On the Entropy of a Random Geometric Graph

Praneeth Kumar Vippathalla, Justin P. Coon, and Mihai-Alin Badiu

***Abstract*—In this paper, we study the entropy of a *hard* random geometric graph (RGG), a commonly used model for spatial networks, where the connectivity is governed by the distances between the nodes. Formally, given a connection range r, a hard RGG G_m on m vertices is formed by drawing m random points from a spatial domain, and then connecting any two points with an edge when they are within a distance r from each other. The two domains we consider are the d-dimensional unit cube $[0,1]^d$ and the d-dimensional unit torus $\mathbb{T}^d$. We derive upper bounds on the entropy $H(G_m)$ for both these domains and for all possible values of r. In a few cases, we obtain an exact asymptotic characterization of the entropy by proving a tight lower bound. Our main results are that $H(G_m) \sim dm\log_2 m$ for $0 < r \leq 1/4$ in the case of $\mathbb{T}^d$ and that the entropy of a one-dimensional RGG on $[0,1]$ behaves like $m\log_2 m$ for all $0 < r < 1$. As a consequence, we can infer that the asymptotic structural entropy of an RGG on $\mathbb{T}^d$, which is the entropy of an unlabelled RGG, is $\Omega((d-1)m\log_2 m)$ for $0 < r \leq 1/4$. For the rest of the cases, we conjecture that the entropy behaves asymptotically as the leading order terms of our derived upper bounds.**

I. Introduction

In many real-world networks, such as wireless, brain and social networks, connectivity is governed by the spatial separation of entities. For instance, devices in a wireless network share a communication link if they are geographically close enough to one another. A random network model that captures the fundamental aspects of these networks, and is widely used in their study, is the random geometric graph (RGG) model. Here, a random graph is generated by scattering m nodes uniformly at random in a spatial domain and drawing an edge between any two nodes based on the distance between them. Over the past few decades, a variety of problems related to RGGs have been studied in detail, aiming to understand the statistics of graph properties [1], [2], the recoverability of the latent spatial embedding [3], and their combinatorial properties [4]. In contrast, we are interested in the information-theoretic compression of these graphs.

In order to develop efficient storage methods for large graph datasets, researchers have recently focused on studying graph compression [5]–[15]. The graph sources are typically non-standard and are unlike independent and identically distributed (i.i.d.) sources, which makes their study particularly challenging. Nevertheless, many works have attempted to characterize Shannon entropy of various graph sources, because of its importance in lossless compression. For a random source X, it is well-known that the minimum expected length of a prefix-free code $l^*(X)$ is within 1 bit from the entropy $H(X)$ [16], i.e.,

$$H(X) \leq l^*(X) < H(X) + 1,$$

and the minimum expected length of a one-to-one code $\ell^*(X)$ is close to the entropy [17], [11, Lem. 1]:

$$H(X) - \log_2(H(X)+1) - \log_2 e \leq \ell^*(X) < H(X).$$

Results on the characterization of entropy are known for Erdős-Rényi (ER) random graphs and their structures [5], stochastic block model (SBM) graphs, [11], random geometric graphs and their structures [10], [12], [13], among many others.

The asymptotic behaviour of the entropy of a *soft* RGG, where the connection rule is probabilistic, is well understood [12]. On the other hand, the entropy of *hard* RGG, where the nodes are connected by an edge iff they are within a connection range r, remains understudied. However, in a prior work [10], it was hinted that the asymptotic behavior of the entropy of a one-dimensional RGG on the unit interval $[0,1]$ is $m\log_2 m$ bits, where m is the number of vertices of the graph. Later, [18] gave an argument for this entropic behavior in a special case of the one-dimensional model, i.e., when the connection range decays linearly to zero. In our current work, we will consider a hard RGG on two domains: d-dimensional unit cube $[0,1]^d$ and the d-dimensional unit torus $\mathbb{T}^d$ with a fixed connection range r. Along with deriving asymptotic upper bounds on the entropy, we will prove under some conditions a matching lower bound, giving our main result

$$H(G_m) \sim dm\log_2 m,$$

when $r \leq \frac{1}{4}$ for an RGG G_m with m vertices on $\mathbb{T}^d$. Finally, we will use this result to infer the structural entropy of an RGG.

II. Preliminaries

Let $G_m = (V,E)$ be a graph with the vertex set $V = [m] := \{1,2,\ldots,m\}$ and the edge set E containing unordered pairs of vertices. We restrict to undirected and simple graphs, where simple means that there are no self-loops or multiple edges between any pair of two vertices. For defining random geometric graphs, we will focus on two spatial domains: the d-dimensional unit cube $[0,1]^d$ and the d-dimensional unit torus $\mathbb{T}^d$. The unit torus is nothing but $[0,1]^d$ with the boundaries wrapped around. In the unit cube, the distance between two

The authors are with the Department of Engineering Science, University of Oxford, OX1 3PJ Oxford, U.K. (Email: praneeth.vippathalla@eng.ox.ac.uk; justin.coon@eng.ox.ac.uk;mihai.badiu@eng.ox.ac.uk). This research was funded in whole or in part by the U. S. Army Research Laboratory and the U. S. Army Research Office (W911NF-22-1-0070 and W911NF-24-2-0102).

points is given by the Euclidean distance, whereas in the unit torus, it is given by the toroidal distance [1], [2], which is defined as

$$d_t(x, y) := \min\{\|x + z - y\| : z \in \mathbb{Z}^d\} \quad \text{for} \quad x, y \in \mathbb{T}^d$$

with $\|\cdot\|$ being the Euclidean norm.

For a fixed connection range $r \geq 0$, we say that a graph G_m is a *hard geometric graph* on $[0,1]^d$ (resp. $\mathbb{T}^d$) if there exist m points $x_1, x_2, \ldots, x_m \in [0,1]^d$ (resp. $\mathbb{T}^d$) such that for any $u, v \in V$, $||x_u - x_v|| \leq r$ (resp. $d_t(x_u, y_v) \leq r$) if and only if $u \sim v$, i.e., u is adjacent to v in G_m. The toroidal metric overcomes some of the problems posed by the boundary effects of the unit cube, i.e., the local connectivity properties of the points near the boundary of $[0,1]^d$ is different from those of the points well within it. In the hard random geometric graph model, we draw m points $X_1, X_2, \ldots, X_m$ independently with uniform distribution on $[0,1]^d$ (resp. $\mathbb{T}^d$) and form the corresponding geometric graph G_m on the vertex set $[m]$. This graph generation process induces a probability distribution P_{G_m} on the set of all possible random geometric graphs $\mathcal{G}_m$.

It is useful to think of G_m in terms of its adjacency matrix representation $(E_{i,j})_{i,j\in[m]}$, where $E_{i,j} = 1$ if i and j are adjacent in G_m, and $E_{i,j} = 0$ otherwise. As the graphs are simple and undirected, this matrix is symmetric with zeros on the diagonal. It means that G_m can equivalently be represented by the collection $\{E_{i,j} : i < j\}$. Hence, the probability of a random geometric graph G_m is completely specified by the joint distribution of $\{E_{i,j} : i < j\}$, and vice versa. In an ER graph model, where we add an edge between two vertices with some probability independently of the rest of the edges, this collection is mutually independent. However, in the RGG model, this collection is correlated because of the underlying geometry. For example, if $u \sim v$ and $u \sim w$, then more often than not we see an edge between v and w.

A. Entropy of G_m

The entropy[1] of an RGG G_m is defined as

$$H(G_m) \triangleq -\sum\nolimits_{g\in\mathcal{G}_m} P_{G_m}(g) \log P_{G_m}(g).$$

As G_m can be uniquely identified with the collection of random variables $\{E_{i,j} : i < j\}$, we have $H(G_m) = H(\{E_{i,j} : i < j\})$. By using various properties of entropy, the following upper and lower bounds were noted in [12]:

$$H(G_m|X_1, \ldots, X_m) \leq H(G_m) \leq \sum\nolimits_{i<j} H(E_{i,j}),$$

where $X_1, \ldots, X_m \in [0,1]^d$ are the random variables corresponding to the locations of the vertices. Though these bounds are useful in determining the behavior of $H(G_m)$ for soft random geometric graphs, they fall short in the case of hard RGGs, which is due to the fact that the lower bound evaluates to zero:

$$0 \leq H(G_m) \leq \frac{m(m-1)}{2} h_2(p_r), \tag{1}$$

[1]Throughout, logarithms are to base 2 unless stated otherwise.

where $h_2(x) := -x\log_2 x - (1-x)\log_2(1-x)$ is the binary entropy function, and p_r is the probability that two random nodes are within a distance of r from each other. For a fixed r, the upper bound in (1) behaves like $m^2/2$ and it is tight when the $E_{i,j}$'s are mutually independent. However, since $E_{i,j}$'s are correlated in an RGG, it is reasonable to say that $m^2/2$ might not be the right behavior. In the subsequent sections, we will try to improve these bounds to characterize the behavior of $H(G_m)$.

III. Entropy of an RGG

A. d-dimensional unit cube $[0,1]^d$

In this section, we present the results on the entropy of an RGG on the d-dimension unit cube $[0,1]^d$. First, we will give an upper bound for all fixed connection ranges $0 < r < \sqrt{d}$. Observe that when r equals 0 or $\sqrt{d}$, the resulting graph is always an empty graph or a complete graph, respectively, whose entropy is trivially zero. Hence, we exclude those two extreme connection ranges.

Theorem 1. *An upper bound on the entropy of an RGG G_m on the d-dimensional unit cube $[0,1]^d$ is given by*

$$H(G_m) \leq \begin{cases} dm\log m + o(m\log m) & \text{if } 0 < r \leq \frac{\sqrt{d}}{2}, \\ [1-\beta(r)]dm\log m + o(m\log m) & \text{if } \frac{\sqrt{d}}{2} \leq r < \sqrt{d}. \end{cases}$$

where $\beta(r)$ is the volume of the ball $B((1/2, 1/2, \ldots, 1/2); r - \sqrt{d}/2) \cap [0,1]^d$.

Proof. See Section IV-A. □

Observe that the leading order term of the upper bound in the second case of Thm. 1 is less than $dm\log m$ and it decreases with r beyond $\sqrt{d}/2$. This refinement is possible due to the fact that when the connection range is larger than $\sqrt{d}/2$, all the random points that lie within $B((1/2, 1/2, \ldots, 1/2); r - \sqrt{d}/2) \cap [0,1]^d$ are connected by edges to the rest of the points, and such connections are significant in a random graph, resulting in entropy loss. The proof of Thm. 1 involves bounding the cardinality of the ensemble of RGGs for all $0 < r < \sqrt{d}$, following the argument of [4] that uses the Warren's theorem (Thm. 7), a result on the number of possible sign patterns of a collection of polynomials. Combining this cardinality bound with entropic inequalities, we obtain a refined bound for $\sqrt{d}/2 \leq r < \sqrt{d}$.

Due to the boundary effects of the domain, the analysis for a lower bound is technically challenging. However, we believe that the leading order terms in Thm. 1 give the correct asymptotic behavior of the entropy. To confirm this, we considered the case of $d = 1$ and proved a matching lower bound, which is presented in the following theorem.

Theorem 2. *The entropy of an RGG G_m on $[0,1]$ is lower bounded as follows:*

$$H(G_m) \geq \begin{cases} m\log m - o(m\log m) & \text{if } 0 < r \leq \frac{1}{2}, \\ 2(1-r)m\log m - o(m\log m) & \text{if } \frac{1}{2} \leq r < 1. \end{cases}$$

The proof of Thm. 2 relies on the order statistics of m uniformly and independently drawn points from $[0,1]$, and it can be found in the longer version [19]. By combining Thm. 1 and 2, and noting that $\beta(r)=2r-1$ for $\frac{1}{2}\le r<1$, we have the following result. The limit of the entropy normalized by $m\log m$ as a function of the connection range r is plotted in Fig. 1.

Theorem 3. *The asymptotic characterization of the entropy of an RGG G_m on $[0,1]$ is given by*

$$H(G_m)\sim\begin{cases} m\log m & \text{if } 0<r\le \frac{1}{2},\\ 2(1-r)m\log m & \text{if } \frac{1}{2}\le r<1.\end{cases} \tag{2}$$

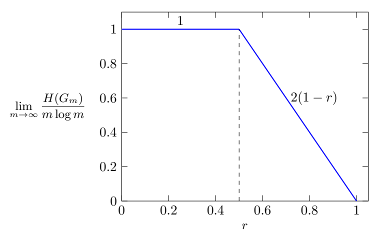

Fig. 1. Limit of $H(G_m)/m\log m$ of a one-dimensional RGG

B. d-dimensional unit torus $\mathbb{T}^d$

Some of the technical difficulties posed by boundary effects of the domain $[0,1]^d$ can be avoided by considering the unit torus $\mathbb{T}^d$, where the local connectivity is statistically the same at every point. Note that the connection range on $\mathbb{T}^d$ cannot be more than $\sqrt{d}/2$. The following theorem presents an entropy upper bound in the case of a unit torus.

Theorem 4. *The entropy of an RGG G_m on $\mathbb{T}^d$ is upper bounded as follows:*

$$H(G_m)\le dm\log m+o(m\log m)\ \text{ for }\ 0<r<\frac{\sqrt{d}}{2}.$$

Even on a unit torus, the leading order term of the upper bound is still $dm\log m$ for all connection ranges $0<r<\sqrt{d}/2$. The basic proof idea of Thm. 4 is similar to that of Thm. 1— bounding the cardinality of the ensemble of RGGs using the Warren's theorem (Thm. 7). However, we need to choose a different collection of polynomials based on the toroidal metric. The proof is given in detail in the longer version [19].

A tight lower bound is presented in the next theorem, which holds for the connection range values $r\le\frac{1}{4}$. This condition is a mere technical choice to simplify the analysis. However, we believe that the same result is valid over the whole interval $0<r<\sqrt{d}/2$.

Theorem 5. *The entropy of an RGG G_m on $\mathbb{T}^d$ is lower bounded as follows:*

$$H(G_m)\ge dm\log m-o(m\log m)\ \text{ for }\ r\le\frac{1}{4}.$$

Proof. See Section IV-B. □

The proof relies on a result on the Crofton cell properties [20] for a particular Boolean model whose focus is on the intersection of randomly placed spheres. For an alternative proof, see [19].

Now our main result can be stated by combining Thm. 4 and 5. It says that the entropy scales like $dm\log m$, which is linear in the dimension d and independent of the connection range r.

Theorem 6. *The asymptotic characterization of the entropy of an RGG G_m on $\mathbb{T}^d$ is given by*

$$H(G_m)\sim dm\log m\ \text{ for }\ r\le\frac{1}{4}. \tag{3}$$

C. Structural Entropy of an RGG

The entropy characterization result (Thm. 6) can be used to deduce how small the structural entropy of an RGG can be. A structure [5] of a graph is its unlabelled version. With applications to compression, the structural entropy of various random graphs has been studied [5], [10], [21]. In the work [10] on one-dimensional RGGs, it was established that the structural entropy $H(S_m)$, where S_m is the structure of an RGG G_m, satisfies the relation $\log e(1-r)m\le H(S_m)\le 2(1-r)m$ for $0<r<1$. By combining this with the result of Thm. 3, we can conclude that for a one-dimensional RGG, the randomness in the graph is dominated by the randomness in the labels. To see this, consider the expansion of the entropy of a general RGG in terms of the structural entropy:

$$H(G_m)=H(G_m,S_m)=H(S_m)+H(G_m|S_m) \tag{4}$$

which follows from the fact that S_m is completely determined by G_m. Notice that when conditioned on a structure S_m, the randomness in the graph G_m comes only from its labeling. Hence, $H(G_m|S_m)$ represent the average conditional randomness in the labeling. It follows from (4) that the entropy of this labeling must be $\Theta(m\log m)$ for a one-dimensional RGG as $H(G_m)=\Theta(m\log m)$ and $H(S_m)=\Theta(m)$, dominating the entropy of the graph. However, this is not the case when $d>1$. Here, the structural entropy dominates as shown in the following result.

Corollary 1. *The structural entropy of an RGG G_m on $\mathbb{T}^d$ is given by $H(S_m)=\Omega((d-1)m\log m)$ for $r\le\frac{1}{4}$.*

Proof. Since $H(G_m|S_m)$ is upper bounded by the log of the cardinality of all possible labelings, which is $m!$, we have from (4) that $H(S_m)=H(G_m)-H(G_m|S_m)\ge H(G_m)-\log(m!)$. The result immediately follows from Stirling's approximation and Thm. 6. □

IV. Proofs

A. Proof of Theorem 1: Upper bound in the case of $[0,1]^d$

The first bound in the theorem follows from the work of McDiarmid and Müller [4] on disk-intersection graphs, where Warren's theorem [22] (stated below for convenience) was used to find an upper bound on the cardinality of the set of possible graphs. For the sake of completeness, we will present that argument applied to random geometric graphs.

Recall that $\mathcal{G}_m$ denotes the set of all hard RGGs with a fixed connection range r. In a hard RGG, two vertices $u, v \in V$ are connected by an edge iff $\|x_u - x_v\| \le r$. Now we consider the degree-2 polynomials of the form $\|X_u - X_v\|^2 - r^2$ for each pair $u, v \in V$. For a fixed pair of vertices u and v, if the sign of the polynomial $\|X_u - X_v\|^2 - r^2$ evaluated at $X_u = x_u$ and $X_v = x_v$ is negative, then there is an edge in the graph G_m between u and v; otherwise, there is no edge.

By arranging the signs of the evaluations of all these polynomials (in a fixed order) at a realization of the node locations $X_1 = x_1, \ldots X_m = x_m$, we get a sign pattern. Let $\mathcal{S}_m \subseteq \{+,-\}^{\binom{m}{2}}$ be the collection of all such sign patterns. In other words, $\mathcal{S}_m$ is the set containing all the sign patterns corresponding to the points $(x_1, \ldots, x_m) \in [0,1]^{dm}$. As each sign pattern is uniquely identified with an RGG, and vice versa, the map from $\mathcal{S}_m$ to $\mathcal{G}_m$ is bijective. Therefore, $|\mathcal{G}_m| = |\mathcal{S}_m|$. We can now bound $|\mathcal{S}_m|$ using the following theorem.

Theorem 7 (Warren [4], [22])**.** *For polynomials $Q_1, Q_2, \ldots, Q_u$ with at most degree k in real variables $z_1, \ldots, z_t$, the number of the sign patterns $(\mathrm{sign}(Q_1), \ldots, \mathrm{sign}(Q_u)) \in \{-,+\}^u$ evaluated over $\mathbb{R}^t \setminus \cup_{i=1}^{u}\{(z_1, \ldots, z_t) : Q_i(z_1, \ldots, z_t) = 0\}$ is bounded above by $\left(\frac{4eku}{t}\right)^t$.*

By applying Theorem 7 with $u = \binom{m}{2}$, $t = dm$ and $k = 2$ for the above degree-2 polynomials, we get

$$|\mathcal{G}_m| = |\mathcal{S}_m| \le \left(\frac{4e \cdot 2 \cdot \binom{m}{2}}{md}\right)^{dm} \le m^{dm} \cdot \left(\frac{4e}{d}\right)^{dm},$$

which is indeed independent of the connection range r. This implies that

$$H(G_m) \le \log|\mathcal{G}_m| \le dm \log m + dm \log C, \tag{5}$$

where $C = \left(\frac{4e}{d}\right)^d$. This bound is true for the whole interval $0 < r < \sqrt{d}$, proving the first part of the theorem. However, we can improve this bound if $\sqrt{d}/2 \le r < \sqrt{d}$, which is presented below.

Let us denote the ball $B((1/2, 1/2, \ldots, 1/2); r - \sqrt{d}/2) \cap [0,1]^d$ around the center of $[0,1]^d$ of radius $r - \sqrt{d}/2$ by $B(r - \sqrt{d}/2)$ and its volume $\mathrm{Vol}(B(r - \sqrt{d}/2))$ by $\beta(r)$. Define $S \triangleq \{i \in [m] : X_i \in B(r - \sqrt{d}/2)\}$. Note that for any $i \in S$, $\|X_v - X_i\| \le r$ for all $v \in [m] \setminus \{i\}$ because every point $y \in [0,1]^d$ is within a distance r from every $x \in B\left(r - \sqrt{d}/2\right)$. Therefore, any edge random variable corresponding to a vertex $i \in S$ is 1. We can think[2] of $G[S]$, the graph with only vertices in S as the *core* of the whole graph $G[V] = G_m$. The graph $G[S]$ is complete with significant number of vertices in it, and it is connected to all the other vertices within $G[V]$. If we know the set S, the only uncertainty in the graph $G[V]$ is due to $G[V \setminus S]$. Using this fact we can bound the entropy:

$$\begin{aligned} H(G_m) = H(G[V]) &\le H(G[V], S) = H(S) + H(G[V] \mid S) \\ &= H(S) + H(G[V \setminus S]|S) \quad (6)\\ &\le \log 2^m + H(G[V \setminus S]|S) \quad (7)\\ &= m + \mathbb{E}_S\left[H(G[V \setminus S]|S)\right] \\ &\le m + \mathbb{E}_S\left[d|V \setminus S| \log|V \setminus S| + |V \setminus S| \log C\right] \quad (8)\\ &\le m + d \cdot \mathbb{E}_S\left[|V \setminus S|\right] \log m + \mathbb{E}_S\left[|V \setminus S|\right] \log C \quad (9)\\ &= m + (1 - \beta(r))m \log m + (1 - \beta(r))m \log C, \quad (10) \end{aligned}$$

[2]We use the notation $G[U]$ to denote the vertex-induced subgraph of $G_m := G[V]$, where $V = [m]$, containing only the vertices in $U \subseteq V$.

where (6) relies on the fact that all the edge random variables in the graph $G[V]$ except for those in $G[V \setminus S]$ are 1; as the cardinality of all possible sets S is 2^m, (7) follows; in (8) and (9), we use the upper bound on the entropy of a random geometric graph in terms of the cardinality of the number graphs (5) with $|V \setminus S|$ vertices, i.e.,

$$\begin{aligned} H(G[V \setminus S]|S = s) &\le \log|\mathcal{G}_{|V \setminus s|}| \\ &\le d|V \setminus s| \log|V \setminus s| + |V \setminus s| \log C \\ &\le d|V \setminus s| \log m + |V \setminus s| \log C; \end{aligned}$$

(10) is due to the fact that $\mathbb{E}_S\left[|V \setminus S|\right] = m - \mathbb{E}_S\left[|S|\right] = m - m\beta(r) = m(1 - \beta(r))$. This shows that

$$H(G_m) \le [1 - \beta(r)]m \log m + [1 + (1 - \beta(r)) \log C]m$$

for $\sqrt{d}/2 \le r < \sqrt{d}$, completing the proof of the theorem.

B. Proof of Theorem 5: Lower Bound in the case of $\mathbb{T}^d$

We begin with the series of equalities involving the differential entropies:

$$\begin{aligned} H(G_m) &= I(G_m; X_1, \ldots, X_m) \quad (11)\\ &= h(X_1, \ldots, X_m) - h(X_1, \ldots, X_m \mid G_m) \\ &= -h(X_1, \ldots, X_m \mid G_m) \quad (12)\\ &= -\sum_{k=1}^{m} h(X_k \mid X_1, \ldots, X_{k-1}, G_m), \end{aligned}$$

where (11) results from the fact that $H(G_m \mid X_1, \ldots, X_m) = 0$, and (12) follows because the X_i's are uniformly distributed and $\mathrm{Vol}([0,1]^d) = 1$, therefore $h(X_1, \ldots, X_m) = 0$. We now obtain a bound by ignoring all connections in G_m that do not relate directly to node k:

$$\begin{aligned} H(G_m) &\ge -\sum_{k=1}^{m} h(X_k \mid X_1, \ldots, X_{k-1}, E_{1k}, \ldots, E_{k-1,k}) \\ &= -\sum_{k=1}^{m} h(X_k \mid X_1, \ldots, X_{k-1}, L_k) \quad (13) \end{aligned}$$

where the first inequality is a consequence of the property that conditioning reduces entropy, and $L_k \triangleq \{i \in [k-1] : E_{ik} = 1\}$, which is a random subset of $\{1, \ldots, k-1\}$. Furthermore, for fixed $X^{k-1} = x^{k-1}$ and $L_k = \ell_k$, the differential entropy is just the logarithm of the volume of the region $\mathcal{R}(X^{k-1}, L_k)$ in $\mathbb{T}^d$ carved out by the intersection of the balls centered at x_i, $i \in L_k$, excluding the portion covered by the balls centered at x_i, $i \in L_k^c$. From this reasoning, we have

$$\begin{aligned} h(X_k \mid X_1, \ldots, X_{k-1}, L_k) &= \mathbb{E}\left[\log\left[\mathrm{Vol}\left(\mathcal{R}(X^{k-1}, L_k)\right)\right]\right] \\ &\le \log \mathbb{E}\left[\mathrm{Vol}\left(\mathcal{R}(X^{k-1}, L_k)\right)\right], \end{aligned}$$

where the second line follows from Jensen's inequality, and

$$\mathcal{R}(X^{k-1}, L_k) = \left[\cap_{i \in L_k} \mathcal{B}_{X_i}(r)\right] \cap \left[\cap_{i \in L_k^c} \mathcal{B}^c_{X_i}(r)\right].$$

Let $\mathcal{I}(X^{k-1},L_k)=\cap_{i\in L_k}\mathcal{B}_{X_i}(r)$ and take $\mathcal{I}(X^{k-1},\emptyset)=\mathbb{T}^d$ by definition. Note that $\mathrm{Vol}\left(\mathcal{R}(X^{k-1},L_k)\right)\leq\mathrm{Vol}\left(\mathcal{I}(X^{k-1},L_k)\right)$ for $L_k\subseteq[k-1]$, which gives

$$h(X_k\mid X_1,\ldots,X_{k-1},L_k)\leq\log\mathbb{E}\left[\mathrm{Vol}\left(\mathcal{I}(X^{k-1},L_k)\right)\right]. \tag{14}$$

Let L denote the cardinality of L_k, i.e., the number of ones in $E_{1k},\ldots,E_{k-1,k}$. It follows from stationarity and symmetry, together with the fact that the volume measure is translation-ally invariant in a torus, that $E_{1k},\ldots,E_{k-1,k}$ are independent and identically distributed Bernoulli random variables with parameter p, where $p=\Pr(X\sim Y)$ for any two points X and Y uniformly distributed on $\mathbb{T}^d$. As a consequence, L has a binomial distribution with parameters $k-1$ and p.

Define $\mathcal{I}(X^\ell)\triangleq\bigcap_{i=1}^{\ell}\mathcal{B}_{X_i}(r)$ for $\ell\geq 1$ and $\mathcal{I}(X^0)\triangleq\mathbb{T}^d$. The variable pairs (X_i,E_{ik}), $1\leq i\leq k-1$, are exchangeable. Hence, the conditional expectation of the volume of $\mathcal{I}(X^{k-1},L_k)$ with respect to L_k depends only on the cardinality of L_k, which is L:

$$\mathbb{E}\left[\mathrm{Vol}\left(\mathcal{I}(X^{k-1},L_k)\right)\mid L_k\right]=\mathbb{E}\left[\mathrm{Vol}\left(\mathcal{I}(X^{L})\right)\mid L\right].$$

Now, by taking the expectation on both sides, we obtain

$$\begin{aligned}\mathbb{E}\left[\mathrm{Vol}\left(\mathcal{I}(X^{k-1},L_k)\right)\right]&=\mathbb{E}\left[\mathbb{E}\left[\mathrm{Vol}\left(\mathcal{I}(X^{k-1},L_k)\right)\mid L_k\right]\right]\\&=\mathbb{E}\left[\mathbb{E}\left[\mathrm{Vol}\left(\mathcal{I}(X^{L})\right)\mid L\right]\right]\\&=\mathbb{E}[v(L)],\end{aligned} \tag{15}$$

where $v(\ell)=\mathbb{E}\left[\mathrm{Vol}\left(\mathcal{I}(X^L)\right)\mid L=\ell\right]$ is the average volume of the intersection region formed by ℓ balls of radius r conditioned on the event that X_k lies in that region. By definition, $v(\ell)=1$ for $\ell=0$.

Conditioned on the event $L=\ell$, where $1\leq\ell\leq k-1$ and the location X_k, the random variables $X^L=(X_1,\ldots,X_L)$ are independently and uniformly distributed on the ball $\mathcal{B}_{X_k}(r)$ centered at X_k. Because of the translation invariance of the volume measure on torus, the average volume $v(\ell)$ is nothing but the average volume of $\mathcal{I}(X^L)$ when the random variables in X^L are independently and uniformly distributed on a ball of radius r around a fixed point of $\mathbb{T}^d$:

$$v(\ell)=\mathbb{E}\left[\mathbb{E}\left[\mathrm{Vol}\left(\mathcal{I}(X^L)\right)\mid L=\ell,X_k\right]\right]=\mathbb{E}[\mathrm{Vol}(\mathcal{I}(\tilde{X}^\ell))],$$

where $\tilde{X}^\ell=(\tilde{X}_1,\ldots,\tilde{X}_\ell)$ are independently and uniformly distributed points in $\mathcal{B}_c(r)$ with $c=(1/2,\ldots,1/2)\in\mathbb{T}^d$.

The asymptotic behavior of the average volume $v(\ell)$ follows from a result of Richey and Sarkar [20, Proposition 1.1] that says that for a Poisson point process $\tilde{X}^\lambda$ with intensity parameter λ in a ball of radius r (say, centered at c) in $\mathbb{R}^d$,

$$\mathbb{E}\left[\mathrm{Vol}\left(\mathcal{I}^\lambda\right)\right]\sim C_{d,r}\cdot\lambda^{-d},$$

where $C_{d,r}$ is a constant that depends only on the dimension d and radius r, and $\mathcal{I}^\lambda=\mathcal{B}_c(r)\cap[\cap_{X\in\tilde{X}^\lambda}\mathcal{B}_X(r)]$. The authors remark that the same is the scaling limit even in the fixed count version with ℓ points, i.e., $\mathbb{E}[\mathrm{Vol}(\mathcal{I}(\tilde{X}^\ell))]\sim C\cdot\ell^{-d}$ for a constant C. In fact, it can be verified by applying the standard de-poissonization argument for the expectation of monotonically decreasing functions[3] [23, Theorem 5.10] that $\mathbb{E}[\mathrm{Vol}(\mathcal{I}(\tilde{X}^\ell))]\leq 2\cdot\mathbb{E}[\mathrm{Vol}(\mathcal{I}^\ell)]$, where we choose the intensity parameter to be ℓ. Therefore, we have $\mathbb{E}[\mathrm{Vol}(\mathcal{I}(\tilde{X}^\ell))]=O(\ell^{-d})$. When $r\leq\frac{1}{4}$, the volume of the intersection $\mathcal{I}(\tilde{X}^\ell)$ in a torus is the same[4] as that in $\mathbb{R}^d$. Therefore, we have

$$v(\ell)=O(\ell^{-d}). \tag{16}$$

Lemma 1. *Let $v(\ell)$ denote the volume of an intersection of ℓ balls independently distributed on $\mathcal{B}_c(r)$ in a torus, where $c=(1/2,\ldots,1/2)$ and $r\leq 1/4$, and let $L\sim\mathsf{Bin}(k-1,p)$. Then, $\mathbb{E}[v(L)]=O(k^{-d})$.*

Proof. The result follows by considering the high probability event $\{L\geq(k-1)\frac{p}{2}\}$ and expanding the expectation:

$$\begin{aligned}\mathbb{E}[v(L)]&=\mathbb{E}[v(L)1\{L<(k-1)\tfrac{p}{2}\}]\\&\qquad+\mathbb{E}[v(L)1\{L\geq(k-1)\tfrac{p}{2}\}]\\&\leq\mathbb{P}\left(L\leq(k-1)\tfrac{p}{2}\right)+C.\left((k-1)\tfrac{p}{2}\right)^{-d} &(17)\\&\leq e^{-(k-1)\frac{p^2}{2}}+C.\left((k-1)\tfrac{p}{2}\right)^{-d} &(18)\\&\leq\tilde{C}.k^{-d}, &(19)\end{aligned}$$

where the first term of (17) is due to the fact that $v(L)\leq 1$ and the second term follows from combining (16), which is $v(l)\leq C\cdot l^{-d}$ for some constant $C>0$, with $l\geq(k-1)\frac{p}{2}$; (18) applies the Hoeffding's inequality; and in (19), $\tilde{C}$ is a large enough constant. □

By using Lemma 1 in (15) and combining it with (14) and (13), we obtain

$$H(G_m)\geq-\sum_{k=1}^{m}\log\left(\tilde{C}.k^{-d}\right)=dm\log m-o(m\log m),$$

completing the proof of the theorem.

V. Conclusion

In this work, we studied the asymptotic behavior of the entropy of a hard random geometric graph on the d-dimensional unit cube and unit torus. For all fixed connection range values, we derived upper bounds on $H(G_m)$. In a few cases, we proved that the entropy asymptotically behaves like $dm\log m$, which was then used to deduce the structural entropy of an RGG. It should be noted that the proof technique of our complete result on $\mathbb{T}^d$ can be extended to the other connection range values and the unit cube. The reason for restricting to a special case is to make the technical analysis simpler. We strongly believe and conjecture that the leading order terms in the upper bounds of Thm. 1 and Thm. 4 give the right behaviour for the entropy asymptotically. Currently we considered only a connection range value r that is fixed; however, it is an interesting open question how the entropy scales when r depends on the number of vertices m. Also, from a practical standpoint, it is of interest to design a compression scheme that attains a compression length close to the entropy $dm\log m$, which we leave for future work.

[3] It is evident that $\mathbb{E}[\mathrm{Vol}(\mathcal{I}(\tilde{X}^\ell))]$ is a non-increasing function of ℓ.

[4] When $r\leq 1/4$ and all the balls contain $c=(1/2,\ldots,1/2)$, they are strictly contained in $[0,1]^d$ without wrapped around the boundaries. So, the region covered by the intersection is exactly the same as that in $\mathbb{R}^d$.

References

[1] M. Penrose, *Random Geometric Graphs*. Oxford: Oxford University Press, 2003.

[2] M. Haenggi, *Stochastic Geometry for Wireless Networks*. Cambridge University Press, 2012.

[3] V. Dani, J. Díaz, T. P. Hayes, and C. Moore, "Reconstruction of random geometric graphs: Breaking the $\Omega(r)$ distortion barrier," *Eur. J. Comb.*, vol. 121, p. 103842, 2024.

[4] C. McDiarmid and T. Müller, "The number of disk graphs," *Eur. J. Comb.*, vol. 35, pp. 413–431, 2014.

[5] Y. Choi and W. Szpankowski, "Compression of graphical structures: Fundamental limits, algorithms, and experiments," *IEEE Trans. Inf. Theory*, vol. 58, no. 2, pp. 620–638, Feb. 2012.

[6] P. Delgosha and V. Anantharam, "Universal lossless compression of graphical data," *IEEE Trans. Inf. Theory*, vol. 66, no. 11, pp. 6962–6976, Nov. 2020.

[7] H. Nikpey, S. Sarkar, and S. S. Bidokhti, "Compression with unlabeled graph side information," in *Proc. IEEE Int. Symp. Inf. Theory (ISIT)*, Taipei, Taiwan, Jun. 2023, pp. 713–718.

[8] M. W. Wafula, P. K. Vippathalla, J. Coon, and M.-A. Badiu, "Rate-distortion function of the stochastic block model," in *Proc. 57th Asilomar Conf. Signals Syst. Comput.*, Pacific Grove, CA, USA, Oct. 2023, pp. 699–703.

[9] R. Bustin and O. Shayevitz, "On lossy compression of directed graphs," *IEEE Trans. Inf. Theory*, vol. 68, no. 4, pp. 2101–2122, Apr. 2022.

[10] M.-A. Badiu and J. P. Coon, "Structural complexity of one-dimensional random geometric graphs," *IEEE Trans. Inf. Theory*, vol. 69, no. 2, pp. 794–812, Sep. 2023.

[11] E. Abbe, "Graph compression: The effect of clusters," in *Proc. 54th Annu. Allerton Conf. Commun. Contr. Comput.*, Monticello, IL, USA, Sep. 2016, pp. 1–8.

[12] J. P. Coon, C. P. Dettmann, and O. Georgiou, "Entropy of spatial network ensembles," *Phys. Rev. E*, vol. 97, p. 042319, Apr. 2018.

[13] M.-A. Badiu and J. P. Coon, "On the distribution of random geometric graphs," in *Proc. IEEE Int. Symp. Inf. Theory (ISIT)*, Vail, CO, USA, Jun. 2018, pp. 2137–2141.

[14] P. K. Vippathalla, M.-A. Badiu, and J. P. Coon, "Graph compression with side information at the decoder," in *Proc. IEEE Int. Symp. Inf. Theory (ISIT)*, Ann Arbor, MI, USA, Jun. 2025, pp. 1–6.

[15] P. K. Vippathalla, M. W. Wafula, M.-A. Badiu, and J. P. Coon, "On the lossy compression of spatial networks," in *Proc. IEEE Int. Symp. Inf. Theory (ISIT)*, Athens, Greece, Jul. 2024, pp. 416–421.

[16] T. M. Cover and J. A. Thomas, *Elements of Information Theory*, 2nd ed. USA: Wiley-Interscience, 2006.

[17] N. Alon and A. Orlitsky, "A lower bound on the expected length of one-to-one codes," *IEEE Trans. Inf. Theory*, vol. 40, no. 5, pp. 1670–1672, Sep. 1994.

[18] J. Paton, H. Hartle, H. Stepanyants, P. van der Hoorn, and D. Krioukov, "Entropy of labeled versus unlabeled networks," *Phys. Rev. E*, vol. 106, p. 054308, Nov. 2022.

[19] P. K. Vippathalla, J. P. Coon, and M.-A. Badiu, "On the entropy of a random geometric graph," 2026. [Online]. Available: https://arxiv.org/abs/2601.10778

[20] J. Richey and A. Sarkar, "Intersections of random sets," *J. Appl. Probab.*, vol. 59, no. 1, p. 131–151, Mar. 2022.

[21] T. Łuczak, A. Magner, and W. Szpankowski, "Asymmetry and structural information in preferential attachment graphs," *Random Struct. Alg.*, vol. 55, no. 3, pp. 696–718, Mar. 2019.

[22] H. E. Warren, "Lower bounds for approximation by nonlinear manifolds," *Trans. Am. Math. Soc.*, vol. 133, no. 1, pp. 167–178, Aug. 1968.

[23] M. Mitzenmacher and E. Upfal, *Probability and Computing: Randomized Algorithms and Probabilistic Analysis*. Cambridge University Press, 2005.