ON THE CLIQUE NUMBER OF NOISY RANDOM GEOMETRIC GRAPHS

MATTHEW KAHLE, MINGHAO TIAN, AND YUSU WANG

ABSTRACT. A geometric graph is a graph constructed in a metric space such that two graph vertices are connected by an edge if and only if they are within a threshold distance away from each other. *Noisy geometric networks* are geometric graphs with *non-geometric edges*; that is, the long-range edges are added to geometric graphs by chance. These graphs have many real-world applications, for example, in biological epidemics and collective social processes.

In this paper, we consider the following noisy geometric network model: First, randomly sample points $X_1, X_2, \cdots, X_n \in \mathbb{R}^d$ independently from a common probability distribution; Then, pick a threshold distance $r(n)$ and construct the geometric graph $G_n = G(X_1, X_2, \cdots, X_n; r(n))$ such that X_i and X_j are adjacent if and only if $\|X_i - X_j\| \leq r(n)$. G_n is also called a *random geometric graph*. Finally, we add random "noise" to this base geometric graph: In particular, pick real numbers $q, p \in [0, 1)$ (potentially functions of n), we construct a (q,p)*-perturbed noisy random geometric graph* $G_n^{q,p}$ where each existing edge in G_n is removed with probability q, while each non-existent edge in G_n is inserted with probability p.

We study the behavior of an important graph property called clique number of $G_n^{q,p}$, denoted by $\omega(G_n^{q,p})$. Specifically, we give asymptotically tight bounds of $\omega\left(G_n^{q,p}\right)$ under several different settings by characterizing the behavior of the so-called *edge clique numbers* (a local version of the clique number). Edge clique number itself can also be of interest, as it provides a refined view of the (global) clique number. To obtain our results, we also develop a novel approach called "well-separated clique-partitions family", which helps to decouple the mingled randomness generated from both random geometric graph and (q, p)-perturbation.

1. INTRODUCTION AND STATEMENT OF RESULTS

A geometric graph is a graph constructed in a metric space such that two graph vertices are connected by an edge if and only if they are within a threshold distance away from each other. *Noisy geometric networks* are geometric graphs with *non-geometric edges*; that is, the long-range edges are added to geometric graphs by chance. These graphs have many real-world applications, e.g., in biological epidemics and collective social processes [22]. See [13] for some recent work on the contagion dynamics on noisy geometric networks. A natural choice of the underlying geometric graph is the so-called random geometric graph $G(\mathcal{X}_n; r) = G_{\mathbb{R}^d}(X_1, X_2, \cdots, X_n; r)$ [20], where $\mathcal{X}_n = \{X_1, X_2, \cdots, X_n\}$ is a set of n independent and identically distributed d-dimensional variables sampled from a common probability distribution ν on $\mathbb{R}^d$, and an edge (X_i, X_j) is added whenever X_i and X_j are within "distance" (e.g., Euclidean distance) $r = r(n) > 0$ to each other. Random geometric graphs can be used to model real-world applications where physical locations of objects involved play an important role (e.g., wireless networks, transportation networks [2, 18]). They form an important family of objects in random graph theory, with a large literature studying their properties [20].

Date: August 22, 2019.

This work is partially supported by National Science Foundation grants DMS-1547357, DMS-1352386, CCF-1740761, and RI-1815697.

A particular important quantity associated to graphs is the so-called *clique number* $\omega(G)$, which is the size of the largest clique contained in a given graph G. The celebrated two-point concentration theorem on clique number $\omega(G(n,p))$ of Erdős–Rényi random graph $G(n,p)$ with fixed $p \in (0,1)$ was first proved by Matula [14] in 1976 using probabilistic method. Around three decades later, Penrose [20], Müller [17] and McDiarmid [15] studied the clique number $\omega\left(G\left(\mathcal{X}_n; r\right)\right)$ of random geometric graph $G(\mathcal{X}_n; r)$ for different ranges of r. For example, Müller showed that the clique number $\omega\left(G(\mathcal{X}_n; r)\right)$ satisfies a similar two-point concentration if $nr^d = o(\log n)$ [17]. However, it has dramatically different behaviors when other ranges of r are chosen. See [15] for the latest results; many of which are obtained using tools like Stein-Chen method, scan statistic and generalized scan statistic. The high dimensional case (when d is very large) was studied in [7].

In this paper, we consider the clique number of a random graph model called *noisy random geometric graph model* previously introduced in [19] (termed "ER-perturbed random geometric graphs" there), which is intuitively obtained by adding non-geometric edges and deleting geometric edges randomly to a random geometric graph. Specifically, we call $G^{q,p}(\mathcal{X}_n; r)$ a (q,p)*-perturbed noisy random geometric graph* if a (q,p)*-perturbation* is added to a random geometric graph $G(\mathcal{X}_n; r)$; that is, each edge in $G(\mathcal{X}_n; r)$ is deleted with a uniform probability q, while each "short-cut" edge between two unconnected nodes u, v is inserted to $G(\mathcal{X}_n; r)$ with a uniform probability p. To the best of our knowledge, although this random graph model is related to the continuum percolation theory [16], the understanding about them so far is still limited: In previous studies, the underlying spaces are typically the plane (called the Gilbert disc model) [4], cubes [6] or tori [10]; the vertices are often chosen as the standard lattices of the space; and the results usually concern the connectivity [5, 21] or diameter [23].

For random geometric graphs sampled from well-behaved probability distributions on nice metric spaces (e.g., from uniform distribution in Euclidean hypercubes), the clique number and the maximum degree are of the same order (Section 6 in [20]). However, this is no longer true for the *noisy* random geometric graphs, since a clique can contain vertices located everywhere in the space. Due to the existence of such arbitrary long-range edges, tools like (generalised) scan statistic cannot be easily directly applied to solve the clique number problem.

1.1. **Some definitions and notations.**

Before we state our main results, we first give the formal definition of the *random geometric graphs* $G(\mathcal{X}_n; r)$.

Definition 1 (Random geometric graph [15])**.** *Given a sequence of independent random points $X_1, X_2, \cdots$ in $\mathbb{R}^d$ sampled from a common probability distribution ν with bounded density function f (that is, for any Borel set $A \subseteq \mathbb{R}^d$, $\nu(A) = \int_A f(x)dx$), and a positive distance $r = r(n) > 0$, we construct a random geometric graph $G(\mathcal{X}_n; r)$ with vertex set $\mathcal{X}_n = \{X_1, \cdots, X_n\}$, where distinct X_i and X_j are adjacent when $\|X_i - X_j\| \leq r$. Here $\|\cdot\|$ may be any norm on $\mathbb{R}^d$.*

Denote σ as the essential supremum of the probability density function f of ν, that is

$$\sigma := \sup\left\{t : \int_{\{y: f(y) > t\}} dx > 0\right\}.$$

We call σ the maximum density of ν. Denote $B_s(x) := \{y \in \mathbb{R}^d : \|y - x\| \leq s\}$ as the ball centered at $x \in \mathbb{R}^d$ of radius s. Also set $\theta = \int_{B_1(\mathbf{o})} dx$, where $\mathbf{o}$ is the origin of $\mathbb{R}^d$; that is, θ is the volume of any radius-1 ball in $\mathbb{R}^d$.

We now introduce our (q,p)*-perturbed noisy random geometric graphs* $G^{q,p}(\mathcal{X}_n; r)$.

Definition 2 ((q,p)-perturbed noisy random geometric graph)**.** *Given a random geometric graph $G(\mathcal{X}_n; r)$ as in Definition 1, the (q,p)-perturbed noisy random geometric graph $G^{q,p}(\mathcal{X}_n; r)$ is obtained by deleting each existing edge in $G(\mathcal{X}_n; r)$ independently with probability q as well as inserting each non-existent edge in $G(\mathcal{X}_n; r)$ independently with probability p.*

Note that the order of applying the above two types of perturbations doesn't matter since they are applied to two disjoint sets respectively. This process can be applied to any graph, and we call it a *(q,p)-perturbation*. The resulting graph $G^{q,p}(\mathcal{X}_n; r)$ is called a *(q,p)-perturbation of $G(\mathcal{X}_n; r)$*, or simply a *noisy random geometric graph*.

Throughout this paper, we use the standard Bachmann-Landau notation (asymptotic notation). That is, for real valued functions $f(n)$ and $g(n)$, as $n \to \infty$, we say

(1) $f(n) = O(g(n))$: $\exists$ two constants $c > 0$ and $n_0 \in \mathbb{N}$ such that $|f(n)| \leq cg(n)$ for all $n \geq n_0$;
(2) $f(n) = o(g(n))$: $\forall \epsilon > 0$, $\exists n_0 \in \mathbb{N}$ such that $|f(n)| < \epsilon g(n)$ for all $n \geq n_0$;
(3) $f(n) = \Omega(g(n))$: $\exists$ two constants $c > 0$ and $n_0 \in \mathbb{N}$ such that $f(n) \geq cg(n)$ for all $n \geq n_0$;
(4) $f(n) = \Theta(g(n))$: $f(n) = O(g(n))$ and $f(n) = \Omega(g(n))$;

We also use the notation $f \ll g$ to mean that $f(n)/g(n) \to 0$ as $n \to \infty$, $f \lesssim g$ to mean that there exists a constant $C > 0$ such that $f(n)/g(n) < C$ for all sufficiently large n, $f \gtrsim g$ to mean that there exists a constant $c > 0$ such that $f(n)/g(n) > c$ for all sufficiently large n, and $f \sim g$ to mean that $f \lesssim g$ and $f \gtrsim g$.

Recall that a *clique* in any graph G is a set of vertices which are pairwise connected. In this paper, we use the standard notation $\omega(G)$ in graph theory to denote the *clique number* of G, which is the largest cardinality of a clique in G.

Many properties of $G(\mathcal{X}_n; r)$ are qualitatively different depending on which distance $r = r(n)$ is chosen. In some sense, the distance r here plays a role similar to the edge-inserting probability $p(n)$ in Erdős–Rényi random graphs $G(n,p)$. Following standard settings in the literature [15, 20], we consider the following three regimes of r, or more precisely, of the quantity nr^d:

I. (*"very sparse"*) $nr^d \leq n^{-\alpha}$ for some fixed $\alpha > 0$;
II. (*"quite sparse"*) $n^{-\epsilon} \ll nr^d \ll \log n$ for all $\epsilon > 0$;
III. (*"dense"*) $\sigma nr^d / \log n \to t \in (0, \infty)$;

We often use the terminology *almost surely* (or a.s.): In particular, if $\xi_1, \xi_2, \cdots$ is a sequence of random variables and $k_1, k_2, \cdots$ is a sequence of positive numbers, then "*a.s.* $\xi_n \geq k_n$" means that $\lim_{n\to\infty} \mathbb{P}[\xi_n \geq k_n] = 1$. The other direction *a.s.* $\xi_n \leq k_n$ is defined similarly. Moreover, *a.s.* $\xi_n \lesssim k_n$ means that there exist $C_1 > 0$ such that $\lim_{n\to\infty} \mathbb{P}[\xi_n \leq C_1 k_n] = 1$. Similarly, we define a.s. $\xi_n \gtrsim k_n$ and a.s. $\xi_n \sim k_n$. We also use the terminology *with high probability* (or w.h.p.): Specifically, if $A_1, A_2, \cdots$ is a sequence of events, then "A_n happens with high probability" means that $\lim_{n\to\infty} \mathbb{P}[A_n] = 1 + o(1)$.

Assumptions and notations for the remainder of the paper. In what follows, unless specified explicitly, we assume the following setting throughout, which we refer to as *the standard-setting-R*:

- The space we consider is the d-dimensional Euclidean space $\mathbb{R}^d$ with a fixed dimension d, equipped with some arbitrary norm $\|\cdot\|$ on $\mathbb{R}^d$.
- $\theta = \int_{B_1(\mathbf{o})} dx$ is the volume of the unit ball $B_1(\mathbf{o}) = \{x \in \mathbb{R}^d : \|x\| \leq 1\}$.
- β is the so-called Besicovitch constant of $(\mathbb{R}^d, \|\cdot\|)$ (see Section 2.2).

- ν is a probability distribution with finite maximum density σ; and $X_1, X_2, \cdots$ are independent random variables sampled from ν.
- $r = (r(1), r(2), \cdots)$ is a sequence of positive real numbers such that $r(n) \to 0$ as $n \to \infty$.
- q and $p = p(n)$ are real numbers between 0 and 1 (for simplicity, we only consider the case when q is a fixed constant).
- $G_n, G_n^{q,p}$ denote the *random geometric graph* $G(X_1, \cdots, X_n; r(n))$ and its (q,p)-*perturbation*, respectively.

For any graph G, let $V(G)$ and $E(G)$ refer to its vertex set and edge set, and let $N_G(u)$ denote the set of neighbors of u in G (i.e. nodes connected to $u \in V(G)$ by edges in $E(G)$). For a subset $W \subseteq \mathbb{R}^d$, we denote the number of indices $i \in \{1, \cdots, n\}$ such that $X_i \in W$ by $\mathcal{N}(W) = \mathcal{N}_n(W)$; that is, $\mathcal{N}(W)$ is the number of points from $\mathcal{X}_n = \{X_1, \ldots, X_n\}$ contained in W.

1.2. **Overview of main results.** We now state the main results of this paper, which concerns the behavior of clique number of (q,p)-perturbed noisy random geometric graphs. To understand the behavior of q-deletion and p-insertions (which have different effects on the clique numbers), we first separate the *insertion-only* case (where the perturbation only has random insertions) and the *deletion-only* case (where the perturbation only has random deletions), and present results for the two cases in Theorem 1 and 2, respectively.

Insertion only. We first consider the clique number of $G_n^{0,p}$, where no edges in G_n are removed and only new edges are added. The graph generated this way can be thought of as the union of a random geometric graph and an Erdős–Rényi random graph. Indeed, in Theorem 1 below, we show the interplay between those two random graphs as $p = p(n)$ increases in different regimes of r.

Theorem 1. *Given a $(0,p)$-perturbed noisy random geometric graph $G_n^{0,p}$ in the standard-setting-R, the following holds:*

(I) *Suppose that $nr^d \leq n^{-\alpha}$ for some fixed $\alpha \in (0, 1/\beta^2]$. Then there exist constants C_1, C_2 such that*

(I.a) *if $p \leq (1/n)^{C_1}$, then a.s.*

$$\omega\left(G_n^{0,p}\right) \sim 1,$$

(I.b) *and if $(1/n)^{C_1} < p \leq C_2$, then a.s.*

$$\omega\left(G_n^{0,p}\right) \sim \log_{1/p} n.$$

(II) *Suppose that for every $\epsilon > 0$, $n^{-\epsilon} \ll nr^d \ll \log n$. Then there exist constants C_1, C_2 such that*

(II.a) *if $p \leq \left(nr^d / \log n\right)^{C_1}$, then a.s.*

$$\omega\left(G_n^{0,p}\right) \sim \frac{\log n}{\log\left(\log n / nr^d\right)},$$

(II.b) *and if $\left(nr^d / \log n\right)^{C_1} < p \leq C_2$, then a.s.*

$$\omega\left(G_n^{0,p}\right) \sim \log_{1/p} n.$$

(III) *Suppose that $\sigma nr^d / \log n \to t \in (0, \infty)$ as $n \to \infty$. Then there exists a constant C_1 such that if $p \leq C_1$, then a.s.*

$$\omega\left(G_n^{0,p}\right) \sim nr^d.$$

For this insertion-only case, one could view the graph $G_n^{0,p}$ as the union of a random geometric graph G_n and an Erdős–Rényi random graph $G(n,p)$ on the same vertex set. The theorem above suggests that intuitively either the clique number from the random geometric graph or the one from the Erdős–Rényi random graph will dominate, depending on regimes of nr^d and p. On the surface, this may not look surprising. However, in general, the clique number of the union $G = G_1 \cup G_2$ of two graphs G_1 and G_2 could be significantly larger than the clique number in each individual graph G_i: Consider for example G_1 is a collection of $\sqrt{n}$ disjoint cliques, each of size $\sqrt{n}$, while G_2 equals to the complement of G_1. The union $G_1 \cup G_2$ is the complete graph and the clique number is n. However, the clique number of G_1 or G_2 is $\sqrt{n}$. Our results suggest that due to the randomness in each of the individual graph we are considering, with high probability such a scenario will not happen and the two types of random graph do not interact strongly.

To prove our technical results, we apply a novel approach using what we call a well-separated clique-partitions family (see section 2.2) to help us to decouple the interaction between the two types of hidden random structures (i.e, random geometric graph, and the $(0,p)$-perturbation).

Deletion only. We now present our main result for the clique number of $G_n^{q,0}$, where we only delete edges in G_n with a fixed edge-deletion probability q. We remark that technically speaking, the deletion-only case is easier to handle than the insertion-only case.

Theorem 2. *Given a $(q,0)$-perturbed noisy random geometric graph $G_n^{q,0}$ in the standard-setting-R with a fixed constant $0 < q < 1$, the following holds:*

(I) *Suppose that $nr^d \le n^{-\alpha}$ for some fixed $\alpha > 0$. Then a.s.*
$$\omega\left(G_n^{q,0}\right) \sim 1$$

(II) *Suppose that $n^{-\epsilon} \ll nr^d \ll \log n$ for all $\epsilon > 0$. Then a.s.*
$$\omega\left(G_n^{q,0}\right) \sim \log \frac{\log n}{\log(\log n / nr^d)}$$

(III) *Suppose that $\sigma nr^d / \log n \to t \in (0,\infty)$. Then a.s.*
$$\omega\left(G_n^{q,0}\right) \lesssim \log\left(nr^d\right)$$

Furthermore, there exists a constant $T > 0$ such that if $\sigma nr^d \ge T \log n$, then a.s.
$$\omega\left(G_n^{q,0}\right) \sim \log\left(nr^d\right)$$

Combined case. The above *insertion-only* and *deletion-only* cases in fact represent key technical challenges. When there are both random insertions and deletions, we can derive some bounds on the clique number by simply combining the above results and some technical lemmas later in the paper together with the monotone property of clique number. For example, we have the following result when nr^d is in the "very sparse" regime.

Corollary 3. *Given a (q,p)-perturbed noisy random geometric graph $G_n^{q,p}$ in the standard-setting-R with a fixed constant $0 < q < 1$ and suppose that $nr^d \le n^{-\alpha}$ for some fixed $\alpha \in (0, 1/\beta^2]$, then there exist constants C_1, C_2 such that*

a) *if $p \le (1/n)^{C_1}$, then a.s.*
$$\omega\left(G_n^{q,p}\right) \sim 1$$

b) *and if $(1/n)^{C_1} < p \le C_2$, then a.s.*
$$\omega\left(G_n^{q,p}\right) \sim \log_{1/p} n$$

The complete list of results can be found in Theorem 26 of Section 5.

2. Preliminaries and well-separated clique-partitions family

In this section, we first state in Section 2.1 some existing results / tools that we will use frequently throughout the paper. We will then define in Section 2.2 a new object called *well-separated clique-partitions family* which we will need later in the arguments.

2.1. **Some standard results and tools.** For the proofs in this paper, we need some bounds on the binomial and Poisson distributions.

Lemma 4 (Lemma 3.6 in [15]). *Let Z be either binomial or Poisson and $k \geq \mu := \mathbb{E}[Z]$. Then*

$$\left(\frac{\mu}{ek}\right)^k \leq \mathbb{P}[Z \geq k] \leq \left(\frac{e\mu}{k}\right)^k$$

Lemma 5 (Chernoff-Hoeffding theorem [20]). *Suppose $n \in \mathbb{N}, \alpha \in (0,1)$ and $0 < k < n$. Let $X \sim Bin(n, \alpha)$ be either a binomial random variable with mean $\mu = n\alpha$ or $X \sim Poisson(\mu)$ be a Poisson random variable with mean $\mu > 0$. If $k \geq \mu$, then*

$$\mathbb{P}[X \geq k] \leq \exp\left(-\mu H\left(\frac{k}{\mu}\right)\right)$$

where $H : [0, \infty] \to [0, \infty)$ is a function defined by $H(0) = 1$ and $H(a) = 1 - a + a \log a$.

One object we use frequently in our proofs is the following *(generalised) scan statistics* defined in [15]. Recall that for any set $U \subseteq \mathbb{R}^d$, $\mathcal{N}(U)$ is the number of points from $\mathcal{X}_n = \{X_1, \ldots, X_n\}$ contained in U.

Definition 3 ((Generalised) scan statistic [15]). *For any set $W \subseteq \mathbb{R}^d$, we define M_W by $M_W := \max_{x \in \mathbb{R}^d} \mathcal{N}(x + rW)$ where $rW = \{rw : w \in W\}$ is the scaled set of W.*

In other words, M_W is the maximum number of points in $\mathcal{X}_n$ in any translate of rW.

2.2. **Well-separated clique-partitions family.** This section discuss one main technique we will use to bound the clique number of $G_n^{q,p}$ from above. The main challenge here is to disentangle two types of randomness — the location of vertices and the (q, p)-perturbation. In particular, our model allows vertices even far away to each other to become connected. To solve this, we develop a novel approach using what we call a *well-separated clique-partitions family* (to be defined shortly) to help us to decouple the interaction between these two types of hidden random structures.

To set up the stage, we first recall the *Besicovitch covering lemma* which has a lot of applications in measure theory [8].

Definition 4 (Packings, covers, and partitions). *(1) A* packing *is a countable collection $\mathcal{B}$ of* pairwise disjoint *closed balls in $\mathbb{R}^d$. Such a collection $\mathcal{B}$ is a* packing w.r.t. a set $P \subset \mathbb{R}^d$ *if the centers of the balls in $\mathcal{B}$ lie in the set P, and it is a* δ-packing *if all of the balls in $\mathcal{B}$ have radius δ.*

(2) A set $\{A_1, \ldots, A_\ell\}$, $A_i \subseteq \mathbb{R}^d$, covers *$P$ if $P \subseteq \bigcup_i A_i$.*

(3) Given a set A, we say that A is partitioned into *$A_1, A_2, \cdots, A_k$, if $A = A_1 \cup \cdots \cup A_k$ and $A_i \cap A_j = \emptyset$ for any $i \neq j$.*

Lemma 6 (Besicovitch Covering Lemma [9]). *There exists a constant $\beta = \beta(d) \in \mathbb{N}$ such that for any set $P \subset \mathbb{R}^d$ and $\delta > 0$, there are β number of δ-packings w.r.t. P, denoted by $\{\mathcal{B}_1, \cdots, \mathcal{B}_\beta\}$, whose union also covers P.*

We call the constant β above the *Besicovitch constant*. Note that β depends only on the dimension d and is **not** dependent of δ.

Definition 5 (Well-separated clique-partitions family). *Given a geometric graph G^* in $(\mathbb{R}^d, \|\cdot\|)$ with vertex set V and edge set E, a family $\mathcal{P} = \{P_i\}_{i\in\Lambda}$, where $P_i \subseteq V$ and Λ is the index set of P_is, forms a* well-separated clique-partitions family *of G^* if:*

(1) $V = \cup_{i\in\Lambda} P_i$.

(2) $\forall i \in \Lambda$, *P_i can be partitioned as $P_i = C_1^{(i)} \sqcup C_2^{(i)} \sqcup \cdots \sqcup C_{m_i}^{(i)}$ where*

(2-a) $\forall j \in [1, m_i]$, *there exist $\bar{v}_j^{(i)} \in V$ such that $C_j^{(i)} \subseteq B_{r/2}\left(\bar{v}_j^{(i)}\right) \cap V$.*

(2-b) *For any $j_1, j_2 \in [1, m_i]$ with $j_1 \neq j_2$, $d_H\left(C_{j_1}^{(i)}, C_{j_2}^{(i)}\right) > r$, where d_H is the Hausdorff distance between two sets in $\mathbb{R}^d$ with respect to norm $\|\cdot\|$.*

We also call $C_1^{(i)} \sqcup C_2^{(i)} \sqcup \cdots \sqcup C_{m_i}^{(i)}$ a clique-partition of P_i (w.r.t. G^*)*, and its* size *(cardinality) is m_i. The* size *of the well-separated clique-partitions family $\mathcal{P}$ is its cardinality $|\mathcal{P}| = |\Lambda|$.*

In the above definition, (2-a) implies that each $C_j^{(i)}$ spans a clique in the geometric graph G^*; thus we call $C_j^{(i)}$ as a *clique* in P_i and $C_1^{(i)} \sqcup C_2^{(i)} \sqcup \cdots \sqcup C_{m_i}^{(i)}$ a clique-partition of P_i. (2-b) means that there are no edges in G^* between any two cliques of P_i; thus later, any edge in its corresponding (q,p)-perturbation $G_n^{q,p}$ between such cliques must come from (q,p)-perturbation (p-insertion). See figure 1.

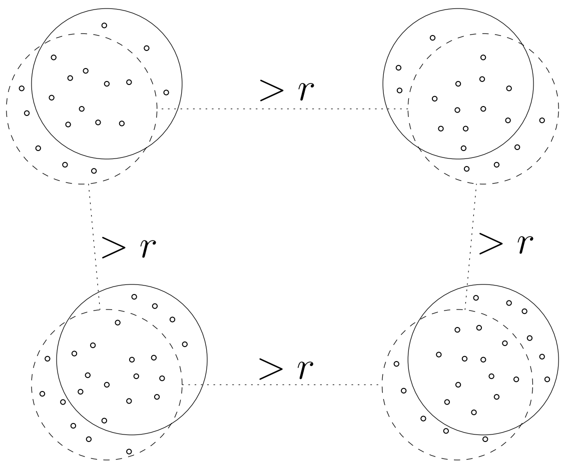

FIGURE 1. Points in the solid balls are P_1, and those in dashed balls are P_2. Each adapts a clique-partition of size $m_1 = m_2 = 4$. Assuming that all nodes in G^* are shown in this figure, then $\mathcal{P} = \{P_1, P_2\}$ forms a well-separated clique-partitions family of G^*.

We have the following existence theorem of a well-separated clique-partitions family with constant size only depending on the dimension d.

Theorem 7. *There exists a well-separated clique-partitions family $\mathcal{P} = \{P_i\}_{i\in\Lambda}$ of any geometric graph G^* with $|\Lambda| \leq \beta^2$, where $\beta = \beta(d)$ is the Besicovitch constant of $\mathbb{R}^d$.*

Proof. To prove the theorem, first imagine we grow an $r/2$-ball around each node in $V \subset \mathbb{R}^d$ (the vertex set of G^*). By Besicovitch Covering Lemma (Lemma 6), we have a family of $(r/2)$-packings w.r.t. V, $\mathcal{B} = \{\mathcal{B}_1, \mathcal{B}_2, \cdots, \mathcal{B}_{\alpha_1}\}$, whose union covers V. Here, the constant α_1 satisfies $\alpha_1 \leq \beta(d)$.

Each $\mathcal{B}_i$ contains a collection of disjoint $r/2$-balls centered at a subset of nodes in V, and let $V_i \subseteq V$ denote the centers of these balls. For any $u, v \in V_i$, we have $\|u - v\| > r$ as otherwise, $B_{r/2}(u) \cap B_{r/2}(v) \neq \emptyset$ meaning that the $r/2$-balls in $\mathcal{B}_i$ are not all pairwise disjoint. Now consider the collection of r-balls centered at all nodes in V_i. Applying Besicovitch Covering Lemma to V_i again with $\delta = r$, we now obtain a family of r-packings w.r.t. V_i, denoted by $\mathcal{D}^{(i)} = \mathcal{D}_1^{(i)} \sqcup \cdots \sqcup \mathcal{D}_{\alpha_2^{(i)}}^{(i)}$, whose union covers V_i. Here, the constant $\alpha_2^{(i)}$ satisfies $\alpha_2^{(i)} \leq \beta(d)$ for each $i \in [1, \alpha_1]$.

Now each $\mathcal{D}_j^{(i)}$ contains a set of disjoint r-balls centered at a subset of nodes $V_j^{(i)} \subseteq V_i$ of V_i. First, we claim that $\bigcup_j V_j^{(i)} = V_i$. This is because that $\mathcal{B}_i$ is an $r/2$-packing which implies that $\|u - v\| > r$ for any two nodes $u, v \in V_i$. In other words, the r-ball around any node from V_i contains no other nodes in V_i. As the union of r-balls $\mathcal{D}_1^{(i)} \sqcup \cdots \sqcup \mathcal{D}_{c_2^{(i)}}^{(i)}$ covers V_i by construction, it is then necessary that each node V_i has to appear as the center in at least one $\mathcal{D}_j^{(i)}$ (i.e, in $V_j^{(i)}$). Hence $\bigcup_j V_j^{(i)} = V_i$.

Now for each vertex set $V_j^{(i)}$, let $P_j^{(i)} \subseteq V$ denote all points from V contained in the $r/2$-balls centered at points in $V_j^{(i)}$. As $\bigcup_j V_j^{(i)} = V_i$, we have $\bigcup_j P_j^{(i)} = \bigcup_{v \in V_i} \left(B_{r/2}(v) \cap V\right)$. It then follows that $\bigcup_{i \in [1, \alpha_1]} \left(\bigcup_{j \in [1, \alpha_2^{(i)}]} P_j^{(i)} \right) = V$ as the union of the family of $r/2$-packings $\mathcal{B} = \{\mathcal{B}_1, \mathcal{B}_2, \cdots, \mathcal{B}_{c_1}\}$ covers all points in V (recall that $\mathcal{B}_i$ is just the set of $r/2$-balls centered at points in V_i).

Clearly, each $P_j^{(i)}$ adapts a clique-partition: Indeed, for each $V_j^{(i)}$, any two nodes in $V_j^{(i)}$ are at least distance $2r$ apart (as the r-balls centered at nodes in V_j^i are disjoint), meaning that the $r/2$-balls around them are more than r (Hausdorff-)distance away. In other words, $\mathcal{P} = \left\{ P_j^{(i)}, i \in [1, \alpha_1], j \in [1, \alpha_2^{(i)}] \right\}$ forms a well-separated clique-partitions family of G^*. Finally, since $\alpha_1, \alpha_2^{(i)} \leq \beta(d) = \beta$, the cardinality of $\mathcal{P}$ is thus bounded by β^2. □

3. Proof of Theorem 1

In this section, we focus on estimating the order of $\omega\left(G_n^{0,p}\right)$, the clique number of $G_n^{0,p}$. Note that for any set $W \subseteq \mathbb{R}^d$, the generalised scan statistic M_W (see Definition 3) is the maximum number of points in the vertex set $\mathcal{X}_n = \{X_1, X_2, \cdots, X_n\}$ in any translate of rW. Set $W_{1/2} := B_{1/2}(\mathbf{o})$ and $W_1 := B_1(\mathbf{o})$ where $\mathbf{o}$ is the origin. It is obvious that $\omega\left(G_n^{0,p}\right) \geq M_{W_{1/2}}$. Thus, the lower bound can be directly derived by using the results related to the generalised scan statistic in [15]. However, getting an upper bound is much more challenging, since unlike $\omega(G_n) \leq M_{W_2}$ holds in G_n, the vertices in a clique of $G_n^{0,p}$ can come from everywhere in the space.

3.1. **Proof of Part (I) — "very sparse" regime.** In this section, we discuss the order of $\omega(G_n^{0,p})$ in the regime $nr^d \leq n^{-\alpha}$ for some fixed $\alpha \in (0, 1/\beta^2]$. First, we define the *long-edges* in $G_n^{0,p}$.

Definition 6 (long-edges)**.** *An edge (u, v) in a $(0, p)-$perturbed noisy random geometric graph $G_n^{0,p}$ is a* long-edge *if and only if $\|u - v\| > 3r$.*

Cliques in $G_n^{0,p}$ can be classified into the following two types:

- Type-I clique: doesn't contain any long-edges
- Type-II clique: contains at least one long-edge

In what follows, we derive upper bounds for each type of cliques separately. The lower bounds are easier to derive and can be found later in Section 3.1.2.

Type-I cliques. The case of Type-I cliques is rather simple to handle: Note that vertices of one Type-I clique are contained within a ball of radius $3r$ centered at some vertex $X_i \in \mathcal{X}_n$. Thus, to bound the size of such clique from above, it suffices to estimate the number of vertices in each of the n number of $3r$-ball centered at some vertex in $\mathcal{X}_n$. Set $W_3 := B_3(\mathbf{o})$. We have the following lemma, which gives a uniform upper bound of number of vertices in any $3r-$ball. It is a simplified variant of Lemma 3.8 in [15]. We include its simple proof for completeness.

Lemma 8. *If* $nr^d \leq n^{-\alpha}$ *then* $\mathbb{P}\left[M_{W_3} \leq \lceil 4/\alpha \rceil\right] = 1 + O(n^{-3})$.

Proof. For some fixed integer k, we have the following inequality.

$$\mathbb{P}\left[M_{W_3} \geq k+1\right] \leq \mathbb{P}\left[\exists i : \mathcal{N}\left(B_{6r}(X_i)\right) \geq k+1\right] \;\leq\; n\mathbb{P}\left[\mathcal{N}\left(B_{6r}(X_1)\right) \geq k+1\right]$$

Furthermore, note that

$$\mathbb{P}\left[\mathcal{N}\left(B_{6r}(X_1)\right) \geq k+1\right] \leq \mathbb{P}\left[Bin\left(n, \sigma\theta(6r)^d\right) \geq k\right] \;\leq\; \left(\frac{e\sigma\theta 6^d (nr)^d}{k}\right)^k \;=\; O(n^{-k\alpha})$$

Recall that σ is the maximum density of ν and $\theta = \int_{B_1(\mathbf{o})} dx$ are introduced in the standard setting-R at the end of Section 1.1. The second inequality holds due to Lemma 4. Now pick $k = \lceil 4/\alpha \rceil$. We then have $\mathbb{P}\left[M_{W_3} \leq \lceil 4/\alpha \rceil\right] = 1 + O(n^{-3})$ as required. □

It then follows that the size of Type-I cliques can be bounded from above by $\lceil 4/\alpha \rceil$ almost surely.

Type-II cliques. Now let's consider the Type-II cliques, which is significantly more challenging to handle. Recall that $W_1 = B_1(\mathbf{o})$. We can use the same argument in Lemma 8 to derive the following lemma which gives a uniform upper bound of the number of points in any r-ball.

Lemma 9. *If* $nr^d \leq n^{-\alpha}$ *then* $\mathbb{P}\left[M_{W_1} \leq \lceil 4/\alpha \rceil\right] = 1 + O\left(n^{-3}\right)$.

We now introduce a local version of the clique number called *edge clique number*.

Definition 7 (Edge clique number)**.** *Given a graph* $G = (V, E)$*, for any edge* $(u, v) \in E$*, its* edge clique number $\omega_{u,v}(G)$ *is defined as*

$$\omega_{u,v}(G) = \text{ the largest size of any clique in } G \text{ containing } (u, v).$$

We are now ready to bound the size of all the type-II cliques in $G_n^{0,p}$. More precisely, the following theorem first bound the edge clique number for all *long edge* (u, v).

Theorem 10. *Given an* $(0, p)$*-perturbed noisy random geometric graph* $G_n^{0,p}$ *in the standard-setting-R and suppose that* $nr^d \leq n^{-\alpha}$ *for some fixed* $\alpha \in (0, 1/\beta^2]$*, then*

(a) *There exist constants* $C_1, C_2 > 0$ *which depend on the Besicovitch constant* β *and* α *such that if*

$$p \leq C_1 \left(1/n\right)^{C_2} \tag{3.1}$$

then, with high probability, for all long-edge (u, v) *in* $G_n^{0,p}$*, its edge clique number* $\omega_{u,v}(G_n^{0,p}) \lesssim 1$.

(b) *There exists a constant* $\xi > 0$ *which depends on the Besicovitch constant* β *and* α *such that if* $(1/n)^{\xi} \leq p < 1$*, then, with high probability, for all long-edge* (u, v) *in* $G_n^{0,p}$*, its edge clique number* $\omega_{u,v}(G_n^{0,p}) \lesssim \log_{1/p} n$.

3.1.1. *Proof of Theorem 10.*

Proof of part (a) of Theorem 10. Given any vertex y, let $B_r^{\mathcal{X}_n}(y) \subseteq \mathcal{X}_n$ denote $B_r(y) \cap \mathcal{X}_n$. Now consider a long-edge (u, v). Set $A_{uv} = \mathcal{X}_n \backslash \big(B_r^{\mathcal{X}_n}(u) \cup B_r^{\mathcal{X}_n}(v)\big)$ and $B_{uv} = B_r^{\mathcal{X}_n}(u) \cup B_r^{\mathcal{X}_n}(v)$. Denote $\tilde{A}_{uv} = A_{uv} \cup \{u\} \cup \{v\}$; easy to check that $V = \tilde{A}_{uv} \cup B_{uv}$.

Let $G|_S$ denote the subgraph of G spanned by a subset S of its vertices. Given any set C, let $\mathrm{C}|_S = \mathrm{C} \cap S$ be the restriction of C to another set S. Now consider a subset of vertices $\mathrm{C} \subseteq \mathcal{X}_n$: obviously, $\mathrm{C} = \mathrm{C}|_{\tilde{A}_{uv}} \cup \mathrm{C}|_{B_{uv}}$.

Set $N_{\max} := \lceil 4/\alpha \rceil$. Denote F to be the event that "for every $v \in \mathcal{X}_n$, the ball $B_r(v) \cap \mathcal{X}_n$ contains at most $N_{\max}$ points"; and F^c denotes the complement event of F. By Lemma 9, we know that, $\mathbb{P}[\mathrm{F}^c] = O(n^{-3})$.

Let $\mathsf{K} \geq 8\beta^2$ be an integer to be determined. By applying the pigeonhole principle and the union bound, we have:

$$
\begin{aligned}
&\mathbb{P}\left[\omega_{u,v}\left(G_n^{0,p}\right) \geq \mathsf{K} \middle| \mathrm{F}\right] \\
&\leq \mathbb{P}\left[\omega_{u,v}\left(G_n^{0,p}|_{\tilde{A}_{uv}}\right) \geq \mathsf{K}/2 \middle| \mathrm{F}\right] + \mathbb{P}\left[\omega_{u,v}\left(G_n^{0,p}|_{B_{uv}}\right) \geq \mathsf{K}/2 \middle| \mathrm{F}\right] \qquad (3.2)
\end{aligned}
$$

Next, we bound the two terms on the right hand side of Eqn. (3.2) separately in Case (i) and Case (ii) below.

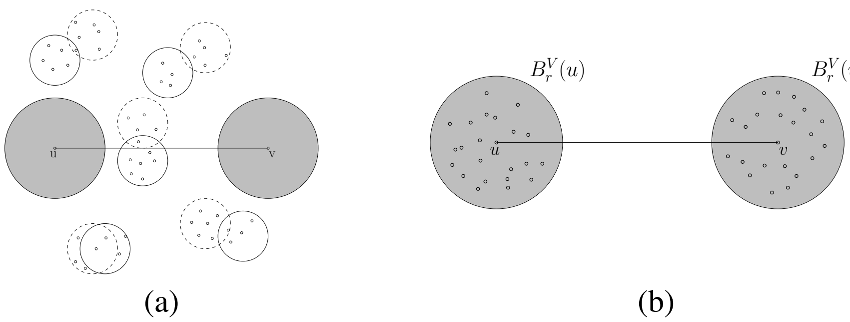

FIGURE 2. (a) A well-separated clique partition $\mathcal{P} = \{P_1, P_2\}$ of A_{uv} — points in the solid balls are P_1, and those in dashed balls are P_2. (b) Points in B_{uv}.

Case (i): bounding the first term in Eqn. (3.2). We apply Theorem 7 for points in A_{uv}. This gives us a well-separated clique-partitions family $\mathcal{P} = \{P_i\}_{i\in\Lambda}$ of A_{uv} with $|\Lambda| \leq \beta^2$ being a constant. See Figure 2 (a). Augment each P_i to $\tilde{P}_i = P_i \cup \{u\} \cup \{v\}$. Suppose there is a clique C in $G_n^{0,p}|_{\tilde{A}_{uv}}$, then as $\bigcup_i \tilde{P}_i = \tilde{A}_{uv}$, we have $C = \bigcup_{i\in\Lambda} C|_{\tilde{P}_i}$, implying that $|C| \leq \sum_{i\in\Lambda} \left|C|_{\tilde{P}_i}\right|$. Hence by applying pigeonhole principle and the union bound, we derive the following inequality:

$$
\mathbb{P}\left[\omega_{u,v}\left(G_n^{0,p}|_{\tilde{A}_{uv}}\right) \geq \mathsf{K}/2 \middle| \mathrm{F}\right] \;\leq\; \sum_{i=1}^{|\Lambda|} \mathbb{P}\left[\omega_{u,v}\left(G_n^{0,p}|_{\tilde{P}_i}\right) \geq \mathsf{K}/(2|\Lambda|) \middle| \mathrm{F}\right] \qquad (3.3)
$$

Now for arbitrary $i \in \Lambda$, consider $G_n^{0,p}|_{\tilde{P}_i}$, the induced subgraph of $G_n^{0,p}$ spanned by vertices in $\tilde{P}_i$. Note, $G_n^{0,p}|_{\tilde{P}_i}$ can be viewed as generated by inserting each edge not in $G_n|_{\tilde{P}_i} \cup \{(u, v)\}$ with probability p. Recall from Definition 5 that each P_i adapts a clique-partition $C_1^{(i)} \sqcup \cdots \sqcup C_{m_i}^{(i)}$, where every $C_j^{(i)}$ is contained in an $r/2$-ball, and all such balls are r-separated (w.r.t Hausdorff distance).

Now fix any $i \in \Lambda$. For simplicity of the argument below, set $m = m_i$, and let $N_j = \left|C_j^{(i)}\right|$ denote the number of points in the j-th cluster $C_j^{(i)}$. Note that obviously, $m \leq |P_i| \leq |V| = n$ for any $i \in \Lambda$. We also know that if event F has already happened, then $N_j \leq N_{\max}$.

Observe that the induced subgraph $G_n^{0,p}|_{\tilde{P}_i}$ consists of a set of cliques (each clique is spanned by some $C_j^{(i)}$ with edges coming from the base random geometric graph G_n), u, v, edge (u, v), and inserted edge between them with insertion probability p (see Figure 3).

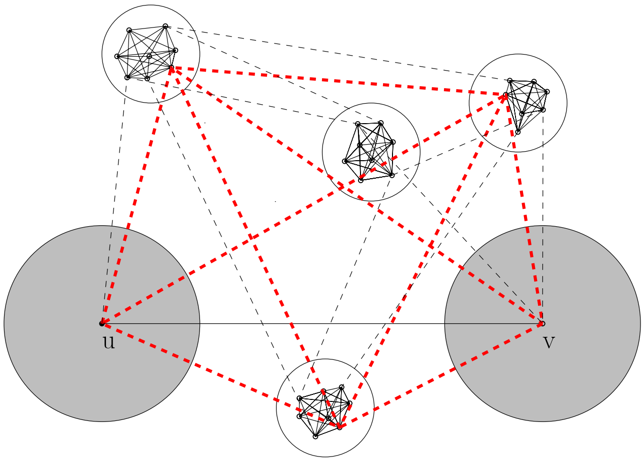

FIGURE 3. The red dashed lines and the edge uv form a possible clique in some well-separated clique partition P_i. The points in the small balls are the nodes falling in $r/2-$balls (and thus they are all pairwise connected in the base random geometric graph G_n). All the dashed lines are the randomly inserted edges (independently with probability p).

Now set $k := \lfloor \mathsf{K}/2|\Lambda| \rfloor - 2$. Since $\mathsf{K} \geq 8\beta^2$, easy to see that $k \geq 1$. For every set S of $k+2$ vertices in this graph $G_n^{0,p}|_{\tilde{P}_i}$, let A_S be the event "*S is clique in $G_n^{0,p}|_{\tilde{P}_i}$ containing (u, v) given* F" and I_S its indicator random variable. Set

$$\mathrm{I} = \sum_{|S|=k+2} \mathrm{I}_S$$

and note that I is the number of cliques of size $(k+2)$ in $G_n^{0,p}|_{\tilde{P}_i}$ containing (u, v) given F. It follows from Markov inequality that:

$$\mathbb{P}\left[\omega_{u,v}\left(G_n^{0,p}|_{\tilde{P}_i}\right) \geq k+2 \middle| \mathrm{F}\right] = \mathbb{P}[\mathrm{I} > 0] \leq \mathbb{E}[\mathrm{I}] \tag{3.4}$$

On the other hand, using linearity of expectation, we have:

$$\mathbb{E}[\mathrm{I}] = \sum_{|S|=k+2} \mathbb{E}[\mathrm{I}_S] = p^{2k} \sum_{\substack{x_1+x_2+\cdots+x_m=k \\ 0\leq x_i \leq N_i}} \binom{N_1}{x_1}\binom{N_2}{x_2}\cdots\binom{N_m}{x_m} p^{(k^2-\sum_{i=1}^m x_i^2)/2}$$

$$\leq p^{2k} \sum_{\substack{x_1+x_2+\cdots+x_m=k \\ 0\leq x_i \leq N_{\max}}} \binom{N_{\max}}{x_1}\binom{N_{\max}}{x_2}\cdots\binom{N_{\max}}{x_m} p^{(k^2-\sum_{i=1}^m x_i^2)/2} \tag{3.5}$$

To estimate this quantity, we have the following lemma:

Lemma 11. *If $1 \leq k \leq N_{\max}$ and p is less than or equal to*

$$\min\left\{\frac{1}{\sqrt{e}}\left(\frac{1}{n^3 m}\right)^{\frac{1}{2k}}\left(\frac{k}{N_{\max}}\right)^{\frac{1}{2}}, \frac{1}{2ek^{\frac{1}{k}}}\left(\frac{1}{n^3m^2}\right)^{\frac{1}{k}}\frac{k}{N_{\max}}, \frac{1}{e^{\frac{4}{k}}}\left(\frac{1}{n^3m^k}\right)^{\frac{4}{k^2}}\left(\frac{k}{N_{\max}}\right)^{\frac{4}{k}}\right\}, \tag{3.6}$$

then we have that $\mathbb{E}[\mathrm{I}] = O(n^{-3})$.

The proof of this lemma is rather technical, and can be found in Appendix A.1.

Note that $\alpha \leq 1/\beta^2$, thus $2N_{\max} = 2\lceil 4/\alpha \rceil \geq 8\beta^2$. Note that if $\mathsf{K} \in [8\beta^2, 2N_{\max}]$, then it is easy to check that the assumption $1 \leq k \leq N_{\max}$ in Lemma 11 holds.

Furthermore, $|\Lambda| \leq \beta^2$ (which is a constant) and $m = |P_i| \leq |V| = n$. One can then verify that there exist constants c_1^a and c_2^a (which depend on the Besicovitch constant β and α), such that if

$$p \leq c_1^a \cdot (1/n)^{c_2^a/\mathsf{K}},$$

then the conditions in Eqn. (3.6) will hold. Thus, combining this with Lemma 11 and Eqn. (3.4), we know that

$$\text{If } 8\beta^2 \leq \mathsf{K} \leq 2N_{\max} \text{ and } p \leq c_1^a \cdot (1/n)^{c_2^a/\mathsf{K}},$$
$$\text{then } \forall i \in \Lambda, \mathbb{P}\left[\omega_{u,v}\left(G_n^{0,p}|_{\tilde{P}_i}\right) \geq k+2 \middle| \mathrm{F}\right] = O(n^{-3}). \tag{3.7}$$

On the other hand, note that

$$\mathbb{P}\left[\omega_{u,v}\left(G_n^{0,p}|_{\tilde{P}_i}\right) \geq \mathsf{K}/(2|\Lambda|) \middle| \mathrm{F}\right] = \mathbb{P}\left[\omega_{u,v}\left(G_n^{0,p}|_{\tilde{P}_i}\right) \geq k+2 \middle| \mathrm{F}\right]$$

As $|\Lambda|$ is a constant, by Eqn. (3.3), we obtain that

$$\text{if } \forall i \in \Lambda, \mathbb{P}\left[\omega_{u,v}\left(G_n^{0,p}|_{\tilde{P}_i}\right) \geq \mathsf{K}/(2|\Lambda|) \middle| \mathrm{F}\right] = O(n^{-3}), \text{ then}$$
$$\mathbb{P}\left[\omega_{u,v}\left(G_n^{0,p}|_{\tilde{A}_{uv}}\right) \geq \mathsf{K}/2 \middle| \mathrm{F}\right] = O(|\Lambda| n^{-3}) = O(n^{-3}). \tag{3.8}$$

It then follows from Eqn. (3.7) and (3.8) that

$$\text{If } 8\beta^2 \leq \mathsf{K} \leq 2N_{\max} \text{ and } p \leq c_1^a \cdot (1/n)^{c_2^a/\mathsf{K}},$$
$$\text{then } \mathbb{P}\left[\omega_{u,v}\left(G_n^{0,p}|_{\tilde{A}_{uv}}\right) \geq \mathsf{K}/2 \middle| \mathrm{F}\right] = O(n^{-3}). \tag{3.9}$$

Finally, suppose $\mathsf{K} > \mathsf{K}_0 = 2N_{\max}$. Using Eqn (3.9), we know that if $p \leq c_1^a \cdot (1/n)^{c_2^a/\mathsf{K}_0}$ and $\mathsf{K} > \mathsf{K}_0$ (in which case note also that $c_1^a \cdot (1/n)^{c_2^a/\mathsf{K}_0} \leq c_1^a \cdot (1/n)^{c_2^a/\mathsf{K}}$), then

$$\mathbb{P}\left[\omega_{u,v}\left(G_n^{0,p}|_{\tilde{A}_{uv}}\right) \geq \mathsf{K}/2 \middle| \mathrm{F}\right] \leq \mathbb{P}\left[\omega_{u,v}\left(G_n^{0,p}|_{\tilde{A}_{uv}}\right) \geq \mathsf{K}_0/2 \middle| \mathrm{F}\right] = O(n^{-3}).$$

Combining this with Eqn. (3.9), we thus obtain that:

$$\text{If } \mathsf{K} \geq 8\beta^2 \text{ and } p \leq \min\left\{c_1^a \cdot (1/n)^{c_2^a/(2N_{\max})},\ c_1^a \cdot (1/n)^{c_2^a/\mathsf{K}}\right\},$$
$$\text{then } \mathbb{P}\left[\omega_{u,v}\left(G_n^{0,p}|_{\tilde{A}_{uv}}\right) \geq \mathsf{K}/2 \middle| \mathrm{F}\right] = O(n^{-3}). \tag{3.10}$$

Case (ii): bounding the second term in Eqn. (3.2). First recall that $B_{uv} = B_r^{\mathcal{X}_n}(u) \cup B_r^{\mathcal{X}_n}(v)$ (see Figure 2 (b)).

On one hand, imagine we now build the following random graph $\tilde{G}_{uv}^{local} = (\tilde{V}, \tilde{E})$: The vertex set $\tilde{V}$ is simply B_{uv}. To construct the edge set $\tilde{E}$, first, add edges between all pairs of distinct vertices in $B_r^{\mathcal{X}_n}(u)$ and do the same thing for $B_r^{\mathcal{X}_n}(v)$; that is, every two vertices in $B_r^{\mathcal{X}_n}(u)$ or $B_r^{\mathcal{X}_n}(v)$ are now connected by an edge. Next, add edge (u, v). Finally, insert each crossing edge (x, y) with $x \in B_r^{\mathcal{X}_n}(u)$ and $y \in B_r^{\mathcal{X}_n}(v)$ with probability p.

On the other hand, consider the graph $G_n^{0,p}|_{B_{uv}}$, the induced subgraph of $G_n^{0,p}$ spanned by vertices in B_{uv}. We can imagine that the graph $G_n^{0,p}|_{B_{uv}}$ was produced by first taking the induced subgraph $G_n|_{B_{uv}}$, and then insert crossing edges (x, y) each with probability p. Since (u, v) is a long-edge, by Definition 6, we know that there are no edges between nodes in $B_r^{\mathcal{X}_n}(u)$ and $B_r^{\mathcal{X}_n}(v)$ in $G_n|_{B_{uv}}$. Since every two vertices in $B_r^{\mathcal{X}_n}(u)$ or $B_r^{\mathcal{X}_n}(v)$ are not necessarily connected by an edge in $G_n|_{B_{uv}}$, we know that

$$\mathbb{P}\left[\omega_{u,v}\left(G_n^{0,p}|_{B_{uv}}\right) \geq \mathsf{K}/2 \middle| \mathrm{F}\right] \leq \mathbb{P}\left[\omega_{u,v}\left(\tilde{G}_{uv}^{local}\right) \geq \mathsf{K}/2 \middle| \mathrm{F}\right] \tag{3.11}$$

Using a similar argument as in case (i) (the missing details can be found in Appendix A.2), we have that there exist constants $c_1^b, c_2^b > 0$ which depend on the Besicovitch constant β and α such that

$$\text{If } \mathsf{K} \geq 8\beta^2 \text{ and } p \leq c_1^b \cdot (1/n)^{c_2^b/\mathsf{K}},$$
$$\text{then } \mathbb{P}\left[\omega_{u,v}\left(\tilde{G}_{uv}^{local}\right) \geq \mathsf{K}/2 \middle| \mathrm{F}\right] = O(n^{-3})$$

Pick $\mathsf{K} = 2N_{\max} = 2\lceil 4/\alpha \rceil \geq 8\beta^2$ (by condition $\alpha \in (0, 1/\beta^2]$). Note that $\mathsf{K} = O(1)$ in this case. Thus, combining the above bound with Eqn. (3.11), (3.10) and (3.2), there exist constants $C_1 = \min\{c_1^a, c_1^b\}$ and $C_2 = \max\{c_2^a, c_2^b\}$ such that if p satisfies conditions in Eqn. (3.1), then

$$\mathbb{P}\left[\omega_{u,v}\left(G_n^{0,p}\right) \geq \mathsf{K}\right] \leq \mathbb{P}\left[\omega_{u,v}\left(G_n^{0,p}\right) \geq \mathsf{K} \middle| \mathrm{F}\right] + \mathbb{P}[\mathrm{F}^c] = O(n^{-3})$$

Finally, by applying the union bound, this means:

$$\mathbb{P}\left[\text{for all long-edge } (u,v),\ \omega_{u,v}\left(G_n^{0,p}\right) \geq \mathsf{K}\right] = O(n^{-1})$$

Thus with high probability, we have that for all long-edge (u, v), $\omega_{u,v}(G_n^{0,p}) = O(1)$ as long as Eqn. (3.1) holds. This completes the proof of Part (a) of Theorem 10.

Proof of part (b) of Theorem 10. We use the same strategy in the proof of part (a). That is, we again try to bound the two terms on the right hand side of Eqn. (3.2) from above respectively. The key difference here is to give an alternative estimate of Eqn. (3.5) in case (i) and its counterpart in case (ii) under the new constraint of p.

For case (i), instead of using Lemma 11, we now use the following lemma, whose proof can be found in Appendix A.3.

Lemma 12. *There exists a constant $C_3 > 0$ depending on the Besicovitch constant β and α such that if $(1/n)^{8/(3N_{\max})} \leq p < 1$ and $\mathsf{K} = C_3\lfloor \log_{1/p} n \rfloor$, then we have that* $\mathbb{E}[\mathrm{I}] = O(n^{-3})$.

Now choose such C_3 as specified in Lemma 12. We know that the following holds.

$$\text{If } (1/n)^{8/(3N_{\max})} \leq p < 1,$$
$$\text{then } \mathbb{P}\left[\omega_{u,v}\left(G_n^{0,p}|_{\tilde{A}_{uv}}\right) \geq C_3\lfloor \log_{1/p} n \rfloor/2 \middle| \mathrm{F}\right] = O(n^{-3}). \tag{3.12}$$

For case (ii), we know that if event F has already happened, then $|B_{uv}| \leq 2N_{\max}$, where $|B_{uv}|$ denotes the cardinality of set B_{uv}. Note that if $(1/n)^{C_3/(4N_{\max}+C_3)} \leq p < 1$, then

$$C_3\lfloor \log_{1/p} n \rfloor/2 \geq 2N_{\max} \geq |B_{uv}|.$$

Hence, we obtain that:

$$\text{If } (1/n)^{C_3/(4N_{\max}+C_3)} \leq p < 1,$$
$$\text{then } \mathbb{P}\left[\omega_{u,v}\left(G_n^{0,p}|_{B_{uv}}\right) \geq C_3\lfloor \log_{1/p} n \rfloor/2 \middle| \mathrm{F}\right] = 0. \tag{3.13}$$

Set $\xi = \min\{8/(3N_{\max}), C_3/(4N_{\max} + C_3)\}$, which is also a constant. Thus, combining Eqn. (3.13), (3.12) and (3.2), we know that if $(1/n)^{\xi} \leq p < 1$, then

$$\mathbb{P}\left[\omega_{u,v}\left(G_n^{0,p}\right) \geq C_3\lfloor \log_{1/p} n \rfloor \middle| \mathrm{F}\right] = O(n^{-3}).$$

Finally, by a similar argument in the proof for Part (a) using the law of total probability and union bound, we can show that with high probability, we have that for any long-edge (u, v), $\omega_{u,v}\left(G_n^{0,p}\right) \lesssim \log_{1/p} n$. This completes the proof of Theorem 10.

3.1.2. *Finishing the proof of Part (I) of Theorem 1.* Based on the discussion of Type-I cliques as well as Theorem 10 for Type-II cliques, we have the following corollary regarding the upper bound of $\omega(G_n^{0,p})$ in the "very sparse" regime.

Corollary 13. *Given a $(0,p)$-perturbed noisy random geometric graph $G_n^{0,p}$ in the standard-setting-R and suppose that $nr^d \le n^{-\alpha}$ for some fixed $\alpha \in (0, 1/\beta^2]$, then*

(i) *there exist two constants $C_1, C_2 > 0$ such that if $p \le C_1 (1/n)^{C_2}$, then a.s.*

$$\omega\left(G_n^{0,p}\right) \lesssim 1$$

(ii) *and there exists a constant $\xi > 0$ such that if $(1/n)^{\xi} \le p < 1$, then a.s.*

$$\omega\left(G_n^{0,p}\right) \lesssim \log_{1/p} n$$

Part (i) of Corollary 13 can be derived by combining Lemma 8 and part (a) of Theorem 10, while part (ii) of Corollary 13 can be derived by combining Lemma 8 and part (b) of Theorem 10.

To derive a lower bound of $\omega(G_n^{0,p})$, we need the following result on the clique number of Erdős–Rényi random graphs (proof can be found in Appendix A.4).

Lemma 14. *For Erdős–Rényi random graph $G(n,p)$ with $(1/n)^{1/11} \le p \le (1/n)^{1/\sqrt[4]{n}}$, we have a.s. $\omega(G(n,p)) > \lfloor \log_{1/p} n \rfloor$.*

Note that p here is no longer a fixed constant as in the standard literature [1, 3], thus the well-known $\omega(G(n,p)) \sim 2\log_{1/p} n$ statement cannot be directly applied here. Not surprisingly, the standard second moment method [1] is used here, but the calculation is different. Details of the proof can be found in Appendix A.4.

Easy to see that $\mathbb{P}[\omega(G_n^{0,p}) \ge K] \ge \mathbb{P}[\omega(G(n,p)) \ge K]$ for any positive integer K. The following corollary of Lemma 14 gives a lower bound of $G_n^{0,p}$ regardless of which regime nr^d belongs to.

Corollary 15. *Given a $(0,p)$-perturbed noisy random geometric graph $G_n^{0,p}$ in the standard-setting-R and suppose that $(1/n)^{1/11} \le p \le (1/n)^{1/\sqrt[4]{n}}$, then we have a.s.*

$$\omega\left(G_n^{0,p}\right) > \lfloor \log_{1/p} n \rfloor$$

Note that $(1/n)^{1/\sqrt[4]{n}} \to 1$ as $n \to \infty$. Thus, there exists a constant $C_3 \in (0,1)$ (very close to 1) such that $C_3 \le (1/n)^{1/\sqrt[4]{n}}$ for sufficiently large n. Also note that the following monotone property holds.

$$\text{For any } S > 0 \text{ and } 0 \le q_1 < q_2 < 1, \mathbb{P}\left[\omega\left(G_n^{0,q_1}\right) \ge S\right] \le \mathbb{P}\left[\omega\left(G_n^{0,q_2}\right) \ge S\right]$$

And by Corollary 13 (b), we know that $\omega\left(G_n^{0,n^{-\xi}}\right) = O\left(\log_{n^{-\xi}} n\right) = O(1)$ a.s.. Easy to see that there exists a constant C_1' such that $p \le (1/n)^{C_1'}$ implies $p \le C_1 (1/n)^{C_2}$. Also notice that $\log_{1/p} n = \Theta(1)$ for $p \in \left((1/n)^{C_1'}, (1/n)^{\xi}\right)$ (if this interval exists). Thus, the lower bound of p in the condition of part (b) of Corollary 13 (which is $(1/n)^{\xi}$) can be extended to $(1/n)^{C_1'}$ and the conclusion still holds. Combining these facts with Corollary 13 and Corollary 15 concludes the proof of part (I) of Theorem 1.

3.2. **Proof of Part (II) — "quite sparse" regime.** In this section, we discuss the order of $\omega(G_n^{0,p})$ in the regime $n^{-\epsilon} \ll nr^d \ll \log n$ for all $\epsilon > 0$. Again, we first derive an upper bound of $\omega(G_n^{0,p})$ by considering two types of cliques (Type-I and Type-II) introduced in Section 3.1. The idea of the proof in this section is very similar to the one in Section 3.1, although the calculations are quite different.

Type-I cliques. Recall that $W_3 = B_3(\mathbf{o})$. The following lemma gives an upper bound of the number of vertices in any $3r-$ball.

Lemma 16. *If $n^{-\epsilon} \ll nr^d \ll \log n$ for all $\epsilon > 0$, then*

$$\mathbb{P}\left[M_{W_3} \le \frac{5\log n}{\log\left(\log n/(\sigma\theta 6^d nr^d)\right)}\right] = 1 + O(n^{-3}).$$

Proof. Set

$$k_n := \frac{5\log n}{\log\left(\log n/(\sigma\theta 6^d nr^d)\right)}.$$

Similar to Lemma 8, we have

$$\begin{aligned}\mathbb{P}\left[M_{W_3} > k_n\right] &\le \mathbb{P}\left[\exists i : \mathcal{N}\left(B_{6r}(X_i)\right) > k_n\right]\\ &\le n\mathbb{P}\left[\mathcal{N}\left(B_{6r}(X_1)\right) > k_n\right]\\ &\le n\mathbb{P}\left[Bin\left(n, \sigma\theta(6r)^d\right) > k_n\right]\end{aligned}$$

Set $\mu := n\sigma\theta(6r)^d$. It is easy to check that $k_n > \mu$ (since $nr^d \ll \log n$) Finally, by Chernoff-Hoeffding theorem (Lemma 5) and note that $H(a) \ge a(\log a - 1)$, combining with $n^{-\epsilon} \ll nr^d \ll \log n$, we have

$$\begin{aligned}&n\mathbb{P}\left[Bin\left(n, \sigma\theta(6r)^d\right) > k_n\right]\\ &\le n\exp\left[-\mu H\left(\frac{k_n}{\mu}\right)\right]\\ &\le n\exp\left[-k_n\left(\log\frac{k_n}{n\sigma\theta(6r)^d} - 1\right)\right]\\ &= n\exp\left[-5\log n\left(1 + \frac{\log 5 - 1 - \log\log\left(\log n/\left(n\sigma\theta(6r)^d\right)\right)}{\log\left(\log n/\left(n\sigma\theta(6r)^d\right)\right)}\right)\right]\\ &< n\cdot n^{-4} = n^{-3}.\end{aligned}$$

The claim then follows. ☐

Type-II cliques. Recall that $W_1 = B_1(\mathbf{o})$. By using a similar argument as the one used for Lemma 16, we can get the following lemma which gives an upper bound for the number of points in any $r-$ball for the regime of nr^d under discussion.

Lemma 17. *If $n^{-\epsilon} \ll nr^d \ll \log n$ for all $\epsilon > 0$, then*

$$\mathbb{P}\left[M_{W_1} \le \frac{5\log n}{\log\left(\log n/(\sigma\theta 2^d nr^d)\right)}\right] = 1 + O(n^{-3})$$

Theorem 18. *Given an $(0,p)$-perturbed noisy random geometric graph $G_n^{0,p}$ in the standard-setting-R and suppose that $n^{-\epsilon} \ll nr^d \ll \log n$ for all $\epsilon > 0$, then*

(a) *there exist constants $C_1, C_2 > 0$ which depend on the Besicovitch constant β such that if*

$$p \le C_1 \cdot \left(nr^d/\log n\right)^{C_2} \tag{3.14}$$

then, with high probability, for all long-edge (u,v) in $G_n^{0,p}$, its edge clique number

$$\omega_{u,v}(G_n^{0,p}) \lesssim \frac{\log n}{\log\left(\log n/nr^d\right)}.$$

(b) *and there exists a constant ξ which depends on the Besicovitch constant β such that if $\left(nr^d/\log n\right)^\xi \le p < 1$, then, with high probability, for all long-edge (u,v) in $G_n^{0,p}$, its edge clique number $\omega_{u,v}(G_n^{0,p}) \lesssim \log_{1/p} n$.*

The proof of the above theorem follows the same flow as the proof of Theorem 10; although detailed derivations differ. The details can be found in Appendix A.5.

3.2.1. *Putting everything together for Part (II) of Theorem 1.* To wrap up all the above results, we have the following corollary regarding the upper bound of $\omega(G_n^{0,p})$ in "quite sparse" regime.

Corollary 19. *Given a $(0,p)$-perturbed noisy random geometric graph $G_n^{0,p}$ in the standard-setting-R and suppose that $n^{-\epsilon} \ll nr^d \ll \log n$ for all $\epsilon > 0$, then*

(i) *there exist two constants $C_1, C_2 > 0$ such that if $p \le C_1\left(nr^d/\log n\right)^{C_2}$, then a.s.*

$$\omega\left(G_n^{0,p}\right) \lesssim \frac{\log n}{\log\left(\log n/nr^d\right)}$$

(ii) *and there exists a constant $\xi > 0$ such that if $\left(nr^d/\log n\right)^\xi \le p < 1$, then a.s.*

$$\omega\left(G_n^{0,p}\right) \lesssim \log_{1/p} n$$

Part (i) can be derived by combining Lemma 16 and part (a) of Theorem 18, while part (ii) can be derived by combining Lemma 16 and part (b) of Theorem 18.

To derive a tight bound of $\omega(G_n^{0,p})$, in addition to Corollary 15, we also need the follwing lemma, which provides a lower bound of $\omega(G_n^{0,p})$.

Lemma 20. *Given a $(0,p)$-perturbed noisy random geometric graph $G_n^{0,p}$ in the standard-setting-R and suppose that $n^{-\epsilon} \ll nr^d \ll \log n$ for all $\epsilon > 0$, then a.s.*

$$\omega\left(G_n^{0,p}\right) \ge \frac{\log n}{2\log\left(\log n/nr^d\right)}$$

Proof of Lemma 20. Note that $\omega\left(G_n^{0,p}\right) \ge M_{W_{1/2}}$ and $M_{W_{1/2}}$ can a.s. be bounded from below by $\log n/\left(2\log\left(\log n/nr^d\right)\right)$. (Pick $\epsilon = 1/2$ in Lemma 3.9 of [15].) □

Finally, combining Lemma 20 with part (i) of Corollary 19 concludes the proof of part (II.a) of Theorem 1. Note that there exists a constant C_1' such that $p \le \left(nr^d/\log n\right)^{C_1'}$ implies $p \le C_1\left(nr^d/\log n\right)^{C_2}$ and if $p \in \left(\left(nr^d/\log n\right)^{C_1'}, \left(nr^d/\log n\right)^\xi\right)$ (if this interval exists), then

$$\log_{1/p} n = \Theta\left(\frac{\log n}{\log\left(\log n/nr^d\right)}\right).$$

Thus we can extend the lower bound of the condition in part (ii) of Corollary 19 to $\left(nr^d/\log n\right)^{C_1'}$ by the same reasoning at the end of the proof for "very sparse" regime. Combining these facts with Corollary 15 and part (ii) of Corollary 19 concludes the proof of (II.b) of Theorem 1.

3.3. **Proof of Part (III) — "dense" regime.** In this section, we discuss the order of $\omega(G_n^{0,p})$ in the regime $\sigma nr^d/\log n \to t \in (0,\infty)$. Again, we first derive an upper bound of $\omega(G_n^{0,p})$ by considering two types of cliques (Type-I and Type-II) introduced in Section 3.1. The idea of the proof in this section is very similar to the one in Section 3.2, thus the proofs are omitted.

Set τ be the smallest real number such that $\tau \ge 2$ and $\tau(\log\tau - 1) \ge 4/(2^d\theta t)$. Since d, t and θ are all given constants, τ is also a constant.

Type-I cliques. The following lemma gives upper bounds of the number of vertices in each $r-$ball and $3r-$ball respectively. Recall that $W_1 = B_1(\mathbf{o})$ and $W_3 = B_3(\mathbf{o})$.

Lemma 21. *If $\sigma nr^d / \log n \to t \in (0, \infty)$, then*

$$\mathbb{P}\left[M_{W_1} \leq \tau 2^d \theta \sigma nr^d\right] = 1 + O(n^{-3})$$
$$\mathbb{P}\left[M_{W_3} \leq \tau 6^d \theta \sigma nr^d\right] = 1 + O(n^{-3})$$

Type-II cliques. The proof of the following technical theorem is almost the same as the proof of part (a) of Theorem 18 thus is omitted.

Theorem 22. *Given a $(0, p)$-perturbed noisy random geometric graph $G_n^{0,p}$ in the standard-setting-R and suppose that $\sigma nr^d / \log n \to t \in (0, \infty)$, then there exists a constant C which depends on the Besicovitch constant β such that if $p \leq C$ then, with high probability, for all long-edge (u, v) in $G_n^{0,p}$, its edge clique number $\omega_{u,v}(G_n^{0,p}) \lesssim nr^d$.*

To derive a tight bound of $\omega(G_n^{0,p})$, in addition to Corollary 15, we also need the following result on lower bound.

Lemma 23. *Given a $(0, p)$-perturbed noisy random geometric graph $G_n^{0,p}$ in the standard-setting-R and suppose that $\sigma nr^d / \log n \to t \in (0, \infty)$, then a.s.*

$$\omega\left(G_n^{0,p}\right) \geq \frac{1}{2} \eta \sigma nr^d$$

where η is the unique solution $x \geq \theta\,(1/2)^d$ to $H\left(x / \left(\theta\,(1/2)^d\right)\right) = 1/\left(\theta\,(1/2)^d\, t\right)$ (recall that function H is defined as $H(a) = 1 - a + a \log a$).

Proof of Lemma 23. Note that $\omega\left(G_n^{0,p}\right) \geq M_{W_{1/2}}$ and $M_{W_{1/2}}$ can be bounded from below by $\eta\sigma nr^d/2$ almost surely (directly by Theorem 1.8 of [15]). □

Finally, combining Lemma 23, Lemma 21 and Theorem 22 concludes the proof.

4. Proof of Theorem 2

In this section, we focus on deriving the order of $\omega\left(G_n^{q,0}\right)$. Note that $\omega\left(G_n^{q,0}\right) \leq M_{W_1}$. Thus, Theorem 2 part (I) is obvious due to Lemma 9. Our proof of the remaining parts of Theorem 2 uses the following lemma following easily from known results in the literature. Recall that ν is the probability distribution defined in Section 1.1.

Lemma 24 (Lemma 3.1 in [20])**.** *For any fixed $\rho > 0$, recall that $W_1 = B_1(\mathbf{o})$ (and thus $rW_1 = B_r(\mathbf{o})$). There exists $N = \Omega(r^{-d})$ disjoint translates $x_1 + rW_1, \cdots, x_N + rW_1$ of rW_1 with $\nu(x_i + rW_1) \geq (1 - \rho)\sigma\theta r^d$ for all $i = 1, \cdots, N$.*

We would like to point out that our proof in this section follows an approach analogous to the proof of Theorem 1.8 in [15]. We only show the proof for the "quite sparse" in the main context. Since we use almost identical technique for the "dense" regime, the proof of the this regime (part (III)) is put in Appendix B.1.

4.1. Proof of Part (II) — "quite sparse" regime.

In this section, we discuss the order of $\omega(G_n^{q,0})$ in the regime $n^{-\epsilon} \ll nr^d \ll \log n$ for all $\epsilon > 0$.

4.1.1. *Deriving upper bound.* We first focus on the upper bound of $\omega\left(G_n^{q,0}\right)$. This is obtained via considering and relating to the random geometric graphs whose nodes are generated by Poisson point process. Let $N \sim Poisson\left((1+\delta)n\right)$ for some $\delta > 0$ (say $\delta = 1/2$).

Note that G_N (random geometric graph on N nodes; recall Definition 1) is a geometric graph (r-neighborhood graph) of the Poisson point process $\mathcal{P}_{(1+\delta)n}$ with intensity $(1+\delta)nf$ [20], where f is the density defined in Section 1.1. Similar to $G_n^{q,0}$, we define a $(q,0)-$perturbation of G_N as $G_N^{q,0}$. Set k_n be an integer to be determined. Now, we have

$$\begin{aligned}\mathbb{P}\left[\omega\left(G_n^{q,0}\right) \geq k_n\right] &\leq \mathbb{P}\left[\omega\left(G_N^{q,0}\right) \geq k_n\right] + \mathbb{P}\left[N \leq n-1\right] \\ &\leq \mathbb{P}\left[\omega\left(G_N^{q,0}\right) \geq k_n\right] + e^{-\gamma n}\end{aligned}$$

for some constant $\gamma > 0$ (depending on δ) by Lemma 5. For $y \in \mathbb{R}^d$ let $\mathbf{X}_y$ be the set of nodes of G_N falling in $B_r(y)$. Let M_y be the number of points falling in $B_r(y)$ spanning a maximum clique in $G_N^{q,0}\mid_{\mathbf{X}_y}$. Define $M := \max_{y\in\mathbb{R}^d} M_y$. Easy to see

$$\mathbb{P}\left[\omega\left(G_N^{q,0}\right) \geq k_n\right] = \mathbb{P}\left[M \geq k_n\right] \tag{4.1}$$

Fix $y \in \mathbb{R}^d$. By the property of Poisson point process, we know $|\mathbf{X}_y| \sim Poisson(\lambda)$ where $\lambda := (1+\delta)n\int_{B_r(y)} f(x)dx$. By using Markov's inequality, we have

$$\begin{aligned}\mathbb{P}\left[M_y \geq k_n\right] &= \sum_{i\geq k_n} \mathbb{P}\left[M_y \geq k_n \middle| |\mathbf{X}_y| = i\right]\mathbb{P}\left[|\mathbf{X}_y| = i\right] \\ &= \sum_{i\geq k_n} \mathbb{P}\left[\text{number of } k_n\text{-cliques in } G_N^{q,0}\mid_{\mathbf{X}_y} \geq 1 \middle| |\mathbf{X}_y| = i\right]\frac{e^{-\lambda}\lambda^i}{i!} \\ &\leq \sum_{i\geq k_n} \mathbb{E}\left[\text{number of } k_n\text{-cliques in } G_N^{q,0}\mid_{\mathbf{X}_y} \middle| |\mathbf{X}_y| = i\right]\frac{e^{-\lambda}\lambda^i}{i!} \\ &\leq \sum_{i\geq k_n} \binom{i}{k_n}(1-q)^{\binom{k_n}{2}}\frac{e^{-\lambda}\lambda^i}{i!} \\ &= \frac{\lambda^{k_n}}{k_n!}(1-q)^{\binom{k_n}{2}} \cdot e^{-\lambda}\sum_{i\geq k_n}\frac{\lambda^{i-k_n}}{(i-k_n)!} \\ &= \frac{\lambda^{k_n}}{k_n!}(1-q)^{\binom{k_n}{2}}\end{aligned}$$

Note that $\lambda \leq (1+\delta)\sigma\theta n r^d$. Thus,

$$\mathbb{P}\left[M_y \geq k_n\right] \leq \frac{\left((1+\delta)\sigma\theta n r^d\right)^{k_n}}{k_n!}(1-q)^{\binom{k_n}{2}}$$

which does not depend on the choice of y. Combining this with (4.1), we have

$$\begin{aligned}\mathbb{P}\left[\omega\left(G_N^{q,0}\right) \geq k_n\right] &\leq \frac{\left((1+\delta)\sigma\theta n r^d\right)^{k_n}}{k_n!}(1-q)^{\binom{k_n}{2}} \\ &< \frac{1}{\sqrt{2\pi}}\left(\frac{(1+\delta)e\sigma\theta n r^d(1-q)^{(k_n-1)/2}}{k_n}\right)^{k_n}\end{aligned}$$

Finally, pick

$$k_n = 2\log_{1/(1-q)}\left(\frac{\log n}{\log\left(\log n/nr^d\right)}\right) + 1.$$

Since $n^{-\epsilon} \ll nr^d \ll \log n$ for all $\epsilon > 0$, easy to see that $k_n \to \infty$. Note that

$$\frac{(1+\delta)e\sigma\theta nr^d(1-q)^{(k_n-1)/2}}{k_n} = \frac{(1+\delta)e\sigma\theta}{k_n} \cdot \frac{\log\left(\log n/nr^d\right)}{\log n/nr^d} \le \frac{C}{k_n}$$

for some constant $C > 0$. Thus, $\mathbb{P}\left[\omega\left(G_N^{q,0}\right) \ge k_n\right] = o(1)$. Hence, we have that a.s.

$$\omega\left(G_n^{q,0}\right) \lesssim \log\frac{\log n}{\log\left(\log n/nr^d\right)}$$

4.1.2. *Deriving the lower bound.* Now we consider the lower bound of $\omega\left(G_n^{q,0}\right)$. We first state the following well-known result on the clique number of Erdős–Rényi random graphs, which plays an important role in proving the lower bound.

Lemma 25. *Suppose $p \in (0,1)$ is a constant. For Erdős–Rényi random graph $G(n,p)$ with $n \to \infty$, we have*

$$\mathbb{P}\left[\omega(G(n,p)) \le \left\lfloor \log_{1/p} n\right\rfloor\right] < e^{-n}$$

This is a direct corollary of the standard $2\log_{1/p} n$ statement (see P. 185 [1]), thus we omit the proof.

Now let $N \sim Poisson\left((1-\delta')n\right)$ for some $\delta' \in (0,1)$ (say $\delta' = 1/2$). Note that G_N is an $r-$neighborhood graph of the Poisson point process $\mathcal{P}_{(1-\delta')n}$ with intensity $(1-\delta')nf$, where f is the density defined in Section 1.1. Similarly, we define a $(q,0)-$perturbation of G_N as $G_N^{q,0}$. Set k_n be an integer to be determined. Now, we have

$$\begin{aligned}\mathbb{P}\left[\omega\left(G_n^{q,0}\right) \le k_n\right] &\le \mathbb{P}\left[\omega\left(G_N^{q,0}\right) \le k_n\right] + \mathbb{P}\left[N \ge n+1\right] \\ &\le \mathbb{P}\left[\omega\left(G_N^{q,0}\right) \le k_n\right] + e^{-\gamma' n}\end{aligned}$$

for some constant $\gamma' > 0$ (depending on δ') by Lemma 5. Now fix some constant $\rho \in (0,1)$ (say $\rho = 1/2$). Recall $W_{1/2} = B_{1/2}(\mathbf{o})$. By Lemma 24, there exist points $x_1, x_2, \cdots, x_m$ with $m = \Omega\left(r^{-d}\right)$ such that the sets $x_i + W_{1/2}$ are disjoint and

$$\nu\left(x_i + W_{1/2}\right) \ge \frac{(1-\rho)\sigma\theta}{2^d} r^d$$

for $i = 1, \cdots, m$ where ν is the probability distribution defined in Section 1.1. Let $\mathbf{X}_i$ be the set of points of G_N falling in $x_i + W_{1/2}$. Then, we have

$$\begin{aligned}\mathbb{P}\left[\omega\left(G_N^{q,0}\right) \le k_n\right] &\le \mathbb{P}\left[\omega\left(G_N^{q,0}\mid_{\mathbf{X}_1}\right) \le k_n, \cdots, \omega\left(G_N^{q,0}\mid_{\mathbf{X}_m}\right) \le k_n\right] \\ &= \prod_{i=1}^m \mathbb{P}\left[\omega\left(G_N^{q,0}\mid_{\mathbf{X}_i}\right) \le k_n\right] \qquad (4.2)\end{aligned}$$

Note that all the points falling in any $r/2$-ball span a complete graph. Thus, for each i, we know the following holds.

$$\mathbb{P}\left[\omega\left(G_N^{q,0}\mid_{\mathbf{X}_i}\right) \le k_n\right] = \mathbb{P}\left[\omega\left(G\left(|\mathbf{X}_i|, 1-q\right)\right) \le k_n\right].$$

Set

$$\Phi_n := \frac{\log n}{2\log\left(\log n/nr^d\right)}$$

which goes to infinty as n grows. Note that $|\mathbf{X}_i| \sim Poisson\left(\tilde{\lambda}\right)$ where

$$\frac{(1-\delta')\sigma\theta nr^d}{2^d} \ge \tilde{\lambda} := (1-\delta')n \cdot \nu(x_i + W_{1/2}) \ge \frac{(1-\rho)(1-\delta')\sigma\theta nr^d}{2^d}. \qquad (4.3)$$

The upper bound follows from the upper bound of volume of balls.

Now pick

$$k_n := \left\lfloor \log_{1/(1-q)} \Phi_n \right\rfloor = \Omega\left(\log \frac{\log n}{\log\left(\log n / nr^d\right)}\right).$$

Let $Q \sim Poisson\left(\tilde{\lambda}/e\right)$. By the law of total probability, we have

$$\begin{aligned}
&\mathbb{P}\left[\omega\left(G\left(|\mathbf{X}_i|, 1-q\right)\right) \le k_n\right] \\
\le&\mathbb{P}\left[|\mathbf{X}_i| \le \Phi_n\right] + \sum_{j=\lceil \Phi_n \rceil}^{\infty} \mathbb{P}\left[\omega\left(G\left(j, 1-q\right)\right) \le k_n\right] \mathbb{P}\left[|\mathbf{X}_i| = j\right] \\
\le&1 - \mathbb{P}\left[|\mathbf{X}_i| \ge \Phi_n + 1\right] + \sum_{j=\lceil \Phi_n \rceil}^{\infty} \mathbb{P}\left[\omega\left(G\left(j, 1-q\right)\right) \le \left\lfloor \log_{\frac{1}{1-q}} j \right\rfloor\right] \frac{e^{-\tilde{\lambda}}\tilde{\lambda}^j}{j!} \\
<&1 - \left(\frac{\tilde{\lambda}}{e(\Phi_n+1)}\right)^{\Phi_n+1} + \sum_{j=\lceil \Phi_n \rceil}^{\infty} e^{-j} \frac{e^{-\tilde{\lambda}}\tilde{\lambda}^j}{j!} \qquad (4.4) \\
=&1 - e^{-(\Phi_n+1)\log\left(e(\Phi_n+1)/\tilde{\lambda}\right)} + e^{-\tilde{\lambda}+\frac{\tilde{\lambda}}{e}} \sum_{j=\lceil \Phi_n \rceil}^{\infty} \frac{e^{-\tilde{\lambda}/e}\left(\tilde{\lambda}/e\right)^j}{j!} \\
\le&1 - e^{-(\Phi_n+1)\log\left(e(\Phi_n+1)/\tilde{\lambda}\right)} + e^{-\tilde{\lambda}+\frac{\tilde{\lambda}}{e}} \cdot \mathbb{P}[Q \ge \Phi_n] \\
<&1 - e^{-(\Phi_n+1)\log\left(e(\Phi_n+1)/\tilde{\lambda}\right)} + e^{-(1-1/e)\tilde{\lambda}} \cdot e^{-\Phi_n \log\left(\Phi_n/\tilde{\lambda}\right)} \qquad (4.5)
\end{aligned}$$

where Eqn. (4.4) and (4.5) hold due to Lemma 4 (note that $\Phi \gg \tilde{\lambda}/e$) and Lemma 25. Routine calculations show that for n large enough, we have

$$\begin{aligned}
(\Phi_n+1)\log\left(e(\Phi_n+1)/\tilde{\lambda}\right) &\le \frac{1}{2}\log n + 1 \\
\Phi_n \log\left(\Phi_n/\tilde{\lambda}\right) &\ge \frac{1}{2}\log n - 1
\end{aligned}$$

and $e^{-(1-1/e)\tilde{\lambda}} \le 1/(2e^2)$ since $nr^d \gg n^{-\epsilon}$ for all $\epsilon > 0$. Thus

$$\mathbb{P}\left[\omega\left(G\left(\mathcal{N}\left(\mathbf{X}_i\right), 1-q\right)\right) \le k_n\right] < 1 - \frac{1}{2e} n^{-1/2}.$$

Plugging this back into Eqn. (4.2), we have

$$\mathbb{P}\left[\omega\left(G_N^{q,0}\right) \le k_n\right] < \left(1 - \frac{1}{2e}n^{-1/2}\right)^m \le e^{-\frac{1}{2e}n^{-1/2}m}$$

Recall $m = \Omega(r^{-d})$ and $nr^d \ll \log n$, thus $n^{-1/2}m = \Omega\left(\sqrt{n}/\log n\right)$. This implies $\mathbb{P}\left[\omega\left(G_N^{q,0}\right) \le k_n\right] = o(1)$. Since we have that $\mathbb{P}\left[\omega\left(G_n^{q,0}\right) \le k_n\right] = o(1)$ with

$$k_n = \Omega\left(\log \frac{\log n}{\log\left(\log n / nr^d\right)}\right),$$

we thus obtain the lower bound in Part (II) of Theorem 2.

5. COMBINED CASE

In this section, we focus on bounding the clique number $\omega\left(G_n^{q,p}\right)$ of $G_n^{q,p}$, for different regimes of nr^d, q and p. Analogously to the monotonicity of the clique number of Erdős–Rényi random graphs [11], we have the following two monotone properties: for any positive integer K,

$$\mathbb{P}\left[\omega\left(G_n^{0,p}\right) \le K\right] \le \mathbb{P}\left[\omega\left(G_n^{q,p}\right) \le K\right] \le \mathbb{P}\left[\omega\left(G_n^{q,0}\right) \le K\right]$$
$$\mathbb{P}\left[\omega\left(G_n^{q,0}\right) \ge K\right] \le \mathbb{P}\left[\omega\left(G_n^{q,p}\right) \ge K\right] \le \mathbb{P}\left[\omega\left(G_n^{0,p}\right) \ge K\right]$$

Combining these properties with Theorem 1, Theorem 2 and technical lemmas (Lemma 25 and Corollary 15), we can derive some of the results showing below (part (I) and part (III.b)). Other results can be derived by carefully choosing K_n in the proof of Theorem 18 and Theorem 22 to fit the corresponding lower bound for different regimes of nr^d. For example, we can set some

$$\mathsf{K}_n = \Theta\left(\log\left(\frac{\log n}{\log\left(\log n/nr^d\right)}\right)\right)$$

(the lower bound in the "quite sparse" regime for deletion-only case; see Theorem 2) in the proof of Theorem 18 to derive part (II.a) of Theorem 26. For these reasons, we omit the proof of the following theorem.

Theorem 26. *Given a (q,p)-perturbed noisy random geometric graph $G_n^{q,p}$ in the standard-setting-R with a fixed constant $0 < q < 1$, the following holds:*

(I) *Suppose that $nr^d \le n^{-\alpha}$ for some fixed $\alpha \in (0, 1/\beta^2]$. Then there exist constants C_1, C_2 such that*

(I.a) *if $p \le (1/n)^{C_1}$, then a.s.*

$$\omega\left(G_n^{q,p}\right) \sim 1$$

(I.b) *and if $(1/n)^{C_1} < p \le C_2$, then a.s.*

$$\omega\left(G_n^{q,p}\right) \sim \log_{1/p} n$$

(II) *Suppose that $n^{-\epsilon} \ll nr^d \ll \log n$ for all $\epsilon > 0$. Then there exist constants C_1, C_2, C_3 such that*

(II.a) *if*

$$p \le (1/n)^{C_1/\log\frac{\log n}{\log(\log n/nr^d)}},$$

then a.s.

$$\omega\left(G_n^{q,p}\right) \sim \log\frac{\log n}{\log\left(\log n/nr^d\right)}$$

(II.b) *and if $\left(nr^d/\log n\right)^{C_2} < p \le C_3$, then a.s.*

$$\omega\left(G_n^{q,p}\right) \sim \log_{1/p} n$$

(III) *There exists a constant $T > 0$ such that if $\sigma nr^d/\log n \to t \in (T, \infty)$, then there exist constant C_1, C_2 such that*

(III.a) *if $p \le (1/n)^{C_1/\log\log n}(\log\log n/\log n)$, then a.s.*

$$\omega\left(G_n^{q,p}\right) \sim \log\left(nr^d\right)$$

(III.b) *and if $0 < p \le C_2$ and $p = \Theta(1)$, then a.s.*

$$\omega\left(G_n^{q,p}\right) \sim \log_{1/p} n$$

6. Concluding remarks

In this paper, we study the behavior of the clique number of noisy random geometric graphs $G_n^{q,p}$. In particular, we give the asymptotic tight bounds for the insertion-only case $G_n^{0,p}$ and the deletion-only case $G_n^{q,0}$ under different assumptions on nr^d (Theorem 1 and Theorem 2, respectively). To obtain these results, we deploy a range of classical and new techniques: For example, we develop a novel approach based on what we call the "well-separated clique-partitions family" to handle the insertion case. Some partial results for the general case $\omega\left(G_n^{q,p}\right)$ are also provided (Theorem 26). We also note that results in our paper can be extended beyond the Euclidean setting: For example, in [12], noisy random geometric graphs generated from points sampled from a well-behaved doubling measure supported on a geodesic space are considered, and behaviors of the edge clique number are investigated.

This work represents a first step towards characterizing properties of the noisy random geometric graphs (which intuitively are generated based on two types of random processes). There are many interesting open problems. For example, the combined case is not yet completely resolved (there are still gaps in the regimes). Also currently we only provide asymptotic tight bounds, and it would be interesting to identify the exact constant for the high order terms too. It will also be interesting to study other quantities beyond the clique number. Finally, we note that the random deletions/insertions can be viewed as "noise" on top of a base graph (which is a random geometric graph in our work). It will be interesting to see whether studies of clique numbers of other quantities can be used to "denoise" the input graph in practical applications (e.g, as in [19]).

References

[1] Noga Alon and Joel H. Spencer, *The probabilistic method*, 4th ed., Wiley Publishing, 2016.

[2] Philippe Blanchard and Dimitri Volchenkov, *Mathematical analysis of urban spatial networks*, Springer Science & Business Media, 2008.

[3] Béla Bollobás, *Random graphs*, Cambridge university press, 2001.

[4] Béla Bollobás and Oliver Riordan, *Percolation*, Cambridge University Press, 2006.

[5] L Booth, J Bruck, M Cook, and M Franceschetti, *Ad hoc wireless networks with noisy links*, Information theory, proceedings. ieee international symposium on, 2003, pp. 386–386.

[6] Don Coppersmith, David Gamarnik, and Maxim Sviridenko, *The diameter of a long-range percolation graph*, Mathematics and computer science ii: Algorithms, trees, combinatorics and probabilities, 2002, pp. 147–159.

[7] Luc Devroye, András György, Gábor Lugosi, Frederic Udina, et al., *High-dimensional random geometric graphs and their clique number*, Electronic Journal of Probability **16** (2011), 2481–2508.

[8] Herbert Federer, *Geometric measure theory*, Springer, 2014.

[9] Juha Heinonen, *Lectures on analysis on metric spaces*, Springer Science & Business Media, 2012.

[10] Svante Janson, Róbert Kozma, Miklós Ruszinkó, and Yury Sokolov, *Bootstrap percolation on a random graph coupled with a lattice*, Electronic Journal of Combinatorics (2016).

[11] Svante Janson, Tomasz Luczak, and Andrzej Rucinski, *Random graphs*, Vol. 45, John Wiley & Sons, 2011.

[12] Matthew Kahle, Minghao Tian, and Yusu Wang, *Local cliques in er-perturbed random geometric graphs*, arXiv preprint arXiv:1810.08383 (2018).

[13] Barbara I Mahler, Ulrike Tillmann, and Mason A Porter, *Analysis of contagion maps on a class of networks that are spatially embedded in a torus*, arXiv preprint arXiv:1812.09806 (2018).

[14] David W Matula, *The largest clique size in a random graph*, Technical Report, 1976.

[15] Colin McDiarmid and Tobias Müller, *On the chromatic number of random geometric graphs*, Combinatorica **31** (2011), no. 4, 423–488.

[16] Ronald Meester and Rahul Roy, *Continuum percolation*, Vol. 119, Cambridge University Press, 1996.

[17] Tobias Müller, *Two-point concentration in random geometric graphs*, Combinatorica **28** (2008), no. 5, 529.

[18] Maziar Nekovee, *Worm epidemics in wireless ad hoc networks*, New Journal of Physics **9** (2007), no. 6, 189.
[19] Srinivasan Parthasarathy, David Sivakoff, Minghao Tian, and Yusu Wang, *A quest to unravel the metric structure behind perturbed networks*, 33rd international symposium on computational geometry, socg 2017, july 4-7, 2017, brisbane, australia, 2017, pp. 53:1–53:16.
[20] Mathew Penrose, *Random geometric graphs*, Oxford University Press, 2003.
[21] Gareth William Peters and Tomoko Matsui, *Theoretical aspects of spatial-temporal modeling*, Springer, 2015.
[22] Dane Taylor, Florian Klimm, Heather A Harrington, Miroslav Kramár, Konstantin Mischaikow, Mason A Porter, and Peter J Mucha, *Topological data analysis of contagion maps for examining spreading processes on networks*, Nature communications **6** (2015), 7723.
[23] Xian Yuan Wu, *Mixing time of random walk on poisson geometry small world*, Internet Mathematics (2017).

APPENDIX A. THE MISSING PROOFS IN SECTION 3

A.1. **The proof of Lemma 11.** Since $1 \leq k \leq N_{\max}$, we know that $p \leq 1$ thus is well-defined as a probability. To estimate the summation on the right hand side of Eqn. (3.5), we consider the quantity $x_{\max} := \max_i\{x_i\}$. We first enumerate all the possible cases of $(x_1, x_2, \cdots, x_m)$ when $x_{\max}$ is fixed, and then vary the value of $x_{\max}$.

Set $h(y) = \max_{x_{\max}=y}\{\sum_{i=1}^m x_i^2\}$ for $y \geq \lceil k/m \rceil$. It is the maximum value of $\sum_{i=1}^m x_i^2$ under the constraint $x_{\max} = y$. Without loss of generality, we assume $x_1 = y$ and $y \geq x_2 \geq x_3 \geq \cdots \geq x_m \geq 0$. We argue that $\arg\max_{x_{\max}=y}\{\sum_{i=1}^m x_i^2\} = \{y, y, \cdots, y, k - ry, 0, \cdots, 0\}$, that is $x_1 = x_2 = \cdots = x_r = y, x_{r+1} = k - ry$ where $r = \lfloor k/y \rfloor$.

To show this, we first consider x_2: if $x_2 = y$, then consider x_3; otherwise, $x_2 < y$, then we search for the largest index j such that $x_j > 0$. Note the fact that if $x \geq y > 0$, then $(x+1)^2 + (y-1)^2 = x^2 + y^2 + 2(x-y) + 2 > x^2 + y^2$. So if we increase x_2 by 1 and decrease x_j by 1, we will enlarge $\sum_{i=1}^m x_i^2$. After we update $x_2 = x_2 + 1, x_j = x_j - 1$, we still get a decreasing sequence $x_1 \geq x_2 \geq \cdots \geq x_m \geq 0$. If we still have $x_2 < y$, then we repeat the same procedure above (by increasing x_2 and decreasing x_j where j is the largest index such that $x_j > 0$). We repeat this process until $x_2 = y$ or $x_1 + x_2 = k$. If it is the former case (i.e, $x_2 = y$), then we consider x_3 and so on. Finally, we will get the sequence $x_1 = x_2 = \cdots = x_r = y, x_{r+1} = k - ry$ where $r = \lfloor k/y \rfloor$ as claimed, and this setting maximizes $\sum_{i=1}^m x_i^2$.

Next we claim that $h(y+1) > h(y)$. The reason is similar to the above. We update the sequence $x_1 = x_2 = \cdots = x_r = y, x_{r+1} = k - ry$ (which corresponding to $h(y)$) from x_1: we increase x_1 by 1; search the largest index s such that $x_s > 0$ and decrease x_s by 1. And then consider x_2 and so on and so forth. This process won't stop until $x_1 = x_2 = \cdots = x_s = y + 1$ and $x_{s+1} = k - s(y+1)$ with $s = \lfloor k/y+1 \rfloor$. Thus $h(y+1) > h(y)$.

By enumerating all the possible values of $x_{\max}$, we split Eqn. (3.5) into three parts as follows (corresponding to the cases when $x_{\max} = k, x_{\max} \in [\lceil (k+1)/2 \rceil, k-1]$ and $x_{\max} \in [\lceil k/m \rceil, \lceil (k+1)/2 \rceil - 1]$) (see the remarks after this equation for how the inequality is derived);

$$p^{2k} \sum_{\substack{x_1+x_2+\cdots+x_m=k \\ x_i \geq 0}} \binom{N_{\max}}{x_1}\binom{N_{\max}}{x_2}\cdots\binom{N_{\max}}{x_m} p^{(k^2-\sum_{i=1}^m x_i^2)/2}$$

$$\text{(A.1)} \quad \leq p^{2k}\binom{N_{\max}}{k} m + p^{2k} \sum_{x_{\max}=\lceil \frac{k+1}{2} \rceil}^{k-1} \left(\binom{m}{1}\binom{N_{\max}}{x_{\max}} \sum_{\substack{\sum_{i=1}^{m-1} y_i = k-x_{\max} \\ 0 \leq y_i \leq x_{\max}}} \binom{N_{\max}}{y_1} \cdots \binom{N_{\max}}{y_{m-1}} p^{x_{\max}(k-x_{\max})} \right) + \binom{mN_{\max}}{k} p^{\frac{(k-1)^2}{4}+2k}$$

The first term on the right hand side of Eqn. (A.1) comes from the fact that if $x_{\max} = k$, then all the possible cases for $(x_1, x_2, \cdots, x_m)$ are $(k, 0, 0, \cdots, 0), \cdots, (0, \cdots, 0, k)$, and there are m cases all together. For each case, the value of each term in the summation is $\binom{N_{\max}}{k}$, giving rise to the first term in Eqn. (A.1).

The third term on the right hand side of Eqn. (A.1) can be derived as follows. First, observe that

$$\sum_{x_{\max}=\lceil \frac{k}{m} \rceil}^{\lceil \frac{k+1}{2} \rceil - 1} \left(\sum_{\substack{x_1+x_2+\cdots+x_m=k \\ x_i \geq 0, \max_i\{x_i\}=x_{\max}}} \binom{N_{\max}}{x_1}\binom{N_{\max}}{x_2}\cdots\binom{N_{\max}}{x_m} \right)$$
$$\leq \sum_{\substack{x_1+x_2+\cdots+x_m=k \\ x_i \geq 0}} \binom{N_{\max}}{x_1}\binom{N_{\max}}{x_2}\cdots\binom{N_{\max}}{x_m} = \binom{mN_{\max}}{k}.$$

On the other hand, as $x_{\max} \leq \lceil (k+1)/2 \rceil - 1 = \lceil (k-1)/2 \rceil$, we have:

$$\frac{k^2 - \sum_{i=1}^m x_i^2}{2} \geq \frac{k^2 - h(x_{\max})}{2} \geq \frac{k^2 - h(\lceil \frac{k-1}{2} \rceil)}{2} \geq \frac{(k-1)^2}{4},$$

where the second inequality uses the fact that $h(y)$ is an increasing function, and the last inequality comes from that $h(\lceil (k-1)/2 \rceil) \leq (\lceil (k-1)/2 \rceil)^2 + (\lceil (k-1)/2 \rceil)^2 + 1 \leq k^2/4 + k^2/4 + 1 = k^2/2 + 1$.

In what remains, it suffices to estimate all the three terms on the right hand side of Eqn. (A.1). We will repeatedly use the well-known combinatorial inequality $\binom{n}{k} < (en/k)^k$.

The first term of Eqn. (A.1): According to the assumptions in Eqn. (3.6), we know

$$p \leq \frac{1}{\sqrt{e}} \left(\frac{1}{n^3 m}\right)^{\frac{1}{2k}} \left(\frac{k}{N_{\max}}\right)^{\frac{1}{2}}.$$

Thus, for the first term of Eqn. (A.1), we have:

$$\text{(A.2)} \qquad p^{2k}\binom{N_{\max}}{k} m < \left(\frac{1}{e^k} \left(\frac{1}{n^3 m}\right) \left(\frac{k}{N_{\max}}\right)^k \right) \left(\frac{eN_{\max}}{k}\right)^k m = \frac{1}{n^3}$$

The second term of Eqn. (A.1): For the second term of Eqn. (A.1), we relax the constraint $x_{\max} \geq y_i \geq 0$ to $y_i \geq 0$. Thus, we have:

$$\sum_{\substack{y_1+\cdots+y_{m-1}=k-x_{\max} \\ x_{\max} \geq y_i \geq 0}} \binom{N_{\max}}{y_1}\cdots\binom{N_{\max}}{y_{m-1}} \leq \sum_{\substack{y_1+\cdots+y_{m-1}=k-x_{\max} \\ y_i \geq 0}} \binom{N_{\max}}{y_1}\cdots\binom{N_{\max}}{y_{m-1}}$$

$$= \binom{(m-1)N_{\max}}{k-x_{\max}} < \left(\frac{e(m-1)N_{\max}}{k-x_{\max}}\right)^{k-x_{\max}} \tag{A.3}$$

Now apply (A.3) to the second term of (A.1), we have (starting from the second line, we replace x_{max} to be j for simplicity):

$$p^{2k} \sum_{x_{\max}=\lceil \frac{k+1}{2} \rceil}^{k-1} \left(\binom{m}{1}\binom{N_{\max}}{x_{\max}} \sum_{\substack{\sum_{i=1}^{m-1} y_i = k-x_{\max} \\ 0 \leq y_i \leq x_{\max}}} \binom{N_{\max}}{y_1}\cdots\binom{N_{\max}}{y_{m-1}} p^{x_{\max}(k-x_{\max})} \right)$$

$$< \sum_{j=\lceil \frac{k+1}{2} \rceil}^{k-1} \left(m \left(\frac{eN_{\max}}{j}\right)^j p^{2k+j(k-j)} \left(\frac{e(m-1)N_{\max}}{k-x_{\max}}\right)^{k-x_{\max}} \right)$$

$$= \sum_{j=\lceil \frac{k+1}{2} \rceil}^{k-1} \left(m^{k-j+1} N_{\max}^k e^k \left(\frac{1}{j}\right)^j \left(\frac{1}{k-j}\right)^{k-j} p^{2k+j(k-j)} \right)$$

$$< \sum_{j=\lceil \frac{k+1}{2} \rceil}^{k-1} \left(m^{k-j+1} N_{\max}^k e^k \left(\frac{2}{k}\right)^k p^{2k+j(k-j)} \right) \tag{A.4}$$

where the last inequality holds due to the inequality of arithmetic and geometric means.

Note that by tedious elementary calculation, we know $[2k+j(k-j)]/k \geq (k-j+1)/2 \geq 1$ when $\lceil (k+1)/2 \rceil \leq j \leq k-1$. Since

$$p \leq \frac{1}{2ek^{\frac{1}{k}}} \left(\frac{1}{n^3m^2}\right)^{\frac{1}{k}} \frac{k}{N_{\max}} < 1$$

by Eqn. (3.6), for each j satisfying $\lceil (k+1)/2 \rceil \leq j \leq k-1$, we have:

$$m^{k-j+1} N_{\max}^k e^k \left(\frac{2}{k}\right)^k p^{2k+j(k-j)}$$

$$\leq m^{k-j+1} N_{\max}^k e^k \left(\frac{2}{k}\right)^k p^{k\left(\frac{k-j+1}{2}\right)}$$

$$\leq m^{k-j+1} N_{\max}^k e^k \left(\frac{2}{k}\right)^k \left(\frac{1}{(2e)^k}\frac{1}{kn^3}\frac{k^k}{N_{\max}^k}\right)^{\frac{k-j+1}{2}} \left(\frac{1}{m}\right)^{k-j+1}$$

$$\leq N_{\max}^k e^k \left(\frac{2}{k}\right)^k \left(\frac{1}{(2e)^k}\frac{1}{kn^3}\frac{k^k}{N_{\max}^k}\right)$$

$$= \frac{1}{kn^3} \tag{A.5}$$

where the inequality on the fourth line holds as $k \leq N_{\max}$ and $(k-j+1)/2 \geq 1$.

The third term of Eqn. (A.1): For the third term of (A.1), note that $(k-1)^2/4+2k > k^2/4$ and by plugging in the condition

$$p \le \frac{1}{e^{\frac{4}{k}}} \left(\frac{1}{n^3 m^k}\right)^{\frac{4}{k^2}} \left(\frac{k}{N_{\max}}\right)^{\frac{4}{k}} < 1,$$

we have

$$\binom{mN_{\max}}{k} p^{\frac{(k-1)^2}{4}+2k} < \left(\frac{emN_{\max}}{k}\right)^k p^{\frac{k^2}{4}} \le \left(\frac{emN_{\max}}{k}\right)^k \frac{1}{e^k}\frac{1}{n^3 m^k}\frac{k^k}{N_{\max}^k} = \frac{1}{n^3} \tag{A.6}$$

Finally, combining (A.2), (A.5) and (A.6), we have:

$$E[\mathrm{I}] \le \frac{1}{n^3} + \frac{k}{2}\cdot\frac{1}{kn^3} + \frac{1}{n^3} = \frac{5}{2n^3}$$

This concludes Lemma 11.

A.2. **The missing details in case (ii) of part (a) of Theorem 10.** Set $N_u := |B_r^{\mathcal{X}_n}(u)|$ and $N_v := |B_r^{\mathcal{X}_n}(v)|$. Let $\tilde{k} := \lfloor \mathsf{K}/2 \rfloor - 2$. Easy to see $\tilde{k} \ge 1$ since $\mathsf{K} \ge 8\beta^2 \ge 8$. For every set S of $(\tilde{k}+2)$ vertices in $\tilde{G}_{uv}^{local}$, let A_S be the event "*S is a clique in $\tilde{G}_{uv}^{local}$ containing edge (u,v) given* F" and Y_S its indicator random variable. Set

$$\mathrm{Y} = \sum_{|S|=\tilde{k}+2} \mathrm{Y}_S.$$

Then Y is the number of cliques of size $(\tilde{k}+2)$ in $\tilde{G}_{uv}^{local}$ containing edge (u,v) given F. Linearity of expectation gives:

$$\mathbb{E}[\mathrm{Y}] = \sum_{|S|=\tilde{k}+2} \mathbb{E}[\mathrm{Y}_S] = \sum_{\substack{x_1+x_2=\tilde{k}\\ 0\le x_1\le N_u-1\\ 0\le x_2\le N_v-1}} \binom{N_u-1}{x_1}\binom{N_v-1}{x_2} p^{(x_1+1)(x_2+1)-1} \tag{A.7}$$

To estimate this quantity, we first prove the following result:

Lemma 27. *If $\tilde{k} \ge 1$ and*

$$p \le \frac{1}{e}\left(\frac{1}{n^3}\right)^{\frac{1}{\tilde{k}}} \frac{\tilde{k}}{N_u+N_v} \tag{A.8}$$

hold, then $\mathbb{E}[\mathrm{Y}] = O\left(n^{-3}\right)$

Proof. Easy to see that if $\tilde{k} > N_u + N_v - 2$, then the summation on the right hand side of Eqn. (A.7) is 0. Now we move on to the case when $\tilde{k} \le (N_u-1)+(N_v-1) < N_u+N_v$. Easy to see $p<1$ in this case. Thus, the right hand side of (A.7) can be bounded from above by:

$$\begin{aligned}
&\sum_{\substack{x_1+x_2=\tilde{k}\\ 0\le x_1\le N_u-1\\ 0\le x_2\le N_v-1}} \binom{N_u-1}{x_1}\binom{N_v-1}{x_2} p^{(x_1+1)(x_2+1)-1}\\
&\le \sum_{i=0}^{\tilde{k}} \binom{N_u}{i}\binom{N_v}{\tilde{k}-i} p^{\tilde{k}+i(\tilde{k}-i)} < \sum_{i=0}^{\tilde{k}} \binom{N_u}{i}\binom{N_v}{\tilde{k}-i} p^{\tilde{k}} = \binom{N_u+N_v}{\tilde{k}} p^{\tilde{k}}\\
&< \left(\frac{e(N_u+N_v)}{\tilde{k}}\right)^{\tilde{k}} p^{\tilde{k}} \le \frac{1}{n^3}
\end{aligned}$$

where the last inequality holds due to condition (A.8). □

Easy to see that there exist two constants c_1^b and c_2^b which depend on the Besicovitch constant β and α, such that

if $\mathsf{K} \geq 8\beta^2$ and $p \leq c_1^b \cdot (1/n)^{c_2^b/\mathsf{K}}$, then the conditions in Eqn. (A.8) will hold.

On the other hand, we have

$$\mathbb{P}\left[\omega_{u,v}\left(\tilde{G}_{uv}^{local}\right) \geq \mathsf{K}/2 \middle| \mathrm{F}\right] = \mathbb{P}[Y > 0] \leq \mathbb{E}[Y]$$

Thus, by Lemma 27, we know that

$$\text{If } \mathsf{K} \geq 8\beta^2 \text{ and } p \leq c_1^b \cdot (1/n)^{c_2^b/\mathsf{K}}, \text{ then } \mathbb{P}\left[\omega_{u,v}\left(\tilde{G}_{uv}^{local}\right) \geq \mathsf{K}/2 \middle| \mathrm{F}\right] = O(n^{-3}) \tag{A.9}$$

A.3. **The proof of Lemma 12.**

Proof. By using a similar argument as in Appendix A.1, it is easy to see that the maximum value of $\sum_{i=1}^m x_i^2$, under the constraints $\sum_{i=1}^m x_i = k$ and $x_i \in [0, N_{\max}]$ for each $i \in [1, m]$, is $rN_{\max}^2 + (k - rN_{\max})^2$ where $r = \lfloor k/N_{\max} \rfloor$. The maximum can be achieved when $(x_1, x_2, \cdots, x_m) = (\underbrace{N_{\max}, \cdots, N_{\max}}_{r}, k - rN_{\max}, \ldots)$. Thus, we have

$$\begin{aligned}
\mathbb{E}[\mathrm{I}] &\leq p^{2k} \sum_{\substack{x_1+x_2+\cdots+x_m=k \\ x_i \geq 0}} \binom{N_{\max}}{x_1}\binom{N_{\max}}{x_2}\cdots\binom{N_{\max}}{x_m} p^{(k^2 - \sum_{i=1}^m x_i^2)/2} \\
&< \left(\sum_{\substack{x_1+x_2+\cdots+x_m=k \\ x_i \geq 0}} \binom{N_{\max}}{x_1}\binom{N_{\max}}{x_2}\cdots\binom{N_{\max}}{x_m} \right) \cdot p^{\frac{k^2 - \left(rN_{\max}^2 + (k - rN_{\max})^2\right)}{2} + 2k} \\
&= \binom{mN_{\max}}{k} p^{\frac{k^2 - \left(rN_{\max}^2 + (k - rN_{\max})^2\right)}{2} + 2k} \\
&< \left(\frac{emN_{\max}}{k}\right)^k p^{k(rN_{\max}+1) - \frac{r(r+1)}{2} N_{\max}^2} \\
&< \left(\frac{emN_{\max}}{k}\right)^k p^{k\left(\left(\frac{k}{N_{\max}} - 1\right)N_{\max} + 1\right) - \frac{1}{2}\frac{k}{N_{\max}}\left(\frac{k}{N_{\max}} + 1\right) N_{\max}^2} && \text{(A.10)} \\
&= \left(\frac{emN_{\max}}{k} p^{\frac{1}{2}k - \frac{3}{2}N_{\max} + 1}\right)^k && \text{(A.11)}
\end{aligned}$$

where Eqn. (A.10) holds since $k/N_{\max} \geq r > k/N_{\max} - 1$.

Pick a constant C_3, which only depends on the Besicovitch constant β and α, such that

$$\frac{C_3 \lfloor \log_{1/p} n \rfloor}{2|\Lambda|} - 3 \geq 16 \log_{1/p} n \geq 6N_{\max}.$$

This can be done since $N_{\max}$ is a constant and $(1/n)^{8/(3N_{\max})} \leq p < 1$ implies $\log_{1/p} n \geq 3N_{\max}/8$. Set $\mathsf{K} = C_3 \lfloor \log_{1/p} n \rfloor$. Recall that $k = \lfloor \mathsf{K}/(2|\Lambda|) \rfloor - 2$, thus $k \geq 16 \log_{1/p} n \geq 6N_{\max}$. Also note that $m \leq n$. Hence, we have the following inequality.

$$\left(\frac{emN_{\max}}{k} p^{\frac{1}{2}k - \frac{3}{2}N_{\max} + 1}\right)^k < \left(\frac{enN_{\max}}{k} p^{\frac{1}{4}k}\right)^k \leq \left(\frac{eN_{\max}}{k} n^{-3}\right)^k = O(n^{-3}) \tag{A.12}$$

Finally, combining (A.11) and (A.12), we have $\mathbb{E}[\mathrm{I}] = O(n^{-3})$. □

A.4. **Proof of Lemma 14.** We will use the standard second moment method to prove this lemma. For completeness, we first state the second moment method. For those who are familiar with this method, our main technical step is to estimate the summation on the right hand side of Eqn. (A.13).

Definition 8 (Symmetric random variables)**.** *We say random variables $Z_1, \cdots, Z_m$, where Z_i is the indicator random variable for event U_i, are* ***symmetric*** *if for every $i \neq j$ there is a measure preserving mapping of the underlying probability space that permutes the m events and sends event U_i to event U_j.*

Let Z be a nonnegative integral-valued random variable, and suppose we have a decomposition $Z = Z_1 + \cdots + Z_m$, where Z_i is the indicator random variable for event U_i and $Z_1, \cdots, Z_m$ are symmetric. For indices i, j, write $i \sim j$ if $i \neq j$ and the events U_i, U_j are not independent. For any fixed index i, we set

$$\Delta^* := \sum_{j \sim i} \mathbb{P}[U_j \mid U_i],$$

and note that by the symmetry of Z_i, Δ^* is independent of the index i (thus we are not denoting it by Δ_i^*).

Theorem 28 (The second moment method [1])**.** *If $\mathbb{E}[Z] \to \infty$ and $\Delta^* = o(\mathbb{E}[Z])$ as $m \to \infty$, then $\mathbb{P}[Z = 0] \to 0$.*

Now we are ready to prove Lemma 14.

Proof of Lemma 14. Set $k = \lfloor \log_{1/p} n \rfloor$. Now consider all the $k-$set S_i of vertices in $G(n, p)$. Let U_i be the event "S_i is a clique" and Z_i its indicator random variable. (All $k-$sets "look the same" so that the Z_i's are symmetric.) I is a finite index set enumerating all the $k-$sets in $G(n, p)$. Set

$$Z = \sum_{i \in I} Z_i$$

so that Z is the number of $k-$cliques in $G(n, p)$. Linearity of Expectation gives:

$$\mathbb{E}[Z] = \sum_{i \in I} \mathbb{E}[Z_i] = \binom{n}{k} p^{\binom{k}{2}}$$

Easy to see that

$$\Delta^* = \sum_{j \sim i} \mathbb{P}[U_j | U_i] = \sum_{\ell=2}^{k-1} \binom{k}{\ell} p^{\binom{k}{2} - \binom{\ell}{2}} \binom{n-k}{k-\ell}$$

Since $k = \lfloor \log_{1/p} n \rfloor$ and $p \leq (1/n)^{1/\sqrt[4]{n}}$, we know that $p^{k-1} > p^k > 1/n$, $k \leq n^{1/4}$ and $\log k / \log n \leq 1/4$. Also note that $p \geq (1/n)^{1/11}$. Easy to see that

$$k + 1 > \log_{1/p} n \geq \frac{1}{\xi} = 11$$

which further implies $k > 10$.

Note that for sufficiently large n, we have $n - k > n/2$. Thus, using $p^{k-1} > 1/n$ as derived earlier, we have:

$$\mathbb{E}[Z] = \binom{n}{k} p^{\binom{k}{2}} > \frac{(n-k)^k}{k^k} p^{\frac{k(k-1)}{2}} > \left(\frac{n}{2k}\right)^k n^{-\frac{k}{2}} = n^{\frac{k}{2}\left(1 - \frac{2\log(2k)}{\log n}\right)} > n^{\frac{k}{4}} \to \infty$$

To apply Theorem 28, it suffices to estimate the term $\Delta^*/\mathbb{E}[Z]$.

$$\frac{\Delta^*}{\mathbb{E}[Z]} = \frac{\sum_{\ell=2}^{k-1} \binom{k}{\ell}\binom{n-k}{k-\ell}(p)^{\binom{k}{2}-\binom{\ell}{2}}}{\binom{n}{k}(p)^{\binom{k}{2}}} = \sum_{\ell=2}^{k-1} \frac{\binom{k}{\ell}\binom{n-k}{k-\ell}}{\binom{n}{k}}(p)^{-\binom{\ell}{2}} \tag{A.13}$$

We estimate the summation on the right hand side term by term. Let

$$g(\ell) := \frac{\binom{k}{\ell}\binom{n-k}{k-\ell}}{\binom{n}{k}}(p)^{-\binom{\ell}{2}}.$$

Note that for $\ell \in [2, k-1]$, we have

$$\begin{aligned} g(\ell) &= \frac{\binom{k}{\ell}\binom{n-k}{k-\ell}}{\binom{n}{k}}(p)^{-\binom{\ell}{2}} \le \frac{\binom{k}{\ell}\frac{(n-k)^{k-\ell}}{(k-\ell)!}}{\frac{(n-k)^k}{k!}} n^{\frac{1}{k}\frac{\ell(\ell-1)}{2}} \le \frac{\binom{k}{\ell}k^\ell}{(n-k)^\ell} n^{\frac{1}{k}\frac{\ell(\ell-1)}{2}} \\ &\le \frac{\left(\frac{ek}{\ell}\right)^\ell k^\ell}{(n-k)^\ell} n^{\frac{1}{k}\frac{\ell(\ell-1)}{2}} = n^{-\frac{\ell \log\left(\frac{\ell(n-k)}{ek^2}\right)}{\log n} + \frac{1}{k}\frac{\ell(\ell-1)}{2}} \end{aligned}$$

Now set

$$h(\ell) = -\frac{\ell \log\left(\frac{\ell(n-k)}{ek^2}\right)}{\log n} + \frac{1}{k}\frac{\ell(\ell-1)}{2};$$

and thus by the above inequality we have $g(\ell) \le n^{h(\ell)}$. We claim that $\forall \ell \in [2, k-1], h(\ell) \le \max\{h(2), h(k-1)\}$. We then then further use $h(2)$ and $h(k-1)$ to derive an upper bound on $g(l)$.

Indeed, by the following direct calculation, we can easily prove this:

Note that its derivative with respect to ℓ is

$$h'(\ell) = -\frac{\log \ell + \log(n-k) - 2\log k}{\log n} + \frac{2\ell - 1}{2k}.$$

Further calculate its second derivative:

$$h''(\ell) = -\frac{1}{\log n}\frac{1}{\ell} + \frac{1}{k}$$

. Note that $\ell_0 = k/\log n$ is the only solution of $h''(\ell) = 0$. Easy to check that $\ell_0 \le k - 1$. Therefore, we have the following two cases:

- **Case (i):** If $\ell_0 < 2$, then $h'(\ell)$ is strictly increasing on $\ell \in [2, k-1]$;
- **Case (ii):** If $\ell_0 \in [2, k-1]$, then $h'(\ell)$ is strictly decreasing on $[2, \ell_0]$ and strictly increasing on $[\ell_0, k-1]$.

Note that

$$h'(2) < -\frac{\log 2 + \log(n/2) - 2\log k}{\log n} + \frac{3}{2k} = -1 + \frac{2\log k}{\log n} + \frac{3}{2k} < -1 + \frac{1}{2} + \frac{3}{20} < 0.$$

Thus in either case $h'(\ell)$ can become 0 at most once within $\ell \in [2, k-1]$, and we have $\max_{\ell \in [2,k-1]} h(\ell) = \max\{h(2), h(k-1)\}$.

Routine calculation shows that (using that $n-k>n/2$), for n large enough:

$$h(2) < -\frac{2\left[\log\left(2(n/2)\right)-1-2\log k\right]}{\log n}+\frac{1}{k} = -2+\frac{1}{k}+\frac{2}{\log n}+\frac{4\log k}{\log n} < -\frac{1}{2},$$

$$\begin{aligned} h(k-1) &< -\frac{(k-1)\left[\log\left(n/2\right)-1-\log k-\log\left(k/(k-1)\right)\right]}{\log n}+\frac{k^2-3k+2}{2k} \\ &< \left[\frac{k^2-3k+2}{2k}-(k-1)\right]+\frac{k(1+\log 2)}{\log n}+\frac{k\log k}{\log n}+\frac{k\log\left(k/(k-1)\right)}{\log n} \\ &< -\frac{1}{2}+\frac{1}{10}-\frac{k}{6} < -\frac{1}{2}. \end{aligned}$$

Thus, $\forall \ell \in [2, k-1]$, we have $g(\ell) < n^{-1/2}$. It then follows that

$$\sum_{\ell=2}^{k-1} g(\ell) < k \cdot n^{-\frac{1}{2}} \leq n^{\frac{1}{4}} \cdot n^{-\frac{1}{2}} = n^{-\frac{1}{4}}.$$

Hence by Eqn (A.13), we have $\Delta^*/\mathbb{E}[Z] < n^{-1/4} \to 0$, and therefore $\mathbb{P}[Z=0] \to 0$ by Theorem 28. □

A.5. **Proof of Theorem 18.**

Proof of part (a). We use the same notation $\tilde{A}_{uv}$ and B_{uv} as in the proof of Theorem 10. Now we set

$$N_{\max} := \frac{5\log n}{\log\left(\log n/(\sigma\theta 2^d n r^d)\right)}.$$

Again, denote F to be the event that "for every $v \in \mathcal{X}_n$, the ball $B_r(v) \cap \mathcal{X}_n$ contains at most $N_{\max}$ points"; and F^c denotes the complement event of F. By Lemma 17, we know that, $\mathbb{P}[\mathrm{F}^c] = O(n^{-3})$.

Let K_n be a positive number to be determined such that $\mathsf{K}_n \to \infty$ as $n \to \infty$. By applying the pigeonhole principle and the union bound, we have:

$$\begin{aligned} &\mathbb{P}\left[\omega_{u,v}\left(G_n^{0,p}\right) \geq \mathsf{K}_n \middle| \mathrm{F}\right] \\ \text{(A.14)} \qquad &\leq \mathbb{P}\left[\omega_{u,v}\left(G_n^{0,p}|_{\tilde{A}_{uv}}\right) \geq \mathsf{K}_n/2 \middle| \mathrm{F}\right] + \mathbb{P}\left[\omega_{u,v}\left(G_n^{0,p}|_{B_{uv}}\right) \geq \mathsf{K}_n/2 \middle| \mathrm{F}\right] \end{aligned}$$

Case (i): bounding the first term in Eqn. (A.14). Applying Theorem 7 for points in A_{uv} gives a well-separated clique-partitions family $\mathcal{P} = \{P_i\}_{i\in\Lambda}$ of A_{uv} with $|\Lambda| \leq \beta^2$ being a constant. Augment each P_i to $\tilde{P}_i = P_i \cup \{u\} \cup \{v\}$. Check Figure 2 (a). Again, by applying pigeonhole principle and the union bound, we have:

$$\text{(A.15)} \qquad \mathbb{P}\left[\omega_{u,v}\left(G_n^{0,p}|_{\tilde{A}_{uv}}\right) \geq \mathsf{K}_n/2 \middle| \mathrm{F}\right] \leq \sum_{i=1}^{|\Lambda|} \mathbb{P}\left[\omega_{u,v}\left(G_n^{0,p}|_{\tilde{P}_i}\right) \geq \mathsf{K}_n/(2|\Lambda|) \middle| \mathrm{F}\right]$$

Now set $k_n := \lfloor \mathsf{K}_n/(2|\Lambda|) \rfloor - 2$. Easy to see that $k_n \to \infty$ as $n \to \infty$. Same as in the proof for Theorem 10, we have

$$\begin{aligned} &\mathbb{P}\left[\omega_{u,v}\left(G_n^{0,p}|_{\tilde{P}_i}\right) \geq k_n+2 \middle| \mathrm{F}\right] \\ \text{(A.16)} \qquad &\leq p^{2k_n} \sum_{\substack{x_1+x_2+\cdots+x_m=k_n \\ 0\leq x_i\leq N_{\max}}} \binom{N_{\max}}{x_1}\binom{N_{\max}}{x_2}\cdots\binom{N_{\max}}{x_m} p^{(k_n^2-\sum_{i=1}^m x_i^2)/2} \end{aligned}$$

where $m \leq n$ is the number of $C_i^{(j)}$ in the clique-partition $\tilde{P}_i$.

If $\mathsf{K}_n \le 2N_{\max}$, then $k_n \in [1, N_{\max}]$. By applying Lemma 11, we have that if $1 \le k_n \le N_{\max}$, then there exist constants c_1^a and c_2^a (which depend on the Besicovitch constant β), such that if

$$p \le c_1^a \cdot \left(\frac{1}{n}\right)^{c_2^a/\mathsf{K}_n} \frac{\mathsf{K}_n}{N_{\max}},$$

then the right hand side of Eqn. (A.16) is $O(n^{-3})$.

Thus, following the same argument in the proof of part (a) of Theorem 10, we have

$$\text{If } \mathsf{K}_n \le 2N_{\max} \text{ and } p \le c_1^a \cdot (1/n)^{c_2^a/\mathsf{K}_n} (\mathsf{K}_n/N_{\max}),$$
$$\text{then } \mathbb{P}\left[\omega_{u,v}\left(G_n^{0,p}|_{\tilde{A}_{uv}}\right) \ge \mathsf{K}_n/2 \middle| \mathrm{F}\right] = O(n^{-3}). \tag{A.17}$$

Finally, suppose $\mathsf{K}_n > \mathsf{K}_0 = 2N_{\max}$. Using Eqn (A.17), we know that if

$$p \le c_1^a \cdot \left(\frac{1}{n}\right)^{c_2^a/\mathsf{K}_0} \frac{\mathsf{K}_0}{N_{\max}} = 2c_1^a \left(\frac{\sigma\theta 2^d \left(nr^d\right)}{\log n}\right)^{\frac{c_2^a}{10}} = \left(2c_1^a \cdot \left(\sigma\theta 2^d\right)^{\frac{c_2^a}{10}}\right) \left(\frac{nr^d}{\log n}\right)^{\frac{c_2^a}{10}}$$

and $\mathsf{K}_n > \mathsf{K}_0$, then

$$\mathbb{P}\left[\omega_{u,v}\left(G_n^{0,p}|_{\tilde{A}_{uv}}\right) \ge \mathsf{K}_n/2 \middle| \mathrm{F}\right] \le \mathbb{P}\left[\omega_{u,v}\left(G_n^{0,p}|_{\tilde{A}_{uv}}\right) \ge \mathsf{K}_0/2 \middle| \mathrm{F}\right] = O(n^{-3}).$$

Set $C_1^a := 2c_1^a \cdot \left(\sigma\theta 2^d\right)^{c_2^a/10}$ and $C_2^a := c_2^a/10$ be two constants. Combining this with Eqn. (A.17), we thus obtain that:

$$\text{If } \mathsf{K}_n \to \infty \text{ and } p \le \min\left\{C_1^a \cdot \left(nr^d/\log n\right)^{C_2^a},\ c_1^a \cdot (1/n)^{c_2^a/\mathsf{K}_n} (\mathsf{K}_n/N_{\max})\right\},$$
$$\text{then } \mathbb{P}\left[\omega_{u,v}\left(G_n^{0,p}|_{\tilde{A}_{uv}}\right) \ge \mathsf{K}_n/2 \middle| \mathrm{F}\right] = O(n^{-3}). \tag{A.18}$$

Case (ii): bounding the second term in Eqn. (A.14). Recall that $B_{uv} = B_r^{\mathcal{X}_n}(u) \cup B_r^{\mathcal{X}_n}(v)$ (see Figure 2 (b)). We again use the notation $\tilde{G}_{uv}^{local}$ defined in the proof of part (a) of Theorem 10. Set $N_u := |B_r^{\mathcal{X}_n}(u)|$ and $N_v := |B_r^{\mathcal{X}_n}(v)|$. Let $\tilde{k}_n := \lfloor \mathsf{K}_n/2 \rfloor - 2$. Easy to see $\tilde{k}_n \ge 1$. Using the same argument as in Case (ii) in the proof of Theorem 10, we have

$$\mathbb{P}\left[\omega_{u,v}\left(\tilde{G}_{uv}^{local}\right) \ge \mathsf{K}_n/2 \middle| \mathrm{F}\right] \le \sum_{\substack{x_1+x_2=\tilde{k}_n \\ 0\le x_1 \le N_u-1 \\ 0 \le x_2 \le N_v - 1}} \binom{N_u-1}{x_1}\binom{N_v-1}{x_2} p^{(x_1+1)(x_2+1)-1} \tag{A.19}$$

By applying Lemma 27, we know that there exist constants c_1^b and c_2^b (which depend on the Besicovitch constant β), such that if $\mathsf{K}_n \to \infty$ and

$$p \le c_1^b \cdot \left(\frac{1}{n}\right)^{c_2^b/\mathsf{K}_n} \frac{\mathsf{K}_n}{N_{\max}},$$

then the right hand side of Eqn. (A.19) is $O(n^{-3})$. That is,

$$\text{if } \mathsf{K}_n \to \infty \text{ and } p \le\ c_1^b \cdot (1/n)^{c_2^b/\mathsf{K}_n} (\mathsf{K}_n/N_{\max}),$$
$$\text{then } \mathbb{P}\left[\omega_{u,v}\left(\tilde{G}_{uv}^{local}\right) \ge \mathsf{K}_n/2 \middle| \mathrm{F}\right] = O(n^{-3}) \tag{A.20}$$

Pick

$$\mathsf{K}_n = 4N_{\max} = \frac{20\log n}{\log\left(\log n/(\sigma\theta 2^d nr^d)\right)} = \frac{20\log n}{\log\left(\log n/nr^d\right)} + \text{const.}.$$

Note that this makes the first term of the constraint on p in Eqn. (A.18) dominate. Thus, combining Eqn. (A.20), (A.18) and (A.14), there exist constants $C_1 = \min\{C_1^a, c_1^b\}$ and $C_2 = \max\{C_2^a, c_2^b/10\}$ such that if p satisfies conditions in Eqn. (3.14), then

$$\mathbb{P}\left[\omega_{u,v}\left(G_n^{0,p}\right) \geq \mathsf{K}_n\right] \leq \mathbb{P}\left[\omega_{u,v}\left(G_n^{0,p}\right) \geq \mathsf{K}_n \middle| \mathrm{F}\right] + \mathbb{P}[\mathrm{F}^c] = O(n^{-3})$$

Finally, by applying the union bound, we derive that with high probability, for each of the $O(n^2)$ long-edge (u, v), its edge clique number

$$\omega_{u,v}(G_n^{0,p}) \lesssim \frac{\log n}{\log\left(\log n/nr^d\right)}$$

as long as Eqn. (3.14) holds. This completes the proof of Part (a) if Theorem 18.

Proof of part (b). We again try to bound the two terms on the right hand side of Eqn. (A.14) from above separately. For case (i), our result relies on the following lemma.

Lemma 29. *There exists a constant $C_3 > 0$ depending on the Besicovitch constant β such that if $\left(nr^d/\log n\right)^{4/15} \leq p < 1$ and $\mathsf{K}_n = C_3\lfloor \log_{1/p} n\rfloor$, then we have*

$$\mathbb{P}\left[\omega_{u,v}\left(G_n^{0,p}|_{\tilde{A}_{uv}}\right) \geq \mathsf{K}_n/2 \middle| \mathrm{F}\right] = O(n^{-3}).$$

Proof. By a similar argument in Appendix A.3, we know that

$$\mathbb{P}\left[\omega_{u,v}\left(G_n^{0,p}|_{\tilde{A}_{uv}}\right) \geq \mathsf{K}_n/2 \middle| \mathrm{F}\right] \leq \frac{1}{\sqrt{2\pi}}\left(\frac{emN_{\max}}{k_n} p^{\frac{1}{2}k_n - \frac{3}{2}N_{\max}+1}\right)^{k_n} \tag{A.21}$$

where $k_n = \lfloor \mathsf{K}_n/(2|\Lambda|)\rfloor - 2$. Pick a constant C_3, which only depends on the Besicovitch constant β, such that

$$\frac{C_3\lfloor \log_{1/p} n\rfloor}{2|\Lambda|} - 3 \geq 16\log_{1/p} n \geq 6N_{\max}.$$

This can be done since we have $\log n/nr^d \to \infty$ and $\left(nr^d/\log n\right)^{4/15} \leq p < 1$ and thus

$$\begin{aligned}\log_{1/p} n \geq \frac{15\log n}{4\log\left(\log n/nr^d\right)} &= \frac{3}{8}\frac{5\log n}{(1/2)\log\left(\log n/nr^d\right)}\\ &> \frac{3}{8}\frac{5\log n}{\log\left(\log n/(\sigma\theta 2^d nr^d)\right)} = \frac{3}{8}N_{\max}.\end{aligned}$$

Set $\mathsf{K}_n = C_3\lfloor \log_{1/p} n\rfloor$. Recall that $k_n = \lfloor \mathsf{K}/(2|\Lambda|)\rfloor - 2$, thus $k_n \geq 16\log_{1/p} n \geq 6N_{\max}$. Finally, we use Eqn. (A.12) (with k being replaced by k_n) to complete the proof. □

Now pick such C_3 in Lemma 29. We know that the following statement holds.

$$\begin{aligned}&\text{If } \left(nr^d/\log n\right)^{4/15} \leq p < 1,\\ &\text{then } \mathbb{P}\left[\omega_{u,v}\left(G_n^{0,p}|_{\tilde{A}_{uv}}\right) \geq C_3\lfloor \log_{1/p} n\rfloor/2 \middle| \mathrm{F}\right] = O(n^{-3}).\end{aligned} \tag{A.22}$$

For case (ii), we know that if event F has already happened, then $|B_{uv}| \leq 2N_{\max}$. Note that if $\left(nr^d/\log n\right)^{C_3/42} \leq p < 1$, then we have

$$\begin{aligned}\frac{C_3\lfloor \log_{1/p} n\rfloor}{2} &> 2\frac{5\log n}{(1/2)\log\left(\log n/nr^d\right)} + \left(\frac{\log n}{\log\left(\log n/nr^d\right)} - \frac{C_3}{2}\right)\\ &> 2\frac{5\log n}{\log\left(\log n/(\sigma\theta 2^d nr^d)\right)} = 2N_{\max} \geq |B_{uv}|.\end{aligned}$$

Set $\xi = \min\{4/15, C_3/42\}$. Hence, we obtain that:

$$\text{(A.23)} \quad \text{If } \left(nr^d/\log n\right)^\xi \le p < 1, \text{ then } \mathbb{P}\left[\omega_{u,v}\left(G_n^{0,p}|_{B_{uv}}\right) \ge C_3\lfloor \log_{1/p} n\rfloor/2 \middle| \mathrm{F}\right] = 0.$$

Thus, combining Eqn. (A.23), (A.22) and (A.14), we know that if $\left(nr^d/\log n\right)^\xi \le p < 1$, then with high probability, for every long-edge (u, v), its edge clique number $\omega_{u,v}\left(G_n^{0,p}\right) \lesssim \log_{1/p} n$. This completes the proof of Theorem 18.

APPENDIX B. THE MISSING PROOFS IN SECTION 4

B.1. **Proof of Part (III) — "dense" regime.** In this section, we discuss the order of $\omega(G_n^{q,0})$ in the regime $\sigma nr^d/\log n \to t \in (0, \infty)$.

Proof of upper bound. We first focus on the upper bound of $\omega\left(G_n^{q,0}\right)$. Let N be a random variable sampled from $Poisson\left((1+\delta)\, n\right)$ for some $\delta > 0$ (say $\delta = 1/2$). Note that G_N is an $r-$neighborhood graph of the Poisson point process $\mathcal{P}_{(1+\delta)n}$ with intensity $(1+\delta)nf$. Completely analogously to the proof of upper bound in Section 4.1, we have

$$\mathbb{P}\left[\omega\left(G_N^{q,0}\right) \ge k_n\right] < \frac{1}{\sqrt{2\pi}}\left(\frac{(1+\delta)e\sigma\theta nr^d(1-q)^{(k_n-1)/2}}{k_n}\right)^{k_n}$$

Finally, pick $k_n = 3\log_{1/(1-q)} nr^d$. Since in the "dense" regime $\sigma nr^d/\log n \to t \in (0, \infty)$, routine calculations show that $\mathbb{P}\left[\omega\left(G_N^{q,0}\right) \ge k_n\right] = o(1)$. Hence, a.s.

$$\omega\left(G_n^{q,0}\right) \lesssim \log\left(nr^d\right)$$

Proof of lower bound. Now let us move on to the lower bound of $\omega\left(G_n^{q,0}\right)$. For this regime, we need slightly stronger condition on the range of t. That is, $\sigma nr^d \ge T\log n$ for some constant $T > 0$ to be determined.

Completely analogously to the proof of lower bound in Section 4.1, let N be a random variable sampled from $Poisson\left((1-\delta')n\right)$ for some $\delta' \in (0, 1)$ (say $\delta' = 1/2$). Set k_n be an integer to be determined. Now, we have

$$\mathbb{P}\left[\omega\left(G_n^{q,0}\right) \le k_n\right] \le \mathbb{P}\left[\omega\left(G_N^{q,0}\right) \le k_n\right] + e^{-\gamma' n}$$

for some constant $\gamma' > 0$ (depending on δ') by Lemma 5. Now fix some constant $\rho \in (0, 1)$ (say $\rho = 1/2$). Recall $W_{1/2} = B_{1/2}(0)$. By Lemma 24, there exist points $x_1, x_2, \cdots, x_m$ with $m = \Omega\left(r^{-d}\right) \ge 1$ such that the sets $x_i + W_{1/2}$ are disjoint and

$$\nu\left(x_i + W_{1/2}\right) \ge \frac{(1-\rho)\sigma\theta}{2^d} r^d$$

for $i = 1, \cdots, m$. Let $\mathbf{X}_i$ be the set of points of G_N falling in $x_i + W_{1/2}$. Then, we have

$$\begin{aligned}\mathbb{P}\left[\omega\left(G_N^{q,0}\right) \le k_n\right] &\le \mathbb{P}\left[\omega\left(G_N^{q,0}\left|_{\mathbf{X}_1}\right.\right) \le k_n, \cdots, \omega\left(G_N^{q,0}\left|_{\mathbf{X}_m}\right.\right) \le k_n\right] \\ &= \prod_{i=1}^m \mathbb{P}\left[\omega\left(G_N^{q,0}\left|_{\mathbf{X}_i}\right.\right) \le k_n\right]\end{aligned}$$

Easy to see that all the points falling in any $r/2$-ball span a clique in G_N. Thus, for each i, we have

$$\mathbb{P}\left[\omega\left(G_N^{q,0}\left|_{\mathbf{X}_i}\right.\right) \le k_n\right] = \mathbb{P}\left[\omega\left(G\left(|\mathbf{X}_i|, 1-q\right)\right) \le k_n\right].$$

Set

$$\Phi_n := \frac{(1-\rho)(1-\delta')\sigma\theta nr^d}{2^{d+1}}$$

which goes to infinity as n grows. Note that $|\mathbf{X}_i| \sim Poisson(\tilde{\lambda})$ where $\tilde{\lambda} := (1-\delta')n \cdot \nu(x_i + W_{1/2}) \geq 2\Phi_n$ (see Eqn. (4.3)). Now pick $k_n := \left\lfloor \log_{1/(1-q)} \Phi_n \right\rfloor = \Omega\left(\log(nr^d)\right)$. By the law of total probability, we have

$$\begin{aligned}
&\mathbb{P}\left[\omega\left(G\left(|\mathbf{X}_i|, 1-q\right)\right) \leq k_n\right] \\
\leq &\mathbb{P}\left[|\mathbf{X}_i| \leq \Phi_n\right] + \sum_{j=\lceil\Phi\rceil}^{\infty} \mathbb{P}\left[\omega\left(G\left(j, 1-q\right)\right) \leq k_n\right] \mathbb{P}\left[|\mathbf{X}_i| = j\right] \\
\leq &\mathbb{P}\left[|\mathbf{X}_i| \leq \frac{\tilde{\lambda}}{2}\right] + \sum_{j=\lceil\Phi\rceil}^{\infty} \mathbb{P}\left[\omega\left(G\left(j, 1-q\right)\right) \leq \left\lfloor \log_{\frac{1}{1-q}} j \right\rfloor\right] \frac{e^{-\tilde{\lambda}}\tilde{\lambda}^j}{j!} \\
< &e^{-\frac{1}{10}\tilde{\lambda}} + \sum_{j=\lceil\Phi\rceil}^{\infty} e^{-j} \frac{e^{-\tilde{\lambda}}\tilde{\lambda}^j}{j!} \qquad \text{(B.1)} \\
< &e^{-\frac{1}{10}\tilde{\lambda}} + e^{-\tilde{\lambda}} \sum_{j=0}^{\infty} \frac{\left(\tilde{\lambda}/e\right)^j}{j!} \\
= &e^{-\frac{1}{10}\tilde{\lambda}} + e^{-\tilde{\lambda}} \cdot e^{\tilde{\lambda}/e} \\
< &2e^{-\frac{1}{10}\tilde{\lambda}}
\end{aligned}$$

Inequality (B.1) holds due to Lemma 5 and Lemma 25. Now set

$$T := \frac{10 \cdot 2^d}{(1-\rho)(1-\delta')\theta}.$$

Note that $\sigma nr^d \geq T \log n$. Then

$$e^{-\frac{1}{10}\tilde{\lambda}} \leq e^{-\frac{(1-\rho)(1-\delta')\theta}{10 \cdot 2^d}(\sigma nr^d)} \leq e^{-\log n} = n^{-1}$$

It follows that $\mathbb{P}\left[\omega\left(G_n^{q,0}\right) \leq k_n\right] = o(1)$ with $k_n = \Omega\left(\log(nr^d)\right)$, which concludes the proof of part (III) of Theorem 2.

DEPARTMENT OF MATHEMATICS, THE OHIO STATE UNIVERSITY, USA
E-mail address: `mkahle@math.osu.edu`

COMPUTER SCIENCE AND ENGINEERING DEPT., THE OHIO STATE UNIVERSITY, USA
E-mail address: `tian.394@osu.edu`

COMPUTER SCIENCE AND ENGINEERING DEPT., THE OHIO STATE UNIVERSITY, USA
E-mail address: `yusu@cse.ohio-state.edu`