Disproving the Limit Law for Random Geometric Graphs

Simi Haber, Tal Hershko, Tobias Müller

September 27, 2022

Abstract

The celebrated theory of random graphs investigates the asymptotic behavior of graph properties in different random graph models. An important class of graph properties are the *first order properties*, which are those expressible by means of first order logic. Studying the asymptotic behavior of first order graph properties is doubly beneficial — it sheds light upon both the underlying behavior of the random graph model and the expressive power of first order logic.

A classic result, discovered independently by Glebskii et al. [4] and Ron Fagin [2], states that in the binomial random graph $G(n, p)$ with a constant p, the limiting probability of every first order property is either 0 or 1. This phenomenon is known as a *Zero-One law*. A similar situation is that of a *Limit law*, in which it is only guaranteed that every first order property has some limiting probability. The existence of Zero-One laws and Limit laws has been well studied for the binomial graph model. In this paper we focus on the *random geometric graph* (RGG) model, which generates a graph by randomly placing n points in a metric space and joining them if they are close. Specifically, we disprove a conjecture which has remained unsolved since 2006 about the existence of a Limit law in this model [18].

1 Introduction

In this paper we study first order properties of random geometric graphs. To handle this interesting blend of probability, graph theory, geometry and logic, we begin by introducing the central definitions and results which stand at the basis of the theory.

1.1 Random Graphs

A random graph over n vertices is defined as a certain random variable, whose values are graphs with a fixed vertex set of size n [10].

Definition 1.1. Let $n \in \mathbb{N}$ and fix a vertex set V_n of size n. Let $\mathcal{G}_n$ be the set of all graphs $G = (V_n, E)$. A *random graph* is a random variable $\mathbf{G}_n : \Omega \to \mathcal{G}_n$ (over some probability space $(\Omega, \mathcal{F}, \mathbb{P})$).

A *random graph model* is simply a method for generating a random graph. It is typically associated with a certain family of random graph distributions. One of the simplest and most well-known models is the *binomial random graph* model $G(n,p)$. It is also commonly known as the Erdős-Rényi model, after Paul Erdős and Alfréd Rényi whose joint work marked the birth of the theory of random graphs.

Definition 1.2. A random graph $\mathbf{G}_n$ is a *binomial Erdős-Rényi graph* with parameter p if its distribution satisfies

$$\mathbb{P}\left(\mathbf{G}_n = G\right) = p^{|E|}(1-p)^{\binom{n}{2}-|E|}$$

for every graph $G = (V_n, E) \in \mathcal{G}_n$. In this case we denote $\mathbf{G}_n \sim G(n,p)$.

It is useful to think of the distribution $G(n,p)$ as a product of $\binom{n}{2}$ Bernoulli distributions $B(p)$. That is, $G(n,p)$ is obtained by letting each edge appear in the graph with probability p and independently of other edges.

Another important family of random graph models, known today as *random geometric graphs* (RGGs), was proposed by Edgar Gilbert in 1961 [3]. This model is obtained by placing n points randomly and independently in some metric space, according to a given probability distribution. A graph is then constructed by regarding those points as the vertices, and connecting two points if their distance is smaller than r (where r is a positive real number). We refer to Penrose [16] as a general reference about RGGs.

Definition 1.3. Fix a metric space (X, d), a probability measure μ on X and a positive real number $r > 0$. Let $v_1, \ldots, v_n$ be n independent and μ-distributed random points in X, and denote $V_n = \{v_1, v_2, \ldots, v_n\}$. Define a random graph $\mathbf{G}_n$ such that for every $i \neq j$,

$$v_i \sim v_j \iff d(v_i, v_j) < r.$$

Here $v_i \sim v_j$ denotes adjacency between the vertices v_i, v_j. $\mathbf{G}_n$ is called a *random geometric graph*, and we write $\mathbf{G}_n \sim G_X(n,r)$.

The theory of random graphs is mostly interested in the *asymptotic behavior* of a model. Hence we consider a sequence of random graphs $\{\mathbf{G}_n\}_{n=1}^{\infty}$. For the binomial model we have $\mathbf{G}_n \sim G(n, p(n))$ and for the geometric model we have $\mathbf{G}_n \sim G_X(n, r(n))$ (note that X, μ are considered fixed and only r depends on n). One typically investigates the asymptotic probabilities of different graph properties.

Definition 1.4. Let $\mathcal{G}$ be the class of all (finite) graphs. A *property* A is a subclass of $\mathcal{G}$ closed under isomorphism.

Definition 1.5. Let $\{\mathbf{G}_n\}_{n=1}^{\infty}$ be a sequence of random graphs and let A be a property of (finite) graphs. We say that A holds *asymptotically almost surely* (a.a.s.) if

$$\lim_{n\to\infty} \mathbb{P}\left(\mathbf{G}_n \in A\right) = 1.$$

We say that A holds *asymptotically almost never* (a.a.n.) if

$$\lim_{n\to\infty} \mathbb{P}\left(\mathbf{G}_n \in A\right) = 0.$$

Remark 1.6. If, for example, we have $\mathbf{G}_n \sim G(n,p)$, it is common to directly write $\mathbb{P}\left(G(n,p) \in A\right)$ instead of $\mathbb{P}\left(\mathbf{G}_n \in A\right)$.

1.2 First Order Logic

We now consider an interesting class of graph properties — properties that can be expressed by a sentence in the first order language of graph theory.

Definition 1.7. The *first order language of graphs*, denoted here by $\mathcal{L}_{\text{graph}}$, is composed of the following symbols:

1. Logical connectives: $\neg, \vee, \wedge, \rightarrow, \leftrightarrow$.
2. Quantifiers: $\forall, \exists$.
3. Variable symbols, which represent vertices.[1]
4. The equality relation $=$ and the adjacency relation $\sim$ (between vertices).

First order sentences are always strings of the symbols listed above (parentheses are also often included for readability). The grammar of first order logic dictates which strings indeed form valid logical formulas; those are known as *well formed formulas* (WFFs).

Definition 1.8. The well formed formulas are defined inductively as follows:

1. $x \sim y$ and $x = y$ are WFFs (for any variable symbols x, y).
2. If φ, ψ are WFFs then $\neg\varphi, \varphi \wedge \psi, \varphi \vee \psi, \varphi \rightarrow \psi, \varphi \leftrightarrow \psi$ are WFFs.
3. If φ is a WFF then $\forall x(\varphi)$ and $\exists x(\varphi)$ are WFFs (for any variable symbol x).

Definition 1.9. A well formed formula is called a *sentence* if every variable it contains always appears quantified.

For example, $x \sim y$ is not a sentence but $\forall x \exists y(x \sim y)$ is.

[1] Variable symbols are typically denoted by Latin lowercase letters such as x, y, z.

Definition 1.10. A graph property A is called a *first order property* if there exists a first order sentence φ which expresses A. That is, the class of graphs that satisfy φ is the exactly the class of graphs defined by A.

For example, the property "the graph contains a triangle" is first order, because the sentence

$$\exists x \exists y \exists z \left(x \sim y \wedge y \sim z \wedge z \sim x \right)$$

expresses it.[2] Also, the property "the graph contains exactly two vertices" is first order, because the sentence

$$\exists x \exists y \left(\neg (x = y) \wedge \forall z \left(z = x \vee z = y \right) \right)$$

expresses it.

In general, many natural graph properties are first order, but certainly not all of them. To name a few examples, it can be proved that connectivity and 2-colorability are not first order properties (see [17], Theorems 2.4.1 and 2.4.2).

We can now define two fundamental phenomena in the theory of random graph logic: Zero-One laws and Limit laws.

Definition 1.11. Let $\{\mathbf{G}_n\}_{n=1}^{\infty}$ be a sequence of random graphs. We say that it satisfies a *Zero-One law* if for every first order property A,

$$\lim_{n\to\infty} \mathbb{P}\left(\mathbf{G}_n \in A\right) \in \{0, 1\}.$$

We say that it satisfies a *Limit law* if for every first order property A,

$$\lim_{n\to\infty} \mathbb{P}\left(\mathbf{G}_n \in A\right) \text{ exists.}$$

1.3 Previous Results

The first Zero-One law was discovered and proved by Glebskii, Kogan, Liagonkii and Talanov in 1969, and independently by Fagin in 1976. Their classic result concerns binomial random graphs with $p = \frac{1}{2}$, and can be easily extended to any constant p.

Theorem 1.12. *Let $p \in (0,1)$ be constant. Then the sequence $\{G(n,p)\}_{n=1}^{\infty}$ satisfies a Zero-One law. That is, for every first order property A, either $\mathbb{P}\left(G(n,p) \in A\right) \to 0$ or $\mathbb{P}\left(G(n,p) \in A\right) \to 1$.*

In their paper from 1988, Joel Spencer and Saharon Shelah prove a wide variety of Zero-One laws for sparse binomial graphs [19].

[2] Note that we shall always assume that $x \sim y$ implies $x \neq y$, as we only consider simple, loopless graphs.

Theorem 1.13. *The sequence* $\{G(n,p(n))\}_{n=1}^{\infty}$ *satisfies a Zero-One law in each of the following cases:*

1. $n^{-1-\frac{1}{k}} \ll p(n) \ll n^{-1-\frac{1}{k+1}}$ *for a certain* $k \in \mathbb{N}$.
2. $n^{-1-\varepsilon} \ll p(n) \ll n^{-1}$ *for every* $\varepsilon > 0$.
3. $n^{-1} \ll p(n) \ll n^{-1}\ln n$.
4. $n^{-1}\ln n \ll p(n) \ll n^{-1+\varepsilon}$ *for every* $\varepsilon > 0$.

Here $f(n) \ll g(n)$ *means* $f(n) = o(g(n))$, *that is,* $\frac{f(n)}{g(n)} \to 0$.

In addition, the following two completing results are proved in [19].

Theorem 1.14. *Consider the sequence* $\{G(n,p(n))\}_{n=1}^{\infty}$ *for* $p(n) = n^{-\alpha}$, $\alpha \in (0,1)$.

1. *If* α *is irrational, then* $\{G(n,p(n))\}_{n=1}^{\infty}$ *satisfies a Zero-One law.*
2. *If* α *is rational, then* $\{G(n,p(n))\}_{n=1}^{\infty}$ *does not satisfy a Zero-One law. Furthermore, not even a Limit law holds: there exists a first order property* A *such that* $\lim_{n\to\infty} \mathbb{P}(G(n,p(n)) \in A)$ *does not exist.*

As an additional reference for the last two theorems, we recommend Spencer's book [17].

The work of Spencer and Shelah, also followed by the paper of Łuczak and Spencer [20], has given us comprehensive knowledge about the logic of binomial random graphs. There has also been research about the logic of other random graph models. Haber and Krivelevich [6] discussed first order Zero-One and Limit laws in the model of random regular graphs, and obtained analogous results to Theorem 1.13 (formulated above). Heinig, Müller, Noy and Taraz [7] proved that a Zero-One law holds for monadic second order logic for the random graph drawn uniformly from all connected graphs in an addable, minor-closed class of graphs. Kleinberg and Kleinberg [12] discussed the limiting behavior of the preferential attachment model, which results in "scale free" graphs that closely approximate complex real-world networks such as large social networks and the Internet. Their results show that a first order Zero-One law holds when the out-degree d of the model is either $d = 1$ or $d = 2$, but not when $d \geq 3$.

As for random geometric graphs, the first key result regarding Zero-One laws was proved by Gregory McColm [13], who studied the case of the 1-dimensional torus $\mathbb{T}^1 = S^1$. In his paper, he proves results similar to Theorem 1.13.

Theorem 1.15. *The sequence* $\{G_{S^1}(n,r(n))\}_{n=1}^{\infty}$ *satisfies a Zero-One law in each of the following cases:*

1. $n^{-1-\frac{1}{k}} \ll r(n) \ll n^{-1-\frac{1}{k+1}}$ *for a certain* $k \in \mathbb{N}$*.*

2. $n^{-1-\varepsilon} \ll r(n) \ll n^{-1}$ *for every* $\varepsilon > 0$*.*

3. $n^{-1} \ll r(n) \ll 1$*.*

4. $r(n) = r$ *for any constant* $r > 0$*.*

At the end of his paper, McColm discusses possible generalizations of his results to higher dimensions. He conjectures that a Zero-One law will still hold for every high-dimensional manifold which is "sufficiently nice" (e.g. it is required to be compact and connected).

However, in their 2006 paper, Joel Spencer and Amit Agarwal [18] show that McColm's conjecture fails, even for the simple case of the 2-dimensional flat torus $\mathbb{T}^2$ with a constant r (and a uniform distribution).[3] They prove the following result.

Theorem 1.16. *Set a constant* $0 < r < 0.1$*. Then the sequence* $\{G_{\mathbb{T}^2}(n,r)\}_{n=1}^{\infty}$ *does not satisfy a first order Zero-One law: there exists a first order property* A *such that the limit* $\lim_{n\to\infty} \mathbb{P}(G_{\mathbb{T}^2}(n,r) \in A)$ *is non-trivial.*

The condition $r < 0.1$ prevents "spurious cases" by ensuring locally Euclidean behavior inside $\mathbb{T}^2$, and could certainly be weakened.

Despite their "negative" result, Spencer and Agarwal end their paper with three positive conjectures. The first and main conjecture is that the sequence $\{G_{\mathbb{T}^2}(n,r)\}_{n=1}^{\infty}$ does actually satisfy a *Limit law*. Intuitively, this conjecture implies that the first order language of graphs is not expressive enough to unravel complex, non-converging behaviors within the graph. Assuming the first conjecture, the second conjecture is that for any $\varepsilon > 0$, there exists an algorithm that approximates the limiting probability of a given input sentence within ε. Finally, their third conjecture aims to extend the scope of "niceness", by claiming that a Zero-One law holds whenever $r = o(1)$ but $r = n^{o(1)}$.

Since the conjectures above had been posed, several new results were obtained regarding random geometric graphs. Müller [15] showed that for $G_n = G_{[0,1]^d}(n, r(n))$ with $nr(n)^d = o(\ln n)$, the probability distribution of the clique number $\omega(G_n)$, the chromatic number $\chi(G_n)$ and several other graph parameters all become concentrated on two consecutive integers. Mcdiarmid and Müller [14] further managed to sharpen the known results about the behavior of $\chi(G_n)$ and its relation to $\omega(G_n)$, considering the "phase change" range $r(n) = (1+o(1))\left(\frac{t\ln n}{n}\right)^{1/d}$ with $t > 0$ constant. Balogh, Bollobás, Krivelevich, Müller and Walters [1] answered a question by Penrose by showing that in the random geometric model, almost every graph becomes Hamiltonian exactly when it first becomes 2-connected.

[3] Recall that the d-dimensional flat torus $\mathbb{T}^d$ is the quotient $\mathbb{R}^d/\mathbb{Z}^d$ with the metric naturally inherited from the Euclidean metric of $\mathbb{R}^d$.

However, no significant results regarding the first order behavior of random geometric graphs were obtained. In particular, Spencer and Agarwal's conjectures have remained open problems.

1.4 Our Results

In this paper we disprove Spencer and Agarwal's main conjecture about the existence of a Limit Law for the random geometric graph.

Theorem 1.17. *There exists a first order property A such that, for any constant $0 < r < 0.1$, the limit*

$$\lim_{n\to\infty} \mathbb{P}\left(G_{\mathbb{T}^2}\left(n,r\right) \in A\right)$$

does not exist.

Note that Theorem 1.17 also makes Spencer and Agarwal's second conjecture irrelevant. Regarding their third conjecture, we point out that our results can be directly generalized to $r = o(1)$ as long as it approaches 0 "sufficiently slowly", which disproves the Limit law within the $r = o(1)$ region as well. See explanation in Section 8. However, this straightforward generalization cannot cover the *entire* region $r = o(1), r = n^{o(1)}$; a full description of the logical behavior of the RGG in the entire region is still an open question.

In the formulation of Theorem 1.17 we choose to concentrate on the two dimensional case, as was also done by Spencer and Agarwal. This choice is made mainly for the sake of concreteness; we believe that our proof should be generalizable to $G_{\mathbb{T}^d}\left(n,r\right)$ with $d > 2$. Other interesting generalizations would be different choices of r (specifically $r = n^{-\alpha}$ for constant α) and other metrics (specifically ℓ^p for $1 \leq p \leq \infty$). While we do not have full answers, we believe that our approach should be useful in these cases as well. Again, see Section 8 for a more elaborated discussion.

To prove Theorem 1.17 we develop the ability to encode arbitrary graph structures within $G_{\mathbb{T}^2}\left(n,r\right)$ by means of first order logic. Then, using graph structures that express an infinitely-alternating property of the size of a graph, we are able to express a first order property whose probability does not converge. Unraveling the graph structures hidden within the random geometric graph is this paper's primary result.

The structure of the paper is as follows. We begin with introducing several preliminary tools in Sections 2 and 3. Section 2 is more geometric in nature. In its first part, we construct first-order formulas which are able to approximate arbitrary (constant) distances between vertices of the RGG. In its second part, we construct an additional first-order formula $\mathrm{WS}(x,y)$ which approximates a distance of at least $n^{-1/6}$ between vertices. Vertices x, y which are at least $n^{-1/6}$ apart are called *well-spaced*. The notion of well-spacedness will be required later to prevent certain overlaps between graph structures.

Section 3, on the other hand, is entirely probabilistic. It introduces a multivariate generalization of a well-known simple result known as Waring's Theorem. This generalization is quite straightforward, yet very effective. It allows us to extract detailed information about a discrete multivariate distribution from its joint moments. We prefer it over other moments methods thanks to its generality and flexibility. For example, it supports the incorporation of errors in the evaluation of the moments. It also only uses the first K moments, where K is some slowly-increasing function of n, instead of all the moments. Due to these virtues, we refer to it as a "flexible" moments method.

In Section 4 we revert back to the random geometric model and define a first order extension formula $S = S(x_1, x_2, x_3; s_1, s_2, s_3, z)$. S is a crucial part of the proof as it shall later be used as the basic "building block" of graph structures within $G_{\mathbb{T}^2}(n, r)$.

In Section 5 we consider random variables which count S-extensions over different vertices, and estimate their joint moments. Then we utilize the flexible moments method from Section 3 to estimate their joint distribution. It turns out that under a certain geometric conditioning, they are approximately independent Poisson variables. This fact is what makes S a suitable basis for the construction of arbitrary graph structures.

Section 6 finishes the proof of our primary result: that first-order logic can be used to express arbitrary graph in the RGG. As promised, the building block of these structures is the first-order extension formula S. At the core of the proof lies a technical result about concentration of certain random variables (closely related to those which were considered in Section 5).

In Section 7 the proof of Theorem 1.17 is completed. The ability to encode arbitrary graph structures is used to construct a first order sentence A that refers to the size of the graph n, in such a way that its probability does not converge as $n \to \infty$. This part is mostly "logical", and hardly depends on the random graph model. Therefore it mostly repeats the arguments of Spencer and Shelah, who disproved the Limit law in $G(n, p)$ (see Theorem 1.14).

Section 8 discusses possible generalizations of our results and poses some of them as conjectures.

The Appendix is composed of two independent subsections. Subsection A.1 is a short, self-contained introduction to the Poissonization technique, a standard technique in the study of RGGs. Subsection A.2 is dedicated to the proof of the Negligibility Theorem (Theorem 5.7), an important technical result about the negligibility of certain "undesirable" events. It is strongly used in Sections 5 and 6.

2 Expressing Distances with First-Order Formulas

From now on we focus on the random geometric graph $G_{\mathbb{T}^2}(n,r)$ with a constant $0 < r < 0.1$. The vertices are $v_1, v_2, \ldots, v_n$, and they are independent and uniformly distributed random points in $\mathbb{T}^2$. Distance between points $a_1, a_2 \in \mathbb{T}^2$ will be denoted $\|a_1 - a_2\|$. For $a \in \mathbb{T}^2$ let $B_r(a)$ denote the (open) ball of radius r around a. Recall that $\mathbb{T}^2$ is the flat torus, therefore it behaves locally like the Euclidean space $\mathbb{R}^2$. We also use the asymptotic notation $f(x) \approx g(x)$ (as $x \to x_0$) to indicate that $\lim_{x \to x_0} \frac{f(x)}{g(x)} = 1$.

As explained in Section 1, first-order formulas (in the language of graph theory) only have access to the adjacency and equality relations, as well as some basic logical ingredients. Our first observation is that in the RGG, first-order formulas are actually able to uncover a lot of information about the underlying geometry of the graph. More explicitly, in this section we will construct first-order formulas which approximate specific *distances* between vertices. This can be considered as the first "hint" towards the expressive power of first-order logic in the RGG.

2.1 Arbitrary Constant Distances

This subsection is dedicated to the approximation of constant distances. It proves the following theorem.

Theorem 2.1. *For every (constant) $\alpha > 0$ and $\varepsilon > 0$, there exists a first order formula $\mathrm{D}(x,y)$ such that a.a.s., for every pair of different vertices x, y,*

1. *If $\|x - y\| < (\alpha - \varepsilon)r$ then $\mathrm{D}(x,y)$ holds.*
2. *If $\|x - y\| > (\alpha + \varepsilon)r$ then $\mathrm{D}(x,y)$ does not hold.*

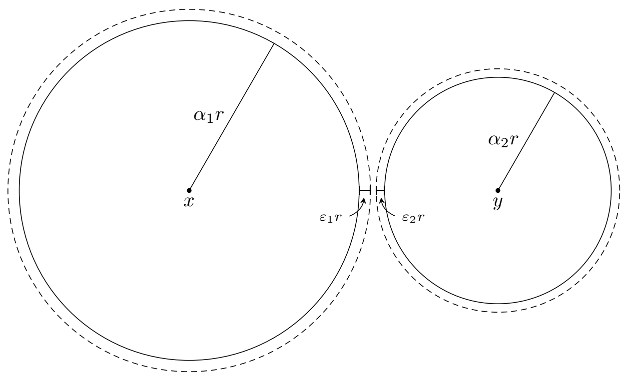
We call a first order formula $\mathrm{D}(x,y)$ which satisfies the conditions of Theorem 2.1 an (α, ε)-*approximator*, or simply an α-approximator when the value of ε is implied. The idea is that it approximates the distance condition $\|x - y\| < \alpha r$. Note that an (α, ε)-approximator is also a (α, ε')-approximator for any $\varepsilon' > \varepsilon$ so we may consider only arbitrarily small values of ε.

The strategy of the proof is as follows. First, we show how, given an α_1-approximator and an α_2-approximator (with sufficiently small ε-s), one can construct an $(\alpha_1 + \alpha_2)$-approximator and an $(\alpha_1 - \alpha_2)$-approximator (with slightly larger ε-s). Second, we explicitly construct a 1-approximator (this is trivially the adjacency relation) and a $\sqrt{3}$-approximator. Finally, we utilize the fact that the set $\{k + \ell\sqrt{3} : k, \ell \in \mathbb{Z}\}$ is dense in $\mathbb{R}$ to approximate arbitrary distances.

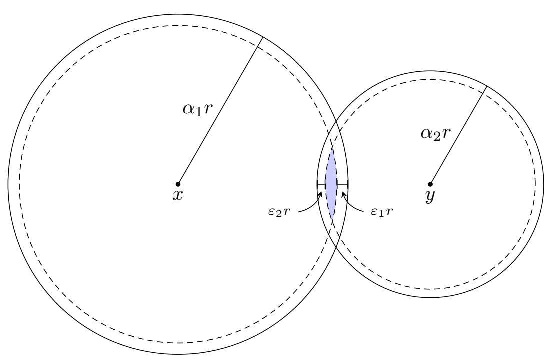
Lemma 2.2. *Assume that* $\mathrm{D}_1(x, y)$ *is an* $(\alpha_1, \varepsilon_1)$*-approximator and that* $\mathrm{D}_2(x, y)$ *is an* $(\alpha_2, \varepsilon_2)$*-approximator, for some* $\alpha_1, \alpha_2 > 0$ *and sufficiently small* $\varepsilon_1, \varepsilon_2 > 0$*. Denote* $\alpha = \alpha_1 + \alpha_2$ *and fix* $\varepsilon > \varepsilon_1 + \varepsilon_2$*. Then the formula*

$$\mathrm{D}(x, y) : \exists z \left(\mathrm{D}_1(x, z) \wedge \mathrm{D}_2(z, y)\right)$$

is an (α, ε)*-approximator.*



(a) The case $\|x - y\| > (\alpha + \varepsilon)r$.



(b) The case $\|x - y\| < (\alpha - \varepsilon)r$. The intersection L is shaded.

Figure 1: Illustration of the two parts of Lemma 2.2.

Proof. Let E_1 be the event that D_1 "fails" as an approximator; that is, the event that the conditions from Theorem 2.1 (with α_1, ε_1) fail for some x, y. Define E_2 similarly for D_2. The assumption that $\mathrm{D}_1, \mathrm{D}_2$ are approximators precisely means that $\mathbb{P}(E_1), \mathbb{P}(E_2) = o(1)$.

First, we show that outside the event $E_1 \cup E_2$, for every x, y, $\|x - y\| > (\alpha + \varepsilon)r$ implies $\neg\mathrm{D}(x, y)$. We do that through contraposition. Fix $x, y \in V$ and assume that $\mathrm{D}(x,y)$ holds. By definition, there exists a vertex z such that $\mathrm{D}_1(x,z)$ and $\mathrm{D}_2(z,y)$. Outside $E_1 \cup E_2$, this implies

$$\|x-z\| \le (\alpha_1+\varepsilon_1)r,$$
$$\|z-y\| \le (\alpha_2+\varepsilon_2)r.$$

The triangle inequality then yields $\|x-y\| \le (\alpha+\varepsilon)r$.

Second, we show that a.a.s., for every x,y, $\|x-y\| < (\alpha-\varepsilon)r$ implies $\mathrm{D}(x,y)$. Fix $x,y \in V$ and assume $\|x-y\| < (\alpha-\varepsilon)r$. Denote $\delta = \varepsilon - \varepsilon_1 - \varepsilon_2$; this is a positive constant, and by our assumptions

$$\|x-y\| \le (\alpha_1-\varepsilon_1)r + (\alpha_2-\varepsilon_2)r - \delta r. \tag{1}$$

Now consider the geometric locus

$$L = B_{(\alpha_1-\varepsilon_1)r}(x) \cap B_{(\alpha_2-\varepsilon_2)r}(y).$$

The area of L is minimal when inequality (1) is an equality, in which case it is still a positive constant since δ is a positive constant. Write $\mathrm{area}(L) \ge c$ for a constant $c > 0$. Then, given that $\|x-y\| < (\alpha-\varepsilon)r$, the probability that L does not contain any vertex z is $O\left((1-c)^n\right)$ (by independence between vertices). This bound decays exponentially, thus by taking the union bound over all $\binom{n}{2}$ pairs $x,y \in V$ we get that a.a.s. for every x,y, $\|x-y\| < (\alpha-\varepsilon)r$ implies that there exists a vertex z with

$$\|x-z\| < (\alpha_1-\varepsilon_1)r,$$
$$\|z-y\| < (\alpha_2-\varepsilon_2)r.$$

Outside the event $E_1 \cup E_2$, this implies that there exists z with $\mathrm{D}_1(x,z) \wedge \mathrm{D}_2(z,y)$, which exactly means that $\mathrm{D}(x,y)$ holds. That finishes the proof. Also see Figure 1, which illustrates the two cases of the proof. ■

Lemma 2.3. *Let $\mathrm{D}_1(x,y)$ be an (α_1,ε_1)-approximator and $\mathrm{D}_2(x,y)$ be an (α_2,ε_2)-approximator, for some $\alpha_1 > \alpha_2 > 0$ and sufficiently small $\varepsilon_1,\varepsilon_2 > 0$. Denote $\alpha = \alpha_1 - \alpha_2$ and fix $\varepsilon > \varepsilon_1 + \varepsilon_2$. Then the formula*

$$\mathrm{D}(x,y) : \forall z \left(\mathrm{D}_2(y,z) \to \mathrm{D}_1(x,z)\right)$$

is an (α,ε)-approximator.

Proof. The proof is very similar to that of the previous lemma. Let E_1, E_2 be as before.

First, we show that outside the event $E_1 \cup E_2$, for every x,y, $\|x-y\| < (\alpha-\varepsilon)r$ implies $\mathrm{D}(x,y)$. Fix $x,y \in V$ and assume $\|x-y\| < (\alpha-\varepsilon)r$. Note that

$$(\alpha-\varepsilon)r \le (\alpha_1-\varepsilon_1)r - (\alpha_2+\varepsilon_2)r.$$

Therefore, by the triangle inequality we have $B_{(\alpha_2+\varepsilon_2)r}(y) \subseteq B_{(\alpha_1-\varepsilon_1)r}(x)$. Now, for every vertex z, outside the event $E_1 \cup E_2$,

$$\begin{aligned}
\mathrm{D}_2(y,z) &\implies \|y-z\| \leq (\alpha_2+\varepsilon_2)r \\
&\implies \|x-z\| \leq (\alpha_1-\varepsilon_1)r \\
&\implies \mathrm{D}_1(x,z).
\end{aligned}$$

This exactly means that $\mathrm{D}(x,y)$ holds.

Second, we show that a.a.s., for every x, y, $\|x-y\| \geq (\alpha+\varepsilon)r$ implies $\neg\mathrm{D}(x,y)$. Fix $x, y \in V$ and assume $\|x-y\| \geq (\alpha+\varepsilon)r$. Denote $\delta = \varepsilon - \varepsilon_1 - \varepsilon_2$; by our assumptions

$$\|x-y\| \geq (\alpha_1+\varepsilon_1)r - (\alpha_2-\varepsilon_2)r + \delta r. \tag{2}$$

Now consider

$$L = \left[B_{(\alpha_1+\varepsilon_1)r}(x)\right]^c \cap B_{(\alpha_2-\varepsilon_2)r}(y).$$

The area of L is minimal when inequality (2) is an equality, in which case it is again a positive constant. Therefore, given that $\|x-y\| < (\alpha-\varepsilon)r$, the probability that L does not contain any vertex z is once again $O\left((1-c)^n\right)$. By the union bound, a.a.s., for every x, y, $\|x-y\| \geq (\alpha+\varepsilon)r$ implies that there exists a vertex z with

$$\begin{aligned}
\|x-z\| &> (\alpha_1+\varepsilon_1)r, \\
\|z-y\| &< (\alpha_2-\varepsilon_2)r.
\end{aligned}$$

Outside the event $E_1 \cup E_2$, this implies that there exists z with $\mathrm{D}_2(y,z) \wedge \neg\mathrm{D}_1(x,z)$, which exactly means that $\neg\mathrm{D}(x,y)$ holds. That finishes the proof. ■

Lemma 2.4.

1. *The formula* $\mathrm{D}_1(x,y) : x \sim y$ *is a* $(1,\varepsilon)$*-approximator for every* $\varepsilon > 0$.
2. *The formula*

$$\mathrm{D}_{\sqrt{3}}(x,y) : \exists z_1, z_2 \left(z_1 \sim x \wedge z_1 \sim y \wedge z_2 \sim x \wedge z_2 \sim y \wedge \neg\left(z_1 \sim z_2\right)\right)$$

(x, y have two non-adjacent common neighbors) is a $(\sqrt{3},\varepsilon)$*-approximator for every* $\varepsilon > 0$.

Proof. Part 1 is trivial; we prove part 2. Fix $\varepsilon > 0$.

First, fix $x, y \in V$ and assume $\|x-y\| > \left(\sqrt{3}+\varepsilon\right)r$. This inequality geometrically implies that $\mathrm{diam}(B_r(x) \cap B_r(y))$ is smaller than r. Therefore $\neg\mathrm{D}_{\sqrt{3}}(x,y)$ holds (even deterministically).

Second, fix $x, y \in V$ and assume $\|x-y\| < \left(\sqrt{3}-\varepsilon\right)r$. Now

$$\mathrm{diam}(B_r(x) \cap B_r(y)) \geq r + \varepsilon'$$

for some constant $\varepsilon' > 0$. Consequently, it is easy to see that there is a constant $c > 0$ for which there are two geometric loci $L_1, L_2 \subseteq B_r(x) \cap B_r(y)$ with

$$\text{area}(L_1), \text{area}(L_2) \geq c$$

and

$$\min_{a_1 \in L_1, a_2 \in L_2} \|a_1 - a_2\| \geq r.$$

Just as in the previous proofs, the probability that L_1 does not contain a vertex z_1 decays exponentially, and the same is true for L_2. By the union bound, a.a.s. for every x, y, $\|x - y\| < \left(\sqrt{3} - \varepsilon\right) r$ implies that there exist vertices z_1, z_2 which are non-adjacent common neighbors of x, y. This exactly means that $\mathrm{D}_{\sqrt{3}}(x, y)$ holds. ■

Remark 2.5. The proof of Lemma 2.4 actually works for ε which is much smaller than a constant. For the proof to work, we only need

$$\text{area}(L_1), \text{area}(L_2) \gg \frac{1}{n}$$

since it still assures that a.a.s. each of them contains a vertex. Indeed, recall that a locus L with area A contains no vertices with probability $(1 - A)^{\Theta(n)} \leq \exp\left(-\Theta(nA)\right)$. Now it can be easily seen that one can always take

$$\text{area}(L_1), \text{area}(L_2) = \Theta\left((\varepsilon')^2\right) = \Theta(\varepsilon^2).$$

Therefore the proof actually works for any $n^{-1/2} \ll \varepsilon \ll 1$. Although right now we do not actually care about non-constant ε, this observation will be useful later, when we handle negligible events in the Appendix (see Proposition A.10).

The following corollary is a direct consequence of the previous lemmas.

Corollary 2.6. *Assume $\beta = k + \ell\sqrt{3}$ is a positive constant for some $k, \ell \in \mathbb{Z}$. Then for every $\varepsilon > 0$ there exists a first order formula $\mathrm{D}(x, y)$ which is a (β, ε)-approximator.*

We are now ready for the proof of the main theorem.

Proof of Theorem 2.1. Fix α and ε. The set $\{k + \ell\sqrt{3} : k, \ell \in \mathbb{Z}\}$ is dense in $\mathbb{R}$, so there exists $\beta > 0$ in that set such that $\beta - \frac{\varepsilon}{2} < \alpha < \beta + \frac{\varepsilon}{2}$. From the previous corollary, there exists a $\left(\beta, \frac{\varepsilon}{2}\right)$-approximator $\mathrm{D}(x, y)$. We show that it is also an (α, ε)-approximator. Indeed,

1. If $\|x - y\| < (\alpha - \varepsilon)r$ then $\|x - y\| < \left(\beta - \frac{\varepsilon}{2}\right) r$, therefore $\mathrm{D}(x, y)$ holds.
2. If $\|x - y\| > (\alpha + \varepsilon)r$ then $\|x - y\| > \left(\beta + \frac{\varepsilon}{2}\right) r$, therefore $\mathrm{D}(x, y)$ does not hold.

That finishes the proof. ■

Finally, sometimes it is more convenient to approximate the equality $\|x-y\| = \alpha r$ instead of the inequality $\|x-y\| < \alpha r$. For that we can use the following corollary.

Lemma 2.7. *For every $\alpha > 0$ and $\varepsilon > 0$, there exists a first order formula* $\mathrm{E}(x,y)$ *such that a.a.s., for every pair x, y,*

1. $\mathrm{E}(x,y)$ *implies* $(\alpha-\varepsilon)r \leq \|x-y\| \leq (\alpha+\varepsilon)r$.
2. $\left(\alpha - \frac{\varepsilon}{2}\right) r \leq \|x-y\| \leq \left(\alpha + \frac{\varepsilon}{2}\right) r$ *implies* $\mathrm{E}(x,y)$.

Proof. Let $\mathrm{D}_*(x,y)$ be an $\left(\alpha - \frac{3\varepsilon}{4}, \frac{\varepsilon}{4}\right)$-approximator and let $\mathrm{D}^*(x,y)$ be an $\left(\alpha + \frac{3\varepsilon}{4}, \frac{\varepsilon}{4}\right)$-approximator. Define

$$\mathrm{E}(x,y) : \mathrm{D}^*(x,y) \wedge \neg \mathrm{D}_*(x,y).$$

Then, from the definitions, it is easy to see that $\mathrm{E}(x,y)$ satisfies the conditions of the corollary. ■

2.2 A Well-Spaced Formula

For technical reasons, we would also want to be able to express with first order logic the fact that two vertices are "not too close" to each other. In this case we say that the vertices are *well-spaced*. A distance of at least $n^{-1/6}$ turns out to be a good fit. The formal statement is as follows.

Theorem 2.8. *For every $\delta > 0$, there exists a first order formula* $\mathrm{WS}(x,y)$ *such that a.a.s., for every pair of different vertices x, y,*

1. *If* $\|x-y\| \leq n^{-1/6-\delta}$ *then* $\mathrm{WS}(x,y)$ *does not hold.*
2. *If* $\|x-y\| \geq n^{-1/6+\delta}$ *then* $\mathrm{WS}(x,y)$ *holds.*

The proof requires some additional geometric manipulations, with the main geometric tool being the *lens* between two points.

Definition 2.9. Let $a_1, a_2 \in \mathbb{T}^2$ be two points. The *lens* between a_1, a_2, denoted $L_{a_1 a_2}$, is defined as the intersection of the two balls of radius r around them:

$$L_{a_1 a_2} = B_r(a_1) \cap B_r(a_2).$$

For a given lens, we define its diameter and width as described in Figure 2.

Remark 2.10. Given two vertices x, y, the property of being inside L_{xy} is first-order expressible: $z \in L_{xy} \iff z \sim x \wedge z \sim y$.

To characterize the size of a lens L_{xy} we choose to use half its width as a parameter. We call this parameter h. Elementary geometry then leads to the following lemma.

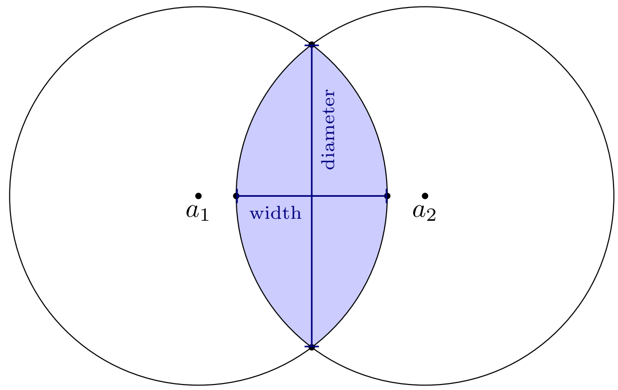


Figure 2: A lens between two points is shaded. The *diameter* is the length of the segment connecting the two intersection points of the circles. The *width* is the length of its perpendicular bisector (which is inside the lens).

Lemma 2.11. *Let* $L = L_{xy}$ *be a lens with* $\mathrm{width}(L) = 2h$*. Assume* $h > 0$ *(the lens is non-empty). Then:*

- $\|x - y\| = 2r - 2h$.
- $\mathrm{diam}(L) = 2\sqrt{2rh - h^2}$.
- $\mathrm{area}(L) = 2r^2 \arccos\left(1 - \frac{h}{r}\right) - 2(r - h)\sqrt{2rh - h^2}$.

As a useful corollary we write the resulting asymptotic estimations.

Corollary 2.12. *As* $h \to 0$*,*

- $\mathrm{diam}(L) \approx 2\sqrt{2} r^{1/2} h^{1/2}$.
- $\mathrm{area}(L) \approx \frac{8\sqrt{2}}{3} r^{1/2} h^{3/2}$.

The area estimation follows from the Taylor expansions

$$\sqrt{1 - x} \underset{x \to 0}{=} 1 - \frac{x}{2} + O(x^2),$$
$$\arccos(1 - x) \underset{x \to 0^+}{=} (2x)^{1/2} + \frac{1}{6\sqrt{2}} x^{3/2} + O(x^{5/2}).$$

Importantly, notice that the diameter asymptotically behaves like the square root of the width.

Lemma 2.13. *For a positive integer* ℓ*, let* $\mathrm{A}_\ell(x, y)$ *be the first-order formula which claims that* L_{xy} *contains exactly* ℓ *vertices. Then, given* $\delta > 0$*, there exists* $\ell \geq 1$ *such that a.a.s., for every pair of vertices* x, y*,*

$$\mathrm{A}_\ell(x, y) \to \mathrm{diam}(L_{xy}) \in \left[n^{-1/3-\delta}, n^{-1/3+\delta}\right].$$

Intuitively, demanding a fixed amount of vertices in L_{xy} means that its area has to be about n^{-1},which explains why its diameter must be about $n^{-1/3}$. Let us write a formal proof.

Proof. We start with the lower bound. Fix x, y; the complement of

$$\mathrm{A}_\ell(x,y) \to \mathrm{diam}(L_{xy}) \geq n^{-1/3-\delta}$$

is the event that $\mathrm{A}_\ell(x,y)$ holds and $\mathrm{diam}(L_{xy}) < n^{-1/3-\delta}$. Let us bound its probability. Given that $\mathrm{diam}(L_{xy}) < n^{-1/3-\delta}$, for some constant C we can write

$$\mathrm{area}(L_{xy}) \leq C\mathrm{diam}(L_{xy})^3 \leq Cn^{-1-3\delta}.$$

Therefore

$$\begin{aligned}
&\mathbb{P}\left(\mathrm{A}_\ell(x,y) \;\middle|\; \mathrm{diam}(L_{xy}) < n^{-1/3-\delta}\right) \\
&\leq \mathbb{P}\left(\exists \text{ at least } \ell \text{ vertices in } L_{xy} \;\middle|\; \mathrm{diam}(L_{xy}) \leq n^{-1/3-\delta}\right) \\
&\leq \binom{n}{\ell}\left(Cn^{-1-3\delta}\right)^\ell \leq n^\ell C^\ell n^{-\ell-3\delta\ell} = C^\ell n^{-3\delta\ell}.
\end{aligned}$$

For a fixed $\delta > 0$ we can choose any $\ell > \frac{2}{3\delta}$ and then the bound above is $o\left(n^{-2}\right)$. Taking a union bound we deduce that a.a.s. for every x, y

$$\mathrm{A}_\ell(x,y) \to \mathrm{diam}(L_{xy}) > n^{-1/3-\delta}.$$

For the upper bound we follow a similar strategy. Again, we fix x, y and now bound the probability of $\mathrm{A}_\ell(x,y)$ and $\mathrm{diam}(L_{xy}) > n^{-1/3+\delta}$. Given that $\mathrm{diam}(L_{xy}) > n^{-1/3+\delta}$, for some constant c we can now write

$$\mathrm{area}(L_{xy}) \geq c\mathrm{diam}(L_{xy})^3 \geq cn^{-1+3\delta}.$$

Thus,

$$\begin{aligned}
&\mathbb{P}\left(\mathrm{A}_\ell(x,y) \;\middle|\; \mathrm{diam}(L_{xy}) > n^{-1/3+\delta}\right) \\
&\leq \mathbb{P}\left(\exists \text{ at least } n-\ell \text{ vertices not in } L_{xy} \;\middle|\; \mathrm{diam}(L_{xy}) \geq n^{-1/3+\delta}\right) \\
&\leq \binom{n}{\ell}\left(1 - cn^{-1+3\delta}\right)^{n-\ell} \leq n^\ell \exp\left(-c(n-\ell)n^{-1+3\delta}\right) \\
&= \exp\left(\ell \ln n - \Omega(n^{3\delta})\right).
\end{aligned}$$

This is definitely $o(n^{-2})$, therefore we can take a union bound again and get the desired result. ■

Lemma 2.13 presents a first order formula which defines a lens whose diameter is about $n^{-1/3}$. That is good, but for well-spacedness we want to define a lens with larger diameter, of about $n^{-1/6}$. To do that we use an additional geometric trick. We mentioned that the diameter of a lens asymptotically behaves like the square root of its width. Therefore, the diameter of a lens is at least $n^{-1/6}$ if we can fit inside it a lens of diameter $n^{-1/3}$ "perpendicularly".

Formally, we define the *angle* between two lenses L, L' as the angle between their diameters, and say that L, L' are *perpendicular* if the angle between their diameters is $\frac{\pi}{2}$. Given four vertices x, y, x', y', we can ensure that L_{xy} and $L_{x'y'}$ are "almost" perpendicular with a first-order formula by using the results of the previous subsection.

Definition 2.14. Let $\varepsilon > 0$, and let $\mathrm{E}_2(x, y)$ and $\mathrm{E}_{\sqrt{2}}(x, y)$ be two first order formulas which approximate the distances $2r$ and $\sqrt{2}r$ within a margin of ε, as in Corollary 2.7. Define the first order formula $\mathrm{EPER}(x, y, x', y')$ (with margin ε) as follows:

$$\mathrm{E}_2(x, y) \wedge \mathrm{E}_2(x', y') \wedge \mathrm{E}_{\sqrt{2}}(x, x') \wedge \mathrm{E}_{\sqrt{2}}(x', y) \wedge \mathrm{E}_{\sqrt{2}}(y, y') \wedge \mathrm{E}_{\sqrt{2}}(y', x).$$

If $\mathrm{EPER}(x, y, x', y')$ holds, we say that the lenses L_{xy} and $L_{x'y'}$ are E-*perpendicular (with margin ε).*

See an illustration of EPER in Figure 3.

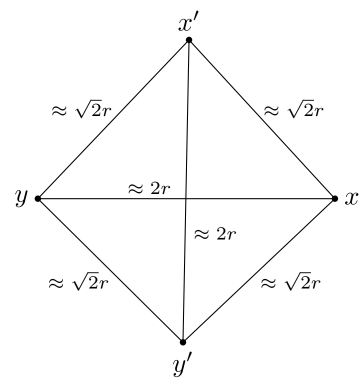


Figure 3: A geometric illustration of the formula $\mathrm{EPER}(x, y, x', y')$.

Remark 2.15. E-perpendicularity indeed a.a.s. implies "almost perpendicularity". More formally, a.a.s., for every x, y, x', y', $\mathrm{EPER}(x, y, x', y')$ (with margin ε) implies that the angle between the diameters of $L_{xy}, L_{x'y'}$ is $\frac{\pi}{2} + O(\varepsilon)$. The idea is that starting with a perfect square and "nudging" each side by $O(\varepsilon)$ also "nudges" the angles by $O(\varepsilon)$. This can be shown formally via the Sine and Cosine Theorems.

With all the logical and geometric ingredients at our disposal, we can formulate a key lemma. It sets the ground for characterizing lenses L_{xy} with diameter at least $n^{-1/6}$ by trying to fit E-perpendicular lenses $L_{x'y'}$ such that $\mathrm{A}_\ell(x',y')$ holds.

Lemma 2.16. *Let $\delta > 0$; then there are $\ell \geq 1$ and $\varepsilon > 0$ such that a.a.s., for every pair of vertices x, y,*

1. *If $\mathrm{diam}(L_{xy}) \leq n^{-1/6-\delta}$ then, for every x', y' such that L_{xy}, $L_{x'y'}$ are E-perpendicular (with margin ε) and $\mathrm{A}_\ell(x',y')$ holds, $L_{x'y'} \cap L_{xy}$ contains less than ℓ vertices.*

2. *If $\mathrm{diam}(L_{xy}) \geq n^{-1/6+\delta}$ and $\|x-y\| \geq \left(2 - \frac{\varepsilon}{2}\right) r$ then there exists a pair x', y' such that $\mathrm{A}_\ell(x',y')$ holds, L_{xy}, $L_{x'y'}$ are E-perpendicular (with margin ε) and $L_{x'y'} \subseteq L_{xy}$.*

Proof. Let $\delta > 0$. Choose ℓ such that $\ell \geq \frac{4}{\delta}$. Notice that ℓ then satisfies the statement of Lemma 2.13 with respect to $\frac{\delta}{2}$ (another use of this requirement will appear soon). In addition, choose $\varepsilon > 0$ sufficiently small, such that a.a.s. E-perpendicularity with margin ε implies that the angle γ between the diameters satisfies $0.49\pi < \gamma < 0.51\pi$.

We start with the first part. We must show that a.a.s. there exist no x, y, x', y' such that $\mathrm{diam}(L_{xy}) \leq n^{-1/6-\delta}$, L_{xy}, $L_{x'y'}$ are E-perpendicular, $\mathrm{A}_\ell(x',y')$ holds and $L_{x'y'} \cap L_{xy}$ contains ℓ vertices. As explained, a.a.s. E-perpendicularity implies $0.49\pi < \gamma < 0.51\pi$ and $\mathrm{A}_\ell(x',y')$ implies

$$\mathrm{diam}(L_{x'y'}) \in \left[n^{-\frac{1}{3}-\frac{\delta}{2}}, n^{-\frac{1}{3}+\frac{\delta}{2}}\right].$$

Therefore, it is sufficient to show that a.a.s. there exist no x, y, x', y' such that

$$\mathrm{diam}(L_{xy}) \leq n^{-1/6-\delta},$$
$$\mathrm{diam}(L_{x'y'}) \in \left[n^{-\frac{1}{3}-\frac{\delta}{2}}, n^{-\frac{1}{3}+\frac{\delta}{2}}\right],$$

L_{xy}, $L_{x'y'}$ form an angle of $0.49\pi < \gamma < 0.51\pi$ and $L_{x'y'} \cap L_{xy}$ contains ℓ vertices.

The idea is that the geometric requirements about the diameters and angle force $\mathrm{area}\,(L_{x'y'} \cap L_{xy})$ to be "too small", so with very high probability it does not contain ℓ vertices. More formally, let w, w' denote the widths of $L_{xy}, L_{x'y'}$ respectively. Each lens is contained in an infinite strip of the same width, parallel to the diameter. As a result, $\mathrm{area}\,(L_{x'y'} \cap L_{xy})$ can be bounded by the area of the intersection of the two strips, with widths w, w' and angle γ between them (which is a parallelogram). This area is trivially $\frac{1}{\sin\gamma} ww'$. The bounds $0.49\pi < \gamma < 0.51\pi$ force $\sin\gamma > 0.99$, and the fact that width $= \Theta\left(\text{diameter}^2\right)$ in small lenses forces

$$w = O\left(n^{-1/3-2\delta}\right),\ w' = O\left(n^{-2/3+\delta}\right).$$

Overall we have

$$\text{area}\left(L_{x'y'} \cap L_{xy}\right) = O\left(n^{-1-\delta}\right).$$

Now, just as in the proof of Lemma 2.13, the probability of having ℓ vertices in $L_{x'y'} \cap L_{xy}$ is $O\left(n^{\ell}\left(n^{-1-\delta}\right)^{\ell}\right) = O\left(n^{-\ell\delta}\right)$. Since $\ell > \frac{4}{\delta}$, this is $o(n^{-4})$, and all that is left is to take the union bound over x, y, x', y'.

The second part is more involved (generally, proving the existence of structures is harder than proving their absence). In its proof we apply a standard technique in the study of RGGs: *Poissonization*. Briefly, Poissonization replaces the random geometric graph $G_{\mathbb{T}^2}(n, r)$, which generates a fixed number of vertices n, with a Poisson random graph, in which the set of vertices is a Poisson point process. The Poisson random graph introduces a crucial property called *spatial independence:* the counts of vertices in disjoint areas of $\mathbb{T}^2$ are independent. See Subsection A.1 in the Appendix for a more detailed introduction of Poissonization.

Let us fix two vertices x, y and show that they satisfy the second part of the lemma with high probability. We need at least $1 - o(n^{-2})$ for the union bound; actually, we will obtain an exponentially high probability of $1 - \exp\left(-n^{\Omega(1)}\right)$. Let us fix the geometric positions of x, y; that is, we fix two points $a_1, a_2 \in \mathbb{T}^2$ and condition on the event $x = a_1, y = a_2$. We prove the desired bound $1 - \exp\left(-n^{\Omega(1)}\right)$ given this conditioning. Eventually, by integration over all possible positions, we obtain a non-conditioned bound of $1 - \exp\left(-n^{\Omega(1)}\right)$ (we use the implicit fact that this bound will be uniform in a_1, a_2).

We apply Poissonization: hold x, y and their positions fixed, and then instead of generating $n - 2$ additional vertices, consider a new random graph which generates the additional vertices as a Poisson point process with intensity $n - 2$. Following the notation from the Appendix, this means we replace $G_{\mathbb{T}^2}(n, r)$ with $G_{\mathbf{a}}(N, r)$ where $N - 2 \sim \text{Pois}(n - 2)$. Corollary A.5 implies that it is sufficient to show that x, y satisfy the second part of the lemma in $G_{\mathbf{a}}(N, r)$ with probability $1 - \exp\left(-n^{\Omega(1)}\right)$.

Assume that the geometric positions of x, y are fixed such that $\text{diam}(L_{xy}) \geq n^{-1/6+\delta}$ and $\|x - y\| \geq \left(2 - \frac{\varepsilon}{2}\right) r$ (otherwise there is nothing to prove). We must prove (with exponentially high probability, in $G_{\mathbf{a}}(N, r)$) the existence of vertices x', y' such that L_{xy}, $L_{x'y'}$ are E-perpendicular (with margin ε), $L_{x'y'} \subseteq L_{xy}$ and $\mathrm{A}_{\ell}(x', y')$ holds. To ensure E-perpendicularity, it is enough to require

$$\|x' - x\|, \|x' - y\| \in \left[\left(\sqrt{2} - \frac{\varepsilon}{2}\right) r, \left(\sqrt{2} + \frac{\varepsilon}{2}\right) r\right],$$
$$\|y' - x\|, \|y' - y\| \in \left[\left(\sqrt{2} - \frac{\varepsilon}{2}\right) r, \left(\sqrt{2} + \frac{\varepsilon}{2}\right) r\right].$$

Note that we omitted the requirements about $\|x - y\|$ and $\|x' - y'\|$, since $\left(2 - \frac{\varepsilon}{2}\right) r \leq \|x - y\| \leq 2r$ is already guaranteed from the way we fixed the positions of x, y and $\left(2 - \frac{\varepsilon}{2}\right) r \leq \|x' - y'\| \leq 2r$ is a.a.s. guaranteed from $\mathrm{A}_{\ell}(x', y')$ (and even with exponentially high probability; see proof of Lemma 2.13). The four requirements above have a clear geometric meaning: they mean that both x' and y' must lie in the intersection of annuli around x, y with radii $\left(\sqrt{2} \pm \frac{\varepsilon}{2}\right) r$. This intersection is comprised of two connected components $C^{(1)}, C^{(2)}$, and x', y' must lie each in a different component. See illustration in Figure 4a. $L_{x'y'} \subseteq L_{xy}$ is an additional requirement about the allowed positions of x', y'.

Let us now describe a way to find suitable x', y' with exponentially high probability. The first step is to find many potential vertices x' in $C^{(1)}$, and the second step is to show that at least one of them can be completed to a pair x', y'. We shall see that spatial independence plays an important part in both steps.

Step 1. Consider the lens L_{xy}; it satisfies $\mathrm{diam}(L_{xy}) \geq n^{-1/6+\delta}$ and therefore

$$\mathrm{width}(L_{xy}) = \Omega\left(n^{-1/3+2\delta}\right).$$

In contrast, $C^{(1)}$ and $C^{(2)}$ are both constant (since we regard ε as a constant and we fixed the positions of x, y). For convenience, in our description we shall refer to the axis parallel to xy as the "horizontal" axis, and to the perpendicular axis as the "vertical" axis (with $C^{(1)}$ being "up" and $C^{(2)}$ being "down"). We begin by creating a copy of L_{xy} shifted upwards by $2r$, denoted $L^{(1)}$ (again, see Figure 4a). Notice that $L^{(1)}$ is completely contained in $C^{(1)}$. We now prove that with probability $1-\exp\left(-n^{\Omega(1)}\right)$, there is a set of $n^{1/2}$ vertices in $L^{(1)}$ which are "spaced": each pair of vertices from the set keeps a horizontal distance of at least $\beta n^{-1/3}$ and a vertical distance of at least $\beta n^{-2/3}$ (where β is some sufficiently large constant). We demand this spacedness in order to avoid overlaps later, when we consider lenses $L_{x'y'}$.

For the proof, divide $L^{(1)}$ into rectangles of dimensions $\beta n^{-1/3} \times \beta n^{-2/3}$ (horizontal dimension first). Consider only quarter of those rectangles, evenly spaced like a grid, as depicted in Figure 4b. Rectangles at the boundary of $L^{(1)}$ can be neglected. The dimensions of $L^{(1)}$ (width and diameter) are $\Omega\left(n^{-1/3+2\delta}\right) \times \Omega\left(n^{-1/6+\delta}\right)$, thus there are $\Omega\left(n^{1/2+3\delta}\right)$ such rectangles. It is sufficient to show that at least $n^{1/2}$ of them contain a vertex. For each rectangle, the distribution of vertices inside is

$$\mathrm{Pois}\left((n-2) \cdot \beta^2 n^{-1}\right) = \mathrm{Pois}\left(\frac{n-2}{n}\beta^2\right)$$

therefore it contains a vertex with probability at least p for some constant $p > 0$. Crucially, these distributions are independent for different rectangles (spatial independence). The probability that at least $n^{1/2}$ of the specified rectangles contain a vertex is then

$$\mathbb{P}\left(\mathrm{Bin}\left(\Omega\left(n^{1/2+3\delta}\right), p\right) \geq n^{1/2}\right).$$

This is indeed $1-\exp\left(-n^{\Omega(1)}\right)$, as can be easily shown from Chernoff's bound.

Step 2. Following step 1, we may assume that there exists a set $x'_1, x'_2, \ldots, x'_t$ of spaced vertices in $L^{(1)}$, where $t = n^{1/2}$. For every $1 \leq i \leq t$, let $\Box_i$ be a square with dimensions $n^{-2/3} \times n^{-2/3}$, whose center is located exactly at a distance of $4r - 2n^{-2/3}$ below x_i', inside $C^{(2)}$. This is where we shall look for y_i'. The distribution of vertices inside $\square_i$ is Pois $\left((n-2) \cdot n^{-4/3}\right)$, therefore it contains a vertex with probability $\Theta\left(n^{-1/3}\right)$.

Suppose it is given that $\square_i$ contains a vertex y_i'. From the choice of positions, the lens $L_{x_i' y_i'}$ has a width of $\Theta\left(n^{-2/3}\right)$, a diameter of $\Theta\left(n^{-1/3}\right)$ and an area of $\Theta\left(n^{-1}\right)$. Moreover, it is "almost perpendicular" to L_{xy}, and contained in L_{xy}. Due to its area, the distribution of vertices inside $L_{x_i' y_i'}$ is Poisson with parameter $\Theta(1)$, and the probability that is contains exactly ℓ vertices is also $\Theta(1)$ (as ℓ is a constant).

Overall, for every $1 \le i \le t$, the probability of finding y_i' such that the pair x_i', y_i' satisfies all the desired properties is $\Theta\left(n^{-1/3}\right)$. Crucially, there is complete independence between the different i-s, because the spacedness of the x_i'-s guarantees that the squares $\square_i$ and also the lenses $L_{x_i' y_i'}$ cannot intersect. Indeed, the x_i'-vertices keep a horizontal distance of $\beta n^{-1/3}$ and a vertical distance of $\beta n^{-2/3}$ and for a sufficiently large β it surpasses the dimensions of the $L_{x_i' y_i'}$ lenses. So the probability that at least one such pair x_i', y_i' exists is

$$
\begin{aligned}
1 - \left(1 - \Theta\left(n^{-1/3}\right)\right)^t &= 1 - \left(1 - \Theta\left(n^{-1/3}\right)\right)^{n^{1/2}} \\
&= 1 - \exp\left(-\Theta\left(n^{-1/3}\right) n^{1/2}\right) \\
&= 1 - \exp(-n^{\Omega(1)}).
\end{aligned}
$$

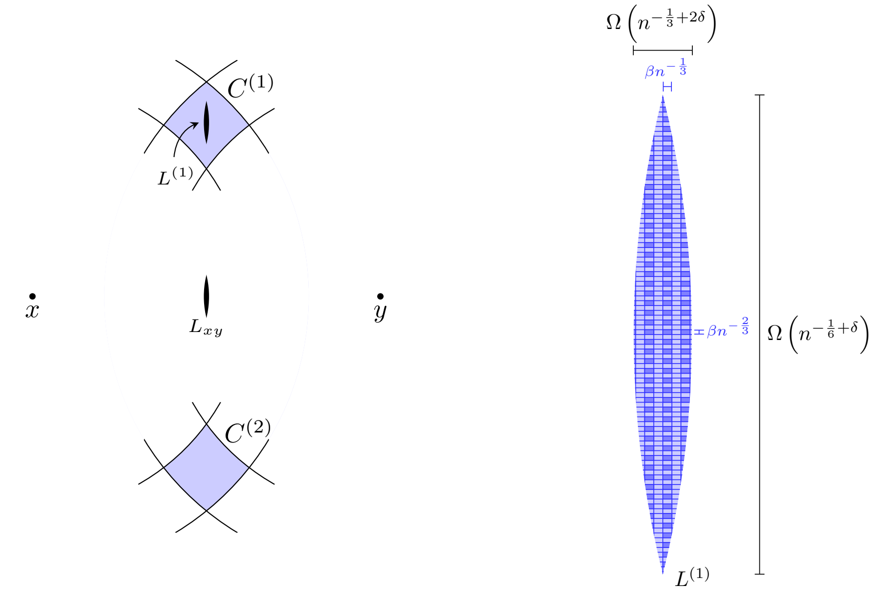
In conclusion, we have shown that there exists a pair x', y' with the desired properties with probability $1 - \exp(-n^{\Omega(1)})$ in the Poisson random geometric graph, which then implies the same conclusion in $G_{\mathbb{T}^2}(n, r)$, and that finishes the proof. ∎

The key lemma leads us to the definition of the following first order formula, which characterizes lenses with diameter of at least $n^{-1/6}$.

Definition 2.17. For a given $\delta > 0$, choose corresponding $\ell \ge 1$ and $\varepsilon > 0$ such that Lemma 2.16 holds and define $\mathrm{B}(x, y)$ as follows. First, let $\mathrm{D}_2(x, y)$ be an $\left(2, \frac{\varepsilon}{2}\right)$-approximator (as in the previous subsection). Then $\mathrm{B}(x, y)$ requires $\neg \mathrm{D}_2(x, y)$ and also the existence of x', y' such that $\mathrm{A}_\ell(x, y)$ holds, $L_{xy}, L_{x'y'}$ are E-perpendicular (with margin ε) and $L_{xy} \cap L_{x'y'}$ contains at least ℓ vertices.

Corollary 2.18. *Let $\delta > 0$ and let $\mathrm{B}(x, y)$ be the corresponding formula, as in the previous definition. Then, a.a.s. for every x, y,*

1. *If $\mathrm{diam}(L_{xy}) \le n^{-1/6-\delta}$ then $\mathrm{B}(x, y)$ does not hold.*
2. *If $\mathrm{diam}(L_{xy}) \ge n^{-1/6+\delta}$ then $\mathrm{B}(x, y)$ holds.*



(a) The vertices x, y in fixed positions, the lens L_{xy} and $C^{(1)}, C^{(2)}$ shaded in light blue. The copy $L^{(1)}$ of L_{xy} in $C^{(1)}$ is also drawn. Note that the illustration greatly exaggerates the size of L_{xy}: in reality, it is asymptotically smaller than the sizes of $C^{(1)}, C^{(2)}$ and should be barely seen.

(b) A close-up on $L^{(1)}$, the copy of L_{xy} inside $C^{(1)}$. As explained in the proof, we divide it into "small" rectangles and consider a quarter of them (shaded in darker blue) to assure spacedness.

Figure 4: Geometric illustrations for the proof of the second part of Lemma 2.16.

Proof. First note that a.a.s. for every x, y, $\neg\mathrm{D}_2(x, y)$ implies $\|x-y\| \geq \left(2-\frac{\varepsilon}{2}\right) r$. The rest follows directly from Lemma 2.16. ■

We are now ready to construct the formula $\mathrm{WS}(x, y)$ and complete the proof of Theorem 2.8. Intuitively, $\mathrm{WS}(x, y)$ claims that $\|x-y\| > n^{-1/6}$ by stating that if x, y are contained in a lens $L_{x'y'}$, then its diameter must be $> n^{-1/6}$.

Proof of Theorem 2.8. Let $\delta > 0$. Take $\mathrm{B}(x, y)$ to be the formula from Definition 2.17 which corresponds to $\frac{\delta}{2}$. That is, a.a.s. for every x, y,

1. $\mathrm{diam}(L_{xy}) \leq n^{-\frac{1}{6}-\frac{\delta}{2}}$ implies $\neg\mathrm{B}(x, y)$.
2. $\mathrm{diam}(L_{xy}) \geq n^{-\frac{1}{6}+\frac{\delta}{2}}$ implies $\mathrm{B}(x, y)$.

Now define $\mathrm{WS}(x,y)$ as the formula claiming that for every x', y', if $x, y \in L_{x'y'}$ then $\mathrm{B}(x',y')$ holds. This is indeed a first order formula. We now prove the two parts of the theorem.

We begin with the second part. Assume $\|x-y\| \geq n^{-1/6+\delta}$. Then for every x', y', if $x, y \in L_{x'y'}$ then

$$\mathrm{diam}(L_{x'y'}) \geq n^{-\frac{1}{6}+\delta} \geq n^{-\frac{1}{6}+\frac{\delta}{2}}$$

which a.a.s implies $\mathrm{B}(x',y')$. Therefore $\mathrm{WS}(x,y)$ holds.

For the first part, assume $\|x-y\| \leq n^{-1/6-\delta}$. We show that a.a.s. there exist x', y' with $x, y \in L_{x'y'}$ and $\mathrm{diam}(L_{x'y'}) \leq n^{-\frac{1}{6}-\frac{\delta}{2}}$. That would end the proof, since a.a.s. $\mathrm{diam}(L_{x'y'}) \leq n^{-\frac{1}{6}-\frac{\delta}{2}}$ implies $\neg\mathrm{B}(x',y')$.

Fix the positions of x, y such that $\|x-y\| \leq n^{-1/6-\delta}$. Let o denote their midpoint. Construct a perpendicular bisector of xy with midpoint o and length $2r - cn^{-1/3-\delta}$ (for some constant $c > 0$). Consider two squares with the two endpoints as centers, of size $\beta n^{-1/3-\delta} \times \beta n^{-1/3-\delta}$ (for some constant $\beta > 0$). The probability that such a square does not contain any vertex is

$$\left(1 - \beta^2 n^{-2/3-2\delta}\right)^{n-2} = \exp\left(-\Theta\left(n^{1/3-2\delta}\right)\right).$$

WLOG we may assume $\delta < \frac{1}{6}$, so this is $\exp\left(-n^{\Omega(1)}\right)$. Therefore, a.a.s. both squares contain vertices. Let x', y' such vertices (one from each square). The right choice of c and a sufficiently small β then guarantee $\mathrm{diam}(L_{x'y'}) \leq n^{-\frac{1}{6}-\frac{\delta}{2}}$, as we once again use the relations $\mathrm{diam} = \Theta\left(\sqrt{\mathrm{width}}\right)$ and $\mathrm{dist} = 2r - \mathrm{width}$.

It is left to explain why $x, y \in L_{x'y'}$; that is, why the distances from x', y' to both x, y are all smaller than r. Take $\|x'-x\|$ for example; the other three cases are symmetric. Let b be the center of the corresponding square. Then $\|x'-b\| \leq \sqrt{2}\beta n^{-1/3-\delta}$. Now, $\|x-b\|$ is the length of an hypotenuse in a right triangle with sides $r - \frac{c}{2}n^{-1/3-\delta}$ and $\leq \frac{1}{2}n^{-1/6-\delta}$. Therefore

$$\begin{aligned}\|x-b\|^2 &= \left[r^2 - \Theta\left(n^{-1/3-\delta}\right)\right] + \Theta\left(n^{-1/3-2\delta}\right)\\ &= r^2 - \Theta\left(n^{-1/3-\delta}\right)\end{aligned}$$

and taking a square root gives $\|x-b\| = r - \Theta\left(n^{-1/3-\delta}\right)$. Again, for a sufficiently small constant β we get

$$\begin{aligned}\|x-x'\| &\leq \|x-b\| + \|b-x'\|.\\ &\leq r - \Theta\left(n^{-1/3-\delta}\right) + \sqrt{2}\beta n^{-1/3-\delta}\\ &< r.\end{aligned}$$

■

3 A Flexible Moments Method

In many probabilistic scenarios, directly evaluating the distribution of a certain random variable X might be highly challenging, while evaluating its moments is significantly simpler. This is usually the case when X counts certain substructures of a random graph, for example. Hence the great importance of methods for extracting information about a distribution via its moments. Two very popular examples in the theory of random graphs are *first and second moment methods* and the *method of moments* (see any introductory text about random graphs, e.g. [10]). The first and second moment methods use $\mathbb{E}(X)$ and $\mathrm{Var}(X)$ to bound $\mathbb{P}(X = 0)$ for a non-negative integer-valued X. The moments method uses *all* the moments to show convergence in distribution.

For our purposes, we will need a moments method that is more informative than the first and second moments methods, yet more general than the method of moments. The current section is therefore dedicated to the development of such a method. We call it a "flexible" moments method because of its generality. The flexible moments method allows us to extract detailed information about a discrete multivariate distribution from its joint moments. It has several important virtues. First, it supports the incorporation of errors in the evaluation of the moments. Second, it only uses the first K moments, where K is some slowly-increasing function of n, instead of all the moments. Finally, it gives concretes bounds on the distribution of specific variables (not just at the limit $n \to \infty$).

Developing this flexible method is surprisingly simple: it is a relatively straightforward generalization of a well-known simple result known as *Waring's Theorem*. Despite its straightforwardness and its multiple advantages, we are unaware of any prior use of this particular method. Due to its generality, we believe in its usefulness to other problems as well, at least in the context of random graphs.

Let us begin by recalling the notion of *factorial moments*.

Definition 3.1. Let X be a non-negative integer-valued random variable. Define its k-th *factorial moment* as

$$\mathbb{E}\left((X)_k\right) = \mathbb{E}\left(X(X-1)(X-2)\dots(X-k+1)\right).$$

That is, the k-th factorial moment of a random variable is the expected value of its k-th falling factorial.

The factorial moments are a useful tool in the study of non-negative integer-valued random variables. Note that they admit a convenient combinatorial meaning: if X counts objects, then $(X)_k$ counts k-tuples of distinct objects. They are also particularly suitable for dealing with Poisson distribution, because of their simple form: if $X \sim \mathrm{Pois}(\lambda)$ then $\mathbb{E}\left((X)_k\right) = \lambda^k$.

Waring's Theorem

We recall Waring's classical theorem. The formulation brought here is actually a bit different from the one most commonly seen in texts; for the more common formulation, we refer to Exercise 1.8.13 from [5].

Theorem 3.2 (Waring's Theorem)**.** *Let X be a non-negative integer-valued random variable. Fix a value $z \geq 0$, and for every $s \geq 0$, let*

$$\Sigma_{z,s} = \sum_{i=0}^{s} (-1)^i \frac{\mathbb{E}\left((X)_{z+i}\right)}{z!i!}.$$

Then $\mathbb{P}\left(X = z\right) \leq \Sigma_{z,s}$ for even values of s and $\mathbb{P}\left(X = z\right) \geq \Sigma_{z,s}$ for odd values of s.

In other words, the theorem presents the bounds

$$\Sigma_{z,s-1} \leq \mathbb{P}\left(X = z\right) \leq \Sigma_{z,s}$$

on $\mathbb{P}\left(X = z\right)$, which apply for every even s. To get a feel for this result, one can try taking $X \sim \text{Pois}(\lambda)$ and observe the resulting bounds (keep in mind that in that case $\mathbb{E}\left((X)_k\right) = \lambda^k$).

Remark 3.3. Notice that the difference between the upper bound and the lower bound is $\frac{1}{z!s!}\mathbb{E}\left((X)_{z+s}\right)$. Therefore better upper bounds on the factorial moments lead to tighter bounds on $\mathbb{P}\left(X = z\right)$. In addition, notice how we can easily control how many moments to use: if all the moments up to K (with $K \geq z$) are known, simply take s to be the maximal even number with $z + s \leq K$. Finally, notice how this bound can naturally incorporate errors in the estimation of the moments. For example, if we know that

$$\left|\mathbb{E}\left((X)_k\right) - \mu_k\right| < \varepsilon_k$$

for every $0 \leq k \leq K$, we obtain

$$\left|\mathbb{P}\left(X = z\right) - \sum_{i=0}^{s} (-1)^i \frac{\mu_{z+i}}{z!i!}\right| \leq \frac{\mu_{z+s}}{z!s!} + \sum_{i=0}^{s} \frac{\varepsilon_{z+i}}{z!i!}.$$

Let us give a short explanation for Waring's Theorem, which will naturally lead us to its multivariate generalization. The key observation is that $\Sigma_{z,s} = \mathbb{E}\left(\pi_{z,s}(X)\right)$ where $\pi_{z,s}(x)$ is the polynomial

$$\pi_{z,s}(x) = \sum_{i=0}^{s} (-1)^j \frac{(x)_{z+i}}{z!i!}.$$

Therefore we can also write

$$\Sigma_{z,s} = \sum_{x \geq 0} \mathbb{P}\left(X = x\right) \pi_{z,s}(x).$$

Now it is sufficient to verify the following formula for the values of $\pi_{z,s}(x)$ at non-negative integer values x:

$$\pi_{z,s}(x) = \begin{cases} 1 & x = z \\ 0 & x \neq z,\ x \leq z+s \\ (-1)^s \binom{x-z-1}{s}\binom{x}{z} & x > z+s \end{cases} .$$

In particular, the values for $x \neq z$ are non-negative when s is even and non-positive when s is odd.

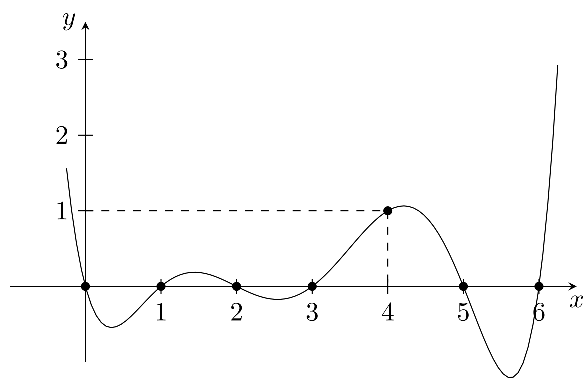


Figure 5: An illustrative example of the graph of $\pi_{z,s}(x)$ for $z = 4$, $s = 2$.

Generalizing to multiple variables

Now let $X_1, X_2, \ldots, X_t$ be a t-tuple of non-negative integer-valued random variables (over the same probability space). Also fix a t-tuple of integers $z_1, z_2, \ldots, z_t$ and $K \geq \max_i z_i$.

Definition 3.4. The $(k_1, k_2, \ldots, k_t)$-th *joint factorial moment* of $X_1, X_2, \ldots, X_t$ is

$$\mathbb{E}\left((X_1)_{k_1}(X_2)_{k_2}\ldots(X_t)_{k_t}\right).$$

Again, the joint factorial moments carry an important combinatorial meaning. If X_i counts objects of type i, then

$$(X_1)_{k_1}(X_2)_{k_2}\ldots(X_t)_{k_t}$$

counts $(k_1, k_2, \ldots, k_t)$-tuples of objects: k_1 distinct objects of type 1, k_2 distinct objects of type 2 and so on.

For the sake of notational convenience, from now on we denote $\mathbf{X} = (X_1, \ldots, X_t)$, $\mathbf{k} = (k_1, \ldots, k_t)$, $\mathbf{z} = (z_1, \ldots, z_t)$ and so on. We also write

$$\mathbb{E}\left((\mathbf{X})_{\mathbf{k}}\right) = \mathbb{E}\left((X_1)_{k_1} \ldots (X_t)_{k_t}\right),$$
$$\mathbb{P}\left(\mathbf{X} = \mathbf{z}\right) = \mathbb{P}\left(X_1 = z_1, \ldots, X_t = z_t\right).$$

$0 \leq \mathbf{k} \leq K$ simply means $0 \leq k_i \leq K$ for every $1 \leq i \leq t$.

Similar to before, our goal is to estimate $\mathbb{P}\left(\mathbf{X} = \mathbf{z}\right)$ by using the joint factorial moments $\mathbb{E}\left((\mathbf{X})_{\mathbf{k}}\right)$ with $0 \leq \mathbf{k} \leq K$. To do that we want to find suitable multivariate polynomials $\pi(\mathbf{x}) = \pi(x_1, \ldots, x_t)$ which will give us upper and lower bounds. This leads to the following definition.

Definition 3.5. Following the notation above, for every $1 \leq i \leq t$ let s_i be the maximal even integer with $z_i + s_i \leq K$. Define the following two polynomials:

$$\begin{aligned} \pi^+(\mathbf{x}) &= \prod_{i=1}^{t} \pi_{z_i,s_i}(x_i), \\ \pi^-(\mathbf{x}) &= \pi^+(\mathbf{x}) - \prod_{i=1}^{t} \left(1 + \frac{(x_i)_{z_i+s_i}}{z_i!s_i!}\right) + 1. \end{aligned}$$

The motivation behind this definition is quite simple. For the upper bound, we can simply take the product of $\pi_{z_i,s_i}(x_i)$ since a product of non-negatives is non-negative. For the lower bound, we subtract the necessary terms to turn all the positive regions of $\pi^+(\mathbf{x})$ to negative.

Note that both polynomials are expressed solely through falling factorials of $x_1, x_2, \ldots, x_t$, and so $\mathbb{E}\left(\pi^-(\mathbf{X})\right)$ and $\mathbb{E}\left(\pi^+(\mathbf{X})\right)$ are indeed linear combinations of $\mathbb{E}\left((\mathbf{X})_{\mathbf{k}}\right)$ with $0 \leq \mathbf{k} \leq K$. Thus, in the following theorem, $\mathbb{P}\left(\mathbf{X} = \mathbf{z}\right)$ is indeed bounded through the joint factorial moments alone.

Theorem 3.6.
$$\mathbb{E}\left(\pi^-(\mathbf{X})\right) \leq \mathbb{P}\left(\mathbf{X} = \mathbf{z}\right) \leq \mathbb{E}\left(\pi^+(\mathbf{X})\right).$$

Proof. It is sufficient to prove that $\pi^-(\mathbf{z}) = \pi^+(\mathbf{z}) = 1$ and that $\pi^+(\mathbf{x})$ is non-negative for $\mathbf{x} \neq \mathbf{z}$ while $\pi^-(\mathbf{x})$ is non-positive for $\mathbf{x} \neq \mathbf{z}$.

Start with $\pi^+(\mathbf{x})$. First,

$$\pi^+(\mathbf{z}) = \prod_{i=1}^{t} \pi_{z_i,s_i}(z_i) = \prod_{i=1}^{t} 1 = 1.$$

Second, since $\pi^+(\mathbf{x})$ is a product of non-negative polynomials $\pi_{z_i,s_i}(x_i)$, it is always non-negative.

Now consider π^-. First,

$$\begin{aligned}\pi^-(\mathbf{z}) &= \pi^+(\mathbf{z}) - \prod_{i=1}^{t}\left(1 + \frac{(z_i)_{z_i+s_i}}{z_i!s_i!}\right) + 1 \\ &= 1 - \prod_{i=1}^{t}(1+0) + 1 \\ &= 1 - 1 + 1 = 1.\end{aligned}$$

Second, consider any $\mathbf{x} \neq \mathbf{z}$. Showing that $\pi^-(\mathbf{x}) \leq 0$ is equivalent to showing that

$$\prod_{i=1}^{t}\left(1 + \frac{(x_i)_{z_i+s_i}}{z_i!s_i!}\right) \geq 1 + \prod_{i=1}^{t}\pi_{z_i,s_i}(x_i).$$

If any x_i is neither z_i nor $> z_i + s_i$, we get $\pi_{z_i,s_i}(x_i) = 0$ and the inequality is trivially true. So assume that every x_i is either z_i or $> z_i + s_i$. $\mathbf{x} \neq \mathbf{z}$ and so the set $I \subseteq [t] = \{1, 2, , \ldots, t\}$ of indices for which $x_i > z_i + s_i$ is non-empty. Note that:

(a) For every $i \in I$, since $\pi_{z_i,s_i-1}(x_i)$ is non-positive,

$$\frac{(x_i)_{z_i+s_i}}{z_i!s_i!} \geq \pi_{z_i,s_i}(x_i).$$

(b) For every $i \notin I$, $x_i = z_i$ and so $\pi_{z_i,s_i}(x_i) = 1$.

Overall, we get

$$\begin{aligned}\prod_{i=1}^{t}\left(1 + \frac{(x_i)_{z_i+s_i}}{z_i!s_i!}\right) &= \prod_{i \in I}\left(1 + \frac{(x_i)_{z_i+s_i}}{z_i!s_i!}\right) \\ &\overset{I \neq \emptyset}{\geq} 1 + \prod_{i \in I}\frac{(x_i)_{z_i+s_i}}{z_i!s_i!} \\ &\overset{\text{(a)}}{\geq} 1 + \prod_{i \in I}\pi_{z_i,s_i}(x_i) \\ &\overset{\text{(b)}}{=} 1 + \prod_{i=1}^{t}\pi_{z_i,s_i}(x_i)\end{aligned}$$

and that concludes the proof. ■

Theorem 3.6 is a direct generalization of Waring's Theorem, and therefore the contents of Remark 3.3 naturally generalize to it as well.

To finish the section, let us intuitively describe how a typical use of Theorem 3.6 might look like. Let $\mathbf{X}$ be a tuple of random variables with a known joint distribution. Let $\mathbf{Y}$ be another tuple of random variables and assume the first joint factorial moments are similar: $\mathbb{E}\left((\mathbf{Y})_{\mathbf{k}}\right)$ is "close" to $\mathbb{E}\left((\mathbf{X})_{\mathbf{k}}\right)$ for every $0 \le \mathbf{k} \le K$.

Since π^+, π^- are linear combinations of the joint factorial moments, conclude that $\mathbb{E}\left(\pi^+(\mathbf{X})\right)$ and $\mathbb{E}\left(\pi^+(\mathbf{Y})\right)$ are also "close", as well as $\mathbb{E}\left(\pi^-(\mathbf{X})\right)$ and $\mathbb{E}\left(\pi^-(\mathbf{Y})\right)$. Now, use Theorem 3.6 to bound

$$\mathbb{E}\left(\pi^-(\mathbf{X})\right) \le \mathbb{P}\left(\mathbf{X} = \mathbf{z}\right) \le \mathbb{E}\left(\pi^+(\mathbf{X})\right),$$
$$\mathbb{E}\left(\pi^-(\mathbf{Y})\right) \le \mathbb{P}\left(\mathbf{Y} = \mathbf{z}\right) \le \mathbb{E}\left(\pi^+(\mathbf{Y})\right).$$

Overall deduce that the difference $\left|\mathbb{P}\left(\mathbf{Y} = \mathbf{z}\right) - \mathbb{P}\left(\mathbf{X} = \mathbf{z}\right)\right|$ must be "close" to the difference

$$\left|\mathbb{E}\left(\pi^+(\mathbf{X})\right) - \mathbb{E}\left(\pi^-(\mathbf{X})\right)\right|.$$

Finally, show that the difference $\left|\mathbb{E}\left(\pi^+(\mathbf{X})\right) - \mathbb{E}\left(\pi^-(\mathbf{X})\right)\right|$ itself is small. That would be true whenever the denominators of $\pi^+ - \pi^-$, which typically contain factorials of K, are much larger than the numerators, which typically contain moments of the K-th order.

The proof of Theorem 5.12 provides a detailed example of such a use, in which $\mathbf{X}$ is distributed like independent Poisson variables.

4 The Extension Formula S

In this section we shall give our most important definition — the definition of the first-order extension formula S for the random geometric graph. This formula will be the building block of the graph structures we intend to express within $G_{\mathbb{T}^2}(n, r)$. S will be defined over a triplet of root vertices denoted x_1, x_2, x_3 and introduce four new vertices denoted s_1, s_2, s_3, z, so in full it should be written as $S = S(x_1, x_2, x_3; s_1, s_2, s_3, z)$. A quadruplet (s_1, s_2, s_3, z) which satisfies this extension will be called a *witness* for the extension (over x_1, x_2, x_3).

The motivation behind the definition of S is *keeping the expected number of witnesses asymptotically constant.* When counting witnesses for an extension formula in a random graph, this is often enough to assure that the asymptotic distribution is Poisson. In Section 5 we will show that in some sense this is also the case for S.

Heuristically, a good way to promise an asymptotically constant expectation is to use a relation between vertices whose probability p is a (negative) rational power of n. Then, the overall expected number of witnesses should behave like $\Theta(n^v p^r)$ where v is the number of new vertices and r is the number of required relations, so we just need to correctly balance v, r. In Subsection 4.1 we describe a first-order relation $\mathrm{UN}(x, y)$ whose probability (given x, y) is $\approx \mathrm{Const} \cdot n^{-2/3}$. In Subsection 4.2 we define S by introducing $v = 4$ new vertices and requiring $r = 6$ UN-relations.

4.1 The UN-Relation

Recall the first-order relation $\mathrm{A}_\ell(x,y)$ from Lemma 2.13. For a given constant ℓ, $\mathrm{A}_\ell(x,y)$ claims that x,y have exactly ℓ common neighbors. For concreteness let us fix $\ell=1$, so it claims that x,y have a unique neighbor. Let us rename it accordingly and denote it $\mathrm{UN}(x,y)$. In Section 2 we intuitively explained how this requirement forces $\mathrm{area}(L_{xy})$ to be about n^{-1}, and therefore $\mathrm{diam}(L_{xy})$ should be around $n^{-1/3}$ and $\mathrm{width}(L_{xy})$ should be around $n^{-2/3}$. Recall that

$$\|x-y\| = 2r - \mathrm{width}(L_{xy})$$

so we intuitively expect $\mathbb{P}\left(\mathrm{UN}(x,y)\right)$ to also be around $n^{-2/3}$. As it turns out, this estimation can be made very precise.

Theorem 4.1. *Let x,y be two vertices in $G_{\mathbb{T}^2}(n,r)$. Then*

$$\mathbb{P}\left(\mathrm{UN}(x,y)\right) \approx Cn^{-2/3}$$

as $n\to\infty$, for some positive constant C (which can be computed explicitly).

Before the proof, let us introduce some useful notation.

For two vertices x,y define the random variable $H_{xy} = r - \frac{\|x-y\|}{2}$. Note that it equals half the width of L_{xy}, so it corresponds to the parameter h we use to describe the geometry of lenses (see Lemma 2.11). Also note that $h\le r$ by definition, and that $h\ge 0$ if and only if L_{xy} is non-empty. Let $f_{xy}(h)$ denote the probability density function of H_{xy}.

Moreover, define the function

$$A(h) = \begin{cases} 2r^2\arccos\left(1-\frac{h}{r}\right) - 2(r-h)\sqrt{2rh-h^2} & h\ge 0 \\ 0 & h<0 \end{cases}. \tag{3}$$

Recall that it describes the area of a lens with parameter h. In Corollary 2.12 we have seen than

$$A(h) \approx C_A r^{1/2} h^{3/2}$$

as $h\to 0$, where $C_A = \frac{8\sqrt{2}}{3}$. Let us denote $B(h) = C_A r^{1/2} h^{3/2}$. Using more careful asymptotics it can actually be shown that

$$A(h) - B(h) = O\left(h^{5/2}\right).$$

Proof of Theorem 4.1. Given that $H_{xy}=h$, the probability that exactly one of the remaining $n-2$ vertices lands inside L_{xy} is $Q(A(h))$, where

$$Q(t) = (n-2)t(1-t)^{n-3}.$$

Integrating over the possible h gives us

$$\mathbb{P}\left(\mathrm{UN}(x,y)\right) = \int_0^r Q(A(h)) f_{xy}(h)\mathrm{d}h.$$

Now, let us define

$$h_n = \left(\frac{\ln \ln n}{C_A \sqrt{r}} \cdot \frac{\ln n}{n} \right)^{2/3} \tag{4}$$

where $C_A = \frac{8\sqrt{2}}{3}$ (this constant comes from Corollary 2.12). We separately estimate $\int_0^r = \int_0^{h_n} + \int_{h_n}^r$.

Let us first show that the second term is negligible:[4]

$$\int_{h_n}^r Q(A(h)) f_{xy}(h) \mathrm{d}h = n^{-\Theta(\ln \ln n)}. \tag{5}$$

We list the following facts:

1. We have

$$A(h_n) \approx B(h_n) = C_A \sqrt{r} \cdot \frac{\ln \ln n}{C_A \sqrt{r}} \cdot \frac{\ln n}{n} = \frac{\ln \ln n \cdot \ln n}{n}.$$

2. $Q(t)$ is monotonically decreasing for $t \geq \frac{1}{n-2}$. This follows from simple differentiation.
3. $A(h_n) > \frac{1}{n-2}$ for sufficiently large n (this follows directly from 1).
4. $A(h)$ is monotonically increasing in h.

Altogether, we deduce that for $h \geq h_n$ we have $Q(A(h)) \leq Q(A(h_n))$, so

$$\int_{h_n}^r Q(A(h)) f_{xy}(h) \mathrm{d}h \leq Q(A(h_n)) \int_{h_n}^r f_{xy}(h) \mathrm{d}h \leq Q(A(h_n)).$$

Now, standard asymptotic estimations give us

$$\begin{aligned} Q(A(h_n)) &\approx \ln \ln n \cdot \ln n \cdot \left(1 - \frac{\ln \ln n \cdot \ln n}{n}\right)^n \\ &\approx \ln \ln n \cdot \ln n \cdot e^{-\ln \ln n \cdot \ln n} \\ &= \ln \ln n \cdot \ln n \cdot n^{-\ln \ln n} = n^{-\Theta(\ln \ln n)}. \end{aligned}$$

Now for the main term:

$$\int_0^{h_n} Q(A(h)) f_{xy}(h) \mathrm{d}h.$$

By computing the density function we get $f_{xy}(h) \underset{h \to 0}{\to} 8\pi r$. Since $h_n \to 0$ we can replace $f_{xy}(h)$ with the constant $8\pi r$ and maintain asymptotic equivalence. Also, since $\lim_{n\to\infty} \frac{n-2}{n} = 1$ and $\lim_{h\to 0^+} (1 - A(h)) = 1$, we can replace

$$Q(t) = (n-2)t(1-t)^{n-3}$$

[4] The bound $n^{-\Theta(\ln \ln n)}$ is much tighter than what we currently need, but we shall use it in Subsection A.2.

with the simpler $nt(1-t)^n$. We get a "cleaner" expression:

$$\int_0^{h_n} Q(A(h)) f_{xy}(h) \mathrm{d}h \approx 8\pi r \int_0^{h_n} nA(h)(1-A(h))^n \mathrm{d}h.$$

To further simplify the integral, we use standard asymptotic estimations to replace $(1-A(h))^n$ with $e^{-nA(h)}$ and then $A(h)$ with $B(h)$. We obtain

$$\int_0^{h_n} nA(h)(1-A(h))^n \mathrm{d}h \approx \int_0^{h_n} nB(h)e^{-nB(h)} \mathrm{d}h.$$

Finally, let us directly evaluate the obtained integral. Explicitly written, it is

$$\int_0^{h_n} nB(h)e^{-nB(h)} \mathrm{d}h = \int_0^{h_n} nC_A r^{1/2} h^{3/2} \exp\left(-nC_A r^{1/2} h^{3/2}\right) \mathrm{d}h.$$

Consider the change of variables $t = nC_A r^{1/2} h^{3/2}$.

$$\mathrm{d}t = \frac{3}{2} nC_A r^{1/2} h^{1/2} \mathrm{d}h$$
$$nC_A r^{1/2} h^{3/2} \mathrm{d}h = \frac{C'}{n^{2/3}} t^{2/3} \mathrm{d}t$$

for some constant C'. Hence

$$\begin{aligned}\int_0^{h_n} nB(h)e^{-nB(h)} \mathrm{d}h &= \frac{C'}{n^{2/3}} \int_0^{nC_A r^{1/2} h_n^{3/2}} t^{2/3} e^{-t} \mathrm{d}t \\ &= \frac{C'}{n^{2/3}} \int_0^{\ln \ln n \cdot \ln n} t^{2/3} e^{-t} \mathrm{d}t.\end{aligned}$$

But

$$\int_0^{\ln \ln n \cdot \ln n} t^{2/3} e^{-t} dt \underset{n\to\infty}{\to} \int_0^{\infty} t^{2/3} e^{-t} dt = \Gamma\left(\frac{5}{3}\right)$$

and so

$$\int_0^{h_n} nB(h)e^{-nB(h)} \mathrm{d}h \approx C'\Gamma\left(\frac{5}{3}\right) n^{-2/3}.$$

In conclusion, we have shown that there exists a constant C such that

$$\mathbb{P}\left(\mathrm{UN}(x,y)\right) \approx \int_0^{h_n} Q(A(h)) f_{xy}(h) \mathrm{d}h \approx Cn^{-2/3}.$$

That finishes the proof. ■

4.2 Defining S

Definition 4.2. Define the following first-order formula:

$$\begin{gathered} S(x_1, x_2, x_3; s_1, s_2, s_3, z) = \\ \mathrm{DIS} \wedge \mathrm{UN}(x_1, s_1) \wedge \mathrm{UN}(x_2, s_2) \wedge \mathrm{UN}(x_3, s_3) \wedge \mathrm{UN}(s_1, z) \wedge \mathrm{UN}(s_2, z) \wedge \mathrm{UN}(s_3, z) \\ \wedge \mathrm{CD}(x_1, z) \wedge \mathrm{CD}(x_2, z) \wedge \mathrm{CD}(x_3, z), \end{gathered}$$

where:

- DIS demands that all seven vertices $x_1, x_2, x_3, s_1, s_2, s_3, z$ are distinct.
- $\mathrm{UN}(x, y)$ is the "Unique Neighbor" relation defined in subsection 4.1.
- $\mathrm{CD}(x, z)$ is the formula $\mathrm{D}_{\sqrt{3}}(x, z) \wedge \neg\mathrm{D}_1(x, z)$, where $\mathrm{D}_1, \mathrm{D}_{\sqrt{3}}$ are the formulas from Lemma 2.4. CD stands for "Comfortable Distances" (see the following remark).

Figure 6 illustrates the S-extension.

Remark 4.3. While the idea behind having four vertices and six UN relations has been explained, the choice to add CD to the definition is not immediately clear. $\mathrm{CD}(x, z)$ forces the distance $\|x - z\|$ to be approximately between r and $\sqrt{3}r$. The idea is that allowing $\|x - z\|$ to be extremely close to either 0 or $4r$ causes a "blow up" that dominates the probability of the extension; the CD conditions prevent that and keep the behavior of S "tamed". This kind of argument is demonstrated in the proof of the Negligibility Theorem (see Subsection A.2 in the Appendix), where we show that certain undesirable overlaps between S-extensions have negligible effect.

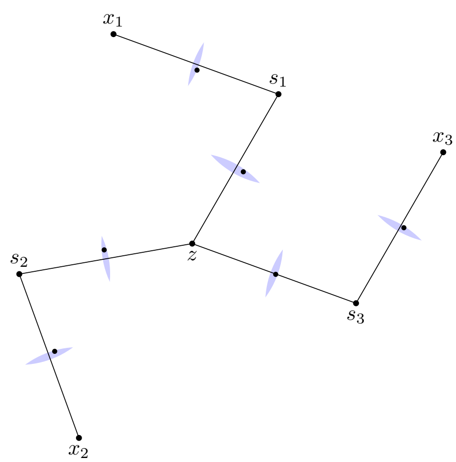


Figure 6: A geometric illustration of the S-structure. The black lines represent UN-relations. For every two UN-related vertices, their lens contains exactly one vertex, as shown.

5 Joint Distribution of Witnesses

In this section we investigate the joint distribution of witnesses for the extension formula S over different triplets of vertices. As it turns out, under a certain geometric conditioning, the counts of witnesses asymptotically behave like independent Poisson variables.

From now until the end of the section, let us fix t distinct triplets of vertices, which we denote $x_1^{(i)}, x_2^{(i)}, x_3^{(i)}$ for $1 \leq i \leq t$. Note that by "distinct triplets" we mean distinct as sets. That is, *triplets may share vertices*, but two triplets may not share all three vertices. Let U denote the set of all vertices from those triplets and let $v = |U|$. Of course, v and t may depend on the underlying n.

The approach taken in this paper for encoding arbitrary graph structures works for very slowly-growing numbers of vertices.[5] We therefore assume $v \leq \ln\ln\ln\ln n$, which also means that $t \leq \binom{v}{3} \leq (\ln\ln\ln\ln n)^3$.

For brevity, we denote $\mathbf{x}^{(i)} = (x_1^{(i)}, x_2^{(i)}, x_3^{(i)})$. We also denote quadruplets of vertices as $\mathbf{q} = (s_1, s_2, s_3, z)$. So $S(\mathbf{x}^{(i)}; \mathbf{q})$ means that the quadruplet $\mathbf{q}$ is a witness for the S-extension over $\mathbf{x}^{(i)}$. Let $Z^{(i)}$ be the random variable which counts those witnesses:

$$Z^{(i)} = \sum_{\mathbf{q}} \mathbb{1}_{\{S(\mathbf{x}^{(i)};\mathbf{q})\}}.$$

We are interested in estimating the joint distribution of $Z^{(1)}, Z^{(2)}, \ldots, Z^{(t)}$. To do that, we estimate their joint factorial moments and then use the flexible moments method from Section 3.

To obtain the promised Poisson-like behavior, it turns out that we must fix the geometric positions of the vertices of U. Intuitively, this act eliminates "malicious" dependency which hides within the geometric configurations of the triplets. This leads us to the following definition.

Definition 5.1. Write $U = \{u_1, u_2, \ldots, u_v\}$. A *geometric configuration* (or simply a *configuration*) of U is an event of the form

$$\{u_1 = a_1, u_2 = a_2, \ldots, u_v = a_v\}$$

where $a_1, a_2, \ldots, a_v \in \mathbb{T}^2$ are fixed points in the torus. For short, we write $\mathbf{u} = (u_1, u_2, \ldots, u_v)$ and $\mathbf{a} = (a_1, a_2, \ldots, a_v)$ so the configuration is written as $\mathbf{u} = \mathbf{a}$. Given such a configuration, for every $1 \leq i \leq t$ the fixed positions of the triplet $x_1^{(i)}, x_2^{(i)}, x_3^{(i)}$ are denoted $a_1^{(i)}, a_2^{(i)}, a_3^{(i)}$.

Throughout this section we will always perform our evaluations given a certain geometric configuration. It turns out to be the most comfortable and natural way to understand the variables $Z^{(1)}, Z^{(2)}, \ldots, Z^{(t)}$. When necessary, we will always be able to *integrate* over all the configurations and obtain general, non-conditioned evaluations (this will happen in Section 6). It is important to keep in mind that in order to do that, our evaluations must remain uniform with respect to the configuration $\mathbf{u} = \mathbf{a}$.

[5]This is also the case in the binomial model; see [19].

Definition 5.2. A configuration $\mathbf{u} = \mathbf{a}$ is called *feasible* if for every $1 \leq i \leq t$, the set $R^{(i)} \subseteq \mathbb{T}^2$ of points which are at "comfortable distances" from $a_1^{(i)}, a_2^{(i)}, a_3^{(i)}$ is non-empty. Formally,

$$R^{(i)} = \left\{ a \in \mathbb{T}^2 \;\middle|\; \left\|a - a_1^{(i)}\right\|, \left\|a - a_2^{(i)}\right\|, \left\|a - a_3^{(i)}\right\| \in \left[r, \sqrt{3}r\right] \right\}.$$

Furthermore, we say that $\mathbf{u} = \mathbf{a}$ is *q-feasible* (for a constant $q > 0$) if $\text{area}(R^{(i)}) \geq q$ for every $1 \leq i \leq t$.

Intuitively, feasibility of $\mathbf{u} = \mathbf{a}$ assures that all the extensions are geometrically possible. q-feasibility prevents "edge cases" and is essential to allow uniform bounds over different configurations.

Definition 5.3. A configuration $\mathbf{u} = \mathbf{a}$ is called *well-spaced* if $\|a_j - a_{j'}\| \geq n^{-1/6}$ for every two different $1 \leq j, j' \leq v$. Furthermore, we say that $\mathbf{u} = \mathbf{a}$ is *strongly well-spaced* if, in addition, for every $1 \leq i \leq t$ the set

$$\left\{ a \in \mathbb{T}^2 \;\middle|\; \left\|a - a_1^{(i)}\right\|, \left\|a - a_2^{(i)}\right\|, \left\|a - a_3^{(i)}\right\| \in \left[2r - 2n^{-1/6}, 2r + 2n^{-1/6}\right] \right\}$$

is empty.

Intuitively, these conditions help to prevent dangerous overlap between different extensions, in a similar fashion to the CD-relations. The exact details can be found in Subsection A.2 (specifically, see the proof of Proposition A.9 and the preceding remark).

5.1 The Joint Factorial Moments

From now and until the end of the section we fix a geometric configuration $\mathbf{u} = \mathbf{a}$, which is strongly well-spaced and q_0-feasible for a small constant q_0 (e.g. $q_0 = \frac{r}{100}$). Our probabilities and expectations will be conditioned by $\mathbf{u} = \mathbf{a}$. To emphasize that, we will denote

$$\begin{aligned} \mathbb{P}_{\mathbf{a}}(A) &= \mathbb{P}(A \mid \mathbf{u} = \mathbf{a}), \\ \mathbb{E}_{\mathbf{a}}(X) &= \mathbb{E}(X \mid \mathbf{u} = \mathbf{a}) \end{aligned}$$

for events A and variables X. In addition, we reuse abbreviated notation from Section 3: we write $\mathbf{Z} = (Z^{(1)}, Z^{(2)}, \ldots, Z^{(t)})$ and consider the factorial moments $\mathbb{E}_{\mathbf{a}}\left((\mathbf{Z})_{\mathbf{k}}\right)$ for tuples $\mathbf{k} = (k_1, k_2, \ldots, k_t)$.

Let us set $K = \ln\ln\ln n$. This value is chosen to grow faster than v, t but still sufficiently slowly. The purpose of this subsection is to estimate the joint factorial moments $\mathbb{E}_{\mathbf{a}}\left((\mathbf{Z})_{\mathbf{k}}\right)$ for every $0 \leq \mathbf{k} \leq K$.

Negligible Bad Situations

Our first step is to list some "bad" situations and to claim that their contribution to the joint moments is negligible. The ability to ignore these situations will make the evaluation of the joint moments much cleaner. We call this result the *Negligibility Theorem* (see Theorem 5.7).

Recall the combinatorial meaning of the joint factorial moments. The variable

$$(\mathbf{Z})_{\mathbf{k}} = \left(Z^{(1)}\right)_{k_1} \cdots \left(Z^{(t)}\right)_{k_t}$$

counts tuples of $k_1 + \cdots + k_t$ quadruplets

$$\mathbf{Q} = \left(\mathbf{q}_1^{(1)}, \ldots, \mathbf{q}_{k_1}^{(1)}, \ldots, \mathbf{q}_1^{(t)}, \ldots, \mathbf{q}_{k_t}^{(t)}\right)$$

such that $\mathbf{q}_1^{(i)}, \ldots, \mathbf{q}_{k_i}^{(i)}$ are all distinct for every $1 \leq i \leq t$ and such that each $\mathbf{q}_j^{(i)}$ satisfies $S\left(\mathbf{x}^{(i)}; \mathbf{q}_j^{(i)}\right)$. We refer to such tuples $\mathbf{Q}$ as $\mathbf{k}$*-tuples of witnesses* for the S-extension (over the triplets $\mathbf{x}^{(i)}$).

The idea of the Negligiblity Theorem is that instead of counting all those $\mathbf{k}$-tuples $\mathbf{Q}$, we may only count "well-behaved" $\mathbf{k}$-tuples. Let us explain what do we mean by well-behaved.

The first thing we would like to do is actually to replace S with a different, non-first-order extension, which we denote S^*. This new extension will be simpler to analyze, yet it turns out that replacing S with S^* causes only a negligible change.

Definition 5.4. Define the following first-order formula:

$$S^*(x_1, x_2, x_3; s_1, s_2, s_3, z) =$$
$$\mathrm{DIS} \wedge \mathrm{UN}^*(x_1, s_1) \wedge \mathrm{UN}^*(x_2, s_2) \wedge \mathrm{UN}^*(x_3, s_3) \wedge \mathrm{UN}^*(s_1, z) \wedge \mathrm{UN}^*(s_2, z) \wedge \mathrm{UN}^*(s_3, z)$$
$$\wedge \mathrm{CD}^*(x_1, z) \wedge \mathrm{CD}^*(x_2, z) \wedge \mathrm{CD}^*(x_3, z)$$

Where:

- DIS demands that all seven vertices $x_1, x_2, x_3, s_1, s_2, s_3, z$ are distinct.
- $\mathrm{UN}^*(x, y)$ demands that $\mathrm{UN}(x, y)$ holds and also that $2r - 2h_n \leq \|x - y\| \leq 2r$ (see Equation 4 for the definition of h_n).
- $\mathrm{CD}^*(x, z)$ demands that $r \leq \|x - z\| \leq \sqrt{3}r$.

Remark 5.5. Note that $\mathrm{UN}^*, \mathrm{CD}^*$are not first order formulas, since they directly refer to the distances. The idea is that they are approximated by the first-order UN, CD. The distance conditions imposed by UN^* and CD^* turn out to be very comfortable in certain situations. Also note that $\mathrm{UN}^*(x, y)$ is equivalent to

$$\mathrm{UN}(x, y) \wedge H_{xy} \in [0, h_n]$$

(H_{xy} was defined in Subsection 4.1).

The second important aspect of good behavior is avoiding overlaps between triplets and their extensions. We summarize it in the following definition.

Definition 5.6. Set $0 \le \mathbf{k} \le K$ and consider a $\mathbf{k}$-tuple

$$\mathbf{Q} = \left(\mathbf{q}_1^{(1)}, \ldots, \mathbf{q}_{k_1}^{(1)}, \ldots, \mathbf{q}_1^{(t)}, \ldots, \mathbf{q}_{k_t}^{(t)}\right)$$

of witnesses for S^* over the triplets $\mathbf{x}^{(i)}$. We say that $\mathbf{Q}$ is *well-behaved* if *none* of the "bad" situations listed below occur. In the following list, i, j are always indices between 1 and t (not necessarily distinct!) and $\mathbf{q}^{(i)}$ always denotes one of $\mathbf{q}_1^{(i)}, \ldots, \mathbf{q}_{k_i}^{(i)}$.

1a. There exist a witness $\mathbf{q}^{(i)}$ and a triplet $\mathbf{x}^{(j)}$ that share a vertex.

1b. There exist different quadruplets $\mathbf{q}^{(i)}, \mathbf{q}^{(j)}$ that share a vertex.

2. Not all the vertices of $\mathbf{Q}$ are well-spaced.

3a. There exist $\mathbf{q}^{(i)}$ and $\mathbf{x}^{(j)}$ such that a vertex from $\mathbf{x}^{(j)}$ is also the witness for a UN-relation of $S(\mathbf{x}^{(i)}; \mathbf{q}^{(i)})$.

3b. There exist $\mathbf{q}^{(i)}, \mathbf{q}^{(j)}$ such that a vertex from $\mathbf{q}^{(j)}$ is also the witness for a UN-relation of $S(\mathbf{x}^{(i)}; \mathbf{q}^{(i)})$.

4. There exist $\mathbf{q}^{(i)}, \mathbf{q}^{(j)}$ such that a UN-relation of $\mathbf{q}^{(i)}$ and a UN-relation of $\mathbf{q}^{(j)}$ both have the same witness.

We can sum up the meaning of situations **1a-b**, **3a-b** and **4** in one sentence. Those are all the situations in which a vertex "plays two roles" at the same time, where the roles are being an x-vertex, being a $\mathbf{q}$-vertex and being the witness for a UN relation. Situation **2** mainly helps to prove the negligibility of the other situations.

Theorem 5.7 (The Negligibility Theorem)**.** *Set $0 \le \mathbf{k} \le K$. Let $\tilde{\mathbf{Z}}_{\mathbf{k}}$ count well-behaved $\mathbf{k}$-tuples of witnesses for S^*. Then*

$$\mathbb{E}_{\mathbf{a}}\left((\mathbf{Z})_{\mathbf{k}}\right) = \mathbb{E}_{\mathbf{a}}\left(\tilde{\mathbf{Z}}_{\mathbf{k}}\right) + n^{-\Omega(1)}.$$

Moreover, the bound $n^{-\Omega(1)}$ on the error term is uniform (with respect to $\mathbf{k}$ and the configuration $\mathbf{u} = \mathbf{a}$).

The proof of the Negligibility Theorem is quite technical and elaborate. However, it employs geometric and asymptotic considerations which will also be relevant in Section 6. To facilitate the flow of reading, we choose to put it in the Appendix; see Subsection A.2.

Estimating the Joint Factorial Moments

Before we get to the joint moments we need one more lemma. Note that we are still under the conditioning $\mathbf{u} = \mathbf{a}$.

Lemma 5.8. *Let $\mathbf{x} = (x_1, x_2, x_3)$ be one of the t fixed triplets and let $\mathbf{q} = (s_1, s_2, s_3, z)$ be a quadruplet of distinct vertices (not from U). Define the six-dimensional random variable*

$$\mathbf{H} = (H_{x_1 s_1}, H_{x_2 s_2}, H_{x_3 s_3}, H_{s_1 z}, H_{s_2 z}, H_{s_3 z}).$$

Then, the variable $\mathbf{H}$ has a joint density function $f_{\mathbf{H}}(\mathbf{h})$ on $[0, h_n]^6$ given the event

$$\mathrm{CD}^*(\mathbf{x}; \mathbf{q}) = \mathrm{CD}^*(x_1, z_1) \wedge \mathrm{CD}^*(x_2, z_2) \wedge \mathrm{CD}^*(x_3, z_3).$$

Moreover, $f_{\mathbf{H}}(\mathbf{h})$ is Lipschitz (and in particular, continuous) and its value at $\mathbf{h} = \mathbf{0}$ is positive.

Proof. To prove that f exists, it is sufficient to prove that there exists a constant C such that for every box $Q \subseteq [0, h_n]^6$,

$$\mathbb{P}_{\mathbf{a}} \left(\mathbf{H} \in Q \mid \mathrm{CD}^*(\mathbf{x}; \mathbf{q}) \right) \leq C \mathrm{vol}(Q).$$

For every $1 \leq i \leq 3$, the probability of $H_{x_i s_i}, H_{s_i z} \in [a, b] \times [c, d]$ is the area of the intersection of the two annuli around x_i, z and radii $2r - 2b < 2r - 2a$ and $2r - 2d < 2r - 2c$ respectively. We invoke the proof of Lemma A.13 from the Appendix, which handles intersections of annuli. If we denote $\Delta = \|x_i - z\|$ then the area in question is

$$2 \iint\limits_J \frac{r_1 r_2}{\Delta \cdot y} \mathrm{d}r_1 \mathrm{d}r_2$$

where $J = [2r - 2b, 2r - 2a] \times [2r - 2d, 2r - 2c]$ and y is a specific function of r_1, r_2 (which also depends on Δ). It is not hard to show that for $a, b, c, d \leq h_n$ and $r \leq \Delta \leq \sqrt{3}r$ there is a constant C' such that $\frac{r_1 r_2}{\eta y} \leq C'$. Therefore, given $\mathrm{CD}^*(x_i, z)$, the probability that s_i lies in the intersection is

$$\leq \mathrm{Const} \cdot \mathrm{vol}([a, b] \times [c, d]).$$

Due to the independence between the s-vertices, we can multiply and get

$$\mathbb{P} \left(\mathbf{H} \in Q \mid \mathrm{CD}^*(\mathbf{x}; \mathbf{q}) \right) \leq \mathrm{Const} \cdot \mathrm{vol}(Q)$$

which proves that $f_{\mathbf{H}}(\mathbf{h})$ exists.

The values of $f_{\mathbf{H}}$ can be calculated with boxes as follows:

$$f_{\mathbf{H}}(h_1, \ldots, h_6) = \lim_{\varepsilon \to 0} \left[\frac{1}{\varepsilon^6} \mathbb{P} \left(H_{x_1 s_1} \in [h_1, h_1 + \varepsilon], \ldots, H_{s_3 z} \in [h_6, h_6 + \varepsilon] \;\middle|\; \mathrm{CD}^*(\mathbf{x}; \mathbf{q}) \right) \right].$$

Substituting $\mathbf{h} = \mathbf{0}$, the conditioned probability again turns into three integrals over $\frac{r_1 r_2}{\Delta_i \cdot y}$ with varying $\Delta_1, \Delta_2, \Delta_3$ which are always in $[r, \sqrt{3}r]$. The same considerations (but now with a lower bound) yield $f_{\mathbf{H}}(\mathbf{0}) > 0$. Moreover, the dependence of $\frac{r_1 r_2}{\Delta y}$ on r_1, r_2 in the relevant region is Lipschitz, which then shows that $f_{\mathbf{H}}(\mathbf{h})$ is itself Lipschitz. ■

The following theorem the main result of the current section.

Theorem 5.9. *There exist positive constants $\lambda_1, \lambda_2, \ldots, \lambda_t$ such that for every $0 \le \mathbf{k} \le K$,*

$$\mathbb{E}_{\mathbf{a}}\left((\mathbf{Z})_{\mathbf{k}}\right) = \lambda_1^{k_1} \lambda_2^{k_2} \ldots \lambda_t^{k_t} + n^{-\Omega(1)}.$$

Moreover, the term $n^{-\Omega(1)}$ is uniform (with respect to $\mathbf{k}$ and the configuration $\mathbf{u} = \mathbf{a}$).

Remark 5.10. As we shall see, the constants $\lambda_1, \lambda_2, \ldots, \lambda_t$ depend on the geometric configuration $\mathbf{u} = \mathbf{a}$, however they can be bound uniformly by constants $0 < \lambda_{\min} < \lambda_{\max}$.

In the proof we employ the following notation. First, recall the function $A(h)$ from Equation (3), which computes the area of the lens L_{xy} given than $H_{xy} = h$. Now, for a vector $\mathbf{h} = (h_1, h_2, \ldots, h_\ell)$, we write

$$\Sigma A(\mathbf{h}) = \sum_{i=1}^{\ell} A(h_i), \quad \Pi A(\mathbf{h}) = \prod_{i=1}^{\ell} A(h_i).$$

For a tuple $\mathbf{k} = (k_1, k_2, \ldots, k_t)$ we also denote $k = k_1 + k_2 + \cdots + k_t$.

Proof of Theorem 5.9. First, from the Negligibility Theorem 5.7, it is sufficient to prove

$$\mathbb{E}_{\mathbf{a}}\left(\tilde{\mathbf{Z}}_{\mathbf{k}}\right) = \lambda_1^{k_1} \lambda_2^{k_2} \ldots \lambda_t^{k_t} + n^{-\Omega(1)}$$

where $\tilde{\mathbf{Z}}_{\mathbf{k}}$ counts *well-behaved* $\mathbf{k}$-tuples

$$\mathbf{Q} = \left(\mathbf{q}_1^{(1)}, \ldots, \mathbf{q}_{k_1}^{(1)}, \ldots, \mathbf{q}_1^{(t)}, \ldots, \mathbf{q}_{k_t}^{(t)}\right)$$

of witnesses for S^*. We prove the asymptotic equivalence

$$\mathbb{E}_{\mathbf{a}}\left(\tilde{\mathbf{Z}}_{\mathbf{k}}\right) \approx \lambda_1^{k_1} \lambda_2^{k_2} \ldots \lambda_t^{k_t}.$$

Keeping track of the error is postponed to the end of the proof. Uniformity with respect to $\mathbf{k}$ and $\mathbf{u} = \mathbf{a}$ follows from the fact that we rely only on the general assumptions about them.

For a $\mathbf{k}$-tuple of quadruplets $\mathbf{Q}$, let $\mathrm{WB}(\mathbf{Q})$ denote the event that it is a well-behaved $\mathbf{k}$-tuple of witnesses for S^*. Then

$$\tilde{\mathbf{Z}}_{\mathbf{k}} = \sum_{\mathbf{Q}} \mathbb{1}_{\{\mathrm{WB}(\mathbf{Q})\}}$$

where the sum is over all $\mathbf{k}$-tuples $\mathbf{Q}$ of $4k$ distinct vertices (not from U). So

$$\mathbb{E}_{\mathbf{a}}\left(\tilde{\mathbf{Z}}_{\mathbf{k}}\right) = \sum_{\mathbf{Q}} \mathbb{P}_{\mathbf{a}}\left(\mathrm{WB}(\mathbf{Q})\right).$$

By symmetry, the probability $\mathbb{P}_{\mathbf{a}}\left(\mathrm{WB}(\mathbf{Q})\right)$ is the same for every $\mathbf{Q}$, let us denote is P. Therefore

$$\mathbb{E}_{\mathbf{a}}\left(\tilde{\mathbf{Z}}_{\mathbf{k}}\right) = (n-v)_{4k} P \approx n^{4k} P.$$

From now on, we fix a $\mathbf{k}$-tuple $\mathbf{Q}$ of $4k$ vertices and estimate the probability P that it is a well-behaved $\mathbf{k}$-tuple of witnesses for S^*.

Let $\mathrm{CD}(\mathbf{Q})$ denote the event that every quadruplet $\mathbf{q}_j^{(i)}$ of $\mathbf{Q}$ satisfies $\mathrm{CD}^*\left(\mathbf{x}^{(i)}; \mathbf{q}_j^{(i)}\right)$. We evaluate P through the probabilistic chain rule:

$$P = \mathbb{P}_{\mathbf{a}}\left(\mathrm{WB}(\mathbf{Q}) \;\middle|\; \mathrm{CD}(\mathbf{Q})\right) \mathbb{P}_{\mathbf{a}}\left(\mathrm{CD}(\mathbf{Q})\right).$$

Start with $\mathbb{P}_{\mathbf{a}}\left(\mathrm{CD}(\mathbf{Q})\right)$. The probability that a given z vertex satisfies the CD^* conditions over $\mathbf{x}^{(i)}$ is the area q_i of the set $R^{(i)}$ (see Definition 5.2). The z vertices are all distinct, therefore independent, and so

$$\mathbb{P}_{\mathbf{a}}\left(\mathrm{CD}(\mathbf{Q})\right) = q_1^{k_1} q_2^{k_2} \dots q_t^{k_t}.$$

Also recall that $\mathbf{u} = \mathbf{a}$ is q_0-feasible, thus $q_i \geq q_0$ for every $1 \leq i \leq t$.

Now for the main term, $\mathbb{P}_{\mathbf{a}}\left(\mathrm{WB}(\mathbf{Q}) \;\middle|\; \mathrm{CD}(\mathbf{Q})\right)$. The idea is to represent this probability as an integral over the $6k$ relevant H-variables. Then we "decompose" the resulting integral into a product of $6k$ one-dimensional integrals.

Given the event $\mathrm{CD}(\mathbf{Q})$, the meaning of the event $\mathrm{WB}(\mathbf{Q})$ is that the $6k$ relevant UN^*-relations are all satisfied, and also that there is no overlap between their $6k$ witnesses: they are all distinct, and none of them is from U or from $\mathbf{Q}$.[6] Recall that $\mathrm{UN}^*(x,y)$ can be understood as the requirement that L_{xy} contains exactly one vertex (the "witness") and that $H_{xy} \in [0, h_n]$.

For every $1 \leq i \leq t$ and $1 \leq j \leq k_i$ let $\mathbf{H}_j^{(i)}$ be the 6-dimensional variable of the relevant H-variables for the extension $S^*\left(\mathbf{x}^{(i)}; \mathbf{q}_j^{(i)}\right)$. Write the $6k$-dimensional variable

$$\mathbf{H} = \left(\mathbf{H}_1^{(1)}, \dots, \mathbf{H}_{k_1}^{(1)}, \dots, \mathbf{H}_1^{(t)}, \dots, \mathbf{H}_{k_t}^{(t)}\right).$$

Given that $\mathbf{H} = \mathbf{h}$ for some $6k$-dimensional vector $\mathbf{h}$, the probability of $\mathrm{WB}(\mathbf{Q})$ is

$$(n-v-4k)_{6k} \cdot \Pi A(\mathbf{h}) \cdot (1-M)^{n-v-10k}$$

where M is the area of the union of the $6k$ lenses.

Two asymptotic estimations make this expression simpler:

[6] It actually also requires that the vertices of $\mathbf{Q}$ are well-spaced, but we know it has negligible effect on the probability so we may ignore it right now.

1. $(n-v-4k)_{6k} \approx n^{6k}$ (easily justified by the logarithmic bounds on v, k).

2. $(1-M)^{n-v-10k} \approx \exp(-n\Sigma A(\mathbf{h}))$. Indeed, we may again assume that the vertices of $\mathbf{Q}$ are well-spaced (from the Negligibility Theorem). Then, Lemma A.15 shows that the areas of the intersections between lenses are $n^{-1\frac{1}{6}+o(1)}$. So

$$M = \Sigma A(\mathbf{h}) + n^{-1\frac{1}{6}+o(1)}.$$

We also know that each $A(h)$ is at most $A(h_n) = \frac{(\ln n)^{O(1)}}{n}$ so the same bound applies for M. These two facts justify the above estimation.

We get the simpler, asymptotically equivalent expression

$$T(\mathbf{h}) = n^{6k} \cdot \Pi A(\mathbf{h}) \cdot \exp\left(-n\Sigma A(\mathbf{h})\right).$$

Integrating over the different values for $\mathbf{H} = \mathbf{h}$, we get

$$\mathbb{P}_{\mathbf{a}}\left(\mathrm{WB}(\mathbf{Q}) \bigm| \mathrm{CD}(\mathbf{Q})\right) = \int\!\!\cdots\!\!\int_{[0,h_n]^{6k}} T(\mathbf{h}) f_{\mathbf{H}}(\mathbf{h}) \mathrm{d}\mathbf{h}$$

where $f_{\mathbf{H}}(\mathbf{h})$ is the joint density function of $\mathbf{H}$ given $\mathrm{CD}(\mathbf{Q})$. It now mainly remains to estimate this integral.

Consider $f_{\mathbf{H}}(\mathbf{h})$ first. The $\mathbf{H}_j^{(i)}$ variables are independent, because $\mathbf{H}_j^{(i)}$ depends only on the vertices $\mathbf{q}_j^{(i)}$. Here we strongly use the underlying conditioning by $\mathbf{u} = \mathbf{a}$; without it, $\mathbf{H}_j^{(i)}$ would also depend on $\mathbf{x}^{(i)}$. Write

$$f_{\mathbf{H}}(\mathbf{h}) = \prod_{i=1}^{t} \prod_{j=1}^{k_i} f_{\mathbf{H}_j^{(i)}}(\mathbf{h}_j^{(i)})$$

where $f_{\mathbf{H}_j^{(i)}}(\mathbf{h}_j^{(i)})$ is the joint density of $\mathbf{H}_j^{(i)}$ given $\mathrm{CD}(\mathbf{Q})$. By Lemma 5.8, this joint density function indeed exists and is also continuous over the region $[0, h_n]^6$. Therefore $f_{\mathbf{H}_j^{(i)}}(\mathbf{h}_j^{(i)}) \approx f_{\mathbf{H}_j^{(i)}}(\mathbf{0})$ uniformly over $[0, h_n]^6$ as $n \to \infty$. The constant $f_{\mathbf{H}_j^{(i)}}(\mathbf{0})$ is positive and depends only on the geometric locations $a_1^{(i)}, a_2^{(i)}, a_3^{(i)}$ (and, in particular, does not depend on j). Denote it by f_i. So we can say that

$$f_{\mathbf{H}}(\mathbf{h}) \approx f_1^{k_1} f_2^{k_2} \dots f_t^{k_t}$$

uniformly over $[0, h_n]^{6k}$ as $n \to \infty$. This allows us to "take the density out of the integral":

$$\int\!\!\cdots\!\!\int_{[0,h_n]^{6k}} T(\mathbf{h}) f_{\mathbf{H}}(\mathbf{h}) \mathrm{d}\mathbf{h} \approx f_1^{k_1} f_2^{k_2} \dots f_t^{k_t} \int\!\!\cdots\!\!\int_{[0,h_n]^{6k}} T(\mathbf{h}) \mathrm{d}\mathbf{h}.$$

The remaining integral naturally decomposes into a product:

$$\underset{[0,h_n]^{6k}}{\int\!\cdots\!\int} n^{6k}\Pi A(\mathbf{h}) \exp\left(-n\Sigma A(\mathbf{h})\right) \mathrm{d}\mathbf{h} = \left(\int_0^{h_n} nA(h)e^{-nA(h)}\mathrm{d}h\right)^{6k}.$$

This integral already came up in the proof of Theorem 4.1, where it was shown that

$$\int_0^{h_n} nA(h)e^{-nA(h)}\mathrm{d}h \approx Cn^{-2/3}$$

for a positive constant C. So overall,

$$\begin{aligned}\mathbb{P}_{\mathbf{a}}\left(\mathrm{WB}(\mathbf{Q}) \,\middle|\, \mathrm{CD}(\mathbf{Q})\right) &\approx f_1^{k_1} f_2^{k_2} \dots f_t^{k_t} \cdot \left(Cn^{-2/3}\right)^{6k} \\ &\approx f_1^{k_1} f_2^{k_2} \dots f_t^{k_t} \cdot C^{6k} n^{-4k}.\end{aligned}$$

In conclusion,

$$\begin{aligned}\mathbb{E}_{\mathbf{a}}\left(\tilde{\mathbf{Z}}_{\mathbf{k}}\right) &\approx n^{4k}P = n^{4k} \cdot \mathbb{P}_{\mathbf{a}}\left(\mathrm{CD}(\mathbf{Q})\right) \cdot \mathbb{P}_{\mathbf{a}}\left(\mathrm{WB}(\mathbf{Q}) \,\middle|\, \mathrm{CD}(\mathbf{Q})\right) \\ &\approx n^{4k} \cdot q_1^{k_1} q_2^{k_2} \dots q_t^{k_t} \cdot f_1^{k_1} f_2^{k_2} \dots f_t^{k_t} \cdot C^{6k} n^{-4k} \\ &= (C^6 q_1 f_1)^{k_1} (C^6 q_2 f_2)^{k_2} \dots (C^6 q_t f_t)^{k_t}.\end{aligned}$$

Denoting $\lambda_i = C^6 q_i f_i$, we obtain the desired estimation

$$\mathbb{E}_{\mathbf{a}}\left(\tilde{\mathbf{Z}}_{\mathbf{k}}\right) \quad \approx \quad \lambda_1^{k_1} \lambda_2^{k_2} \dots \lambda_t^{k_t}.$$

The only thing remaining is to explain why the error is $n^{-\Omega(1)}$. Since

$$\lambda_1^{k_1} \lambda_2^{k_2} \dots \lambda_t^{k_t} = e^{O(k)} = n^{o(1)},$$

writing

$$\mathbb{E}_{\mathbf{a}}\left(\tilde{\mathbf{Z}}_{\mathbf{k}}\right) = \lambda_1^{k_1} \lambda_2^{k_2} \dots \lambda_t^{k_t} + n^{-\Omega(1)}$$

is equivalent to writing

$$\mathbb{E}_{\mathbf{a}}\left(\tilde{\mathbf{Z}}_{\mathbf{k}}\right) = \left(1 + n^{-\Omega(1)}\right) \lambda_1^{k_1} \lambda_2^{k_2} \dots \lambda_t^{k_t}.$$

So we need to explain why all the transitions involving asymptotic equivalence $\approx$ indeed admit a multiplicative term of $1 + n^{-\Omega(1)}$ (and also notice that their amount is poly-logarithmic). We shall do that briefly for some key transitions.

$(n-v)_{4k} \approx n^{4k}$: Note that

$$\frac{(n-v-4k)^{4k}}{n^{4k}} \leq \frac{(n-v)_{4k}}{n^{4k}} \leq 1.$$

But

$$\frac{(n-v-4k)^{4k}}{n^{4k}} = \left(1 - \frac{v+4k}{n}\right)^{4k} = 1 + O\left(\frac{k(v+4k)}{n}\right) = 1 + n^{-\Omega(1)}.$$

$f_{\mathbf{H}}(\mathbf{h}) \approx f_1^{k_1} f_2^{k_2} \dots f_t^{k_t}$: The 6-dimensional joint density function $f_{\mathbf{H}_j^{(i)}}(\mathbf{h}_j^{(i)})$ is Lipschitz (from Lemma 5.8), so

$$\left| f_{\mathbf{H}_j^{(i)}}(\mathbf{h}_j^{(i)}) - f_i \right| \le \text{Const} \cdot \sqrt{6} h_n$$

for every $\mathbf{h}_j^{(i)} \in [0, h_n]^6$. f_i is a positive constant, therefore

$$f_{\mathbf{H}_j^{(i)}}(\mathbf{h}_j^{(i)}) = (1 + O(h_n)) f_i = \left(1 + n^{-\Omega(1)}\right) f_i.$$

$(1-M)^{n-v-10k} \approx \exp\left(-n\Sigma A(\mathbf{h})\right)$: again, we rely on the two estimations

$$M = \Sigma A(\mathbf{h}) + n^{-1\frac{1}{6}+o(1)}$$

and $M = \frac{(\ln n)^{O(1)}}{n}$. There are a three steps here:

$$(1-M)^{n-v-10k} \approx (1-M)^n \overset{(*)}{\approx} e^{-Mn} \approx e^{-n\Sigma A(\mathbf{h})}.$$

In step $(*)$, for example, the multiplicative error is

$$1 + O(nM^2) = 1 + n^{-\Omega(1)}.$$

$\int_0^{h_n} n\alpha(h) e^{-n\alpha(h)} \mathrm{d}h \approx C n^{-2/3}$: Carefully follow the proof of Theorem 4.1. ■

As mentioned in Remark 5.10, the exact values of $\lambda_1, \lambda_2, \dots, \lambda_t$ from Theorem 5.9 depends on the geometric configuration $\mathbf{u} = \mathbf{a}$. However, as the following proposition states, there are *universal bounds* on those values.

Proposition 5.11. *There exist positive constants* $\lambda_{\min}, \lambda_{\max}$*, independent of* $\mathbf{u} = \mathbf{a}$*, such that*

$$\lambda_{\min} \le \lambda_1, \lambda_2, \dots, \lambda_t \le \lambda_{\max}$$

for every q_0*-feasible configuration* $\mathbf{u} = \mathbf{a}$*.*

Proof. Recall that $\lambda_i = C^6 q_i f_i$.

C comes from the integral $\int_0^{h_n} n\alpha(h) e^{-n\alpha(h)} \mathrm{d}h$ and does not depend on $\mathbf{u} = \mathbf{a}$ in any way.

q_i is the area of $R^{(i)}$ which is trivially bounded by 1 from above, and also bounded by q_0 from below by our assumption.

f_i comes from the joint density function of $\mathbf{H}_j^{(i)}$. Its evaluation comes down to integrals of the form

$$\iint_J \frac{r_1 r_2}{\Delta y} \mathrm{d}r_1 \mathrm{d}r_2$$

as mentioned in Lemma 5.8. The integrand $\frac{r_1 r_2}{\Delta y}$ can be bounded by positive constants from above and below, which do not depend on $r \le \Delta \le \sqrt{3}r$. Therefore there are positive constants $f_{\min}, f_{\max}$ that universally bound f_i from both sides.

Overall

$$C^6 q_0 f_{\min} \le C^6 q_i f_i \le C^6 f_{\max}$$

so defining $\lambda_{\min} = C^6 q_0 f_{\min}$ and $\lambda_{\max} = C^6 f_{\max}$ ends the proof. ■

5.2 The Joint Distribution

With a good estimation of the joint factorial moments, it is now the time to exploit the flexible moments method from Section 3.

Theorem 5.12. *Let $\lambda_1, \lambda_2, \ldots, \lambda_t$ be the constants from Theorem 5.9. Let $K = \ln\ln\ln n$ as always. Then for every t-tuple of integers $\mathbf{z} = (z_1, \ldots, z_t) \le \ln\ln\ln\ln n$,*

$$\mathbb{P}_{\mathbf{a}}(\mathbf{Z} = \mathbf{z}) = \prod_{i=1}^{t} \left[e^{-\lambda_i} \frac{\lambda_i^{z_i}}{z_i!} \right] + \frac{e^{O(K)}}{K!}.$$

Moreover, the term $\frac{e^{O(K)}}{K!}$ is uniform (with respect to $\mathbf{z}$ and the configuration $\mathbf{u} = \mathbf{a}$).

Proof. Let $X^{(1)}, X^{(2)}, \ldots, X^{(t)}$ be t independent random variables with $X^{(i)} \sim \mathrm{Pois}(\lambda_i)$. Write $\mathbf{X} = \left(X^{(1)}, X^{(2)}, \ldots, X^{(t)}\right)$. Note that for every $\mathbf{z}$,

$$\mathbb{P}(\mathbf{X} = \mathbf{z}) = \prod_{i=1}^{t} \left[e^{-\lambda_i} \frac{\lambda_i^{z_i}}{z_i!} \right]$$

so we need to prove

$$\mathbb{P}_{\mathbf{a}}(\mathbf{Z} = \mathbf{z}) = \mathbb{P}(\mathbf{X} = \mathbf{z}) + \frac{e^{O(K)}}{K!}$$

for $0 \le \mathbf{z} \le \ln\ln\ln\ln n$.

We do it with the flexible moments method. For every $0 \le \mathbf{k} \le K$, by Theorem 5.9 we know that

$$\mathbb{E}_{\mathbf{a}}\left((\mathbf{Z})_{\mathbf{k}}\right) = \lambda_1^{k_1} \lambda_2^{k_2} \ldots \lambda_t^{k_t} + n^{-\Omega(1)}.$$

In addition, for every $\mathbf{k}$,

$$\mathbb{E}\left((\mathbf{X})_{\mathbf{k}}\right) = \lambda_1^{k_1}\lambda_2^{k_2}\dots\lambda_t^{k_t}.$$

The flexible moments method bounds probabilities through the joint factorial moments:

$$\begin{aligned}\mathbb{E}\left(\pi^{-}(\mathbf{X})\right) \leq \mathbb{P}\left(\mathbf{X}=\mathbf{z}\right) \leq \mathbb{E}\left(\pi^{+}(\mathbf{X})\right),\\ \mathbb{E}_{\mathbf{a}}\left(\pi^{-}(\mathbf{Z})\right) \leq \mathbb{P}_{\mathbf{a}}\left(\mathbf{Z}=\mathbf{z}\right) \leq \mathbb{E}_{\mathbf{a}}\left(\pi^{+}(\mathbf{Z})\right),\end{aligned}$$

where $\pi^{+}(\mathbf{x}), \pi^{-}(\mathbf{x})$ are the suitable multivariate polynomials from Definition 3.5. Recall that they are indeed linear combinations of the joint falling factorials: we can write

$$\begin{aligned}\pi^{+}(\mathbf{x}) &= \sum_{0\leq\mathbf{k}\leq K}\alpha_{\mathbf{k}}(\mathbf{x})_{\mathbf{k}},\\ \pi^{-}(\mathbf{x}) &= \sum_{0\leq\mathbf{k}\leq K}\beta_{\mathbf{k}}(\mathbf{x})_{\mathbf{k}}\end{aligned}$$

with coefficients $\alpha_{\mathbf{k}}, \beta_{\mathbf{k}}$. The coefficients can be easily computed, but for the proof we are satisfied with the simple bound $|\alpha_{\mathbf{k}}|, |\beta_{\mathbf{k}}| \leq 1$, which directly follows from the definition. We therefore get

$$\begin{aligned}\left|\mathbb{E}_{\mathbf{a}}\left(\pi^{+}(\mathbf{Z})\right) - \mathbb{E}\left(\pi^{+}(\mathbf{X})\right)\right| &\leq \sum_{0\leq\mathbf{k}\leq K}|\alpha_{\mathbf{k}}|\left|\mathbb{E}_{\mathbf{a}}\left((\mathbf{Z})_{\mathbf{k}}\right) - \mathbb{E}\left((\mathbf{X})_{\mathbf{k}}\right)\right|\\ &= n^{-\Omega(1)}\cdot\sum_{0\leq\mathbf{k}\leq K}|\alpha_{\mathbf{k}}| = n^{-\Omega(1)}\end{aligned}$$

and similarly for π^{-}. It remains to bound $|\mathbb{E}\left(\pi^{+}(\mathbf{X})\right) - \mathbb{E}\left(\pi^{-}(\mathbf{X})\right)|$.

$$\begin{aligned}\left|\mathbb{E}\left(\pi^{+}(\mathbf{X})\right) - \mathbb{E}\left(\pi^{-}(\mathbf{X})\right)\right| &= \mathbb{E}\left[\prod_{i=1}^{t}\left(1+\frac{(X_i)_{z_i+s_i}}{z_i!s_i!}\right)-1\right]\\ &= \mathbb{E}\left[\sum_{\emptyset\neq I\subseteq[t]}\left(\prod_{i\in I}\frac{(X_i)_{z_i+s_i}}{z_i!s_i!}\right)\right]\\ &= \sum_{\emptyset\neq I\subseteq[t]}\left[\prod_{i\in I}\frac{(\lambda_i)^{z_i+s_i}}{z_i!s_i!}\right]\\ &= \prod_{i=1}^{t}\left(1+\frac{(\lambda_i)^{z_i+s_i}}{z_i!s_i!}\right)-1.\end{aligned}$$

Here s_i is either $K-z_i$ or $K-z_i-1$ (depending on parity). Since we have

$$K = \ln\ln\ln n,\ z_i \leq \ln\ln\ln\ln n,\ t \leq (\ln\ln\ln\ln n)^3$$

we can write

$$\begin{aligned}\prod_{i=1}^{t}\left(1+\frac{(\lambda_i)^{z_i+s_i}}{z_i!s_i!}\right)-1 &= O\left(\sum_{i=1}^{t}\frac{(\lambda_i)^{z_i+s_i}}{z_i!s_i!}\right)\\ &= O\left(\sum_{i=1}^{t}\frac{e^{O(K)}}{K!}\right)\\ &= \frac{e^{O(K)}}{K!}.\end{aligned}$$

Overall

$$\left|\mathbb{P}_{\mathbf{a}}\left(\mathbf{Z}=\mathbf{z}\right)-\mathbb{P}\left(\mathbf{X}=\mathbf{z}\right)\right| \le n^{-\Omega(1)}+\frac{e^{O(K)}}{K!}=\frac{e^{O(K)}}{K!}$$

and that finishes the proof. ∎

6 Expressing Arbitrary Graph Structures

In this section we prove the most important result in the paper: that first-order logic can unravel arbitrary graph structures (with a slowly-growing number of vertices) within the random geometric graph. This result is the key to the construction of a first order sentence with non-converging probability.

To state this result formally, we begin by explaining how we encode graph structures in $G_{\mathbb{T}^2}(n,r)$. We use the extension formula S as our building block.

Definition 6.1. 1. Let $\mathbf{x}=(x_1,x_2,x_3)$ be a fixed triplet of distinct vertices in $G_{\mathbb{T}^2}(n,r)$. Define the (random) set of vertices $U_{\mathbf{x}}\subseteq V$ as follows:

$$U_{\mathbf{x}}=\left\{z \mid \exists s_1,s_2,s_3.\ S(\mathbf{x};s_1,s_2,s_3,z)\right\}.$$

2. Let $U\subseteq V$ be a set of vertices and let $w\in V\setminus U$ be another fixed vertex. Define $H_w(U)$ as the (random) graph whose vertex set is U and whose edge set is

$$\left\{\{u_1,u_2\} \mid \exists\mathbf{q}.\ S(u_1,u_2,w;\mathbf{q})\right\}.$$

We see that S can be used to define both a vertex set and an edge set. This is somewhat analogous to the way Spencer and Shelah encode graph structures in $G(n,p)$ (for $p=n^{-\alpha}$ with rational $0<\alpha<1$) with a suitable extension formula (see Sub-subsection 8.3.1 in Spencer [17]).

Recall that to successfully count witnesses for S-extensions over a set of vertices, two geometric conditions are required: well-spacedness and feasibility. Here it will be more comfortable to replace feasibility with a simpler, stronger condition.

Definition 6.2. We say that a set of vertices U in $G_{\mathbb{T}^2}(n,r)$ is *nicely positioned* if its vertices are well-spaced and its diameter is at most $\frac{r}{4}$.

$\frac{r}{4}$ is simply a comfortable arbitrary choice. Note that $\mathrm{diam}(U) \le \frac{r}{4}$ assures q_0-feasibility for $q_0 = 0.63\pi r^2$. Indeed, if the distances between three vertices x_1, x_2, x_3 are all less than $\frac{r}{4}$, then the corresponding locus R from Definition 5.2 contains the annulus around x_1 with radii $\frac{5}{4}r, \left(\sqrt{3} - \frac{1}{4}\right) r$, and the area of this annulus is

$$\left[\left(\sqrt{3} - \frac{1}{4}\right)^2 - \left(\frac{5}{4}\right)^2\right] \pi r^2 > 0.63\pi r^2.$$

We can now formulate the main theorem.

Theorem 6.3 (Main Theorem)**.** *Fix $1 \le v \le \ln\ln\ln\ln n$.*

1. *A.a.s., there exists a triplet $\mathbf{x} = (x_1, x_2, x_3)$ such that $|U_{\mathbf{x}}| = v$ and $U_{\mathbf{x}}$ is nicely positioned.*
2. *A.a.s., for every nicely positioned set of vertices U with $|U| = v$, and for every possible graph H with v vertices, there exists a vertex $w \in V \setminus U$ such that $H_w(U)$ is isomorphic to H.*

In particular, the following is true. A.a.s., for every possible graph H on v vertices, there exist four vertices x_1, x_2, x_3, w such that $H_w(U_{\mathbf{x}}) \cong H$.

The theorem shows that the first order extension S can be used to express any graph structure H with at most $\ln\ln\ln\ln n$ vertices in $G_{\mathbb{T}^2}(n, r)$. In Section 7 we will see how it is used to finish the proof of Theorem 1.17.

Subsection 6.1 is dedicated to the proof of Theorem 6.3. As we shall see, the proof crucially relies on the concentration of certain random variables. Proving this concentration theorem is a bit more involved, and we handle it separately in Subsection 6.2.

6.1 Proving the Main Theorem

Although the two parts of Theorem 6.3 may appear different, they actually rely on the same core ideas: in both cases we need to find vertices such that counting the witnesses over them gives us certain specific values. We prove the second part, which is harder, and then explain the minor adjustments needed for a proof of the first part.

Our first step is to reformulate (the second part of) the theorem more conveniently. We get rid of quantifiers and fix a geometric configuration of U.

Theorem 6.4. *Fix $1 \le v \le \ln\ln\ln\ln n$. Also fix a set of vertices U with $|U| = v$, a graph structure H on U, and a geometric configuration $\mathbf{u} = \mathbf{a}$ of U which is nicely positioned. Then, given $\mathbf{u} = \mathbf{a}$, with probability $1 - \exp\left(-\Omega(\ln^2 n)\right)$ there exists a vertex w such that $H_w(U) = H$. Moreover, the asymptotic bound on the probability is uniform: it can be made independent of the choices of U, H and the configuration $\mathbf{u} = \mathbf{a}$.*

Let us explain how to obtain Theorem 6.3 part 2 from Theorem 6.4. First consider the set U and the graph H fixed and integrate over the nicely positioned configurations $\mathbf{u} = \mathbf{a}$. Uniformity of the bounds is applied to maintain a bound of $\exp\left(-\Omega(\ln^2 n)\right)$ on the probability that the appropriate w does not exist, given a specific set U which is known to be nicely positioned and a specific graph structure H. Then, take a union bound over all $2^{\binom{v}{2}}$ graph structures H and over all $\binom{n}{v}$ sets of vertices U. This is where the exponential bound on the probability becomes crucial. We get

$$2^{\binom{v}{2}}\binom{n}{v} \cdot \exp\left(-\Omega(\ln^2 n)\right) = \exp\left(\ln n \cdot O(\ln\ln\ln\ln n)\right) \cdot \exp\left(-\Omega(\ln^2 n)\right) = o(1).$$

So indeed, a.a.s. we can find a suitable w for every U, H.

Let us now briefly review the course of the proof of Theorem 6.4. To obtain the strong probabilistic bound we use the following trick. We consider geometric configurations of *all* the vertices in the random graph, which we call *complete configurations*. Given a complete configuration, the existence of a suitable vertex w is deterministic; we call a complete configuration that determines the existence of such a w a *good* configuration. We would like to show that with very high probability, the vertices of the RGG are configured in a good way. To do that, we give a sufficient condition for being a good configuration, based on the values of certain random variables, closely related to $Z^{(1)}, Z^{(2)}, \dots, Z^{(t)}$ from the previous section. A concentration result is applied to these variables to show that with very high probability, we can replace them with their expected values, which turns out to guarantee that the configuration is good.

From now on we fix $1 \le v \le \ln\ln\ln\ln n$. We also fix a set of vertices U with $|U| = v$, a graph structure H on U, and a geometric configuration $\mathbf{u} = \mathbf{a}$ which is nicely positioned. From now on, everything is conditioned by $\mathbf{u} = \mathbf{a}$ unless stated otherwise. Again, we use the notations $\mathbb{P}_{\mathbf{a}}(\cdot)$ and $\mathbb{E}_{\mathbf{a}}(\cdot)$ for probabilities and expectations in the RGG, conditioned by $\mathbf{u} = \mathbf{a}$.

Definition 6.5. Write $V = \{v_1, v_2, \dots, v_n\}$. A *complete configuration* is an event of the form

$$\{v_1 = b_1, v_2 = b_2, \dots, v_n = b_n\}$$

where $b_1, b_2, \dots, b_n \in \mathbb{T}^2$ are fixed points. Again, we use the abbreviated notation $\mathbf{v} = \mathbf{b}$.

Remark 6.6. Since we fixed the configuration $\mathbf{u} = \mathbf{a}$, the only relevant complete configurations are those who "agree" with it: every $u \in U$ must be assigned the same position by $\mathbf{v} = \mathbf{b}$ and by $\mathbf{u} = \mathbf{a}$. Consequently, when we discuss complete configurations from now on we always assume that they agree with $\mathbf{u} = \mathbf{a}$.

Definition 6.7. A complete configuration $\mathbf{v} = \mathbf{b}$ is called *good* if, given the configuration, there exists a vertex w such that $H_w(U) = H$.

Note that this is well-defined: a complete configuration determines all the information about the RGG, including the S-extensions and thus the existence of a suitable w.

To show that a complete configuration is good we use a *probabilistic method* argument. Suppose that we fix $\mathbf{v} = \mathbf{b}$. The idea is to choose w randomly and uniformly among $V \setminus U$, and then consider the probability that it satisfies $H_w(U) = H$ (which is a probability over an entirely new probability space!). If we manage to show that this probability is positive, then surely a suitable w exists, which means that the configuration is good.

For notational convenience, let us enumerate all *pairs* of vertices from U:

$$\binom{U}{2} = \{\mathbf{u}^{(1)}, \mathbf{u}^{(2)}, \dots, \mathbf{u}^{(t)}\}$$

where now $t = \binom{v}{2}$.

Definition 6.8. Given a complete configuration $\mathbf{v} = \mathbf{b}$, let W be a random vertex which is uniformly distributed over all $n - v$ vertices of $V \setminus U$. For every $1 \le i \le t$, define $X^{(i)}$ as the number of witnesses for the S-extension over the triplet $(\mathbf{u}^{(i)}, W)$. Notice that (given the configuration $\mathbf{v} = \mathbf{b}$) each $X^{(i)}$ is a function of W.

We use the notation $\mathbb{P}_{\mathbf{b}}(\cdot)$ for probabilities in this new probability space. To prove that a configuration is good, we would like to show that

$$\mathbb{P}_{\mathbf{b}}\left(H_W(U) = H\right) > 0.$$

Importantly, notice that the event $H_W(U) = H$ can be expressed through the values of the variables $X^{(i)}$: it occurs if and only if $X^{(i)} \ge 1$ whenever $\mathbf{u}^{(i)}$ is an edge of H and $X^{(i)} = 0$ whenever $\mathbf{u}^{(i)}$ is not an edge of H. It becomes a question about the joint distribution of $X^{(1)}, X^{(2)}, \dots, X^{(t)}$.

This is a good time to invoke the flexible moments method. First, for compatibility with the method, we replace $X^{(i)} \ge 1$ with $X^{(i)} = 1$:

$$\begin{aligned}\mathbb{P}_{\mathbf{b}}\left(H_W(U) = H\right) &= \mathbb{P}_{\mathbf{b}}\left(\substack{X^{(i)} \ge 1 \text{ when } \mathbf{u}^{(i)} \in E(H) \\ X^{(i)} = 0 \text{ when } \mathbf{u}^{(i)} \notin E(H)}\right) \\ &\ge \mathbb{P}_{\mathbf{b}}\left(\substack{X^{(i)} = 1 \text{ when } \mathbf{u}^{(i)} \in E(H) \\ X^{(i)} = 0 \text{ when } \mathbf{u}^{(i)} \notin E(H)}\right).\end{aligned}$$

Of course, it is sufficient to show that this smaller probability is positive. In this case, the variables are $\mathbf{X} = \left(X^{(1)}, \dots, X^{(t)}\right)$, the values $\mathbf{z}$ are all 0 or 1 and we set $K = \ln\ln\ln n$ as usual. Let $\pi^-(\mathbf{x})$ be the suitable multivariate polynomial, as defined in Definition 3.5. We get the lower bound

$$\mathbb{P}_{\mathbf{b}}\left(H_W(U) = H\right) \ge \mathbb{E}_{\mathbf{b}}\left(\pi^-(\mathbf{X})\right). \tag{6}$$

$\mathbb{E}_{\mathbf{b}}\left(\pi^-(\mathbf{X})\right)$ is a specific linear combination of the joint factorial moments: we can write

$$\mathbb{E}_{\mathbf{b}}\left(\pi^-(\mathbf{X})\right) = \sum_{0 \le \mathbf{k} \le K} \beta_{\mathbf{k}} \mathbb{E}_{\mathbf{b}}\left((\mathbf{X})_{\mathbf{k}}\right)$$

for coefficients $\beta_{\mathbf{k}}$ which can be explicitly computed.

Naturally, our next step is to investigate the joint factorial moments $\mathbb{E}_{\mathbf{b}}\left((\mathbf{X})_{\mathbf{k}}\right)$ for $0 \le \mathbf{k} \le K$. To do that we introduce the following definition.

Definition 6.9. For every $0 \leq \mathbf{k} \leq K$, define the variable $Y_{\mathbf{k}}$ in $G_{\mathbb{T}^2}(n, r)$ as the number of tuples

$$\left(w, \mathbf{q}_1^{(1)}, \ldots, \mathbf{q}_{k_1}^{(1)}, \ldots, \mathbf{q}_1^{(t)}, \ldots, \mathbf{q}_{k_t}^{(t)}\right)$$

such that w is a vertex and $\mathbf{q}_1^{(i)}, \ldots, \mathbf{q}_{k_i}^{(i)}$ are distinct witnesses for the S-extension over $(\mathbf{u}^{(i)}, w)$, for every $1 \leq i \leq t$.

Note that $Y_{\mathbf{k}}$ is a random variable in the RGG. Given a complete configuration $\mathbf{v} = \mathbf{b}$, the value of $Y_{\mathbf{k}}$ is determined, and we denote it $Y_{\mathbf{k}}\big|_{\mathbf{b}}$.

Lemma 6.10. *Given a complete configuration* $\mathbf{v} = \mathbf{b}$*, for every* $0 \leq \mathbf{k} \leq K$*,*

$$\mathbb{E}_{\mathbf{b}}\left((\mathbf{X})_{\mathbf{k}}\right) = \frac{1}{n-v} Y_{\mathbf{k}}\big|_{\mathbf{b}}.$$

Proof. By definition,

$$\mathbb{E}_{\mathbf{b}}\left((\mathbf{X})_{\mathbf{k}}\right) = \frac{1}{n-v} \sum_{w \in V \setminus U} \left(X^{(1)}(w)\right)_{k_1} \cdots \left(X^{(t)}(w)\right)_{k_t}.$$

Here $X^{(i)}(w)$ is the number of witnesses for S over $(\mathbf{u}^{(i)}, w)$ given $\mathbf{v} = \mathbf{b}$. Note that, for a given w, the term

$$\left(X^{(1)}(w)\right)_{k_1} \cdots \left(X^{(t)}(w)\right)_{k_t}$$

counts the number of $\mathbf{k}$-tuples of witnesses

$$\left(\mathbf{q}_1^{(1)}, \ldots, \mathbf{q}_{k_1}^{(1)}, \ldots, \mathbf{q}_1^{(t)}, \ldots, \mathbf{q}_{k_t}^{(t)}\right).$$

Therefore

$$Y_{\mathbf{k}}\big|_{\mathbf{b}} = \sum_{w \in V \setminus U} \left(X^{(1)}(w)\right)_{k_1} \cdots \left(X^{(t)}(w)\right)_{k_t}$$

and that finishes the proof. ∎

From the lemma, we can rewrite our lower bound from Equation (6) in terms of $Y_{\mathbf{k}}$:

$$\mathbb{P}_{\mathbf{b}}\left(H_W(U) = H\right) \geq \sum_{0 \leq \mathbf{k} \leq K} \beta_{\mathbf{k}} \mathbb{E}_{\mathbf{b}}\left((\mathbf{X})_{\mathbf{k}}\right) = \frac{1}{n-v} \sum_{0 \leq \mathbf{k} \leq K} \beta_{\mathbf{k}} Y_{\mathbf{k}}\big|_{\mathbf{b}}.$$

This is where concentration steps in. We shall now estimate the general expected values $\mathbb{E}_{\mathbf{a}}\left(Y_{\mathbf{k}}\right)$, and claim that the variables $Y_{\mathbf{k}}$ are highly concentrated around their expectations. The following simple definition and lemma reduce $\mathbb{E}_{\mathbf{a}}\left(Y_{\mathbf{k}}\right)$ into a more familiar form.

Definition 6.11. Set an arbitrary vertex $w_0 \in V \setminus U$. For every $1 \leq i \leq t$, define a random variable $Z^{(i)}$ to be the number of witnesses for the S-extension over $(\mathbf{u}^{(i)}, w_0)$ in $G_{\mathbb{T}^2}(n, r)$.

The variables $Z^{(i)}$ are essentially the same as the variables that we studied in Section 5: they count witnesses for the S-extension over fixed triplets.

Lemma 6.12. *For every* $0 \leq \mathbf{k} \leq K$,

$$\mathbb{E}_{\mathbf{a}}(Y_{\mathbf{k}}) = (n-v)\mathbb{E}_{\mathbf{a}}\left((\mathbf{Z})_{\mathbf{k}}\right).$$

Proof. $(\mathbf{Z})_{\mathbf{k}}$ counts $\mathbf{k}$-tuples of quadruplets

$$\left(\mathbf{q}_1^{(1)}, \ldots, \mathbf{q}_{k_1}^{(1)}, \ldots, \mathbf{q}_1^{(t)}, \ldots, \mathbf{q}_{k_t}^{(t)}\right)$$

such that $\mathbf{q}_i^{(i)}, \ldots, \mathbf{q}_{k_i}^{(i)}$ are distinct witnesses for S over $(\mathbf{u}^{(i)}, w_0)$. Now it is simply left to notice that $Y_{\mathbf{k}}$ is the sum of $n-v$ random variables with the same distribution as $(\mathbf{Z})_{\mathbf{k}}$ (by symmetry between the vertices of $V \setminus U$). ■

Finally, we formulate the concentration result. Its proof is presented in Subsection 6.2.

Theorem 6.13 (The Concentration Theorem)**.** *For every* $0 \leq \mathbf{k} \leq K$,

$$\mathbb{P}_{\mathbf{a}}\left(|Y_{\mathbf{k}} - \mathbb{E}_{\mathbf{a}}(Y_{\mathbf{k}})| = \Omega\left(\frac{\mathbb{E}_{\mathbf{a}}(Y_{\mathbf{k}})}{(\ln n)^{1000}}\right)\right) = \exp\left(-\Omega(\ln^2 n)\right).$$

More explicitly, there exists a constant $C > 0$ *such that*

$$\mathbb{P}_{\mathbf{a}}\left(|Y_{\mathbf{k}} - \mathbb{E}_{\mathbf{a}}(Y_{\mathbf{k}})| \geq \frac{C}{(\ln n)^{1000}}\mathbb{E}_{\mathbf{a}}(Y_{\mathbf{k}})\right) = \exp\left(-\Omega(\ln^2 n)\right).$$

Again, both C *and the bound are uniform with respect to* $\mathbf{k}$ *and the nicely positioned* $\mathbf{u} = \mathbf{a}$.

We are now ready to prove Theorem 6.4.

Proof of Theorem 6.4. From Theorem 6.13, with probability $1 - \exp\left(-\Omega(\ln^2 n)\right)$, the RGG is configured in such a way that the concentration inequality

$$|Y_{\mathbf{k}} - \mathbb{E}_{\mathbf{a}}(Y_{\mathbf{k}})| \leq \frac{C}{(\ln n)^{1000}}\mathbb{E}_{\mathbf{a}}(Y_{\mathbf{k}}) \tag{7}$$

holds for every $0 \leq \mathbf{k} \leq K$ (where C is some positive constant).

Fix a complete configuration $\mathbf{v} = \mathbf{b}$ such that (7) holds for every $0 \leq \mathbf{k} \leq K$. Let us show that it is a good configuration; that is, a configuration which admits a vertex w with $H_w(U) = H$. That will finish the proof.

As already explained, it is sufficient to prove that $\mathbb{P}_{\mathbf{b}}\left(H_W(U)=H\right)>0$. We have already seen the following lower bound:

$$\mathbb{P}_{\mathbf{b}}\left(H_W(U)=H\right)\geq\frac{1}{n-v}\sum_{0\leq\mathbf{k}\leq K}\beta_{\mathbf{k}}Y_{\mathbf{k}}\big|_{\mathbf{b}}.$$

From concentration, we can write

$$\sum_{\mathbf{k}}\beta_{\mathbf{k}}Y_{\mathbf{k}}\big|_{\mathbf{b}}=\sum_{\mathbf{k}}\beta_{\mathbf{k}}\mathbb{E}_{\mathbf{a}}\left(Y_{\mathbf{k}}\right)+O\left(\frac{1}{(\ln n)^{1000}}\sum_{\mathbf{k}}\left|\beta_{\mathbf{k}}\right|\mathbb{E}_{\mathbf{a}}\left(Y_{\mathbf{k}}\right)\right).$$

From Lemma 6.12 we can further write

$$\frac{1}{n-v}\sum_{\mathbf{k}}\beta_{\mathbf{k}}Y_{\mathbf{k}}\big|_{\mathbf{b}}=\sum_{\mathbf{k}}\beta_{\mathbf{k}}\mathbb{E}_{\mathbf{a}}\left((\mathbf{Z})_{\mathbf{k}}\right)+O\left(\frac{1}{(\ln n)^{1000}}\sum_{\mathbf{k}}\left|\beta_{\mathbf{k}}\right|\mathbb{E}_{\mathbf{a}}\left((\mathbf{Z})_{\mathbf{k}}\right)\right). \quad (8)$$

Let us separately handle the main term and the error term.

Start with the main term. Recall the definition of the coefficients $\beta_{\mathbf{k}}$, from which

$$\sum_{\mathbf{k}}\beta_{\mathbf{k}}\mathbb{E}_{\mathbf{a}}\left((\mathbf{Z})_{\mathbf{k}}\right)=\mathbb{E}_{\mathbf{a}}\left(\pi^{-}\left(\mathbf{Z}\right)\right).$$

Here the polynomial $\pi^{-}(\mathbf{x})$ is adjusted to the values of a binary vector $\mathbf{z}\in\{0,1\}^t$, which is determined by the graph structure H. From the flexible moments method,

$$\mathbb{E}_{\mathbf{a}}\left(\pi^{-}(\mathbf{Z})\right)\leq\mathbb{P}_{\mathbf{a}}\left(\mathbf{Z}=\mathbf{z}\right)\leq\mathbb{E}_{\mathbf{a}}\left(\pi^{+}(\mathbf{Z})\right).$$

The variables $Z^{(i)}$ count extensions over the triplets $(\mathbf{u}^{(i)},w_0)$, therefore the configuration $\mathbf{u}=\mathbf{a}$ fixes the position of all the vertices from these triplets, except w_0. This is easy to handle:

- The probability that w_0 lands in a position which keeps $U\cup\{w_0\}$ strongly well-spaced and q_0-feasible is bounded from below by a constant $p_0>0$.
- Given a nicely positioned configuration of $U\cup\{w_0\}$, which we denote $\mathbf{u}'=\mathbf{a}'$, we can apply the results of Section 5 for $\mathbf{Z}$. In particular, from Theorem 5.12 we claim that there is an independent constant $c>0$ such that $\mathbb{P}_{\mathbf{a}'}\left(\mathbf{Z}=\mathbf{z}\right)>c^t$, and from the flexible moments method we claim that

  $$\left|\mathbb{E}_{\mathbf{a}'}\left(\pi^{+}(\mathbf{Z})\right)-\mathbb{E}_{\mathbf{a}'}\left(\pi^{-}(\mathbf{Z})\right)\right|=\frac{e^{O(K)}}{K!}.$$

  So overall $\mathbb{E}_{\mathbf{a}'}\left(\pi^{-}(\mathbf{Z})\right)\geq c^t-\frac{e^{O(K)}}{K!}=(1-o(1))c^t$.
- Now, integrate over all possible positions of w_0 which keep $U\cup\{w_0\}$ strongly well-spaced and q_0-feasible, and obtain the lower bound

  $$\mathbb{E}_{\mathbf{a}}\left(\pi^{-}(\mathbf{Z})\right)\geq(p_0-o(1))c^t. \quad (9)$$

Now for the error term. We can similarly use Theorem 5.9 and Proposition 5.11 to claim that

$$\mathbb{E}_{\mathbf{a}}\left((\mathbf{Z})_{\mathbf{k}}\right) \leq \lambda_{\max}^{k} + n^{-\Omega(1)}$$

(where $k = k_1 + \cdots + k_t$). Combining with the bounds $|\beta_{\mathbf{k}}| \leq 1$, we get

$$O\left(\frac{1}{(\ln n)^{1000}} \sum_{\mathbf{k}} |\beta_{\mathbf{k}}| \, \mathbb{E}_{\mathbf{a}}\left((\mathbf{Z})_{\mathbf{k}}\right)\right) = O\left(\frac{K^t \lambda_{\max}^{Kt}}{(\ln n)^{1000}}\right).$$

This is definitely $o(c^t)$ for any constant c (recall that $K = \ln\ln\ln n$ and that $t = \binom{v}{2}$, $v \leq \ln\ln\ln\ln n$). Finally, substituting back in Equation (8), we get

$$\begin{aligned}\mathbb{P}_{\mathbf{b}}\left(H_W(U) = H\right) &\geq \frac{1}{n-v} \sum_{0 \leq \mathbf{k} \leq K} \beta_{\mathbf{k}} Y_{\mathbf{k}}\big|_{\mathbf{b}} \\ &\geq (p_0 - o(1))c^t > 0\end{aligned}$$

which is what we wanted. ■

Adjustments for the First Part

Let us explain how to adjust the above proof to the first part of Theorem 6.3. To find a triplet $\mathbf{x}$ with $|U_{\mathbf{x}}| = v$, we arbitrarily fix a pair of vertices $\{x_1, x_2\}$ and look for x_3 such that the number of witnesses for S over x_1, x_2, x_3 is exactly v. Here $\{x_1, x_2\}$ functions like U from before (but now there is only one couple instead of t) and x_3 functions like w. We follow the exact same strategy with the straightforward adjustments. For example:

- A "good" complete configuration is now one which determines the existence of a suitable x_3.
- Instead of t variables $X^{(1)}, \ldots, X^{(t)}$ we have one variable X. We are now interested in showing $\mathbb{P}_{\mathbf{b}}(X = v) > 0$.
- The flexible moments method is now applied on a single variable, so it simply reduces to Waring's Theorem. To approximate the moments of X, we now introduce the variables Y_k for $0 \leq k \leq K$.
- Similarly, instead of t variables $Z^{(1)}, \ldots, Z^{(t)}$ we have one variable Z, for which the results of Section 5 apply.
- In the analogue of Equation (9), the lower bound now takes the form

$$\mathbb{E}_{\mathbf{a}}\left(\pi^{-}(Z)\right) \geq (p_0 - o(1)) \frac{\lambda_{\min}^{v}}{v!}.$$

It remains to explain how to promise the existence of $U_{\mathbf{x}}$ which not only has exactly v vertices, but is also nicely positioned. This is actually quite easy given what we know so far.

Recall that "nicely positioned" requires well-spacedness and diameter $\leq \frac{r}{4}$. For the second requirement, we can just restrict ourselves to x_1, x_2, x_3 such that the corresponding locus R from Definition 5.2 satisfies $\operatorname{diam}(R) \leq \frac{r}{4}$. Indeed, the vertices of $U_{\mathbf{x}}$ are the z-vertices of a witness $\mathbf{q}$ for $S(x_1, x_2, x_3; \mathbf{q})$, so by definition they must land in R. This requirement only adds a constant factor (the requirement about x_1, x_2, x_3 forces them to an area of $\Theta(1)$) and other than that the proof remains the same.

Regarding well-spacedness, the idea is that situations in which $U_{\mathbf{x}}$ is not well-spaced are negligible. This is a consequence of the Negligibility Theorem 5.7. Here is a more detailed explanation. Suppose that x_1, x_2 are fixed, and say that x_3 is "forbidden" if $U_{\mathbf{x}}$ is not well-spaced. Replace the variable X with a variable X' which is the same, except it ignores any situation in which x_3 is forbidden. Just as we have

$$\mathbb{E}_{\mathbf{b}}((X)_k) = \frac{1}{n-2} Y_k\big|_{\mathbf{b}},\ \mathbb{E}_{\mathbf{a}}\left(Y_k\right) = (n-2)\mathbb{E}_{\mathbf{a}}\left((Z)_k\right),$$

we can also write

$$\mathbb{E}_{\mathbf{b}}((X')_k) = \frac{1}{n-2} Y'_k\big|_{\mathbf{b}},\ \mathbb{E}_{\mathbf{a}}\left(Y'_k\right) = (n-2)\mathbb{E}_{\mathbf{a}}\left((Z')_k\right)$$

where Y'_k, Z' are the same as Y_k, Z, except they ignore any forbidden x_3. Now,

$$\left|\frac{1}{n-2}\mathbb{E}_{\mathbf{a}}\left(Y_k\right) - \frac{1}{n-2}\mathbb{E}_{\mathbf{a}}\left(Y'_k\right)\right| = \left|\mathbb{E}_{\mathbf{a}}\left((Z)_k\right) - \mathbb{E}_{\mathbf{a}}\left((Z')_k\right)\right| = n^{-\Omega(1)}$$

where the last bound follows from the Negligibility Theorem. By Markov's inequality we deduce[7] that a.a.s.

$$\left|\frac{1}{n-2} Y_k - \frac{1}{n-2} Y'_k\right| = n^{-\Omega(1)}.$$

This means that a.a.s., the lower bound on $\mathbb{P}_{\mathbf{b}}(X = v)$ is also valid for $\mathbb{P}_{\mathbf{b}}(X' = v)$, except for an error term of $n^{-\Omega(1)}$. So $\mathbb{P}_{\mathbf{b}}(X' = v)$ is also positive and there exists a suitable, non-forbidden x_3.

Remark 6.14. By "playing" with the allowed positions of x_1, x_2, x_3, the same considerations can promise the existence of well-spaced sets $U_{\mathbf{x}}$ with exactly v vertices in any specific region of $\mathbb{T}^2$. In particular, for every constant $\varepsilon > 0$ we can make the following claim. A.a.s., for every ball $B \subseteq \mathbb{T}^2$ with radius ε, there exist x_1, x_2, x_3 such that $U_{\mathbf{x}}$ is well-spaced, has exactly v vertices and is contained in B. We will use this fact in Section 7.

[7] It is worth noting that here we lose the strong bound of $\exp\left(-\Omega(\ln^2 n)\right)$ to the weaker $n^{-\Omega(1)}$. This is not a problem, as we do not need the strong bound in the first part — there is no union bound to overcome.

6.2 Proving the Concentration Result

We turn to a proof of the Concentration Theorem 6.13. It is divided into three steps. The preparation step applies Poissonization and replaces $Y_{\mathbf{k}}$ with a similar variable which can be written as the sum of "nearly independent" indicators. The remaining two steps separately prove the lower bound and the upper bound, which together compose the Concentration Theorem.

Following the previous subsection, we fix $1 \le v \le \ln\ln\ln\ln n$, and also fix a set U of v vertices and a configuration $\mathbf{u} = \mathbf{a}$ which is nicely positioned (see Definition 6.2). $\mathbf{u}^{(1)}, \mathbf{u}^{(2)}, \ldots, \mathbf{u}^{(t)}$ is an enumeration of all the pairs of vertices from U, with $t = \binom{v}{2}$.

6.2.1 Preparation

Our first step is to apply the Poissonization technique (see Subsection A.1 in the Appendix). We consider U and its configuration $\mathbf{u} = \mathbf{a}$ fixed, and then, instead of randomly generating $n - v$ additional vertices, we introduce a new random graph which generates the additional vertices as a Poisson point process with intensity $n - v$. Following the notation from the Appendix, this means we replace $G_{\mathbb{T}^2}(n, r)$ with $G_{\mathbf{a}}(N, r)$.

We intend to apply Corollary A.5, in order to reduce the Concentration Theorem in $G_{\mathbb{T}^2}(n, r)$ to the analogous result in $G_{\mathbf{a}}(N, r)$.

Definition 6.15. For every $0 \le \mathbf{k} \le K$, define the variable $\tilde{Y}_{\mathbf{k}}$ as the number of tuples

$$\left(w, \mathbf{q}_1^{(1)}, \ldots, \mathbf{q}_{k_1}^{(1)}, \ldots, \mathbf{q}_1^{(t)}, \ldots, \mathbf{q}_{k_t}^{(t)}\right)$$

in $G_{\mathbf{a}}(N, r)$ such that w is a vertex and $\mathbf{q}_i^{(i)}, \ldots, \mathbf{q}_{k_i}^{(i)}$ are distinct witnesses for the S-extension over $(\mathbf{u}^{(i)}, w)$, for every $1 \le i \le t$.

That is, $\tilde{Y}_{\mathbf{k}}$ is defined just like $Y_{\mathbf{k}}$, but in the Poisson random graph.

Proposition 6.16. *Suppose that in $G_{\mathbf{a}}(N, r)$, $\tilde{Y}_{\mathbf{k}}$ is concentrated for every $0 \le \mathbf{k} \le K$:*

$$\mathbb{P}\left(\left|\tilde{Y}_{\mathbf{k}} - \mathbb{E}\tilde{Y}_{\mathbf{k}}\right| = \Omega\left(\frac{\mathbb{E}\tilde{Y}_{\mathbf{k}}}{(\ln n)^{1000}}\right)\right) = \exp\left(-\Omega(\ln^2 n)\right).$$

Then in $G_{\mathbb{T}^2}(n, r)$ (given $\mathbf{u} = \mathbf{a}$), $Y_{\mathbf{k}}$ is similarly concentrated for every $0 \le \mathbf{k} \le K$:

$$\mathbb{P}_{\mathbf{a}}\left(\left|Y_{\mathbf{k}} - \mathbb{E}_{\mathbf{a}}\left(Y_{\mathbf{k}}\right)\right| = \Omega\left(\frac{\mathbb{E}_{\mathbf{a}}\left(Y_{\mathbf{k}}\right)}{(\ln n)^{1000}}\right)\right) = \exp\left(-\Omega(\ln^2 n)\right).$$

Proof. Corollary A.5 from the Appendix immediately implies

$$\mathbb{P}_{\mathbf{a}}\left(\left|Y_{\mathbf{k}} - \mathbb{E}\tilde{Y}_{\mathbf{k}}\right| = \Omega\left(\frac{\mathbb{E}\tilde{Y}_{\mathbf{k}}}{(\ln n)^{1000}}\right)\right) = \exp\left(-\Omega(\ln^2 n)\right). \tag{10}$$

That is, $Y_{\mathbf{k}}$ is concentrated around $\mathbb{E}\tilde{Y}_{\mathbf{k}}$. To replace $\mathbb{E}\tilde{Y}_{\mathbf{k}}$ with $\mathbb{E}_{\mathbf{a}}\left(Y_{\mathbf{k}}\right)$ we show that the two expected values are "very close". Let E be the event (in $G_{\mathbb{T}^2}(n,r)$) whose probability is bounded in equation (10). Write

$$\mathbb{E}_{\mathbf{a}}\left(Y_k\right) = \mathbb{E}_{\mathbf{a}}\left(Y_k \mathbb{1}_{E^c}\right) + \mathbb{E}_{\mathbf{a}}\left(Y_k \mathbb{1}_{E}\right).$$

For the first term, by definition of E, we can write (for some positive constant C)

$$\mathbb{P}_{\mathbf{a}}(E^c)\left(1 - \frac{C}{(\ln n)^{1000}}\right)\mathbb{E}\tilde{Y}_k \leq \mathbb{E}_{\mathbf{a}}\left(Y_k \mathbb{1}_{E^c}\right) \leq \mathbb{P}_{\mathbf{a}}(E^c)\left(1 + \frac{C}{(\ln n)^{1000}}\right)\mathbb{E}\tilde{Y}_k.$$

For the second term, recall that $Y_{\mathbf{k}}$ counts $(4k+1)$-tuples of vertices in $G_{\mathbb{T}^2}(n,r)$, therefore we always have $Y_{\mathbf{k}} \leq n^{4k+1}$, so

$$\mathbb{E}_{\mathbf{a}}\left(Y_k \mathbb{1}_{E}\right) \leq n^{4k+1}\mathbb{P}_{\mathbf{a}}(E).$$

Using the estimation $\mathbb{P}_{\mathbf{a}}(E) = \exp\left(-\Omega(\ln^2 n)\right)$, the fact that $k \leq tK$ and the standard bounds on t, K, we can overall write

$$\mathbb{E}_{\mathbf{a}}\left(Y_{\mathbf{k}}\right) = \left(1 + O\left(\frac{1}{(\ln n)^{1000}}\right)\right)\mathbb{E}\tilde{Y}_k. \tag{11}$$

The desired result easily follows. ■

Following last the proposition, we now focus on proving concentration of $\tilde{Y}_{\mathbf{k}}$ in the Poisson random graph $G_{\mathbf{a}}(N,r)$. Since from now on we are interested only in the variables $\tilde{Y}_{\mathbf{k}}$, for notational simplicity we rename them as $Y_{\mathbf{k}}$. So, from now on, $Y_{\mathbf{k}}$ actually denotes the variable in $G_{\mathbf{a}}(N,r)$ from Definition 6.15.

The next step is another standard technique: *discretization* of the random geometric graph. This means we divide the torus $\mathbb{T}^2$ into "very small" squares. For our purposes, "very small" will be $\exp\left(-\ln^2 n\right)$.

Definition 6.17. Let $M = \left\lceil \exp\left(\ln^2 n\right)\right\rceil$ and denote $\varepsilon = \frac{1}{M}$. Divide the torus $\mathbb{T}^2$ into M^2 squares of size $\varepsilon \times \varepsilon$ and denote them $\square_{ij}$ (with indices $1 \leq i,j \leq M$). Note that they are all mutually disjoint (perhaps except the sides, but their measure is 0).

Proposition 6.18. *With probability* $1 - \exp\left(-\Omega(\ln^2 n)\right)$*, each of the* M^2 *squares contains at most one vertex of* $G_{\mathbf{a}}(N,r)$*.*

Proof. First, consider a square $\square$ which contains a vertex from U. It does not contain any other vertex from U (because U is well-spaced). The number of non-U vertices it contains is distributed like $\mathrm{Pois}(\lambda)$ for $\lambda = (n-v)\varepsilon^2$. A simple asymptotic estimation shows that $\square$ contains non-U vertices with probability $O(\lambda) = O(n\varepsilon^2)$. A union bound over the v relevant squares yields

$$O(vn\varepsilon^2) = \exp\left(-\Omega(\ln^2 n)\right).$$

Now consider all the remaining squares. For a square $\square$, the probability that it contains two non-U vertices is $O(\lambda^2)$. A union bound over all $M^2 - v$ remaining squares yields

$$O(M^2) \cdot O\left(n^2\varepsilon^4\right) = O\left(n^2\varepsilon^2\right) = \exp\left(-\Omega(\ln^2 n)\right).$$

That finishes the proof. ∎

The intuitive meaning of Proposition 6.18 is that, given our discretization, we can "identify" vertices with the squares that contain them without ambiguity. Now, the idea is to discretize $Y_{\mathbf{k}}$ accordingly, by counting squares instead of vertices. This yields an approximating random variable $Y_{\mathbf{k}}^{[\varepsilon]}$. Its important advantage over $Y_{\mathbf{k}}$ is that, while $Y_{\mathbf{k}}$ counts witnesses by going over the *vertices*, $Y_{\mathbf{k}}^{[\varepsilon]}$ counts witnesses by going over *regions* of $\mathbb{T}^2$. The spatial independence property then allows the decomposition of $Y_{\mathbf{k}}^{[\varepsilon]}$ into a sum of "almost independent" indicators, which is the ideal form for concentration results.

Definition 6.19. The discretized S-extension is a relation between seven squares, which we denote

$$S^{[\varepsilon]}\left(\square_{x_1}, \square_{x_2}, \square_{x_3}; \square_{s_1}, \square_{s_2}, \square_{s_3}, \square_z\right).$$

and define as follows. First, $S^{[\varepsilon]}$ requires the that each of the seven squares contains a vertex. Second, it requires

$$\begin{aligned} &\mathrm{UN}^{[\varepsilon]}\left(\square_{x_1}, \square_{s_1}\right) \wedge \mathrm{UN}^{[\varepsilon]}\left(\square_{x_2}, \square_{s_2}\right) \wedge \mathrm{UN}^{[\varepsilon]}\left(\square_{x_3}, \square_{s_3}\right) \\ \wedge &\mathrm{UN}^{[\varepsilon]}\left(\square_{s_1}, \square_{z}\right) \wedge \mathrm{UN}^{[\varepsilon]}\left(\square_{s_2}, \square_{z}\right) \wedge \mathrm{UN}^{[\varepsilon]}\left(\square_{s_3}, \square_{z}\right). \end{aligned}$$

Here $\mathrm{UN}^{[\varepsilon]}\left(\square_x, \square_y\right)$ is the discretized UN-relation. It requires that the distance between the centers b_x, b_y of the squares is between $2r - 2h_n$ and $2r$, and also that the lens $L_{b_x b_y}$ contains exactly one vertex. Third, it requires

$$\mathrm{CD}^{[\varepsilon]}\left(\square_{x_1}, \square_z\right) \wedge \mathrm{CD}^{[\varepsilon]}\left(\square_{x_2}, \square_z\right) \wedge \mathrm{CD}^{[\varepsilon]}\left(\square_{x_3}, \square_z\right).$$

Here $\mathrm{CD}^{[\varepsilon]}\left(\square_x, \square_z\right)$ is the discretized CD-relation. It requires that the distance between the centers b_x, b_z is between r and $\sqrt{3}r$.

Again, we use abbreviated notation:

$$\square_{\mathbf{x}} = \left(\square_{x_1}, \square_{x_2}, \square_{x_3}\right), \square_{\mathbf{q}} = \left(\square_{s_1}, \square_{s_2}, \square_{s_3}, \square_z\right)$$

and write $S^{[\varepsilon]}\left(\square_{\mathbf{x}}; \square_{\mathbf{q}}\right)$. When this extension holds we say that $\square_{\mathbf{q}}$ is a *witness* for $S^{[\varepsilon]}$ over $\square_{\mathbf{x}}$.

Definition 6.20. For every $0 \leq \mathbf{k} \leq K$, define $Y_{\mathbf{k}}^{[\varepsilon]}$ to be the random variable which counts the number of tuples

$$\left(\square_w, \square_{\mathbf{q}_1^{(1)}}, \ldots, \square_{\mathbf{q}_{k_1}^{(1)}}, \ldots, \square_{\mathbf{q}_1^{(t)}}, \ldots, \square_{\mathbf{q}_{k_t}^{(t)}}\right)$$

such that $\square_w$ is a square and $\square_{\mathbf{q}_1^{(i)}}, \ldots, \square_{\mathbf{q}_{k_i}^{(i)}}$ are distinct witnesses for the $S^{[\varepsilon]}$-extension over $(\square_{\mathbf{u}^{(i)}}, \square_w)$ for every $1 \le i \le t$. Here $\square_{\mathbf{u}^{(i)}}$ is the pair of squares which contain the pair of vertices $\mathbf{u}^{(i)}$.

We employ the standard notation $k = k_1 + \cdots + k_t$. While $Y_{\mathbf{k}}$ counts $(4k+1)$-tuples of vertices, $Y_{\mathbf{k}}^{[\varepsilon]}$ counts $(4k+1)$-tuples of squares.

Lemma 6.21. *For every* $0 \le \mathbf{k} \le K$*,*

$$\mathbb{E}\left|Y_{\mathbf{k}}^{[\varepsilon]} - Y_{\mathbf{k}}\right| = \exp\left(-\Omega(\ln^2 n)\right).$$

Proof Sketch. The difference between $Y_{\mathbf{k}}^{[\varepsilon]}$ and $Y_{\mathbf{k}}$ comes from S-extensions $S(\mathbf{x};\mathbf{q})$ which do not correspond to $S^{[\varepsilon]}$-extensions $S^{[\varepsilon]}(\square_{\mathbf{x}};\square_{\mathbf{q}})$, or vice versa. Let us understand how such a situation occurs.

Consider vertices $\mathbf{x}, \mathbf{q}$ which satisfy $S(\mathbf{x};\mathbf{q})$. Let $\square_{\mathbf{x}}, \square_{\mathbf{q}}$ denote the corresponding squares which contain them. What would prevent them from satisfying $S^{[\varepsilon]}(\square_{\mathbf{x}};\square_{\mathbf{q}})$?

One possibility is that $\mathrm{CD}(x_1, z)$ holds but $\mathrm{CD}^{[\varepsilon]}(\square_{x_1}, \square_z)$ does not hold. Let b_{x_1}, b_z be the centers of the respective squares. The distances $\|b_{x_1} - x_1\|, \|b_z - z\|$ are at most $\sqrt{2}\varepsilon$. Therefore if

$$\|x_1 - z\| \in \left[r, \sqrt{3}r\right] \text{ but } \|b_{x_1} - b_z\| \notin \left[r, \sqrt{3}r\right],$$

this forces x_1 to be in an $O(\varepsilon)$-wide annulus around z. Another possibility is when $L_{x_1 s_1}$ contains a vertex but $L_{b_{x_1} b_{s_1}}$ does not contain a vertex. Again, $\|b_{x_1} - x_1\|, \|b_{s_1} - s_1\| \le \sqrt{2}\varepsilon$, thus the difference $L_{x_1 s_1} \setminus L_{b_{x_1} b_{s_1}}$, which is supposed to contain a vertex, has an area of $O(\varepsilon)$.

We see that every possibility forces a requirement which adds a factor of $O(\varepsilon)$ to the probability of the extension. Our situation highly resembles the Negligibility Theorem, only here the added factor is $O(\varepsilon)$ instead of $n^{-\Omega(1)}$. The same considerations allow us to bound the contribution to the expected value by $O(\varepsilon) = \exp\left(-\Omega(\ln^2 n)\right)$. See Proposition A.11 as illustration of those the considerations. ■

The immediate conclusion from the lemma is that it suffices to prove concentration for $Y_{\mathbf{k}}^{[\varepsilon]}$ instead of $Y_{\mathbf{k}}$. Indeed, Markov's inequality yields $\mathbb{P}(Y_{\mathbf{k}}^{[\varepsilon]} \ne Y_{\mathbf{k}}) = \exp\left(-\Omega(\ln^2 n)\right)$, so the replacement of $Y_{\mathbf{k}}^{[\varepsilon]}$ with $Y_{\mathbf{k}}$ keeps the same asymptotic bounds.

In conclusion, the preparation step reduces the original Concentration Theorem to the following result: for every $0 \le \mathbf{k} \le K$,

$$\mathbb{P}\left(\left|Y_{\mathbf{k}}^{[\varepsilon]} - \mathbb{E}Y_{\mathbf{k}}^{[\varepsilon]}\right| = \Omega\left(\frac{\mathbb{E}Y_{\mathbf{k}}^{[\varepsilon]}}{(\ln n)^{1000}}\right)\right) = \exp\left(-\Omega(\ln^2 n)\right). \tag{12}$$

The remaining two steps prove Equation (12) by separately dealing with the lower bound and the upper bound.

6.2.2 The Lower Bound

In this step we prove

$$\mathbb{P}\left(Y_{\mathbf{k}}^{[\varepsilon]} \leq \left(1 - \frac{C}{(\ln n)^{1000}}\right)\mathbb{E}Y_{\mathbf{k}}^{[\varepsilon]}\right) = \exp\left(-\Omega(\ln^2 n)\right)$$

for every $0 \leq \mathbf{k} \leq K$. Actually, our proof will give us a much stronger bound on the probability: $\exp\left(-n^{\Omega(1)}\right)$.

Our strategy is to use Theorem 10 from Janson [8]. Its formulation is repeated here.

Theorem 6.22. *Let $\{I_i\}_{i\in\mathcal{I}}$ be a finite family of indicator random variables (over the same probability space) and let $S = \sum_{i\in\mathcal{I}} I_i$. Let Γ be a dependency graph*[8] *for $\{I_i\}_{i\in\mathcal{I}}$. Denote $i \sim j$ for $i, j \in \mathcal{I}$ if they are adjacent in Γ. Define:*

- $p_i = \mathbb{E}(I_i) = \mathbb{P}(I_i = 1)$.
- $\mu = \mathbb{E}(S) = \sum_{i\in\mathcal{I}} p_i$.
- $\delta_i = \sum_{j:j\sim i} p_j$ *and* $\delta = \max_{i\in\mathcal{I}} \delta_i$.
- $\Delta = \sum_{\{i,j\}:i\sim j} \mathbb{E}(I_i I_j)$. *Note that this is a sum over unordered pairs.*

Then, for every $0 \leq a \leq 1$,

$$\mathbb{P}\left(S \leq a\mu\right) \leq \exp\left(-\min\left\{(1-a)^2\frac{\mu^2}{8\Delta + 2\mu}, (1-a)\frac{\mu}{6\delta}\right\}\right). \qquad (13)$$

To apply the theorem on $Y_{\mathbf{k}}^{[\varepsilon]}$, we decompose it into a sum of indicators as follows. Let the index set $\mathcal{I}$ be the set of all $(4k+1)$-tuples of squares

$$T = \left(\square_w, \square_{\mathbf{q}_1^{(1)}}, \ldots, \square_{\mathbf{q}_{k_1}^{(1)}}, \ldots, \square_{\mathbf{q}_1^{(t)}}, \ldots, \square_{\mathbf{q}_{k_t}^{(t)}}\right).$$

For every tuple $T \in \mathcal{I}$, define the indicator $\mathbb{1}_T$ to be 1 when $\square_{\mathbf{q}_1^{(i)}}, \ldots, \square_{\mathbf{q}_{k_i}^{(i)}}$ are distinct witnesses for the $S^{[\varepsilon]}$-extension over $(\square_{\mathbf{u}^{(i)}}, \square_w)$ for every $1 \leq i \leq t$. Then $Y_{\mathbf{k}}^{[\varepsilon]} = \sum_{T\in\mathcal{I}} \mathbb{1}_T$.

Recall that the $S^{[\varepsilon]}$-extension is composed of several requirements. We can divide them into two kinds: *distance requirements* about the distances between squares, and *vertex requirements* about the existence of vertices in certain loci (either one of the squares or one of the lenses). Given a tuple of squares T, the distance requirements of all the $S^{[\varepsilon]}$-extensions are already determined by T itself, so $\mathbb{1}_T$ only depends on the vertex requirements. Let us say that a tuple T

[8] This means a graph with vertex set $\mathcal{I}$ such that if A and B are two disjoint subsets of $\mathcal{I}$ and Γ contains no edge between A and B then $\{I_i\}_{i\in A}$ and $\{I_i\}_{i\in B}$ are independent.

is *valid* if it satisfies all the relevant distance requirements. For non-valid tuples T the indicator $\mathbb{1}_T$ is trivially 0, so we may ignore them and sum over the valid tuples.

Let us fix a valid T and examine $\mathbb{1}_T$ more closely. Denote by $L_1, L_2, \ldots, L_{6k}$ the $6k$ lenses which are relevant for the $S^{[\varepsilon]}$-extensions. Each L_i is a fixed locus in $\mathbb{T}^2$. Then $\mathbb{1}_T$ equals 1 if and only if every square of T contains a vertex, and every lens L_i contains exactly one vertex. We can therefore define the dependency graph as follows. $T \sim T'$ for two valid tuples T, T' if one of the squares of T or its relevant lenses intersects one of the squares of T' or its relevant lenses. The spatial independence property (Proposition A.6) shows that this is indeed a dependency graph on $\mathcal{I}$.

Let us apply Theorem 6.22 on $Y_{\mathbf{k}}^{[\varepsilon]}$ with this dependency graph. We set

$$a = 1 - \frac{C}{(\ln n)^{1000}}$$

so the LHS of (13) becomes

$$\mathbb{P}\left(Y_{\mathbf{k}}^{[\varepsilon]} \leq \left(1 - \frac{C}{(\ln n)^{1000}}\right)\mathbb{E}Y_{\mathbf{k}}^{[\varepsilon]}\right).$$

For the RHS, we shall prove the following asymptotic estimations on μ, δ, Δ:

1. $\mu = n^{1+o(1)}$.
2. $\delta = n^{1-\Omega(1)}$.
3. $\Delta = n^{2-\Omega(2)}$.

Substituting these estimations in equation (13), we get

$$\frac{\mu^2}{8\Delta + 2\mu} = \frac{n^{2-o(1)}}{n^{2-\Omega(1)} + n^{1-\Omega(1)}} = n^{\Omega(1)},$$
$$\frac{\mu}{6\delta} = \frac{n^{1-o(1)}}{n^{1-\Omega(1)}} = n^{\Omega(1)}.$$

In addition, $1 - a = \frac{1}{(\ln n)^{1000}} = n^{-o(1)}$. Overall

$$\min\left\{(1-a)^2\frac{\mu^2}{8\Delta + 2\mu}, (1-a)\frac{\mu}{6\delta}\right\} = n^{\Omega(1)}$$

and the RHS is indeed $\exp\left(-n^{\Omega(1)}\right)$.

It therefore remains to prove the asymptotic estimations on μ, δ, Δ.

Estimating μ. This is straightforward:

$$\mu = \mathbb{E}\left(Y_{\mathbf{k}}^{[\varepsilon]}\right) = (1 + o(1))\mathbb{E}\left(Y_{\mathbf{k}}\right) = (1 + o(1))(n - v)\mathbb{E}_{\mathbf{a}}\left((\mathbf{Z})_{\mathbf{k}}\right).$$

Here $\mathbb{E}_{\mathbf{a}}\left((\mathbf{Z})_{\mathbf{k}}\right)$ a joint moment just like in Section 5, and we know that it is $e^{\Theta(k)} = n^{o(1)}$, so overall $\mu = n^{1+o(1)}$.

Estimating δ. Recall that $\delta = \max_{T\in\mathcal{I}} \delta_T$, where

$$\delta_T = \sum_{T':T'\sim T} \mathbb{E}\left(\mathbb{1}_{T'}\right).$$

That is, δ_T is the expected number of tuples T' which satisfy the k relevant $S^{[\varepsilon]}$-relations and also $T' \sim T$. Without the requirement $T' \sim T$, this sum simply becomes $\mu = \mathbb{E}\left(Y_{\mathbf{k}}^{[\varepsilon]}\right)$ which is $n^{1+o(1)}$. So we need to show that the additional requirement $T' \sim T$ contributes a factor of $n^{-\Omega(1)}$ to the expectation. This leads us once more to the considerations of the Negligibility Theorem. They apply in the exact same way, except now we handle tuples of squares instead of tuples of vertices.

For the sake of concreteness, let us demonstrate the concept with a single quadruplet. Fix a triplet of squares $\square_{\mathbf{x}}$ and let us bound the expected number of quadruplets $\square_{\mathbf{q}}$ which satisfy $S^{[\varepsilon]}\left(\square_{\mathbf{x}};\square_{\mathbf{q}}\right)$. The main step is to count the number of quadruplets $\square_{\mathbf{q}}$ which satisfy the relevant distance requirements.

- $\square_z$ has $O(M^2)$ options to be chosen with "comfortable distances" from $\square_{x_1}, \square_{x_2}, \square_{x_3}$.
- $\square_{s_1}$ has $O(M^2 h_n^2)$ options to be chosen with the correct distances from $\square_{x_1}, \square_z$. Indeed, its center must lie in the intersection of two annuli with radii $2r - 2h_n, 2r$, whose centers are "comfortable distances" apart, and the area of such an intersection is $O(h_n^2)$ (see Lemma A.13).
- Each of $\square_{s_2}, \square_{s_3}$ also has $O(M^2 h_n^2)$ valid options for the same reasons.

So overall we count $O\left((M^2)^4 h_n^6\right)$ quadruplets. For each one of these quadruplets, the probability that it also satisfies the vertex requirements can be simply bounded by the probability that each of the four squares contain a vertex, which is $O\left((n\varepsilon^2)^4\right)$ (we allow ourselves to ignore the lenses here). In conclusion, the expected number of quadruplets $\square_{\mathbf{q}}$ with $S^{[\varepsilon]}\left(\square_{\mathbf{x}};\square_{\mathbf{q}}\right)$ bounded by

$$O\left((M^2)^4 h_n^6\right) \cdot O\left((n\varepsilon^2)^4\right) = O\left(n^4 h_n^6\right) = (\ln n)^{O(1)}.$$

Now suppose that we fix a lens L with width $\Theta(n^{-2/3})$ "in the background", and ask for the expected number of quadruplets $\square_{\mathbf{q}}$ which satisfy the extension $S^{[\varepsilon]}\left(\square_{\mathbf{x}};\square_{\mathbf{q}}\right)$, and also that one of its six relevant lenses intersects L. Here L plays the role of a lens from the fixed tuple T. Geometric considerations show that this additional requirement forces $\square_z$ to be in a locus with area $n^{-\Omega(1)}$ (as explained in detail in Subsection A.2). So now we count quadruplets $O\left((M^2)^4 h_n^6 \cdot n^{-\Omega(1)}\right)$ quadruplets, with the same bound $O\left((n\varepsilon^2)^4\right)$ on the probability for each one, which yields

$$(\ln n)^{O(1)} \cdot n^{-\Omega(1)} = n^{-\Omega(1)}.$$

To handle the general case, with multiple quadruplets, we consider the factor added by each quadruplet separately; again this idea is explained in detail in Subsection A.2. Now, the count of tuples T' which satisfy the distance requirements and $T' \sim T$ is

$$O\left((M^2)^{4k+1} h_n^{6k} \cdot n^{-\Omega(1)}\right)$$

and for every such T', we have $\mathbb{E}\left(\mathbb{1}_{T'}\right) = O\left((n\varepsilon^2)^{4k+1}\right)$. So overall

$$\begin{aligned}
\delta_T &= O\left((M^2)^{4k+1} h_n^{6k} \cdot n^{-\Omega(1)}\right) \cdot O\left((n\varepsilon^2)^{4k+1}\right) \\
&= O\left((\ln n)^{O(4k)} n^{-4k} \cdot n^{4k+1} \cdot n^{-\Omega(1)}\right) \\
&= n^{1-\Omega(1)}.
\end{aligned}$$

Estimating Δ. Recall that

$$\Delta = \sum_{\{T,T'\}:T\sim T'} \mathbb{E}(\mathbb{1}_T \mathbb{1}_{T'})$$

(summing over unordered pair of tuples). Rewrite it as

$$\Delta = \frac{1}{2} \sum_T \left[\sum_{T':T'\sim T} \mathbb{E}(\mathbb{1}_T \mathbb{1}_{T'}) \right].$$

Let us again fix a valid T and estimate $\Delta_T = \sum_{T':T'\sim T} \mathbb{E}(\mathbb{1}_T \mathbb{1}_{T'})$. This sum is exactly like δ_T, but with $\mathbb{E}(\mathbb{1}_{T'})$ replaced by $\mathbb{E}(\mathbb{1}_T \mathbb{1}_{T'})$.

We begin by separately handling tuples T' which share squares with T. Their number is $\mathrm{PL} \cdot \left(M^2\right)^{4k}$ (here PL stands for a poly-logarithmic factor; it is polynomial in K, t), and each summand is

$$\mathbb{E}(\mathbb{1}_T \mathbb{1}_{T'}) \le \mathbb{E}(\mathbb{1}_T) = O\left((n\varepsilon^2)^{4k+1}\right).$$

Therefore their contribution to Δ_T is

$$O\left(\mathrm{PL} \cdot n^{4k+1} \left(M^2\right)^{4k} \left(\varepsilon^2\right)^{4k+1}\right) = O\left(\mathrm{PL} \cdot n^{4k+1} \varepsilon^2\right) = \exp\left(-\Omega(\ln^2 n)\right).$$

Now consider tuples T' with $T' \sim T$ which do not share squares with T. As we already explained, counting the number of tuples T' which satisfy the distance requirements and $T' \sim T$ gives

$$O\left((M^2)^{4k+1} h_n^{6k} \cdot n^{-\Omega(1)}\right).$$

Now, $\mathbb{E}(\mathbb{1}_T \mathbb{1}_{T'})$ can be bounded by the probability that the $2(4k+1)$ squares of T and T' all contain vertices, thus it is $O\left((n\varepsilon^2)^{2(4k+1)}\right)$. Overall

$$\begin{aligned}
\Delta_T &= O\left((M^2)^{4k+1} h_n^{6k} \cdot n^{-\Omega(1)}\right) \cdot O\left((n\varepsilon^2)^{2(4k+1)}\right) \\
&= O\left(\left(\varepsilon^2\right)^{4k+1} (\ln n)^{O(4k)} n^{-4k} \cdot n^{2(4k+1)} \cdot n^{-\Omega(1)}\right) \\
&= O\left(\left(\varepsilon^2\right)^{4k+1} \cdot n^{4k+1} \cdot n^{1-\Omega(1)}\right).
\end{aligned}$$

Finally, summing over valid tuples T,

$$\begin{aligned}\Delta &= O\left((M^2)^{4k+1} h_n^{6k}\right) \cdot O\left(\left(\varepsilon^2\right)^{4k+1} \cdot n^{4k+1} \cdot n^{1-\Omega(1)}\right) \\ &= O\left((\ln n)^{O(4k)} n^{-4k} \cdot n^{4k+1} \cdot n^{1-\Omega(1)}\right) = n^{2-\Omega(1)}.\end{aligned}$$

6.2.3 The Upper Bound

In this step we prove

$$\mathbb{P}\left(Y_{\mathbf{k}}^{[\varepsilon]} \geq \left(1 + \frac{C}{(\ln n)^{1000}}\right) \mathbb{E}Y_{\mathbf{k}}^{[\varepsilon]}\right) = \exp\left(-\Omega(\ln^2 n)\right)$$

for every $0 \leq \mathbf{k} \leq K$. Our strategy is to use Theorem 2.1 from Janson and Ruciński [9]. Its formulation is repeated here.

Theorem 6.23. *Let $\{I_i\}_{i\in\mathcal{I}}$ a finite family of random variables and let $S = \sum_{i\in\mathcal{I}}$. Let Γ be a dependency graph for $\{I_i\}_{i\in\mathcal{I}}$ and denote $i \sim j$ if they are adjacent in $\mathcal{I}$. Define:*

- $\mu = \mathbb{E}(S) = \sum_{i\in\mathcal{I}} \mathbb{E}(I_i)$.
- $X_i = \sum_{j:j\sim i} I_j$.

Then, for every $\delta > 0$ and $\rho > 0$,

$$\mathbb{P}\left(S \geq (1+\delta)\mu\right) \leq \left(1 + \frac{\delta}{2}\right)^{-\rho} + \sum_{i\in\mathcal{I}} \mathbb{P}\left(X_i > \frac{\delta\mu}{2\rho}\right). \tag{14}$$

Again, we intend to use the theorem for $Y_{\mathbf{k}}^{[\varepsilon]}$, with its decomposition into indicators $Y_{\mathbf{k}}^{[\varepsilon]} = \sum_T \mathbb{1}_T$ and their dependency graph. We set $\delta = \frac{C}{(\ln n)^{1000}}$ and $\rho = n^\gamma$, for a certain $0 < \gamma < 1$ which will be specified later. The LHS of (14) becomes

$$\mathbb{P}\left(Y_{\mathbf{k}}^{[\varepsilon]} \geq \left(1 + \frac{C}{(\ln n)^{1000}}\right) \mathbb{E}Y_{\mathbf{k}}^{[\varepsilon]}\right).$$

The first summand in the RHS is

$$\left(1 + \frac{C}{2\,(\ln n)^{1000}}\right)^{-n^\gamma} = \exp\left(-\Omega\left(\frac{n^\gamma}{(\ln n)^{1000}}\right)\right) = \exp\left(-n^{\Omega(1)}\right).$$

It remains to bound the second summand, which is

$$\sum_T \mathbb{P}\left(X_T > \frac{\delta\mu}{2\rho}\right)$$

where T runs over valid $(4k+1)$-tuples of squares and $X_T = \sum_{T':T'\sim T} \mathbb{1}_{T'}$. Notice that $\mathbb{E}(X_T)$ is precisely δ_T from the lower bound. As it turns out, for an upper bound, the expected value of X_T is not sufficient; we need more direct information about the distribution of X_T. This makes our task a bit more involved.

We shall now prove that

$$\mathbb{P}\left(X_T > \frac{\delta\mu}{2\rho}\right) = \exp\left(-\Omega\left(\ln^3 n\right)\right) \tag{15}$$

for every fixed (valid) tuple of squares T (uniformly over T). That will finish the proof. Indeed, we know the number of valid tuples T: it is

$$O\left(M^{4k+1}h_n^{6k}\right) = \exp\left(O\left(\ln^2 n \cdot \ln\ln\ln\ln n\right)\right).$$

Therefore

$$\begin{aligned}\sum_T \mathbb{P}\left(X_T > \frac{\delta\mu}{2\rho}\right) &= O\left(M^{4k+1}h_n^{6k}\right)\cdot\exp\left(-\Omega\left(\ln^3 n\right)\right)\\ &= \exp\left(O\left(\ln^2 n\cdot\ln\ln\ln\ln n\right) - \Omega\left(\ln^3 n\right)\right)\\ &= \exp\left(-\Omega\left(\ln^3 n\right)\right).\end{aligned}$$

In particular, this bound is also $\exp\left(-\Omega\left(\ln^2 n\right)\right)$, which is exactly what we need.

So, let us fix a valid $(4k+1)$-tuple of squares T; our goal is to prove (15). We continue the tradition of ignoring the lenses in our bounds: let us replace the indicator $\mathbb{1}_{T'}$ with $\tilde{\mathbb{1}}_{T'}$, which only requires that each of the squares of T' contains a vertex. Define $\tilde{X}_T = \sum_{T':T'\sim T} \tilde{\mathbb{1}}_{T'}$. Obviously $\tilde{X}_T \geq X_T$ so it is sufficient to prove

$$\mathbb{P}\left(\tilde{X}_T > \frac{\delta\mu}{2\rho}\right) = \exp\left(-\Omega\left(\ln^3 n\right)\right).$$

The term $\frac{\delta\mu}{2\rho}$ can be written as follows:

$$\frac{\delta\mu}{2\rho} = \frac{Cn^{1+o(1)}}{2\left(\ln n\right)^{1000} n^{\gamma}} = n^{1-\gamma+o(1)}.$$

Let us compare it to the expectation $\tilde{\delta}_T = \mathbb{E}\tilde{X}_T$. Our bound on δ_T from the lower bound step was actually a bound on $\tilde{\delta}_T$, so we have $\tilde{\delta}_T = n^{1-\Omega(1)}$. Setting γ to be a sufficiently small constant, we get

$$\tilde{\delta}_T = n^{-\Omega(1)}\cdot\frac{\delta\mu}{2\rho}$$

so we can write

$$\mathbb{P}\left(\tilde{X}_T > \frac{\delta\mu}{2\rho}\right) \leq \mathbb{P}\left(\left|\tilde{X}_T - \tilde{\delta}_T\right| \geq n^{1-\gamma+o(1)}\right). \tag{16}$$

In words, $\tilde{X}_T > \frac{\delta\mu}{2\rho}$ implies that $\tilde{X}_T$ deviates from its expectation $\tilde{\delta}_T$ by a multiplicative factor of $n^{\Omega(1)}$.

Equation (16) reduces our task into the proof of another concentration result, this time for $\tilde{X}_T$. Our strategy is to use Kim and Vu's result about concentration of multivariate polynomials [11]. To understand how "multivariate polynomials" come into play we rewrite $\tilde{X}_T$ as follows. For a square $\square$, let $I_\square$ be the indicator of the event that $\square$ contains a vertex. Then

$$\tilde{X}_T = \sum_{T' : T' \sim T} \left[\prod_{\square \in T'} I_\square \right].$$

This expression is a multivariate polynomial in the variables $\left\{ I_{\square_{ij}} \right\}_{1 \le i,j \le M}$.

Kim and Vu's result

We present Kim and Vu's result in a somewhat simplified version which is more suited for our situation.

Assume that $\{I_i\}_{i \in \mathcal{I}}$ is a finite set of *independent* Bernoulli random variables. Let $\mathcal{P}$ be a set of subsets of $\mathcal{I}$, all with the same cardinality k. Kim and Vu give a concentration inequality for the random variable

$$X = \sum_{P \in \mathcal{P}} \left[\prod_{i \in P} I_i \right].$$

To formulate it we need to introduce some definitions.

For a given set $A \subseteq \mathcal{I}$ with cardinality $|A| \le k$ we define

$$\mathcal{P}_A = \{ P \in \mathcal{P} \mid A \subseteq P \}.$$

In addition, we associate it with a random variable:

$$X_{\mathcal{P}_A} = \sum_{P \in \mathcal{P}_A} \left[\prod_{i \in P \setminus A} I_i \right].$$

For every $0 \le j \le k$ let

$$E_j = \max_{A \subseteq \mathcal{I}, |A| = j} \mathbb{E}\left(X_{\mathcal{P}_A} \right).$$

Also define

$$E = \max_{j \ge 0} E_j, \quad E' = \max_{j \ge 1} E_j.$$

Notice that $E_0 = \mathbb{E}X$, thus $E = \max\{E', \mathbb{E}X\}$.

Kim and Vu's main result is the following theorem.

Theorem 6.24. *In the above setting, for every $\lambda > 1$,*

$$\mathbb{P}\left(|X - \mathbb{E}X| > \frac{8^k}{\sqrt{k!}} \sqrt{E \cdot E'} \cdot \lambda^k \right) = O\left(e^{-\lambda + (k-1)\ln|\mathcal{I}|} \right).$$

Back to concentration of $\tilde{X}_T$

Let us see how Theorem 6.24 applies for $\tilde{X}_T$. In our case:

- $\mathcal{I}$ is the set of all M^2 squares.
- $\mathcal{P}$ is the set of all valid tuples T' with $T' \sim T$. It is a set of subsets of $\mathcal{I}$, all with cardinality $4k+1$.
- As explained, with the above $\mathcal{I}, \mathcal{P}$ we have $X = \tilde{X}_T$.

For a set of squares $A \subseteq \mathcal{I}$ with $|A| = j$, $\mathcal{P}_A$ becomes the set of valid tuples T' with $T' \sim T$ which contain all the squares of A. $X_{\mathcal{P}_A}$ then sums over $\mathcal{P}_A$ and counts the tuples for which each of the remaining $4k+1-j$ squares contains a vertex.

Let $E' = \max_{j \geq 1} E_j$ as in Kim-Vu. The key technical result is the following lemma.

Lemma 6.25. $E' = n^{1-\Omega(1)}$.

Before proving the lemma, let us explain how it is used to finish the proof. Theorem 6.24 tells us that for every $\lambda > 1$,

$$\mathbb{P}\left(\left|\tilde{X}_T - \tilde{\delta}_T\right| > \frac{8^{4k+1}}{\sqrt{(4k+1)!}}\sqrt{E \cdot E'} \cdot \lambda^{4k+1}\right) = O\left(e^{-\lambda + 4k\ln(M^2)}\right).$$

From Lemma 6.25, the fact that $E = \max\{E', \tilde{\delta}_T\}$ and the bound $\tilde{\delta}_T = n^{1-\Omega(1)}$,

$$\sqrt{E \cdot E'} = O\left(n^{1-c}\right)$$

for some (absolute) positive constant c. Setting $\lambda = \ln^3 n$ and recalling that $k \leq K \cdot t \leq (\ln\ln\ln n)^2$, we also have

$$\frac{8^{4k+1}}{\sqrt{(4k+1)!}}\lambda^{4k+1} = n^{o(1)}.$$

In addition,

$$e^{-\lambda + 4k\ln(M^2)} = e^{-\ln^3 n + 4k \cdot 2\ln^2 n} = \exp\left(-\Omega(\ln^3 n)\right).$$

Overall

$$\mathbb{P}\left(\left|\tilde{X}_T - \tilde{\delta}_T\right| \geq n^{1-c+o(1)}\right) = \exp\left(-\Omega(\ln^3 n)\right).$$

By setting the constant γ to be smaller than c and recalling Equation (16) we obtain

$$\mathbb{P}\left(\tilde{X}_T > \frac{\delta\mu}{2\rho}\right) = \exp\left(-\Omega\left(\ln^3 n\right)\right)$$

which is what we wanted to prove.

Finally, we are left with the proof of Lemma 6.25. Recall that

$$E' = \max_{j\geq 1} E_j = \max_{A\subseteq \mathcal{I}, |A|\geq 1} \mathbb{E}\left(X_{\mathcal{P}_A}\right).$$

So let us fix a set of squares A with $|A| = j$, $1 \leq j \leq 4k+1$ and prove $\mathbb{E}\left(X_{\mathcal{P}_A}\right) = n^{1-\Omega(1)}$ (uniformly over A). With A fixed, we need to estimate

$$\mathbb{E}\left(X_{\mathcal{P}_A}\right) = \sum_{T':T'\sim T, T'\supseteq A} \mathbb{E}\left(\prod_{\square\in T'\setminus A} I_\square\right).$$

The indicators $I_\square$ are independent, and each summand is the expectation of the product of $4k+1-j$ of them, therefore each summand is $O\left((n\varepsilon^2)^{4k+1-j}\right)$. The main task is therefore to count the number of summands, i.e. the number of valid $(4k+1)$-tuples T' with $T' \sim T$ such that $T' \supseteq A$. A bit surprisingly, it turns out we can achieve the desired bound even when we ignore the condition $T' \sim T$ and count all valid tuples T' with $T' \supseteq A$. The idea is that having a square from a quadruplet $\square_{\mathbf{q}}$ fixed forces the other squares to be in a specific locus with small area. Counting all possibilities for the new $4k+1-j$ squares while taking into account these geometric requirements lead us to a bound of

$$O\left((M^2)^{4k+1-j} n^{4k-j} \cdot n^{-\Omega(1)}\right).$$

The precise details repeat (once more) considerations from the Negligibility Theorem and introduce no novelty; we leave it as an exercise to the interested reader. Overall we get

$$\mathbb{E}\left(X_{\mathcal{P}_A}\right) = O\left((M^2)^{4k+1-j} n^{4k-j} \cdot n^{-\Omega(1)}\right) \cdot O\left((n\varepsilon^2)^{4k+1-j}\right) = n^{1-\Omega(1)}$$

which, as explained, proves Lemma 6.25.

7 Disproving the Limit Law

The results of Section 6 indicate that the language of first order logic is far more expressive in the random geometric model than previously conjectured. In this section we demonstrate the power of our results by showing how they allow the construction of a first order sentence A with no limiting probability.

Given what we have so far, the remaining work is mostly logical, and hardly depends on the random graph model. Therefore, most of this section borrows directly from the work of Spencer and Shelah [19, 17], who disproved the Limit law for $G(n, p)$ (with $p = n^{-\alpha}$, $0 < \alpha < 1$ irrational).

7.1 Existential Second Order Logic and Arithmetization

The general idea is to construct a first order sentence A which makes an infinitely-alternating statement about the size of the graph. Generally, first order logic cannot directly refer to the size of the model in such a way. Existential second order logic, however, definitely can.

Recall that second order logic is built upon first order logic and adds the ability to quantify variables that represent *relations* (and not just *elements*, as in first order logic). Existential second order (ESO) logic allows only existential quantification $\exists$ over relational variables. An ESO sentence takes the form

$$\exists R_1^{k_1} \ldots \exists R_t^{k_t} (\varphi)$$

where each R_i is an k_i-ary relation and φ is a first order sentence that may use the relational symbols $R_1, \ldots, R_t$. We define its *maximal arity* to be $K = \max\{k_1, \ldots, k_t\}$.

Example 7.1. The following ESO sentence expresses the property "the graph is not connected":

$$\exists R^1 \left[\exists x(R(x)) \wedge \exists y(\neg R(y)) \wedge \forall x \forall y \left(R(x) \wedge \neg R(y) \to \neg x \sim y\right)\right].$$

In words, it asserts the existence of an unary relation R which divides the graph into two non-empty and non-connected parts. Note that this property cannot be expressed with a first-order sentence (e.g. see [17], Theorem 2.4.1).

The following theorem shows how to "convert" an ESO sentence into a first-order sentence, at the price of altering the size of the model in a controlled way.

Theorem 7.2. *Let A be an existential second sentence with maximal arity K. Then there exists a first order sentence A^* and there exists a function $f : \mathbb{N} \to \mathbb{N}$ which is $f(n) = O(n^K)$ as $n \to \infty$, such that:*

1. *If A holds for a finite graph H with n vertices, then A^* holds for a certain extension G of H with m vertices, such that $m \leq f(n)$.*
2. *If A^* holds for a finite graph G with m vertices, then A holds for a certain induced subgraph H of G with n vertices, such that $m \leq f(n)$.*

The conversion from A to A^* is done by "encoding" the relations $R_1, R_2, \ldots, R_t$ which A uses as certain structures in the graph, which requires an extension with $O(n^K)$ new vertices. For a detailed proof we refer to [17], Claim 8.2.2.

The next step is to use ESO logic to express arithmetic properties of the size of the graph n. This step is known as *arithmetization*. In particular, we are interested in an infinitely-alternating property which is also "insensitive" to a replacement of n with $\ln \ln \ln \ln n$ (as such a replacement will be forced by the results of Section 6). This naturally leads us to the *tower function* and its inverse, *the iterated logarithm function*.

Definition 7.3. The *tower function* $T : \mathbb{N} \to \mathbb{N}$ is defined recursively as follows:

$$\begin{aligned} T(0) &= 1, \\ T(n+1) &= 2^{T(n)}. \end{aligned}$$

The *iterated logarithm function* $\log^* : \mathbb{N} \to \mathbb{N}$ is defined as follows:

$$\log^*(n) = \min\{k \geq 1 \mid T(k) \geq n\}.$$

That is, $\log^*(n)$ is the first number k such that $T(k)$ surpasses n.

Example 7.4. The first values of the iterated logarithm function are:

$$\log^*(n) = \begin{cases} 1 & 1 \leq n \leq 2 \\ 2 & 2 < n \leq 4 \\ 3 & 4 < n \leq 16 \\ 4 & 16 < n \leq 65536 \\ 5 & 65536 < n \leq 2^{65536} \end{cases}.$$

Generally, $\log^*(n) = k$ if and only if

$$\underbrace{2^{2^{\cdot^{\cdot^{\cdot^{2}}}}}}_{k-1} = T(k-1) < n \leq T(k) = \underbrace{2^{2^{\cdot^{\cdot^{\cdot^{2}}}}}}_{k}.$$

Lemma 7.5. *Let n, n' be natural numbers such that*

$$\frac{1}{2} \ln\ln\ln\ln n \leq n' \leq n.$$

Then

$$\log^*(n) - 5 \leq \log^*(n') \leq \log^*(n).$$

Proof. Directly from the definition of $\log^*(n)$. ∎

We are now ready to describe an ESO sentence BigGap, which expresses an infinitely-alternating, "insensitive" property of the size of the graph n. The property is precisely that $\log^*(n)$ is equivalent to one of $1, 2, \ldots, 50$ modulo 100. The construction of BigGap as an ESO sentence is elaborated in [17], Sub-subsection 8.3.3. Let us briefly repeat the idea here.

First, we construct an ESO sentence Arith as follows. Arith begins by asserting the existence of four binary relations $x < y, D(x, y), E(x, y), T(x, y)$ and one unary relation $\text{Mod}(x)$. Then it elaborates a list of requirements about these relations. One set of requirements makes $x < y$ a total order, which induces a numbering $1, 2, \ldots, n$ on the vertices. Another set of requirements guarantees that $D(x, y)$ if and only if $y = 2x$ (with respect to the numbering), and similarly that $E(x, y)$ if and only if $y = 2^x$, $T(x, y)$ if and only if $y = T(x)$ and $\text{Mod}(x)$ if and only if $x \equiv 1, 2, \ldots, 50 \pmod{100}$.

BigGap asserts the existence of the same relations and makes the same set of requirements as in Arith, and additionally uses these relations to require

$$\log^*(n) \equiv 1, 2, \ldots, 50 \pmod{100}$$

where n is the size of the model.

Apply Theorem 7.2 to convert Arith, BigGap into first order sentences Arith* and BigGap*.

Proposition 7.6. *Let G be a graph with m vertices, as assume that it satisfies* Arith**.*

- *If* $\log^* m$ *is one of* $2, 3, \ldots, 50$ *modulo* 100*,* BigGap* *holds for G.*
- *If* $\log^* m$ *is one of* $52, 53, \ldots, 99$ *modulo* 100*,* BigGap* *does not hold for G.*

Proof. From Theorem 7.2, there exists an induced subgraph H of G such that Arith holds for H. Moreover, BigGap* holds for G if and only if BigGap holds for H. Since the maximal arity of the sentences is 2,

$$n = |V(H)| = \Omega(\sqrt{m}).$$

In particular,

$$\log^*(m) - 1 \leq \log^*(n) \leq \log^*(m).$$

- If $\log^* m$ is one of $2, 3, \ldots, 50$ modulo 100 then $\log^*(n)$ is one of $1, 2, \ldots, 50$ and so BigGap holds for H and BigGap* holds for G.
- If $\log^* m$ is one of $52, 53, \ldots, 99$ modulo 100 then $\log^*(n)$ is one of $51, 52, \ldots, 99$ and so BigGap does not hold for H and BigGap* does not hold for G.

■

7.2 A First Order Sentence With No Limiting Probability

At this point it mainly remains to piece together all that we have already seen.

As Proposition 7.6 shows, the first order BigGap* is able to make an infinitely-alternating statement about the size of the graph, given that the graph satisfies Arith*. Therefore we only need to express a graph structure H inside $G_{\mathbb{T}^2}(n, r)$ which satisfies Arith* and has a size comparable to n (in our case $\ln \ln \ln \ln n$). The sentence "H satisfies BigGap*" will then be a first order sentence with no limiting probability.

Theorem 6.3 shows that indeed, we can express any graph structure H with $\ln\ln\ln\ln n$ vertices as $H_w(U_{\mathbf{x}})$ by using four vertices x_1, x_2, x_3, w. The sentence "$H_w(U_{\mathbf{x}})$ satisfies BigGap*" is first-order expressible, *given that we have the right vertices* x_1, x_2, x_3, w.

The only remaining problem is that we seem to have no direct way to express the right vertices x_1, x_2, x_3, w with first order logic. This is because we have no direct way to claim that $U_{\mathbf{x}}$ has $\ln\ln\ln\ln n$ vertices. To solve that we need one final trick. We would like to choose not just any four vertices x_1, x_2, x_3, w such that $H_w(U_{\mathbf{x}})$ satisfies Arith*, but those for which the size of $U_{\mathbf{x}}$ is *maximal*. To make it work, we need to be able to compare the sizes of sets of vertices by first order means. Theorem 6.3 assures that this is indeed possible when the number of vertices is slowly-growing.

Definition 7.7. Let U_1, U_2 be two sets of vertices. We say that a graph G on $U_1 \cup U_2$ *proves* $|U_2| > |U_1|$ if:

- Every vertex from $U_1 \setminus U_2$ is connected to exactly one vertex from $U_2 \setminus U_1$.
- There exists a vertex from $U_2 \setminus U_1$ which is not connected to any vertex from $U_1 \setminus U_2$.

Definition 7.8. Let U_1, U_2 be two sets of vertices in the random geometric graph $G_{\mathbb{T}^2}(n, r)$. We say that U_2 is *pseudo-bigger* than U_1 if there exists a vertex w such that $H_w(U_1 \cup U_2)$ proves $|U_2| > |U_1|$.

If U_2 is pseudo-bigger than U_1 then indeed $|U_2| > |U_1|$. Theorem 6.3 shows that if $|U_1 \cup U_2| \leq \ln\ln\ln\ln n$ and also the vertices of $U_1 \cup U_2$ are nicely positioned (see Definition 6.2) then the opposite is also true. Luckily, the property of being nicely positioned is also expressible with first-order logic (at least to a good approximation). Indeed, recall that it requires well-spacedness and diameter $\leq \frac{r}{4}$. But in Section 2 we constructed a first-order formula which approximates well-spacedness, and first-order formulas which approximate any constant distance ($\frac{r}{4}$ in particular).

Definition 7.9. Suppose that U is a set of vertices in $G_{\mathbb{T}^2}(n, r)$, such that the property $v \in U$ is first-order expressible. Define the first-order sentence $\mathrm{NicePos}[U]$ as

$$\forall u_1, u_2 \in U \left[\mathrm{WS}(u_1, u_2) \wedge \forall u_1, u_2 \left(\mathrm{D}(u_1, u_2)\right)\right].$$

Here $\mathrm{WS}(x, y)$ is the formula from Theorem 2.8 (for a sufficiently small constant $\delta > 0$) and $\mathrm{D}(x, y)$ is the formula from Theorem 2.1 for an sufficiently small ε and $\alpha = \frac{1}{4}$.

Lemma 7.10. *A.a.s., for every set of vertices U in $G_{\mathbb{T}^2}(n, r)$, the following is true.*

- *If* $\mathrm{NicePos}[U]$ *holds then* $\mathrm{diam}(U) \leq \left(\frac{1}{4} + \varepsilon\right) r$ *and* $\|u_1 - u_2\| \geq n^{-\frac{1}{6} - \delta}$ *for every* $u_1, u_2 \in U$.

- *If* $\operatorname{diam}(U) \le \left(\frac{1}{4} - \varepsilon\right) r$ *and* $\|u_1 - u_2\| \ge n^{-\frac{1}{6}+\delta}$ *for every* $u_1, u_2 \in U$, *then* $\mathrm{NicePos}[U]$ *holds.*

The lemma is a direct consequence of the results from Section 2. For sufficiently small ε, δ, the fact that we have $\frac{r}{4} \pm \varepsilon r$ and $n^{-\frac{1}{6}\pm\delta}$ instead of $\frac{r}{4}$ and $n^{-\frac{1}{6}}$ makes no essential difference; we may allow the definition of "nicely positioned" to be slightly "stretched" accordingly.

We are now finally ready to define our first order sentence with no limiting probability.

Definition 7.11. Let A be the sentence that claims the existence of vertices x_1, x_2, x_3, w such that:

1. Arith^* holds for $H_w(U_{\mathbf{x}})$.
2. $\mathrm{NicePos}[U_{\mathbf{x}}]$ holds.
3. If x_1', x_2', x_3', w' are also vertices such that Arith^* holds for $H_{w'}(U_{\mathbf{x}'})$, then $U_{\mathbf{x}'}$ is *not* pseudo-bigger than $U_{\mathbf{x}}$.
4. BigGap^* holds for $H_w(U_{\mathbf{x}})$.

Note that A is indeed a first order sentence: the graph structure $H_w(U_{\mathbf{x}})$, the sentences Arith^* and BigGap^*, the sentence $\mathrm{NicePos}[U_{\mathbf{x}}]$ and being pseudo-bigger — they are all first order expressible. Theoretically A could be written down explicitly as a formal first order sentence in the language of graph theory, though it would be awfully long.

Theorem 7.12. *The sentence A has no limiting probability in the random geometric model: the limit*

$$\lim_{n\to\infty} \mathbb{P}\left(G_{\mathbb{T}^2}(n,r) \in A\right)$$

does not exist.

Proof. First, we restrict to a sequence of n such that $\log^* n$ is 99 modulo 100, and show that the limiting probability of A along this sequence is 0. To do that, we show that a.a.s., if x_1, x_2, x_3, w are four vertices such that the first three conditions from Definition 7.11 hold, then the fourth condition must fail.

A.a.s., for every x_1, x_2, x_3 such that $U_{\mathbf{x}}$ is nicely positioned, there exist x_1', x_2', x_3' such that $|U_{\mathbf{x}'}| = \frac{1}{2} \ln\ln\ln\ln n$ and $U_{\mathbf{x}} \cup U_{\mathbf{x}'}$ is also nicely positioned (from the same considerations as Remark 6.14). So, given x_1, x_2, x_3 such that the first three conditions from Definition 7.11 hold, we may assume the existence of a suitable triplet x_1', x_2', x_3' as above. From Theorem 6.3 we may also assume that for every set of nicely positioned vertices U with $|U| \le \ln\ln\ln\ln n$ and for every graph structure H on U there exists a vertex w with $H_w(U) = H$.

Since $U_{\mathbf{x}'}$ is nicely positioned, there exists a vertex w' such that Arith* holds for $H_{w'}(U_{\mathbf{x}'})$. If $|U_{\mathbf{x}}| < |U_{\mathbf{x}'}|$ then $U_{\mathbf{x}'}$ is pseudo-bigger than $U_{\mathbf{x}}$; indeed, $U_{\mathbf{x}} \cup U_{\mathbf{x}'}$ is nicely positioned and its size is at most $2\,|U_{\mathbf{x}'}| = \ln\ln\ln\ln n$. But that contradicts the second condition. Therefore

$$t = |U_{\mathbf{x}}| \geq |U_{\mathbf{x}'}| = \frac{1}{2}\ln\ln\ln\ln n.$$

By Lemma 7.5, $\log^* t$ is one of $94, 95, \ldots, 99$ modulo 100, and from Proposition 7.6 BigGap* does not hold for $H_w(U_{\mathbf{x}})$.

Second, we restrict to a sequence of n such that $\log^* n$ is 49 modulo 100, and show that the limiting probability of A along this sequence is 1. Choose x_1, x_2, x_3, w such that NicePos$[U_{\mathbf{x}}]$ holds, $H_w(U_{\mathbf{x}})$ satisfies Arith* and $|U_{\mathbf{x}}|$ is the maximal possible. From Theorem 6.3 we know that a.a.s. such vertices do exist, and that

$$t = |U_{\mathbf{x}}| \geq \ln\ln\ln\ln n.$$

By definition, the first two conditions from Definition 7.11 hold. Since pseudo-bigger implies bigger, the third condition also holds. Finally, by Lemma 7.5 $\log^* t$ is one of $44, 45, \ldots, 49$ modulo 100, and from Proposition 7.6 BigGap* holds for $H_w(U_{\mathbf{x}})$. That finishes the proof. ■

8 Further Generalizations and Conjectures

There are several interesting approaches for generalizing our results, some are relatively straightforward while others would require non-trivial adjustments.

First, recall Spencer and Agarwal's third conjecture [18]: that a Zero-One law holds for $G_{\mathbb{T}^2}(n, r(n))$ whenever $r(n) = o(1)$ but $r(n) = n^{o(1)}$. We can disprove this conjecture as well by straightforward generalization of our results from the case when r is an arbitrarily small constant to the case when $r(n) \to 0$ "sufficiently slowly". The idea is that the entire proof eventually relies on asymptotic estimations of distances, expectations and probabilities. These asymptotic estimations, specifically those which involve r, are "flexible": if we make sure that $r(n) \to 0$ sufficiently slowly, they are all still able to serve the same purpose. No part of the proof relies on a phenomenon that is essentially unique to the constant r case, in a way which cannot be "smoothed out" to $r(n) \to 0$ accordingly.

Let us give a more detailed explanation. For concreteness let us set $r(n) = \frac{1}{\ln\ln\ln\ln\ln\ln\ln n}$. Theorem 2.1, for example, still holds, only now it does not approximate constant distances but constant multiples of $r(n)$. The same proof applies: we can still add and subtract distances and express the distances $r(n)$ and $\sqrt{3}r(n)$. The fact that now some geometric loci have an area of $\Theta\left(r(n)^{-\mathrm{const}}\right)$ instead of $\Theta(1)$ does not damage the proof, as the difference is insignificant relative to the number of vertices n. In Section 4, following the proof of Theorem 4.1 reveals that the probability of the UN-relation becomes $Cr(n)n^{-2/3}$. Consequently, in Section 5 the joint distribution of witnesses over different triplets will still be approximately independent Poisson variables with parameters $\lambda_1, \lambda_2, \ldots, \lambda_t$, but now the parameters λ_i will not be constant, but $\Theta\left(r(n)^{-\text{const}}\right)$. However, $r(n)$ is chosen in such a way that the resulting estimations of probabilities are not harmed by this change. The main example is the proof of Theorem 6.3, in which we bound $\mathbb{P}_{\mathbf{a}}(\mathbf{Z} = \mathbf{z})$ from below by c^t. Now, instead of a constant, c will also be $\Theta\left(r(n)^{-\text{const}}\right)$; but when we compare it to much smaller error term, it makes no difference, and we still obtain the desired positive lower bound.

Notice that this argument does not apply for the entire region where $r(n) \to 0$ and $r(n) = n^{o(1)}$. Taking $r(n) = \frac{1}{(\ln n)^{10}}$, for example, breaks at least some of the estimations. Investigating the logical behavior of the RGG in the entire region would probably require more flexible adaptations of our ideas (and careful attention to details). We leave it as an open question.

The next natural generalization is the case of higher dimensions: $G_{\mathbb{T}^d}(n, r)$ for $d > 2$. We believe that our work can also be generalized to handle this case as well. The main difference is that now the probability of the UN-relation becomes $\approx Cn^{-\frac{2}{d+1}}$. This means that we must redefine the S-extension, in order to maintain the balance between the number of vertices and the numbers of UN-relations. Changing the definition of S also means that several other technical details (mainly the proof of the Negligibility Theorem) would have to be adjusted. However, we believe that these are just technicalities, and the essential ideas remain the same. To further justify our belief, note that this generalization is similar to a generalization that Spencer and Shelah [19] make. When they disprove the Limit law for $p = n^{-\alpha}$ with $0 < \alpha < 1$ rational, they actually illustrate the proof only for $\alpha = \frac{1}{7}$ (or $\alpha = \frac{1}{3}$ in Spencer's book [17]), and claim that redefining the extension accordingly handles the general case. In the RGG, changing the dimension becomes analogous to changing the value of α (although with some added technical difficulties).

Another interesting case is when $G_{\mathbb{T}^d}(n, r(n))$ with $r(n) = n^{-\beta}$. The analogue of Theorem 4.1 shows that the probability of the UN-relation is

$$\approx Cr^{d-1}n^{-\frac{2}{d+1}} = Cn^{-\frac{2}{d+1}-(d-1)\beta}.$$

To maintain the analogy to the binomial case $p(n) = n^{-\alpha}$, where it is assumed that $0 < \alpha < 1$, here we assume

$$0 < \frac{2}{d+1} + (d-1)\beta < 1$$

which implies $0 < \beta < \frac{1}{d+1}$. Again, we expect that when the exponent is rational, it is possible to find a suitable extension formula with asymptotically constant expected number of witnesses, and can therefore serve as a basis for encoding arbitrary graph structures. However, the generalization is less straightforward, and may require significant adjustments to many of the asymptotic estimations. We pose the following conjecture.

Conjecture 8.1. *Consider the d-dimensional RGG $G_{\mathbb{T}^d}(n, r(n))$ with $r(n) = n^{-\beta}$, $0 < \beta < \frac{1}{d+1}$.*

1. *If β is irrational, the Zero-One law holds.*
2. *If β is rational, the Limit law fails: there exists a first-order A with no limiting probability.*

Finally, an additional interesting generalization would be to investigate other metrics on the torus. Let us consider the ℓ^p metric for $1 \leq p \leq \infty$. For concreteness, let us fix $d = 2$ and a constant r. The analogue of Theorem 4.1 shows that the probability of the UN-relation is now $\approx Cn^{-\frac{p}{p+1}}$ (which becomes $\approx Cn^{-1}$ when $p = \infty$). As always, we expect the logical behavior to depend on the rationality of the exponent: rational exponents should allow the construction of a suitable first-order extension, which serves to disprove a Limit law. This line of reasoning leads to a surprising prediction: the logical behavior of the RGG with ℓ^p should depend on the rationality of p! It would be very interesting to study the nature of this phenomenon in depth.

A Appendix

A.1 Poissonization

Poissonization is a standard technique in the study of random geometric graphs. We use it twice throughout the paper: in the proof of Lemma 2.16 and in the proof of the Concentration Theorem 6.13. For completeness, we briefly review it here.

Poissonization replaces the random geometric graph $G_{\mathbb{T}^2}(n, r)$, which generates a fixed number of vertices, with a Poisson random graph, in which the set of vertices is a Poisson point process. Here are the relevant definitions.

Definition A.1. Let N be a random variable with $N \sim \text{Pois}(\lambda)$, and let $\{v_i\}_{i \in \mathbb{N}}$ an infinite set of random points in $\mathbb{T}^2$ which are uniformly and independently distributed (while also being independent of N). Then the random set of points $\mathcal{P}_\lambda = \{v_1, v_2, \ldots, v_N\}$ is called a *Poisson point process* with intensity λ.

Definition A.2. A *Poisson random graph* with intensity λ and distance parameter r (in the torus $\mathbb{T}^2$), denoted here by $G_{\mathbb{T}^2}(\mathcal{P}_\lambda, r)$, is a random graph whose vertex set is a Poisson point process $\mathcal{P}_\lambda$, and whose edges are defined by $v_i \sim v_j \iff \|v_i - v_j\| < r$.

In the paper, we actually need a slight variation of the Poisson random graph, which also incorporates a fixed set of vertices "in the background".

Definition A.3. Fix $1 \leq v \leq \ln\ln\ln\ln n$ (n is an underlying parameter). Also fix a set U of v vertices and a geometric configuration $\mathbf{u} = \mathbf{a}$ of U (see Definition 5.1). We define a random geometric graph as follows. Its vertex set is $U \cup \mathcal{P}_\lambda$, where $\mathcal{P}_\lambda$ is a Poisson point process with intensity $\lambda = n - v$. Its edges are defined geometrically as expected: $v_i \sim v_j \iff \|v_i - v_j\| < r$. We denote this random geometric graph $G_{\mathbf{a}}(N, r)$, where N is the number of vertices, which is a random variable with $N - v \sim \text{Pois}(n - v)$. Note that the expected number of vertices is $\mathbb{E}(N) = n$.

Remark A.4. Note that the distribution of $G_{\mathbf{a}}(N, r)$ conditioned by the event $N = n$ is identical to the distribution of the standard $G_{\mathbb{T}^2}(n, r)$, conditioned by the configuration $\mathbf{u} = \mathbf{a}$.

The last remark serves as a bridge between $G_{\mathbf{a}}(N, r)$ and $G_{\mathbb{T}^2}(n, r)$. Here is an important corollary.

Corollary A.5. *Let P be a property of geometric graphs (it may depend on the positions of the vertices). Suppose that*

$$\mathbb{P}\left(G_{\mathbf{a}}(N, r) \in P\right) = n^{-\omega(1)}.$$

Then

$$\mathbb{P}\left(G_{\mathbb{T}^2}(n, r) \in P \mid \mathbf{u} = \mathbf{a}\right) = n^{-\omega(1)}.$$

That is, if something happens with "very small" probability in the Poisson random graph, it also happens with "very small" probability in the standard RGG.

Proof. From the remark,

$$\mathbb{P}\left(G_{\mathbb{T}^2}(n, r) \in P \mid \mathbf{u} = \mathbf{a}\right) \leq \frac{\mathbb{P}\left(G_{\mathbf{a}}(N, r) \in P\right)}{\mathbb{P}(N = n)}.$$

By Stirling's approximation (and the bound on v), $\mathbb{P}(N = n) = \Theta\left(n^{-1/2}\right)$ and the corollary easily follows. ∎

The reason that Poisson random graphs are so useful is a crucial property called *spatial independence*. We formalize it for $G_{\mathbf{a}}(N, r)$ in the following proposition.

Proposition A.6. *Let $G_{\mathbf{a}}(N, r)$ be as in Definition A.3. For a (measurable) set $A \subseteq \mathbb{T}^2$, let $\nu(A)$ count the number of vertices that land inside A, excluding vertices of U. Then:*

1. *$\nu(A) \sim \text{Pois}((n - v) \cdot m(A))$ where $m(A)$ is the Lebesgue measure of A.*
2. *For any finite collection of pairwise-disjoint sets $A_1, A_2, \ldots, A_k$, the variables $\nu(A_1), \nu(A_2), \ldots, \nu(A_k)$ are independent.*

A.2 Proof of the Negligibility Theorem

This subsection is dedicated to the proof of Theorem 5.7. The theorem states that certain situations, in which the witnesses for the extensions do not "behave well", have negligible contribution to the joint factorial moments. The considerations of the proof are also relevant for similar situations from Section 6. The different parts of the proof rely on different geometric lemmas, which we choose to concentrate at the end of this subsection.

Recall the setting of the theorem: we fix a strongly well-spaced geometric configuration $\mathbf{u} = \mathbf{a}$, which fixes the positions of the t given triplets $\mathbf{x}^{(i)}$, $1 \le i \le t$. For $0 \le \mathbf{k} \le K$, the joint factorial moment $\mathbb{E}_{\mathbf{a}}\left((\mathbf{Z})_{\mathbf{k}}\right)$ counts $\mathbf{k}$-tuples of quadruplets

$$\mathbf{Q} = \left(\mathbf{q}_1^{(1)}, \ldots, \mathbf{q}_{k_1}^{(1)}, \ldots, \mathbf{q}_1^{(t)}, \ldots, \mathbf{q}_{k_t}^{(t)}\right)$$

such that $\mathbf{q}_1^{(i)}, \ldots, \mathbf{q}_{k_i}^{(i)}$ are all distinct quadruplets for every $1 \le i \le t$ and such that each $\mathbf{q}_j^{(i)}$ satisfies $S\left(\mathbf{x}^{(i)}; \mathbf{q}_j^{(i)}\right)$. Also recall that $K = \ln\ln\ln n$ and that $t \le (\ln\ln\ln\ln n)^3$. Definition 5.6 and the replacement of S with S^* both define a list of all the "bad" situations which must be considered.

From now on let us fix $0 \le \mathbf{k} \le K$. All our estimations will be uniform in $\mathbf{k}$. $\mathbf{Q}$ will always denote a $\mathbf{k}$-tuple of quadruplets

$$\mathbf{Q} = \left(\mathbf{q}_1^{(1)}, \ldots, \mathbf{q}_{k_1}^{(1)}, \ldots, \mathbf{q}_1^{(t)}, \ldots, \mathbf{q}_{k_t}^{(t)}\right)$$

such that $\mathbf{q}_1^{(i)}, \ldots, \mathbf{q}_{k_i}^{(i)}$ are all distinct quadruplets for every $1 \le i \le t$. We denote $k = k_1 + \cdots + k_t$.

Before handling the other situations, we must first be able to replace the UN-relations with the stronger UN*-relations.

Proposition A.7. *The expected number of* $\mathbf{k}$*-tuples* $\mathbf{Q}$ *of witnesses for the* S*-extension (over the triplets* $\mathbf{x}^{(i)}$*), such that it contains a pair of vertices which satisfies the* UN *relation but not the* UN* *relation, is* $n^{-\Omega(1)}$*.*

Proof. Let us fix a $\mathbf{k}$-tuple of vertices $\mathbf{Q}$. We prove that it contains a pair which satisfies UN but not UN* with probability $n^{-\Theta(\ln\ln n)}$. Then, the expectation is the sum of those probabilities over all possible $\mathbf{Q}$, which is $n^{\Theta(Kt)} = n^{o(\ln\ln n)}$, so overall we have

$$\mathbb{E}_{\mathbf{a}}(\ldots) = n^{o(\ln\ln n)} \cdot n^{-\Theta(\ln\ln n)} = n^{-\Omega(1)}.$$

To prove the bound on the probability for a given $\mathbf{Q}$, note that for a pair of vertices x, y, satisfying $\mathrm{UN}(x, y)$ but not $\mathrm{UN}^*(x, y)$ means that

$$\mathrm{UN}(x, y) \wedge H_{xy} \in [h_n, r].$$

However, Equation (5) from Section 4 shows that this happens with probability $n^{-\Theta(\ln\ln n)}$. This bound holds for each of the $6k$ relevant pairs, and a simple union bound over them finishes the proof. ■

Remark A.8. We can now shed more light on the purpose of the *strong* well-spacedness condition from Definition 5.3. Let $\mathbf{q} = (s_1, s_2, s_3, z)$ satisfy $S(\mathbf{x}^{(i)}; \mathbf{q})$. Following the last proposition, assume that all each of the six UN-relation is also a UN*-relation. Then, given that the triplet $\mathbf{x}^{(i)}$ is strongly well-spaced, at least two of s_1, s_2, s_3 must be themselves well-spaced. Indeed, assume otherwise:

$$\|s_1 - s_2\|, \|s_2 - s_3\|, \|s_3 - s_1\| \leq n^{-1/6}.$$

The UN* relations tell us that

$$\|x_1 - s_1\|, \|x_2 - s_2\|, \|x_3 - s_3\| \in [2r - 2h_n, 2r].$$

Since $h_n = o\left(n^{-1/6}\right)$, the distance of each s-vertex from each x-vertex is in

$$[2r - 2n^{-1/6}, 2r + 2n^{1/6}]$$

which contradicts strong well-spacedness.

In the following key step, we simultaneously handle situations **1a-b** and **2** from Definition 5.6.

Proposition A.9. *The expected number of* $\mathbf{k}$*-tuples* $\mathbf{Q}$ *of witnesses for the* S*-extension such that either **1a, 1b** or **2** occur is* $n^{-\Omega(1)}$.

Proof. We handle only situations **1b** and **2**; situation **1a** can be handled very similarly (actually it is even simpler). So, our task is to bound the expected number of $\mathbf{k}$-tuples $\mathbf{Q}$ of witnesses in which there are two distinct quadruplets which share a vertex, or two different vertices with distance $\leq n^{-1/6}$.

In this proof only, for the sake of brevity, if two vertices are not well-spaced we say that they are *badly close* to each other.

First, from the previous proposition, we may consider $\mathbf{k}$-tuples in which all the $6k$ UN-relations are also UN*-relations. Actually, we show that the expectation is $n^{-\Omega(1)}$ even when we "relax" our conditions, and replace each UN$^*(x, y)$ with the simple distance requirement $H_{xy} \in [0, h_n]$. This idea will repeat in all later steps, and it is the main reason we replaced UN with UN*. Let $S_{\text{rel}}(\mathbf{x}; \mathbf{q})$ denote the relaxed extension.

We fix a specific pattern of equalities and badly close vertices between the vertices of a $\mathbf{k}$-tuple $\mathbf{Q}$ and show that the contribution of $\mathbf{k}$-tuples with this pattern is (uniformly) $n^{-\Omega(1)}$. Summing over all patterns still leaves a bound of $n^{-\Omega(1)}$, since the number of patterns is poly-logarithmic.

So, let us fix the pattern and consider the expected number of $\mathbf{k}$-tuples $\mathbf{Q}$ with this pattern. The strategy is to "expose" each of the k quadruplets one by one. For every exposed quadruplet $\mathbf{q}$, we bound the factor that it contributes to the overall expectation. The idea is that each new vertex of $\mathbf{q}$ contributes a factor of $\approx n$ to the expectation, while the probabilities that all the necessary requirements are satisfied contribute an $o(1)$-factor. We show that any case of shared vertices or badly close vertices allows the probabilities to "win" over the vertices by a factor of $n^{-\Omega(1)}$.

WLOG we can arrange the order of exposure such that the the first two quadruplets already share vertices or have badly close vertices. Let handle those two quadruplets first.

We expose the first quadruplet. Denote it $\mathbf{q} = (s_1, s_2, s_3, z)$ and the corresponding triplet $\mathbf{x} = (x_1, x_2, x_3)$. It introduces four new vertices, which together contribute a factor of $O(n^4)$. There are still no previous quadruplets to share vertices with, and the UN* relations assure that z is not badly close to any of s_1, s_2, s_3. So there are two cases to consider.

Case 1. $\mathbf{q}$ has no badly closed vertices. So the only requirement on $\mathbf{q}$ is $S_{\mathrm{rel}}(\mathbf{x}; \mathbf{q})$. It introduces the requirements

$$H_{x_1 s_1}, H_{x_2 s_2}, H_{x_3 s_3}, H_{s_1 z}, H_{s_2 z}, H_{s_3 z} \in [0, h_n]$$
$$\wedge \mathrm{CD}(x_1, z) \wedge \mathrm{CD}(x_2, z) \wedge \mathrm{CD}(x_3, z),$$

Let us bound the probability of those requirements by $O(h_n^6)$.

- z must satisfy $\mathrm{CD}(x_1, z) \wedge \mathrm{CD}(x_2, z) \wedge \mathrm{CD}(x_3, z)$. The probability that is does is $O(1)$.
- Given $\mathrm{CD}(x_1, z)$, the probability that s_1 satisfies $H_{x_1 s_1}, H_{s_1 z} \in [0, h_n]$ is $O(h_n^2)$ by Lemma A.13 part 1. Indeed: this event geometrically means that s_1 is inside the intersection of the two annuli around x_1, z with radii $2r - 2h_n$ and $2r$.
- The same is true for s_2 and s_3: each contributes a factor of $O(h_n^2)$.

Overall we get a factor of $O\left(n^4 h_n^6\right)$. Recall that

$$h_n = \Theta\left(\left(\frac{\ln \ln n \cdot \ln n}{n}\right)^{2/3}\right)$$

so $O\left(n^4 h_n^6\right) = (\ln n)^{O(1)}$.

Case 2. $\mathbf{q}$ has two badly close s vertices. WLOG assume that s_1, s_2 are badly close. We bound the probability as follows. The idea is to show that the requirements force z to be inside a geometric locus with very small area, and handle s_1, s_2, s_3 like Case 1. Indeed, the $S_{\mathrm{rel}}(\mathbf{x}; \mathbf{q})$ requirement and the badly close requirement imply that

$$\|x_1 - s_1\|, \|x_2 - s_2\|, \|z - s_1\|, \|z - s_2\| \in [2r - 2h_n, 2r],$$
$$\|s_1 - s_2\| \le n^{-1/6}.$$

It is not hard to see that such distances force z to be at distance $O(n^{-1/6})$ from one of the two intersection points between the circles of radius $2r$ around x_1, x_2. This forces z to lie inside the union of two annuli whose widths are $O(n^{-1/6})$. The area of this union L is itself $O(n^{-1/6})$. The rest is the same as before:

- z satisfies the CD-relations and also lies inside L with probability $O\left(n^{-1/6}\right)$.
- Given that, the probability that s_1 satisfies $H_{x_1 s_1}, H_{s_1 z} \in [0, h_n]$ is again $O(h_n^2)$, and the same is true for s_2, s_3.

Overall we get a factor of

$$O(n^4 n^{-1/6} h_n^6) = (\ln n)^{O(1)} \cdot n^{-1/6}.$$

Now expose the second quadruplet. Denote it $\mathbf{q}' = (s'_1, s'_2, s'_3, z')$ and the corresponding triplet $\mathbf{x} = (x'_1, x'_2, x'_3)$. There are many different patterns for how $\mathbf{q}'$ may shares vertices with $\mathbf{q}$ or break well-spacedness; let us consider several representative cases.

Case 1. $\mathbf{q}'$ shares no vertices with $\mathbf{q}$ and has no badly close vertices. We proceed like Case 1 for $\mathbf{q}$ and get a factor of $(\ln n)^{O(1)}$. This case is possible only if $\mathbf{q}$ already has two badly close vertices, so in this case $\mathbf{q}$ already contributed a factor of $n^{-\Omega(1)}$.

Case 2. $\mathbf{q}'$ shares no vertices with $\mathbf{q}$ and s'_1, s'_2 are badly close to each other. We proceed like Case 2 for $\mathbf{q}$ and get a factor of $(\ln n)^{O(1)} \cdot n^{-1/6}$.

Case 3. $\mathbf{q}'$ shares no vertices with $\mathbf{q}$ and z' is badly close to a vertex of $\mathbf{q}$. The vertices contribute $O(n^4)$. Let us bound the relevant probability.

- z' must satisfy the CD-relations and also be badly close to one of 4 given vertices, which it does with probability $O\left(n^{-1/3}\right)$.
- Given that, the probability that s'_1 satisfies $H_{x'_1 s'_1}, H_{s'_1 z'} \in [0, h_n]$ is again $O(h_n^2)$ by Lemma A.13 part 1. The same is true for s'_2, s'_3.

Overall we get a factor of

$$O(n^4 n^{-1/3} h_n^6) = (\ln n)^{O(1)} \cdot n^{-1/3}.$$

Case 4. s'_2, s'_3, z' are shared which $\mathbf{q}$. The new vertex s'_1 contributes $O(n)$.

- If the existing vertex z' does not satisfy the CD-conditions with $\mathbf{x}'$, it simply zeroes out the probability.
- Otherwise, the probability that s'_1 satisfies $H_{x'_1 s'_1}, H_{s'_1 z'} \in [0, h_n]$ is again $O(h_n^2)$.

Overall we get a factor of $O(nh_n^2) = (\ln n)^{O(1)} \cdot n^{-1/3}$.

Case 5. s_1', s_2', s_3' are shared with $\mathbf{q}$. The new vertex z_1' contributes $O(n)$. To bound the probability, we further subdivide into two sub-cases.

Sub-case 5.1. s_1', s_2', s_3' are all well-spaced. In that case, the probability that z' satisfies $H_{s_1'z'}, H_{s_2'z'}, H_{s_3'z'} \in [0, h_n]$ is $O\left(n^{1/6}h_n^2\right)$ by Lemma A.14. Overall we get a factor of

$$O(n \cdot n^{1/6}h_n^2) = (\ln n)^{O(1)} \cdot n^{-1/6}.$$

Sub-case 5.2. Two of s_1', s_2', s_3' are not well-spaced. This means that it was Case 2 for $\mathbf{q}$, so it already contributed $n^{-\Omega(1)}$, and we can now be satisfied with a bound of $(\ln n)^{O(1)}$. From the assumption that $\mathbf{x}'$ is strongly well-spaced, we still know that at least two of s_1', s_2', s_3' are well-spaced (see Remark A.8). WLOG assume that s_1', s_2' are well-spaced. z' must satisfy $H_{s_1'z'}, H_{s_2'z'} \in [0, h_n]$, which it does with probability $O\left(h_n^{3/2}\right)$ by Lemma A.13 part 3. Overall we indeed get a factor of

$$O(n \cdot h_n^{3/2}) = (\ln n)^{O(1)}.$$

Case 6. All the vertices of $\mathbf{q}'$ are shared with $\mathbf{q}$. We recalculate the factor added by $\mathbf{q}$ and $\mathbf{q}'$ and show that it is $n^{-\Omega(1)}$. The vertices contribute $O(n^4)$. Now we bound the probability that they satisfy $S_{\text{rel}}(\mathbf{x}; \mathbf{q})$ and $S_{\text{rel}}(\mathbf{x}'; \mathbf{q}')$. First, notice that we must have $z = z'$; indeed, quadruplets must have a vertex at distances $\approx 2r$ from all the other three, and there can be only one. There are different cases about which of s_1, s_2, s_3 equal which of s_1', s_2', s_3', and also whether $\mathbf{x}' = \mathbf{x}$ or not, but all cases are handled in the same way. Take the case of $s_1 = s_1', s_2 = s_2', s_3 = s_3'$ and $\mathbf{x}' \neq \mathbf{x}$ for example. From $\mathbf{x}' \neq \mathbf{x}$ we may WLOG assume that $x_1 \neq x_1'$. From the assumption that the configuration $\mathbf{u} = \mathbf{a}$ is well-spaced we know that x_1, x_1' are well-spaced. The probability is bounded as follows.

- The probability that s_1 satisfies $H_{x_1s_1}, H_{x_1's_1} \in [0, h_n]$ is $O(h_n^{3/2})$ by Lemma A.13 part 3.
- The probability that z satisfies $H_{s_1z} \in [0, h_n]$ and also the CD-conditions is $O(h_n)$.
- Given all that, the probability that s_2 and s_3 satisfy $H_{x_2s_2}, H_{s_2z} \in [0, h_n]$ and $H_{x_3s_3}, H_{s_3z} \in [0, h_n]$ is $O(h_n^4)$, once more by Lemma A.13 part 1.

Overall we get a factor of

$$O\left(n^4 h_n^{6+1/2}\right) = (\ln n)^{O(1)} \cdot n^{-1/3}.$$

In conclusion, we have shown that any pattern of shared vertices and badly close vertices between $\mathbf{q}$, $\mathbf{q}'$ implies that together they contribute a factor of $n^{-\Omega(1)}$. Finally, we expose all the remaining quadruplets, one by one. It is now sufficient to show that each of them contributes a factor of at most $(\ln n)^{O(1)}$. The final bound on the expectation will then be

$$(\ln n)^{O(k)} \cdot n^{-\Omega(1)} = n^{-\Omega(1)}$$

as desired. Indeed, each of the new quadruplets can be handled by division to cases in the exact same way as $\mathbf{q}'$. In what we called Case 5, we can always continue like Sub-case 5.2. In what we called Case 6, no new vertices are added so we may actually ignore the new requirements and just take 1 as a bound. ■

With Proposition A.9 at hand, we can handle much more easily the remaining situations, which are the replacement of CD with CD* and situations **3a-b** and **4a-b** from Definition 5.6. We have already encountered the general idea: each new quadruplet generally contributes a factor of $(\ln n)^{O(1)}$, but an additional "bad situation" requirement adds a factor of $n^{-\Omega(1)}$ according to some geometric considerations.

Proposition A.10. *The expected number of* $\mathbf{k}$*-tuples* $\mathbf{Q}$ *of witnesses for the* S*-extension, such that* $\mathbf{Q}$ *contains a pair of vertices which satisfies the* CD*-relation but not the* CD***-relation, or vice versa, is* $n^{-\Omega(1)}$.

Proof. From the previous propositions, we may count only $\mathbf{k}$-tuples $\mathbf{Q}$ in which all $4k$ vertices are distinct. We also replace each $S(\mathbf{x};\mathbf{q})$ requirement with $S_{\text{rel}}(\mathbf{x};\mathbf{q})$ like before. Again, we expose the k quadruplets of $\mathbf{Q}$ one by one and bound the factor that they contribute to the expectation.

Suppose that we expose a new quadruplet $\mathbf{q} = (s_1, s_2, s_3, z)$. Denote the corresponding triplet $\mathbf{x} = (x_1, x_2, x_3)$. If the only requirement is $S_{\text{rel}}(\mathbf{x};\mathbf{q})$, we obtain the same bound on the factor contributed by $\mathbf{q}$:

$$O\left(n^4 h_n^6\right) = (\ln n)^{O(1)}.$$

Now assume the additional requirement that CD is not equivalent to CD* in $\mathbf{q}$. We show that the contributed factor becomes $n^{-\Omega(1)}$. WLOG we consider the requirement

$$\neg\left(\mathrm{CD}(x_1, z) \leftrightarrow \mathrm{CD}^*(x_1, z)\right).$$

Recall Remark 2.5. Let us take, for example, $\varepsilon = n^{-1/4}$. The remark claims that with probability $1 - \exp\left(-\Theta\left(n^{1/4}\right)\right)$, if

$$\|x_1 - z\| \notin \left[(\sqrt{3} - \varepsilon)r, (\sqrt{3} + \varepsilon)r\right]$$

then $\mathrm{CD}(x_1, z) \leftrightarrow \mathrm{CD}^*(x_1, z)$. We can therefore bound the probability that $S_{\text{rel}}(\mathbf{x};\mathbf{q})$ holds and also

$$\|x_1 - z\| \in \left[(\sqrt{3} - \varepsilon)r, (\sqrt{3} + \varepsilon)r\right]. \tag{17}$$

We do it in the standard way.

- z must satisfy the CD-relations and also Equation (17), which it does with probability $O(\varepsilon)$.
- Given that, the probability that s_1 satisfies $H_{x_1 s_1}, H_{s_1 z} \in [0, h_n]$ is again $O(h_n^2)$ by Lemma A.13 part 1. The same is true for s_2, s_3.

Overall we get a factor of

$$O(n^4 \varepsilon h_n^6) = (\ln n)^{O(1)} \cdot n^{-1/4}.$$

Such a factor is contributed by at least one of the exposed $\mathbf{q}$ (by assumption) so the overall expectation is indeed $n^{-\Omega(1)}$, as desired. ■

Now for situations **3a-b** from Definition 5.6, in which a vertex from a triplet $\mathbf{x}$ or a quadruplet $\mathbf{q}$ also functions as a witness to a UN-relation.

Proposition A.11. *The expected number of* $\mathbf{k}$*-tuples* $\mathbf{Q}$ *of witnesses for the* S*-extension such that either* ***3a*** *or* ***3b*** *occur is* $n^{-\Omega(1)}$.

Proof. Again, we count only $\mathbf{k}$-tuples $\mathbf{Q}$ in which all $4k$ vertices are distinct and replace each $S(\mathbf{x}; \mathbf{q})$ with $S_{\text{rel}}(\mathbf{x}; \mathbf{q})$. We expose the k quadruplets one by one. WLOG we can arrange the order of exposure such that situations **3a** or **3b** already occur in the first two quadruplets. We show that together the two quadruplets contribute a factor of $n^{-\Omega(1)}$. Every additional quadruplet will contribute at most $O(n^4 h_n^6) = (\ln n)^{O(1)}$ so the overall expectation would then be $n^{-\Omega(1)}$, as desired.

We expose the first quadruplet. Denote it $\mathbf{q} = (s_1, s_2, s_3, z)$ and the corresponding triplet $\mathbf{x} = (x_1, x_2, x_3)$. There are still no previous quadruplets, so occurrence of **3a** or **3b** means that a vertex from $\mathbf{x}$ or from $\mathbf{q}$ must be a witness for one of the six UN-relations of $\mathrm{S}(\mathbf{x}; \mathbf{q})$. Many cases can be ruled out by simple distances considerations: for instance, s_2 can never be a witness for $\mathrm{UN}(s_1, z)$, because then $\|s_2 - z\|$ would have to be both $2r - o(1)$ and $r - o(1)$. There are essentially two different possible cases, which we are covered by the two representative cases below.

Case 1. s_2 is the witness for $\mathrm{UN}(x_1, s_1)$. We bound the probability as follows.

- s_1 satisfies $H_{x_1 s_1 \in [0, h_n]}$ with probability $O(h_n)$.
- Given that, s_2 is inside $L_{x_1 s_1}$ with probability $O(h_n^{3/2})$ (the area of the lens).
- Given that, z satisfies $H_{s_1 z}, H_{s_2 z} \in [0, h_n]$ with probability $O(h_n^2)$ by Lemma A.13 part 1.

- Given that z also satisfies the CD-relations, s_3 satisfies $H_{x_3 s_3}, H_{s_3 z} \in [0, h_n]$ with probability $O(h_n^2)$ by the same lemma.

Overall we get a factor of

$$O\left(n^4 h_n^{6\frac{1}{2}}\right) = (\ln n)^{O(1)} \cdot n^{-1/3}.$$

Case 2. x_2 is the witness for $\text{UN}(x_1, s_1)$. We show that the requirements force z to be inside a geometric locus with very small area, and handle s_1, s_2, s_3 like in the standard way. Indeed, the requirements include

$$x_2 \in L_{x_1 s_1}, \|x_1 - s_1\|, \|s_1 - z\| \in [2r - 2h_n, 2r].$$

It is not hard to see that such distances force s_1 to be at distance $O(n^{-1/3})$ from the point a such that x_2 is the midpoint of $x_1 a$. Therefore z to be at distance $2r + O(n^{-1/3})$ from this a. So:

- z satisfies the CD-relations and also keeps a distance $2r + O(n^{-1/3})$ from a with probability $O(n^{-1/3})$.
- Each of s_1, s_2, s_3 contributes $O(h_n^2)$ as usual.

Overall we get a factor of

$$O\left(n^4 \cdot n^{-1/3} \cdot h_n^6\right) = (\ln n)^{O(1)} \cdot n^{-1/3}.$$

Now expose the second quadruplet $\mathbf{q}' = (s_1', s_2', s_3', z')$. Denote the corresponding triplet $\mathbf{x}' = (x_1', x_2', x_3')$. Again, there are several ways in which **3a** or **3b** can occur, and we handle two representative cases.

Case 1. s_3 is the witness for $\text{UN}(x_1', s_1')$. Continue like Case 2 for $\mathbf{q}$ (note that the position of s_3 is considered fixed because $\mathbf{q}$ was already exposed).

Case 2. x_3 is the witness for $\text{UN}^*(s_1', z')$. Again, we show how it forces z' to be inside a very small geometric locus. Here,

$$x_3 \in L_{s_1' z'}, \|s_1' - z'\| \in [2r - 2h_n, 2r]$$

force $\|z' - x_3\| = 2r + O(n^{-1/3})$, which puts z' in a locus with area $O(n^{-1/3})$. So again we get

$$O\left(n^4 \cdot n^{-1/3} \cdot h_n^6\right) = (\ln n)^{O(1)} \cdot n^{-1/3}.$$

■

Finally, we are left with situation **4** from Definition 5.6, in which two different UN-relations have the same witness.

Proposition A.12. *The expected number of* $\mathbf{k}$*-tuples* $\mathbf{Q}$ *of witnesses for the* S*-extension such that situation 4 occurs is* $n^{-\Omega(1)}$.

Proof. Once more, we count only $\mathbf{k}$-tuples $\mathbf{Q}$ in which all $4k$ vertices are distinct and replace each $S(\mathbf{x};\mathbf{q})$ with $S_{\text{rel}}(\mathbf{x};\mathbf{q})$. Moreover, we can disregard $\mathbf{k}$-tuples $\mathbf{Q}$ which are not well-spaced, and also disregard the situation in which a witness for one of the $6k$ relevant UN-relations is already a vertex from $\mathbf{Q}$ or from one of the triplets $\mathbf{x}^{(i)}$.

As usual, we expose the k quadruplets of $\mathbf{Q}$ one by one. This time, however, we bound the factor that each one contributes by the general bound $O(n^4 h_n^6) = (\ln n)^{O(1)}$. So, after exposing all the quadruplets we get a bound of $(\ln n)^{O(k)} = n^{o(1)}$. The $n^{-\Omega(1)}$ factor now comes *after* the exposure of all the quadruplets, by considering the witnesses for the UN-relations.

Take the $6k$ lenses of the $6k$ pairs of vertices which are required to satisfy a UN-relation. Let us denote them

$$L_1, L_2, \ldots, L_{6k}.$$

The UN-relations are satisfied when each of those lenses contains exactly one vertex. Situation **4** occurs when one vertex is contained in two of those lenses. Define I to be the union of all intersections between two lenses:

$$I = \bigcup_{1\le i,j\le 6k,\ i\neq j} (L_i \cap L_j).$$

Then, the probability of situation **4** is bounded by the probability that I contains a vertex. As explained, we may ignore vertices from $\mathbf{Q}$ or from the triplets $\mathbf{x}^{(i)}$, and look only for the remaining $n-\Theta(k)$ vertices in I (the vertices which are not already exposed). We also explained why we may assume that the vertices of $\mathbf{Q}$ are well-spaced. Now, Lemma A.15 tells us that the area of each intersection $L_i \cap L_j$ is $n^{-1-\Omega(1)}$. A simple union bound over $O(k^2)$ intersections shows that $\text{area}(I) = n^{-1-\Omega(1)}$. Now, a union bound over the $n-\Theta(k)$ remaining vertices shows that the probability of having at least one vertex in I is

$$(n-\Theta(k))\, n^{-1-\Omega(1)} = n^{-\Omega(1)}.$$

Overall, situation **4** does contribute a factor of $n^{-\Omega(1)}$ to the expectation, and that finishes the proof. ■

The Geometric Lemmas

We state the geometric lemmas that we used and briefly explain their proofs.

Lemma A.13. *Let* $a_1, a_2 \in \mathbb{T}^2$ *be two points and denote* $\Delta = \|a_1 - a_2\|$. *Let* A *be the area of the intersection between the two annuli with radii* $2r - 2h_n, 2r$ *around* a_1 *and* a_2.

1. *If* $C_1 r \le \Delta \le C_2 r$ *for constants* $0 < C_1 < C_2 < 4$ *then* $A = O\left(h_n^2\right)$.
2. *If* $n^{-1/6} \le \Delta \le Cr$ *for a constant* $C < 4$ *then* $A = O\left(n^{1/6} h_n^2\right)$.
3. *If* $\Delta \ge n^{-1/6}$ *then* $A = O\left(h_n^{3/2}\right)$.

Before the proof, let us provide some geometric intuition. For part 1, the the condition on Δ forces the two annuli to intersect with an angle α which is bounded away from 0 and from π. So their intersection is asymptotically equivalent to the intersection of two "strips" of width $2h_n$ with angle α, which is

$$\frac{(2h_n)^2}{\sin\alpha} = O(h_n^2).$$

For part 2, note that now the condition on Δ only assures $\alpha = \Omega\left(n^{-1/6}\right)$ and that α is bounded away from π. So now the area is asymptotically equivalent to

$$\frac{(2h_n)^2}{\sin\alpha} = O\left(n^{1/6} h_n^2\right).$$

For part 3, the idea is that the maximal area is obtained for $\Delta = 4r - 2h_n$, as the intersection turns into a lens with width $2h_n$. Its area is then $O(h_n^{3/2})$.

Let us present a more formal approach.

Proof of Lemma A.13. We can replace $\mathbb{T}^2$ with $\mathbb{R}^2$ because the local behavior is Euclidean. Let us first tackle a more general problem: take any $a_1, a_2 \in \mathbb{R}^2$ and let $\Delta = \|a_1 - a_2\|$. Consider the annulus with center a_1 and radii $\alpha_1 < \beta_1$ and the annulus with center a_2 and radii $\alpha_2 < \beta_2$. What is the area of their intersection?

The idea is to define new coordinates based on the two distances from a_1 and from a_2. WLOG assume that they are the points $\left(\pm\frac{\Delta}{2}, 0\right)$. Let $\Phi_\Delta(r_1, r_2)$ be the function that receives $r_1, r_2 > 0$ and returns the unique point b in the upper half plane with $\|b - a_1\| = r_1$ and $\|b - a_2\| = r_2$ (if one exists). Since the function transforms a rectangle into half an intersection of annuli around a_1 and a_2, change of variables gives us that the area is

$$2 \iint\limits_{[\alpha_1,\beta_1]\times[\alpha_2,\beta_2]} |\det D\Phi_\Delta| \, \mathrm{d}r_1 \mathrm{d}r_2.$$

By straightforward calculations, we can show that

$$\Phi_\Delta(r_1, r_2) = (x, y) = \left(\frac{r_1^2 - r_2^2}{2\Delta}, \sqrt{\frac{r_1^2 + r_2^2}{2} - \frac{\Delta^2}{4} - \frac{(r_1^2 - r_2^2)^2}{4\Delta^2}}\right),$$

$$|\det D\Phi_\Delta| = \frac{r_1 r_2}{\Delta y}.$$

Let us return to our case. The area of the intersection is

$$2 \iint\limits_{[2r-2h_n,2r]^2} \frac{r_1 r_2}{\Delta y} \mathrm{d}r_1 \mathrm{d}r_2$$

so it only remains to bound $\frac{r_1 r_2}{\Delta y}$. It is not hard to show that for constants $0 < C_1 < C_2 < 4$,

1. $C_1 r \leq \Delta \leq C_2 r$ implies $\frac{r_1 r_2}{\Delta y} \leq \mathrm{Const}$.
2. $n^{-1/6} \leq \Delta \leq C_2 r$ implies $\frac{r_1 r_2}{\Delta y} \leq \mathrm{Const} \cdot n^{1/6}$.
3. $C_1 r \leq \Delta \leq 4r$ implies $\frac{r_1 r_2}{\Delta y} \leq \mathrm{Const} \cdot \frac{1}{\sqrt{h_n}}$.

Together these inequalities prove the lemma. ■

Lemma A.14. *Let s_1, s_2, s_3 be three points in $\mathbb{T}^2$ and assume that they are well-spaced. Let $A \subseteq \mathbb{T}^2$ be the area of the intersection between the three annuli around them with radii $2r - 2h_n, 2r$. Then $A = O\left(n^{1/6} h_n^2\right)$.*

Proof. If the intersection is empty then the bound holds trivially. So assume there is a point z in the intersection. s_1, s_2, s_3 all lie inside the annulus with radii $2r - 2h_n, 2r$ around z. It is easy to see that at least one of

$$\|s_1 - s_2\|, \|s_2 - s_3\|, \|s_3 - s_1\|$$

must be at most $\sqrt{3}r$. But they are all $\geq n^{1/6}$ by assumption. So we can bound the area A through Lemma A.13 part 2 and obtain the desired bound $O\left(n^{1/6} h_n^2\right)$. ■

Lemma A.15. *Let s_1, y_1, s_2, y_2 be four points in $\mathbb{T}^2$. Assume*

$$\|s_1 - y_1\|, \|s_2 - y_2\| \in [2r - 2h_n, 2r]$$

and also that s_1, s_2 are well-spaced. Then

$$\mathrm{Area}\left(L_{s_1 y_1} \cap L_{s_2 y_2}\right) = n^{-1\frac{1}{6}+o(1)}.$$

Proof Sketch. If $L_{s_1 y_1} \cap L_{s_2 y_2} = \emptyset$ then we are done. Assume that the intersection is non-empty. Let m_1, m_2 be the midpoints of $s_1 y_1, s_2 y_2$ (respectively). They are obviously contained in $L_{s_1 y_1}, L_{s_2 y_2}$ (respectively). The diameters of the lens are $O\left(\sqrt{h_n}\right)$ and so

$$\|m_1 - m_2\| = O\left(\sqrt{h_n}\right).$$

Now we can show that the angle α between $s_1 y_1$ and $s_2 y_2$ is $\alpha = \Theta(n^{-1/6})$, as described in Figure 7. Finally, we bound the area of the intersection of the lenses by the area of the intersection of the infinite strips that contain them, with the same widths, perpendicular to s_1y_1 and s_2y_2 respectively. This area is at most

$$\frac{(2h_n)^2}{\sin(\alpha)} = O\left(n^{1/6}h_n^2\right) = (\ln n)^{O(1)} \cdot n^{-1\frac{1}{6}} = n^{-1\frac{1}{6}+o(1)}.$$

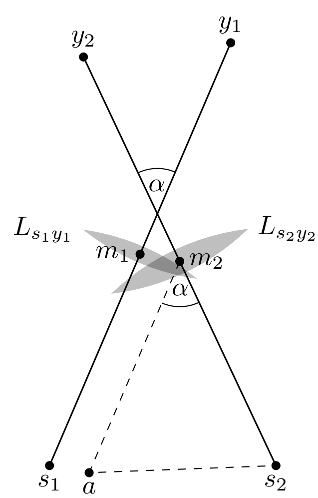


Figure 7: Illustration of the lemma. The two lenses are shaded. To show that $\alpha = \Theta(n^{-1/6})$, we can "copy" the segment s_1m_1 to a parallel segment of the same length am_2. Now consider the triangle am_2s_2. It is easily shown that $\|a_2 - m_2\|, \|s_2 - m_2\| = r - O(h_n)$ and that $\|a - s_2\| = \Theta(n^{-1/6})$. The cosine theorem then yields $\alpha = \Theta(n^{-1/6})$.

■

References

[1] József Balogh, Béla Bollobás, Michael Krivelevich, Tobias Müller, and Mark Walters. Hamilton cycles in random geometric graphs. *Annals of Applied Probability*, 21:1053–1072, 2011.

[2] Ronald Fagin. Probabilities on finite models. *The Journal of Symbolic Logic*, 41(1):50–58, 1976.

[3] Edgar Gilbert. Random plane networks. *Journal of the Society for Industrial and Applied Mathematics*, 9(4):533–543, 1961.

[4] Yu. V. Glebskii, D. I. Kogan, M. I. Liogon'kii, and V. A. Talanov. Range and degree of realizability of formulas in the restricted predicate calculus. *Cybernetics*, 5:142–154, 1969.

[5] Geoffrey R. Grimmett and David R. Stirzaker. *Probability and Random Processes*. Oxford University Press, third edition, 2001.

[6] Simi Haber and Michael Krivelevich. The logic of random regular graphs. *Journal of Combinatorics*, 1:389–440, 2010.

[7] Peter Heinig, Tobias Müller, Marc Noy, and Anusch Taraz. Logical limit laws for minor-closed classes of graphs. *Journal of Combinatorial Theory Series B*, 130:158–206, 2018.

[8] Svante Janson. New versions of suen's correlation inequality. *Random Structures and Algorithms*, 13(3-4):467–483, 1998.

[9] Svante Janson and Andzrej Ruciński. The deletion method for upper tail estimates. *Combinatorica*, 24(2):615–640, 2004.

[10] Svante Janson, Tomasz Łuczak, and Andrzej Rucinski. *Random Graphs*. Uppsala University, 2000.

[11] Jeong Han Kim and Van H. Vu. Concentration of multivariate polynomials and its applications. *Combinatorica*, 20:417–434, 2000.

[12] Robert D. Kleinberg and Jon M. Kleinberg. Isomorphism and embedding problems for infinite limits of scale-free graphs. *Society for Industrial and Applied Mathematics*, pages 277–286, 2005.

[13] Gregory McColm. First order zero-one laws for random graphs on the circle. *Random Structures and Algorithms*, 14:239–266, 1999.

[14] Colin Mcdiarmid and Tobias Müller. On the chromatic number of random geometric graphs. *Combinatorica*, 31:423–488, 2011.

[15] Tobias Müller. Two-point concentration in random geometric graphs. *Combinatorica*, 28:529–545, 2008.

[16] Matthew D. Penrose. *Random Geometric Graphs*. Oxford University Press, 2003.

[17] Joel Spencer. *The Strange Logic of Random Graphs*. Springer, 2001.

[18] Joel Spencer and Amit Agarwal. Undecidable statements and the zero one law in random geometric graphs. 2006.

[19] Joel Spencer and Saharon Shelah. Zero-one laws for sparse random graphs. *Journal of the American Mathematical Society*, 1:97–115, 1988.

[20] Tomasz Łuczak and Joel Spencer. When does the zero-one law hold? *Journal of the American Mathematical Society*, 4(3):451–468, 1991.