequality gives that

$$\mathbb{P}[|\hat{E}_s| - \mathbb{E}[|\hat{E}_s|] \ge n^\beta / \log n] \le n^{p - 2\beta + o(1)}.$$

For any admissible value of p, this function vanishes as n increases, and $\mathbb{P}[|\hat{E}_s| \le \tilde{\mu}_s + n^\beta / \log n] > 1 - \varepsilon$. Substituting this into (5.5) gives us

$$\mathbb{P}[\mathscr{C}] \le \exp(-n^{2\beta - p - 6a}),$$

completing the proof. $\square$

Next, assume that $\mathscr{D} \ge \log n$. In this regime, we replace Talagrand's inequality with the celebrated Azuma–Hoeffding inequality.

THEOREM 5.9 (Azuma–Hoeffding inequality [4, 22]). *Let $\{Z_0, Z_1, \ldots, Z_n\}$ be a martingale sequence with $|Z_k - Z_{k-1}| < c_k$ for all k. Then*

$$\mathbb{P}[|Z_n - Z_0| > t] \le 2 \exp\left(-\frac{t^2}{2 \sum_{k=1}^n c_k^2}\right).$$

We wish to prove the following lemma, bounding the probability of the event $\mathscr{C}$.

LEMMA 5.10.   *Assume that $\mathscr{D} \geq \log n$. Then, for all n sufficiently large,*

$$\mathbb{P}[\mathscr{C}] \leq \exp\left(-\frac{n^{2\beta}}{m^d \mathscr{D}^3 n^{6a}}\right).$$

We note that a naive application of Azuma–Hoeffding to the martingale given by conditioning on the value of $\hat{X}_I$ would give a bound on the probability of $\mathscr{C}$ which depends on the *fourth* power of $\mathscr{D}^{-1}$, not the third as in the lemma—an inferior bound. Thus, we need to be more careful in this analysis.

PROOF.   We partition A_I into sets of measure $1/n$; formally, let $\{F_{I,t}\}$, for natural $t \leq \lceil \mathscr{D} \rceil$ be a collection of disjoint subsets of A_I such that $\lambda(F_{I,t}) = 1/n$ for every $t \leq \lceil \mathscr{D} \rceil - 1$, and

$$\bigcup_t F_{I,t} = A_I.$$

Note that the measure of the final $F_{I,t}$ will be strictly smaller than $1/n$, unless $\mathscr{D}$ is an integer. We define $W_{I,t} = |\chi(F_{I,t})|$.

Clearly, $\sum_t W_{I,t} = X_I$. Define $\overline{|E_s|}$ as (yet another!) truncation of $|E_s|$. Specifically, let $\overline{W_{I,t}} = W_{I,t} \cdot 1_{W_{I,t} < n^a/2}$, and define $\overline{|E_s|}$ by replacing each X_I in the definition of $|E_s|$ by $\sum_t \overline{W_{I,t}}$. Note that $\overline{|E_s|}$ is a function of $m^d \lceil \mathscr{D} \rceil$ independent random variables. Letting $\mathscr{F}'_\ell$ be the σ-algebra generated by the first ℓ $W_{I,t}$'s (enumerated arbitrarily), we once again have a martingale sequence $Z'_\ell = \mathbb{E}[\overline{|E_s|} \mid \mathscr{F}'_\ell]$. We also have that

$$|Z'_\ell - Z'_{\ell-1}| \leq |\tilde{N}_I| \cdot \lceil \mathscr{D} \rceil n^{2a}.$$

Thus, Azuma–Hoeffding implies that

$$\mathbb{P}[\overline{|E_s|} - \tilde{\mu}_s > n^\beta/2] \leq \mathbb{P}[\overline{|E_s|} - \mathbb{E}[\overline{|E_s|}] > n^\beta/2]$$

$$\leq 2\exp\left(-\frac{n^{2\beta}}{8m^d \lceil \mathscr{D} \rceil^3 |\tilde{N}_I|^2 n^{4a}}\right),$$

where the first line follows since $\tilde{\mu}_s > \mathbb{E}[\overline{|E_s|}]$. Using the uniform bound on $|\tilde{N}_I|$, we deduce that

$$\mathbb{P}[\overline{|E_s|} - \tilde{\mu}_s > n^\beta/2] \leq \exp\left(-\frac{n^{2\beta + o(1)}}{m^d \mathscr{D}^3 n^{4a}}\right)$$

$$\leq \exp\left(-\frac{n^{2\beta}}{m^d \mathscr{D}^3 n^{5a}}\right).$$

By partitioning,

(5.6) $$\mathbb{P}[\mathscr{C}] \leq \mathbb{P}[\overline{|E_s|} - \tilde{\mu}_s > n^\beta/2] + \mathbb{P}[\mathscr{C}, \overline{|E_s|} - \tilde{\mu}_s < n^\beta/2].$$

The second event on the right-hand side implies the event

$$\mathscr{E} := \{|\hat{E}_s| - \overline{|E_s|} > n^\beta/2\}.$$

The lemma will follow if we can produce a good upper bound on the probability of the event $\mathscr{E}$.

The difference between the random variables $|\hat{E}_s|$ and $\overline{|E_s|}$ is given by configurations in which at least $W_{I,t}$ is larger than $n^a/2$. In fact,

$$|\hat{E}_s| - \overline{|E_s|} < \sum_{(I,t)} \left[(W_{I,t} \cdot 1_{W_{I,t}>n^a/2}) \cdot \sum_{J:\rho(I,J)\leq s} (X_J \cdot 1_{X_J < \lceil \mathscr{D} \rceil n^a}) \right].$$

While the random variables in the expression above are far from independent, we can replace the second sum over the X_J's by $|\tilde{N}_I| \cdot \lceil \mathscr{D} \rceil n^a$, the upper bound imposed on it by the indicator random variables involved. Therefore,

$$\mathbb{P}[\mathscr{E}] \leq \mathbb{P}\left[\sum_{(I,t)} (W_{I,t} \cdot 1_{W_{I,t}>n^a/2}) > \frac{Cn^\beta}{\mathscr{D}n^a} \right].$$

To bound this final probability, we can directly bound the exponential moment of $W_{I,t} \cdot 1_{W_{I,t}>n^a/2}$:

$$\mathbb{E}\left[\exp(W_{I,t} \cdot 1_{W_{I,t}>n^a/2}) \right] \leq 1 + \sum_{k>n^a/2} \frac{e^k}{k!} \leq 1 + \exp(-n^a).$$

The first inequality follows because $W_{I,t}$ is a Poisson random variable of mean 1 (or possibly less than 1, if we pick the small $W_{I,t}$ in each I), while the second can deduced by using Stirling's approximation and explicitly summing.

Applying a Chernoff strategy, we find that

$$\mathbb{P}\left[\sum_{(I,t)} (W_{I,t} \cdot 1_{W_{I,t}>n^a/2}) > \frac{Cn^\beta}{\mathscr{D}n^a} \right] \leq (1 + \exp(-n^a))^{m^d(\mathscr{D}+1)} \cdot \exp\left(-\frac{Cn^\beta}{\mathscr{D}n^a} \right).$$

Using the standard approximation $(1 + x) \leq e^x$, we find that the prefactor is bounded by 2 for all n sufficiently large.

Substituting the bounds into (5.6) gives

$$\mathbb{P}[\mathscr{C}] \leq \exp\left(-\frac{n^{2\beta}}{m^d \mathscr{D}^3 n^{5a}} \right) + 2\exp\left(-\frac{Cn^\beta}{\mathscr{D}n^a} \right).$$

The first term vanishes like $\exp(-n^{1-9a+o(1)})$, whereas the second vanishes as $\exp(-n^{1-7a+o(1)})$. Therefore, we conclude that

$$\mathbb{P}[\mathscr{C}] \leq \exp\left(-\frac{n^{2\beta}}{m^d \mathscr{D}^3 n^{6a}} \right),$$

as required. $\square$

5.4. *Reducing to deterministic inequalities.*

The final probabilistic step of this proof involves bounding the probability of a rather complicated set of simultaneous inequalities. Luckily, instead of dealing with the event itself, we will control it using the events $\mathscr{A}$, $\mathscr{B}_\xi$, and $\mathscr{C}$, whose probability we controlled previously.

Recall that

$$V(\mathfrak{W}) = \frac{1}{q} \sum_{I \in \mathfrak{W}} X_I.$$

We also recall from the outline that, for any $\mathfrak{W} \subset T$, we set

$$(5.7) \qquad Q(\mathfrak{W}) := \frac{2}{q^2} \sum_{I \in \mathfrak{W}} \left(\binom{X_I}{2} + \frac{1}{2} \sum_{\substack{J \in \tilde{N}_I \cap \mathfrak{W} \\ J \neq I}} X_I X_J \right).$$

This counts the number of edges with both endpoints in A_I's with $I \in \mathfrak{W}$, normalized b $q^2/2$; this choice will avoid many unnecessary factors in our later analysis. This immediately implies that, for any $\mathfrak{W} \subset T$,

$$Q(\mathfrak{W}) \leq [V(\mathfrak{W})]^2.$$

Using this notation, we can formulate Jensen's inequality in the following way.

LEMMA 5.11.   *For any $\mathfrak{W} \subset T$,*

$$\frac{1}{q} \sum_{I \in \mathfrak{W}} Y_I \geq V(\mathfrak{W}) \left[\log\left(\frac{q}{w}\right) + \log V(\mathfrak{W}) - \log\left(\frac{|\mathfrak{W}|}{\tilde{\tau}_s}\right) - 1 \right].$$

PROOF.   This is a direct application of Jensen's inequality to the convex function $Y_I = X_I(\log(X_I/\mathscr{D}) - 1) + \mathscr{D}$. It implies that

$$\sum_{I \in \mathfrak{W}} Y_I \geq q V(\mathfrak{W}) \left[\log\left(\frac{qV(\mathfrak{W})}{|\mathfrak{W}|\mathscr{D}}\right) - 1 \right] + \mathscr{D}|\mathfrak{W}|.$$

Dividing through by q, using the definition of w, and ignoring the positive additive term $\mathscr{D}|\mathfrak{W}|/q$ gives the bound above. $\square$

We now begin the part of the analysis where we will obtain the universal constant $\tilde{\varepsilon}_0$ in the statement of Theorem 3.1. For that, we first introduce a parameter $\xi > 0$. We will eventually define $\tilde{\varepsilon}$ in terms of ξ. In the following, wherever we say "ξ sufficiently small," we mean "$\xi \leq \xi_0$ for some universal constant ξ_0." The meaning of "n sufficiently large" will remain as it is.

Recall that $\mathscr{T}_M := \max\{I : X_I \geq M\}$, and let $\mathfrak{I} := \{1, 2, \ldots, \mathscr{T}_M\}$. Given $\xi > 0$ and a constant C which does not depend on n, we define $\mathscr{H}_\xi$ to be the event that the following four inequalities hold:

$$(5.8) \qquad\qquad\qquad\qquad \mathscr{T}_M < n^\alpha,$$

$$(5.9) \qquad\qquad\qquad \frac{1}{q} \sum_{I \in \mathfrak{I}} Y_I \leq \log(q/w) - 1 + \xi,$$

$$(5.10) \qquad\qquad\qquad\qquad V(\mathfrak{I}) < C,$$

$$(5.11) \qquad\qquad\qquad\qquad Q(\mathfrak{I}) \geq 1 - \frac{\xi}{\log n}.$$

PROPOSITION 5.12.   *Define $\mathscr{H}_\xi$ as above. Then, for all n sufficiently large and ξ sufficiently small,*

$$\mathbb{P}\big[\mathscr{H}_\xi^c \cap \mathscr{L}_n(\tilde{\delta})\big] \leq 3\exp\big(-q[\log(q/w) - 1 + \xi/2]\big).$$

PROOF.   Suppose that $\mathscr{A}^c \cap \mathscr{B}_\xi^c \cap \mathscr{C}^c \cap \mathscr{L}_n(\tilde{\delta})$ implies $\mathscr{H}_\xi$. If this were true, then the union bound would imply that

$$\mathbb{P}\big[\mathscr{H}_\xi^c \cap \mathscr{L}_n(\tilde{\delta})\big] \leq \mathbb{P}[\mathscr{A}] + \mathbb{P}[\mathscr{B}_\xi] + \mathbb{P}[\mathscr{C}].$$

The right-hand side of the proposition above can be expressed as $e^{-n^{p/2+o(1)}}$. Therefore, for all sufficiently large n, Proposition 5.3 implies that

$$\mathbb{P}[\mathscr{A}] \cdot \exp(q[\log(q/w) - 1 + \xi/2]) \leq \exp(-n^{p/2+a/3} + n^{p/2+o(1)}) < 1,$$

where the final inequality follows because the first exponent is larger than $p/2$ for all admissible values of p. Similar analysis of exponents and Lemma 5.8 tell us that, if $\mathscr{D} < \log n$ and n is large,

$$\mathbb{P}[\mathscr{C}] \cdot \exp(q[\log(q/w) - 1 + \xi/2] \leq \exp(-n^{2\beta - p - 6a} + n^{p/2 + o(1)}) < 1.$$

When $\mathscr{D} > \log n$, $\mathbb{P}[\mathscr{C}] < \exp(-n^{1 - 10a - o(1)})$ by Lemma 5.10, and, for all n sufficiently large,

$$\mathbb{P}[\mathscr{C}] \cdot \exp(q[\log(q/w) - 1 + \xi/2]) \leq \exp(-n^{1 - 10a} + n^{p/2 + o(1)}) < 1.$$

Finally, Proposition 5.5 bounds the probability of $\mathscr{B}_\xi$ by the right-hand side. We conclude that, if $\mathscr{A}^c \cap \mathscr{B}_\xi^c \cap \mathscr{C}^c \cap \mathscr{L}_n(\tilde{\delta})$ implies $\mathscr{H}_\xi$,

$$\mathbb{P}[\mathscr{H}_\xi^c \cap \mathscr{L}_n(\tilde{\delta})] \leq 3 \exp(-q[\log(q/w) - 1 + \xi/2]).$$

For the remainder of this proof, we condition on the event $\mathscr{A}^c \cap \mathscr{B}_\xi^c \cap \mathscr{C}^c \cap \mathscr{L}_n(\tilde{\delta})$. The event $\mathscr{A}^c$ automatically implies (5.8). With this bound on $\mathscr{T}_M$, (5.9) is immediate from $\mathscr{B}_\xi^c$. Next, we apply Lemma 5.11 to the sum of the Y_I's over $I \in \mathfrak{J}$ to deduce that

$$V(\mathfrak{J})\left[\log\left(\frac{q\tilde{\tau}_s}{w \cdot \mathscr{T}_M}\right) - 1\right] + V(\mathfrak{J})\log V(\mathfrak{J}) \leq \log(q/w) - 1 + \xi.$$

If $V(\mathfrak{J}) \leq 1$, we have a (much better than needed) bound on $V(\mathfrak{J})$. Otherwise, $V(\mathfrak{J})\log V(\mathfrak{J})$ is positive, and we conclude that

$$V(\mathfrak{J}) \leq \frac{\log(q/w) - 1 + \xi}{\log(q\tilde{\tau}_s/(w \cdot \mathscr{T}_M)) - 1}.$$

By (5.8) and the definitions of the variables q, w and α,

$$(5.12) \qquad\qquad\qquad\qquad \frac{q\tilde{\tau}_s}{w \cdot \mathscr{T}_M} \geq n^{a/2 + o(1)}.$$

Therefore, the denominator grows at least as a constant multiple of $\log n$. Meanwhile, $q/w \leq n^{1 - p/2 + o(1)}$. This proves (5.10).

For the final stipulation, we use the event $\mathscr{L}_n(\tilde{\delta}) \cap \mathscr{C}^c$. By $\mathscr{C}^c$,

$$(5.13) \qquad\qquad\qquad\qquad (q^2/2) \cdot Q(\mathfrak{J}^c) - \tilde{\mu}_s \leq n^\beta.$$

By the occurrence of $\mathscr{L}_n(\tilde{\delta})$,

$$\frac{q^2}{2}(Q(\mathfrak{J}) + Q(\mathfrak{J}^c)) + \left(\sum_{I \in \mathfrak{J}} \sum_{J \in \tilde{N}_I \cap \mathfrak{J}^c} X_I X_J\right) \geq (1 + \tilde{\delta})\tilde{\mu}_s,$$

where the first term counts edges with both endpoints either in or outside the bulk, whereas the sum counts the number of edges with exactly one endpoint in the bulk. Using the upper bound on $Q(\mathfrak{J}^c)$ given by (5.13), we deduce that

$$(5.14) \qquad\qquad Q(\mathfrak{J}) + \frac{2}{q^2}\left(\sum_{I \in \mathfrak{J}} \sum_{J \in \tilde{N}_I \cap \mathfrak{J}^c} X_I X_J\right) \geq 1 - \frac{2n^\beta}{q^2}.$$

By definition, $X_J \leq M$ whenever $J \notin \mathfrak{J}$ and, therefore,

$$\frac{2}{q^2}\left(\sum_{I \in \mathfrak{J}} \sum_{J \in \tilde{N}_I \cap \mathfrak{J}^c} X_I X_J\right) \leq \frac{2|\tilde{N}_I|M}{q} \cdot V(\mathfrak{J}).$$

Recalling (5.4), we see that

$$\frac{M}{q} = n^{\max\{a - p/2 + o(1),\, p/2 - 1 + a + o(1)\}}.$$

The exponent is always negative and, therefore, for any admissible s-graded model, $M/q < \xi/(2C \log n)$ for sufficiently large n. By (5.10) and Lemma 5.1, $V(\mathfrak{J})$ and $|\tilde{N}_I|$ are uniformly bounded in n. Similarly, we can increase n sufficiently to ensure that $(2n^\beta)/q^2$ is also bounded above by $\xi/(2 \log n)$ (which is possible thanks to the definition of β). Substituting the bounds into (5.14) shows that $\mathscr{A}^c \cap \mathscr{B}^c_\xi \cap \mathscr{C}^c \cap \mathscr{L}_n(\tilde{\delta})$ implies $\mathscr{H}_\xi$, completing the proof. □

5.5. *Controlling the linear sum.*   The rest of the section is dedicated to analyzing configurations in $\mathscr{H}_\xi$. We will condition on this event, that is, assume that the four inequalities in the definition hold, and show that subsets with certain properties exist. We emphasize that the statement will hold for *any* positive integer s, asymptotically in n. The number ξ will be chosen to be sufficiently small for certain estimates to hold.

Define $\mathscr{T}_V$ by

$$(5.15) \qquad \mathscr{T}_V := \min\left\{k : V(\{1, \ldots, k\}) > 1 - \frac{2\xi}{\log n}\right\}.$$

A priori, the set $\mathfrak{T} := \{1, 2, \ldots, \mathscr{T}_V\}$ may include some elements of the bulk. The next lemma proves that not only are all these indices away from bulk, but, in fact, restricting our attention to $\mathfrak{T}$ does not force us to ignore too many edges.

LEMMA 5.13.   *Define $\mathfrak{T}$ as above, and assume that $\mathscr{H}_\xi$ holds. Then, for all n sufficiently large and ξ sufficiently small, the following holds*:

$$\tilde{\tau}_s(1 - \xi^{1/2}) \le \mathscr{T}_V \le \mathscr{T}_M,$$

$$Q(\mathfrak{T}) \ge 1 - \psi(\mathscr{T}_V)$$

and

$$1 - \frac{2\xi}{\log n} \le V(\mathfrak{T}) \le 1 + \phi(\mathscr{T}_V),$$

where

$$\phi(x) = \min\left\{\frac{C[\log(x/\tilde{\tau}_s) + \xi]}{\log n}, \frac{2}{x}\right\}$$

and

$$\psi(x) = \min\left\{\frac{C[1 + \log(x/\tilde{\tau}_s)]}{\log n}, \frac{C'}{x}\right\} + \frac{\xi}{\log n}$$

for some constants C and C' independent of n.

The exact forms of ϕ and ψ are chosen to make the proof more transparent. The important feature of the functions are that ϕ and ψ decrease for large x, providing better bounds whenever $\mathscr{T}_V$ is large. Since we have no *a priori* bound for this cardinality, this will be crucial for later analysis. Furthermore, for any positive x and sufficiently large value of n,

$$(5.16) \qquad \max\{\psi(x), \phi(x)\} \le \frac{1}{\sqrt{\log n}}.$$

PROOF OF LEMMA 5.13.   Since $Q(\mathfrak{J}) \ge 1 - \xi/\log n$ (by (5.11)), we know that

$$V(\mathfrak{J}) \ge \sqrt{Q(\mathfrak{J})} \ge 1 - \xi/\log n,$$

which immediately implies $\mathscr{T}_V \leq \mathscr{T}_M$. Since $Y_I \geq 0$, the upper bound (5.9) in the definition of $\mathscr{H}_\xi$ can be applied to elements of $\mathfrak{T}$. Applying Lemma 5.11 to this set, we deduce that

$$(5.17) \qquad V(\mathfrak{T})\left[\log\left(\frac{q}{w}\right) + \log V(\mathfrak{T}) - \log\left(\frac{\mathscr{T}_V}{\tilde{\tau}_s}\right) - 1\right] \leq \log\left(\frac{q}{w}\right) + \xi - 1.$$

By definition, $V(\mathfrak{T})$ is at least $1 - 2\xi/\log n$. Noting that

$$\log\left(1 - \frac{2\xi}{\log n}\right) \geq \frac{-4\xi}{\log n}$$

for all sufficiently large n, we can conclude that

$$-\left(1 - \frac{2\xi}{\log n}\right)\log\left(\frac{\mathscr{T}_V}{\tilde{\tau}_s}\right) \leq \xi + \frac{2\xi\log(q/w)}{\log n} + \frac{C\xi}{\log n}.$$

We recall that $q/w \leq n^{1-p/2+o(1)}$, and therefore there exists a constant C such that

$$-\log\left(\frac{\mathscr{T}_V}{\tilde{\tau}_s}\right) \leq C\xi.$$

Inverting the negative logarithm gives

$$\mathscr{T}_V \geq \tilde{\tau}_s \exp(-C\xi) \geq \tilde{\tau}_s(1 - C\xi) \geq \tilde{\tau}_s\left(1 - \xi^{1/2}\right),$$

where the final inequality holds for all sufficiently small ξ, as required.

We prove the upper bound on $V(\mathfrak{T})$ by proving a bound that holds for all values of $\mathscr{T}_V$, and then improving it in the case $\mathscr{T}_V < \log n$. We observe that the definition of the ordering of the X_I's implies that $X_{\mathscr{T}_V}$ is equal to the minimum of the set $\{X_1, X_2, \ldots, X_{\mathscr{T}_V}\}$, and thus

$$(5.18) \qquad \frac{X_{\mathscr{T}_V}}{q} \leq \frac{V(\mathfrak{T})}{\mathscr{T}_V}.$$

Furthermore, by minimality of $\mathscr{T}_V$,

$$V\left(\mathfrak{T} \setminus \{X_{\mathscr{T}_V}\}\right) \leq 1 - \frac{2\xi}{\log n}.$$

Recall that $\tilde{\tau}_s \geq 2^d$ (from (3.8)) and, therefore, $\mathscr{T}_V > (1 - \xi^{1/2})\tilde{\tau}_s$ implies that $\mathscr{T}_V \geq 2$ for any ξ sufficiently small. Since $V(\mathfrak{T}) = V(\mathfrak{T} \setminus \{X_{\mathscr{T}_V}\}) + X_{\mathscr{T}_V}/q$, we deduce that

$$V(\mathfrak{T}) \leq \frac{1 - (2\xi)/(\log n)}{1 - 1/\mathscr{T}_V} \leq 1 + \frac{2}{\mathscr{T}_V} - \frac{2\xi}{\log n}.$$

Ignoring the negative contribution $2\xi/\log n$ gives the desired bound.

Whenever $\mathscr{T}_V \geq \log n$, an explicit computation will show that $\phi(\mathscr{T}_V) = 2/\mathscr{T}_V$ for all sufficiently large n. Thus, to complete the bound on $V(\mathfrak{T})$, we may assume that $\mathscr{T}_V < \log n$. We now return to (5.17). If $V(\mathfrak{T}) \leq 1$, we are done. Otherwise, $V(\mathfrak{T})\log V(\mathfrak{T})$ is positive, and thus

$$V(\mathfrak{T}) \leq \frac{\log(q/w) - 1 + \xi}{\log(q/w) - \log(\mathscr{T}_V/\tilde{\tau}_s) - 1}$$

$$= 1 + \frac{\log(\mathscr{T}_V/\tilde{\tau}_s) + \xi}{\log(q/w) - \log(\mathscr{T}_V/\tilde{\tau}_s) - 1}$$

$$\leq 1 + \frac{C[\log(\mathscr{T}_V/\tilde{\tau}_s) + \xi]}{\log n},$$

where we use the upper bound $\mathscr{T}_V \leq \mathscr{T}_M$ and (5.12) to ensure that the denominator is bounded below by $[(a/3)\log n - 1] > (a/4)\log n$ for all sufficiently large n. This gives half the desired upper bound on $V(\mathfrak{T})$ under the assumption $\mathscr{T}_V < \log n$.

The lower bound on $Q(\mathfrak{T})$ will follow a similar strategy. We begin by noting that an algebraic manipulation will prove that, for any set $\mathfrak{W} \subset T$,

$$Q(\mathfrak{W}) = \frac{1}{q^2}\left(\sum_{I \in \mathfrak{W}} X_I\left[\sum_{J \in \tilde{N}_I \cap \mathfrak{W}} X_J\right]\right) - \frac{V(\mathfrak{W})}{q}.$$

Let $\mathfrak{Z} = \mathfrak{I} \setminus \mathfrak{T}$, and observe that

$$Q(\mathfrak{I}) - Q(\mathfrak{T}) \leq \frac{1}{q^2} \cdot \left(\sum_{I \in \mathfrak{I}} X_I\left[\sum_{J \in \tilde{N}_I \cap \mathfrak{I}} X_J\right] - \sum_{I \in \mathfrak{T}} X_I\left[\sum_{J \in \tilde{N}_I \cap \mathfrak{T}} X_J\right]\right)$$

$$= \frac{1}{q^2} \cdot \left(\sum_{I \in \mathfrak{Z}} X_I\left[\sum_{J \in \tilde{N}_I \cap \mathfrak{I}} X_J\right] + \sum_{I \in \mathfrak{T}} X_I\left[\sum_{J \in \tilde{N}_I \cap \mathfrak{Z}} X_J\right]\right).$$

We decompose the first sum into

$$\sum_{I \in \mathfrak{Z}} X_I\left[\sum_{J \in \tilde{N}_I \cap \mathfrak{I}} X_J\right] = \sum_{I \in \mathfrak{Z}} X_I\left[\sum_{J \in \tilde{N}_I \cap \mathfrak{Z}} X_J\right] + \sum_{I \in \mathfrak{Z}} X_I\left[\sum_{J \in \tilde{N}_I \cap \mathfrak{T}} X_J\right]$$

$$= \sum_{I \in \mathfrak{Z}} X_I\left[\sum_{J \in \tilde{N}_I \cap \mathfrak{Z}} X_J\right] + \sum_{I \in \mathfrak{T}} X_I\left[\sum_{J \in \tilde{N}_I \cap \mathfrak{Z}} X_J\right],$$

where we can flip the order of summation in the final equality thanks to the symmetry of ρ. Substituting this in, we find that

(5.19) $$Q(\mathfrak{I}) - Q(\mathfrak{T}) \leq \frac{2}{q^2}\left(\sum_{I \in \mathfrak{I}} X_I\left[\sum_{J \in \tilde{N}_I \cap \mathfrak{Z}} X_J\right]\right).$$

By ordering of the X_I's, we have that $X_I \leq X_{\mathscr{T}_V}$ for any $I \in \mathfrak{Z}$. Therefore,

$$\frac{1}{q}\sum_{J \in \tilde{N}_I \cap \mathfrak{Z}} X_J \leq \frac{|\tilde{N}_I|V(\mathfrak{T})}{\mathscr{T}_V},$$

where we reuse (5.18). Substituting this into (5.19), we see that

$$Q(\mathfrak{I}) - Q(\mathfrak{T}) \leq \frac{2|\tilde{N}_I|V(\mathfrak{I}) \cdot V(\mathfrak{T})}{\mathscr{T}_V}.$$

Thanks to $\mathscr{H}_\xi$, $V(\mathfrak{I})$ is uniformly bounded in n (by (5.10)) and $Q(\mathfrak{I}) \geq 1 - \xi/\log n$ (by (5.11)). The uniform bound on $|\tilde{N}_I|$ from Lemma 5.1 proves half of the lower bound on $Q(\mathfrak{T})$.

Again, if $\mathscr{T}_V \geq \log n$, $\psi(\mathscr{T}_V) = C'/\mathscr{T}_V + \xi/\log n$ for all sufficiently large n. Thus, we may assume that $\mathscr{T}_V < \log n$ for the rest of the proof. Suppose that we were given the bound $V(\mathfrak{Z}) \leq C[1 + \log(\mathscr{T}_V/\tilde{\tau}_s)]/\log n$. From (5.19), we know that

$$Q(\mathfrak{I}) - Q(\mathfrak{T}) \leq 2V(\mathfrak{I}) \cdot V(\mathfrak{Z}) \leq \frac{C[1 + \log(\frac{\mathscr{T}_V}{\tilde{\tau}_s})]}{\log n},$$

where we use (5.10) to replace $V(\mathfrak{I})$ by a uniform constant. Thus, it is sufficient to prove the upper bound on $V(\mathfrak{Z})$.

To do so, we return to (5.11), and apply Lemma 5.11 to $\mathfrak{T}$ *without* ignoring the contribution of elements of $\mathfrak{Z}$. This allows us to deduce that

(5.20)
$$V(\mathfrak{T})\left[\log\left(\frac{q}{w}\right) + \log V(\mathfrak{T}) - \log\left(\frac{\mathscr{T}_V}{\tilde{\tau}_s}\right) - 1\right] + \frac{1}{q}\sum_{I \in \mathfrak{Z}} Y_I$$

$$\leq \log(q/w) - 1 + \xi.$$

Thanks to the upper bound on $\mathscr{T}_V$, we know that the bracketed term is positive and increasing in $V(\mathfrak{T})$. Since $V(\mathfrak{T}) \geq 1 - 2\xi/\log n$, some careful calculations imply that, for all sufficiently large n and sufficiently small ξ,

$$V(\mathfrak{T})\left[\log\left(\frac{q}{w}\right) + \log V(\mathfrak{T}) - \log\left(\frac{\mathscr{T}_V}{\tilde{\tau}_s}\right) - 1\right]$$

$$\geq \log(q/w) - 1 - \log\left(\frac{\mathscr{T}_V}{\tilde{\tau}_s}\right) - \frac{3\xi}{\log n} - \frac{3\xi \log(q/w)}{\log n}.$$

Substituting this into (5.20) gives that

$$\frac{1}{q}\sum_{I \in \mathfrak{Z}} Y_I \leq \xi + \frac{3\xi}{\log n} + \log\left(\frac{\mathscr{T}_V}{\tilde{\tau}_s}\right) + \frac{3\xi \log(q/w)}{\log n}$$

$$\leq C\xi + \log\left(\frac{\mathscr{T}_V}{\tilde{\tau}_s}\right),$$

for some uniform C and all sufficiently large n. Another application of Lemma 5.11—this time, to $\mathfrak{Z}$—implies that

$$V(\mathfrak{Z})\log\left(\frac{q\tilde{\tau}_s}{w \cdot \mathscr{T}_M}\right) + V(\mathfrak{Z})(\log V(\mathfrak{Z}) - 1) \leq C\xi + \log\left(\frac{\mathscr{T}_V}{\tilde{\tau}_s}\right),$$

where we bound $|\mathfrak{Z}|$ by $\mathscr{T}_M$. The function $x[\log(x) - 1]$ is bounded below by -1 for any positive x; applying this bound to $V(\mathfrak{Z})[\log V(\mathfrak{Z}) - 1]$ and rearranging the previous inequality algebraically, we find that, for all sufficiently large n,

$$V(\mathfrak{Z}) \leq \frac{C\xi + \log(\frac{\mathscr{T}_V}{\tilde{\tau}_s}) + 1}{\log[q\tilde{\tau}_s/(w \cdot \mathscr{T}_M)]} \leq \frac{C\xi + \log(\frac{\mathscr{T}_V}{\tilde{\tau}_s}) + 1}{(a/3) \cdot \log n} \leq \frac{C[1 + \log(\frac{\mathscr{T}_V}{\tilde{\tau}_s})]}{\log n},$$

where use (5.12) to get the penultimate bound. This completes the proof. $\quad\square$

Recall that, for any $\mathfrak{W} \subset T$, $Q(\mathfrak{W}) \leq V(\mathfrak{W})^2$. In a sense, the content of Lemma 5.13 is that this bound is nearly right for $\mathfrak{T}$. To make this precise, we define

$$P_I(\mathfrak{W}) := \frac{1}{q}\sum_{J \in \mathfrak{W}, \rho(I,J) > s} X_J.$$

Note that the sum is over indices whose distance *exceeds* s.

COROLLARY 5.14. *Assume $\mathscr{H}_\xi$ holds, and that $\mathscr{T}_V$ and $\mathfrak{T}$ are defined as above. Then, for all n sufficiently large and ξ sufficiently small,*

$$\frac{1}{q}\sum_{I \in \mathfrak{T}} X_I P_I(\mathfrak{T}) \leq 3\phi(\mathscr{T}_V) + \psi(\mathscr{T}_V),$$

where $\phi(\cdot)$ and $\psi(\cdot)$ are defined as in Lemma 5.13.

PROOF.   We observe that

$$V(\mathfrak{T})^2 - Q(\mathfrak{T}) = \frac{1}{q} \sum_{I \in \mathfrak{T}} X_I P_I(\mathfrak{T}) + \frac{V(\mathfrak{T})}{q},$$

where the additive factor of $V(\mathfrak{T})/q$ comes from the fact that $n^2 - 2\binom{n}{2} = n$. By Lemma 5.13, we conclude that

$$\frac{1}{q} \sum_{I \in \mathfrak{T}} X_I P_I(\mathfrak{T}) \le V(\mathfrak{T})^2 - Q(\mathfrak{T})$$

$$\le \left(1 + \phi(\mathscr{T}_V)\right)^2 - \left(1 - \psi(\mathscr{T}_V)\right) \le 3\phi(\mathscr{T}_V) + \psi(\mathscr{T}_V).$$

This completes the proof.   □

5.6. *Removing lower order terms.*   Before proceeding, consider the situation in which we assume that both ϕ and ψ vanish, and that $\xi = 0$. This implies that $P_I(\mathscr{T}_V)$ must be zero for every $I \le \mathscr{T}_V$, since the sum in Corollary 5.14 is made up of nonnegative terms. Thus, $\mathfrak{T}$ would have diameter at most s. Thanks to the lower bound on $\mathscr{T}_V$ in Lemma 5.13, the set would be a maximal clique set.

Of course, ϕ, ψ, and ξ are nonzero, so we cannot apply this argument to $\mathfrak{T}$ directly. We further truncate the set to deal with this difficulty. Define

(5.21) $$\mathscr{T}_P := \max\left\{k \le \mathscr{T}_V : X_k \ge \frac{\xi q}{\tilde{\tau}_s}\right\},$$

where we set $\mathscr{T}_P = 0$ if the set on the right is empty. We denote the set $\{1, \ldots, \mathscr{T}_P\}$ by $\mathfrak{P}$. The following lemma establishes bounds on $V(\mathfrak{P})$ and the sum of the Y_I's in $\mathfrak{P}$. At the end of this section, we will use these bounds to deduce some geometric properties of $\mathfrak{P}$.

LEMMA 5.15.   *Assume that $\mathscr{H}_\xi$ holds, and define $\mathscr{T}_P$ and $\mathfrak{P}$ as above. Then, for sufficiently small ξ and sufficiently large n,*

$$1 - \xi^{1/2} < V(\mathfrak{P}) \le 1 + \phi(\mathscr{T}_V)$$

and

$$\frac{1}{q} \sum_{I \in \mathfrak{P}} Y_I < V(\mathfrak{P})\left(\log(q/w) - 1\right) + \xi^{1/2}.$$

The stipulation on the sum of the Y_I's in $\mathfrak{P}$ is a slight (but essential) improvement on the naive inclusion bound given by (5.9) in the definition of $\mathscr{H}_\xi$.

PROOF.   The upper bound on $V(\mathfrak{P})$ follows from the inclusion $\mathfrak{P} \subset \mathfrak{T}$ and Lemma 5.13. Define

$$\mathfrak{L}_1 := \left\{I \in \mathfrak{T} : X_I < \frac{\xi q}{\tilde{\tau}_s \log(\mathscr{T}_V)}\right\}$$

and

$$\mathfrak{L}_2 := \left\{I \in \mathfrak{T} : \frac{\xi q}{\tilde{\tau}_s \log(\mathscr{T}_V)} \le X_I < \frac{\xi q}{\tilde{\tau}_s}\right\}.$$

Clearly, $\mathfrak{L}_1$, $\mathfrak{L}_2$ and $\mathfrak{P}$ form a partition of $\mathfrak{T}$. Proving the lemma is tantamount to proving good upper bounds on $V(\mathfrak{L}_1)$ and $V(\mathfrak{L}_2)$, as well as good lower bounds on the sum of the Y_I's in both sets.

To bound $V(\mathfrak{L}_1)$, we first need to bound $P_I(\mathfrak{L}_1)$ from below for any $I \in \mathfrak{T}$. The worst case scenario is that the distance restriction removes the $|\tilde{N}_I|$ largest elements of $\mathfrak{L}_1$. Therefore,

$$P_I(\mathfrak{L}_1) \geq V(\mathfrak{L}_1) - \frac{1}{q}|\tilde{N}_I|\Big(\max_{J \in \mathfrak{L}_1} X_J\Big) > V(\mathfrak{L}_1) - \frac{\xi|\tilde{N}_I|}{\tilde{\tau}_s \log(\mathcal{T}_V)}.$$

Since $P_I(\mathfrak{W}) \leq P_I(\mathfrak{W}')$ whenever $\mathfrak{W} \subset \mathfrak{W}'$, we see that Corollary 5.14 implies that

$$\frac{1}{q} \sum_{I \in \mathfrak{T}} X_I P_I(\mathfrak{L}_1) \leq 3\phi(\mathcal{T}_V) + \psi(\mathcal{T}_V).$$

Replacing $P_I(\mathfrak{L}_1)$ with its minimum and recalling that $|\tilde{N}_I|$ and $\tilde{\tau}_s$ are uniformly bounded in n (by Lemma 5.1), we see that

$$\Big(V(\mathfrak{L}_1) - \frac{C\xi}{\log(\mathcal{T}_V)}\Big)V(\mathfrak{T}) < 3\phi(\mathcal{T}_V) + \psi(\mathcal{T}_V).$$

Using the (very suboptimal) lower bound of $1/2$ for $V(\mathfrak{T})$ (which follows from Lemma 5.13 and n sufficiently large), we conclude that

$$(5.22) \qquad V(\mathfrak{L}_1) < 6\phi(\mathcal{T}_V) + 2\psi(\mathcal{T}_V) + \frac{C\xi}{\log(\mathcal{T}_V)},$$

for some C independent of n. Repeating this analysis with $\mathfrak{L}_2$ yields the inequality

$$V(\mathfrak{L}_2) < 6\phi(\mathcal{T}_V) + 2\psi(\mathcal{T}_V) + C\xi.$$

Since both $\phi(x)$ and $\psi(x)$ are bounded above by $1/(\log n)^{1/2}$ (from (5.16)) and $\mathcal{T}_V > (1 - \xi^{1/2})\tilde{\tau}_s$ (by Lemma 5.13), we get

$$(5.23) \qquad \max\{V(\mathfrak{L}_1), V(\mathfrak{L}_2)\} < C\xi.$$

Combining the previous bounds and the lower bound on $V(\mathfrak{T})$ from Lemma 5.13, we find that, for all sufficiently small values of ξ,

$$V(\mathfrak{P}) = V(\mathfrak{T}) - V(\mathfrak{L}_1) - V(\mathfrak{L}_2)$$

$$(5.24) \qquad\qquad \geq 1 - 2C\xi - \frac{2\xi}{\log n}$$

$$\geq 1 - \xi^{1/2}.$$

This establishes the lower bound on $V(\mathfrak{P})$.

We now turn to bounding $\sum_{I \in \mathfrak{P}} Y_I$. Suppose that

$$(5.25) \qquad \frac{1}{q} \sum_{I \in \mathfrak{L}_i} Y_I > V(\mathfrak{L}_i)\big(\log(q/w) - 1\big) - 2\xi^{2/3}.$$

We observe that (5.24) and Lemma 5.13 imply that

$$V(\mathfrak{P}) + (2\xi)/\log n \geq 1 - V(\mathfrak{L}_1) - V(\mathfrak{L}_2).$$

By inclusion, we may apply (5.9) to $\mathfrak{T}$, partition the elements into $\mathfrak{P}$, $\mathfrak{L}_1$ and $\mathfrak{L}_2$ and substitute the above bounds to get the following set of deductions:

$$\frac{1}{q} \sum_{I \in \mathfrak{P}} Y_I \leq \big(\log(q/w) - 1\big) + \xi - \frac{1}{q} \sum_{I \in \mathfrak{L}_1 \cup \mathfrak{L}_2} Y_I$$

$$< \big(1 - V(\mathfrak{L}_1) - V(\mathfrak{L}_2)\big)\big(\log(q/w) - 1\big) + 4\xi^{2/3} + \xi$$

$$< V(\mathfrak{P})\big(\log(q/w) - 1\big) + \frac{2\xi}{\log n}\big(\log(q/w) - 1\big) + 5\xi^{2/3}$$

$$< V(\mathfrak{P})\big(\log(q/w) - 1\big) + \xi^{1/2},$$

where the final inequality holds for sufficiently small ξ. Thus, we only need to show (5.25).

By applying Lemma 5.11 to $\mathfrak{L}_i$, we know that

$$(5.26) \qquad \frac{1}{q} \sum_{I \in \mathfrak{L}_i} Y_I \geq V(\mathfrak{L}_i)(\log(q/w) - 1) + V(\mathfrak{L}_i) \log V(\mathfrak{L}_i) - V(\mathfrak{L}_i) \log(|\mathfrak{L}_i|),$$

where we ignored the positive term $V(\mathfrak{L}_i) \log \tilde{\tau}_s$. Using (5.23), we find that $V(\mathfrak{L}_i) \log V(\mathfrak{L}_i)$ is bounded below by $C\xi \log[C\xi] > -\xi^{2/3}$.

Controlling $V(\mathfrak{L}_1) \log(|\mathfrak{L}_1|)$ is quite straightforward: since inclusion implies that $|\mathfrak{L}_1| \leq \mathscr{T}_V$, we can use (5.22) to see that, for sufficiently large n,

$$V(\mathfrak{L}_1) \log(|\mathfrak{L}_1|) \leq \left(6\phi(\mathscr{T}_V) + 2\psi(\mathscr{T}_V) + \frac{C\xi}{\log(\mathscr{T}_V)}\right) \log(\mathscr{T}_V)$$

$$\leq \max_{2 \leq x \leq \mathscr{T}_M} \left\{ \log x \cdot (6\phi(x) + 2\psi(x)) \right\} + C\xi,$$

where we use the fact that $2 \leq \tilde{\tau}_s(1 - \xi^{1/2}) \leq \mathscr{T}_V \leq \mathscr{T}_M$. Recalling the definitions of ϕ and ψ, we have that

$$6\phi(x) + 2\psi(x) = \min\left\{ \frac{6C[\log(x/\tilde{\tau}_s) + \xi]}{\log n}, \frac{12}{x} \right\}$$

$$+ \min\left\{ \frac{2C[1 + \log(x/\tilde{\tau}_s)]}{\log n}, \frac{2C'}{x} \right\} + \frac{\xi}{\log n},$$

and thus, it can be shown that, for some (different) uniform constant C,

$$\log x \cdot (6\phi(x) + 2\psi(x)) \leq C \cdot \min\left\{ \frac{\log x(1 + \log x)}{\log n}, \frac{\log x}{x} \right\} + \frac{\xi \log x}{\log n}.$$

We wish to prove a uniform, x-independent bound on the minimum above (for any $x \geq 2$). The first function in the minimum is increasing; meanwhile, the second function in the minimum is larger than the first one on $[2, e]$ (for all sufficiently large n), and is a decreasing function of x on (e, ∞). Thus, for any $2 \leq x \leq \mathscr{T}_M$, the minimum is bounded above by $\log y(1 + \log y)/\log n$ for any y that satisfies $(1 + \log y)/\log n \geq 1/y$. The value $y = \log n$ is one such value. Combining the two estimates gives

$$\log x \cdot (6\phi(x) + 2\psi(x)) \leq \frac{C[\log \log n + (\log \log n)^2]}{\log n} + \frac{\xi \log x}{\log n}.$$

Thus, we find that

$$\max_{2 \leq x \leq \mathscr{T}_M} \log(x) \cdot (6\phi(x) + 2\psi(x))$$

$$(5.27)$$

$$\leq \frac{C[\log \log n + (\log \log n)^2]}{\log n} + \frac{\xi \log \mathscr{T}_M}{\log n}.$$

Using (5.8) to bound $\mathscr{T}_M$, we deduce that, for all sufficiently large n, $V(\mathfrak{L}_1) \log(|\mathfrak{L}_1|) < C\xi$ for some uniform constant C; this, in turn, is bounded by $\xi^{2/3}$ for all sufficiently small ξ.

To control $V(\mathfrak{L}_2) \log(|\mathfrak{L}_2|)$, we must be slightly more careful. For any $I \in \mathfrak{L}_2$, we know that

$$P_I(\mathfrak{L}_2) \geq \frac{1}{q} \sum_{J \in \mathfrak{L}_2, \rho(I,J) > s} \left(\min_{J \in \mathfrak{L}_2} X_J \right)$$

$$\geq \sum_{J \in \mathfrak{L}_2, \rho(I,J) > s} \frac{\xi}{\tilde{\tau}_s \log(\mathscr{T}_V)}$$

$$\geq \frac{\xi(|\mathfrak{L}_2| - |\tilde{N}_I|)}{\tilde{\tau}_s \log(\mathscr{T}_V)},$$

where we use the lower bound defining $\mathfrak{L}_2$. By inclusion, $P_I(\mathfrak{L}_2) < P_I(\mathfrak{T})$, and Corollary 5.14 allows us to conclude that

$$V(\mathfrak{T})\left[\frac{\xi(|\mathfrak{L}_2| - |\tilde{N}_I|)}{\tilde{\tau}_s \log(\mathscr{T}_V)}\right] \leq \frac{1}{q}\sum_{I \in \mathfrak{T}} X_I P_I(\mathfrak{L}_2)$$

$$\leq 3\phi(\mathscr{T}_V) + \psi(\mathscr{T}_V).$$

Solving for $|\mathfrak{L}_2|$, we deduce that

$$|\mathfrak{L}_2| \leq |\tilde{N}_I| + \frac{\tilde{\tau}_s}{\xi} \cdot \log(\mathscr{T}_V) \cdot (6\phi(\mathscr{T}_V) + 2\psi(\mathscr{T}_V)),$$

where we bound $V(\mathfrak{T})$ from below by $1/2$ by Lemma 5.13. Referring back to (5.27), we see that

$$|\mathfrak{L}_2| \leq |\tilde{N}_I| + \frac{C\tilde{\tau}_s \cdot [\log\log n + (\log\log n)^2]}{\xi \log n} + \frac{\tilde{\tau}_s \log \mathscr{T}_M}{\log n}.$$

Using (5.8) and Lemma 5.1 to bound $|\tilde{N}_I|$ and $\tilde{\tau}_s$ from above, this proves that $|\mathfrak{L}_2|$ is uniformly bounded in n. Appealing to (5.23) a final time,

$$V(\mathfrak{L}_2)\log(|\mathfrak{L}_2|) < C\xi < \xi^{2/3}.$$

This completes the proof. $\square$

As promised, we now show that $\mathfrak{P}$ has the desired geometric properties.

LEMMA 5.16. *For all n sufficiently large and ξ sufficiently small,*

$$\text{diam}(\mathfrak{P}) \leq s \quad \text{and} \quad \mathscr{T}_P > \tilde{\tau}_s(1 - \xi^{1/3}),$$

that is, the set $\mathfrak{P}$ is a $\xi^{1/3}$-almost maximal clique set.

PROOF. Assume that there exists a pair of indices $I^*, J^* \in \mathfrak{P}$ such that $\rho(I^*, J^*) > s$. Then

$$\frac{1}{q}\sum_{I \in \mathfrak{T}} X_I P_I(\mathfrak{T}) \geq \frac{X_{I^*} X_{J^*}}{q^2} \geq \frac{\xi^2}{\tilde{\tau}_s^2},$$

where the final bound is from the definition of $\mathfrak{P}$. By Corollary 5.14, we know the left-hand quantity cannot exceed $3\phi(\mathscr{T}_V) + \psi(\mathscr{T}_V)$, which is bounded above by $4/(\log n)^{1/2}$—a clear contradiction for all sufficiently large n. Thus, the diameter $\mathfrak{P}$ is at most s.

Combining Lemmas 5.11 and 5.15 give

$$V(\mathfrak{P})(\log(q/w) - 1) + V(\mathfrak{P})\left(\log V(\mathfrak{P}) - \log\left(\frac{\mathscr{T}_P}{\tilde{\tau}_s}\right)\right)$$

$$< V(\mathfrak{P})(\log(q/w) - 1) + \xi^{1/2}$$

and, therefore,

$$\mathscr{T}_P > \tilde{\tau}_s \cdot V(\mathfrak{P})\exp\left(-\frac{\xi^{1/2}}{V(\mathfrak{P})}\right) > \tilde{\tau}_s \cdot V(\mathfrak{P})\left(1 - \frac{\xi^{1/2}}{V(\mathfrak{P})}\right)$$

using the standard estimate $e^{-x} \geq 1 - x$. Combining this with the lower bound on $V(\mathfrak{P})$ from Lemma 5.15 forces

$$\mathscr{T}_P > \tilde{\tau}_s(1 - \xi^{1/3})$$

for all sufficiently large values of n and small values of ξ. $\square$

5.7. *Convex analysis.*   We are nearly done with the proof: all that remains is to show that, for most $I \in \mathfrak{P}$, X_I is close to $q/\tilde{\tau}_s$, and then to formally prove the theorem. The essential additional information we are now armed with is an *upper* bound on $\mathscr{T}_P$—namely, $\tilde{\tau}_s$, as $\mathfrak{P}$ has diameter at most s, and $\tilde{\tau}_s$ is the largest possible cardinality for such a set of indices.

LEMMA 5.17.   *Let $\mathfrak{P}$ be as above, and assume that $\mathscr{H}_\xi$ holds. Define*

$$\mathfrak{B} := \left\{ I : \left| \frac{X_I \tilde{\tau}_s}{q} - 1 \right| < \xi^{1/5} \right\} \quad and \quad \mathfrak{C} = \mathfrak{P} \setminus \mathfrak{B}.$$

Then, for all sufficiently large n and ξ sufficiently small,

$$|\mathfrak{B}| > (1 - 10\xi^{1/10})\tilde{\tau}_s, \qquad |\mathfrak{C}| < 9\xi^{1/10}\tilde{\tau}_s \quad and \quad V(\mathfrak{C}) < 10\xi^{1/10}.$$

PROOF.   We consider the Taylor expansion of Y_I around the value $q/\tilde{\tau}_s$. Explicitly, we let $f(x) = x(\log(x/\mathscr{D}) - 1) + \mathscr{D}$, and by Taylor's theorem,

$$Y_I = f(X_I) = f\left(\frac{q}{\tilde{\tau}_s}\right) + f'\left(\frac{q}{\tilde{\tau}_s}\right)\left(X_I - \frac{q}{\tilde{\tau}_s}\right) + \frac{f''(L(X_I))}{2}\left(X_I - \frac{q}{\tilde{\tau}_s}\right)^2,$$

where $L(X_I)$ is some number between X_I and $q/\tilde{\tau}_s$. Differentiating $f(x)$ explicitly and simplifying algebraically, we see that

$$Y_I = \mathscr{D} - \frac{q}{\tilde{\tau}_s} + X_I \log(q/w) + \frac{1}{2L(X_I)}\left(X_I - \frac{q}{\tilde{\tau}_s}\right)^2.$$

Next, we sum over $\mathfrak{P}$ and use the upper bound on $\sum_{I \in \mathfrak{P}} Y_I$ from Lemma 5.15 to deduce that

$$\frac{1}{q} \sum_{I \in \mathfrak{P}} \left[\mathscr{D} - \frac{q}{\tilde{\tau}_s} + X_I \log(q/w) + \frac{1}{2L(X_I)}\left(X_I - \frac{q}{\tilde{\tau}_s}\right)^2 \right]$$

$$\leq V(\mathfrak{P})(\log(q/w) - 1) + \xi^{1/2}.$$

We ignore the positive term $\mathscr{D}$ on the left-hand side. From Lemma 5.16, the diameter of $\mathfrak{P}$ is at most s, and hence $\mathscr{T}_P \leq \tilde{\tau}_s$. Thus,

$$(5.28) \qquad \frac{1}{q} \sum_{I \in \mathfrak{P}} \frac{1}{2L(X_I)}\left(X_I - \frac{q}{\tilde{\tau}_s}\right)^2 \leq \frac{\mathscr{T}_P}{\tilde{\tau}_s} - V(\mathfrak{P}) + \xi^{1/2} \leq 2\xi^{1/2},$$

where the final inequality follows thanks to the lower bound on $V(\mathfrak{P})$ from Lemma 5.15.
   Now, define

$$\mathfrak{W}_1 := \{ I \in \mathfrak{P} : X_I \geq (1 + \xi^{1/5})q/\tilde{\tau}_s \}$$

and

$$\mathfrak{W}_2 := \{ I \in \mathfrak{P} : X_I \leq (1 - \xi^{1/5})q/\tilde{\tau}_s \}.$$

We recall that $\mathfrak{C} = \mathfrak{P} \setminus \mathfrak{B} = \mathfrak{W}_1 \cup \mathfrak{W}_2$. On $\mathfrak{W}_1$, the function $1/L(X_I)$ is bounded below by $1/X_I$. Thus, (5.28) implies that

$$\frac{|\mathfrak{W}_1|}{q} \cdot \min_{I \in \mathfrak{W}_1} \left\{ \frac{1}{2X_I}\left(X_I - \frac{q}{\tilde{\tau}_s}\right)^2 \right\} \leq 2\xi^{1/2}.$$

The function $x \mapsto (x - q/\tilde{\tau}_s)^2/(2x)$ is strictly increasing on the interval $[(q/\tilde{\tau}_s)(1 + \xi^{1/5}), \infty)$, and always convex. Using the first fact, we find that

$$|\mathfrak{W}_1| \cdot \left(\frac{\xi^{2/5}}{2\tilde{\tau}_s \cdot (1 + \xi^{1/5})} \right) \leq 2\xi^{1/2}.$$

This implies that $|\mathfrak{W}_1| < 5\xi^{1/10}\tilde{\tau}_s$. To control $V(\mathfrak{W}_1)$, we apply (the standard version of) Jensen's inequality to the convex function $x \mapsto (x - q/\tilde{\tau}_s)^2/(2x)$. Algebraically manipulating the resulting expression gives

$$\frac{1}{2V(\mathfrak{W}_1)}\left(V(\mathfrak{W}_1) - \frac{|\mathfrak{W}_1|}{\tilde{\tau}_s}\right)^2 \le \frac{1}{q}\sum_{I \in \mathfrak{W}_1}\frac{1}{2X_I}\left(X_I - \frac{q}{\tilde{\tau}_s}\right)^2 \le 2\xi^{1/2},$$

where the final bound follows from (5.28). Since the left-hand side is increasing in $V(\mathfrak{W}_1)$ whenever $V(\mathfrak{W}_1) \ge |\mathfrak{W}_1|/\tilde{\tau}_s$, we can conclude that

$$V(\mathfrak{W}_1) \le |\mathfrak{W}_1|/\tilde{\tau}_s + 2\xi^{1/4} < 6\xi^{1/10}.$$

For $\mathfrak{W}_2$, we can bound $1/L(X_I)$ from below by $\tilde{\tau}_s/q$. A final appeal to (5.28) gives that

$$\frac{|\mathfrak{W}_2|}{q} \cdot \min_{I \in \mathfrak{W}_2}\left\{\frac{\tilde{\tau}_s}{2q}\left(X_I - \frac{q}{\tilde{\tau}_s}\right)^2\right\} \le 2\xi^{1/2}.$$

We can bound the minimum from below by assuming that some X_I realizes the upper bound that defines $\mathfrak{W}_2$. In this case, it is immediate that $|\mathfrak{W}_2| \le 4\xi^{1/10}\tilde{\tau}_s$. Finally, $X_I \le q/\tilde{\tau}_s$ for each $I \in \mathfrak{W}_2$ and, therefore, $V(\mathfrak{W}_2) < 4\xi^{1/10}$. Putting these terms together, we find that

$$V(\mathfrak{C}) = V(\mathfrak{W}_1) + V(\mathfrak{W}_2) < 10\xi^{1/10}.$$

We also see that

$$|\mathfrak{C}| = |\mathfrak{W}_1| + |\mathfrak{W}_2| < 9\xi^{1/10}\tilde{\tau}_s.$$

The lower bound on $|\mathfrak{B}|$ follows from

$$|\mathfrak{B}| = \mathcal{T}_P - |\mathfrak{C}| > \tilde{\tau}_s(1 - \xi^{1/3}) - 9\xi^{1/10}\tilde{\tau}_s > \tilde{\tau}_s(1 - 10\xi^{1/10}),$$

where the penultimate inequality was proved in Lemma 5.16. $\square$

We have completed the proof of the difficult assertion in Theorem 3.1; all that is left is to ensure the second stipulation holds.

PROOF OF THEOREM 3.1.   We recall the definition of $\mathcal{G}_{n,\tilde{\delta}}(\tilde{\varepsilon})$: there must a pair of sets $\mathfrak{B}$ and $\mathfrak{C}$ such that

(a) $\mathfrak{B}$ is $\tilde{\varepsilon}$-almost maximal clique set such that $I \in \mathfrak{B}$ implies

$$\left|\frac{\tilde{\tau}_s X_I}{q} - 1\right| < \tilde{\varepsilon},$$

(b) $\mathfrak{C}$ satisfies

$$|\mathfrak{C}| < \tilde{\varepsilon}\tilde{\tau}_s \quad \text{and} \quad V(\mathfrak{C}) < \tilde{\varepsilon},$$

and

(c) whenever $J \in (\mathfrak{B} \cup \mathfrak{C})^c$,

$$X_J < \frac{\tilde{\varepsilon} \cdot q}{\tilde{\tau}_s}.$$

We set $\xi = (\tilde{\varepsilon}/10)^{10}$. Then, whenever $\mathcal{H}_\xi$ holds, the sequence of assertions given by Lemmas 5.13, 5.15, 5.16 and 5.17 assure us that, for n sufficiently large and $\tilde{\varepsilon}$ sufficiently small (interpreted according to our stated conventions), the sets $\mathfrak{B}$ and $\mathfrak{C}$ of Lemma 5.17 satisfy the first and second conditions of $\mathcal{G}_{n,\tilde{\delta}}(\tilde{\varepsilon})$.

To show the final condition holds, we must show that all elements of $T \setminus \mathfrak{P}$ are small. If $\mathscr{T}_V > \mathscr{T}_P$, then $X_{\mathscr{T}_P+1} < \xi q / \tilde{\tau}_s < \tilde{\varepsilon} q / \tilde{\tau}_s$. This implies the upper bound holds for every element outside of $\mathfrak{P}$, thanks to the ordering of the elements.

We are left with the scenario in which $\mathscr{T}_V = \mathscr{T}_P$. By definition (recall (5.15) and (5.21)), this means that $\mathfrak{P} = \mathfrak{T}$ and, therefore, $\mathscr{T}_V \le \tilde{\tau}_s$. Formally, it is still possible that $X_{\mathscr{T}_V+1} \ge \xi q / \tilde{\tau}_s$. By Lemma 5.11 and Lemma 5.13, we deduce that

$$\frac{1}{q} \sum_{I \in \mathfrak{T}} Y_I \ge V(\mathfrak{T})\left(\log(q/w) + \log(V(\mathfrak{T})) - \log\left(\frac{\mathscr{T}_V}{\tilde{\tau}_s}\right) - 1\right)$$

$$\ge \left(1 - \frac{2\xi}{\log n}\right)\left(\log(q/w) - 1 - \frac{3\xi}{\log n}\right),$$

where the contribution of the term including the logarithm of $\mathscr{T}_V$ is nonnegative, and thus can be safely ignored. Note that, since we assumed $q > 3w$, the right-hand side above is positive for all sufficiently large n. Therefore, (5.9) implies that

$$\frac{1}{q} \sum_{I \in \mathfrak{I} \setminus \mathfrak{T}} Y_I \le [\log(q/w) - 1 + \xi] - \left(1 - \frac{2\xi}{\log n}\right)\left[\log(q/w) - 1 - \frac{3\xi}{\log n}\right]$$

$$\le \frac{4\xi \log(q/w)}{\log n} + \xi$$

$$\le C\xi.$$

As before, the final inequality follows because $\log(q/w) \le C \log n$ for some C.

Suppose $X_{\mathscr{T}_V+1} \ge \tilde{\varepsilon} q / \tilde{\tau}_s = 10 \xi^{1/10} q / \tilde{\tau}_s$. Then we find that

$$Y_{\mathscr{T}_V+1} = X_{\mathscr{T}_V+1}(\log(X_{\mathscr{T}_V+1}/\mathscr{D}) - 1) + \mathscr{D}$$

$$\ge \xi q(\log(q/w) + \log(10\xi^{1/10}) - 1) + \mathscr{D}.$$

If we divide through by q, we find that this expression still grows with n, whereas the earlier upper bound is uniformly bounded. This is a contradiction, and we get the upper bound $X_{\mathscr{T}_V+1} < \tilde{\varepsilon} q / \tilde{\tau}_s$.

The above computation implies that any configuration in $\mathscr{H}_{(\tilde{\varepsilon}/10)^{10}}$ is also in $\mathscr{G}_{n,\tilde{\delta}}(\tilde{\varepsilon})$. Taking complements and intersecting with $\mathscr{L}_n(\tilde{\delta})$, we conclude that

$$\mathbb{P}[\mathscr{G}_{n,\tilde{\delta}}(\tilde{\varepsilon})^c \cap \mathscr{L}_n(\tilde{\delta})] \le \mathbb{P}[\mathscr{H}_{(\tilde{\varepsilon}/10)^{10}}^c \cap \mathscr{L}_n(\tilde{\delta})]$$

$$\le 3 \exp(-q[\log(q/w) - 1 + (\tilde{\varepsilon}/10)^{10}/2]),$$

where the final inequality is Proposition 5.12. $\square$

6. Moving from discrete to continuous. This section has three parts: first, we prove several estimates that show the discrete setting of the s-graded model approximates the continuous geometry of $\mathbb{T}^d$. We then prove a proposition relating the random geometric graph structural event $\mathscr{F}_n(\varepsilon)$ with $\mathscr{G}_{n,\tilde{\delta}}(\tilde{\varepsilon})$, given by the structure Theorem 3.1 on the s-graded model, with appropriately chosen parameters. The second part is probabilistic, where we find a tight lower bound on the event that the number of edges in the random geometric graph exceeds its mean. We then assume Theorem 3.1 and deduce Theorem 2.1 by choosing $\tilde{\delta}$ and $\tilde{\varepsilon}$ judiciously and employing a correlation inequality.

6.1. *Geometric lemmas.* Recall that (as in (5.2)), for any $K \subset \mathbb{T}^d$, we define $\mathfrak{R}(K)$ and $\mathfrak{O}(K)$ to be the maximal (resp., minimal) subsets of T such that

(6.1) $$\mathfrak{U}(\mathfrak{R}(K)) \subset K \quad \text{and} \quad K \setminus K' \subset \mathfrak{U}(\mathfrak{O}(K)),$$

where K' is some subset of K of Lebesgue measure 0. Recall that $\lambda(\cdot)$ will be used to denote Lebesgue measure of sets. We begin by showing that this operation does not alter the measure of convex subsets S of diameter at most r by very much. Formally, we have the following.

PROPOSITION 6.1. *Fix $\varepsilon > 0$, and let S be a convex subset of diameter at most $2r$. Then there exists an s_0 that depends only on ε, the dimension and the choice of norm, such that whenever $s > s_0$, the inner hull $\mathfrak{R}(S)$ satisfies*

$$\lambda(S) - \frac{\varepsilon^2 \tau}{64} < \lambda\big(\mathfrak{U}(\mathfrak{R}(S))\big) \le \lambda(S).$$

Similarly, the corresponding inequality

$$\lambda(S) \le \lambda\big(\mathfrak{U}(\mathfrak{O}(S))\big) < \lambda(S) + \frac{\varepsilon^2 \tau}{64}$$

hold for the outer hull $\mathfrak{O}(S)$.

Note that, for any fixed S, this result follows from the continuity of Lebesgue measure. The essential part of this proposition is that the choice of s_0 is uniform for all convex subsets S that satisfy the assumptions of the proposition.

Instead of proving the proposition directly, we consider the following minor modification.

LEMMA 6.2. *Let S be a convex set as in Proposition 6.1. Define*

$$(\partial S)_l := \big\{ y : \|\partial S - y\| \le l \big\},$$

where $\|B - x\|$ is shorthand for $\inf_{b \in B} \|b - x\|$ for any set B. Then there exists an s_0, depending on ε, the dimension, and the norm, such that $s > s_0$ implies

$$\lambda\big((\partial S)_{\varsigma/m}\big) < \frac{\varepsilon^2 \tau}{64},$$

where ς is the diameter of a unit square (as in (5.3)).

This lemma implies Proposition 6.1, since

$$S \subset \mathfrak{U}(\mathfrak{R}(S)) \cup (\partial S)_{\varsigma/m}$$

and

$$\mathfrak{U}(\mathfrak{O}(S)) \subset S \cup (\partial S)_{\varsigma/m}.$$

Taking the Lebesgue measure of both sides and using subadditivity gives the two nontrivial bounds in Proposition 6.1.

PROOF OF LEMMA 6.2. Heuristically, the volume of $(\partial S)_{\varsigma/m}$ should be commensurate with the product of ς/m with the surface area of S. Since S has diameter at most $2r$ and is a convex set (and, therefore, its boundary cannot be too convoluted), this surface area ought to be bounded above by Cr^{d-1}. To formalize this loose heuristic, we turn to the tools of geometric measure theory. Although this approach is standard in that field, we include a detailed proof for completeness.

Consider a function $f : \mathbb{R}^d \to \mathbb{R}$ that is Lipschitz with respect to the Euclidean distance, and a Borel set A. The Euclidean coarea formula [15], page 248, states that, with the functions as above,

$$\int_A \|Df(x)\|_2 \, dx = \int_{-\infty}^{\infty} H^{d-1}(A \cap f^{-1}(y)) \, dy,$$

where $\| \cdot \|_2$ is the Euclidean norm, and H^{d-1} is the Hausdorff measure on the surface (effectively the surface area) and Df is the gradient of f, which exists since f is almost everywhere differentiable.

The natural choice for our analysis is $f(x) = \|\partial S - x\|$. We begin by proving that f is Lipshitz with respect to the Euclidean norm. We recall the classical fact that all norms are equivalent in finite dimensional space, that is, there exists two positive constants c and C such that, for all x and y,

$$c\|x - y\| \le \|x - y\|_2 \le C\|x - y\|.$$

Pick two points x and y, and let $a \in \partial S$ be the point such that $f(x) = \|x - a\|$ (this point exists because ∂S is closed). Then

$$f(y) - f(x) \le \|y - a\| - \|x - a\| \le \|x - y\| \le C\|x - y\|_2.$$

Similarly, $f(x) - f(y) \le C\|x - y\|_2$.

Thus, f is differentiable almost everywhere. Pick an x where the function is differentiable, and let a be as before. Then, for any $t \in (0, 1)$,

$$f(x + t(a - x)) \le f(x) - t\|a - x\|,$$

by the properties of norms. Subtracting $f(x)$ from both sides, dividing by t and letting $t \to 0$, we get

$$\langle a - x, Df \rangle \le -\|a - x\|,$$

where $\langle \cdot, \cdot \rangle$ is the Euclidean inner product (or in this case, the directional derivative). Applying the Cauchy–Schwarz inequality, we conclude that

$$\|a - x\| \le \|a - x\|_2 \|Df\|_2,$$

which implies, by the equivalence of norms, that

$$\|Df\|_2 \ge \frac{\|a - x\|}{\|a - x\|_2} \ge c.$$

Letting $A = \{x : f(x) \le \varsigma/m\}$, we apply the Euclidean coarea formula to deduce that

(6.2) $$\lambda[(\partial S)_{\varsigma/m}] \le C \int_0^{\varsigma/m} H^{d-1}(\{y : \|\partial S - y\| = z\}) \, dz,$$

where C is some (possibly different) universal constant. Note that, for sufficiently small z and T convex, the set $\{y : \|\partial T - y\| = z\}$ has two connected components: one inside T and the other outside of it. We wish to show that both are boundaries of convex sets. The outer one is the boundary of the affine sum of S and the ball of radius z, which is known to be convex. The internal one is the boundary of

$$S^{(z)} := \{x : x \in S, \|\partial S - x\| \ge z\}.$$

We claim that this set is also convex: if it were not, we could find $x, y \in S^{(z)}$ such that $w = tx + (1 - t)y \notin S^{(z)}$ for some $t \in (0, 1)$. Let v be the minimal length vector such that $w + v \in \partial S$. Then, since $\|v\| < z$ by definition, we can find an ε sufficiently small such that

$w + (1 + \varepsilon)v \notin S$, while $x + (1 + \varepsilon)v$, $y + (1 + \varepsilon)v$ are both in S, contradicting convexity of S.

We complete the proof using Cauchy's surface area formula [28], page 55: for any convex S,

$$H^{d-1}(\partial S) = C_d \int_{u \in \mathbb{S}^{d-1}} \lambda_{d-1}(S|u^{\perp}) \, du,$$

where $\mathbb{S}^{d-1}$ is the $d - 1$-dimensional unit ball in $\mathbb{R}^d$, λ_{d-1} is the $d - 1$-dimensional Lebesgue measure, $S|u^{\perp}$ is the projection of S onto the $d - 1$ dimensional subspace perpendicular to u, and C_d is a constant used to ensure the Hausdorff and Lebesgue measures are compatible.

We apply this formula to the two convex sets $(S)_z$ and $S^{(z)}$ (discussed above) whose boundaries give the two connected components of $\{y : \|\partial S - y\| = z\}$. The projection of either set onto a $d - 1$-dimensional subset has diameter at most $2(r + z)$ and, therefore, by the triangle inequality, is included in some ball of diameter $4(r + z)$. Thus, inclusion implies that

$$\lambda_{d-1}((S)_z|u^{\perp}) \leq C_d(r + z)^{d-1} \quad \text{and} \quad \lambda_{d-1}(S^{(z)}|u^{\perp}) \leq C_d(r + z)^{d-1}$$

for some constant C_d. Substituting this into (6.2) gives

$$\lambda[(\partial S)_{\varsigma/m}] \leq C \int_r^{r+(\varsigma/m)} u^{d-1} \, du = C[(r + \varsigma/m)^d - r^d].$$

Increasing s, we can ensure that $[1 + \varsigma/(rm)]^d < 1 + (2d\varsigma)/(rm)$ and, therefore,

$$\lambda[(\partial S)_{\varsigma/m}] \leq \frac{Cr^{d-1}}{m} \leq \frac{C\tau}{rm} < \frac{C\tau}{s - r},$$

where the value of C may change from one inequality to the next. Increasing s, we get the desired result. $\square$

Next, we prove a lemma that proves that almost maximal clique sets are, essentially, discretized versions of balls of diameter r.

LEMMA 6.3. *Fix $\varepsilon > 0$. Then there exists s_0 such that, for any $s > s_0$, r sufficiently small and $s^{-1/20}$-almost maximal clique set $\mathfrak{B}$ in the s-graded model, there exists a set of indices $\mathfrak{H} \subset T$ with $|\mathfrak{H}| < \varepsilon \cdot \tilde{\tau}_s$ and B, a ball of diameter r in $\mathbb{T}^d$, such that*

$$B \subset \mathfrak{U}(\mathfrak{B} \cup \mathfrak{H}) \quad \text{and} \quad \mathfrak{U}(\mathfrak{B}) \subset (B)_{\varepsilon r}.$$

The choice of the term $s^{-1/20}$ is somewhat arbitrary—the above result will follow for any function which vanishes as s grows. We framed the lemma as we did since the function $s^{-1/20}$ is used in the proof of Theorem 2.1 below.

PROOF. To prove this lemma, we go through the abstract framework of Hausdorff convergence of subsets of a metric space. Consider an abstract metric space X imbued with metric ι, and, for any $S \subset X$, define the l-fattening of S as before, using the metric ι to measure distance. For any two $A, B \subset X$, the Hausdorff distance is defined as

$$\iota_H(A, B) := \inf\{l : A \subset (B)_l, B \subset (A)_l\}.$$

If X is compact in the topology defined by ι, the space of closed subsets of X makes a compact space with respect to this metric [36], page 294.

Recall that $T_s = \{1, 2, \ldots, m\}^d$ is a set of indices, where we add the subscript s to emphasize the dependence on this variable. Let $\{\mathfrak{B}_s\}_{s \geq 2}$ be a sequence of $s^{-1/20}$-almost maximal clique sets, where $\mathfrak{B}_s \subset T_s$. Define A to be some ball of radius $r(1 + 2\varsigma)$ in $\mathbb{T}^d$. For any

set S of *diameter* of at most $r(1 + 2\varsigma)$, there exists a translate of S which lies completely in A—this can be done by simply translating any point of S to the center of A. Thus, there is a translate of $\mathfrak{U}(\mathfrak{B}_s)$ that is a subset of A, since

$$\text{diam}(\mathfrak{U}(\mathfrak{B}_s)) \leq r + \frac{2\varsigma}{m} \leq r\left(1 + \frac{2\varsigma}{s-r}\right) < r(1 + 2\varsigma),$$

whenever s is sufficiently large and r sufficiently small.

Let $\tilde{A}$ and $\tilde{B}_s$ be the sets A and the translate of $\mathfrak{U}(\mathfrak{B}_s)$ that are inside A, scaled by $1/r$ and embedded in $\mathbb{R}^d$ (which is possible as long as r is sufficiently small to not "notice" the torroidal geometry of $\mathbb{T}^d$). Thus, $\tilde{A}$ is a ball of radius $1 + 2\varsigma$. By Lemma 5.1, we know that

$$\text{diam}(\tilde{B}_s) \leq 1 + \frac{C}{s} \quad \text{and} \quad \lambda(\tilde{B}_s) \geq \frac{\tau(1 - s^{-1/20})}{r^d} = \frac{\nu(1 - s^{-1/20})}{2^d},$$

with the final inequality being the definition of τ.

Since all the $\tilde{B}_s$ are closed sets embedded in a single compact metric space (namely, $\tilde{A}$), we can extract a subsequence $\tilde{B}_{s_k}$ which converges to some set $\tilde{B}$ in the Hausdorff metric. Passing to the limit, we see that $\tilde{B}$ must have diameter of at most 1 and measure at least $\nu/2^d$. Quoting the isodiametric inequality again [6], page 94,

$$\lambda(\tilde{B}) \leq \nu\left(\frac{\text{diam}\,\tilde{B}}{2}\right)^d,$$

where equality holds if and only if $\tilde{B}$ is a ball. This implies that $\tilde{B}$ is, in fact, the ball of diameter 1 (up to sets of measure zero). Indeed, if we take an arbitrary subsequence of $\{\tilde{B}_s\}$, we can extract a convergent sub-subsequence whose limit will have diameter 1 and measure $\nu/2^d$—meaning any sub-subsequential limit is some ball of diameter 1. Letting $\mathscr{B}$ be the set of all balls of diameter in A, we find that

$$\lim_{s \to \infty} \inf_{\tilde{B} \in \mathscr{B}} \iota_H(\tilde{B}_s, \tilde{B}) = 0.$$

Therefore, there is an s_0, such that, for any $s > s_0$, there exists some ball $\tilde{B} \in \mathscr{B}$ such that

$$\iota_H(\tilde{B}_s, \tilde{B}) < \varepsilon/(16d).$$

Scaling by r, we find that, for any $s > s_0$, there is a ball B of diameter r in $\mathbb{T}^d$, such that

$$B \subset (\mathfrak{U}(\mathfrak{B}_s))_{\varepsilon r/(16d)} \quad \text{and} \quad \mathfrak{U}(\mathfrak{B}_s) \subset (B)_{\varepsilon r/(16d)}.$$

The second statement implies the required inclusion of $\mathfrak{U}(\mathfrak{B})$ in a $(B)_{\varepsilon r}$. For the other direction, we note that $(\mathfrak{U}(\mathfrak{B}_s))_{\varepsilon r/(16d)} \subset (B)_{\varepsilon r/8d}$. Set $\mathfrak{H} = \mathfrak{O}(B_{\varepsilon r/8d}) \setminus \mathfrak{B}_s$. Since $(B)_{\varepsilon r/8d}$ is convex, we can use Proposition 6.1 and Lemma 5.1 to ensure that

$$|\mathfrak{O}((B)_{\varepsilon r/8d})| \leq m^d \cdot \tau\left[\left(1 + \frac{\varepsilon}{4d}\right)^d + \varepsilon^2/64\right] < \tilde{\tau}_s(1 + \varepsilon/2),$$

where the final inequality holds for sufficiently large s and sufficiently small ε. $\mathfrak{B}_s$ has cardinality at least $\tilde{\tau}_s(1 - s^{-1/20})$; if s_0 is sufficiently large to ensure $s^{-1/20} < \varepsilon/2$, we find that $|\mathfrak{H}| < \varepsilon\tilde{\tau}_s$, and inclusion guarantees that

$$B \subset \mathfrak{U}(\mathfrak{B}_s \cup \mathfrak{H}).$$

This completes the proof. $\square$

6.2. *Relating the s-graded model and random geometric graph structure theorems.* Using the geometric information derived in the previous section, we wish to prove the following proposition.

PROPOSITION 6.4. *There is some $\varepsilon_0 > 0$ such that the following holds for any $\varepsilon < \varepsilon_0$. Take any $\delta > 0$ and $\tilde{\delta} \in [(1 - \varepsilon/16)\delta, \delta]$. Let $\mathscr{F}_n(\varepsilon)$ and $\mathscr{G}_{n,\tilde{\delta}}(s^{-1/20})$ be the events described in Theorems* 2.1 *and* 3.1, *respectively. Then there exist n_0 and s_0 depending only on ε and δ, such that if $n \geq n_0$ and $s \geq s_0$, then*

$$\mathscr{G}_{n,\tilde{\delta}}(s^{-1/20}) \subset \mathscr{F}_n(\varepsilon).$$

PROOF. Set $\tilde{\varepsilon} = s^{-1/20}$ and s_0 and n_0 to be the sufficiently large to ensure that Lemma 5.1 holds for any $s > s_0$ and $n \geq n_0$ (further conditions on s may be imposed on later in the proof). Assume $\mathscr{G}_{n,\tilde{\delta}}(\tilde{\varepsilon})$ occurs, and let $\mathfrak{B}$ and $\mathfrak{C}$ be the sets described in Theorem 3.1. Then, by definition,

$$(6.3) \qquad \left| \frac{\tilde{\tau}_s X_I}{q} - 1 \right| < \tilde{\varepsilon} \quad \forall I \in \mathfrak{B},$$

$$(6.4) \qquad |\mathfrak{C}| < \tilde{\varepsilon} \cdot \tilde{\tau}_s \quad \text{and} \quad \sum_{I \in \mathfrak{C}} X_I < \tilde{\varepsilon} q,$$

and

$$(6.5) \qquad X_I < \frac{\tilde{\varepsilon} q}{\tilde{\tau}_s} \quad \forall I \notin \mathfrak{B} \cup \mathfrak{C}.$$

Since the above bounds are in terms of q, whereas $\mathscr{F}_n(\varepsilon)$ is defined using $\sqrt{2\delta\mu}$, we begin by bounding $\sqrt{2\delta\mu}/q$ from above and below. $\tilde{\delta} \leq \delta$ by definition, and $\mu \geq \tilde{\mu}_s(1 + C/s)^{-1}$ by Lemma 5.1. Therefore,

$$\frac{\sqrt{2\delta\mu}}{q} = \sqrt{\frac{\delta\mu}{\tilde{\delta}\tilde{\mu}_s}} \geq \left(1 + \frac{C}{s}\right)^{-1/2} \geq (1 - \tilde{\varepsilon}),$$

where the final inequality holds for all s sufficiently large. Similarly, $\mu \leq \tilde{\mu}_s$ and $\tilde{\delta} > \delta(1 - \varepsilon/16)$ implies that

$$\frac{\sqrt{2\delta\mu}}{q} = \sqrt{\frac{\delta\mu}{\tilde{\delta}\tilde{\mu}_s}} \leq \left(1 - \frac{\varepsilon}{16}\right)^{-1/2} \leq 1 + \varepsilon/8.$$

Putting this together, we see that, for all sufficiently large s,

$$(6.6) \qquad (1 - \tilde{\varepsilon}) \leq \frac{\sqrt{2\delta\mu}}{q} \leq 1 + \varepsilon/8.$$

We now set s to be sufficiently large so that, when we apply Lemma 6.3 to $\mathfrak{B}$ to produce B, a ball of radius r, and $\mathfrak{H} \subset T$ with $B \subset \mathfrak{U}(\mathfrak{B} \cup \mathfrak{H})$, we can be certain that $|\mathfrak{H}| < (\varepsilon/8) \cdot \tilde{\tau}_s$ and that $\mathfrak{B} \subset (B)_{(\nu\varepsilon^2 r)/(d \cdot 2^{d+3})}$ (the slightly odd constants are chosen to make later computations simpler). We also require that $\tilde{\varepsilon} < \varepsilon^2/16$. Our goal is to show that B will satisfy the conditions of $\mathscr{F}_n(\varepsilon)$.

For any $S \subset \mathbb{T}^d$, it is straightforward to see that

$$\sum_{I \in \mathfrak{R}(S)} X_I \leq |\chi(S)| \leq \sum_{I \in \mathfrak{O}(S)} X_I.$$

To prove the first condition of $\mathscr{F}_n(\varepsilon)$, it is sufficient to show that, for any convex $S \subset B$,

$$\sum_{I \in \mathfrak{R}(S)} X_I > \left(\frac{\lambda(S)}{\tau} - \varepsilon\right)\sqrt{2\delta\mu} \quad \text{and} \quad \sum_{I \in \mathfrak{D}(S)} X_I < \left(\frac{\lambda(S)}{\tau} + \varepsilon\right)\sqrt{2\delta\mu}.$$

For the upper bound, we use Proposition 6.1 and Lemma 5.1 to see that

$$|\mathfrak{D}(S)| = m^d \lambda(\mathfrak{U}(\mathfrak{D}(S))) < \left(\frac{\lambda(S)}{\tau} + \frac{\varepsilon^2}{64}\right) m^d \tau \leq \left(\frac{\lambda(S)}{\tau} + \frac{\varepsilon^2}{64}\right)\tilde{\tau}_s.$$

We also know that, for any $\mathfrak{W} \subset T$,

(6.7) $$\sum_{I \in \mathfrak{W}} X_I \leq \sum_{I \in \mathfrak{C}} X_I + |\mathfrak{W}| \cdot \max_{I \notin \mathfrak{C}} X_I < \left(\tilde{\varepsilon} + \frac{|\mathfrak{W}|(1+\tilde{\varepsilon})}{\tilde{\tau}_s}\right)q,$$

using (6.3), (6.4) and (6.5). Applying this to $\mathfrak{D}(S)$ gives that

$$\sum_{I \in \mathfrak{D}(S)} X_I < \left[\tilde{\varepsilon} + \left(\frac{\lambda(S)}{\tau} + \frac{\varepsilon^2}{64}\right)(1 + \tilde{\varepsilon})\right]q$$

$$\leq \left[\frac{\tilde{\varepsilon}}{1 - \tilde{\varepsilon}} + \left(\frac{\lambda(S)}{\tau} + \frac{\varepsilon^2}{64}\right) \cdot \frac{1 + \tilde{\varepsilon}}{1 - \tilde{\varepsilon}}\right]\sqrt{2\delta\mu},$$

where the final inequality is (6.6). Since $\lambda(S)/\tau \leq 1$ (as $S \subset B$) and $\tilde{\varepsilon} < \varepsilon^2/16$, the right-hand side is smaller than $[\lambda(S)/\tau + \varepsilon]\sqrt{2\delta\mu}$ for all ε sufficiently small, as required.

To get a lower bound on the sum of the X_I's in $\mathfrak{R}(S)$, we observe that $S \subset B$ implies that $\mathfrak{R}(S) \subset \mathfrak{B} \cup \mathfrak{H}$ and, therefore,

$$|\mathfrak{R}(S) \cap \mathfrak{B}| \geq |\mathfrak{R}(S)| - |\mathfrak{H}|$$

$$> \left(\frac{\lambda(S)}{\tau} - \frac{\varepsilon^2}{64}\right)m^d \tau - \frac{\varepsilon \cdot \tilde{\tau}_s}{8}$$

$$\geq \left(\frac{\lambda(S)}{(1 + \tilde{\varepsilon})\tau} - \frac{\varepsilon^2}{64(1 + \tilde{\varepsilon})} - \frac{\varepsilon}{8}\right)\tilde{\tau}_s,$$

using Proposition 6.1 and Lemma 5.1. By definition, $X_I \geq [q(1 - \tilde{\varepsilon})]/\tilde{\tau}_s$ for any $I \in \mathfrak{B}$ and, therefore,

$$\sum_{I \in \mathfrak{R}(S)} X_I > \sum_{I \in \mathfrak{R}(S) \cap \mathfrak{B}} X_I$$

$$> \frac{1 - \tilde{\varepsilon}}{1 + \tilde{\varepsilon}}\left[\frac{\lambda(S)}{\tau} - \frac{\varepsilon^2}{64} - \frac{\varepsilon}{4}\right]q$$

$$> \frac{1 - \tilde{\varepsilon}}{(1 + \tilde{\varepsilon})(1 + \varepsilon/8)}\left[\frac{\lambda(S)}{\tau} - \frac{\varepsilon^2}{64} - \frac{\varepsilon}{4}\right]\sqrt{2\delta\mu}.$$

Reusing the bounds $\lambda(S) \leq \tau$ and $\tilde{\varepsilon} < \varepsilon^2/16$ gives that, for all ε sufficiently small, the final lower bound is smaller than $[\lambda(S)/\tau - \varepsilon]\sqrt{2\delta\mu}$, as required.

Next, let S' be a convex set disjoint from B with diameter at most r and $\lambda(S') > \varepsilon\tau$. A slightly more involved version of (6.7) gives that

$$\sum_{I \in \mathfrak{D}(S')} X_I < \sum_{I \in \mathfrak{C}} X_I + |\mathfrak{D}(S') \cap \mathfrak{B}| \cdot \max_{I \in \mathfrak{B}} X_I + |\mathfrak{D}(S') \cap (\mathfrak{B} \cup \mathfrak{C})^c| \cdot \max_{I \in (\mathfrak{B} \cup \mathfrak{C})^c} X_I$$

$$\leq \left[\tilde{\varepsilon} + \frac{|\mathfrak{D}(S') \cap \mathfrak{B}| \cdot (1 + \tilde{\varepsilon})}{\tilde{\tau}_s} + \frac{|\mathfrak{D}(S') \cap (\mathfrak{B} \cup \mathfrak{C})^c| \cdot \tilde{\varepsilon}}{\tilde{\tau}_s}\right]q.$$

Since $S' \cap B = \varnothing$, we know that $\mathfrak{U}(\mathfrak{R}(S')) \cap B = \varnothing$. We have assumed above that $\mathfrak{B} \subset (B)_{(\nu \varepsilon^2 r)/(d \cdot 2^{d+3})}$ and, therefore,

$$\mathfrak{U}(\mathfrak{R}(S') \cap \mathfrak{B}) \subset (B)_{(\nu \varepsilon^2 r)/(d \cdot 2^{d+3})} \setminus B.$$

Taking the measure of both sides and multiplying by m^d to get cardinality bounds, we find that

$$|\mathfrak{R}(S') \cap \mathfrak{B}| \leq m^d r^d \left[\left(1 + \frac{\nu \varepsilon^2}{d 2^{d+3}}\right)^d - 1 \right] < \frac{\varepsilon^2 m^d \tau}{4} \leq \varepsilon^2 \tilde{\tau}_s / 4.$$

Thus,

$$|\mathfrak{O}(S') \cap \mathfrak{B}| \leq |\mathfrak{O}(S') \setminus \mathfrak{R}(S')| + |\mathfrak{R}(S') \cap \mathfrak{B}| \leq (\varepsilon^2/32 + \varepsilon^2/4)\tilde{\tau}_s < (\varepsilon^2/2)\tilde{\tau}_s,$$

using Proposition 6.1 to bound the difference between the inner and outer hulls of the convex set S'. Bounding $|\mathfrak{O}(S') \cap (\mathfrak{B} \cup \mathfrak{C})^c|$ by $|\mathfrak{O}(S')|$, using inequality (6.6) to move from q to $\sqrt{2\delta\mu}$, and appealing to Proposition 6.1 one final time, we find that

$$\sum_{I \in \mathfrak{O}(S')} X_I < \left[\tilde{\varepsilon} + \frac{\varepsilon^2(1 + \tilde{\varepsilon})}{2} + \left(\frac{\lambda(S')}{\tau} + \frac{\varepsilon^2}{32}\right)\tilde{\varepsilon} \right] q$$

$$\leq \left[\frac{\tilde{\varepsilon}\tau}{\lambda(S')} + \frac{2\varepsilon^2 \tau}{3\lambda(S')} + \left(1 + \frac{\varepsilon^2 \tau}{32\lambda(S')}\right)\tilde{\varepsilon} \right](1 - \tilde{\varepsilon})^{-1} \cdot \frac{\lambda(S') \cdot \sqrt{2\delta\mu}}{\tau}.$$

Since $\lambda(S') > \varepsilon\tau$ by assumption, and $\tilde{\varepsilon} < \varepsilon^2/16$, the product of bracketed term and $(1 - \tilde{\varepsilon})^{-1}$ is bounded by ε for all sufficiently small values of ε. This completes the proof. $\square$

6.3. *Theorem* 3.1 *implies Theorem* 2.1. The next lemma establishes a lower bound on the event $\{|E| > (1 + \delta)\mu\}$—the conditioning event in Theorem 2.1.

LEMMA 6.5. *Fix* $\delta > 0$, *and let* $H_n(\delta) := \{|E| > (1 + \delta)\mu\}$. *Then, setting* $z = \max\{p/4, 3p/4 - 1/2\}$, *there exists* n_0 *such that* $n > n_0$ *implies that*

$$\mathbb{P}[H_n(\delta)] \geq \exp\left(-\sqrt{2\delta\mu}\left[\log\left(\frac{\sqrt{2\delta\mu}}{n \cdot \tau}\right) - 1\right] - Cn^z \log n\right),$$

where C *is an absolute constant independent of* n *and* δ.

PROOF. Fix B, a ball of diameter r, and let H' be the event that there are at least $\lceil \sqrt{2\delta\mu} + n^z \rceil$ vertices in B. Since the number of vertices in B is a Poisson random variable of mean $n\tau$, we can get very straightforward lower bounds on the probability of H':

$$\mathbb{P}[H'] \geq \mathbb{P}[\mathrm{Poisson}(n\tau) = \lceil \sqrt{2\delta\mu} + n^z \rceil]$$

$$= \frac{\exp(-n\tau) \cdot (n\tau)^{\lceil \sqrt{2\delta\mu} + n^z \rceil}}{(\lceil \sqrt{2\delta\mu} + n^z \rceil)!}.$$

By Stirling's approximation, it follows that, for sufficiently large n,

$$\mathbb{P}[H'] \geq \exp\left[-n\tau - \lceil \sqrt{2\delta\mu} + 2n^z \rceil \left(\log\left[\frac{\lceil \sqrt{2\delta\mu} + n^z \rceil}{n\tau}\right] - 1\right) - C \log n \right],$$

for some universal constant C. Thanks to our judicious choice of z, $n\tau \ll n^z \log n \ll \sqrt{2\delta\mu}$; therefore, by increasing n we can find a C independent of n that guarantees

(6.8) $$\mathbb{P}[H'] \geq \exp\left(-\sqrt{2\delta\mu}\left[\log\left(\frac{\sqrt{2\delta\mu}}{n \cdot \tau}\right) - 1\right] - Cn^z \log n\right).$$

If H' occurs, there exists a clique with at least $\delta\mu + n^z\sqrt{\delta\mu}$ edges in it (since $0 < z < p/2$, this quantity is a lower bound for the number of edges in a clique of $\lceil\sqrt{2\delta\mu}+n^z\rceil$ vertices for all sufficiently large n). Conditional on H', $H_n(\delta)$ occurs if the number of edges with at most one endpoint in B exceeds $\mu - n^z\sqrt{\delta\mu}$. Let $|E|'$ be the number of edges with *no* endpoints in B. Letting $1_{i,j}$ be the indicator of an edge between vertices i and j, we can see that

$$\mathbb{E}(|E|') = \mathbb{E}\left[\binom{N}{2}\mathbb{E}(1_{1,2}\cdot 1_{v_1,v_2\notin B}\mid N)\right]$$

$$= \frac{n^2}{2}\mathbb{P}(\{\|v_1 - v_2\|\leq r\}\cap\{v_1, v_2\notin B\}),$$

where N is the total number of point in the torus, as before, and the probability measure in the second equality is given by the uniform process. For notational convenience, let $1^B_{i,j}$ be the indicator of the event $\{\|v_i - v_j\|\leq r\}\cap\{v_i, v_j\notin B\}$, and μ^B be its expectation under the measure of the uniform process (by symmetry, this is independent of the indices i and j). If v_1 is at a distance greater than r from B, the second condition holds trivially. For a fixed B, the probability that v_1 is within distance r of B is a constant multiple of r^d. Thus,

$$\mu^B \geq (1 - Cr^d)vr^d,$$

for some C that depends only on the norm and the dimension. Thus, the expected value of $|E|'$ is bounded below by $\mu(1 - Cr^d)$. Therefore, by definition of z, we have the inequality

$$(6.9) \qquad \mu - \sqrt{\delta\mu}\cdot n^z \leq \mathbb{E}[|E|'] - \frac{\sqrt{\delta\mu}\cdot n^z}{4}.$$

We now need a variance estimate for $|E|'$:

$$\mathrm{Var}(|E|') = \mathbb{E}(\mathrm{Var}(|E|'\mid N)) + \mathrm{Var}(\mathbb{E}(|E|'\mid N)).$$

We have already calculated the expectation of $|E|'$ given N above; since μ^B does not depend on N, we deduce that

$$\mathrm{Var}(\mathbb{E}(|E|'\mid N)) = (\mu^B)^2\,\mathrm{Var}\left[\binom{N}{2}\right].$$

A standard calculation will show that the variance of $\binom{N}{2}$ is $n^3 + n^2/2$. Meanwhile,

$$\mu^B \leq \mathbb{P}(\|v_1 - v_2\|\leq r) = vr^d.$$

Combining these facts gives

$$\mathrm{Var}(\mathbb{E}(|E|'\mid N)) \leq Cr^{2d}n^3,$$

for some universal constant C.

Next, we estimate the expression $\mathbb{E}(\mathrm{Var}(|E|'\mid N))$. We can write this variance as

$$\mathrm{Var}(|E|'\mid N) = \mathbb{E}\left[\left(\sum_{1\leq i<j\leq N}[1^B_{i,j} - \mu^B]\right)^2\,\Big|\,N\right].$$

We now decompose this sum into three sums by distributing the square: one sum over pairs of the form (i, j), (k, l) with four distinct indices, one with pairs of the form (i, j), (i, k) where one index repeats and the final over perfect squares of terms involving (i, j). The expectation of the first one is zero, as the event that i, j form an edge with both endpoints outside of B is completely independent of the same event occurring over distinct vertices k, l. For a fixed choice of (i, j) and (i, k), we can bound

$$\mathbb{E}[1^B_{i,j}\cdot 1^B_{i,k}] \leq \mathbb{P}[\|v_i - v_j\|\leq r, \|v_i - v_k\|\leq r] = (vr^d)^2,$$

where the first inequality follows by removing the requirement that the vertices lie outside of B, and thus increasing the probability. There are $N(N-1)(N-2)$ ways to choose a pair of indices that overlap in exactly one entry. Thus,

$$\sum (1_{i,j}^B - \mu^B)(1_{i,k}^B - \mu^B) = \sum (1_{i,j}^B \cdot 1_{i,k}^B - (\mu^B)^2) \le C' r^{2d} N^3,$$

for some universal constant C'. Again, this overestimates the value of this sum dramatically, but is sufficient for our purposes. Finally, the contribution of terms of the form $(1_{i,j}^B - \mu^B)^2$ to the sum is exactly $\binom{N}{2}(\mu_B - \mu_B^2)$, which is bounded above by $C'' r^d N^2$. Combining these results, taking expectations over N, and then adding the contribution of the variance of the expectation from before, we conclude that

$$(6.10) \qquad \text{Var}(|E|') \le C''' (r^d n^2 + r^{2d} n^3),$$

for yet another universal constant C'''. $r^d n^2$ grows as $n^{p+o(1)}$, while $r^{2d} n^3$ grows as $n^{2p-1+o(1)}$. Thus, the variance of $|E|'$ is $n^{p+o(1)}$ if $p \le 1$, and $n^{2p-1+o(1)}$ when $p > 1$.

By Chebyshev's inequality and (6.9),

$$\mathbb{P}[|E|' < \mu - \sqrt{\delta\mu} \cdot n^z] \le \mathbb{P}\left[|E|' < \mathbb{E}[|E|'] - \frac{\sqrt{\delta\mu} \cdot n^z}{4}\right]$$

$$\le \frac{16\,\text{Var}(|E|')}{\delta\mu \cdot n^{2z}}.$$

Regardless of the value of p, this quantity vanishes as $n^{-p/2+o(1)}$ and, therefore, with probability $1 - \varepsilon$, $|E|'$ exceeds $\mu - \sqrt{\delta\mu} \cdot n^z$ for all sufficiently large n.

To conclude, we note that

$$\mathbb{P}[H_n(\delta)] \ge \mathbb{P}[H_n(\delta)|H'] \cdot \mathbb{P}[H']$$

$$\ge \mathbb{P}[|E|' \ge \mu - \sqrt{2\delta\mu} \cdot n^z] \cdot \mathbb{P}[H']$$

$$\ge (1-\varepsilon)\mathbb{P}[H'].$$

Substituting (6.8) completes the proof. $\square$

PROOF OF THEOREM 2.1. Fix $\delta, \varepsilon \in (0,1)$. In this proof, whenever we say "s sufficiently large", we mean "$s \ge s_0$ for some s_0 that depends only on δ and ε, and the universal constant $\tilde{\varepsilon}_0$ from Theorem 3.1", and whenever we say "n sufficiently large," we mean "$n \ge n_0$ for some n_0 that depends only on ε, δ, and our choice of s."

Let $\tilde{\delta}_0 = (1 - \varepsilon/16)\delta$ and $\tilde{\Delta}_0 = \delta$. Define $\tilde{\delta}$ to satisfy

$$\tilde{\delta}\tilde{\mu}_s = \delta\mu(1 - 1/(\log n)^2).$$

Although $\tilde{\delta}$ will depend on n, we have that $\tilde{\delta}_0 \le \tilde{\delta} \le \tilde{\Delta}_0$ for s and n sufficiently large (where we use Lemma 5.1 to bound $\mu/\tilde{\mu}_s$ from above and below). This allows us to apply Theorem 3.1. We also define

$$D_n := \{|E_s| - |E| > (1 + \tilde{\delta})\tilde{\mu}_s - (1 + \delta)\mu\}$$

$$= \left\{|E_s| - |E| > \tilde{\mu}_s - \mu - \frac{\delta\mu}{(\log n)^2}\right\}.$$

The definitions are chosen in such a way that $\{D_n \cap H_n(\delta)\}$ imply the event $\mathscr{L}_n(\tilde{\delta})$.

Let $\mathscr{F}_n(\varepsilon)$ be the event described in the statement of Theorem 2.1, that is, the existence of a ball B housing the giant clique. Our goal is to prove that

$$\lim_{n\to\infty} \mathbb{P}[\mathscr{F}_n(\varepsilon)^c | H_n(\delta)] = 0.$$

By Proposition 6.4, $\mathscr{G}_{n,\tilde{\delta}}(s^{-1/20})$ implies the event $\mathscr{F}_n(\varepsilon)$ whenever $\tilde{\delta} \in [(1 - \varepsilon/16)\delta, \delta]$, n and s are sufficiently large, and ε is sufficiently small. Then, taking complements, we find that

$$
\begin{aligned}
\mathbb{P}[\mathscr{F}_n(\varepsilon)^c | H_n(\delta)] &\leq \mathbb{P}[\mathscr{F}_n(\varepsilon)^c \cap D_n | H_n(\delta)] + \mathbb{P}[D_n^c | H_n(\delta)] \\
&\leq \frac{\mathbb{P}[\mathscr{G}_{n,\tilde{\delta}}(s^{-1/20})^c \cap D_n \cap H_n(\delta)]}{\mathbb{P}[H_n(\delta)]} + \mathbb{P}[D_n^c | H_n(\delta)] \\
&\leq \frac{\mathbb{P}[\mathscr{G}_{n,\tilde{\delta}}(s^{-1/20})^c \cap \mathscr{L}_n(\tilde{\delta})]}{\mathbb{P}[H_n(\delta)]} + \mathbb{P}[D_n^c | H_n(\delta)].
\end{aligned}
$$

We begin with the second term. To analyze it, we will use an instance of the celebrated FKG correlation inequality. First stated in the context of finite lattice systems [16], we will use a version that first appears in [23]. There is a natural partial ordering on configurations of χ: we say $\omega \prec \omega'$ if $\omega \subset \omega'$, that is, every point of ω is also in ω'. An event A is increasing with respect to this ordering if $\omega \prec \omega'$ and $\omega \in A$ implies $\omega' \in A$. Heuristically, A is increasing adding points to the configuration only makes A more likely to occur.

LEMMA 6.6 (The FKG inequality [23]).   *Let A and B increasing events. Then*

$$
\mathbb{P}[A | B] \geq \mathbb{P}[A].
$$

It is clear that $H_n(\delta)$ is an increasing event. In addition, we can write $|E_s| - |E|$ as a sum over pairs of points in the Poisson point process whose distance exceeds r, but whose corresponding indices satisfy $\rho(I, J) \leq s$. Thus, D_n is also an increasing event, and Lemma 6.6 allows us to deduce that

$$
\mathbb{P}[D_n | H_n(\delta)] \geq \mathbb{P}[D_n].
$$

Taking complements, we find that

$$
\mathbb{P}[\mathscr{F}_n(\varepsilon)^c | H_n(\delta)] \leq \frac{\mathbb{P}[\mathscr{G}_{n,\tilde{\delta}}(s^{-1/20})^c \cap \mathscr{L}_n(\tilde{\delta})]}{\mathbb{P}[H_n(\delta)]} + \mathbb{P}[D_n^c].
$$

It is easy to see that $\{|E_s| > \tilde{\mu}_s - \delta\mu/[2(\log n)^2]\}$ and $\{|E| < \mu + \delta\mu/[2(\log n)^2]\}$ imply D_n and, therefore, the union bound tells us that

$$
\mathbb{P}[D_n^c] \leq \mathbb{P}\left[|E_s| \leq \tilde{\mu}_s - \frac{\delta\mu}{2\log^2 n}\right] + \mathbb{P}\left[|E| \geq \mu + \frac{\delta\mu}{2\log^2 n}\right].
$$

Recalling (2.1) and (3.6), we can see that

$$
\begin{aligned}
\operatorname{Var}[|E|] &= \mu(1 + 2\nu n r^d) = \max\{n^{p+o(1)}, n^{2p-1+o(1)}\}, \\
\operatorname{Var}[|E_s|] &\leq 16|\tilde{S}_I||\tilde{N}_I|^2 m^d \cdot \max\{\mathscr{D}^2, \mathscr{D}^3\} = \max\{n^{p+o(1)}, n^{2p-1+o(1)}\},
\end{aligned}
$$

where we use the fact that $nr^d = n^{p-1+o(1)}$ (which follows from (2.3)) for the first line, and equations (3.1), (3.2) and the bounds on $|\tilde{S}_I|$ and $|\tilde{N}_I|$ from Lemma 5.1 to get control of the variance of $|E_s|$. Thus, Chebyshev's inequality gives that

$$
\mathbb{P}[D_n^c] \leq \frac{4\log^4 n(\operatorname{Var}[|E|] + \operatorname{Var}[|E_s|])}{\delta^2\mu^2} = n^{\max\{-p, -1\}+o(1)}.
$$

As n grows, this vanishes for all admissible p.

To complete the proof, we must show that, for s sufficiently large,

$$\lim_{n\to\infty} \frac{\mathbb{P}[\mathscr{G}_{n,\tilde{\delta}}(s^{-1/20})^c \cap \mathscr{L}_n(\tilde{\delta})]}{\mathbb{P}[H_n(\delta)]} = 0$$

as n grows. We will now use Theorem 3.1, which holds since $\tilde{\delta}$ is bounded above and below by $\delta(1 - \varepsilon/16)$ and δ, respectively. Reusing the convention $q = (2\tilde{\delta}\tilde{\mu}_s)^{1/2}$ and $w = \tilde{\tau}_s \cdot \mathscr{D}$, the quantitative bound of Theorem 3.1 and Lemma 6.5 gives us

$$\frac{\mathbb{P}[\mathscr{G}_{n,\tilde{\delta}}(s^{-1/20})^c \cap \mathscr{L}_n(\tilde{\delta})]}{\mathbb{P}[H_n(\delta)]} \le \exp\left(-q\left[\log\left(\frac{q}{w}\right) - 1\right] - \frac{q}{2}\left(\frac{1}{10 \cdot s^{1/20}}\right)^{10}\right.$$
$$\left. + \sqrt{2\delta\mu}\left[\log\left(\frac{\sqrt{2\delta\mu}}{n \cdot \tau}\right) - 1\right] + Cn^z \log n\right).$$

Here, the choice $\tilde{\delta}\tilde{\mu}_s = \delta\mu(1 - 1/(\log n)^2)$ becomes essential. Careful algebra will show that the first order terms, that is, those of order $\sqrt{2\delta\mu} \cdot \log n$ will cancel perfectly. In fact,

$$\sqrt{2\delta\mu}\left[\log\left(\frac{\sqrt{2\delta\mu}}{n \cdot \tau}\right) - 1\right] - q\left[\log\left(\frac{q}{w}\right) - 1\right] = q\log\left(\frac{w}{n\tau}\right) + q\log\left(\frac{\sqrt{2\delta\mu}}{q}\right)$$
$$+ (\sqrt{2\delta\mu} - q)\log\left(\frac{\sqrt{2\delta\mu}}{en\tau}\right).$$

By the choice of $\tilde{\delta}$, $\sqrt{2\delta\mu}/q = (1 - 1/(\log n)^2)^{-1/2}$, and hence, the bound $\log(1 - x) \ge -2x$ for all sufficiently small x implies that

$$\log\left(\frac{\sqrt{2\delta\mu}}{q}\right) = -\frac{1}{2}\log\left(1 - \frac{1}{(\log n)^2}\right) \le \frac{1}{(\log n)^2},$$

for all sufficiently large n. Similarly, $\sqrt{2\delta\mu} - q \le 2q/(\log n)^2$ for all sufficiently large n and s. Since $\sqrt{2\delta\mu}/en\tau = n^{(2-p)/2+o(1)}$ (which follows from (2.3)), we conclude that

$$\sqrt{2\delta\mu}\left[\log\left(\frac{\sqrt{2\delta\mu}}{n \cdot \tau}\right) - 1\right] - q\left[\log\left(\frac{q}{w}\right) - 1\right] \le q\left[\log\left(\frac{w}{n\tau}\right) + \frac{3-p}{2\log n}\right].$$

Recalling that $w/(n\tau) = \tilde{\tau}_s/(m^d\tau)$, we are left with

(6.11)
$$\frac{\mathbb{P}[\mathscr{G}_{n,\tilde{\delta}}(s^{-1/20})^c \cap \mathscr{L}_n(\tilde{\delta})]}{\mathbb{P}[H_n(\delta)]}$$
$$\le \exp\left(q\left[\log\left(\frac{\tilde{\tau}_s}{m^d\tau}\right) - \frac{1}{2 \cdot 10^{10}s^{1/2}} + \frac{3-p}{2\log n}\right] + Cn^z \log n\right).$$

From Lemma 5.1, we can see that $\tilde{\tau}_s/(m^d\tau) \le (1 + C/s)$, and thus, for sufficiently large values of s,

$$\log\left(\frac{\tilde{\tau}_s}{m^d\tau}\right) - \frac{1}{2 \cdot 10^{10}s^{1/2}} \le \frac{C}{s} - \frac{1}{2 \cdot 10^{10}s^{1/2}} \le -\frac{1}{4 \cdot 10^{10}s^{1/2}}.$$

For all sufficiently large s and n, we now see that the bracketed term of (6.11) will be negative. Since $z < p/2$, we now see that the exponent approaches negative infinity as n grows, and we deduce that

$$\lim_{n\to\infty} \frac{\mathbb{P}[\mathscr{G}_{n,\tilde{\delta}}(s^{-1/20})^c \cap \mathscr{L}_n(\tilde{\delta})]}{\mathbb{P}[H_n(\delta)]} = 0,$$

as required. $\quad\square$

7. Proof of the upper tail large deviation principle. We now prove Theorem 2.2, which claims that the function

$$I(x) := \left(\frac{2-p}{2}\right)\sqrt{2x}$$

is the upper tail rate function for the random variable $|E|$ with speed $s(n) = \sqrt{\mu}\log n$. Recall that we restrict our attention to subsets of the interval $(0, \infty)$, as our result only holds for events in which $|E|$ exceeds its expectation.

Instead of proving Theorem 2.2 directly, we will prove the following proposition instead.

PROPOSITION 7.1.   *Recall that, for any $\delta > 0$,*

$$H_n(\delta) = \{|E| > (1+\delta)\mu\} = \left\{\frac{|E| - \mu}{\mu} > \delta\right\}.$$

Then, for any $\delta > 0$ fixed,

(7.1) $$\lim_{n\to\infty} \frac{\log \mathbb{P}[H_n(\delta)]}{\sqrt{\mu}\log n} = -I(\delta).$$

The equivalence of this proposition to Theorem 2.2 is standard, but its proof is straightforward and we include it for completeness.

PROOF THAT PROPOSITION 7.1 IMPLIES THEOREM 2.2.   Pick F to be a closed subset of $(0, \infty)$, and let $a_F > 0$ be its leftmost endpoint. Since $I(x)$ is increasing, its infimum over F occurs at a_F. Furthermore, $F \subset [a_F, \infty)$ and, therefore, for any $\varepsilon > 0$ satisfying $(a_F - \varepsilon) > 0$,

$$\mathbb{P}\left[\frac{|E| - \mu}{\mu} \in F\right] \leq \mathbb{P}\left[\frac{|E| - \mu}{\mu} \in (a_F - \varepsilon, \infty)\right] = \mathbb{P}[H_n(a_F - \varepsilon)].$$

Taking the logarithm, dividing by $\sqrt{\mu}\log n$, applying (7.1) and letting ε go to zero gives the upper bound for F.

Next, take G open in $(0, \infty)$ and pick $b \in G$. For some $\varepsilon > 0$, we know that $(b - \varepsilon, b] \in G$. Therefore,

$$\mathbb{P}\left[\frac{|E| - \mu}{\mu} \in G\right] \geq \mathbb{P}\left[\frac{|E| - \mu}{\mu} \in (b - \varepsilon, b]\right] = \mathbb{P}[H_n(b - \varepsilon)] - \mathbb{P}[H_n(b)].$$

Applying (7.1) twice, we deduce that, for any $\tilde{\varepsilon} > 0$, there is an n sufficiently large to ensure that

$$\mathbb{P}\left[\frac{|E| - \mu}{\mu} \in G\right] \geq \exp\left(-(1 + \tilde{\varepsilon}) \cdot I(b - \varepsilon) \cdot \sqrt{\mu}\log n\right)$$

$$- \exp\left(-(1 - \tilde{\varepsilon}) \cdot I(b) \cdot \sqrt{\mu}\log n\right).$$

Picking $\tilde{\varepsilon}$ sufficiently small (as a function of ε) ensures that the second term is smaller than half the first term. Taking logarithms, dividing by $\sqrt{\mu}\log n$, and taking ε to zero establishes the lower bound on the probability of $\{(|E| - \mu)/\mu \in G\}$, and establishes Theorem 2.2.   □

PROOF OF PROPOSITION 7.1.   The statement that, for any $\varepsilon > 0$, there exists n sufficiently large to ensure that

$$\frac{\log \mathbb{P}[H_n(\delta)]}{\sqrt{\mu}\log n} > -(1 + \varepsilon)I(\delta)$$

is a direct consequence of Lemma 6.5. Thus, it will be sufficient an upper bound on the probability of $H_n(\delta)$.

Fix $\varepsilon > 0$. For an arbitrary pair of events A and B, assume that, conditional on A, the event B occurs with probability at least $1 - \varepsilon$. This implies that

$$\mathbb{P}[A] \leq \left(\frac{1}{1 - \varepsilon}\right)\mathbb{P}[B].$$

By Theorem 2.1, there exists a sufficient large n such that conditioning on $H_n(\delta)$ implies that the random geometric graph has a clique of size at least $\sqrt{2\delta\mu}(1 - \varepsilon)$ with probability at least $1 - \varepsilon$. This means that, for any s, the s-graded model includes a maximal clique set $\mathfrak{P} \subset T$ with at least as many vertices as the clique of the random geometric graph. Since every maximal clique set has $\tilde{\tau}_s$ indices, there can be at most $m^{d\tilde{\tau}_s}$ distinct maximal clique sets; this is an egregious overcount, but we have no need for finer control. Thus, by the union bound, the probability that there exists a maximal clique set with $\sqrt{2\delta\mu}(1 - \varepsilon)$ vertices is bounded above by $m^{d\tilde{\tau}_s}$ times the probability that a single one has the same property. The number of vertices in a maximal clique set is distributed as a Poisson random variable of mean $\tilde{\tau}_s \cdot \mathscr{D} = w$. Therefore, the chain of implication allows us to conclude that

$$\mathbb{P}[H_n(\delta)] \leq \left(\frac{m^{d\tilde{\tau}_s}}{1 - \varepsilon}\right)\mathbb{P}[\text{Poisson}(w) > \sqrt{2\delta\mu}(1 - \varepsilon)].$$

Let $v := \sqrt{2\delta\mu}(1 - \varepsilon)$. Applying the Chernoff bounds of Poisson random variables (see Lemma 5.2) to the right-hand side above gives

$$\mathbb{P}[H_n(\delta)] \leq \left(\frac{m^{d\tilde{\tau}_s}}{1 - \varepsilon}\right)\exp\left(-v\left[\log\left(\frac{v}{w}\right) - 1\right] - w\right)$$
$$\leq \exp\left(-(1 - 2\varepsilon)\sqrt{2\delta\mu}\log\left[\frac{\sqrt{\mu}}{w}\right]\right),$$

where the second inequality follows for all sufficiently large n by noting that all the missing terms vanish in comparison to $\sqrt{\mu} \cdot \log n$, and can therefore be absorbed at the cost of changing ε to 2ε. By the definitions of μ, p and w,

$$\frac{\sqrt{\mu}}{w} = n^{(2-p)/2+o(1)}.$$

Therefore, for any $\eta > 0$, there exists an n sufficiently large to ensure that

$$\frac{1}{\sqrt{\mu}\log n}\log\mathbb{P}[H_n(\delta)] \leq -(1 - 2\varepsilon)\left(\frac{2 - p}{2} + \eta\right)\sqrt{2\delta}.$$

Since ε and η are arbitrary, we conclude the desired upper bound. $\square$

REFERENCES

[1] ALON, N. and SPENCER, J. H. (2008). *The Probabilistic Method*, 3rd ed. *Wiley-Interscience Series in Discrete Mathematics and Optimization*. Wiley, Hoboken, NJ. MR2437651 https://doi.org/10.1002/9780470277331

[2] AUGERI, F. (2018). Nonlinear large deviation bounds with applications to traces of Wigner matrices and cycles counts in Erdos–Renyi graphs. Preprint. Available at arXiv:1810.01558.

[3] AVIN, C. and ERCAL, G. (2007). On the cover time and mixing time of random geometric graphs. *Theoret. Comput. Sci.* **380** 2–22. MR2330638 https://doi.org/10.1016/j.tcs.2007.02.065

[4] AZUMA, K. (1967). Weighted sums of certain dependent random variables. *Tohoku Math. J.* (2) **19** 357–367. MR0221571 https://doi.org/10.2748/tmj/1178243286

[5] BHATTACHARYA, B. B., GANGULY, S., LUBETZKY, E. and ZHAO, Y. (2017). Upper tails and independence polynomials in random graphs. *Adv. Math.* **319** 313–347. MR3695877 https://doi.org/10.1016/j.aim.2017.08.003

[6] BURAGO, J. D. and ZALGALLER, V. A. (1980). *Geometric Inequalities*. "Nauka" Leningrad. Otdel., Leningrad. MR0602952

[7] CHATTERJEE, S. (2012). The missing log in large deviations for triangle counts. *Random Structures Algorithms* **40** 437–451. MR2925306 https://doi.org/10.1002/rsa.20381

[8] CHATTERJEE, S. (2016). An introduction to large deviations for random graphs. *Bull. Amer. Math. Soc.* (*N.S.*) **53** 617–642. MR3544262 https://doi.org/10.1090/bull/1539

[9] CHATTERJEE, S. (2017). A note about the uniform distribution on the intersection of a simplex and a sphere. *J. Topol. Anal.* **9** 717–738. MR3684622 https://doi.org/10.1142/S1793525317500224

[10] CHATTERJEE, S. and DEMBO, A. (2016). Nonlinear large deviations. *Adv. Math.* **299** 396–450. MR3519474 https://doi.org/10.1016/j.aim.2016.05.017

[11] COOK, N. A. and DEMBO, A. (2018). Large deviations of subgraph counts for sparse Erdos–Renyi graphs. Preprint. Available at arXiv:1809.11148.

[12] DEMARCO, B. and KAHN, J. (2012). Upper tails for triangles. *Random Structures Algorithms* **40** 452–459. MR2925307 https://doi.org/10.1002/rsa.20382

[13] DEMBO, A. and ZEITOUNI, O. (2010). *Large Deviations Techniques and Applications. Stochastic Modelling and Applied Probability* **38**. Springer, Berlin. MR2571413 https://doi.org/10.1007/978-3-642-03311-7

[14] ELDAN, R. (2018). Gaussian-width gradient complexity, reverse log-Sobolev inequalities and nonlinear large deviations. *Geom. Funct. Anal.* **28** 1548–1596. MR3881829 https://doi.org/10.1007/s00039-018-0461-z

[15] FEDERER, H. (1969). *Geometric Measure Theory. Die Grundlehren der Mathematischen Wissenschaften* **153**. Springer, New York. MR0257325

[16] FORTUIN, C. M., KASTELEYN, P. W. and GINIBRE, J. (1971). Correlation inequalities on some partially ordered sets. *Comm. Math. Phys.* **22** 89–103. MR0309498

[17] GILBERT, E. N. (1961). Random plane networks. *J. Soc. Indust. Appl. Math.* **9** 533–543. MR0132566

[18] GOEL, A., RAI, S. and KRISHNAMACHARI, B. (2005). Monotone properties of random geometric graphs have sharp thresholds. *Ann. Appl. Probab.* **15** 2535–2552. MR2187303 https://doi.org/10.1214/105051605000000575

[19] GRIMMETT, G. (1999). *Percolation*, 2nd ed. *Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences]* **321**. Springer, Berlin. MR1707339 https://doi.org/10.1007/978-3-662-03981-6

[20] HAFNER, R. (1972). The asymptotic distribution of random clumps. *Computing* (*Arch. Elektron. Rechnen*) **10** 335–351. MR0329136 https://doi.org/10.1007/bf02242245

[21] HAREL, M., MOUSSET, F. and SAMOTIJ, W. (2019). Upper tails via high moments and entropic stability. Preprint. Available at arXiv:1904.08212.

[22] HOEFFDING, W. (1963). Probability inequalities for sums of bounded random variables. *J. Amer. Statist. Assoc.* **58** 13–30. MR0144363

[23] JANSON, S. (1984). Bounds on the distributions of extremal values of a scanning process. *Stochastic Process. Appl.* **18** 313–328. MR0770197 https://doi.org/10.1016/0304-4149(84)90303-X

[24] JANSON, S. (2004). Large deviations for sums of partly dependent random variables. *Random Structures Algorithms* **24** 234–248. MR2068873 https://doi.org/10.1002/rsa.20008

[25] JANSON, S., OLESZKIEWICZ, K. and RUCIŃSKI, A. (2004). Upper tails for subgraph counts in random graphs. *Israel J. Math.* **142** 61–92. MR2085711 https://doi.org/10.1007/BF02771528

[26] JANSON, S. and RUCIŃSKI, A. (2002). The infamous upper tail **20** 317–342. MR1900611 https://doi.org/10.1002/rsa.10031

[27] KIM, J. H. and VU, V. H. (2000). Concentration of multivariate polynomials and its applications. *Combinatorica* **20** 417–434. MR1774845 https://doi.org/10.1007/s004930070014

[28] KLAIN, D. A. and ROTA, G.-C. (1997). *Introduction to Geometric Probability. Lezioni Lincee.* [*Lincei Lectures*]. Cambridge Univ. Press, Cambridge. MR1608265

[29] LUBETZKY, E. and ZHAO, Y. (2017). On the variational problem for upper tails in sparse random graphs. *Random Structures Algorithms* **50** 420–436. MR3632418 https://doi.org/10.1002/rsa.20658

[30] MEESTER, R. and ROY, R. (1996). *Continuum Percolation. Cambridge Tracts in Mathematics* **119**. Cambridge Univ. Press, Cambridge. MR1409145 https://doi.org/10.1017/CBO9780511895357

[31] MÜLLER, T. (2008). Two-point concentration in random geometric graphs. *Combinatorica* **28** 529–545. MR2501248 https://doi.org/10.1007/s00493-008-2283-3

[32] NAGAEV, S. V. (1979). Large deviations of sums of independent random variables. *Ann*. *Probab.* **7** 745–789. MR0542129

[33] PENROSE, M. (2003). *Random Geometric Graphs*. *Oxford Studies in Probability* **5**. Oxford Univ. Press, Oxford. MR1986198 https://doi.org/10.1093/acprof:oso/9780198506263.001.0001

[34] PENROSE, M. D. (2002). Focusing of the scan statistic and geometric clique number. *Adv. in Appl. Probab.* **34** 739–753. MR1938940 https://doi.org/10.1239/aap/1037990951

[35] PENROSE, M. D. and YUKICH, J. E. (2005). Normal approximation in geometric probability. In *Stein's Method and Applications*. *Lect. Notes Ser. Inst. Math. Sci. Natl. Univ. Singap.* **5** 37–58. Singapore Univ. Press, Singapore. MR2201885 https://doi.org/10.1142/9789812567673_0003

[36] PETERSEN, P. (2006). *Riemannian Geometry*, 2nd ed. *Graduate Texts in Mathematics* **171**. Springer, New York. MR2243772

[37] SCHREIBER, T. and YUKICH, J. E. (2005). Large deviations for functionals of spatial point processes with applications to random packing and spatial graphs. *Stochastic Process*. *Appl*. **115** 1332–1356. MR2152378 https://doi.org/10.1016/j.spa.2005.03.007

[38] TALAGRAND, M. (1995). Concentration of measure and isoperimetric inequalities in product spaces. *Inst. Hautes Études Sci*. *Publ. Math*. **81** 73–205. MR1361756

[39] TALAGRAND, M. (2003). *Spin Glasses*: *A Challenge for Mathematicians. Cavity and Mean Field Models. Ergebnisse der Mathematik und Ihrer Grenzgebiete*. 3. *Folge. A Series of Modern Surveys in Mathematics* [*Results in Mathematics and Related Areas*. 3*rd Series*. *A Series of Modern Surveys in Mathematics*] **46**. Springer, Berlin. MR1993891