Compression Limits of Soft Random Geometric Graphs

Oliver Baker*, Praneeth Kumar Vippathalla†, Justin P. Coon†, Carl P. Dettmann*, and Mihai-Alin Badiu†

* School of Mathematics, University of Bristol

† Department of Engineering Science, University of Oxford

Abstract

Random geometric graphs (RGGs) are commonly used to model spatial graph data where connectivity between the nodes is governed by the distance between them in an embedded space. One class of RGGs that is of interest is soft RGGs, where the connectivity rule is given by a probabilistic function of the distance. With a motivation to compress spatial graph data, we consider soft random geometric graphs and study their information-theoretic compression limits. We address the lossless and lossy compression of soft RGGs in two regimes, namely dense and sparse configurations. We provide an asymptotic characterization of the entropy and the information-distortion function, and determine the rate of convergence as the number of nodes goes to infinity. Our main result is that if the connection function p is Hölder continuous, then the rate of convergence is given by $\mathcal{O}\left(\frac{\log n}{n}\right)$ in dense regimes and $\mathcal{O}\left(\frac{\log n}{ns_n \log(1/s_n)}\right)$ in sparse regimes, where n is the number of vertices and s_n is the sparsity parameter.

I. Introduction

Many real-world networks or graphs are spatial in nature. Examples include brain, wireless, and social networks. For instance, two devices in a wireless network share a communication link

The work of Praneeth Kumar Vippathalla, Justin P. Coon, and Mihai-Alin Badiu (Email: praneeth.vippathalla@eng.ox.ac.uk; justin.coon@eng.ox.ac.uk;mihai.badiu@eng.ox.ac.uk) 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). Oliver Baker is supported by the EPSRC Centre for Doctoral Training in Computational Statistics and Data Science (COMPASS). This was presented in part at the 2024 IEEE International Symposium on Information Theory (ISIT 2024), Athens, Greece.

if they are within a certain distance of each other. Similarly, two neurons are connected through a synapse if they are sufficiently close to each other. In addition, graph-structured data, in which entities are connected by edges based on shared properties, are also commonly in use today. Such data sets can contain billions of nodes and trillions of edges; hence their compression, and graph compression in general, is a particularly interesting and useful task.

In this work, we focus on the fundamental compression limits of soft random geometric graphs (SRGG), a model that is commonly used to represent spatial networks. In a SRGG, n points (nodes) are scattered uniformly and independently in a bounded domain in $\mathbb{R}^d$, and two points are connected by an edge with a probability that is a function of the distance between the points. We address both lossless and lossy compression of such graphs.

The main objectives of graph compression, in general, are to store a graph efficiently by utilizing less memory or to speed up an algorithm by running it on the compressed version. Recently, there have been various works addressing different aspects of graph compression. In [1], [2], the compression of marked graphs was studied using the framework of Bordenave-Caputo entropy, providing universal compression algorithms for both dense and sparse regimes. There have also been a number of works on different graph models focusing on lossless graph compression, notable works include [3]–[8]. Entropy results have been characterized for Erdős-Rényi (ER) graphs [3], stochastic block graphs [9], small-world graphs [8] and classical random geometric graphs (with a deterministic connection rule) [10]–[12]. Limiting results and bounds on SRGG entropy were first treated in [13], [14].

Due to storage limitations, it is also of interest to compress a large graph in a lossy way. For example, upon reconstruction, we can allow a particular graph feature to be distorted within prescribed limits. A natural framework in which to study this problem is classical rate-distortion theory. As graph sources are arbitrary and the distortion functions could be complicated, the study of lossy compression is a challenging task. For instance, [15] considered the lossy compression problem of an ER random graph with a distortion measure on the degree of the nodes. In that work, lower and upper bounds on the asymptotic rate of compression were derived. The rate-distortion function for a source generating stochastic block graphs was studied in [16] for cases where side information in the form of community labels was present and not. As it is difficult to extend techniques used for non-spatial studies to SRGGs, we adopt an approach that uses the information-distortion function. This is a natural analogue of entropy in the context of lossy compression. Preliminary results on SRGGs that exploit the information-distortion function framework can be found in Vippathalla et al's earlier work [17].

This paper significantly advances the work reported in [17]. We first study the entropy of a soft random geometric graph by characterizing its asymptotic behavior for both dense and sparse regimes. The proof uses tools from the theory of graphons along with the discretization of the underlying domain. The main result is that the normalized entropy approaches the limiting quantity as $\mathcal{O}\left(\frac{\log n}{n}\right)$ and $\mathcal{O}\left(\frac{\log n}{ns_n \log(1/s_n)}\right)$, in dense and sparse regimes respectively, if the connection function p is Hölder continuous. Then, analogously, for the lossy compression, we study the information-distortion function with Hamming distortion measure. For this, we first identify the limiting quantity for the normalized information distortion function and then show that the rate of convergence is still $\mathcal{O}\left(\frac{\log n}{n}\right)$ and $\mathcal{O}\left(\frac{\log n}{ns_n \log(1/s_n)}\right)$, in dense and sparse regimes respectively, if the connection function p is Hölder continuous.

The rest of the paper is organized as follows. Section II provides the preliminaries by formally defining SRGGs and graphons. In Section III, we present the results on the entropy of an SRGG in the dense and sparse regimes. Section IV contains the results on the characterization of the information-distortion function of an SRGG in these two regimes. The proofs of all results are given in Section V and VI, and Section VII concludes the paper.

II. Preliminaries

A. Random Geometric Graphs

Let $G_n = (V, E)$ be a graph with the vertex set $V = \{1, 2, \ldots, n\}$ and the edge set E which is a collection of unordered pairs of vertices. Throughout, we use the notation $[n]$ to denote $\{1, 2, \ldots, n\}$ for any positive integer n. For simplicity, we restrict ourselves to undirected and simple graphs, where simple means that there are no self loops or multiple edges. With $\mathcal{G}_n$ denoting the set of all simple and undirected labelled graphs on n vertices, a *random graph* model is specified by a probability distribution P_{G_n} on $\mathcal{G}_n$.

Random geometric graphs are commonly used to describe random spatial networks where there is an underlying hidden spatial embedding of the nodes. In a *random geometric graph model*, vertices are in a d-dimensional Euclidean space and the edge connectivity between a pair of nodes is governed by the distance between them. Formally, the vertices are spatially embedded in a bounded set $\mathcal{K} \subset \mathbb{R}^d$ with unit volume with respect to the Lebesgue measure. We note that our results can be extended to $\mathcal{K}$ with any strictly positive volume by rescaling the arguments of the connection function. Let $K \triangleq \sup_{x,y\in\mathcal{K}} \|x - y\|_2$ be the diameter of the domain $\mathcal{K}$. For a given connectivity function $p_n : [0, K] \to [0, 1]$ and a fixed positive integer n, a *soft random geometric graph* (SRGG) is obtained as follows. We first draw n points $\mathbf{X} = (X_1, X_2, \ldots, X_n)$ uniformly at random from $\mathcal{K}$ and independently of each other. The distance between two nodes i and j is given by $R_{ij} \triangleq \|X_i - X_j\|_2$. We then form a graph G_n on the vertex set $V = [n]$ by drawing an edge between a pair of vertices $i < j$ with probability $p_n(R_{ij})$. This induces a probability distribution P_{G_n} on $\mathcal{G}_n$.

It is often convenient to view a graph in terms of its adjacency matrix representation. The $(i, j)^{\text{th}}$ entry E_{ij} of an adjacency matrix indicates the presence or absence of an edge between vertices i and j in the graph, i.e., $E_{ij} = 1$ if i and j are adjacent, and $E_{ij} = 0$ otherwise. As the graphs are simple and undirected, this matrix is symmetric with zeros on the diagonal. It means that G_n can equivalently be represented by the upper triangular entries E_{ij}, $i < j$. Hence, the probability of a random geometric graph G_n is completely specified by the joint distribution of the collection $\{E_{ij} : i < j\}$, and vice versa.

B. Dense and Sparse Regimes

We consider two regimes of interest, namely the dense and sparse regimes, based on the average density of edges in a random graph. We consider a connectivity function p_n of the form $s_n \cdot p$, where $\{s_n\}$ is positive, non-increasing sequence of real numbers with $s_n \leq 1$ for every n, and p is a function of distance r and it is independent of n. The average density of edges is given by

$$\mathbb{E}\left[\frac{1}{\binom{n}{2}} \sum_{i<j} \mathbb{1}_{\{E_{ij}=1\}}\right] = \mathbb{E}\left[p_n(R)\right] = s_n \mathbb{E}\left[p(R)\right],$$

where $R \triangleq \|X - Y\|_2$ with X and Y being independent and uniformly distributed random variables over $\mathcal{K}$. In the *dense regime*, s_n is chosen to be 1 to maintain a constant density[1] of edges. As density must go to zero in the *sparse regime*, the sequence of s_n's is chosen to satisfy $s_n \to 0$. The interesting special case is when $ns_n \to \infty$, which ensures that each node has infinitely many neighbors in the limit $n \to \infty$.

C. Graphons

The theory of graphons is useful in the study of the limit of theorems of the random geometric graphs. For a probability space $(\mathcal{X}, \mathcal{F}, \nu)$, a graphon is a symmetric, measurable function

[1] A constant density of edges is obtained for any $s_n \in \Omega(1)$ such that $s_n \leq 1$, and we choose $s_n = 1$ for convenience.

$W : \mathcal{X}^2 \to [0, 1]$. Graphons are a relatively recent development, useful for studying graph limits and large graphs, which is motivated by the following fact, which we state informally here. We refer the interested reader to [18] for more detail. It can be shown that there exists a suitable metric, called the *cut metric* which may be defined for both finite graphs and graphons, such that the limit of any sequence of graphs that converges in the cut metric converges to a graphon. The uniqueness of the limit holds up to 'equivalence', where we say two graphons are equivalent if they have cut distance zero. Thus, technically the cut metric is a proper metric only on the quotient space of graphons. We will not provide further detail here as we do not explicitly use this property in this paper, but this justifies the use of graphons to study the limiting behaviour of large graphs.

We will instead use the fact that for a given graphon W on $[0, 1]$ (we will generalise to arbitrary $\mathcal{X}$ in the next paragraph), it is simple to construct a sequence of graphs that converges to W almost surely in cut distance. Define the W *random graph* as follows. Let $X_1, ..., X_n$ be independent and uniformly distributed points on $[0, 1]$, then G_n is the graph with n vertices, and edge set formed by including each edge (i, j) independently with probability $W(X_i, X_j)$. Then it can be shown that the (random) sequence $\{G_n\}$ converges almost surely in cut distance to W. We denote the ensemble of W random graphs as $G(n, W)$.

It is clear from this construction that a W random graph is a generalisation of the SRGG on $[0, 1]$, where the SRGG is a W random graph where for $x, y \in [0, 1]$, $W(x, y) = p(\|x - y\|)$. This can be extended to SRGGs on general $\mathcal{X} \subset \mathbb{R}^d$ using what is known as the graphon representation theorem [19], which states that every graphon W_1 on a probability space $(\mathcal{X}, \mathcal{F}_1, \nu)$ is equivalent to a graphon W_2 on the probability space $([0, 1], \mathcal{F}_2, \lambda)$, where λ represents the Lebesgue measure (i.e. the limit of a W random graph on $[0, 1]$ with uniformly distributed nodes). Thus, graphons may be used to study SRGGs on general domains with general node distributions, and this gives a useful way to prove results about general domains using results on $[0, 1]$. However, the representation of a general graphon on $[0, 1]$ under the Lebesgue measure does not necessarily retain its continuity properties, and so we will need to deal with graphons on general spaces separately [20]. There have been several results on the limiting behaviour of W random graph entropy using graphon entropy, which we will state in the next section. We will then prove quantitative results related to these limits for finite n, for which we will focus on Hölder continuous graphons, defined as follows.

Definition 1 (Hölder continuous graphon)**.** We say a graphon $W : \mathcal{K}^2 \to [0,1]$ is α-Hölder continuous, or Hölder continuous with exponent α if there exists an L (which we call the Hölder constant) such that for every pair of points $(x,y),(x',y') \in \mathcal{K}^2$

$$|W(x,y) - W(x',y')| \leq L\|(x,y) - (x',y')\|_2^{\alpha}. \tag{1}$$

III. Entropy of an SRGG

Entropy is a fundamental quantity in the lossless compression of a random source. For a random graph G_n with a probability distribution P_{G_n} on $\mathcal{G}_n$, the *entropy* is given by

$$H(G_n) \triangleq \sum_{g_n \in \mathcal{G}_n} P_{G_n}(g_n) \log \frac{1}{P_{G_n}(g_n)}. \tag{2}$$

where the logarithm is to base 2. It is well-known that the minimum expected length of a prefix-free code l^* is within 1 bit from the entropy [21], i.e.,

$$H(G_n) \leq l^* < H(G_n) + 1, \tag{3}$$

and we also know that the minimum expected length of a one-to-one code ℓ^* is also close to the entropy [22], [9, Lem. 1]:

$$H(G_n) - \log(H(G_n)+1) - \log e \leq \ell^* < H(G_n). \tag{4}$$

A characterization of the entropy will provide a limit on the minimum average number of bits needed to represent a random graph. Unlike an independent and identically distributed (i.i.d.) source, the underlying edge random variables of a graph source, in general, are correlated. So, it is important to understand the scaling behavior of entropy as the number of vertices grow. The following subsections address the asymptotic behavior of the entropy of soft RGGs.

A. Limit Theorems

We may now state the results that describe the entropy of the SRGG in the $n \to \infty$ limit. The limit in the dense regime comes as a Corollary of Theorem D.5 in [19].

Theorem 1 ([19])**.** *The entropy of the SRGG in the dense regime satisfies*

$$\lim_{n\to\infty} \frac{H(G_n)}{\binom{n}{2}} = \mathbb{E}[h_2(p(R))]. \tag{5}$$

As a high-level summary, the proof uses the graphon representation theorem to work with a graphon on $[0,1]$ instead of $\mathcal{K}$. The space $[0,1]$ is discretised into $m \in \mathbb{N}$ intervals, and then $H(G_n)$ is bounded by conditioning on this discretisation. The entropy of the discretisation is asymptotically sub-leading order, and so after dividing by $\binom{n}{2}$, taking the $n \to \infty$ limit, followed by taking the $m \to \infty$ limit to increase the resolution of the discretisation until we reach conditioning on the exact node locations, the limit is given by the entropy-per-edge of G_n conditioned on the node locations, or equivalently the inter-node distances.

The limit in the sparse regime comes as a consequence of Proposition 2 in [2]. Using largely the same proof strategy as above, they study specifically 'normalised graphons' which are graphons $\widetilde{W} : \widetilde{\mathcal{K}}^2 \to \mathbb{R}_{\geq 0}$ that satisfy $\int\int_{\widetilde{\mathcal{K}}^2} \widetilde{W}(x,y)dxdy = 1$, where we have used the temporary notation $\widetilde{\mathcal{K}}$ to refer to a subset of $\mathbb{R}^d$ of any positive volume. Their proof can be easily adapted to our setting, which yields the following result. We also note that in [2], the logarithm is taken to base e, so our result has an extra factor of $1/\ln 2$ to account for this change of base.

Theorem 2 ([2])**.** *Let s_n be a positive non-increasing sequence of real numbers converging to 0, bounded above by 1. Additionally assume $ns_n \to \infty$ as $n \to \infty$. Then the entropy of the SRGG with connection function $p_n = s_n \cdot p$ in the sparse regime satisfies*

$$\lim_{n\to\infty} \frac{H(G_n) - \binom{n}{2} s_n \log(1/s_n)\mathbb{E}[p(R)]}{\binom{n}{2} s_n} = \frac{1}{\ln 2}\mathbb{E}[p(R)] - \mathbb{E}[p(R)\log p(R)]. \tag{6}$$

B. Bounds for Finite Graphs

After stating the limit theorems in the previous section, it is natural to ask how fast the convergence occurs. That is, for a finite n, how can we bound the difference between the normalized entropy and the limit? For the dense regime, we have the following result.

Theorem 3. *Let G_n be an SRGG in the dense regime with connection function p, which has Hölder exponent α. Then for every $n \geq 2$, define*

$$\Delta_n^{\text{dense}} := \frac{1}{\binom{n}{2}} \left(H(G_n) - \binom{n}{2} \mathbb{E}[h_2(p(R))] \right) \tag{7}$$

Then we have

$$0 \leq \Delta_n^{\text{dense}} \leq \frac{dn \log m}{\binom{n}{2}} + a_1 \frac{\log m}{m^\alpha} + a_2 \frac{1}{m^\alpha} + a_3 \frac{1}{m} + a_4 \frac{1}{m^2} \tag{8}$$

where $m = m(n) = Cn^\gamma$ for some $\gamma > 0$ where $C \geq L^{1/\alpha}\ell\sqrt{2d}$ is a constant, and a_i, $i \in \{1,2,3,4\}$ are constants dependent on d, α and $\mathcal{K}$ defined by $a_1 = \frac{\alpha L}{\ln 2}(\ell\sqrt{2d})^\alpha$, $a_2 =$

$L(\ell\sqrt{2d})^\alpha\left(1-\log\left(L(\ell\sqrt{2d})^\alpha\right)\right)$, $a_3 = 2|\delta\mathcal{K}|\ell$ and $a_4 = |\delta\mathcal{K}|^2\ell^2$, where L is the Hölder constant of p, ℓ is the side length of the largest d-cube covering $\mathcal{K}$, and $|\delta\mathcal{K}|$ is the $(d-1)$-dimensional surface area of $\mathcal{K}$.

Proof. See Section V. □

The following theorem gives an analogous result in the sparse regime.

Theorem 4. *Let G_n be an SRGG in the sparse regime with connection function $p_n = s_n \cdot p$, where p has Hölder exponent α, and $ns_n \to \infty$. Then for every $n \geq 2$, define*

$$\Delta_n^{\mathit{sparse}} := \frac{H(G_n) - \binom{n}{2}s_n\log(1/s_n)\mathbb{E}[p(R)] - \binom{n}{2}\frac{1}{\ln 2}s_n\mathbb{E}[p(R)] + \binom{n}{2}s_n\mathbb{E}[p(R)\log p(R)]}{\binom{n}{2}s_n\log(1/s_n)} \quad (9)$$

Then we have

$$-\frac{s_n\mathbb{E}[p(R)^2] + s_n^2\mathbb{E}[p(R)^3]}{2\ln(2)\log(1/s_n)} \leq \Delta_n^{\mathit{sparse}} \leq \frac{nd\log m}{\binom{n}{2}s_n\log(1/s_n)} + a_1\frac{\alpha\log m + \log(1/s_n)}{m^\alpha\log(1/s_n)}$$
$$+a_2\frac{1}{m^\alpha\log(1/s_n)} + a_3\frac{\frac{1}{\ln 2}+\log(1/s_n)}{m\log(1/s_n)} + a_4\frac{\frac{1}{\ln 2}+\log(1/s_n)}{m^2\log(1/s_n)} \quad (10)$$

where a_i, $i \in \{1,...,4\}$ are defined in Theorem 3, and $m = m(n) = Cn^\gamma$ for some $\gamma > 0$ where $C \geq L^{1/\alpha}\ell\sqrt{2d}$ is a constant.

Proof. See Section V. □

Remark 1. *We point out that, in both Theorems, taking $\gamma = \frac{1}{\alpha}$ gives the optimal asymptotic behaviour of the bound, that is $\Delta_n^{dense} = \mathcal{O}\left(\frac{\log n}{n}\right)$, and $\Delta_n^{sparse} = \mathcal{O}\left(\frac{\log n}{ns_n\log(1/s_n)}\right)$, which goes to 0 if $ns_n \to \infty$.*

There has recently been work on quantifying the *structural entropy* of random graphs, which is the entropy of the unlabelled graph S_n (e.g. [11], [23]). Since, for an n node graph, the uniformly random labels have entropy $\log n! \sim n\log n$, and the entropy of the SRGG with $s_n \ll 1$ is $H(G_n) \sim \binom{n}{2}s_n\log(1/s_n)\mathbb{E}[p(R)]$, we have that the structural entropy $H(S_n)$ is asymptotically equivalent to the topological entropy $H(G_n)$ if $\binom{n}{2}s_n\log(1/s_n) \gg n\log n$, which is equivalent to $ns_n \to \infty$. This follows from the relation $H(G_n) - n\log n \leq H(S_n) \leq H(G_n)$. Therefore, in the sparse regime the topological entropy is equivalent to the structural entropy.

IV. Information-Distortion Function of an SRGG

Similar to entropy in lossless compression, the information-distortion function plays a fundamental role in the theory of lossy compression. Given a distortion function $d_n : \mathcal{G}_n \times \mathcal{G}_n \to [0,\infty)$

and a distortion level D_n, which specifies the cost of representing a graph with another graph, the *information-distortion function* of a random graph G_n is defined as

$$I_{G_n}(D_n) \triangleq \min_{P_{\hat{G}_n|G_n}:\mathbb{E}[d_n(G_n,\hat{G}_n)]\leq D_n} I(G_n;\hat{G}_n). \tag{11}$$

It is well-known from the classical rate-distortion theory [21, Ch. 13] that for an i.i.d. source $\{X^n\}_{n\geq 1}$ and a per-letter distortion function d, the minimum rate of compression $R(D)$ at an (average) distortion level D is given by

$$R(D) = \min_{P_{\hat{X}|X}:\mathbb{E}[d(X,\hat{X})]\leq D} I(X;\hat{X}). \tag{12}$$

The right-hand side of (12) is the first order information-distortion function of the i.i.d. source $\{X^n\}_{n\geq 1}$. On the other hand, the distribution P_{G_n} of a random graph need not have the desired independence structure. For such arbitrary sources [24, Ch. 5], the information-distortion function can be used to study the lossy compression.

Consider the following variable-length lossy compression of random graphs. An encoder $\phi_n : \mathcal{G}_n \to \{0,1\}^*$ maps a graph to a codeword, which is a binary string of finite (but arbitrary) length. Here $\{0,1\}^*$ denotes the set of all binary strings of finite length. We also assume that the encoder satisfies the prefix condition. A decoder $\psi_n : \{0,1\}^* \to \mathcal{G}_n$ maps a codeword to a graph. The encoder and decoder pair is designed such that they satisfy the distortion constraint $\mathbb{E}[d_n(G_n, \psi_n(\phi_n(G_n)))] \leq D_n$ for a distortion level D_n. Let $l_n^*(D_n)$ denote the least average codeword length $\mathbb{E}[l(\phi_n(G_n))]$. We can easily show that

$$I_{G_n}(D_n) \leq l_n^*(D_n). \tag{13}$$

This follows from the observation that for a prefix code [21, Thm. 5.4.1]

$$l_n^*(D_n) \geq H(\phi_n(G_n)), \tag{14}$$

which is further lower bounded by $H(\psi_n(\phi_n(G_n))) \geq I(\psi_n(\phi_n(G_n)); G_n)$, where the reconstructed graph satisfies the average distortion condition, yielding (13). Though an upper bound on the minimum optimal length $l_n^*(D_n)$ in terms of $I_{G_n}(D_n)$ and D_n is not known for a fixed n, asymptotically optimal results for the rate are known.

Define the asymptotic rate of the compression scheme for the distortion level $D_n = \binom{n}{2}D$ to be

$$R(D) = \limsup_{n\to\infty} \frac{1}{\binom{n}{2}} l_n^*\big(\tbinom{n}{2}D\big). \tag{15}$$

A result on arbitrary sources [24, Thm. 5.7.1 and Remark 5.7.3] applied to graphs yield

$$R(D) = \limsup_{n\to\infty} \frac{1}{\binom{n}{2}} I_{G_n}\big(\tbinom{n}{2}D\big), \tag{16}$$

under some reasonable assumptions on the distortion function and by allowing a stochastic encoder instead of deterministic encoder. It was also noted in [24, Cor. 5.7.1] that the stochastic encoder can be implemented without loss of optimality by randomizing between two deterministic encoders.

The information-distortion function is also useful in the regimes where the order of the distortion levels $\{D_n\}_{n\geq 1}$ is not the same as that of the source entropy. For example, in the result (16), a notion of rate is defined with respect to the scaling $\binom{n}{2}$ with $D_n = \binom{n}{2}D$. But we can consider other regimes that are of interest, say $D_n = Dn\log n$ and $H(G_n) = \Theta(\binom{n}{2})$, where the right scaling would be $\binom{n}{2}$. For the cases where we may not be able to define a useful notion of rate, we can still work with $I_{G_n}(D_n)$ owing to the inequality (13).

The effect of distortion on the information-distortion function can be made explicit through the Lagrangian dual formulation

$$I_{G_n}(D_n) = \sup_{\lambda\geq 0} \inf_{P_{\hat{G}_n|G_n}} I(G_n;\hat{G}_n) + \lambda\left[\mathbb{E}[d_n(G_n,\hat{G}_n)] - D_n\right] \tag{17}$$

For instance, in the case of $H(G_n) = \Theta(n^2)$ and $D_n = Dn\log n$, the distortion constraint contributes only to the $n\log n$ term of the expression $H(G_n) - H(G_n|\hat{G}_n) + \lambda\left[\mathbb{E}[d_n(G_n,\hat{G}_n)] - D_n\right]$ with the corresponding coefficient quantifying the effect of distortion on compression. With the above motivation, we study the information-distortion function of random geometric graphs.

A. Hamming Distortion Measure

While there are various distortion measures that one can work with depending on the context, in this preliminary work, we consider the Hamming distortion measure d_n between two graphs G_n and $\hat{G}_n$ that counts the number of pairs of vertices whose connectivity in G_n differs from that in $\hat{G}_n$, i.e.,

$$d_n(G_n,\hat{G}_n) = \sum\nolimits_{i<j} \mathbb{1}(E_{ij} \neq \hat{E}_{i,j}). \tag{18}$$

The Hamming distortion penalizes a reconstructed graph $\hat{G}_n$ that has many pairs of vertices with mismatched connectivity relative to the original graph G_n.

B. Main Results

In this section, we will present our result on the characterization of the information-distortion function (11) of a random geometric graph with the Hamming distortion measure. For establishing the result, we first study the information-distortion function with conditioning on the node locations $\mathbf{X}$, which is defined as

$$I_{G_n|\mathbf{X}}(D_n) \triangleq \inf_{P_{\hat{G}_n|G_n,\mathbf{X}}:\mathbb{E}[d_n(G_n,\hat{G}_n)]\leq D_n} I(G_n;\hat{G}_n|\mathbf{X}). \tag{19}$$

This quantity arises when $\mathbf{X}$ is available as side information to both the encoder and decoder. It is worth noting that the information-distortion functions with and without conditioning satisfy the following inequality.

Lemma 1 ([25])**.** *For a pair $(\mathbf{X}, G_n)$ and $D_n \geq 0$,*

$$I_{G_n|\mathbf{X}}(D_n) \leq I_{G_n}(D_n) \leq I_{G_n|\mathbf{X}}(D_n) + I(G_n;\mathbf{X}). \tag{20}$$

Before we present our result, let us define a few quantities, which will simplify the notation later on. Recall $R \triangleq \|X - Y\|_2$, where X and Y are independent and uniformly distributed random variables over $\mathcal{K}$. For $D_n \geq 0$ and a connectivity function $p_n : [0, K] \to [0, 1]$, define a function $q_n^* : [0, K] \to [0, 1]$ as follows:

$$q_n^*(r) = \min\{p_n(r), 1 - p_n(r), \mu_n\} \tag{21}$$

with μ_n chosen to satisfy the condition

$$\binom{n}{2}\mathbb{E}[q_n^*(R)] = D_n \text{ if } D_n \leq \binom{n}{2}\mathbb{E}[\min\{p_n(R), 1 - p_n(R)\}],$$

and we set $\mu_n = \frac{1}{2}$ otherwise.

Theorem 5. *Let G_n be a random geometric graph with $\mathbf{X}$ being the node locations and $p_n : [0, K] \to [0, 1]$ being the connectivity function. The information-distortion function of G_n conditioned on node locations $\mathbf{X}$ at the Hamming distortion D_n is given by*

$$I_{G_n|\mathbf{X}}(D_n) = \binom{n}{2}\Big[\mathbb{E}\big[h_2\big(p_n(R)\big)\big] - \mathbb{E}\big[h_2\big(q_n^*(R)\big)\big]\Big], \tag{22}$$

where $q_n^* : [0, K] \to [0, 1]$ is as defined in (21)*.*

Proof. See Section VI. □

Lemma 2. *1) Let G_n be an RGG in the dense regime with a connection function p that is Hölder continuous with the exponent α. Then for every $n \geq 2$,*

$$\frac{1}{\binom{n}{2}} I(G_n; \mathbf{X}) = \Delta_n^{\textit{dense}}, \tag{23}$$

where $\Delta_n^{\textit{dense}}$ is as defined in (7) and it is upper bounded as in (8).

2) Let G_n be an RGG in the sparse regime with the connection function $p_n = s_n \cdot p$, where p is Hölder continuous with exponent α and $ns_n \to \infty$. Then for every $n \geq 2$,

$$\frac{1}{\binom{n}{2} s_n \log(1/s_n)} I(G_n; \mathbf{X}) = \Delta_n^{\textit{sparse}} + O(s_n), \tag{24}$$

where $\Delta_n^{\textit{sparse}}$ is as defined in (9) and it is upper bounded as in (10).

Proof. Both the statements can be argued by expressing the mutual information in terms of entropy functions. By applying the chain rule of entropy and noting that E_{ij} is independent of the rest of the random variables given X_i and X_j for all $i < j$, we obtain

$$H(G_n|\mathbf{X}) = \sum_{i<j} H(E_{ij}|X_i, X_j) = \binom{n}{2} \mathbb{E}\big[h_2(p_n(R))\big], \tag{25}$$

where the last equality uses the fact that for $i < j$, conditioned on X_i and X_j, E_{ij} is a Bernoulli random variable with parameter $p_n(\|X_i - X_j\|_2) = p_n(R_{ij})$, so $H(E_{ij}|X_i, X_j) = \mathbb{E}\left[h_2(p_n(R_{ij}))\right]$. As the distribution of R_{ij} is the same for all vertex pairs, we replace it with the random variable $R = \|X - Y\|_2$, which is the distance between two random points X and Y uniformly distributed in $\mathcal{K}$. Therefore, we have

$$I(G_n; \mathbf{X}) = H(G_n) - H(G_n|\mathbf{X}) = H(G_n) - \binom{n}{2} \mathbb{E}\big[h_2(p_n(R))\big] \tag{26}$$

In the dense regime, p_n is not a function of n. So, the first statement immediately follows from the definition of Δ_n^{dense} given in (7) and by applying Thm. 3.

In the sparse regime, p_n is of the form $s_n \cdot p$ where s_n is a sequence of real numbers converging to 0. Here,

$$I(G_n; \mathbf{X}) = H(G_n) - \binom{n}{2} \mathbb{E}\big[h_2(p_n(R))\big] \tag{27}$$

$$= H(G_n) - \binom{n}{2} \mathbb{E}\big[s_n \log(1/s_n) p(R) + s_n p(R)/\ln 2 - s_n p(R) \log p(R)\big] + O\left(\binom{n}{2} s_n^2\right), \tag{28}$$

where the last equality follows from the Taylor series expansion (50). By using the definition of Δ_n^{sparse} given in (9) and by applying Thm. 4, we obtain the second statement. □

By combining Lemma 1, Theorem 5 and Lemma 2, we immediately obtain a characterization of $I_{G_n}(D_n)$.

Theorem 6. *Let G_n be an RGG in the dense regime with a connection function p that is Hölder continuous with the exponent α. Then for every $n \geq 2$,*

$$0 \leq \frac{1}{\binom{n}{2}} \Big[I_{G_n}(D_n) - \Big[\mathbb{E}\big[h_2\big(p(R)\big)\big] - \mathbb{E}\big[h_2\big(q_n^*(R)\big)\big] \Big] \Big] \leq \Delta_n^{\textit{dense}}, \tag{29}$$

where $q_n^* : [0, K] \to [0, 1]$ is defined as in (21) with $p_n = p$, and $\Delta_n^{\textit{dense}}$ is given in Theorem 3.

Theorem 7. *Let G_n be an RGG in the sparse regime with the connection function $p_n = s_n \cdot p$, where p is Hölder continuous with exponent α and $ns_n \to \infty$. Then for every $n \geq 2$,*

$$0 \leq \frac{1}{\binom{n}{2} s_n \log(1/s_n)} \Big[I_{G_n}(D_n) - \Big[\mathbb{E}\big[h_2\big(p_n(R)\big)\big] - \mathbb{E}\big[h_2\big(q_n^*(R)\big)\big] \Big] \Big] \leq \Delta_n^{\textit{sparse}} + O(s_n), \tag{30}$$

where $q_n^* : [0, K] \to [0, 1]$ is defined as in (21), and $\Delta_n^{\textit{sparse}}$ is given in Theorem 4.

Remark 2. *As with the entropy, in both Theorems 6 and 7, taking $\gamma = \frac{1}{\alpha}$ gives the optimal asymptotic behaviour of the bound, that is $\Delta_n^{dense} = \mathcal{O}\left(\frac{\log n}{n}\right)$, and $\Delta_n^{sparse} = \mathcal{O}\left(\frac{\log n}{ns_n \log(1/s_n)}\right)$, which goes to 0 if $ns_n \to \infty$.*

V. Proofs of Theorem 3 and 4

Lemma 3. *For $x, y \in [0, 1]$,*

$$|h_2(x) - h_2(y)| \leq h_2(|x - y|). \tag{31}$$

Proof. Without loss of generality, we can assume $0 < x < y < 1$ due to symmetry. First, rewrite $h(y) - h(x)$ in terms of the derivative of h_2.

$$h_2(y) - h_2(x) = \int_0^x h_2'(t)dt + \int_x^{y-x} h_2'(t)dt + \int_{y-x}^y h_2'(t)dt - \int_0^x h_2'(t)dt \tag{32}$$

$$= \int_0^{y-x} h_2'(t)dt - \int_0^x h_2'(t)dt + \int_{y-x}^y h_2'(t)dt. \tag{33}$$

With a change of variables, we can rewrite this as

$$h_2(y) - h_2(x) = h_2(y - x) - \int_0^x [h_2'(t) - h_2'(t + y - x)]dt \tag{34}$$

$$\leq h_2(y - x) \tag{35}$$

since h_2 is concave, and so its derivative is decreasing. Now, pick $a, b \in [0, 1]$, and set $x = \min\{a, b, 1 - a, 1 - b\}$ and $y = x + |a - b|$. Note that this implies that $x < y$ and $x < 1 - y$.

If $x = a$, $y = a + b - a = b$, and $x < 1 - y$ by the construction of x. If $x = 1 - a$, $y = 1 - a + a - b = 1 - b$, so again $x < 1 - y$. The same works by symmetry if $x = b$ or $1 - b$. Then because $x < y$ and $x < 1 - y$ we have $h(y) \geq h(x)$,

$$|h_2(a) - h_2(b)| = h_2(y) - h_2(x) \leq h_2(y - x) \tag{36}$$

by above, and then because $y > x$, finally

$$|h_2(a) - h_2(b)| \leq h_2(|b - a|) \tag{37}$$

completing the proof. □

Proof of Theorem 3. The proof follows the proof of Theorem 1 from [19], and we bound the approximation error in each step. First, since conditioning reduces entropy, we condition on the set of inter-node distances $\{R_{ij}\}_{i<j}$ to give

$$H(G_n) \geq H(G_n|\{R_{ij}\}_{i<j}) = \sum_{i<j} H(E_{ij}|R_{ij}) = \binom{n}{2}\mathbb{E}[h_2(p(R))]. \tag{38}$$

For this proof we will need to use the graphon formulation of the SRGG. Let $W : \mathcal{K}^2 \to [0, 1]$ be defined for $x, y \in \mathcal{K}$ as

$$W(x, y) = p(\|x - y\|). \tag{39}$$

Then the SRGG on n-nodes with connection function p is the same as the W random graph on n nodes, which we denote by $G(n, W)$. We now construct an upper bound using the W random graph formalism. Cover $\mathcal{K}$ with m^d disjoint, equally sized d-cubes of side length ℓ/m. We have two cases, first, when $\mathcal{K}$ can be perfectly covered by d-cubes, and second where it cannot. We will deal with the first case now, and the second case will follow by a minor modification afterwards.

For each node i, denote by $M_i \subset \mathcal{K}$ the box that contains the position of X_i. Let $M = (M_1, ..., M_n)$ be the vector collecting the boxes of each node. Then, we can write

$$H(G_n) = H(G(n, W)) \leq H(G(n, W)|M) + H(M). \tag{40}$$

Now since the nodes of the SRGG are uniformly distributed on $\mathcal{K}$, and M_i is a deterministic function of X_i, we have $H(M) \leq n \log m^d = dn \log m$. Now, by the independence bound we can write

$$H(G(n, W)|M) \leq \sum_{i<j} H(E_{ij}|M_i, M_j). \tag{41}$$

To express the right-hand-side in integral form, we first must introduce the block-constant approximation of W. For $x \in M_i \cap \mathcal{K}$, $y \in M_j \cap \mathcal{K}$, $\overline{W}_m(x,y)$ is given by

$$\overline{W}_m(x,y) = \mathbb{E}[W(X,Y)|X \in M_i \cap \mathcal{K}, Y \in M_j \cap \mathcal{K}]. \tag{42}$$

That is, $\overline{W}_m(x,y)$ is the average value of W taken over all pairs (X,Y) where X is in the same box as x and Y is in the same box as y. Then

$$\mathbb{P}(E_{ij} = 1|M_i = m_i, M_j = m_j) = \overline{W}_m(x_i, x_j), \tag{43}$$

where $x_i \in m_i$, and $x_j \in m_j$. So it is clear by the definition of the binary entropy function that

$$H(E_{ij}|M_i, M_j) = \int\int_{\mathcal{K}^2} h_2(\overline{W}_m(x,y))dxdy, \tag{44}$$

and therefore we have

$$\binom{n}{2}\int\int_{\mathcal{K}^2} h_2(W(x,y))dxdy \leq H(G_n) \leq \binom{n}{2}\int\int_{\mathcal{K}^2} h_2(\overline{W}_m(x,y))dxdy + nd\log m \tag{45}$$

which implies

$$\begin{aligned} 0 &\leq H(G_n) - \binom{n}{2}\int\int_{\mathcal{K}^2} h_2(W(x,y))dxdy \\ &\leq \binom{n}{2}\int\int_{\mathcal{K}^2} h_2(\overline{W}_m(x,y))dxdy - \binom{n}{2}\int\int_{\mathcal{K}^2} h_2(W(x,y))dxdy + nd\log m. \end{aligned} \tag{46}$$

So to get the result, we need to bound

$$\begin{aligned} &\left|\int\int_{\mathcal{K}^2} h_2(\overline{W}_m(x,y))dxdy - \int\int_{\mathcal{K}^2} h_2(W(x,y))dxdy\right| \\ &\qquad\leq \int\int_{\mathcal{K}^2} |h_2(\overline{W}_m(x,y)) - h_2(W(x,y))|dxdy \\ &\qquad\leq \int\int_{\mathcal{K}^2} h_2\left(\left|\overline{W}_m(x,y) - W(x,y)\right|\right)dxdy \end{aligned} \tag{47}$$

where we have used first the linearity of integration and second Lemma 3. Since $\overline{W}_m$ is the mean of W on each box $M_i \times M_j = m_i \times m_j$, we have that the difference between $\overline{W}_m$ and W evaluated at a pair of points (x,y) is upper bounded by the maximum difference of W evaluated on the box, that is,

$$|\overline{W}_m(x,y) - W(x,y)| \leq \sup_{(x,y),(x',y') \in m_i \times m_j} |W(x,y) - W(x',y')| \tag{48}$$

$$\leq \sup_{(x,y),(x',y') \in m_i \times m_j} L\|(x,y) - (x',y')\|_2^\alpha = L\left(\frac{\ell\sqrt{2d}}{m}\right)^\alpha, \tag{49}$$

where we have used the Hölder continuity of W, and that the points furthest away from each other on $m_i \times m_j$ lie on opposite corners. Next, note that the Taylor expansion of the binary entropy function gives

$$\begin{aligned} h_2(p) &= -p\log p + \frac{p}{\ln 2} - \frac{1}{\ln 2}\sum_{k=2}^{\infty}\left(\frac{1}{k-1} - \frac{1}{k}\right)p^k \\ &\le -p\log p + \frac{p}{\ln 2}, \end{aligned} \tag{50}$$

which is a non-decreasing function of p on $[0,1]$. Then using (49), and provided that $m = m(n)$ so that $m(1) \ge L^{1/\alpha}\ell\sqrt{2d}$,

$$\begin{aligned} h_2(|\overline{W}_m(x,y) - W(x,y)|) &\le -L\left(\frac{\ell\sqrt{2d}}{m}\right)^{\alpha}\log L\left(\frac{\ell\sqrt{2d}}{m}\right)^{\alpha} + \frac{L}{\ln 2}\left(\frac{\ell\sqrt{2d}}{m}\right)^{\alpha} \\ &\le \alpha L\left(\frac{\ell\sqrt{2d}}{m}\right)^{\alpha}\log m + L\left(\frac{\ell\sqrt{2d}}{m}\right)^{\alpha}\left(\frac{1}{\ln 2} - \log\left(L(\ell\sqrt{2d})^{\alpha}\right)\right) \\ &= a_1\frac{\log m}{m^{\alpha}} + a_2\frac{1}{m^{\alpha}}. \end{aligned} \tag{51}$$

So,

$$\int\int_{\mathcal{K}^2} h_2\left(\left|\overline{W}_m(x,y) - W(x,y)\right|\right)dxdy \le \left(a_1\frac{\log m}{m^{\alpha}} + a_2\frac{1}{m^{\alpha}}\right)\int\int_{\mathcal{K}^2} dxdy.$$

Combining everything, we have

$$0 \le \frac{1}{\binom{n}{2}}H(G_n) - \mathbb{E}[h_2(p(R))] \le \frac{dn\log m}{\binom{n}{2}} + a_1\frac{\log m}{m^{\alpha}} + a_2\frac{1}{m^{\alpha}} \tag{52}$$

We now must deal with the case where $\mathcal{K}$ cannot be perfectly covered. In this case, there will be $\mathcal{O}(m^{d-1})$ boxes that overlap the boundary of $\mathcal{K}$. For this case, write

$$\mathcal{K} \subset \bar{\mathcal{K}}_m \cup \delta\mathcal{K}_m \tag{53}$$

where $\bar{\mathcal{K}}_m$ is the set of boxes M_i that are completely contained in $\mathcal{K}$, and $\delta\mathcal{K}_m$ is the set of boxes that overlap a boundary. Then, if we take the convention that $W(x,y) = \overline{W}_m(x,y) = 0$ when $(x,y) \notin \mathcal{K}$,

$$\begin{aligned} \int\int_{\mathcal{K}^2} h_2\left(\left|\overline{W}_m(x,y) - W(x,y)\right|\right)dxdy = &\sum_{\substack{i,j \\ (M_i,M_j)\in\bar{\mathcal{K}}_m^2}} \int\int_{M_i\times M_j} h_2\left(\left|\overline{W}_m(x,y) - W(x,y)\right|\right)dxdy \\ &+ \sum_{\substack{i,j \\ (M_i,M_j)\in\delta\mathcal{K}_m^2}} \int\int_{M_i\times M_j} h_2\left(\left|\overline{W}_m(x,y) - W(x,y)\right|\right)dxdy \end{aligned}$$

$$+2 \sum_{\substack{i,j \\ (M_i \times M_j) \in \bar{\mathcal{K}}_m \times \delta\mathcal{K}_m}} \int\int_{M_i \times M_j} h_2\left(\left|\overline{W}_m(x,y) - W(x,y)\right|\right) dxdy \tag{54}$$

The first sum can be bounded the same as the case where $\mathcal{K}$ can be perfectly tiled. For the second two cases, we cannot guarantee Hölder continuity, and so the best bound we have is a constant bound of 1. Let $|\delta\mathcal{K}|$ be the surface area of $\mathcal{K}$. Then, since we can lower bound the volume of the boxes M_i by $(\ell/m)^{d-1}$, we require at most $|\delta\mathcal{K}|(m/\ell)^{d-1}$ boxes to cover the boundary of $\mathcal{K}$. Let $|\delta\mathcal{K}_m|$ denote the number of elements in $\delta\mathcal{K}_m$. Then we can bound

$$\sum_{\substack{i,j \\ (M_i,M_j) \in \delta\mathcal{K}_m^2}} \int\int_{M_i \times M_j} h_2\left(\left|\overline{W}_m(x,y) - W(x,y)\right|\right) dxdy \leq \sum_{\substack{i,j \\ (M_i,M_j) \in \delta\mathcal{K}_m^2}} \int\int_{M_i \times M_j} dxdy$$
$$\leq |\delta\mathcal{K}_m|^2 \left(\frac{\ell}{m}\right)^{2d} \leq |\delta\mathcal{K}| \left(\frac{\ell}{m}\right)^2 \tag{55}$$

and similarly

$$2 \sum_{\substack{i,j \\ (M_i,M_j) \in \delta\bar{\mathcal{K}}_m \times \mathcal{K}_m}} \int\int_{M_i \times M_j} h_2\left(\left|\overline{W}_m(x,y) - W(x,y)\right|\right) dxdy \leq \frac{2|\delta\mathcal{K}|\ell}{m}. \tag{56}$$

Then, define $a_3 = |\delta\mathcal{K}|^2\ell^2$, and $a_4 = 2|\delta\mathcal{K}|\ell$. This finally results in the bound

$$0 \leq \frac{1}{\binom{n}{2}} H(G_n) - \mathbb{E}[h_2(p(R))] \leq \frac{dn \log m}{\binom{n}{2}} + a_1 \frac{\log m}{m^\alpha} + a_2 \frac{1}{m^\alpha} + a_3 \frac{1}{m} + a_4 \frac{1}{m^2} \tag{57}$$

Finally, note that we can take m to be a function of n to tune the upper bound. For $0 < \alpha \leq 1$, lowest order is achieved when $m(n) \propto n^{1/\alpha}$. The fact that Δ_n^{dense} is non-negative follows immediately from the fact that $\binom{n}{2}\Delta_n^{\text{dense}} = H(G_n) - H(G_n|\mathbf{X})$ is the mutual information $I(G_n; \mathbf{X})$ which is clearly non-negative. □

Proof of Theorem 4. As in the proof of Theorem 3, by conditioning on the node locations, and the box containing each node we arrive at the following bound:

$$H(G_n) \leq H(G_n|M) + nd \log m. \tag{58}$$

We also have

$$H(G_n|M) \leq \binom{n}{2} \mathbb{E}[h_2(\mathbb{E}[s_n W(X,Y)|M])] \tag{59}$$
$$= \mathbb{E}[h_2(s_n \overline{W}_m(X,Y))]$$
$$= \mathbb{E}\Big[s_n \log(1/s_n) \overline{W}_m(X,Y) + \frac{s_n}{\ln 2} \overline{W}_m(X,Y) - s_n \overline{W}_m(X,Y) \log \overline{W}_m(X,Y)$$

$$-\frac{1}{\ln 2}\sum_{k=2}^{\infty}\left(\frac{1}{k-1}-\frac{1}{k}\right)(s_n\overline{W}_m(X,Y))^k\Bigg],$$

where W and $\overline{W}_m$ are defined as in (39) and (42). Recall that $p_n(\cdot) = s_n p(\cdot)$, and so it is clear that $p_n(\|x-y\|) = s_n W(x,y)$, and similarly the 'averaged' graphon is given by $s_n\overline{W}_m(x,y)$ in the sparse case. From (58), it follows that

$$\begin{aligned}\Delta_n^{\text{sparse}} &= \frac{H(G_n) - \binom{n}{2}s_n\log(1/s_n)\mathbb{E}[p(R)] - \binom{n}{2}\frac{s_n}{\ln 2}\mathbb{E}[p(R)] + \binom{n}{2}s_n\mathbb{E}[p(R)\log p(R)]}{\binom{n}{2}s_n\log(1/s_n)}\\ &\leq \frac{H(G_n|M) + nd\log m - \binom{n}{2}s_n\log(1/s_n)\mathbb{E}[p(R)] - \binom{n}{2}\frac{s_n}{\ln 2}\mathbb{E}[p(R)] + \binom{n}{2}s_n\mathbb{E}[p(R)\log p(R)]}{\binom{n}{2}s_n\log(1/s_n)}\end{aligned} \tag{60}$$

By switching into the graphon notation and noting that $p(R) = p(\|X-Y\|) = W(X,Y)$, the quantity of interest, which is the right-hand side of (60) without the $nd\log m$ term, can be bounded as follows:

$$\begin{aligned}&\frac{H(G_n|M) - \binom{n}{2}s_n\log(1/s_n)\mathbb{E}[p(R)] - \binom{n}{2}\frac{s_n}{\ln 2}\mathbb{E}[p(R)] + \binom{n}{2}s_n\mathbb{E}[p(R)\log p(R)]}{\binom{n}{2}s_n\log(1/s_n)}\\
&\leq \frac{\mathbb{E}[h_2(s_n\overline{W}_m(X,Y))] - s_n\log(1/s_n)\mathbb{E}[W(X,Y)] - s_n\mathbb{E}[\frac{1}{\ln 2}W(X,Y) - W(X,Y)\log W(X,Y)]}{s_n\log(1/s_n)}\\
&\leq \frac{\mathbb{E}[s_n\log(1/s_n)\overline{W}_m(X,Y) + \frac{s_n}{\ln 2}\overline{W}_m(X,Y) - s_n\overline{W}_m(X,Y)\log\overline{W}_m(X,Y)]}{s_n\log(1/s_n)}\\
&\qquad - \frac{\mathbb{E}[s_n\log(1/s_n)W(X,Y) + \frac{s_n}{\ln 2}W(X,Y) - s_nW(X,Y)\log W(X,Y)]}{s_n\log(1/s_n)}\\
&\leq \frac{1}{s_n\log(1/s_n)}\int\int_{\mathcal{K}^2}|s_n\log(1/s_n)(\overline{W}_m(x,y) - W(x,y)) + \frac{s_n}{\ln 2}(\overline{W}_m(x,y) - W(x,y))\\
&\qquad - s_n(\overline{W}_m(x,y)\log\overline{W}_m(x,y) - W(x,y)\log W(x,y))|dxdy\\
&= \frac{1}{s_n\log(1/s_n)}\int\int_{\mathcal{K}^2}\left|\frac{s_n}{\ln 2}W(x,y) - s_nW(x,y)\log(s_nW(x,y))\right.\\
&\qquad \left. - \frac{s_n}{\ln 2}\overline{W}_m(x,y) - s_n\overline{W}_m(x,y)\log(s_n\overline{W}_m(x,y))\right|dxdy\\
&\leq \frac{1}{s_n\log(1/s_n)}\int\int_{\mathcal{K}^2}\frac{s_n}{\ln 2}|W(x,y) - \overline{W}_m(x,y)|\\
&\qquad - s_n|W(x,y) - \overline{W}_m(x,y)|\log(s_n|W(x,y) - \overline{W}_m(x,y)|)dxdy\end{aligned} \tag{61}$$

where we have used the fact that $\left|\frac{x}{\ln 2} - x\log x - \frac{y}{\ln 2} + y\log y\right| \leq \frac{1}{\ln 2}|x-y| - |x-y|\log|x-y|$ for $x, y \in [0,1]$. Now, we perform the same decomposition of the domain $\mathcal{K} \subset \bar{\mathcal{K}}_m \cup \delta\mathcal{K}_m$ as before, and note that $g(x) = \frac{x}{\ln 2} - x \log x$ is an increasing function on $[0, 1]$. By combining this with the Hölder bound, and provided that $m = m(n)$ with $m(1) \geq L^{1/\alpha}\ell\sqrt{2d}$,

$$|W(x,y) - \overline{W}_m(x,y)| \leq \min\left\{1, L\left(\frac{\ell\sqrt{2d}}{m}\right)^{\alpha}\right\} = L\left(\frac{\ell\sqrt{2d}}{m}\right)^{\alpha}, \tag{62}$$

we have that (61) is

$$\begin{aligned}
\int\int_{\bar{\mathcal{K}}_m^2} g(s_n|W(x,y) - \overline{W}_m(x,y)|)dxdy &\leq \int\int_{\bar{\mathcal{K}}_m^2} g\left(s_n L\left(\frac{\ell\sqrt{2d}}{m}\right)^{\alpha}\right)dxdy \\
&= g\left(s_n L\left(\frac{\ell\sqrt{2d}}{m}\right)^{\alpha}\right)\int\int_{\bar{\mathcal{K}}_m^2} dxdy \\
&\leq g\left(s_n L\left(\frac{\ell\sqrt{2d}}{m}\right)^{\alpha}\right) \\
&= s_n\frac{L}{\ln 2}\left(\frac{\ell\sqrt{2d}}{m}\right)^{\alpha} - s_n L\left(\frac{\ell\sqrt{2d}}{m}\right)^{\alpha}\log\left(s_n L\left(\frac{\ell\sqrt{2d}}{m}\right)^{\alpha}\right) \\
&= a_1 s_n\frac{\log(m^{\alpha}/s_n)}{m^{\alpha}} + a_2 s_n\frac{1}{m^{\alpha}},
\end{aligned}$$

where $a_1 = \frac{L}{\ln 2}(\ell\sqrt{2d})^{\alpha}$ and $a_2 = L(\ell\sqrt{2d})^{\alpha} - L(\ell\sqrt{2d})^{\alpha}\log(L(\ell\sqrt{2d})^{\alpha})$, and we have used the fact that $\int\int_{\bar{\mathcal{K}}_m^2} dxdy \leq \int\int_{\mathcal{K}^2} dxdy = 1$. Then, for the integrals over $\delta K_m \times \bar{\mathcal{K}}_m$ and $\bar{\mathcal{K}}^2$ we again do not have Hölder continuity. However we have once again, $g(x) \leq 1$ for $x \in [0, 1]$ and

$$\begin{aligned}
\int\int_{\mathcal{K}^2} g(s_n|W(x,y) - \overline{W}_m(x,y)|)dxdy = &\sum_{(M_i\times M_j)\in\bar{\mathcal{K}}_m^2}\int\int_{M_i\times M_j} g(s_n|W(x,y) - \overline{W}_m(x,y)|)dxdy \\
&+ \sum_{(M_i\times M_j)\in\delta\mathcal{K}_m^2}\int\int_{M_i\times M_j} g(s_n|W(x,y) - \overline{W}_m(x,y)|)dxdy \\
&+2\sum_{(M_i\times M_j)\in\bar{\mathcal{K}}_m\times\delta\mathcal{K}_m}\int\int_{M_i\times M_j} g(s_n|W(x,y) - \overline{W}_m(x,y)|)dxdy.
\end{aligned} \tag{63}$$

Next, note that since g is an increasing function on $[0, 1]$, we have $g(s_n|\overline{W}_m(x,y) - W(x,y)|) \leq g(s_n) = s_n\left(\frac{1}{\ln 2} - \log(s_n)\right)$. So by combining everything, we have

$$\begin{aligned}
&\int\int_{\mathcal{K}^2} g(s_n|W(x,y) - \overline{W}_m(x,y)|)dxdy \\
&\leq a_1 s_n\frac{\log(m^{\alpha}/s_n)}{m^{\alpha}} + a_2 s_n\frac{1}{m^{\alpha}} + a_3 s_n\left(\frac{1}{\ln 2} - \log(s_n)\right)\frac{1}{m} + a_4 s_n\left(\frac{1}{\ln 2} - \log(s_n)\right)\frac{1}{m^2},
\end{aligned}$$

where $a_3 = |\delta\mathcal{K}|\left(\frac{\ell}{m}\right)^2$, $a_4 = 2|\delta\mathcal{K}|\frac{\ell}{m}$, and we have used the same bound to count the number of boxes required to cover the boundary of $\mathcal{K}$ as in the proof of Theorem 3. We then divide by $s_n \log(1/s_n)$ to give

$$\Delta_n^{\text{sparse}} \leq \frac{a_1 s_n \frac{\log(m^\alpha/s_n)}{m^\alpha} + a_2 s_n \frac{1}{m^\alpha} + a_3 s_n \left(\frac{1}{\ln 2} - \log(s_n)\right)\frac{1}{m} + a_4 s_n \left(\frac{1}{\ln 2} - \log(s_n)\right)\frac{1}{m^2}}{s_n \log(1/s_n)} + \frac{nd\log m}{\binom{n}{2} s_n \log(1/s_n)}$$

$$= \frac{nd\log m}{\binom{n}{2} s_n \log(1/s_n)} + a_1 \frac{\alpha \log m + \log(1/s_n)}{m^\alpha \log(1/s_n)} + a_2 \frac{1}{m^\alpha \log(1/s_n)} + a_3 \frac{\frac{1}{\ln 2} + \log(1/s_n)}{m \log(1/s_n)} + a_4 \frac{\frac{1}{\ln 2} + \log(1/s_n)}{m^2 \log(1/s_n)}$$

Then, setting $m \propto n^{1/\alpha}$, we get

$$\Delta_n^{\text{sparse}} = \mathcal{O}\left(\frac{\log n}{n s_n \log(1/s_n)}\right) \tag{64}$$

which will go to 0 if $s_n \log(1/s_n) \gg \frac{\log n}{n}$, which we can see is implied by $ns_n \to \infty$. Finally, for the lower bound, note

$$0 \leq \frac{H(G_n) - H(G_n|\mathbf{X})}{\binom{n}{2} s_n \log(1/s_n)} = \Delta_n^{\text{sparse}} + \frac{\mathbb{E}\left[\sum_{k=2}^\infty \left(\frac{1}{k-1} - \frac{1}{k}\right)(s_n p(R))^k\right]}{\ln(2) s_n \log(1/s_n)} \tag{65}$$

$$\Rightarrow \Delta_n^{\text{sparse}} \geq -\frac{\mathbb{E}\left[\sum_{k=2}^\infty \left(\frac{1}{k-1} - \frac{1}{k}\right)(s_n p(R))^k\right]}{\ln(2) s_n \log(1/s_n)}$$

$$= -\frac{\mathbb{E}\left[\frac{1}{2}(s_n p(R))^2 + \sum_{k=3}^\infty \left(\frac{1}{k-1} - \frac{1}{k}\right)(s_n p(R))^k\right]}{\ln(2) s_n \log(1/s_n)}$$

$$\geq -\frac{\mathbb{E}\left[\frac{1}{2}(s_n p(R))^2 + (s_n p(R))^3 \sum_{k=3}^\infty \left(\frac{1}{k-1} - \frac{1}{k}\right)\right]}{\ln(2) s_n \log(1/s_n)}$$

$$= -\frac{s_n \mathbb{E}[p(R)^2] + s_n^2 \mathbb{E}[p(R)^3]}{2\ln(2)\log(1/s_n)}$$

as required. □

VI. Proof of Theorem 5

By conditioning on the node locations, the random geometric graph G_n becomes an inhomogeneous Erdős-Rényi graph, meaning that the conditional edge random variables are independent with different edge probabilities. With this observation, we can follow the approach used in [21, Thm. 10.3.1] to characterize $I_{G_n|\mathbf{X}}(D_n)$.

A. Lower Bound on $I_{G_n|\mathbf{X}}(D_n)$

Consider a conditional joint distribution $P_{\hat{G}_n|G_n,\mathbf{X}}$ satisfying the constraint $\mathbb{E}[d_n(G_n,\hat{G}_n)] \leq D_n$. For this distribution, we have

$$\begin{aligned}
I(G_n;\hat{G}_n|\mathbf{X}) &= H(G_n|\mathbf{X}) - H(G_n|\hat{G}_n,\mathbf{X}) \\
&\overset{(a)}{=} \sum_{i<j} H(E_{ij}|X_i,X_j) - H(G_n|\hat{G}_n,\mathbf{X}) \\
&\geq \sum_{i<j} H(E_{ij}|X_i,X_j) - \sum_{i<j} H(E_{ij}|\hat{E}_{ij},X_i,X_j) \\
&= \sum_{i<j} \mathbb{E}_R\big[H(E_{ij}|R_{ij}=R) - H(E_{ij}\oplus\hat{E}_{ij}|\hat{E}_{ij},R_{ij}=R)\big] \\
&\overset{(b)}{\geq} \sum_{i<j} \mathbb{E}_R\big[\max\big\{H(E_{ij}|R_{ij}=R) - H(E_{ij}\oplus\hat{E}_{ij}|R_{ij}=R),0\big\}\big] \\
&\overset{(c)}{=} \sum_{i<j} \mathbb{E}\big[\max\big\{h_2(p_n(R)) - h_2(q_{ij}(R)),0\big\}\big] \\
&\geq \mathbb{E}\Big[\max\Big\{\binom{n}{2}h_2(p_n(R)) - \sum_{i<j} h_2(q_{ij}(R)),0\Big\}\Big] \\
&\overset{(d)}{\geq} \binom{n}{2}\mathbb{E}\big[h_2(p_n(R))\big] - \binom{n}{2}\mathbb{E}\Big[\min\Big\{h_2(q_n(R)),h_2(p_n(R))\Big\}\Big]
\end{aligned}$$

In (a), we use the conditional independence of the edge random variables. In the inequality (b), the bound $H(E_{ij}|R_{ij}=r) - H(E_{ij}\oplus\hat{E}_{ij}|R_{ij}=r)$ is useful only when it is greater than 0 because we already have $H(E_{ij}|R_{ij}=r) - H(E_{ij}\oplus\hat{E}_{ij}|\hat{E}_{ij},R_{ij}=r) = I(E_{ij};\hat{E}_{ij}|R_{ij}=r) \geq 0$. In (c), we define

$$q_{i,j}(r) \triangleq \min\Big\{\mathbb{P}(E_{ij}\oplus\hat{E}_{ij}=1|R_{ij}=r), \mathbb{P}(E_{ij}\oplus\hat{E}_{ij}=0|R_{ij}=r)\Big\}$$

for $r \in [0,K]$ and $i<j$. The inequality (d), where we set $q_n(r) \triangleq \sum_{i<j} \frac{q_{i,j}(r)}{\binom{n}{2}}$ for $r \in [0,K]$, follows immediately from Jensen's inequality and the concavity of the binary entropy function:

$$\sum_{i<j} h_2(q_{i,j}(r)) \leq \binom{n}{2} h_2(q_n(r)).$$

It immediately follows from the distortion condition that $q_n(r)$ satisfies the condition, $\binom{n}{2}\mathbb{E}[q_n(R)] \leq D_n$.

By minimizing the lower bound in (d) over all functions $q_n(r)$ satisfying the condition $\binom{n}{2}\mathbb{E}[q_n(R)] \leq D_n$, we get

$$I_{G_n|\mathbf{X}}(D_n) \geq \binom{n}{2}\Big[\mathbb{E}\left[h_2(p_n(R))\right] - \max_{\substack{q_n:[0,K]\to[0,1]:\\ \binom{n}{2}\mathbb{E}[q_n(R)]\leq D_n}} \mathbb{E}\left[\min\left\{h_2(q_n(R)),h_2(p_n(R))\right\}\right]\Big]. \tag{66}$$

The optimization problem in (66) is equivalent to solving the following problem:

$$A^* \triangleq \max_{\substack{q_n:[0,K]\to[0,1]: \\ q_n(r)\le \bar{p}_n(r),\ \binom{n}{2}\mathbb{E}[q_n(R)]\le D_n}} \mathbb{E}\left[h_2(q_n(R))\right], \tag{67}$$

where we use the notation $\bar{f}(r)$ to denote $\min\{f(r), 1-f(r)\}$ for a function $f:[0,K]\to[0,1]$. Let us consider two separate cases to solve this optimization problem by finding an optimizer q_n^*. In the case of $\binom{n}{2}\mathbb{E}[\bar{p}_n(R)] \le D_n$,

$$A^* = \mathbb{E}\left[h_2(p_n(R))\right]$$

with $q_n^*(r) = \bar{p}_n(r)$ because $\mathbb{E}\left[h_2(q_n(R))\right] \le \mathbb{E}\left[h_2(p_n(R))\right]$ for every $q_n(r) \le \bar{p}_n(r)$. Here, $q_n^*(r) = \min\{\bar{p}_n(r), \mu_n\}$ with $\mu_n = \frac{1}{2}$.

In the case of $D_n \le \binom{n}{2}\mathbb{E}[\bar{p}_n(r)]$. Consider the function $q_n^*(r) = \min\{\bar{p}_n(r), \mu_n\}$ with μ_n chosen such that $\binom{n}{2}\mathbb{E}[q_n^*(R)] = D_n$. It is possible to choose such a μ_n depending on the value of D_n because $\binom{n}{2}\mathbb{E}[\min\{\bar{p}_n(r), \mu_n\}]$ is a continuous[2] and non-decreasing function of $\mu_n \in [0, \frac{1}{2}]$ with the maximum value being $\binom{n}{2}\mathbb{E}[\bar{p}_n(r)]$, which is greater than D_n.

We will now show that $A^* = \mathbb{E}\left[h_2(q_n^*(R))\right]$, i.e., $\mathbb{E}\left[h_2(q_n(R))\right] \le \mathbb{E}\left[h_2(q_n^*(R))\right]$ for any q that satisfies the constraints in the optimization problem (67). Assume that $D_n > 0$ (which implies that $\mu_n > 0$). On the other hand, if $D_n = 0$ then $A^* = 0$ because $\mathbb{E}\left[h_2(q_n(R))\right] \le h_2(\mathbb{E}\left[q_n(R)\right]) \le h_2\left(D_n/\binom{n}{2}\right) = 0$.

As a consequence of the concavity of the binary function, we have the following relation:

$$h_2(q_n(r)) - h_2(q_n^*(r)) \le [q_n(r) - q_n^*(r)]h_2'(q_n^*(r)),$$

for any $q_n(r) \in [0,1]$ and $q_n^*(r) \in (0,1)$ with $h_2'(q_n^*(r)) = \log_2\left(\frac{1-q_n^*(R)}{q_n^*(R)}\right)$. If $q_n^*(r) = 0$ then $\bar{p}_n(r) = 0$ (because $\mu_n > 0$), which means that $q_n(r) \le \bar{p}_n(r) = 0$. By using these observations, we can write

$$\mathbb{E}\left[h_2(q_n(R)) - h_2(q_n^*(R))\right] \le \mathbb{E}\left[[q_n(R) - q_n^*(R)]\log_2\left(\frac{1-q_n^*(R)}{q_n^*(R)}\right)\mathbb{1}_{q_n^*(R)>0}\right]. \tag{68}$$

Let $\mathcal{A} = \{r : \mu_n \le \bar{p}_n(r)\}$. On $\mathcal{A}$, $q_n^*(r) = \mu_n$, and on $\mathcal{A}^c$, $q_n^*(r) = \bar{p}_n(r)$. Since $\binom{n}{2}\mathbb{E}[q_n(R)] \le D_n = \binom{n}{2}\mathbb{E}[q_n^*(R)]$, we have

$$0 \ge \mathbb{E}[q_n(R) - q_n^*(R)]$$

[2]As $\min\{\bar{p}_n(r), \mu_n\}$ is bounded from above by $\bar{p}_n(r)$, which is an integrable function, the continuity immediately follows by applying the dominated convergence theorem (DCT) for any convergent sequence $\mu_n \to \mu$.

$$= \mathbb{E}\left[[q_n(R) - q_n^*(R)]\mathbb{1}_{q_n^*(R)>0}\right]$$
$$= \mathbb{E}\left[[q_n(R) - q_n^*(R)]\mathbb{1}_{q_n^*(R)>0}\mathbb{1}_{A}\right] + \mathbb{E}\left[[q_n(R) - q_n^*(R)]\mathbb{1}_{q_n^*(R)>0}\mathbb{1}_{\mathcal{A}^c}\right]. \tag{69}$$

Using the above observations, we can bound (68) as follows:

$$\begin{aligned}
&\mathbb{E}\left[h_2(q_n(R)) - h_2(q_n^*(R))\right] \\
&\le \mathbb{E}\left[[q_n(R) - q_n^*(R)]\log_2\left(\frac{1-q_n^*(R)}{q_n^*(R)}\right)\mathbb{1}_{q_n^*(R)>0}\right] \\
&= \mathbb{E}\left[[q_n(R) - q_n^*(R)]\log_2\left(\frac{1-q_n^*(R)}{q_n^*(R)}\right)\mathbb{1}_{q_n^*(R)>0}\mathbb{1}_{\mathcal{A}}\right] \\
&\qquad\qquad + \mathbb{E}\left[[q_n(R) - q_n^*(R)]\log_2\left(\frac{1-q_n^*(R)}{q_n^*(R)}\right)\mathbb{1}_{q_n^*(R)>0}\mathbb{1}_{\mathcal{A}^c}\right] \\
&= \mathbb{E}\left[[q_n(R) - q_n^*(R)]\log_2\left(\frac{1-\mu_n}{\mu_n}\right)\mathbb{1}_{q_n^*(R)>0}\mathbb{1}_{\mathcal{A}}\right] \\
&\qquad\qquad + \mathbb{E}\left[[q_n(R) - q_n^*(R)]\log_2\left(\frac{1-\bar{p}_n(R)}{\bar{p}_n(R)}\right)\mathbb{1}_{q_n^*(R)>0}\mathbb{1}_{\mathcal{A}^c}\right] \\
&= \log_2\left(\frac{1-\mu_n}{\mu_n}\right)\mathbb{E}\left[[q_n(R) - q_n^*(R)]\mathbb{1}_{q_n^*(R)>0}\mathbb{1}_{\mathcal{A}}\right] \\
&\qquad\qquad + \mathbb{E}\left[[q_n(R) - q_n^*(R)]\log_2\left(\frac{1-\bar{p}_n(R)}{\bar{p}_n(R)}\right)\mathbb{1}_{q_n^*(R)>0}\mathbb{1}_{\mathcal{A}^c}\right] \\
&\overset{(e)}{\le} -\log_2\left(\frac{1-\mu_n}{\mu_n}\right)\mathbb{E}\left[[q_n(R) - q_n^*(R)]1_{q_n^*(R)>0}\mathbb{1}_{\mathcal{A}^c}\right] \\
&\qquad\qquad + \mathbb{E}\left[[q_n(R) - q_n^*(R)]\log_2\left(\frac{1-\bar{p}_n(R)}{\bar{p}_n(R)}\right)\mathbb{1}_{q_n^*(R)>0}\mathbb{1}_{\mathcal{A}^c}\right] \\
&= \mathbb{E}\left[[q_n(R) - q_n^*(R)]\left[\log_2\left(\frac{1-\bar{p}_n(R)}{\bar{p}_n(R)}\right) - \log_2\left(\frac{1-\mu_n}{\mu_n}\right)\right]\mathbb{1}_{q_n^*(R)>0}\mathbb{1}_{\mathcal{A}^c}\right] \\
&= \mathbb{E}\left[[q_n(R) - \bar{p}_n(R)]\left[\log_2\left(\frac{1-\bar{p}_n(R)}{\bar{p}_n(R)}\right) - \log_2\left(\frac{1-\mu_n}{\mu_n}\right)\right]\mathbb{1}_{q_n^*(R)>0}\mathbb{1}_{\mathcal{A}^c}\right] \\
&\le 0,
\end{aligned}$$

where (e) follows from (69), and the last inequality is due to the fact that $q_n(r) \le \bar{p}_n(r)$ and the fact that on the event A^c, $\bar{p}_n(r) \le \mu_n$, which implies that $\log_2\left(\frac{1-\bar{p}_n(r)}{\bar{p}_n(r)}\right) \ge \log_2\left(\frac{1-\mu_n}{\mu_n}\right)$ as the derivative of binary entropy function is non-increasing. This shows that $A^* = \mathbb{E}\left[h_2(q_n^*(R))\right]$, with $q_n^*(r)$ being an optimizer. Hence we have

$$I_{G_n|\mathbf{X}}(D_n) \ge \binom{n}{2}\left[\mathbb{E}\left[h_2(p_n(R))\right] - \mathbb{E}\left[h_2(q_n^*(R))\right]\right], \tag{70}$$

where $q^* : [0,K] \to [0,1]$ and $q_n^*(r)$ is defined as above.

B. Upper Bound on $I_{G_n|\mathbf{X}}(D_n)$:

Let $q_n^*(r)$ be the optimizer in (70). Consider the conditional probability distribution $P_{\hat{G}_n|G_n,\mathbf{X}}$ (or equivalently, $P_{\{\hat{E}_{ij}\}_{i<j}|\{E_{ij}\}_{i<j},\mathbf{X}}$) of the form

$$P_{\{\hat{E}_{ij}\}_{i<j}|\{E_{ij}\}_{i<j},\mathbf{X}} = \prod_{i<j} \frac{P_{E_{ij}|\hat{E}_{ij},X_i,X_j} P_{\hat{E}_{ij}|X_i,X_j}}{P_{E_{ij}|X_i,X_j}},$$

where $P_{E_{ij}|\hat{E}_{ij},X_i,X_j}(\cdot|\cdot,x_i,x_j)$ is a binary symmetric channel with the crossover probability $q_n^*(r_{ij})$, $P_{\hat{E}_{ij}|X_i,X_j}(1|x_i,x_j) = \frac{p_n(r_{ij})-q_n^*(r_{ij})}{1-2q_n^*(r_{ij})}$ and $P_{\hat{E}_{ij}|X_i,X_j}(0|x_i,x_j) = \frac{1-p_n(r_{ij})-q_n^*(r_{ij})}{1-2q_n^*(r_{ij})}$, which are non-negative by virtue of the way the $q_n^*(r_{ij})$'s are defined. We can easily see that this distribution satisfies the distortion criterion:

$$\mathbb{E}[d_n(G_n,\hat{G}_n)] = \sum_{i<j} \mathbb{P}(E_{ij} \oplus \hat{E}_{ij} = 1) = \sum_{i<j} \mathbb{E}\big[q_n^*(R)\big] = D_n,$$

which implies that

$$I_{G_n|\mathbf{X}}(D_n) \le I(G_n;\hat{G}_n|\mathbf{X}) = \binom{n}{2} \left[\mathbb{E}[h_2(p_n(R))] - \mathbb{E}[h_2(q_n^*(R))]\right]. \tag{71}$$

By combining (71) and (70), we prove Theorem 5.

□

VII. Conclusion

In this work, we studied the fundamental limits of lossless and lossy compression of soft random geometric graphs. We considered two regimes of interest, namely dense and sparse. In both these regimes, we gave an asymptotic characterization of the entropy of an SRGG. The proof involves the use of tools from the theory of graphons along with the discretization of the underlying domain. The main result is that the normalized entropy approaches the limiting quantity as $\mathcal{O}\left(\frac{\log n}{n}\right)$ and $\mathcal{O}\left(\frac{\log n}{ns_n \log(1/s_n)}\right)$, in dense and sparse regimes respectively, if the connection function p is Hölder continuous. Similarly, we derive an asymptotic characterization of the information-distortion function of an SRGG with Hamming distortion measure. As with entropy, the normalized information-distortion function also converges to its limiting quantity as $\mathcal{O}\left(\frac{\log n}{n}\right)$ and $\mathcal{O}\left(\frac{\log n}{ns_n \log(1/s_n)}\right)$, in dense and sparse regimes respectively, if the connection function p is Hölder continuous. A future direction for the lossy compression work would be to consider various distortion measure that take into account graph features such as clustering, transitivity and so on. Another interesting future research direction is to quantify the *structural* entropy rate of the SRGG in the regime where ns_n is constant, or even decreasing. This will likely require different proof techniques, as the idea of discretising the space now introduces too much entropy into the upper bound (see the proof of Theorems 3 and 4). In the regime $ns_n \to 0$, the expected degree of each node tends to 0, and the graph is disconnected with high probability (since there is a node with degree 0 with high probability), and so it would also be interesting to understand the influence of the connectivity of the graph on the entropy.

REFERENCES

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

[2] P. Delgosha and V. Anantharam, "A universal lossless compression method applicable to sparse graphs and heavy-tailed sparse graphs," *IEEE Transactions on Information Theory*, vol. 69, no. 2, pp. 719–751, 2022.

[3] 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.

[4] 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.

[5] T. Łuczak, A. Magner, and W. Szpankowski, "Compression of preferential attachment graphs," in *Proc. IEEE Int. Symp. Inf. Theory (ISIT)*, Paris, France, Jul. 2019, pp. 1697–1701.

[6] K. Turowski, A. Magner, and W. Szpankowski, "Compression of dynamic graphs generated by a duplication model," *Algorithmica*, vol. 82, no. 9, pp. 2687–2707, Apr. 2020.

[7] 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.

[8] I. Kontoyiannis, Y. Lim, K. Papakonstantinopoulou, and W. Szpankowski, "Symmetry and the entropy of small-world structures and graphs," in *Proc. IEEE Int. Symp. Inf. Theory (ISIT)*, Melbourne, Australia, Jul. 2021, pp. 3026–3031.

[9] 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.

[10] 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.

[11] ——, "Structural complexity of one-dimensional random geometric graphs," *IEEE Trans. Inf. Theory*, vol. 69, no. 2, pp. 794–812, Sep. 2023.

[12] 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

[13] J. P. Coon, "Topological uncertainty in wireless networks," in *2016 IEEE Global Communications Conference (GLOBECOM)*. IEEE, 2016, pp. 1–6.

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

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

[16] 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.

[17] 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.

[18] L. Lovász, *Large networks and graph limits*. American Mathematical Soc., 2012, vol. 60.

[19] S. Janson, “Graphons, cut norm and distance, couplings and rearrangements,” *New York Journal of Mathematics Monographs*, 2013. [Online]. Available: https://nyjm.albany.edu/m/2013/4v.pdf

[20] S. Janson and S. Olhede, “Can smooth graphons in several dimensions be represented by smooth graphons on [0, 1]?” *Examples and Counterexamples*, vol. 1, p. 100011, 2021.

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

[22] 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.

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

[24] T. S. Han, *Information-spectrum methods in information theory*. Berlin, Germany: Springer-Verlag, 2003.

[25] R. Gray, “A new class of lower bounds to information rates of stationary sources via conditional rate-distortion functions,” *IEEE Trans. Inf. Theory*, vol. 19, no. 4, pp. 480–489, Jul. 1973.