Random Geometric Complexes and Graphs on Riemannian Manifolds in the Thermodynamic Limit

Publisher:

note finali coverpage

(Article begins on next page)

20 May 2024

RANDOM GEOMETRIC COMPLEXES AND GRAPHS ON RIEMANNIAN MANIFOLDS IN THE THERMODYNAMIC LIMIT

ANTONIO LERARIO AND RAFFAELLA MULAS

ABSTRACT. We investigate some topological properties of random geometric complexes and random geometric graphs on Riemannian manifolds in the thermodynamic limit. In particular, for random geometric complexes we prove that the normalized counting measure of connected components, counted according to isotopy type, converges in probability to a deterministic measure. More generally, we also prove similar convergence results for the counting measure of types of components of each k–skeleton of a random geometric complex. As a consequence, in the case of the 1–skeleton (i.e. for random geometric graphs) we show that the empirical spectral measure associated to the normalized Laplace operator converges to a deterministic measure.

1. INTRODUCTION

1.1. **Random geometric complexes.** The subject of random geometric complexes has recently attracted a lot of attention, with a special focus on the study of expectation of topological properties of these complexes [Kah11b, Pen03b, NSW08, YSA17, BM15, BK18][1] (e.g. number of connected components, or more generally Betti numbers). In a recent paper [ALL20], Auffinger, Lerario and Lundberg have imported methods from [NS16, SW19] for the study of finer properties of these random complexes, namely the distribution of the homotopy types of the connected components of the complex. Before moving to the content of the current paper, we discuss the main ideas from [ALL20] and introduce some terminology.

Let (M, g) be a compact, Riemannian manifold of dimension m. We normalize the metric g in such a way that
$$\mathrm{vol}(M) = 1.$$
We denote by $\hat{B}(x, r) \subset M$ the Riemannian ball centered at x of radius $r > 0$ and we construct a *random M–geometric complex* in the *thermodynamic regime* as follows. We let $\{p_1, \ldots, p_n\}$ be a set of points independently sampled from the uniform distribution on M, we fix a positive number $\alpha > 0$, and we consider:
$$\mathcal{U}_n := \bigcup_{k=1}^n \hat{B}(p_k, r) \quad \text{where} \quad r := \alpha n^{-1/m}. \tag{1.1}$$
The choice of such r is what defines the so-called *critical* or *thermodynamic regime*[2] and it's the regime where topology is the richest [ALL20, Kah11a]. We say that $\mathcal{U}_n$ is a *random*

---

[1]This list is by no mean complete, see [BK18] for a survey and a more complete set of references!

[2]Quoting the Introduction from [ALL20]: random geometric complexes are studied *within three main phases or regimes based on the relation between density of points and radius of the neighborhoods determining the complex: the subcritical regime (or "dust phase") where there are many connected components with little topology, the critical regime (or "thermodynamic regime") where topology is the richest (and where the percolation threshold appears), and the supercritical regime where the connectivity threshold appears. The thermodynamic regime is seen to have the most intricate topology.*
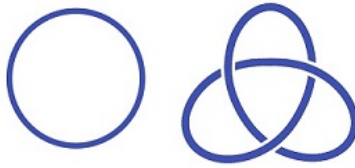
FIGURE 1. The unknot and the trefoil knot are homotopy equivalent but they are not isotopic.

M–*geometric complex* and the name is motivated by the fact that, for n large enough, $\mathcal{U}_n$ is homotopy equivalent to its Čech complex, as we shall see in Lemma 2.4 below.

Auffinger, Lerario and Lundberg [ALL20] proved that, in the case when $\mathrm{vol}\, M = 1$, the normalized counting measure of connected components of such complexes, counted according to homotopy type, converges in probability to a deterministic measure. That is,

$$\tilde{\Theta}_n := \frac{1}{b_0(\mathcal{U}_n)} \sum \delta_{[u]} \xrightarrow[n\to\infty]{\mathbb{P}} \tilde{\Theta}, \tag{1.2}$$

where the sum is over all connected components u of $\mathcal{U}_n$, $[u]$ denotes their homotopy type and b_0 is the zero-th Betti number, therefore $b_0(\mathcal{U}_n)$ is the number of connected components of $\mathcal{U}_n$. In (1.2) the measure $\tilde{\Theta}_n$ is a *random* probability measure on the countable set of all possible homotopy types of connected geometric complexes and the convergence is in probability with respect to the total variation distance (see Section 4 for more precise definitions). The support of the limiting deterministic measure $\tilde{\Theta}$ equals the set of all homotopy types for Euclidean geometric complexes of dimension $m = \dim M$. Roughly speaking, (1.2) tells that, for every fixed homotopy type $[u]$ of connected geometric complexes, denoting by $\mathcal{N}_n([u])$ the random variable "number of connected components of $\mathcal{U}_n$ which are in the homotopy equivalence class $[u]$", there is a convergence of the random variable $\mathcal{N}_n([u])/b_0(\mathcal{U}_n)$ to a constant $c_{[u]}$ as $n \to \infty$ (the convergence is in L^1 and $c_{[u]} > 0$ if and only if $[u]$ contains a $\mathbb{R}^m$–geometric complex).

1.2. **Isotopy classes of geometric complexes.** We now move to the content of the current paper. Our first goal is to include the results of [ALL20] into a more general framework which allows to make even more refined counts (e.g. according to the type of the embedding of the components, or on the structure of their skeleta, or on the property of containing a given motif[3]). The first result that we prove is that (1.2) still holds if we consider *isotopy classes* instead of homotopy classes: intuitively, two complexes are isotopic if the vertices of one can be moved continuously to the vertices of the other without ever changing the combinatorics of the intersection of the corresponding balls (see Definition 2.5). From now on we will always make the assumption that our complexes are *nondegenerate*, i.e. that the boundaries of the balls defining them intersect transversely (see Definition 2.3); our random geometric complexes will be nondegenerate with probability one, and the notion of isotopic nondegenerate complexes coincides with the one from differential topology. In Theorem 4.2 we show that

$$\Theta_n \xrightarrow[n\to\infty]{\mathbb{P}} \Theta,$$

---

[3]A *motif* in a graph (or more generally in a complex) is a recurrent and statistically significant sub-graph or pattern.

where Θ_n is defined in a similar way as $\tilde{\Theta}_n$ above, with isotopy classes instead of homotopy classes Interestingly, the limiting measure depends only on α on the dimension of M.[4]. To appreciate the difference with the results from [ALL20]: the unknot and the trefoil knot in $\mathbb{R}^3$ (Figure 1) are homotopy equivalent but they are not isotopic, and with positive probability there are connected M–geometric complexes whose embedding looks like these two knots (see Proposition 4.5 below); Theorem 4.2 is able to distinguish between them, whereas the construction from [ALL20] is not.

1.3. **A cascade of measures.** Theorem 4.2 contains in a sense the richest possible information on the topological structure of our geometric complexes and the convergence of many other counting measures can be deduced from it. To explain this idea, we consider the space

$$(\mathcal{G}/\cong) := \{\text{isotopy classes of connected geometric complexes}\}$$

and we put an equivalence relation ρ on $\mathcal{G}/\cong$ (the relation can be for example: two isotopy classes are the same if their k–skeleta are isomorphic, or if they contain the same number of a given motif). Then the natural map $\psi : (\mathcal{G}/\cong) \to (\mathcal{G}/\cong)/\rho$ defines the random pushforward measure $\psi_*\Theta_n$ on $(\mathcal{G}/\cong)/\rho$ and Theorem 4.2 implies that $\psi_*\Theta_n \to \psi_*\Theta$.

This idea can be used to produce a "cascade" of random relevant measures. Consider in fact the following diagram of maps and spaces:

$$
\begin{array}{ccc}
& \xrightarrow{\phi} & \\
\mathcal{G}/\cong \xrightarrow{\varphi} & \mathcal{G}/\simeq \longrightarrow & \mathcal{G}/\sim \\
\llap{\scriptstyle \varphi^{(k)}}\searrow & \downarrow & \\
& \mathcal{G}^{(k)}/\simeq &
\end{array}
$$

where the spaces are:

$$(\mathcal{G}/\simeq) := \{\text{isomorphism classes of connected geometric Čech complexes}\}$$
$$\left(\mathcal{G}^{(k)}/\simeq\right) := \{\text{isomorphism classes of components of the } k\text{–skeleton of Čech complexes}\}$$
$$(\mathcal{G}/\sim) := \{\text{homotopy classes of connected geometric complexes}\}$$

and the maps are the natural "forgetful" maps. For example, the map φ takes the isotopy class of a nondegenerate complex and associates to it its homotopy class; the map $\varphi^{(k)}$ associates to it the isomorphism class of its k–skeleton (it is well defined since isotopic complexs have isomorphic Čech complexes). Then for all the pushforward measures defined by these maps we have convergence in probability with respect to the total variation distance (see Section 4) and as $n \to \infty$

$$\phi_*\Theta_n \to \phi_*\Theta, \quad \varphi_*\Theta_n \to \varphi_*\Theta \quad \text{and} \quad \varphi_*^{(k)}\Theta_n \to \varphi_*^{(k)}\Theta.$$

---

[4]In the rest of the paper we will consider α as fixed from the very beginning and omit it from the notation; the study of the dependence of the various objects on $\alpha > 0$ is an interesting problem, on which for now we cannot say much.

1.4. **Random Geometric Graphs.** Of special interest is the case of random geometric graphs: vertices of a random M–geometric graph Γ_n are the points $\{p_1, \ldots, p_n\}$ and we put an edge between p_i and p_j if and only if $i \neq j$ and $\hat{B}(p_i, r) \cap \hat{B}(p_j, r) \neq \emptyset$. Using the above language, a random M–geometric graph is the 1–skeleton of the Čech complex associated to the complex $\mathcal{U}_n$.

To every random M–geometric graph Γ_n we can associate the measure:

$$\varphi_*^{(1)}\Theta_n = \frac{1}{b_0(\Gamma_n)} \sum \delta_\gamma \to \varphi_*^{(1)}\Theta, \tag{1.3}$$

where the sum is over all connected components of Γ_n and γ denotes their isomorphism class (as graphs). There is an interesting fact regarding the random variable $b_0(\Gamma_n)$ appearing in (1.3): it is the same random variable as $b_0(\mathcal{U}_n)$ (the number of components of the random graph and of the random complex are the same), and in [ALL20] it is proved that there exists a constant β (depending on the parameter α in (1.1)) such that:

$$\frac{b_0(\Gamma_n)}{n} = \frac{b_0(\mathcal{U}_n)}{n} \xrightarrow{L^1} \beta.$$

The existence of this limit also follows from [GTT19], where the authors establish a limit law in the thermodynamic regime for Betti numbers of random geometric complexes built over possibly inhomogeneous Poisson point processes in Euclidean space, including the case when the point process is supported on a submanifold. Moreover, we note that for a related model of random graphs (the Poisson model on $\mathbb{R}^m$, see Section 1.6 below) Penrose [Pen03b] has proved that there exists a constant β (depending on the parameter α in (1.1)) such that the normalized component count converges to a constant in L^2. In fact, as we will see below, related to our M–geometric model there is a way to construct a corresponding $\mathbb{R}^m$–geometric model, which is in a sense the rescaled limit of the Riemannian one, and the limit constants for the two models are the same.

In fact the limit measure $\varphi_*^{(1)}\Theta$ also comes from the rescaled Euclidean limit and it is supported on connected $\mathbb{R}^m$–geometric graphs. For a given m, the set of such graphs is not easy to describe, but in the case $m = 1$ they can be characterized by a result of Roberts [Rob69], and from this result we can deduce a description of the support of the limit measure in (1.3) (see Corollary 5.4 and Section 5 for more details).

Remark 1.1 (Related work on random geometric graphs). The general theory of random graphs has been founded in 1959 by Erdös and Rényi, who proposed a model of random graph $G(n, p)$ where the number of vertices is fixed to be n and each pair of distinct vertices is joined by an edge with probability p, independently of other edges [ER59, ER60, ER61a, ER61b, DJ10]. Later on, other models have been proposed in the literature [DGK17], as for instance the Barabási–Albert scale-free network model [BA99] and the Watts-Strogatz small-world network model [WS98]. For general references on random graphs, the reader is referred to [Bol85, Chu97, CL06, CdV98, Dur07, ES74, JLR00, Kol99, M.85]. Here we focus on the random geometric graph model and we refer to [Gil61, Pen03a, Wal11] for more literature on this topic. Applications of random geometric graphs can be found, for instance, in the contexts of wireless networks, epidemic spreading, city growth, power grids, protein-protein interaction networks [DGK17].

1.5. **The spectrum of a random geometric graph.** When talking about a graph, a natural associated object to look at is its normalized Laplace operator, see Section 6. It is known that the spectrum of the (symmetric) normalized Laplace operator for graphs encodes

important information about the graphs [Chu97]. For example, it tells us how many connected components a graph has; it tells whether a graph is bipartite and whether it is complete; it tells us how difficult it is to partition the vertex set of a graph into two disjoint sets V_1 and V_2 such that the number of edges between V_1 and V_2 is as small as possible and such that the *volume* of both V_1 and V_2, i.e. the sum of the degrees of their vertices, is as big as possible. Therefore, the normalized Laplace operator gives a partition of graphs into families and *isospectral graphs* share important common features. Since, furthermore, the computation of the eigenvalues can be performed with tools from linear algebra, such operator is a very powerful and used tool in graph theory and data analytics.

In the context of random M–geometric graphs, the convergence of the counting measure in (1.3) can be used to deduce the existence of a limit measure for the spectrum of the normalized Laplace operator for random geometric graphs. More specifically, we define the *empirical spectral measure* of a graph as the normalized counting measure of eigenvalues of the normalized Laplace operator and we prove that there exists a deterministic measure μ on the real line such that (Theorem 7.4)

$$\mu_{\Gamma_n} := \frac{1}{n} \sum_{i=1}^{n} \delta_{\lambda_i(\Gamma_n)} \xrightarrow[n \to \infty]{*} \mu. \tag{1.4}$$

Here, $\lambda_1(\Gamma_n), \ldots, \lambda_n(\Gamma_n)$ are the eigenvalues of the normalized Laplace operator of Γ_n and the convergence in (1.4) means that for every continuous function $f : [0, 2] \to \mathbb{R}$ we have:

$$\lim_{n \to \infty} \mathbb{E} \int_{[0,2]} f \, d\mu_{\Gamma_n} = \int_{[0,2]} f \, d\mu.$$

The measure μ in (1.4) is far from trivial and we don't have yet a clear understanding of it: we know it is supported on the interval $[0, 2]$, but for example it is not absolutely continuous with respect to Lebesgue measure (in fact $\mu(\{0\}) = \beta > 0$).

Remark 1.2. Interestingly, [DJ10] studies the convergence of μ_{Γ_n} as $n \to \infty$ in the case where Γ_n is a $G(n, p)$ random graphs and the eigenvalues are the ones of the non-normalized Laplacian or the ones of the adjacency matrix. In particular, it is shown that in such context, under suitable conditions, μ_{Γ_n} converges to the semi-circle law if associated to the adjacency matrix and it converges to the free convolution of the standard normal distribution if associated to the non-normalized Laplacian.

Remark 1.3. In [GJLS16], Gu, Jost, Liu and Stadler introduce the notion of *spectral class* of a family of graphs. Given a Radon measure ρ on $[0, 2]$ and a sequence $(\Gamma_n)_{n \in \mathbb{N}}$ of graphs with $\#(V(\Gamma_n)) = n$, they say that this sequence belongs to the spectral class ρ if $\mu_{\Gamma_n} \xrightarrow{*} \rho$ as $n \to \infty$. We can interpret (1.4) as saying that our family of random geometric graphs $(\Gamma_n)_n$ belongs to the spectral class μ (in a probabilistic sense).

Remark 1.4 (Related work on spectral theory). Similarly to the spectrum of the normalized Laplace operator, also the spectra of the non-normalized Laplacian matrix (defined in Section 6) and the one of the adjacency matrix have been widely studied. We refer the reader to [Chu97, Pup08] for general references on spectral graph theory. We refer to [BLW76, JM19] for applications of spectral graph theory in chemistry and we refer to [Eva92, EE92, Eva83, MF91, Nov98, RB88, RDD90] for applications in theoretical physics and quantum mechanics. For references on spectral graph theory of (not necessarily geometric) random graphs, we refer to [CL06, DJ10, DGK17, NGB15]. In [DGK17], in particular, the eigenvalues of the adjacency matrix for random gometric graphs are studied using numerical

and statistical methods. Remarkably, it is shown that random geometrix graphs are statistically very similar to the other random graph models we have mentioned above: Erdős-Rényi random graphs, Barabási–Albert scale-free networks, Watts-Strogatz small-world networks. On the other hand, in [NGB15], it is shown that symmetric structures abundantly occur in random geometric graphs, while the same doesn't hold for the other random graph models. Our main results on spectral graph theory for random geometric graphs, Theorem 7.4 and Proposition 7.5 below, follow the same general idea as [DGK17] and [NGB15], in the sense that we are interested on the limiting spectrum of large random geometric graphs. The main difference is that [DGK17] is focused on the adjacency matrix, [NGB15] gives a focus on the non-normalized Laplacian and we focus on the normalized Laplacian. Therefore the final implications differ very much.

1.6. **The Euclidean Poisson model.** As we already observed, in [ALL20], the proof of (1.2) is based on a *rescaling limit* idea. Namely, one can fix $R > 0$ and a point $p \in M$, and study the limit structure of the random complexes inside the ball $\hat{B}(p, Rn^{-1/d})$. The random geometric complex obtained as $n \to \infty$ can then be described as follows. Let $P := \{p_1, p_2, \ldots\}$ be a set of points sampled from the standard spatial Poisson distribution in $\mathbb{R}^m$. For $\alpha > 0$, let

$$\mathcal{P} := \bigcup_{p \in P} B(p, \alpha)$$

and let

$$\mathcal{P}_R := \{\text{connected components of } \mathcal{P} \text{ entirely contained in the interior of } B(0, R)\}.$$

For the random complex $\mathcal{P}_R$, one can define completely analogue measures, where now the parameter is $R > 0$, and all the above discussion applies also to this model (this is discussed throughout the paper).

Structure of the paper. This paper is structured as follows. In Section 2 we discuss (deterministic) M–geometric complexes and, in particular, we define and see some properties of the set $\mathcal{G}/\cong$ of isotopy classes of connected, nondegenerate M–geometric complexes. In Section 3 we discuss random M–geometric complexes; in Section 4 we prove (1.2). Morevorer, in Section 5 we define and see some properties of geometric graphs; in Section 6 we recall the definition of the normalized Laplace operator $\hat{L}$ for graphs and we prove some properties of the spectral measure in the case of geometric graphs. Finally, in Section 7 we prove (1.4).

2. Geometric complexes

Throughout this paper we fix a Riemannian manifold (M, g) of dimension m.

Definition 2.1 (M–geometric complex and its skeleta)**.** Let $p_1, \ldots, p_n$ be points in M and fix $r \geq 0$. We define a M–*geometric complex* as

$$\mathcal{U}(\{p_1, \ldots, p_n\}, r) := \bigcup_{k=1}^{n} \hat{B}(p_k, r)$$

$$= \{x \in M : d_M(x, \{p_1, \ldots, p_n\}) \leq r\}.$$

For $\mathcal{U} := \mathcal{U}(\{p_1, \ldots, p_n\}, r)$, we also let

$$\check{C}(\mathcal{U}) := \check{C}(\{p_1, \ldots, p_n\}, r)$$
$$:= \text{ nerve of the cover } \{\hat{B}(p_k, r)\}_{k=1}^n$$

and we let

$$\check{C}^{(k)}(\mathcal{U}) := \check{C}^{(k)}(\{p_1, \ldots, p_n\}, r)$$
$$:= k - \text{skeleton of } \check{C}(\{p_1, \ldots, p_n\}, r).$$

In particular, we call $\check{C}^{(1)}(\{p_1, \ldots, p_n\}, r)$ a M–*geometric graph*.

Remark 2.2. In order to avoid unnecessary complications, in the sequel we will always assume that the injectivity radius[5] inj(M) of M is strictly positive (which is true if M is compact or if $M = \mathbb{R}^m$ with the flat metric) and that

$$0 < r \leq \text{inj}(M).$$

This requirement ensures that for every point $p \in M$ the set

$$\partial \hat{B}(p, r) = \{x \in M : d(x, p_k) = r\}$$

is smooth (in fact it is the image of the sphere of radius r in the tangent space at p under the exponential map, which is a diffeomorphism on $B_{T_pM}(0, \text{inj}(M))$). Observe also that for $r \leq \text{inj}(M)$ the ball $\hat{B}(p, r)$ is contractible, but not necessarily geodesically convex.

Definition 2.3. We say that $\mathcal{U}(\{p_1, \ldots, p_n\}, r)$ is *nondegenerate* if for each $J = \{j_1, \ldots, j_l\} \in \binom{n}{l}$ the intersection

$$\bigcap_{k=1}^l \hat{B}(p_{j_k}, r) \quad \text{is transversal.} \tag{2.1}$$

The next result is classical and relates the homotopy of a geometric complex to the one of its associated Čech complex.

Lemma 2.4 (Nerve Lemma)**.** *If M is compact, there exists $\rho > 0$ such that, for each $r \leq \rho$,*

$$\mathcal{U}(\{p_1, \ldots, p_n\}, r) \sim \check{C}(\{p_1, \ldots, p_n\}, r),$$

i.e. they are homotopy equivalent.

Proof. For the proof in this setting, see [ALL20, Lemma 6.1]. □

Definition 2.5 (Isotopy classes of connected geometric complexes)**.** Let $p_1, \ldots, p_n$ and $q_1, \ldots, q_n$ be points in M and let $r_0, r_1 \geq 0$ such that

$$\mathcal{U}_0 := \mathcal{U}(\{p_1, \ldots, p_n\}, r_0) \quad \text{and} \quad \mathcal{U}_1 := \mathcal{U}(\{q_1, \ldots, q_n\}, r_1)$$

---

[5]Recall that the injectivity radius inj$_p(M)$ of M at one point p is defined to be the largest radius of a ball in the tangent space T_pM on which the exponential map $\exp_p : T_pM \to M$ is a diffeomorphism and the injectivity radius of M is defined as the infimum of the injectivity radii at all points:

$$\text{inj}(M) = \inf_{p \in M} \text{inj}_p(M).$$

are nondegenerate M–geometric complexes. We say that $\mathcal{U}_0$ and $\mathcal{U}_1$ are *(rigidly) isotopic* and we write $\mathcal{U}_0 \cong \mathcal{U}_1$ if there exists an isotopy of diffeomorphisms $\varphi_t : M \to M$ with $\varphi_0 = \mathrm{id}_M$ and a continuous function $r(t) > 0$, for $t \in [0,1]$, such that:

- For each $t \in [0,1]$, $\mathcal{U}(\{\varphi_t(p_1), \ldots, \varphi_t(p_n)\}, r(t))$ is nondegenerate,
- $r(0) = r_0$,
- $r(1) = r_1$ and
- $\varphi_1(p_1) = q_1, \ldots, \varphi_1(p_n) = q_n$.

Remark 2.6 (Isotopy classes and discriminants). The definition of two complexes being (rigidly) isotopic is very reminiscent of the notion of rigid isotopy from algebraic geometry, where the "regular" deformations are those which do not intersect some discriminant. We can make this analogy more precise. For every $n \in \mathbb{N}$ consider the smooth manifold

$$H_n := \underbrace{M \times \cdots \times M}_{n \text{ many times}} \times (0, \mathrm{inj}(M)), \tag{2.2}$$

together with the *discriminant*

$$\Sigma_n := \{(p_1, \ldots, p_n, r) \mid \mathcal{U}(\{p_1, \ldots, p_n\}, r) \text{ is degenerate}\}. \tag{2.3}$$

The set Σ_n is closed since its complement R_n is defined by the (finitely many) transversality conditions (2.1). Adopting this point of view, isotopy classes of nondegenerate M–geometric complexes built using n many balls are labeled by the connected components of $R_n := H_n \backslash \Sigma_n$ (the complement of a discriminant). In fact, given a nondegenerate complex $\mathcal{U}(\{p_1, \ldots, p_n\}, r)$, then $(p_1, \ldots, p_n, r) \in R_n$ (because it is nondegenerate) and viceversa every point in R_n correspond to a nondegenerate complex. Moreover, a nondegenerate isotopy of two nondegenerate complexes defines a curve between the corresponding points of R_n, and this curve is entirely contained in R_n; the two corresponding parameters must therefore lie in the same connected component of R_n; viceversa, because for an open set of a manifold connected and path connected are equivalent, every two points in the same component R_n can be joined by an arc all contained in R_n and give therefore rise to isotopic complexes.

Definition 2.7. We define the set

$$\mathcal{G}(M) := \{\text{connected, nondegenerate } M\text{–geometric complexes}\}$$

and use the notation $\mathcal{G} := \mathcal{G}(M)$ when M is given. We also let

$$(\mathcal{G}/\cong) := \{\text{isotopy classes of connected, nondegenerate } M\text{–geometric complexes}\}.$$

Remark 2.8. Observe that, by definition, each class $[\mathcal{U}] \in \mathcal{G}/\cong$ keeps also the information on the way $\mathcal{U}$ is embedded in M. In particular, it might be that to two nondegenerate M–geometric complexes $\mathcal{U}_1$ and $\mathcal{U}_2$ there correspond isomorphic Čech complexes $\check{C}(\mathcal{U}_1) \simeq \check{C}(\mathcal{U}_2)$, but at the same time the complexes $\mathcal{U}_1$ and $\mathcal{U}_2$ themselves are not rigidly isotopic (see Figure 2).

Remark 2.9. For each M of dimension m, $\mathcal{G}(M)/\cong \; \supset \mathcal{G}(\mathbb{R}^m)/\cong$.

Theorem 2.10. $\mathcal{G}/\cong$ *is a countable set.*

Proof. We first partition $\mathcal{G}/\cong$ into countably many sets. For every $n \in \mathbb{N}$ we consider the set:

$$(\mathcal{G}_n/\cong) := \{\text{classes of complexes in } \mathcal{G}/\cong \text{ which are built using } n \text{ many balls}\},$$

and we need to prove that this set is countable. We have already seen (Remark 2.6) that isotopy classes of nondegenerate complexes which are built using n many balls are in one-to-one correspondence with the connected components of $R_n = H_n \backslash \Sigma_n$ (these sets are defined

by (2.2) and (2.3)). The function "number of connected components of a nondegenerate complex" is constant on each component of R_n and consequently the number of isotopy classes of *connected* and nondegenerate complexes (i.e. the cardinality of $\mathcal{G}_n/\cong$) is smaller than the number of components of R_n:

$$\#\big(\mathcal{G}_n/\cong\big) \leq \#\big(\{\text{connected components of } R_n\}\big).$$

We are therefore reduced to prove that R_n has at most countably many components. To this end we write R_n as the disjoint union of its components (each of which is an open set in H_n):

$$R_n = \bigsqcup_{\alpha \in A} C_\alpha. \tag{2.4}$$

We cover now the manifold H_n with countably many manifold charts $\{(V_j, \varphi_j)\}_{j\in\mathbb{N}}$ with $\varphi_j : V_j \xrightarrow{\sim} \mathbb{R}^m$. For every $j \in \mathbb{N}$ we consider also the decomposition of the open set $R_n \cap V_j$ into its connected components:

$$R_n \cap V_j = \bigsqcup_{\beta \in B_j} C_{\beta,j}.$$

Since each C_α from (2.4) is the union of elements of the form

$$C_\alpha = \bigcup_{j \in N_\alpha,\, \beta \in B_j} C_{\beta,j}$$

with the index set j running over the countable set $N_\alpha \subset \mathbb{N}$, it is therefore enough to prove that for every $j \in \mathbb{N}$ the index set B_j is countable, i.e. that the number of connected components of $R_n \cap V_j$ is countable. Observe now that, since φ_j is a diffeomorphism between V_j and $\mathbb{R}^m$, then the number of connected components of $R_n \cap V_j$ is the same of the number of connected components of $\varphi_j(R_n \cap V_j)$, which is an open subset of $\mathbb{R}^m$. Since in each component of $\varphi_j(R_n \cap V_j)$ we can pick a point with rational coordinates, it follows that the number of such components is countable, and this concludes the proof.

$\square$

Remark 2.11. Observe that the key point of the proof of Theorem 2.10 is showing that the number of connected components of an open set in a differentiable manifold is countable.

Definition 2.12. We let

$$(\mathcal{G}/\sim) := \{\text{homotopy classes of connected, nondegenerate } M\text{–geometric complexes}\}$$

and we let

$$\big(\mathcal{G}^{(k)}/\simeq\big) := \{\text{isomorphism classes of } k\text{–skeleta of connected } M\text{–geometric complexes}\}.$$

In particular,

$$\mathcal{G}^{(1)}/\simeq = \{\text{isomorphism classes of connected } M\text{–geometric graphs}\}.$$

Remark 2.13. In order to appreciate the difference between the classes in $(\mathcal{G}/\cong)$, $(\mathcal{G}/\sim)$ and $\big(\mathcal{G}^{(k)}/\simeq\big)$, we look at Figure 2. Here, we have three $\mathbb{R}^3$–geometric complexes, given by the union of balls in $\mathbb{R}^3$, that have three different shapes. All three complexes are homotopy equivalent to each other and they are all pairwise not isotopic. Moreover, while the first one and the second one have isomorphic 1–skeleta, the 1–skeleton of the third complex is not isomorphic to the other ones.

FIGURE 2. The three $\mathbb{R}^3$–geometric complexes described in Remark 2.13. From left to right, we call them $\mathcal{U}_1, \mathcal{U}_2$ and $\mathcal{U}_3$ and we assume the number of balls and their combinatorics needed to to describe the first two are the same. Then $[\mathcal{U}_1] = [\mathcal{U}_2] = [\mathcal{U}_3]$ in $\mathcal{G}/\sim$, because they are all homotopy equivalent, $[\mathcal{U}_1] = [\mathcal{U}_2] \neq [\mathcal{U}_3]$ in $\mathcal{G}/\simeq$ (the first two give rise to same Čech complexes, which is however different from the last one) and $[\mathcal{U}_1] \neq [\mathcal{U}_2] \neq [\mathcal{U}_3] \neq [\mathcal{U}_1]$ in $\mathcal{G}/\cong$ (they are all pairwise non-isotopic).

Remark 2.14. There are natural forgetful maps

$$\phi : \mathcal{G}/\cong \longrightarrow \mathcal{G}/\sim$$
$$[\mathcal{U}] \longmapsto [\mathcal{U}]$$

and

$$\varphi^{(k)} : \mathcal{G}/\cong \longrightarrow \mathcal{G}^{(k)}/\simeq$$
$$[\mathcal{U}] \longmapsto [\check{C}^{(k)}(\mathcal{U})].$$

Definition 2.15 (Component counting function)**.** Given a nondegenerate geometric complex $\mathcal{U} \subset M$, a topological subspace $Y \subset M$ and a class $w \in \mathcal{G}/\cong$, we define

$$\mathcal{N}(\mathcal{U}; w) := \#\big(\text{components of } \mathcal{U} \text{ of type } w\big);$$

$$\mathcal{N}(\mathcal{U}, Y; w) := \#\big(\text{components of } \mathcal{U} \text{ of type } w \text{ entirely contained in the interior of } Y\big);$$

$$\mathcal{N}^*(\mathcal{U}, Y; w) := \#\big(\text{components of } \mathcal{U} \text{ of type } w \text{ intersecting } Y\big).$$

In the paper [NS16] Nazarov and Sodin have introduced a powerful tool (the "integral geometry sandwitch") for localizing the count of the number of components of the zero set of random waves in a Riemannian manifold. This tool has been used by Sarnak and Wigman [SW19] for the study of distribution of components type of the zero set of random waves on a Riemannian manifold, and it has been adapted to geometric complexes in [ALL20]. We recall here this tool, stated in the language of this paper.

Theorem 2.16 (Analogue of the Integral Geometry Sandwich)**.** *The following two estimates are true:*

(1) (The local case) *Let $\mathcal{U}$ be a generic geometric complex in $\mathbb{R}^m$ and fix $w \in \mathcal{G}(\mathbb{R}^m)/\cong$. Then for $0 < r < R$*

$$\int_{B_{R-r}} \frac{\mathcal{N}(\mathcal{U}, B(x,r); w)}{\mathrm{vol}(B_r)} \, \mathrm{d}x \leq \mathcal{N}(\mathcal{U}, B_R; w) \leq \int_{B_{R+r}} \frac{\mathcal{N}^*(\mathcal{U}, B(x,r); w)}{\mathrm{vol}(B_r)} \, \mathrm{d}x.$$

(2) (The global case) *Let $\mathcal{U}$ be a generic geometric complex in a compact Riemannian manifold M and fix $w \in \mathcal{G}(M)/\cong$. Then for every $\varepsilon > 0$ there exists $\eta > 0$ such that for every $r < \eta$:*

$$(1-\varepsilon) \int_M \frac{\mathcal{N}(\mathcal{U}, B(x,r); w)}{\text{vol}(B_r)} \, dx \leq \mathcal{N}(\mathcal{U}, M; w) \leq (1+\varepsilon) \int_M \frac{\mathcal{N}^*(\mathcal{U}, B(x,r); w)}{\text{vol}(B_r)} \, dx.$$

Proof. The proof of both statements is exactly the same as in [SW19], after noticing that the only property needed on the counting functions is that the considered topological spaces only have finitely many components, and these components are counted according to a specific type (here are selected according to isotopy type, but we could consider instead any function that partitions the set of components and count only the components belonging to a given class). $\square$

3. Random geometric complexes (thermodynamic regime)

Definition 3.1 (Riemannian case)**.** Let M be a compact Riemannian manifold and consider a set of points $\{p_1, \ldots, p_n\}$ independently sampled from the uniform distribution on M. Fix a positive number $\alpha > 0$, let $r := \alpha n^{-1/m}$ and let

$$\mathcal{U}_n := \mathcal{U}(\{p_1, \ldots, p_n\}, r).$$

We say that $\mathcal{U}_n$ is a *random M–geometric complex*. The choice of such r is what defines the so-called *critical* or *thermodynamic regime*.

Definition 3.2 (Euclidean Poisson case)**.** Let $P := \{p_1, p_2, \ldots\}$ be a set of points sampled from the standard spatial Poisson distribution in $\mathbb{R}^m$. For $\alpha > 0$, let

$$\mathcal{P} := \bigcup_{p \in P} B(p, \alpha)$$

and, for $R > 0$, let

$$\mathcal{P}_R := \{\text{connected components of } \mathcal{P} \text{ entirely contained in the interior of } B(0, R)\}.$$

Remark 3.3. With probability 1, we have that $\#\big(P \cap B(0, R)\big)$ is finite. To see this, observe that

$$\mathbb{P}\{\#\big(P \cap B(0, R)\big) \geq l\} = \sum_{k \geq l} \frac{\text{vol}(B(0, R))}{k!} e^{-\text{vol}(B(0,R))}$$

and

$$\mathbb{P}\{\#\big(P \cap B(0, R)\big) = \infty\} \leq \mathbb{P}\{\#\big(P \cap B(0, R)\big) \geq l\} \xrightarrow{l \to \infty} 0.$$

From now on, we will only consider nondegenerate complexes, without further mentioning this assumption. This is not reductive, since our random complexes are nondegenerate with probability one.

4. Random measures

We fix the following notation. Given a set A with a fixed sigma algebra (omitted from the notation), we denote by:

$$\mathcal{M}(A) := \{\text{measures on } A\} \quad \text{and} \quad \mathcal{M}^1(A) := \{\text{probability measures on } A\}.$$

Definition 4.1. Let $\mathcal{U} \subset M$ be a finite geometric complex and let

$$\mathcal{U} = \mathcal{U}^1 \sqcup \ldots \sqcup \mathcal{U}^{b_0(\mathcal{U})}$$

be its decomposition into connected components. We define $\Theta_{\mathcal{U}} \in \mathcal{M}^1(\mathcal{G}/\cong)$ as

$$\Theta_{\mathcal{U}} := \frac{1}{b_0(\mathcal{U})} \sum_{j=1}^{b_0(\mathcal{U})} \delta_{[\mathcal{U}^j]}.$$

Observe that the measure $\Theta_{\mathcal{U}}$ just defined is a probability measure. We also endow $\mathcal{M}^1(\mathcal{G}/\cong)$ with the *total variation distance*:

$$d_{\mathrm{tv}}(\Theta_1, \Theta_2) := \sup_{A \subset \mathcal{G}/\cong} |\Theta_1(A) - \Theta_2(A)|.$$

When $\mathcal{U}$ is a random geometric complex, $\Theta_{\mathcal{U}}$ is a random variable with values in the metric space $(\mathcal{M}^1(\mathcal{G}/\cong), d_{\mathrm{tv}})$. In this context, recall the notion of *convergence in probability*:

$$\Theta_n \xrightarrow{\mathbb{P}} \Theta \iff \forall \varepsilon, \lim_{n \to \infty} \mathbb{P}(d_{\mathrm{tv}}(\Theta_n, \Theta) \geq \varepsilon) = 0.$$

Using the previous notation, we set

$$\Theta_n := \Theta_{\mathcal{U}_n} \qquad \text{and} \qquad \Theta_R := \Theta_{\mathcal{P}_R}.$$

Theorem 4.2. *There exists a probability measure* $\Theta \in \mathcal{M}^1(\mathcal{G}/\cong)$ *such that*

(1) $\Theta_n \xrightarrow[n \to \infty]{\mathbb{P}} \Theta \qquad and \qquad \Theta_R \xrightarrow[R \to \infty]{\mathbb{P}} \Theta$

(2) $\mathrm{supp}(\Theta) = \mathcal{G}(\mathbb{R}^m)/\cong.$

We shall see the proof of Theorem 4.2 in Section 4.1. As a first corollary we recover the results from [ALL20].

Corollary 4.3 (Theorem 1.1 and Theorem 1.3 from [ALL20])**.** *Consider the forgetful map*

$$\phi : \mathcal{G}/\cong \longrightarrow \mathcal{G}/\sim$$
$$[\mathcal{U}] \longmapsto [\mathcal{U}].$$

We have that

$$\phi_* \Theta_n \xrightarrow{\mathbb{P}} \phi_* \Theta \qquad and \qquad \phi_* \Theta_R \xrightarrow{\mathbb{P}} \phi_* \Theta.$$

Also, $\mathrm{supp}\, \phi_* \Theta = \mathcal{G}(\mathbb{R}^m)/\sim.$

As a second corollary we see that, because Theorem 4.2 keeps track of fine properties of the geometric complex, we can use other forgetful maps and obtain information on the limit distribution of the components type of each k–skeleton.

Corollary 4.4. *For each* $k \in \mathbb{N}$, *consider the forgetful map from Remark 2.14:*

$$\varphi^{(k)} : \mathcal{G}/\cong \longrightarrow \mathcal{G}^{(k)}/\simeq$$
$$[\mathcal{U}] \longmapsto [\check{C}^{(k)}(\mathcal{U})].$$

We have that

$$\varphi_*^{(k)} \Theta_n \xrightarrow{\mathbb{P}} \varphi_*^{(k)} \Theta \qquad and \qquad \varphi_*^{(k)} \Theta_R \xrightarrow{\mathbb{P}} \varphi_*^{(k)} \Theta.$$

Also, $\mathrm{supp}\, \varphi_*^{(k)} \Theta = \mathcal{G}^{(k)}(\mathbb{R}^m)/\simeq.$

Proposition 4.5 (Existence of all isotopy types)**.** *Let $\mathcal{U}$ be a nondegenerate geometric complex in $\mathbb{R}^m$ and let $\alpha > 0$. Let $(\mathcal{U}_n)_n$ be a sequence of random M–geometric complexes constructed using α. There exist n_0, R, $c > 0$ (depending on Γ and α but not on M) such that for every $p \in M$ and for every $n \geq n_0$:*

$$\mathbb{P}\{\mathcal{U}_n \cap \hat{B}(p, Rn^{-1/m}) \cong \mathcal{U}\} > c.$$

Proof. The proof is similar to the proof of [ALL20, Proposition 1.2]. Here we sketch this proof for the sake of completeness, pointing out what is the main difference with [ALL20].

Assume that $\mathcal{U} \subset B(0, R')$ is constructed using balls of radius $r = 1$:

$$\mathcal{U} = \bigcap_{k=1}^{\ell} B(y_k, 1),$$

set $R = \alpha R'$ and consider the sequence of maps:

$$\psi_n : \hat{B}(p, Rn^{-1/m}) \xrightarrow{\exp_p^{-1}} B_{T_pM}(0, Rn^{-1/m}) \xrightarrow{\text{dilation}} B_{T_pM}(0, R') \simeq B(0, R')$$

For $n > $ large enough the map ψ_n becomes a diffeomorphism and we denote by φ_n its inverse. [ALL20, Proposition 6.2] implies that there exist $\epsilon_0 > 0$ and $n_0 > 0$ such that if $\|\tilde{y}_k - y_k\| \leq \epsilon_0$ for every $k = 1, \ldots, \ell$, then for $n \geq n_0$ the two complexes $\bigcup_{k=1}^{\ell} \hat{B}(\varphi_n(\tilde{y}_k), \alpha n^{-1/m})$ and $\mathcal{U}$ are isomorphic. In fact, because $\mathcal{U}$ is nondegenerate, possibly choosing $\epsilon_0 > 0$ even smaller, we can make sure that the complex $\bigcup_{k=1}^{\ell} \hat{B}(\varphi_n(\tilde{y}_k), \alpha n^{-1/m})$ belong to the same rigid isotopy class of $\mathcal{U}$, because this is an open condition.

One then proceeds considering the event:

$$E_n = \left\{ \exists I_\ell \in \binom{n}{\ell} \; : \; \forall j \in I_\ell \quad p_j \in \psi_n^{-1}(B(y_j, \epsilon)), \text{ and } \quad \forall j \notin I_\ell \quad p_j \in \hat{B}(p, (R+\alpha)n^{-1/m})^c) \right\}.$$

Observe that:

$$E_n \implies \mathcal{U}_n \cap \hat{B}(p, Rn^{-1/m}) \cong \mathcal{U},$$

and in particular, in order to get the conclusion, it is enough to estimate from below the probability of E_n. This is done in the last lines of the proof of [ALL20, Proposition 1.2].

$\square$

Example 4.6. Let C_{137} be the cycle with 137 vertices and 137 edges. By Proposition 4.5 there exist n_0, R and $c > 0$ (depending on C_{137} and α but not on M) such that for every $p \in M$ and for every $n \geq n_0$:

$$\mathbb{P}\{\Gamma_n \cap \hat{B}(p, Rn^{-1/m}) \cong C_{137}\} > c.$$

4.1. **Proof of Theorem 4.2.** We split the statement of Theorem 4.2 into two parts. The first part, Theorem 4.8, states that the random measure Θ_R converges in probability to a deterministic measure $\Theta \in \mathcal{M}^1(\mathcal{G}/\cong)$ supported on the set $\mathcal{G}(\mathbb{R}^m)/\cong$. The second part, Theorem 4.12, states that also the random measure Θ_n converges in probability to Θ.

4.1.1. *The local model.*

Proposition 4.7. *For every $w \in \mathcal{G}(\mathbb{R}^m)/\cong$ there exists a constant $c_w > 0$ such that the random variable*

$$c_{R,w} := \frac{\mathcal{N}(\mathcal{P}_R, w)}{\mathrm{vol}(B(0,R))}$$

converges to c_w in L^1 and almost surely as $R \to \infty$.

Proof of Proposition 4.7. Following the proof of [ALL20, Proposition 2.1] with $\mathcal{N}(\mathcal{P}_R, \tau, w)$ instead of $\mathcal{N}(\mathcal{P}_R, \tau, \gamma)$ and with the application of Theorem 2.16 instead of [ALL20, Theorem 6.6], one proves that there exists a constant c_w such that

$$\frac{\mathcal{N}(\mathcal{P}_R, B(0,R); w)}{\mathrm{vol}(B(0,R))} \longrightarrow c_w.$$

Since $\mathcal{P}_R \subset B(0,R)$, this implies that

$$\frac{\mathcal{N}(\mathcal{P}_R, w)}{\mathrm{vol}(B(0,R))} \longrightarrow c_w.$$

We now have to prove that $c_w > 0$. Since $w \in \mathcal{G}(\mathbb{R}^m)/\cong$, given $\mathcal{U} \in w$ there exist $\beta > 0$ and $y_1, \ldots, y_n \in \mathbb{R}^m$ such that

$$\mathcal{U} \cong \mathcal{U}(\{y_1, \ldots, y_n\}, \beta).$$

Let R_1 be such that $\mathcal{U} \subset B(0, R_1)$ and choose r such that $r\beta = \alpha$, where α is the constant that we used for constructing the random complex $\mathcal{P}$. We can then rescale $\mathcal{U}$ in $B(0, r \cdot R_1)$ so that it is constructed on radius α. Now, since $\mathcal{U}$ is nondegenerate, there exists $\varepsilon > 0$ such that, if for every i we have that $\|y_i - y_i'\| < \varepsilon$, then the complex

$$\mathcal{U}'(\{y_1', \ldots, y_n'\}, \alpha)$$

is isotopic to $\mathcal{U}$. Take $K_R = k \cdot \mathrm{vol}(B(0,R))$ disjoint balls $\{B(y_j, r \cdot R_1)\}_{j=1,\ldots,K_R}$ inside $B(0,R)$ with $c > 0$. Then

$$\frac{\mathcal{N}(\mathcal{P}_R, B(0,R); w)}{\mathrm{vol}(B(0,R))} \geq \frac{1}{\mathrm{vol}(B(0,R))} \cdot \sum_{j=1}^{K_R} \mathcal{N}(\mathcal{P}_R, B(y_j, r \cdot R_1); w).$$

Therefore

$$\mathbb{E}\left(\frac{\mathcal{N}(\mathcal{P}_R, B(0,R); w)}{\mathrm{vol}(B(0,R))}\right) \geq \mathbb{E}\left(\frac{1}{\mathrm{vol}(B(0,R))} \cdot \sum_{j=1}^{K_R} \mathcal{N}(\mathcal{P}_R, B(y_j, r \cdot R_1); w)\right)$$

$$= \frac{K_R}{\mathrm{vol}(B(0,R))} \cdot \mathbb{E}\left(\mathcal{N}(\mathcal{P}_R, B(y, r \cdot R_1); w)\right),$$

since the random variables $\mathcal{N}(\mathcal{P}_R, B(y_j, r \cdot R_1); w)$ are identically distributed by the fact that the balls are disjoint. Now,

$$\frac{K_R}{\mathrm{vol}(B(0,R))} \geq \frac{k \cdot \mathrm{vol}(B(0,R))}{\mathrm{vol}(B(0,R))} = k > 0$$

and

$$\mathbb{E}\left(\mathcal{N}(\mathcal{P}_R, B(y, r \cdot R_1); w)\right) > 0.$$

Therefore, $c_w > 0$. $\qquad\square$

Proposition 4.7 allows us to deduce the following theorem.

Theorem 4.8. *There exists $\Theta \in \mathcal{M}^1(\mathcal{G}/\cong)$ such that*

$$\Theta_R \xrightarrow[R \to \infty]{\mathbb{P}} \Theta.$$

Also, $\operatorname{supp}(\Theta) = \mathcal{G}(\mathbb{R}^m)/\cong.$

Proof. The proof is the same as the one of [ALL20, Theorem 1.3], replacing the homotopy type counting function with the isotopy type one. □

4.1.2. *Riemannian case.*

Theorem 4.9. *Let $p \in M$. For every $\delta > 0$ and for $R > 0$ sufficiently big there exists n_0 such that for every $w \in \mathcal{G}(\mathbb{R}^m)/\cong$ and for $n \geq n_0$:*

$$\mathbb{P}\{\mathcal{N}(\mathcal{P}_R, B(0, R); w) = \mathcal{N}(\mathcal{U}_n, \hat{B}(p, Rn^{-1/m}); w)\} \geq 1 - \delta.$$

Proof. The proof is the same as the one of [ALL20, Theorem 3.1], with the following differences:

- In point (3), instead of considering the homotopy equivalence between the unions of the balls, we consider the isotopy equivalences between their k–skeleta. This is allowed because, at the end of the proof of point (3), it is proved that the combinatorics of the covers are the same.
- After assuming point (1), point (2) and the modified point (3), we can say that the k–skeleta of the two unions of balls are isotopic and also the unions of all the components entirely contained in $B(0, R)$ (respectively $\hat{B}(p, Rn^{-1/m})$) have isotopic k–skeleta. In particular, the number of components of a given isotopy class w is the same for both sets with probability at least $1 - \delta$.

□

Corollary 4.10. *For each $w \in \mathcal{G}(\mathbb{R}^m)/\cong$, $\alpha > 0$, $x \in M$ and $\varepsilon > 0$, we have*

$$\lim_{R \to \infty} \limsup_{n \to \infty} \mathbb{P}\left\{ \left| \frac{\mathcal{N}(\mathcal{U}_n, B(x, Rn^{-1/d}); w)}{\operatorname{vol}(B(0, R))} - c_w \right| > \varepsilon \right\} = 0.$$

Proof. The proof is the same as the one of [ALL20, Corollary 3.2], with the application of Theorem 4.9 and Proposition 4.7 instead of [ALL20, Theorem 3.1] and [ALL20, Proposition 2.1]. □

Theorem 4.11. *For every $w \in \mathcal{G}(\mathbb{R}^m)/\cong$, the random variable*

$$c_{n,w} := \frac{\mathcal{N}(\mathcal{U}_n; w)}{n}$$

converges in L^1 to $c_w \cdot \operatorname{vol}(M)$, where c_w is the constant appearing in Proposition 4.7. In particular, this implies that the random variable

$$c_n := \frac{\mathcal{N}(\mathcal{U}_n)}{n},$$

i.e. when we consider all components with no restriction on their type, converges in L^1 to c, where $c = \sum_{w \in \mathcal{G}(\mathbb{R}^m)/\cong} c_w > 0$.

Proof. The proof is the same as the one of [ALL20, Theorem 4.1], with the application of Theorem 2.16 instead of [ALL20, Theorem 6.7] and the application of Corollary 4.10 instead of [ALL20, Corollary 3.2]. □

Theorem 4.12. *The measure* $\Theta \in \mathcal{M}^1(\mathcal{G}/\cong)$ *appearing in Theorem 4.8 is such that*

$$\Theta_n \xrightarrow[n\to\infty]{\mathbb{P}} \Theta.$$

Proof. The proof is the same as the one of [ALL20, Theorem 1.1], with the following differences:

  – We apply Theorem 4.11 instead of [ALL20, Theorem 4.1];
  – We use $w \in \mathcal{G}(\mathbb{R}^m)/\cong$ instead of $\gamma \in \mathcal{G}$, Θ_n instead of $\hat{\mu}_n$ and $\mathcal{G}(M)/\cong$ instead of $\hat{\mathcal{G}}$.

□

5. Geometric graphs

We specialize the previous discussion to the case $k = 1$, and consider

$$\mathcal{G}^{(1)} = \{\text{isomorphism classes of connected, nondegenerate } M\text{–geometric graphs}\}.$$

Remark 5.1. The set of M–geometric graphs defined using closed balls equals the set of M–geometric graphs defined using open balls. To see this, assume that a geometric graph $\Gamma = (V(\Gamma), E(\Gamma))$ is defined using closed balls of radius r. Then, for each pair of distinct vertices (p_i, p_j),

$$(p_i, p_j) \in E(\Gamma) \iff d(p_i, p_j) \leq r.$$

Now, choose $\varepsilon \geq 0$ small enough that, for each pair of distinct vertices (p_i, p_j),

$$(p_i, p_j) \in E(\Gamma) \iff d(p_i, p_j) < r + \varepsilon.$$

Therefore, Γ can be constructed as a M–geometric graph using open balls of radius $r + \varepsilon$. The inverse implication is analogous.

In the case $M = \mathbb{R}^m$, the problem of describing the set $\mathcal{G}^{(1)}(\mathbb{R}^m)$ is equivalent to asking which graphs are realizable as $\mathbb{R}^m$–geometric graphs in a given dimension m. There is a vast literature about this problem and, commonly, geometric graphs realizable in dimension m are called m–*sphere graphs* while the minimal dimension m such that a given graph is a m–sphere graph is called its *sphericity*. In [Mae84] it is proved that every graph has finite sphericity; in [KM12] the authors prove that the problem of deciding, given a graph Γ, whether Γ is a m–sphere is NP-hard for all $m > 1$. We can also observe that, for each $m > 0$, there are graphs that are not m–sphere graphs. To see this, consider the *kissing number* $k(m)$ in dimension m, defined as the number of non-overlapping unit spheres that can be arranged such that they each touch a common unit sphere. Consider the star graph with a central vertex connected to n external vertices, where $n > k(m)$. In order to have a realization of dimension m of this graph, we need a central sphere that touches n spheres which do not touch each other. Since $n > k(m)$, this is not possible.

Example 5.2. In dimension 2, the kissing number is 6, as shown in Figure 3. Therefore, any star graph S_n on $n+1$ vertices, with $n > 6$, is not realizable in dimension 2 as a sphere graph.
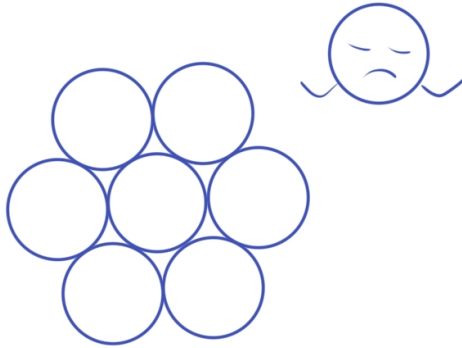
FIGURE 3. 7 kissing spheres in dimension 2. The eighth sphere doesn't know where to go.
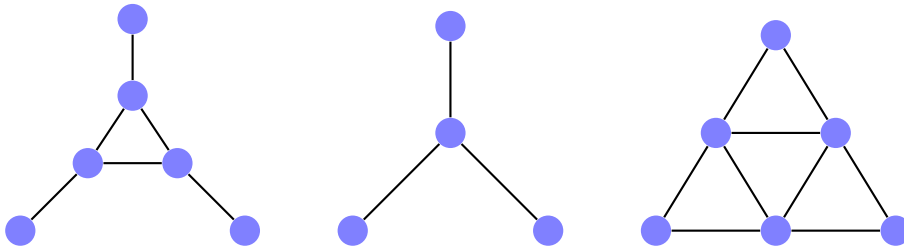

FIGURE 4. Together with the cycles of length at least four, these are the non-admissible induced subgraphs for unit interval graphs on the line.

In the particular case of $m = 1$, 1–sphere graphs are called *indifference graphs*, *unit interval graphs* and there are many characterizations of such graphs [LB63, Rob69, Jac92, GOn96, Weg04, Mer08]. A classical characterization is due to Roberts and Wegner [Rob69, Weg04, Jac92] and it characterizes unit interval graphs by the absence of certain forbidden subgraphs, this is recalled in Theorem 5.3.

Theorem 5.3 (Roberts and Wegner). *A graph is a unit interval graph, i.e. it's an element of $\mathcal{G}^{(1)}(\mathbb{R})$, if and only if it does not contain any cycle of length at least four and any of the graphs shown in Figure 4 as induced subgraph.*

As a consequence, we get the following corollary.

Corollary 5.4. *The support of the measure $\varphi_*^{(1)}\Theta$ defined in Corollary 4.4 is given by all graphs that do not contain any cycle of length at least four and any of the graphs shown in Figure 4 as induced subgraph.*

6. NORMALIZED LAPLACIAN OF A GRAPH AND ITS SPECTRUM

We now fix a graph Γ on n vertices $v_1, \ldots, v_n$ and we recall the definition of the (symmetric) normalized Laplace matrix, together with other common matrices associated to graphs. We shall then define the *spectrum* and the *spectral measure* associated to these matrices, and show some properties.

Definition 6.1 (Matrices associated to a graph)**.** Let A be the *adjacency matrix* of Γ; let $D := \mathrm{diag}(\deg v_1, \ldots, \deg v_n)$ be the *degree matrix*; let $L := D - A$ be the *non-normalized Laplacian matrix* and let $\hat{L} := I_n - D^{-1/2}AD^{-1/2}$ be the *symmetric normalized Laplacian matrix*.

Definition 6.2 (Spectrum of a matrix)**.** Given $Q \in \mathrm{Sym}(n, \mathbb{R})$ let $\mathrm{spec}(Q)$ be the spectrum of Q, i.e. the collection of its eigenvalues repeated with multiplicity,

$$\lambda_1(Q) \leq \ldots \leq \lambda_n(Q).$$

We define the *empirical spectral measure of Q* as

$$\mu_Q := \frac{1}{n}\sum_{i=1}^{n} \delta_{\lambda_i(Q)}.$$

Definition 6.3 (Spectrum of a graph)**.** We define *spectrum* of Γ, $\mathrm{spec}(\Gamma)$, as the spectrum of $\hat{L}$ and we write it as

$$\lambda_1(\Gamma) \leq \ldots \leq \lambda_n(\Gamma).$$

We also define

$$s(\Gamma) := \sum_{i=1}^{n} \delta_{\lambda_i(\Gamma)}$$

and the *spectral measure of Γ* as

$$\mu_\Gamma := \mu_{\hat{L}} = \frac{1}{n}\sum_{i=1}^{n} \delta_{\lambda_i(\Gamma)}.$$

Recall that, for every $i = 1, \ldots, n$, $\lambda_i(\Gamma) \in [0, 2]$ [Chu97, Equation (1.1) and Lemma 1.7]. In particular, this implies that $s(\Gamma) \in \mathcal{M}([0, 2])$ and $\mu_\Gamma \in \mathcal{M}^1([0, 2])$.

Theorem 6.4. *Let $(\Gamma_{1,n})_n$ and $(\Gamma_{2,n})_n$ be two sequences of graphs such that, for every n, $(\Gamma_{1,n})_n$ and $(\Gamma_{2,n})_n$ are two graphs on n nodes that differ at most by c edges. Denote by $\mu_{1,n}$ and $\mu_{2,n}$ the spectral measures associated to one of the matrices A, D, L, $\hat{L}$. Then*

$$\mu_{1,n} - \mu_{2,n} \overset{*}{\rightharpoonup} 0,$$

where $\overset{*}{\rightharpoonup}$ denotes the weak star convergence, i.e. for each $f \in C_c^0(\mathbb{R}, \mathbb{R})$

$$\left| \int_\mathbb{R} f \, \mathrm{d}\mu_{1,n} - \int_\mathbb{R} f \, \mathrm{d}\mu_{2,n} \right| \to 0.$$

We shall prove Theorem 6.4 in Section 6.1.

Remark 6.5. In the case of $\hat{L}$, we have convergence in total variation distance for "connected sum" of complete graphs, but not for paths, as we shall see in Section 6.2.

Remark 6.6. Theorem 2.8 from [GJLS16] tells that, if two families $(\Gamma_{1,n})_n$ and $(\Gamma_{2,n})_n$ differ by at most c edges *and* their corresponding spectral measures have weak limits, then they belong to the same spectral class (the notion of spectral class of a family of graph is recalled in Remark 1.3). In this sense our previous Theorem 6.4 can be considered as an analogue of [GJLS16, Theorem 2.8]: the difference of the spectral measure of two families of graphs $(\Gamma_{1,n})_n$ and $(\Gamma_{2,n})_n$ differing by at most a finite number c of edges, goes to zero weakly (without the assumption that the corresponding spectral measures have weak limits).

6.1. **Proof of Theorem 6.4.**

6.1.1. *Preliminaries.* Given $Q \in \mathrm{Sym}(n, \mathbb{R})$, we define the *1–Shatten norm of Q* as

$$\|Q\|_{S^1} := \sum_{i=1}^{n} |\lambda_i(Q)|.$$

The *Weilandt-Hoffman inequality* [Tao12, Exercise 1.3.6] holds:

$$\sum_{i=1}^{n} |\lambda_i(Q_1) - \lambda_i(Q_2)| \leq \|Q_1 - Q_2\|_{S^1}.$$

Proposition 6.7. *Let $Q_1, Q_2 \in \mathrm{Sym}(n, \mathbb{R})$ such that*

$$\|Q_1 - Q_2\|_{S^1} \leq C.$$

Then, for each $f \in C_c^0(\mathbb{R}, \mathbb{R})$ and for each $\varepsilon > 0$, there exists $\delta > 0$ such that

$$\left| \int_{\mathbb{R}} f \, \mathrm{d}\mu_{Q_1} - \int_{\mathbb{R}} f \, \mathrm{d}\mu_{Q_2} \right| \leq \varepsilon + \frac{2 \sup |f|}{\delta n} \cdot C.$$

Proof. Denote by $\{\lambda_i^{(1)}\}_{i=1}^{n}$ and $\{\lambda_i^{(2)}\}_{i=1}^{n}$ the eigenvalues of Q_1 and Q_2 respectively. Then

$$\int_{\mathbb{R}} f \, \mathrm{d}\mu_{Q_1} - \int_{\mathbb{R}} f \, \mathrm{d}\mu_{Q_2} = \frac{1}{n} \sum_{i=1}^{n} f(\lambda_i^{(1)}) - f(\lambda_i^{(2)}),$$

therefore

$$\left| \int_{\mathbb{R}} f \, \mathrm{d}\mu_{Q_1} - \int_{\mathbb{R}} f \, \mathrm{d}\mu_{Q_2} \right| \leq \frac{1}{n} \sum_{i=1}^{n} |f(\lambda_i^{(1)}) - f(\lambda_i^{(2)})|.$$

Now, since $f \in C_c^0(\mathbb{R}, \mathbb{R})$, f is uniformly continuous and given $\varepsilon > 0$ there exists $\delta = \delta(f)$ such that

$$|\lambda_1 - \lambda_2| \leq \delta \qquad \Longrightarrow \qquad |f(\lambda_1) - f(\lambda_2)| \leq \varepsilon.$$

Therefore, since by Equation 6.1.1 and by hypothesis we have that

$$\sum_{i=1}^{n} |\lambda_i^{(1)} - \lambda_i^{(2)}| \leq \|Q_1 - Q_2\|_{S^1} \leq C,$$

it follows that

$$|\{|\lambda_i^{(1)} - \lambda_i^{(2)}| > \delta\}| \leq \frac{C}{\delta}.$$

Therefore,

$$
\begin{aligned}
\left| \int_{\mathbb{R}} f \, \mathrm{d}\mu_{Q_1} - \int_{\mathbb{R}} f \, \mathrm{d}\mu_{Q_2} \right| &\leq \frac{1}{n} \sum_{i=1}^{n} |f(\lambda_i^{(1)}) - f(\lambda_i^{(2)})| \\
&= \frac{1}{n} \sum_{|\lambda_i^{(1)} - \lambda_i^{(2)}| < \delta} |f(\lambda_i^{(1)}) - f(\lambda_i^{(2)})| + \frac{1}{n} \sum_{|\lambda_i^{(1)} - \lambda_i^{(2)}| \geq \delta} |f(\lambda_i^{(1)}) - f(\lambda_i^{(2)})| \\
&\leq \frac{1}{n} \sum_{|\lambda_i^{(1)} - \lambda_i^{(2)}| < \delta} \varepsilon + \frac{1}{n} \sum_{|\lambda_i^{(1)} - \lambda_i^{(2)}| \geq \delta} |f(\lambda_i^{(1)})| + |f(\lambda_i^{(2)})| \\
&\leq \frac{1}{n} \cdot \varepsilon \cdot |\{|\lambda_i^{(1)} - \lambda_i^{(2)}| < \delta\}| + \frac{1}{n} \cdot 2 \sup |f| \cdot |\{|\lambda_i^{(1)} - \lambda_i^{(2)}| \geq \delta\}| \\
&\leq \frac{1}{n} \cdot \varepsilon \cdot n + \frac{1}{n} \cdot \frac{2 \sup |f|}{\delta} \cdot C
\end{aligned}
$$

$$\leq \varepsilon + \frac{2 \sup |f|}{\delta n} \cdot C.$$

$\square$

6.1.2. *Applications to graphs.*

Lemma 6.8. *Let* Γ_1, Γ_2 *be two graphs with* $V(\Gamma_1) = V(\Gamma_2)$ *that differ by at most* C*–many edges. Then,*

$$\|A_1 - A_2\|_{S^1} \leq 4C$$
$$\|D_1 - D_2\|_{S^1} \leq 4C^2$$
$$\|L_1 - L_2\|_{S^1} \leq 4C^2$$
$$\|\hat{L}_1 - \hat{L}_2\|_{S^1} \leq 2C \cdot \sqrt{2} \cdot \sqrt{n-1}.$$

Proof. Observe that any of the matrices

$$\Delta_1 := A_1 - A_2, \qquad \Delta_2 := D_1 - D_2, \qquad \Delta_3 := L_1 - L_2$$

consists of all zeros except for at most $4C$ entries, all of which entries are bounded by a constant (it is 1 for Δ_1 and C for Δ_2 and Δ_3). Therefore, each $\Delta_i \in \mathrm{Sym}(n, \mathbb{R})$ for $i = 1, 2, 3$ has rank at most $4C$ and all its eigenvalues are zero, except for at most $4C$ of them. It follows that, for $i = 1, 2, 3$,

$$
\begin{aligned}
\|\Delta_i\|_{S^1} &= \sum_{j=n-2C+1}^{n} |\lambda_j(\Delta_i)| \\
&= \langle (|\lambda_{n-2C+1}|, \ldots, |\lambda_n|), (1, \ldots, 1) \rangle \\
&\leq \sum_{j=n-2C+1}^{n} \left(\lambda_j(\Delta_i)^2 \right)^{1/2} \sqrt{4C} \\
&= 2C^{1/2} \|\Delta_i\|_F \\
&= 2C^{1/2} \Big(\sum_{j,k} (\Delta_i)_{jk}^2 \Big)^{1/2} \\
&= 2C^{1/2} (4C\tilde{C}^2)^{1/2} \\
&\leq 4C\tilde{C}_i,
\end{aligned}
$$

where

$$
\tilde{C}_i = \begin{cases} 1 & \text{if } i = 1 \\ C & \text{if } i = 2, 3. \end{cases}
$$

Similarly, $\Delta_4 := \hat{L}_1 - \hat{L}_2$ consists of all zeros except for at most $2C(n-1)$ entries, all of which are bounded by 1, and it has rank at most $4C$. Therefore,

$$
\begin{aligned}
\|\Delta_4\|_{S^1} &= \sum_{j=1}^{n} |\lambda_j(\Delta_4)| \\
&\leq \sum_{j=1}^{n} \left(\lambda_j(\Delta_4)^2 \right)^{1/2} \cdot (4C)^{1/2}
\end{aligned}
$$

$$= \Big(\sum_{j,k}(\Delta_4)_{jk}^2\Big)^{1/2} \cdot 2C^{1/2}$$

$$\leq \Big(\sum_{1}^{2C(n-1)} 1\Big)^{1/2} \cdot 2C^{1/2}$$

$$= 2C \cdot \sqrt{2} \cdot \sqrt{n-1}.$$

$\square$

As a corollary, we can prove Theorem 6.4.

Proof of Theorem 6.4. We prove that, for each $f \in C_c^0(\mathbb{R}, \mathbb{R})$ and for each $\varepsilon > 0$,

$$\lim_{n\to\infty} \Big| \int_{\mathbb{R}} f \, \mathrm{d}\mu_{1,n} - \int_{\mathbb{R}} f \, \mathrm{d}\mu_{2,n} \Big| \leq \varepsilon.$$

Let $c_1 := 4C$, $c_2 := c_3 := 4C^2$ and $c_4 := 2C \cdot \sqrt{2}$. By Lemma 6.8, we have that

$$\|\Delta_i\|_{S^1} \leq c_i$$

for each $i = 1, 2, 3$. By Proposition 6.7, there exists $\delta > 0$ such that

$$\Big| \int_{\mathbb{R}} f \, \mathrm{d}\mu_{1,n} - \int_{\mathbb{R}} f \, \mathrm{d}\mu_{2,n} \Big| \leq \varepsilon + \frac{2 \sup |f|}{\delta n} \cdot c_i.$$

Therefore,

$$\lim_{n\to\infty} \Big| \int_{\mathbb{R}} f \, \mathrm{d}\mu_{1,n} - \int_{\mathbb{R}} f \, \mathrm{d}\mu_{2,n} \Big| \leq \varepsilon.$$

Similarly, by Lemma 6.8 we have that

$$\|\Delta_4\|_{S^1} \leq c_4 \sqrt{n-1}.$$

By Proposition 6.7, there exists $\delta > 0$ such that

$$\Big| \int_{\mathbb{R}} f \, \mathrm{d}\mu_{1,n} - \int_{\mathbb{R}} f \, \mathrm{d}\mu_{2,n} \Big| \leq \varepsilon + \frac{2 \sup |f|}{\delta n} \cdot c_4 \sqrt{n-1}.$$

Therefore,

$$\lim_{n\to\infty} \Big| \int_{\mathbb{R}} f \, \mathrm{d}\mu_{1,n} - \int_{\mathbb{R}} f \, \mathrm{d}\mu_{2,n} \Big| \leq \varepsilon.$$

$\square$

6.2. **Strong convergence for complete graphs.**

Lemma 6.9. *Given $N \in \mathbb{N}^+$, let K_N and K_N' be two complete graphs on N nodes. Let $K_N \sqcup K_N'$ be their disjoint union and let $\Gamma_N := K_N \bigcup_{c \text{ edges}} K_N'$ be their union together with c edges (v_i, v_i') where $v_i \in K_N$ and $v_i' \in K_N'$, for $i = 1, \ldots, c$. Let also $\mu_{K_N \sqcup K_N'}$ and μ_{Γ_N} be the spectral measures of these two graphs. Then,*

$$\mu_{K_N \sqcup K_N'} = \frac{1}{N} \cdot \delta_0 + \Big(\frac{N-1}{N}\Big) \cdot \delta_{\frac{N}{N-1}}$$

and

$$\mu_{\Gamma_N} = \frac{1}{2N} \cdot \delta_0 + \frac{1}{2N} \cdot \sum_{i=1}^{2c+1} \delta_{a_i} + \Big(\frac{N-1-c}{N}\Big) \cdot \delta_{\frac{N}{N-1}}$$

for some $a_i \in (0, 2)$.

Remark 6.10. In order to prove Lemma 6.9, we make the following observation. It is easy to see that the spectrum of the symmetric normalized Laplacian matrix $\hat{L} = I_n - D^{-1/2}AD^{-1/2}$ equals the spectrum of the *random walk normalized Laplacian matrix* $\tilde{L} := I_n - D^{-1}A$. Moreover, for a graph with vertex set V, $\tilde{L}$ can be seen as an operator from the set $\{f : V \to \mathbb{R}\}$ to itself. We shall work on this operator for proving Lemma 6.9 and, for a graph Γ, we shall use the simplified notation L^Γ in order to indicate the random walk normalized Laplace operator for Γ.

Proof of Lemma 6.9. Since the spectrum of $K_N \sqcup K'_N$ is given by 0 with multiplicity 2 and $\frac{N}{N-1}$ with multiplicity $2(N-1)$, we have that

$$\mu_{K_N \sqcup K'_N} = \frac{1}{2N}\left(2 \cdot \delta_0 + 2(N-1) \cdot \delta_{\frac{N}{N-1}}\right).$$

In order to prove the second part of the lemma, we shall find $2(N-1-c)$ functions on $V(\Gamma_N)$ that are eigenfunctions for the normalized Laplace operator with eigenvalue $\frac{N}{N-1}$ and are orthogonal to each other. In particular, by the symmetry of Γ_N, it suffices to find $N-1-c$ such functions that are 0 on the vertices of K'_N.

Observe that K_{N-1-c} is a subgraph of $K_N \setminus \{v_1, \ldots, v_c\}$ that has $N-1-c$ eigenfunctions $f_1, \ldots, f_{N-1-c}$ for the largest eigenvalue. These are orthogonal to each other and orthogonal to the constants, therefore

$$0 = \sum_{v \in V(K_{N-1-c})} \deg_{K_{N-1-c}}(v)f_i(v)f_j(v) = \sum_{v \in V(K_{N-1-c})} f_i(v)f_j(v)$$

and

$$\sum_{v \in V(K_{N-1-c})} \deg_{K_{N-1-c}}(v)f_i(v) = \sum_{v \in V(K_{N-1-c})} f_i(v) = 0$$

for each $i, j \in \{1, \ldots, N-1-c\}$. Now, for $i \in \{1, \ldots, N-1-c\}$, let $\tilde{f}_i$ be the function on $V(K_N)$ that is equal to zero on $v_1, \ldots, v_c$ and is equal to f_i otherwise. Then, $\tilde{f}_1, \ldots, \tilde{f}_{N-1-c}$ are orthogonal to each other and orthogonal to the constants because

$$\sum_{v \in V(K_N)} \deg_{K_N}(v)\tilde{f}_i(v)\tilde{f}_j(v) = \sum_{v \in V(K_{N-1-c})} (N-1)\tilde{f}_i(v)\tilde{f}_j(v)$$

$$= \sum_{v \in V(K_{N-1-c})} \tilde{f}_i(v)\tilde{f}_j(v)$$

$$= \sum_{v \in V(K_{N-1-c})} f_i(v)f_j(v)$$

$$= 0$$

and

$$\sum_{v \in V(K_N)} \tilde{f}_i(v) = \sum_{v \in V(K_{N-1-c})} f_i(v) = 0.$$

Since for complete graphs any function that is orthogonal to the constants is an eigenfunction for $\frac{N}{N-1}$, we have that $\tilde{f}_1, \ldots, \tilde{f}_{N-1-c}$ are (pairwise orthogonal) eigenfunctions for $\frac{N}{N-1}$.

Analogously, for $i \in \{1, \ldots, N-1-c\}$, let now $\hat{f}_i$ be the function on $V(\Gamma_N)$ that is equal to zero on $K'_N \cup \{v_1, \ldots, v_c\}$ and is equal to f_i otherwise. It's then easy to see that also these

functions are orthogonal to each other and orthogonal to the constants. Now, for each i and for each $v \in \Gamma_N$ with $\tilde{f}_i(v) = 0$, we have that $v \in K_{N-1-c}$, therefore

$$
\begin{aligned}
L^{\Gamma_N} \hat{f}_i(v) &= \hat{f}_i(v) - \frac{1}{\deg_{\Gamma_N}(v)} \sum_{w \sim v} \hat{f}_i(w) \\
&= \tilde{f}_i(v) - \frac{1}{\deg_{K_N}(v)} \sum_{w \sim v \text{ in } K_N} \tilde{f}_i(w) \\
&= L^{K_N} \tilde{f}_i(v) \\
&= \frac{N}{N-1} \cdot \tilde{f}_i(v) \\
&= \frac{N}{N-1} \cdot \hat{f}_i(v).
\end{aligned}
$$

This proves that the functions $\hat{f}_i$'s are $N-1-c$ orthogonal eigenfunctions of the Laplace Operator in Γ_N for the eigenvalue $\frac{N}{N-1}$. Since they are all 0 on K'_N, by symmetry we can also get $N-1-c$ eigenfunctions for $\frac{N}{N-1}$ on Γ_N that are 0 on K_N and therefore are orthogonal to the first $N-1-c$ functions. This implies that the multiplicity of $\frac{N}{N-1}$ for Γ_N is at least $2(N-1-c)$. Therefore,

$$
\mu_{\Gamma_N} = \frac{1}{2N} \cdot \delta_0 + \frac{1}{2N} \cdot \sum_{i=1}^{2c+1} \delta_{a_i} + \left(\frac{N-1-c}{N} \right) \cdot \delta_{\frac{N}{N-1}}
$$

for some $a_i \in (0, 2)$. $\qquad\square$

Corollary 6.11. *The total variation distance between the probability measures $\mu_{K_N \sqcup K'_N}$ and μ_{Γ_N} defined in the previous lemma is*

$$
\sup_{A \subseteq [0,2] \text{ measurable}} \left| \mu_{K_N \sqcup K'_N}(A) - \mu_{\Gamma_N}(A) \right| = \frac{2c+1}{2N}.
$$

In particular, if $c = o(N)$, the total variation distance tends to zero for $N \to \infty$.

Example 6.12. The previous corollary doesn't hold in general. As a counterexample, take two copies of the path on N vertices, P_N and P'_N. Their union via one external edge can be for example the path on $2N$ vertices, and

$$
\mu_{P_{2N}} = \frac{1}{2N} \left(\sum_{k=0}^{2N-1} \delta_{1 - \cos \frac{\pi k}{2N-1}} \right),
$$

while

$$
\mu_{P_N \sqcup P'_N} = \frac{1}{N} \left(\sum_{k=0}^{N-1} \delta_{1 - \cos \frac{\pi k}{N-1}} \right).
$$

The total variation distance between these two measures does not tend to zero for $N \to \infty$.

7. Random geometric graphs

Specializing Corollary 4.4 to the case $k = 1$, we get

$$
\varphi_*^{(1)} \Theta_n \xrightarrow{\mathbb{P}} \varphi_*^{(1)} \Theta \qquad \text{and} \qquad \varphi_*^{(1)} \Theta_R \xrightarrow{\mathbb{P}} \varphi_*^{(1)} \Theta,
$$

with $\operatorname{supp} \varphi_*^{(1)} \Theta = \mathcal{G}^{(1)}(\mathbb{R}^m)$.

Definition 7.1. We define the random geometric graphs

$$\Gamma_n := \check{C}^{(1)}(\mathcal{U}_n) \qquad \text{and} \qquad \Gamma_R := \check{C}^{(1)}(\mathcal{P}_R)$$

and we associate the empirical spectral measures

$$\mu_n := \mu_{\Gamma_n} = \frac{1}{n}\sum_{i=1}^{n} \delta_{\lambda_i(\Gamma_n)} \qquad \text{and} \qquad \mu_R := \mu_{\Gamma_R} = \frac{1}{\#(V(\Gamma_R))}\sum_{i=1}^{\#(V(\Gamma_R))} \delta_{\lambda_i(\Gamma_R)}.$$

Since the spectrum of a graph Γ is finite, we can rewrite the two random measures μ_n and μ_R above as follows. First, the set of all possible isomorphism classes of graphs is countable and therefore there exists a sequence $\{x_\ell\}_{\ell \in \mathbb{N}} \subset [0,2]$ such that:

$$\mu_n = \sum_{\ell=1}^{\infty} c_{\ell,n}\delta_{x_\ell} \quad \text{and} \quad \mu_R = \sum_{\ell=1}^{\infty} c_{\ell,R}\delta_{x_\ell},$$

where the random variables $c_{\ell,n}$ and $c_{\ell,R}$ are defined by:

$$c_{\ell,n} = \frac{1}{n}\#\{\text{components } \Gamma_{n,j} \text{ of } \Gamma_n = \sqcup\Gamma_{n,j} \text{ such that } x_\ell \text{ belongs to the spectrum of } \Gamma_{n,j}\}$$

and

$$c_{\ell,R} = \frac{1}{\#(V(\Gamma_R))}\#\{\text{components } \Gamma_{R,j} \text{ of } \Gamma_R = \sqcup\Gamma_{R,j} \text{ such that } x_\ell \text{ belongs to the spectrum of } \Gamma_{R,j}\}.$$

In the above definitions of the coefficients $c_{\ell,n}$ and $c_{\ell,R}$, the components should be counted "with multiplicity": if x_ℓ belongs to the spectrum of $\Gamma_{n,j}$ (respectively $\Gamma_{R,j}$) with some multiplicity $m(x_\ell)$, then the component $\Gamma_{n,j}$ (respectively $\Gamma_{R,j}$) is counted $m(x_\ell)$ times.

We will need the following Lemma. We will need the following Lemma.

Lemma 7.2. *For every $\delta > 0$ there exists $L > 0$ such that*

$$\mathbb{E}\sum_{\ell \geq L} c_{\ell,n} < \frac{\delta}{4} \quad \text{and} \quad \mathbb{E}\sum_{\ell \geq L} c_{\ell,R} < \frac{\delta}{4}.$$

Proof. Let $A \subset \{x_\ell\}_{\ell \in \mathbb{N}}$ be any subset. Then:

$$\sum_{\ell} \mathbb{E}c_{\ell,n} = \mathbb{E}\sum_{\ell} c_{\ell,n} \leq 1 \quad \text{and} \quad \sum_{\ell} \mathbb{E}c_{\ell,R} = \mathbb{E}\sum_{\ell} c_{\ell,R} \leq 1.$$

In particular the two series $\mathbb{E}\sum_{\ell} c_{\ell,n}$ and $\mathbb{E}\sum_{\ell} c_{\ell,R}$ converge and therefore the existence of such L is clear (as the tails of the series must be arbitrarily small). $\square$

Corollary 7.3. *For every $\ell \in \mathbb{N}$ there exists constants $c_\ell \geq 0$ such that we have the following convergence of random variables:*

$$c_{\ell,n} \xrightarrow{L^1} c_\ell \quad \text{and} \quad c_{\ell,R} \xrightarrow{L^1} c_\ell.$$

The constants c_ℓ are positive if and only if x_ℓ belongs to the spectrum of a $\mathbb{R}^m$-geometric graph. Moreover the measure

$$\mu := \sum_{\ell \in \mathbb{N}} c_\ell \delta_{x_\ell} \tag{7.1}$$

is a probability measure on $\mathbb{R}$ with support contained in $[0,2]$.

Proof. The convergence in L^1 of the random variales $c_{\ell,n}$ and $c_{\ell,R}$ follows from their description as "number of components such that x_ℓ belongs to the spectrum of the random geometric graph": in fact for every $\ell \geq 0$ we can introduce the counting function:

$$\mathcal{N}(\Gamma, \ell) = \#\{\text{components of } \Gamma \text{ for which } x_\ell \text{ belongs to their spectrum}\}.$$

With this notation we have:

$$c_{\ell,n} = \frac{\mathcal{N}(\Gamma_n, \ell)}{n} \quad \text{and} \quad c_{\ell,R} = \frac{\mathcal{N}(\Gamma_R, \ell)}{\text{vol}(B(0,R))} = \frac{\mathcal{N}(\Gamma_R, \ell)}{\#(V(\Gamma_R))} \cdot \frac{\#(V(\Gamma_R))}{\text{vol}(B(0,R))}.$$

The convergence of $c_{\ell,n}$ follows the exact same proof of Theorem 4.11 and the convergence of $c_{\ell,R}$ proceeds as in Proposition 4.7. The measure μ is well defined and the fact that it is a probability measure follows from the same proof as [ALL20, Proposition 6.4], using the above Lemma 7.2 as a substitute of [ALL20, Lemma 6.3]. $\qquad\square$

Theorem 7.4. *Let $\mu \in \mathcal{M}^1([0,2])$ be the measure defined in (7.1). Then*

$$\mu_n \xrightarrow[n\to\infty]{*} \mu \qquad and \qquad \mu_R \xrightarrow[R\to\infty]{*} \mu.$$

i.e. $\forall f \in C^0([0,2], \mathbb{R})$ we have $\mathbb{E} \int f \, d\mu_n \to \int f \, d\mu$ and $\mathbb{E} \int f \, d\mu_R \to \int f \, d\mu$

Proof. The proof of the statement for μ_n and μ_R is the same: we do it for μ_R. Let $f \in C^0([0,2], \mathbb{R})$, and fix $\varepsilon > 0$. Apply Lemma 7.2 with the choice of $\delta = \varepsilon/\sup|f|$, and get the corresponding set $F = \{\ell_1, \ldots, \ell_a\} \subset \mathbb{N}$. Also, from the convergence of the series $\sum c_\ell$ we get the existence of a finite set F' such that $\sum_{\ell \notin F} c_\ell < \varepsilon/\sup|f|$. Define $F'' = F \cup F''$ (this is still a finite set) and:

$$\left| \mathbb{E} \int_{[0,2]} f \, d\mu_R - \int_{[0,2]} f \, d\mu \right| = \left| \mathbb{E} \sum_\ell c_{\ell,R} f(x_\ell) - \sum_\ell c_\ell f(x_\ell) \right|$$

$$= \left| \mathbb{E} \sum_{\ell \in F''} c_{\ell,R} f(x_\ell) + \mathbb{E} \sum_{\ell \notin F''} c_{\ell,R} f(x_\ell) - \sum_{\ell \in F''} c_\ell f(x_\ell) - \sum_{\ell \notin F''} c_\ell f(x_\ell) \right|$$

$$\leq \left| \mathbb{E} \sum_{\ell \in F''} c_{\ell,R} f(x_\ell) - \sum_{\ell \in F''} c_\ell f(x_\ell) \right|$$

$$+ \left| \mathbb{E} \sum_{\ell \notin F''} c_{\ell,R} f(x_\ell) \right| + \left| \sum_{\ell \notin F''} c_\ell f(x_\ell) \right|$$

$$\leq \left| \sum_{\ell \in F''} \mathbb{E}(c_{\ell,R} - c_\ell) f(x_\ell) \right| + \sup|f| \cdot \left| \sum_{\ell \notin F''} \mathbb{E} c_{\ell,R} \right| + \sup|f| \cdot \left| \sum_{\ell \notin F''} c_\ell \right|$$

$$\leq \sup|f| \sum_{\ell \in F''} \mathbb{E}|c_{\ell,R} - c_\ell| + \sup|f| \frac{\delta}{4} + \sup|f| \frac{\delta}{4}$$

$$\leq \mathbb{E}|c_{\ell,R} - c_\ell| + \frac{\varepsilon}{2} \longrightarrow \frac{\varepsilon}{2} \quad \text{as } R \to \infty,$$

where in the last line we have used the L^1 convergence of $c_{\ell,R} \to c_\ell$. Since this is true for every $\varepsilon > 0$, it follows that:

$$\lim_{R\to\infty} \mathbb{E} \int_{[0,2]} f \, d\mu_R = \int_{[0,2]} d\mu.$$

$\square$

Proposition 7.5. *The measure μ appearing in Theorem 7.4 has the following properties:*

(1) *μ is not absolutely continuous with respect to Lebesgue;*
(2) *$\lim_{n \to \infty} \frac{1}{n} \cdot \mathbb{E} \# \big(eigenvalues\ of\ \Gamma_n\ in\ [a, b] \big) = \mu([a, b]) > 0;$*
(3) *$\mu(\{0\}) = \beta.$*

Proof. We start with point (3): we have that $\mu(\{0\}) = c_{\ell_0}$ where $0 = x_{\ell_0}$ and c_{ℓ_0} is the L^1-limit of $c_{\ell_0, R}$, which is the random variable:

$$c_{\ell_0, R} = \#\{\text{components of } \Gamma_R \text{ containing } 0 \text{ in their spectrum}\} = b_0(\Gamma_R),$$

and therefore (3) follows from the definition of β.

Point (1) also follows immediately, since $\beta_0 > 0$ and μ charges positively sets of Lebesgue measure zero (hence it cannot be absolutely continuous with respect to Lebesghe measure).

For the proof of point (2) we argue exactly as in the proof of Theorem 7.4, by replacing f with $\chi_{[a,b]}$ (the characteristic function of the interval $[a, b]$) and observing that the only property of f that we have used is its boundedness. $\square$

References

[ALL20] A. Auffinger, A. Lerario, and E. Lundberg. Topologies of random geometric complexes on Riemannian manifolds in the thermodynamic limit. *IMRN, to appear*, 2020.

[BA99] A. Barabási and R. Albert. Emergence of scaling in random networks. *Science*, 286:509–512, 1999.

[BK18] Omer Bobrowski and Matthew Kahle. Topology of random geometric complexes: a survey. *Journal of Applied and Computational Topology*, 1(3):331–364, Jun 2018.

[BLW76] N. L. Biggs, E. K. Lloyd, and R. J. Wilson. Graph theory. *Clarendon, Oxford*, pages 1736–1936, 1976.

[BM15] Omer Bobrowski and Sayan Mukherjee. The topology of probability distributions on manifolds. *Probab. Theory Related Fields*, 161(3-4):651–686, 2015.

[Bol85] B. Bollobás. Random graphs. *Academic Press, London*, 1985.

[CdV98] Y. Colin de Verdière. Spectres de graphes. Cours spécialisés. *Soc. Math. de France, Paris*, 4, 1998.

[Chu97] F. R. K. Chung. *Spectral graph theory*, volume 92 of *CBMS Regional Conference Series in Mathematics*. Published for the Conference Board of the Mathematical Sciences, Washington, DC; by the American Mathematical Society, Providence, RI, 1997.

[CL06] F. Chung and L. Lu. Complex graphs and networks, cbms regional conference series in mathematics. *Conf. Board Math. Sci., Washington*, 92, 2006.

[DGK17] C. P. Dettmann, O. Georgiou, and G. Knight. Spectral statistics of random geometric graphs. *Europhysical Letters*, 118(1), 2017.

[DJ10] X. Ding and T. Jiang. Spectral distributions of adjacency and Laplacian matrices of random graphs. *The Annals of Applied Probability*, 20(6):2086–2117, 2010.

[Dur07] R. Durrett. Random graph dynamics. *Cambridge Univ. Press, Cambridge*, 2007.

[EE92] S. N. Evangelou and E. N. Economou. Spectral density singularities, level statistics, and localization in sparse random matrices. *Phys. Rev. Lett.*, 68:361–364, 1992.

[ER59] P. Erdős and A. Rényi. On random graphs. I. *Publicationes Mathematicae*, 6:290–297, 1959.

[ER60] P. Erdős and A. Rényi. On the evolution of random graphs. *Publ. Math. Inst. Hung. Acad. Sci.*, 5:17–61, 1960.

[ER61a] P. Erdős and A. Rényi. On the evolution of random graphs. *Bull. Inst. Internat. Statist.*, 38:343–347, 1961.

[ER61b] P. Erdős and A. Rényi. On the strength of connectedness of a random graph. *Acta Mathematica Academiae Scientiarum Hungaricae*, 12:261–267, 1961.

[ES74] P. Erdős and J. Spencer. Probabilistic methods in combinatorics. *Academic Press, New York*, 1974.

[Eva83] S. N. Evangelou. Quantum percolation and the Anderson transition in dilute systems. *Phys. Rev. B*, 27:1397–1400, 1983.

[Eva92] S. N. Evangelou. A numerical study of sparse random matrices. *J. Stat. Phys.*, 69:361–383, 1992.

[Gil61]   E. N. Gilbert. Random plane networks. *J. Soc. Indust. Appl. Math.*, 9(4):533–543, 1961.

[GJLS16]   Jiao Gu, Jürgen Jost, Shiping Liu, and Peter F. Stadler. Spectral classes of regular, random, and empirical graphs. *Linear Algebra Appl.*, 489:30–49, 2016.

[GOn96]   M. Gutierrez and L. Oubiña. Metric characterizations of proper interval graphs and tree-clique graphs. *J. Graph Theory*, 21(2):199–205, 1996.

[GTT19]   Akshay Goel, Khanh Duy Trinh, and Kenkichi Tsunoda. Strong law of large numbers for Betti numbers in the thermodynamic regime. *Journal of Statistical Physics*, 174(4):865–892, Feb 2019.

[Jac92]   Z. Jackowski. A new characterization of proper interval graphs. *Discrete Math.*, 105(1-3):103–109, 1992.

[JLR00]   S. Janson, T. Luczak, and A. Rucinski. Random graphs. *Wiley, New York*, 2000.

[JM19]   J. Jost and R. Mulas. Hypergraph Laplace operators for chemical reaction networks. *Advances in Mathematics*, 351:870–896, 2019.

[Kah11a]   M. Kahle. Random geometric complexes. *Discrete Comput. Geom.*, 45(3):553–573, 2011.

[Kah11b]   Matthew Kahle. Random geometric complexes. *Discrete Comput. Geom.*, 45(3):553–573, 2011.

[KM12]   R. J. Kang and T. Müller. Sphere and dot product representations of graphs. *Discrete Comput. Geom.*, 47(3):548–568, 2012.

[Kol99]   V. F. Kolchin. Random graphs. Encyclopedia of mathematics and its applications. *Cambridge Univ. Press, Cambridge*, 53, 1999.

[LB63]   C. G. Lekkerkerker and J. C. Boland. Representation of a finite graph by a set of intervals on the real line. *Fund. Math.*, 51:45–64, 1962/1963.

[M.85]   Palmer. E. M. Graphical evolution: An introduction to the theory of random graphs. *Wiley, Chichester*, 1985.

[Mae84]   H. Maehara. Space graphs and sphericity. *Discrete Appl. Math.*, 7(1):55–64, 1984.

[Mer08]   G. B. Mertzios. A matrix characterization of interval and proper interval graphs. *Appl. Math. Lett.*, 21(4):332–337, 2008.

[MF91]   A. D. Mirlin and Y. V. Fyodorov. Universality of level correlation function of sparse random matrices. *J. Phys. A*, 24:242273–2286, 1991.

[NGB15]   A. Nyberg, T. Gross, and K. E. Bassler. Mesoscopic structures and the Laplacian spectra of random geometric graphs. *Journal of Complex Networks*, 3(4):543–551, 2015.

[Nov98]   S. P. Novikov. Schrödinger operators on graphs and symplectic geometry. *In The Arnoldfest. Field Institute, Toronto*, 2, 1998.

[NS16]   F. Nazarov and M. Sodin. Asymptotic laws for the spatial distribution and the number of connected components of zero sets of Gaussian random functions. *Zh. Mat. Fiz. Anal. Geom.*, 12(3):205–278, 2016.

[NSW08]   Partha Niyogi, Stephen Smale, and Shmuel Weinberger. Finding the homology of submanifolds with high confidence from random samples. *Discrete Comput. Geom.*, 39(1-3):419–441, 2008.

[Pen03a]   M. Penrose. Random geometric graphs. *Oxford studies in probability. Oxford University Press*, 2003.

[Pen03b]   Mathew Penrose. *Random geometric graphs*, volume 5 of *Oxford Studies in Probability*. Oxford University Press, Oxford, 2003.

[Pup08]   T. Puppe. Spectral graph drawing: a survey. *VDM, Verlag*, 2008.

[RB88]   G. J. Rodgers and A. J. Bray. Density of states of a sparse random matrix. *Phys. Rev. B*, 37(3):3557–3562, 1988.

[RDD90]   G. J. Rodgers and C. De Dominicis. Density of states of sparse random matrices. *J. Phys. A*, 23:1567–1573, 1990.

[Rob69]   F. S. Roberts. Indifference graphs. In *Proof Techniques in Graph Theory (Proc. Second Ann Arbor Graph Theory Conf., Ann Arbor, Mich., 1968)*, pages 139–146. Academic Press, New York, 1969.

[SW19]   Peter Sarnak and Igor Wigman. Topologies of nodal sets of random band-limited functions. *Comm. Pure Appl. Math.*, 72(2):275–342, 2019.

[Tao12]   T. Tao. *Topics in random matrix theory*, volume 132 of *Graduate Studies in Mathematics*. American Mathematical Society, Providence, RI, 2012.

[Wal11]   M. Walters. Surveys in combinatorics 2011. london mathematical society lecture note series, 392. *University Press, Cambridge*, pages 365–401, 2011.

[Weg04]   G. Wegner. Eigenschaften der Nerven homologisch-einfacher Familien im $\mathbb{R}^n$. *Ph.D. thesis, Göttingen University, Germany*, 2004.

[WS98]   D. J. Watts and S. H. Strogatz. Collective dynamics of 'small-world' networks. *Nature*, 393(6684):409–10, 1998.

[YSA17] D. Yogeshwaran, Eliran Subag, and Robert J. Adler. Random geometric complexes in the thermo-
         dynamic regime. *Probab. Theory Related Fields*, 167(1-2):107–142, 2017.

SISSA, Trieste, Italy

E-mail address: `lerario@sissa.it`

Max Planck Institute for Mathematics in the Sciences, Leipzig, Germany

E-mail address: `raffaella.mulas@mis.mpg.de`