RANDOMLY WEIGHTED $d-$COMPLEXES: MINIMAL SPANNING ACYCLES AND PERSISTENCE DIAGRAMS

PRIMOZ SKRABA, GUGAN THOPPE*, AND D. YOGESHWARAN

ABSTRACT. A weighted $d-$complex is a simplicial complex of dimension d in which each face is assigned a real-valued weight. We derive three key results here concerning persistence diagrams and minimal spanning acycles (MSAs) of such complexes. First, we establish an equivalence between the MSA face-weights and *death times* in the persistence diagram. Next, we show a novel stability result for the MSA face-weights which, due to our first result, also holds true for the death and birth times, separately. Our final result concerns a perturbation of a mean-field model of randomly weighted $d-$complexes. The $d-$face weights here are perturbation of some i.i.d. distribution while all the lower-dimensional faces have a weight of 0. If the perturbations decay sufficiently quickly, we show that suitably scaled extremal nearest face-weights, face-weights of the $d-$MSA, and the associated death times converge to an inhomogeneous Poisson point process. This result completely characterizes the extremal points of persistence diagrams and MSAs. The point process convergence and the asymptotic equivalence of three point processes are new for any weighted random complex model, including even the non-perturbed case. Lastly, as a consequence of our stability result, we show that Frieze's $\zeta(3)$ limit [Fri85] for random minimal spanning trees and the recent extension to random MSAs by Hino and Kanazawa [HK19] also hold in suitable noisy settings.

1. INTRODUCTION

Broadly, there are two parts to this paper. The first part concerns weighted simplicial complexes. This study significantly deepens the understanding of the relationship of minimal spanning acycles in such complexes to associated persistence diagrams and also to what we refer to as "nearest face" distances. The second part looks at a specific "mean-field" model of complexes with random weights and, in parallel, also considers its perturbations. We

(PS) SCHOOL OF MATHEMATICAL SCIENCES, QUEEN MARY UNIVERSITY OF LONDON, UK & JOZEF STEFAN INSTITUTE, LJUBLJANA, SLOVENIA. Research supported by ARRS project N1-0058.

(GT) DEPT. OF COMPUTER SCIENCE AND AUTOMATION, INDIAN INSTITUTE OF SCIENCE, BENGALURU, INDIA; PAST: FACULTY OF ELECTRICAL ENGINEERING, TECHNION-ISRAEL INSTITUTE OF TECHNOLOGY, HAIFA, ISRAEL. Research supported by URSAT, ERC Grant 320422.

(DY) STATISTICS AND MATHEMATICS UNIT, INDIAN STATISTICAL INSTITUTE, BENGALURU, INDIA. Research supported by DST-INSPIRE faculty award.

E-mail addresses: `primoz.skraba@ijs.si, gugan.thoppe@gmail.com, d.yogesh@isibang.ac.in`.

1991 *Mathematics Subject Classification.* Primary : 60C05, 05E45 Secondary : 60G70, 60B99, 05C80 .

Key words and phrases. Random complexes, Persistent diagrams, Minimal spanning acycles, Point processes, Weak convergence, Stability.

*Corresponding Author. Postal Address: Dept. of Computer Science and Automation, Indian Institute of Science, Bengaluru, Karnataka 560012, India; Tel.: +91 996-906-6600; Fax:+91-80-23602911.

refer to these complexes as randomly weighted $d-$complexes or, simply, weighted random complexes. Our results completely characterize the extremal behaviour of the persistence diagram and the nearest face distances associated with such complexes and then, using the above relationships, also of their minimal spanning acycles.

The motivation for this work comes from the much more studied scenario of weighted graphs, the $1-$dimensional analogue of weighted simplicial complexes, and their random counterparts. A weighted graph can either be viewed in its entirety or as a process wherein it is sequentially built by adding edges in an order dictated by their weights. Taking the former perspective, a minimal spanning acycle corresponds to a minimal spanning tree, while the nearest face distances are basically the nearest neighbour distances. The other viewpoint helps interpret the persistence diagram associated with a graph; informally, it is a record of the "death times", i.e., the weight values of those edges that connect a priori disjoint components.

The fact that connectivity and nearest neighbour distances are intertwined can be seen from the earliest work itself on random graphs by Erdős and Rényi [ER59]. In fact, three years earlier, Kruskal had proposed his algorithm for constructing a minimal spanning tree. The edge weights of this tree are precisely the times at which components get connected in the process type description of the weighted graph. Hence, that work can be viewed as the first to implicitly exploit the connections between minimal spanning trees and persistence diagrams. Indeed, the notion of persistence diagrams did not exist then, but one interpretation of Kruskal's algorithm is via persistence diagrams and this relation is made more clear in this paper. This implicit relationship was also used later in the seminal work of Frieze [Fri85]. On the other hand, connections between the largest nearest neighbour distances and the longest edges of a minimal spanning tree on randomly weighted graphs have played a key role in [Hen82, ST86, AR02, Pen97, HR05]. In [Pen97], it was shown that the extremal nearest neighbour distances coincided with that of extremal edge-weights in a Euclidean random minimal spanning tree. Such a result is crucial to understanding the connectivity threshold for random geometric graphs (see [Pen03, Chapter 13]). More complete accounts of such connections can be found in [Bol01, vdH16, JLR00, Pen03, FK16].

Recent applications in topological data analysis have motivated the extension of the above results to random complexes. While higher dimensional analogues of connectivity thresholds have already been studied [LM06, MW09, Kah14a], this work generalizes some of these later results to the level of persistence diagrams and minimal spanning acycles. Before delving into the background and details of our results, we summarize our main contributions. Note that a weighted $d-$complex is a simplicial complex with dimension d in which each face is assigned a real-valued weight. Throughout, we will assume that this weight function is monotone, i.e., the weight of any face is always larger than that of its sub-faces. Such a weighted complex can also be viewed as a process wherein one adds faces in the order dictated by their weights. Because the weight function is monotone, any intermediate construction is also a simplicial complex. With this dual perspective, one can infer properties about minimal spanning acycles from the death and birth times in the persistence diagram and vice versa.

The section numbers in brackets below indicate where one can find a detailed description of the corresponding contribution.

Key Contributions:

(1) We first provide a simplicial analogue of Kruskal's algorithm that can be used for finding minimal spanning acycles. Comparing this algorithm with the incremental algorithm used to build a persistence diagram, we establish an equivalence between the face-weights of minimal spanning acycles and the death times; a similar result also holds true for the birth times. This result significantly enhances the connection between minimal spanning acycles and persistence diagrams. In fact, one of the theorems in Hiraoka and Shirai [HS17, Theorem 1.1] now becomes a simple corollary of our result. (Section 1.2).

(2) Next, we establish a new stability result for minimal spanning acycles. Because of the equivalence above, this result then automatically applies to the death and birth times as well. Unlike existing stability results for persistence diagrams which concern the multiset of birth-death pairs, our result specifically relates the changes in the set of deaths and, separately, in the set of births to the changes in the face weights. We believe this result can play a crucial role in proving results for randomly weighted complexes with certain dependencies between the different face weights. (Section 1.3)

(3) Our final key result concerns randomly weighted $d-$complexes and suitable noisy perturbations of them, including those with dependencies. If the perturbations decay sufficiently fast, we show that appropriately scaled versions of the following three point processes: (a) nearest face distances, (b) death times in the persistence diagram, and (c) face-weights of the minimum spanning acycle – converge weakly to the same Poisson point process in vague topology. Derivation of (a) and (b) involves use of the method of factorial moments, cohomology theory, and our stability result. On the other hand, going from (b) to (c) is a simple application of our first result. However, unlike in our second contribution, notice that this time we exploit the equivalence in the other direction, i.e., we go from a result on death times to a result on face weights of the minimal spanning acycle. An important paradigm in topological data analysis is that extremal points of a persistence diagram encode meaningful topological information about the underlying structure. Viewed in this light, our result completely characterizes the extremal points of the persistence diagram of these weighted random complexes; this is new even in the non-noisy scenario. (Section 1.4).

(4) We conclude by providing another application of our stability result. Namely, the lifetime sum of persistence diagrams converge for randomly weighted $d-$complexes with noisy weights; this generalizes the $\zeta(3)$-limit for random minimal spanning trees by Frieze [Fri85] and the recent extension to random spanning acycles by Hino and Kanazawa [HK19]. (Section 1.4).

ORGANISATION OF PAPER: The rest of this section quickly introduces simplicial complexes, minimal spanning acycles, persistent homology and then provides more precise statements of some of our main results as well as place them in context. The next section - Section 2 - gives in detail the necessary topological (Section 2.1) and probabilistic preliminaries (Section 2.2)[1]. Section 3 is exclusively devoted to studying various properties of minimal spanning acycles, algorithms to find them, their connection to persistence diagrams, and our crucial stability result. Finally, in Section 4, we study weighted random complexes and prove our point process convergence results. In subsection 4.1, the weights are independent and identically distributed (i.i.d.) uniform $[0, 1]$ random variables while, in subsection 4.2, the weights are either i.i.d. with a more general distribution $\mathscr{F}$ or a perturbation of the same. In the Appendix, we give proofs of two results needed for the main part of the paper and a brief explanation on the method of factorial moments.

1.1. **Simplicial complexes and minimal spanning acycles:** We begin by defining a (simplicial) complex, which is a higher dimensional analogue of a graph.

Definition 1.1. *An (abstract) simplicial complex $\mathcal{K}$ on a finite ground set V is a collection of subsets of V such that if $\sigma_1 \in \mathcal{K}$ and $\emptyset \neq \sigma_2 \subset \sigma_1$, then $\sigma_2 \in \mathcal{K}$ as well. The elements of $\mathcal{K}$ are called* simplices or faces *and the dimension of a simplex σ is $|\sigma| - 1$, where $|\cdot|$ means cardinality. A $d-$face of $\mathcal{K}$ is a face of $\mathcal{K}$ with dimension d.*

Given a complex $\mathcal{K}$ and $d \geq 0$, we denote the $d-$faces of $\mathcal{K}$ by $\mathcal{F}^d(\mathcal{K})$ and its $d-$skeleton by $\mathcal{K}^d$ (i.e., the sub-complex of $\mathcal{K}$ consisting of all faces of dimension at most d). We use σ, τ to denote faces and the dimension of the face shall not be explicitly mentioned unless required. A graph is a complex that consists only of 0-faces and 1-faces, or in other words, the 1-skeleton of a complex is a graph. Associated to each simplicial complex is a collection of non-negative integers denoted $\beta_0(\mathcal{K}), \beta_1(\mathcal{K}), \ldots$, called the Betti numbers[2] (see Section 2.1 for detailed definitions) which are a measure of connectivity of the simplicial complex. Informally, the $d-$th Betti number counts the number of $(d + 1)$-dimensional holes in the complex or equivalently the number of independent non-trivial cycles formed by $d-$faces. Two points to note at the moment are: (i) $\beta_0(\mathcal{K})$ is one less than the number of connected components in the graph formed by 0-faces and 1-faces and (ii) if the dimension of $\mathcal{K}$ (maximum of dimension of faces) is d, then $\beta_j(\mathcal{K}) = 0$ for all $j \geq d + 1$.

The Betti numbers described above are closely connected to spanning acycles. For example, the spanning tree of a graph on a vertex set V can be described in topological terms as a set of edges S such that $\beta_0(V \cup S) = \beta_1(V \cup S) = 0$, i.e., $V \cup S$ is connected and has no cycles. The following higher-dimensional generalization by Kalai [Kal83] is then natural.

Definition 1.2 (Spanning and Maximal acycle). *Consider a complex $\mathcal{K}$ of dimension at least d, $d \geq 1$. A subset S of $d-$faces is said to be* spanning *if $\beta_{d-1}(\mathcal{K}^{d-1} \cup S) = 0$ and*

---

[1]In our effort to make this paper reasonably self-contained as well as accessible to the trio of probabilists, combinatorialists and topologists, we have erred on the side of including too much detail rather than terseness.

[2]Throughout the paper, we work with reduced Betti numbers defined using field coefficients.

4

an accle *if* $\beta_d(\mathcal{K}^{d-1} \cup S) = 0$; *it is called a* spanning acycle *if it has both the properties.* *Separately, a subset S of $d-$faces is called a* maximal acycle *if it is an acycle and maximal with respect to (w.r.t.) the inclusion of $d-$faces.*

Though this definition of a spanning acycle merely replaces appropriate indices in the definition of a spanning tree, what is not obvious is that this is a *good* higher-dimensional generalization of a spanning tree. This work of ours is the first to formally ascertain that several key properties of a spanning tree naturally extend to a spanning acycle as well; see our results in Section 3.

An alternative but more explicit algebraic description of a spanning tree is that it consists of a set of columns which form a basis for the column space of the incidence matrix or the boundary matrix; i.e., the matrix ∂_1 whose rows are indexed by vertices and columns by edges and its $i, j-$th entry is 1 if the vertex i belongs to the edge j and 0 otherwise. For simplicity, we assume throughout this paper that the underlying field $\mathbb{F} = \mathbb{Z}_2$ here, i.e., all vector spaces involved are $\mathbb{Z}_2$-vector spaces. It is well-known that the space of bases for these vector spaces form a matroid. Such a description also holds for spanning acycles. While we never explicitly work with this latter description in this paper, it, however, implicitly underpins many of our proof ideas. We shall explicitly point this out whenever that is the case.

If we assign weights to the faces, we obtain a weighted complex $\mathcal{K}$. Now, setting $w(S) := \sum_{\sigma \in S} w(\sigma)$ for a subset S of simplices, we can naturally define a minimal spanning acycle as a spanning acycle S with minimum weight $w(S)$. Since we deal with only finite complexes, the existence of a minimal spanning acycle is guaranteed once a spanning acycle exists. We shall denote a minimal spanning acycle by M_d or simply M when the dimension is clear. Though Kalai's definition of a spanning acycle and enumeration of number of spanning acycles (a generalization of Cayley's formula for spanning trees) is more than three decades old, it is receiving increased attention in the last few years [BBC14, CCK15, DKM09, IKR$^+$11, DKM16, KKL19, KR14, HS17, HS16, Lyo09, LNPR19, MNRR15]. In Section 3, we prove some fundamental properties for minimal spanning acycles: existence, uniqueness, cut property and a simplicial Kruskal's algorithm. Here we would also like to emphasize that properties of spanning acycles are preserved under simplicial isomorphisms but not necessarily under homotopy equivalence.

We would like to highlight that some more fundamental properties of the minimal spanning acycles can be found in an earlier version of our article (see [STY17]); we do not include them here since they are not used elsewhere in the paper. Of these, we would like to point the reader to two interesting results which are not known for matroids in general. First is an inclusion-exclusion identity for the cardinality of maximal acycles which is derived using the Mayer-Vietoris exact sequence from algebraic topology. Second, we provide a generalization of Jarník-Prim-Dijkstra's algorithm to spanning acycles. In fact, we need the complex to be 'hypergraph connected' for the Prim's algorithm to work and it is not obvious what is the analogous notion of 'hypergraph connectivity' in general matroids. As part of the proof, we also show that a spanning acycle is 'hypergraph connected', again by using the Mayer-Vietoris sequence.

1.2. **Persistence diagrams and minimal spanning acycles.** We now preview the connection between persistence diagrams and acycles. Let $\mathcal{K}$ be a weighted complex such that the real-valued weight function w is monotone. Then, $\mathcal{K}(t) := w^{-1}(-\infty, t]$ is a simplicial complex for all $t \in \mathbb{R}$ and we will refer to $\{\mathcal{K}(t) : t \in \mathbb{R}\}$ as the filtration induced by w on $\mathcal{K}$.

Let $d \geq 0$ and suppose that $\beta_d(\mathcal{K}) = 0$. Let $\beta_d(t) = \beta_d(\mathcal{K}(t))$. We remark that $\beta_d(t)$ is a jump function. The times of positive jumps (counted with multiplicity) are birth times $\mathcal{B} = \{B_i\}$ of the persistence diagram and the times of negative jumps (counted with multiplicity) are death times $\mathcal{D} = \{D_i\}$. The correct way to count multiplicity will be made clear in Definition 2.1. However, if the weight function is injective, then there is no multiplicity. The non-expert reader may assume weight functions to be injective for ease of understanding the results in the introduction.

Formally, a persistence diagram corresponding to dimension d is the multiset of the points $\{(B_i, D_i)\}$. Note that it is not only a record of the birth and death times, but importantly also of the pairing of a birth with its corresponding death. A persistence diagram is useful for understanding the evolution of topology of a filtration. See Figure 1 for persistence diagrams of two weighted random complexes - the uniformly weighted random $d-$complexes (see Section 4.1) and Erdős-Rényi clique complexes. The aforementioned persistence diagram would be referred to as the persistence diagram of $\mathrm{H}_d(\mathcal{K})$ whenever we wish to avoid ambiguities about the dimension and the underlying complex. In this paper, we shall focus only on their two projections - birth and death times. Though not everything can be inferred from these projections, a crucial quantity that can be understood from these projections is the *lifetime sum* $L_d(\mathcal{K}) := \sum_i (D_i - B_i)$, which by Fubini's theorem also equals $\int_0^\infty \beta_d(t) \mathrm{d}t$ ([HS17, (1.4)]). We now present the first of our main theorems that connects persistence diagrams to minimal spanning acycles. Here and elsewhere, when the underlying complex $\mathcal{K}$ is clear we shall drop it from all our notations.

Theorem 1.3. *Let $\mathcal{K}$ be a weighted $d-$complex with $\beta_{d-1}(\mathcal{K}) = 0$. Let $\mathcal{D}$ be the point-set of death times in the persistence diagram of the $\mathrm{H}_{d-1}(\mathcal{K})$ with the canonical filtration[3] induced by the weights. Similarly, let $\mathcal{B}$ be the point-set of birth times in the persistence diagram of $\mathrm{H}_d(\mathcal{K})$. Then, we have that*

$$\mathcal{D} = \{w(\sigma) : \sigma \in M\} \quad and \quad \mathcal{B} = \{w(\sigma) : \sigma \in \mathcal{F}^d \backslash M\},$$

where M is a $d-$minimal spanning acycle of $\mathcal{K}$ and $\mathcal{F}^d$ are the $d-$simplices of $\mathcal{K}$.

This result reveals a stronger connection between persistence diagrams and minimal spanning acycles than what is known in literature. If $\mathcal{K}$ is a weighted $d-$complex with $\beta_{d-1}(\mathcal{K}) = \beta_{d-2}(\mathcal{K}) = 0$ then, as a corollary of the above theorem, we obtain the following relation

$$(1.1) \quad L_{d-1} = \sum_i (D_i - B_i) = \sum_{\sigma \in M_d} w(\sigma) - \sum_{\sigma \in \mathcal{F}^{d-1} \backslash M_{d-1}} w(\sigma) = w(M_d) + w(M_{d-1}) - w(\mathcal{F}^{d-1}).$$

---

[3]Again, this terminology will be explained in Section 2.1.2 and is required only for non-injective weight functions.
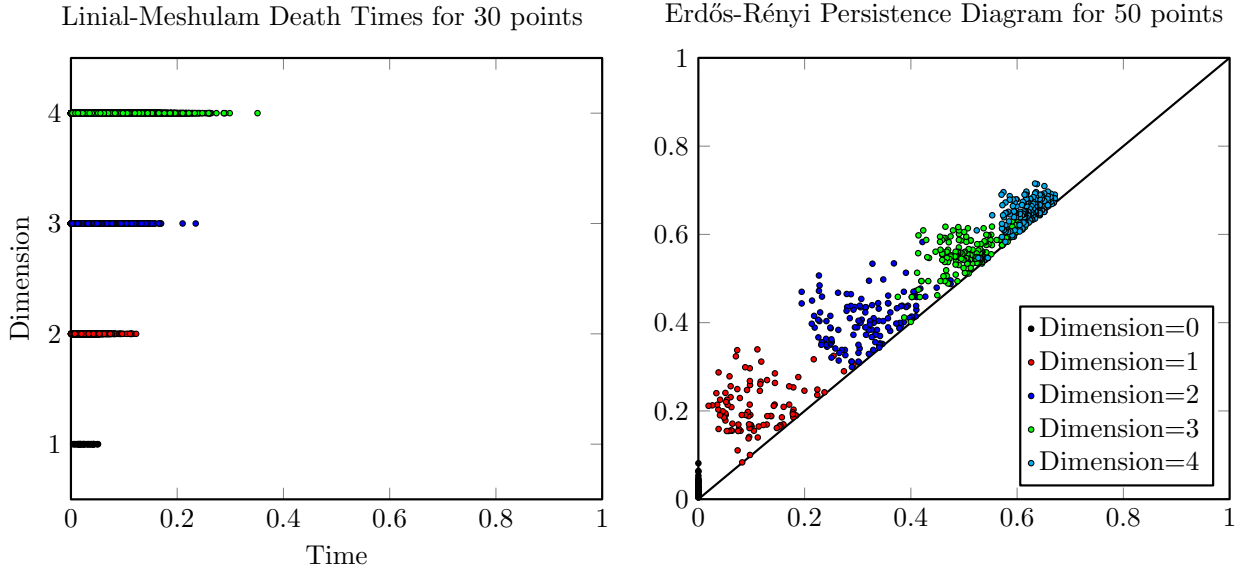
FIGURE 1. (Left) The death times corresponding to the uniformly weighted random $d-$complex built on 30 points in different dimensions. (Right) The persistence diagram for an Erdős-Rènyi clique complex for 50 points.

For $d = 1$ (assuming $\mathcal{K}^0 \subset w^{-1}(0)$), the above relation is well known and, for $d \geq 2$, this relation was derived recently in [HS17, Theorem 1.1] using different techniques. This latter paper and, in particular, their derivation of (1.1) served as our stimulus to investigate minimal spanning acycles. Apart from its striking simplicity, we believe Theorem 1.3 can be useful in studying either of them using the other. In fact, this result is frequently used in this paper. Much of the complexity in understanding persistent homology arises from the pairing of birth and death times. The above result is useful in understanding death or birth times individually and, in certain cases, this shall yield useful information (e.g., lifetime sum) even without the knowledge of the pairings. The proof of the above theorem and some of its consequences can be found in Section 3.3.

1.3. **Stability of birth and death times.** Stability results (e.g. [EH10, Section VIII.2], [CCSG+09, CDSO14, CSEH07, CSEHM10]) are an important cog in the wheel of topological data analysis and provide a theoretical justification for the robustness of persistent homology. While L_∞ stability (or *bottleneck stability*) is the most standard form of stability proven for persistence diagrams, L_p stability for $p \geq 0$ requires restrictive assumptions that are not widely applicable. Using simplicial version of Kruskal's algorithm and the correspondence (Theorem 1.3), we prove the following stability result separately for the birth and death times with minimal assumptions. The usefulness of this stability result will become apparent in Section 1.4

Theorem 1.4. *Let $\mathcal{K}$ be a finite complex with two weight functions f, f'; both of which induce a filtration on $\mathcal{K}$. Let $\mathcal{B}_f = \{B_i\}$ and $\mathcal{D}_f = \{D_i\}$ be the respective birth and death times in the*

$H_d(\cdot)$ *and* $H_{d-1}(\cdot)$ *persistence diagrams of* f. *Similarly, define* $\mathcal{B}_{f'} = \{B'_i\}$ *and* $\mathcal{D}_{f'} = \{D'_i\}$ *w.r.t.* f'. *Let* Π_D *be the set of bijections from* $\mathcal{D}_f$ *to* $\mathcal{D}_{f'}$ *and, similarly, let* Π_B *be the set of bijections from* $\mathcal{B}_f$ *to* $\mathcal{B}_{f'}$. *Then, for any* $p \in \{0, \ldots, \infty\}$,

$$\max\{\inf_{\pi \in \Pi_D} \textstyle\sum_i |D_i - \pi(D_i)|^p, \inf_{\pi \in \Pi_B} \sum_i |B_i - \pi(B_i)|^p\} \leq \sum_{\sigma \in \mathcal{F}^d} |f(\sigma) - f'(\sigma)|^p,$$

For $p = \infty$ *and a sequence* $\{x_i\}_{i \geq 1}$, *in the usual manner,* $\sum_i |x_i|^p$ *should be read as* $\sup_i |x_i|$.

As part of the proof (see Section 3.4) of the above stability result, we show that *on a fixed simplicial complex* changing weights of m ($m \geq 1$) faces can change at most m death times and m birth times by the difference between the weights on the faces[4]. One might suspect that the L_∞ stability in the above theorem can be deduced from the bottleneck stability of persistence diagrams by a projection argument. This is, however, not the case due to the fact that the diagonal plays a special role in the definition of bottleneck stability of persistence diagrams, but for point processes on $\mathbb{R}$ there is no equivalent to the diagonal.

1.4. **Weighted random complexes.**

Having offered a teaser to our deterministic results, we now turn to a preview of the probabilistic results. Whereas there is a rich recent literature on deterministic aspects of spanning acycles (see in Section 1.1) and random complexes (see below), the literature is sparser on weighted random complexes or random minimal spanning acycles. The probabilistic model of interest to us is the one introduced by Linial and Meshulam [LM06] and then extended by Meshulam and Wallach [MW09]. This model, called the random d−complex and denoted by $Y_{n,d}(p)$, consists of all faces on n vertices (i.e., ground set $V = [n] := \{1, \ldots, n\}$) with dimension at most $(d-1)$ and each d−face is included with probability p independently. $Y_{n,1}(p)$ is the classical Erdős-Rènyi graph on n vertices with edge-connection probability p. Like Erdős-Rènyi graph is a mean-field model of pairwise interactions, the random d−complex can be considered as a model of higher-order interactions. This model has spawned a rich literature in the recent years [LM06, MW09, CCFK12, CF15a, CF15b, LP16]. Although we focus on the random d−complex, we alert the reader of the existence of a richer theory of random complexes and topological data analysis [Car14, Kah14b, BK14, Kah17, CFK12].

The focus of many studies on random d−complexes has been the two non-trivial Betti numbers of the complex: $\beta_{d-1}(\cdot)$ and $\beta_d(\cdot)$. The starting part of our study is the following fine phase transition result for $\beta_{d-1}(Y_{n,d}(p))$.

Lemma 1.5. [Ste69], [KP14, Theorem 1.10] *Fix* $d \geq 1$. *Consider* $Y_{n,d}(p_n)$ *with*

$$(1.2) \qquad p_n = \frac{d \log n + c - \log(d!)}{n}$$

for some fixed $c \in \mathbb{R}$. *Then, as* $n \to \infty$, $\beta_{d-1}(Y_{n,d}(p_n)) \Rightarrow Poi(e^{-c})$, *where* $Poi(\lambda)$ *stands for the Poisson random variable with mean* λ *and* $\Rightarrow$ *denotes convergence in distribution.*

The proof of this result proceeds as follows: First, it is shown that $N_{n,d-1}(p_n) \Rightarrow \mathrm{Poi}(e^{-c})$, where $N_{n,d-1}(p)$ denotes the number of isolated $(d-1)$−faces in $Y_{n,d}(p)$. Then, for p_n as

---

[4]By fixing the underlying space, we can ensure the cardinalities of birth and death times remain the same.

chosen, it is established that $N_{n,d-1}(p_n)$ completely determines the behaviour of the $(d-1)$−th Betti number (see also Appendix C). Building upon this relation, one also has that $\mathbb{P}\{\beta_{d-1}(Y_{n,d}(p_n)) = 0\} \to 1$ if $np_n - d\log n \to \infty$ and $\mathbb{P}\{\beta_{d-1}(Y_{n,d}(p_n)) = 0\} \to 0$ if $np_n - d\log n \to -\infty$. These were proven by Erdős and Rényi [ER59] in 1959 for $d = 1$, much later by Linial and Meshulam [LM06] in 2006 for $d = 2$ and shortly thereafter in 2009 for $d \geq 3$ by Meshulam and Wallach [MW09].

One of the goals of this paper is to generalize Lemma 1.5 first to the level of persistence diagrams and then to that of minimal spanning acycles of randomly weighted d−complexes. Before providing the actual statements, we give a formal definition of these weighted complexes.

Definition 1.6. *Let $d \geq 1$ be some integer. Consider n vertices and let $\mathcal{K}_n^d$ be the complete d−skeleton on them. Let $\phi' : \mathcal{K}_n^d \to [0,1]$ be the weight function with the following properties:*
  (1) $\phi'(\sigma) = 0$ for $\sigma \in \bigcup_{i=0}^{d-1} \mathcal{F}^i$, and
  (2) $\phi'(\sigma) = \phi(\sigma) + \epsilon_n(\sigma)$ for $\sigma \in \mathcal{F}^d$.
Here, $\{\phi(\sigma) : \sigma \in \mathcal{F}^d\}$ are real valued i.i.d. random variables with (cumulative) distribution function $\mathscr{F} : \mathbb{R} \to [0,1]$ perturbed respectively by $\{\epsilon_n(\sigma) : \sigma \in \mathcal{F}^d\}$. The latter are another set of real valued random variables not necessarily identically distributed or independent of each other or $\phi(\sigma)$'s. The randomly weighted d−complex $\mathcal{L}'_{n,d}$ is the simplicial complex $\mathcal{K}_n^d$ weighted by ϕ'. Associated with $\mathcal{L}'_{n,d}$ is the canonical simplicial process given by the filtration $\{\mathcal{L}'_{n,d}(t) : t \in \mathbb{R}\}$, where $\mathcal{L}'_{n,d}(t) = \{\sigma : \phi'(\sigma) \leq t\}$.

For ease of use, we shall write $\sigma \in \mathcal{L}'_{n,d}$ to mean $\sigma \in \mathcal{K}_n^d$. Similarly, $\mathcal{F}^i(\mathcal{L}'_{n,d})$ shall mean $\mathcal{F}^i(\mathcal{K}_n^d)$ and so on. Finally, let $\|\epsilon_n\|_\infty := \max_{\sigma \in \mathcal{F}^d(\mathcal{L}'_{n,d})} |\epsilon_n(\sigma)|$.

Our key result concerning randomly weighted d−complexes is that if the perturbations decay sufficiently fast, then suitably scaled point processes related to the nearest face distances, weights of the faces in the d−minimal spanning acycle, and death times in the associated persistence diagram all converge to the same inhomogeneous Poisson point process. The proof crucially relies upon Theorems 1.3 and 1.4.

Formally, we consider the following *three scaled point processes* on $\mathbb{R}$.

  (1) (Extremal) nearest face distances, i.e., $\mathscr{P}_{n,d}^{C'} := \{n\mathscr{F}(C'(\sigma)) - d\log n + \log(d!) : \sigma \in \mathcal{F}^{d-1}\}$, where, for $\sigma \in \mathcal{F}^{d-1}$,

$$(1.3) \qquad C'(\sigma) := \min_{\tau \in \mathcal{F}^d, \tau \supset \sigma} \phi'(\tau).$$

  (2) (Extremal) death times in H_{d-1}, i.e., $\mathscr{P}_{n,d}^{D'} := \{n\mathscr{F}(D_i') - d\log n + \log(d!)\}$, where $\{D_i'\}$ is the set of death times in the persistence diagram of H_{d-1} (see Definition 2.2).
  (3) (Extremal) face weights in M', i.e., $\mathscr{P}_{n,d}^{M'} := \{n\mathscr{F}(\phi'(\sigma)) - d\log n + \log(d!) : \sigma \in M'\}$, where M' is a d−minimal spanning acycle in $\mathcal{L}'_{n,d}$ (see (2.9)).

Observe that the scaling used in the definitions of each of $\mathscr{P}_{n,d}^{C'}$, $\mathscr{P}_{n,d}^{D'}$, and $\mathscr{P}_{n,d}^{M'}$ pushes quantities less than the $d\log(n)/n$ threshold to $-\infty$, asymptotically. In that sense, asymptotically, the three processes only consider the extremal values, i.e., those that are above this threshold. The reason for transforming weights, as will be seen below, is that it yields a

limiting point process independent of $\mathscr{F}$. If we think of the weighted complex as a dynamic complex with simplices being added at times equal to their weights, then the transformation by $\mathscr{F}$ is nothing but a time-change.

At first glance, these are three distinct point processes on $\mathbb{R}$ and there are no obvious reasons why they ought to be connected. However, by applying Theorem 1.3, we get $\mathscr{P}_{n,d}^{M} = \mathscr{P}_{n,d}^{D}$ and, from Corollary 3.13 that we establish later, it follows that $\mathscr{P}_{n,d}^{C'} \subset \mathscr{P}_{n,d}^{M'}$. A natural guess based on this would be that a similar relation holds amongst the three processes asymptotically as well. Surprisingly, the below result shows that the three processes in fact have the same asymptotic behaviour.

Theorem 1.7. *Suppose that $\mathscr{F}$ is Lipschitz continuous. If $n\|\epsilon_n\|_\infty \to 0$ in probability, then each of $\mathscr{P}_{n,d}^{C'}, \mathscr{P}_{n,d}^{D'},$ and $\mathscr{P}_{n,d}^{M'}$ converges vaguely in distribution to $\mathscr{P}_{poi}$, where $\mathscr{P}_{poi}$ is the Poisson point process with intensity $e^{-x}\mathrm{d}x$ on $\mathbb{R}$.*

Since the $(d-1)-$faces have zero weights, the birth times in the persistence diagram of H_{d-1} are all zero. Hence, if $\|\epsilon_n\|_\infty = 0$ and $\mathscr{F}$ is the distribution function of $U[0,1]$, then $\mathscr{P}_{n,d}^{D'}((c,\infty)) = \beta_{d-1}(Y_{n,d}(p_n))$ for p_n and c as in Lemma 1.5. Thus, a point process convergence for $\mathscr{P}_{n,d}^{D'}$ in this special case implies Lemma 1.5 as a corollary. This and more follows from the above result. See Figure 1(a) for simulations of $\mathscr{P}_{n,d}^{D'}$ for $d = 1, 2, 3, 4$ in the above special case.

To the best of our knowledge, a point process convergence result as above is not known even for complete graphs with i.i.d. uniform $[0, 1]$-weights, which might be considered as a mean-field model for random metric spaces. For random geometric graphs, such a point process convergence result for extremal edge weights of the minimal spanning tree was proven in [Pen97, HR05]. These results were important to understand the connectivity of random geometric graphs. However, reversing the scenario, we have gone from results on connectivity (i.e., $\mathrm{H}_k(\cdot)$ persistence diagrams) to those for minimal spanning acycles.

The above weak convergence result along with the continuous mapping theorem yields asymptotics of various statistics of $\mathscr{P}_{n,d}^{D'}$. Our result could be useful in deriving asymptotics for extremes of other summary statistics of persistence diagrams such as persistence landscapes [Bub15], homological scaffolds [PET$^+$14] or accumulative persistence function [BM19].

As for our proof, we first deal with the case when $\mathscr{F}$ is the distribution function of $U[0, 1]$ and $\epsilon_n(\sigma) \equiv 0$. We use the factorial moment method to show convergence of the first point process and then use cohomology theory to show that this is a good enough approximation for the second point process. This yields convergence of the second point process . Finally, this along with Theorem 1.3 gives the convergence of the third point process (see Section 4.1). This approach is inspired by those of [LM06, MW09, KP14]. Next, we extend this result to the case of the more general i.i.d. weights. Finally, we complete the proof of Theorem 1.7 by using our stability result (Theorem 1.4) as well as showing that the topology of bottleneck distance between Radon counting measures is stronger than vague topology (see Section 4.2).

We now present one more powerful consequence of our stability result. While it is believed that introducing weak dependencies between the random variables should not affect the

asymptotics, it is often difficult to prove such a statement rigorously. As we again illustrate, our stability result helps bridge this gap in certain situations. In particular, given an arbitrary random complex, it enables one to translate certain limit theorems to noisy variants of this complex once the same has been shown in the noiseless setting.

Consider $\mathcal{L}'_{n,d}$ from Definition 1.6 and suppose that $\mathscr{F}$ is the distribution function of $U[0,1]$. Further, let $\text{Dgm}(\mathcal{K}^d_n, \phi') = \{(0, D'_i)\}$ be the H_{d-1} persistence diagram. Let us define the (weighted) lifetime sums for $\alpha \geq 0$ as

$$(1.4) \qquad (L'_{n,d-1})^\alpha = \sum_i (D'_i)^\alpha.$$

To begin with, suppose that $\|\epsilon_n\|_\infty = 0$ for all $n \geq 1$. In such a case, we denote the weighted random complex by $\mathcal{U}_{n,d}$ and the corresponding lifetime sum by $L_{n,d}$. Then, it follows from a remarkable recent result by Hino-Kanazawa ([HK19, Theorem 4.11]) that, for $\alpha > 0$

$$(1.5) \qquad n^{-(d-\alpha)}\mathbb{E}[(L_{n,d-1})^\alpha] \to I^\alpha_{d-1},$$

where I^α_{d-1} is an explicitly defined constant (see [HK19, (4.10)] for the definition of constants and [HK19, Section 4.4] for more concrete expressions). In the special case of $d = 1, \alpha = 1$, this is the famed result of Frieze [Fri85] for random minimal spanning trees with $I^1_0 = \zeta(3)$ where ζ is the Riemann-zeta function. Further, I^p_0 for $p \in \{1, 2, \ldots\}$ are shown to be linear combinations of $\zeta(3), \zeta(4)$, etc. Using our stability result, we now extend this result to the noisy case. The proof can be found in Section 4.2.

Corollary 1.8. *Fix a $p, d \in \{1, 2, \ldots\}$. Assume that $\mathscr{F}$ is the distribution function of $U[0,1]$ and that $\sup_{\sigma \in \mathcal{F}^d} \mathbb{E}[|\epsilon_n(\sigma)|^p] = o(n^{-(p+1)})$. Then, with I^p_{d-1} as defined in (1.5), we have that*

$$n^{-(d-p)}\mathbb{E}[(L')^p_{n,d-1}] \to I^p_{d-1}.$$

2. Preliminaries

We describe here the basic notions of simplicial homology, persistent homology, and point processes. We remark that, in an earlier version of the paper (see [STY17, Appendix B]), we have rephrased our topological notions in the language of matrices for an alternative and computationally convenient viewpoint.

2.1. **Topological notions.** We point out that we shall always choose our coefficients from a field $\mathbb{F}$. In this regard, 0 stands for additive identity, 1 stands for multiplicative identity and -1 for the additive inverse of 1. An often convenient choice in computational topology is $\mathbb{F} = \mathbb{Z}_2$ in which case $1 = -1$.

2.1.1. *Simplicial Homology.* For a good introduction to algebraic topology, see [Hat02], and for simplicial complexes and homology, see [EH10, Mun84].

Let $\mathcal{K}$ be a simplicial complex (see Definition 1.1). We assume throughout that all our simplicial complexes are defined over a finite set V. The 0-faces of $\mathcal{K}$ are also called as *vertices.* When obvious, we shall omit the reference to the underlying complex $\mathcal{K}$ in the

notation. A d−simplex σ is often represented as $[v_0, \ldots, v_d]$ to explicitly indicate the subset of V generating the simplex σ.

An *orientation* of a d−simplex is given by an ordering of the vertices and denoted by $[v_0, \ldots, v_d]$. Two orderings induce the same orientation if and only if they differ by an even permutation of the vertices. In other words, for a permutation π on $[d]$,

$$[v_0, \ldots, v_d] = (-1)^{sgn(\pi)}[v_{\pi(0)}, \ldots, v_{\pi(d)}],$$

where $sgn(\pi)$ denotes the sign of the permutation π. We assume that each simplex in our complex is assigned a specific orientation (i.e., ordering).

Let $\mathbb{F}$ be a field. A simplicial d−chain is a formal sum of oriented d−simplices $\sum_i c_i \sigma_i, c_i \in \mathbb{F}$. The free abelian group generated by the d−chains is called the d−th chain group and is denoted by $C_d(\mathcal{K})$. Formally,

$$C_d(\mathcal{K}) := \left\{ \sum_i c_i \sigma_i : c_i \in \mathbb{F}, \sigma_i \in \mathcal{F}^d(\mathcal{K}) \right\}.$$

Clearly, $C_d(\mathcal{K})$ is a $\mathbb{F}$-vector space. We shall set $C_{-1} = \mathbb{F}$ and $C_d = 0$ for $d = -2, -3, \ldots$. For a vector space, let $\beta(\cdot)$ denote its rank. Thus, $\beta(C_d) = f_d$ for $d \geq 0$. For $d \geq 1$, we define the boundary operator $\partial_d : C_d \longrightarrow C_{d-1}$ first on each d−simplex using

$$\partial_d([v_0, \ldots, v_d]) = \sum_{i=0}^{d}(-1)^i[v_0, \ldots, \hat{v_i}, \ldots, v_d],$$

and then extend it linearly on C_d. Above, $\hat{v_i}$ denotes that v_i is to be omitted. ∂_0 is defined by setting $\partial_0([v]) = 1$ for all $v \in \mathcal{F}^0$. It can be verified that ∂_d is a linear map of vector spaces and more importantly that $\partial_{d-1} \circ \partial_d = 0$ for all $d \geq 1$, i.e., boundary of a boundary is zero. When the context is clear, we will drop the dimension d from the subscript of ∂_d.

Note that the free abelian group of d−chains is defined only using $\mathcal{F}^d(\mathcal{K})$. When we use a subset $S \subset \mathcal{F}^d(\mathcal{K})$ of d−faces rather than the entire collection of d−faces to generate the free abelian group, we shall use $C_d(S)$ to denote the corresponding free abelian (sub)group of d−chains. In other words, $C_d(S) = C_d(\mathcal{K}^{d-1} \cup S)$.

The d−th *boundary space* denoted by B_d is im ∂_{d+1} and the d−th *cycle space* Z_d is ker ∂_d. Elements of Z_d are called *cycles* or d−*cycles* to be more specific. The d−dimensional (reduced)[5] homology group is then defined as the quotient group

(2.1) $$\mathrm{H}_d = \frac{Z_d}{B_d}.$$

Again, since we are working with field coefficients, B_d, Z_d and H_d are all $\mathbb{F}$−vector spaces. The bases of these vector spaces form a matroid ([Oxl03, Wel76]). This implies that certain concepts such as the span of a generating set and properties such as the exchange property automatically hold. While it is not necessary for understanding our results, a familiarity with matroids is helpful.

---

[5]Reduced is used to refer to the convention that $C_{-1} = \mathbb{F}$ instead of $C_{-1} = 0$.

The $d-$th *Betti number* of the complex $\beta_d(\mathcal{K})$ is defined to be the rank of the vector space H_d. Respectively, let $b_d(\mathcal{K}) := \beta(B_d)$ and $z_d(\mathcal{K}) := \beta(Z_d)$ denote the ranks of the $d-$th boundary and $d-$th cycle spaces, respectively. Thus, we have that $\beta_d = z_d - b_d$. Note that we drop the adjective reduced henceforth, but all our homology groups and Betti numbers are indeed reduced ones. Some authors prefer to use $\tilde{H}_d$ and $\tilde{\beta}_d$ to denote reduced homology groups and Betti numbers respectively, but we refrain from doing so for notational convenience. However, under such a notation, we note that $\beta_d - \tilde{\beta}_d = 1[d = 0]$. This gives an easy way to translate results for reduced Betti numbers to Betti numbers and vice-versa. We denote the Euler-Poincaré characteristic by χ and the Euler-Poincaré formula holds as follows:

$$(2.2) \qquad \chi(\mathcal{K}) = \sum_{j=0}^{\infty} (-1)^j f_j(\mathcal{K}) = 1 + \sum_{j=0}^{\infty} (-1)^j \beta_j(\mathcal{K}).$$

An important property of homology groups that is often of use is the following: If $\mathcal{K}_1, \mathcal{K}_2$ are two complexes such that the function $h : \mathcal{K}_1^0 \to \mathcal{K}_2^0$ is a *simplicial map* (i.e., $\sigma_1 = [v_0, \ldots, v_d] \in \mathcal{K}_1$ implies that $h(\sigma_1) = [h(v_0), \ldots, h(v_d)] \in \mathcal{K}_2$ for all $d \geq 0$), then there exists an homomorphism $h_* : \mathrm{H}_d(\mathcal{K}_1) \to \mathrm{H}_d(\mathcal{K}_2)$ called the *induced homomorphism* between the homology groups. If $\mathcal{K}_1 \subseteq \mathcal{K}_2$, then a natural simplicial map is the inclusion map from $\mathcal{K}_1$ to $\mathcal{K}_2$. The case of multiple inclusions now brings us to persistent homology.

2.1.2. *Persistent Homology.* A *filtration* of a simplicial complex $\mathcal{K}$ is a sequence of subcomplexes $\{\mathcal{K}(t) : t \in \overline{\mathbb{R}}\}$ satisfying

$$\emptyset = \mathcal{K}(-\infty) \subseteq \mathcal{K}(t_1) \subseteq \mathcal{K}(t_2) \subseteq \mathcal{K}(\infty) = \mathcal{K}$$

for all $-\infty < t_1 \leq t_2 < \infty$. Put differently, the filtration $\{\mathcal{K}(t) : t \in \overline{\mathbb{R}}\}$ describes how to build $\mathcal{K}$ by adding collections of simplices at a time. For more complete introduction and survey of persistent homology, see [EH10, Car09, Car14].

We now describe the natural filtration associated with *weighted simplicial complexes*. Consider a simplicial complex $\mathcal{K}$ weighted by $w : \mathcal{K} \to \mathbb{R}$ satisfying $w(\sigma) \leq w(\tau)$, whenever $\sigma, \tau \in \mathcal{K}$ and $\sigma \subset \tau$. Functions having this property are called *monotonic functions* in [EH10, Chapter VIII]. As w is monotone, $\{\mathcal{K}(t) : t \in \overline{\mathbb{R}}\}$ with $\mathcal{K}(t) := w^{-1}(-\infty, t]$ forms a *sublevel set* filtration of $\mathcal{K}$. Further note that w induces a partial order on the faces of $\mathcal{K}$. Assuming axiom of choice, this partial order can always be extended to a total order [Szp30]. Let $<_l$ denote one such total order. We make the standing assumption that for a given weight function w, the same total order $<_l$ is chosen and used throughout the paper.

One can now view the above sublevel set filtration associated with $(\mathcal{K}, w)$ in a dynamic fashion: as the parameter t evolves over $\overline{\mathbb{R}}$, $\mathcal{K}$ gets built one face at a time respecting the total order $<_l$. In this way, with the addition of faces, the topology of $\mathcal{K}$ evolves. Clearly

$$(2.3) \qquad \mathcal{K}(\sigma^-) := \{\sigma_1 \in \mathcal{K} : w(\sigma_1) <_l w(\sigma)\}$$

denotes the complex right before the face σ is to be added. Thus, given a monotonic weight function w, we can construct a filtration with respect to the chosen total order $<_l$. We

shall call this filtration the *canonical filtration* associated with the total order $<_l$ or a *linear filtration* of the weight function w.

To track the changes in topology, akin to the definition for homology given in (2.1), we define the (t_1, t_2)-*persistent homology group* as the quotient group

$$\mathrm{H}_d^{t_1, t_2} = \frac{Z_d^{t_1}}{Z_d^{t_1} \cap B_d^{t_2}}, \qquad t_1 \leq t_2.$$

The information for all pairs (t_1, t_2) can be encoded in a unique interval representation called a *persistence barcode* [ZC05] or equivalently a *persistence diagram* [CSEH07]. Before giving the definition, we first note that for a finite simplicial complex endowed with a total ordering $<_l$, we can reindex the filtration by assigning a natural number to each simplex. We refer to this as a *discrete weight* $w_{\mathbb{N}}$ corresponding to the monotonic weight function w i.e., $w_{\mathbb{N}}(\sigma) < w_{\mathbb{N}}(\tau)$ iff $w(\sigma) <_l w(\tau)$. Note that there is a bijection between total orders $<_l$ and weight function $w_{\mathbb{N}}$. Thus, the discrete weight has a natural, well-defined projection π back to the original function values,

$$\pi(i) \mapsto w(\sigma) | w_{\mathbb{N}}(\sigma) = i$$

Definition 2.1. *Given a simplicial complex $\mathcal{K}$ with a monotonic function w and the corresponding discrete weight $w_{\mathbb{N}} : \mathcal{K} \to \mathbb{N}$, the $d-$th persistence diagram $\mathrm{Dgm}(\mathcal{K}, w_{\mathbb{N}})$ is the multiset of points in the extended grid $\overline{\mathbb{N}}^2$ such that the each point (i, j) in the diagram represents a distinct class (i.e., a topological feature) in $\mathrm{H}_d(w_{\mathbb{N}}^{-1}(-\infty, t])$ for all $t \in [i, j)$ and is not a class in $t \notin [i, j)$. The persistence diagram $\mathrm{Dgm}(\mathcal{K}, w)$ is then defined as the projection of the multiset of points under π, i.e., $(i, j) \in \mathrm{Dgm}(\mathcal{K}, w_{\mathbb{N}})$ iff $(\pi(i), \pi(j)) \in \mathrm{Dgm}(\mathcal{K}, w)$.*

This differs from the typical definition of a persistence diagram, where the existence and uniqueness of the persistence diagram is defined in terms of an algebraic decomposition into interval modules see [CDSGO16, CB15]. For technical reasons, this approach generally discards the points on the diagonal, i.e., topological features which are both born and die at time t. In the above definition, the total order guarantees that there are no points on the diagonal of the discrete filtration. However, since we deal with the restricted setting of piece-wise constant functions on finite simplicial complexes, we do not lose any information; indeed, we keep more of the chain level information. We then transform the persistence diagram back to the original monotone function. After the transformation, points may lie on the diagonal and, as we shall see, we do require these points.

Our definition is used implicitly in [ZC05], which first identified the algebraic decomposition as a consequence of the structure theorem of finitely generated modules over a principle ideal domain. This applies in this setting since the homology groups of finite simplicial complexes are always finitely generated. Therefore, we could have equivalently defined the diagram using the decomposition directly as done in Corollary 3.1 in [ZC05], as the modified Smith Normal Form of the boundary operator [SVJ13]. We believe that our definition is more accessible to a non-algebraic audience and is included for completeness. But more important for us are birth and death times defined below.

Definition 2.2. *The* death times *(respectively* birth times*) of the filtration associated with* $(\mathcal{K}, w)$ *are equal to the multiset of y-coordinates (x-coordinates) of points in* $\mathrm{Dgm}(\mathcal{K}, w)$.

We now discuss the notion of negative and positive faces which are vital to our proofs.

Lemma 2.3. *([DE93, Section 3]) Let* $\mathcal{K}$ *be a simplicial complex on vertex set* V *and* $\sigma \subset V$ *be a set of cardinality* $d+1$ *in* V *for some* $d \geq 0$. *Additionally, assume that* $\sigma \notin \mathcal{K}$ *but* $\partial \sigma \in C_{d-1}(\mathcal{K})$. *Then,* $\beta_j(\mathcal{K} \cup \sigma) = \beta_j(\mathcal{K})$ *for all* $j \notin \{d-1, d\}$. *Further, one and only one of the following two statements hold:*

(1) $\beta_{d-1}(\mathcal{K} \cup \sigma) = \beta_{d-1}(\mathcal{K}) - 1$ *and* $\beta_d(\mathcal{K} \cup \sigma) = \beta_d(\mathcal{K})$.
(2) $\beta_d(\mathcal{K} \cup \sigma) = \beta_d(\mathcal{K}) + 1$ *and* $\beta_{d-1}(\mathcal{K} \cup \sigma) = \beta_{d-1}(\mathcal{K})$.

From the definition of the cycle and boundary spaces, the above two numbered statements can be interpreted equivalently in the following manner which shall be useful for us:

$$(2.4) \quad \beta_{d-1}(\mathcal{K} \cup \sigma) = \beta_{d-1}(\mathcal{K}) - 1 \quad \Leftrightarrow \quad b_{d-1}(\mathcal{K} \cup \sigma) = b_{d-1}(\mathcal{K}) + 1 \Leftrightarrow \partial \sigma \notin \partial(C_d(\mathcal{K})),$$

$$(2.5) \quad \beta_d(\mathcal{K} \cup \sigma) = \beta_d(\mathcal{K}) + 1 \quad \Leftrightarrow \quad z_d(\mathcal{K} \cup \sigma) = z_d(\mathcal{K}) + 1 \Leftrightarrow \partial \sigma \in \partial(C_d(\mathcal{K})).$$

Definition 2.4 (Positive and Negative faces). *Let* $\mathcal{K}$ *be a complex with vertex set* V *and* $\sigma \subset V$ *be a set of cardinality* $d+1$ *for some* $d \geq 0$. *Further assume that* $\sigma \notin \mathcal{K}$ *but* $\partial \sigma \in C_{d-1}(\mathcal{K}) \neq 0$. *Such a* σ *is called a* negative face *w.r.t.* $\mathcal{K}$ *if* $\beta_{d-1}(\mathcal{K} \cup \sigma) = \beta_{d-1}(\mathcal{K}) - 1$, *it is called a* positive face *if it is not negative, i.e.,* $\beta_d(\mathcal{K} \cup \sigma) = \beta_d(\mathcal{K}) + 1$.

This is useful for understanding how the topology evolves in the filtration associated with w (recall (2.3)). If σ is a $d-$face, then Lemma 2.3 shows that the relationship between the topology of the setup before and after addition of σ is as follows: (i) $\beta_j(\mathcal{K}(\sigma^-) \cup \sigma) = \beta_j(\mathcal{K}(\sigma^-))$ for all $j \notin \{d, d-1\}$, (ii) one and exactly one of the following is true:

$$(2.6) \qquad\qquad \beta_{d-1}(\mathcal{K}(\sigma^-) \cup \sigma) = \beta_{d-1}(\mathcal{K}(\sigma^-)) - 1$$

or

$$(2.7) \qquad\qquad \beta_d(\mathcal{K}(\sigma^-) \cup \sigma) = \beta_d(\mathcal{K}(\sigma^-)) + 1.$$

As in Definition 2.4, when (2.6) holds (respectively (2.7) holds) σ will be called a negative face (positive face) w.r.t. the natural filtration of $(\mathcal{K}, w)$. We emphasize that the total order $<_l$ uniquely determines the label of faces as either positive or negative. The above discussion can be neatly converted to an algorithm to generate birth and death times of the persistence diagram with respect to a given linear filtration of the weight function w.

The above algorithm is a simplification of the persistence algorithm in [ELZ02, Fig. 5] which also used negative and positive simplices. The simplification in our algorithm essentially lies in turning a blind eye to the information about the pairing between the birth and death times. The equivalence of negative faces with death times (and hence positive faces with birth times) was established in [ZC05, Fig. 9]. These algorithms extended the incremental algorithm for computing Betti numbers in [DE93]. We summarize the algorithm, especially for ease for future reference, as follows : Let σ be a $d-$face in $\mathcal{K}$.

$$(2.8) \qquad w(\sigma) \in \mathcal{D}_{d-1} \Leftrightarrow (2.6) \text{ holds or, alternatively, } w(\sigma) \in \mathcal{B}_d \Leftrightarrow (2.7) \text{ holds}$$

Algorithm 1 Incremental Persistence Algorithm

Input: $\mathcal{K}, w$
Main Procedure:
$F = \mathcal{F}$ (set of all faces in $\mathcal{K}$)
while $F \neq \emptyset$

- remove a face σ with minimum weight (w.r.t $<_l$) from F. Set $d = dim(\sigma)$.
- if σ is negative w.r.t. $\mathcal{K}(\sigma^-)$, i.e., if $\beta_{d-1}(\mathcal{K}(\sigma^-) \cup \sigma) = \beta_{d-1}(\mathcal{K}(\sigma^-)) - 1$ then add $w(\sigma)$ to $\mathcal{D}_{d-1}$ else add $w(\sigma)$ to $\mathcal{B}_d$.

Output $\mathcal{B}_d, \mathcal{D}_d$, for all $d \geq 0$.

We end this subsection reiterating a remark with respect to our proofs.

Remark 2.5. *As already explained, if $\mathcal{K}$ is a weighted simplicial complex, there is a unique total ordering of the faces if the weight function is injective. Otherwise, it is only a partial ordering. However, this partial ordering can be extended to a total order. This correspondence between monotonic weights and total orders shall be used to simplify many of our proofs. We shall often prove many statements for weighted simplicial complexes with unique weights and appeal to this correspondence in extending the proof to general monotonic weight functions. Equivalently, one can prove results for $w_\mathbb{N}$ and then use the natural projection π to obtain the corresponding result for monotonic weight function w.*

2.1.3. *Spanning acycles.* As made clear in the title, the other key object of our study is the *spanning acycle*, which has been already introduced in Definition 1.2. We now discuss the definition in more detail. Apart from being more restrictive than that in [HS17, Kal83], our definition differs from that of [HS17] in its use of field coefficients over integer coefficients. Clearly, in the case of $d = 1$, S is a *minimal spanning tree* on the graph $\mathcal{K}^1$. Strictly speaking, the above definition is that of a $d-$spanning acycle but since in most cases the dimension d will be clear from the context, we shall not always explicitly refer to the dimension d. Recall that for any $S \subseteq \mathcal{K}$, $w(S) = \sum_{\sigma \in S} w(\sigma)$ denotes the weight of S. Denoting the set of $d-$spanning acycles of $\mathcal{K}$ by $\mathcal{S}^d(\mathcal{K})$, $S_0 \in \mathcal{S}^d(\mathcal{K})$ is a *minimal spanning acycle* if

$$(2.9) \qquad\qquad w(S_0) = \min_{S \in \mathcal{S}^d(\mathcal{K})} w(S).$$

Spanning trees and more generally connectivity in the case of graphs can be extended in a multitude of ways to higher-dimensions. Betti numbers and acycles represent one possible (and indeed a very satisfying) way to generalize to higher dimensions. Another common generalization is via the notion of a *hypergraph*. In this context, one can define a hypergraph on a simplicial complex by considering all the faces as hyper-edges. We will not use hypergraph connectivity in this paper, but we only remark that studying the hypergraph connectivity of spanning acycles yields interesting results.

Remark 2.6. *We would like to highlight one more interpretation of the spanning acycles before continuing. As will no doubt be known or obvious to experts in the field, an alternative*

*view of **a spanning acycle is as a basis for the space of boundaries**. Indeed, Algorithm 1 maintains a basis and, for insertion, checks whether the boundary of a simplex is in the span of the current basis or not. If it is linearly independent, the simplex (or more accurately its weight) is added to the list of death times, otherwise it is added to the set of birth times.*

2.2. **Probabilistic notions.** We give here a brief introduction to point processes on $\mathbb{R}$. For a more detailed reading on weak convergence of point processes, we refer the reader to [Res13, Chapter 3]. Let $\mathcal{B}(\mathbb{R})$ be the Borel $\sigma-$algebra of subsets in $\mathbb{R}$.

A *point measure* on $\mathbb{R}$ is a map from $\mathcal{B}(\mathbb{R})$ to the set of natural numbers, i.e., it is a Radon (locally-finite) counting measure. A point measure m is represented as $m(\cdot) = \sum_{i=1}^{\infty} \delta_{x_i}(\cdot)$, for some countable but locally-finite collection of points $\{x_i\}$ in $\mathbb{R}$ and where $\delta_x(\cdot)$ denotes the delta measure at x. Alternatively, we define the support of the point measure m, denoted by $supp(m)$ as the multi-set $\{x_i\}$. A point measure is *simple* if $m(\{x\}) \leq 1$ for each $x \in \mathbb{R}$. Let $M_p(\mathbb{R})$ denote the set of all point measures on $\mathbb{R}$. Also, let $C_c^+(\mathbb{R})$ denote the set of all continuous, non-negative functions $f : \mathbb{R} \to \mathbb{R}$ with compact support. For $f \in C_c^+(\mathbb{R})$ and $m = \sum_{i=1}^{\infty} \delta_{x_i} \in M_p(\mathbb{R})$, define

$$(2.10) \qquad\qquad m(f) := \int_{\mathbb{R}} f \, dm = \sum_{i=1}^{\infty} f(x_i).$$

Let $m_n, m \in M_p(\mathbb{R})$. We will say that m_n converges *vaguely* to m, denoted $m_n \xrightarrow{v} m$, if for every $f \in C_c^+(\mathbb{R})$, $m_n(f) \to m(f)$. Using this notion of vague convergence, one defines the vague topology on $M_p(\mathbb{R})$. That is, a subset of $M_p(\mathbb{R})$ is vaguely closed if it includes all its limit points w.r.t. vague convergence. The sub-base for this topology consists of open sets of the form

$$\{m \in M_p(\mathbb{R}) : m(f) \in (s,t)\}$$

for $f \in C_c^+(\mathbb{R})$, $s, t \in \mathbb{R}$, $s < t$.

A *point process* $\mathscr{P}$ on $\mathbb{R}$ is a random variable taking values in the space $(M_p(\mathbb{R}), \mathcal{M}_p(\mathbb{R}))$ where $\mathcal{M}_p(\mathbb{R})$ denotes the Borel σ-algebra generated by the vague topology. A point process is called *simple* if $\mathscr{P}(\{x\}) \leq 1$ a.s. for all x, i.e., if it is supported on simple point measures. An oft-used example of a point process is the Poisson point process.

Definition 2.7. *Let $\mu : \mathbb{R} \to [0, \infty)$ be locally integrable ($\int_A \mu(x) dx < \infty$ for all bounded $A \subset \mathbb{R}$). A point process $\mathscr{P}$ on $\mathbb{R}$ is said to be a* Poisson point process *with intensity function μ if the following two properties hold.*

    (1) *For disjoint $A_1, \ldots, A_m \in \mathcal{B}(\mathbb{R})$, $\mathscr{P}(A_1), \ldots, \mathscr{P}(A_m)$ are independent.*
    (2) *For any $A \subseteq \mathbb{R}$, $\mathscr{P}(A)$ is a Poisson random variable with mean $\int_A \mu(x) dx$.*

Definition 2.8. *Let $\mathscr{P}_n, \mathscr{P}$ be point processes on $\mathbb{R}$, not necessarily defined on the same probability space. We will say that $\mathscr{P}_n$ converges weakly to $\mathscr{P}$, denoted $\mathscr{P}_n \Rightarrow \mathscr{P}$, if*

$$\mathbb{E}[f(\mathscr{P}_n)] \to \mathbb{E}[f(\mathscr{P})]$$

for all continuous and bounded $f : (M_p(\mathbb{R}), \mathcal{M}_p(\mathbb{R})) \to \mathbb{R}$. This is equivalent to saying

$$\lim_{n \to \infty} \mathbb{P}\{\mathscr{P}_n \in A\} = \mathbb{P}\{\mathscr{P} \in A\}$$

for all $A \in \mathcal{M}_p(\mathbb{R})$ such that $\mathbb{P}\{\mathscr{P} \in \partial A\} = 0$. Here ∂A denotes the boundary of A.

An alternative topology on $M_p(\mathbb{R})$ that arises naturally in computational topology is the so-called *bottleneck distance d_B*. Note that we require a modified definition for point measures in $\mathbb{R}$ rather than the more standard definition for persistence diagrams (e.g. [CCSG+09, EH10]).

Definition 2.9. *For $m_1, m_2 \in M_p(\mathbb{R})$*

$$d_B(m_1, m_2) := \inf_\gamma \sup_{x:supp(m_1)} |x - \gamma(x)|,$$

where the infimum is over all possible bijections $\gamma : supp(m_1) \to supp(m_2)$ between the multi-sets. If no bijection exists, set $d_B(m_1, m_2) = \infty$.

Though this is not a metric in the classical sense, taking $\min\{d_B, 1\}$ we obtain a metric on $M_p(\mathbb{R})$. More importantly, the topology induced by d_B and $\min\{d_B, 1\}$ are the same. We shall prove in Lemma A.2 that this topology is stronger than that of vague topology.

3. Minimal spanning acycles

Our main goal here is to derive Theorems 1.3 and 1.4. Additionally, we introduce several relevant combinatorial properties of (minimal) spanning acycles. To avoid tedium, we do not always single out the results for the case of minimal spanning tree, i.e., the $d = 1$ case; since these are classical results, one can refer to [CLR09, Wik16] for graph-theoretic (and expectedly simpler) proofs. Some of these results are direct consequences of the fact that the space of boundaries is a vector space and, hence, allows for a natural matroid to be defined. Others, such as Kruskal's algorithm are folklore, but are included here for completeness as they do not appear in the literature for acycles.

3.1. **Basic Properties.** Our first aim here is to show that a d-spanning acycle exists if and only if $\beta_{d-1}(\mathcal{K}) = 0$. We establish this via a series of results. As introduced in Definition 1.2, a maximal acycle is a natural analogous notion of a spanning forest, however note that we mainly focus on a spanning acycle in the paper. We begin by showing that if a spanning acycle exists for a complex $\mathcal{K}$, then $\beta_{d-1}(\mathcal{K}) = 0$. If S is a spanning acycle, then $\beta_{d-1}(\mathcal{K}^{d-1} \cup S) = 0$ by definition. That this extends to the complete skeleton follows from the following corollary of Lemma 2.3.

Corollary 3.1. *Let $\mathcal{K}$ be a simplicial complex with $S_1 \subset S_2 \subset \mathcal{F}^d$. Then, for any $j \geq d$,*

$$\beta_{d-1}(\mathcal{K}^{d-1}) \geq \beta_{d-1}(\mathcal{K}^{d-1} \cup S_1) \geq \beta_{d-1}(\mathcal{K}^{d-1} \cup S_2) \geq \beta_{d-1}(\mathcal{K}^d) = \beta_{d-1}(\mathcal{K}^j) = \beta_{d-1}(\mathcal{K}) \geq 0.$$

Proof. By Lemma 2.3, adding a d-simplex either increases β_d or decreases β_{d-1}. Since the inequalities only concern β_{d-1}, $S_1 \subset S_2 \subset \mathcal{F}^d$ implies the first 3 inequalities. The equalities $\beta_{d-1}(\mathcal{K}^d) = \beta_{d-1}(\mathcal{K}^j) = \beta_{d-1}(\mathcal{K})$ follow from the property of simplicial complexes that for

18

every simplex, all of its faces must be contained in the complex. Hence, adding higher than d-dimensional simplices cannot change β_{d-1}. $\qquad\square$

It remains to show that $\beta_{d-1}(\mathcal{K}) = 0$ implies the existence of a spanning acycle. We omit the case where $\mathcal{F}^d$ is empty as the $d-$spanning acycle is simply the empty set in this case. We begin by proving the following fact, which states that a positive simplex remains positive under simplex addition and a negative simplex remains negative under deletion. This is nothing but a restatement that the span of a basis is non-decreasing under the addition of elements. It however will be useful in the proof of correctness of Kruskal's algorithm.

Lemma 3.2. *Let $\mathcal{K}$ be a simplicial complex, $\sigma \in \mathcal{K}$ be a $d-$face, and let $S_1 \subseteq S_2 \subseteq \mathcal{F}^d$ be such that $\sigma \notin S_2$. If σ is a positive face w.r.t. $\mathcal{K}^{d-1} \cup S_1$, then σ is a positive face w.r.t $\mathcal{K}^{d-1} \cup S_2$. Conversely, if σ is a negative face w.r.t. $\mathcal{K}^{d-1} \cup S_2$, then σ is a negative face w.r.t. $\mathcal{K}^{d-1} \cup S_1$.*

Proof. If σ is a positive simplex w.rt. $\mathcal{K} \cup S_1$, then (2.5) implies $\partial\sigma \in \partial(C_d(\mathcal{K} \cup S_1))$. Since the $\partial(C_d(\mathcal{K} \cup S_1)) \subseteq \partial(C_d(\mathcal{K} \cup S_2))$, it follows from (2.5) that σ is a positive simplex for $\mathcal{K} \cup S_1$ as well. The second statement is simply the contrapositive, since a simplex must be either positive or negative. $\qquad\square$

The second fact we need is a characterization of positive simplices – if a set of k-simplices do not decrease β_{k-1}, then they are positive.

Lemma 3.3. *Let $S \subseteq \mathcal{F}^d$ be such that $\beta_{d-1}(\mathcal{K}^{d-1} \cup S) = \beta_{d-1}(\mathcal{K})$. Then any $\sigma \in \mathcal{F}^d \setminus S$ is a positive face w.r.t. $\mathcal{K}^{d-1} \cup S$. In particular, this holds when S is a maximal acycle.*

Proof. From Corollary 3.1, $\beta_{d-1}(\mathcal{K}^{d-1} \cup S \cup \sigma) = \beta_{d-1}(\mathcal{K})$ for any $\sigma \in \mathcal{F}^d \setminus S$. Hence, from Lemma 2.3, σ is positive w.r.t. $\mathcal{K}^{d-1} \cup S$. The case of maximal acycles follows from Definition 1.2. $\qquad\square$

Together, the above results imply that we can always find a simplex which will decrease the $(d-1)$-th Betti number whenever it is greater than that of the whole complex.

Lemma 3.4. *Let $\mathcal{K}$ be a simplicial complex. For all $S \subseteq \mathcal{F}^d$, if $\beta_{d-1}(\mathcal{K}^{d-1} \cup S) = m > \beta_{d-1}(\mathcal{K})$ then there exists a $\sigma \in \mathcal{F}^d \setminus S$ such that $\beta_{d-1}(\mathcal{K}^{d-1} \cup S \cup \sigma) = m - 1$.*

Proof. Let $S \subseteq \mathcal{F}^d$ such that $\beta_{d-1}(\mathcal{K}^{d-1} \cup S) = m > \beta_{d-1}(\mathcal{K})$ and suppose that all $\sigma \in \mathcal{F}^d \setminus S$ are positive faces w.r.t. $\mathcal{K}^{d-1} \cup S$. Let $S_1 \subseteq \mathcal{F}^d$ be such that $S_1 \supseteq S$. Then, from Lemma 3.2, we have that any $\sigma \in \mathcal{F}^d \setminus S_1$ is positive w.r.t. $\mathcal{K}^{d-1} \cup S_1$. From this, we have $\beta_{d-1}(\mathcal{K}^{d-1} \cup S_1) = \beta_{d-1}(\mathcal{K}^{d-1} \cup S)$ for any $S_1 \supseteq S$. Taking $S_1 = \mathcal{F}^d$, we obtain the necessary contradiction that $\beta_{d-1}(\mathcal{K}) = \beta_{d-1}(\mathcal{K}^{d-1} \cup S)$. Hence, there is a negative face $\sigma \in \mathcal{F}^d \setminus S$ w.r.t. $\mathcal{K}^{d-1} \cup S$. $\qquad\square$

As an acycle corresponds to a basis in matroid, every maximal acycle of d-faces has a constant cardinality (i.e., the rank of the space of boundaries). We shall prove this independently below. Let $\gamma_d(\mathcal{K}) := f_d(\mathcal{K}^d) - \beta_d(\mathcal{K}^d)$. Evaluating $\chi(\mathcal{K}^d) - \chi(\mathcal{K}^{d-1})$ by applying the

Euler-Poincaré formula (2.2) and then using the definition of $\gamma_d(\mathcal{K})$ yields

(3.1) $$\gamma_d(\mathcal{K}) = \beta_{d-1}(\mathcal{K}^{d-1}) - \beta_{d-1}(\mathcal{K}^d) = \beta_{d-1}(\mathcal{K}^{d-1}) - \beta_{d-1}(\mathcal{K}),$$

where the latter equality follows from Corollary 3.1. Another use of Euler-Poincaré formula yields the following result.

Lemma 3.5. *([DKM16, Proposition 2.13]) For a simplicial complex $\mathcal{K}$ and a subset $S \subseteq \mathcal{F}^d$ of $d-$faces, any two of the following three statements imply the third.*

(1) $\beta_{d-1}(\mathcal{K}^{d-1} \cup S) = \beta_{d-1}(\mathcal{K})$.
(2) $\beta_d(\mathcal{K}^{d-1} \cup S) = 0$.
(3) $|S| = \gamma_d(\mathcal{K})$.

Proof. By applying the Euler-Poincaré formula to $(-1)^d[\chi(\mathcal{K}^{d-1} \cup S) - \chi(\mathcal{K}^d)]$ and rearranging the terms, we derive the identity:

(3.2) $$\beta_d(\mathcal{K}^{d-1} \cup S) + \gamma_d(\mathcal{K}) - |S| - \beta_{d-1}(\mathcal{K}^{d-1} \cup S) + \beta_{d-1}(\mathcal{K}^d) = 0.$$

Separately, from Corollary 3.1, we have $\beta_{d-1}(\mathcal{K}^d) = \beta_{d-1}(\mathcal{K})$. From this, the desired result is easy to see. $\qquad\square$

From the above Lemma, we also have that the cardinality of every maximal acycle is $\gamma_d(\mathcal{K})$. We can now prove the existence result for spanning acycles.

Lemma 3.6. *For a simplicial complex $\mathcal{K}$, if $\beta_{d-1}(\mathcal{K}) = 0$, then there exists a spanning acycle.*

Proof. From (3.1), we obtain the identity

$$\beta_{d-1}(\mathcal{K}^{d-1}) = \gamma_d(\mathcal{K}) + \beta_{d-1}(\mathcal{K}).$$

Since $\gamma_d(\mathcal{K}) \geq 0$, it follows from Lemma 3.4 that, starting with an empty set, we can inductively construct a set $S \subset \mathcal{F}^d$ such that $|S| = \gamma_d(\mathcal{K})$ and $\beta_{d-1}(\mathcal{K}^{d-1} \cup S) = \beta_{d-1}(\mathcal{K})$. By Lemma 3.5, it follows that $\beta_d(\mathcal{K}^{d-1} \cup S) = 0$ implying that S is an acycle with $|S| = \gamma_d(\mathcal{K})$, as desired. If $\beta_{d-1}(\mathcal{K}) = 0$, then this S is also a spanning acycle. $\qquad\square$

We now provide a condition for uniqueness of minimal spanning acycles and, towards deriving the same, we first establish the exchange property of spanning acycles.

Lemma 3.7 (Exchange property). *Let $S \subset \mathcal{F}^d$ be a spanning acycle of a simplicial complex $\mathcal{K}$ and let $\sigma \in \mathcal{F}^d \setminus S$. Then, for any d-face $\sigma_1 \in S$ such that σ_1 is part of a d-cycle containing σ, $S \cup \sigma \setminus \sigma_1$ is also a spanning acycle.*

Proof. By Lemma 3.3, $\sigma \in \mathcal{F}^d \setminus S$ is a positive face w.r.t. $\mathcal{K}^{d-1} \cup S$. So $\beta_d(\mathcal{K} \cup S \cup \sigma) = 1$. Let $\mathscr{C}$ be the d-cycle in $S \cup \sigma$. Clearly, $\sigma \in \mathscr{C}$. So $\sum_{\tau \in \mathscr{C} \cap S} a_\tau \partial \tau = -\partial \sigma$ for some collection of non-zero $\mathbb{F}-$valued coefficients $\{a_\tau\}$.

Suppose that for $\sigma_1 \in S$, $S \cup \sigma \setminus \sigma_1$ is not a spanning acycle. Then by Lemma 2.3, we obtain that $\beta_d(\mathcal{K}^{d-1} \cup S \cup \sigma \setminus \sigma_1) = \beta_{d-1}(\mathcal{K}^{d-1} \cup S \cup \sigma \setminus \sigma_1) = 1$. Let $\mathscr{C}_1$ be the d-cycle in $\mathcal{K}^{d-1} \cup S \cup \sigma \setminus \sigma_1$. Clearly, $\mathscr{C}_1 \not\subset S$ as S is a spanning acycle. Hence $\sigma \in \mathscr{C}_1$ and we

20

derive that for some collection of non-zero $a'_\tau \in \mathbb{F}$, $\sum_{\tau \in (\mathscr{C}_1 \cap S)} a'_\tau \partial \tau = -\partial \sigma$. Setting $a_\tau = 0$ for $\tau \in \mathscr{C}_1 \setminus \mathscr{C}$ and similarly for a'_τ, we derive that

$$\sum_{\tau \in (\mathscr{C} \cup \mathscr{C}_1) \cap S} (a'_\tau - a_\tau) \partial \tau = \sum_{\tau \in (\mathscr{C}_1 \cap S)} a_\tau \partial \tau - \sum_{\tau \in (\mathscr{C} \cap S)} a_\tau \partial \tau = \partial \sigma - \partial \sigma = 0.$$

But since S is a spanning acycle, the above implies that $\forall \tau \in (\mathscr{C} \cup \mathscr{C}_1) \cap S$, $a_\tau = a'_\tau$ and hence $\mathscr{C}_1 = \mathscr{C}$. So, we have that $\sigma_1 \notin \mathscr{C}$ if $S \cup \sigma \setminus \sigma_1$ is not a spanning acycle. By contraposition, we have that if $\sigma_1 \in \mathscr{C}$, then $S \cup \sigma \setminus \sigma_1$ is a spanning acycle. $\qquad\square$

Lemma 3.8 (Uniqueness). *Let $\mathcal{K}$ be a simplicial complex weighted by $w : \mathcal{K} \to \mathbb{R}$ which is injective on $\mathcal{F}^d$. If a minimal spanning acycle exists, then it must be unique.*

Proof. Suppose that S and M are two distinct minimal spanning acycles. Let σ be the d-face with least weight such that $\sigma \in S \bigtriangleup M$ and without loss of generality, assume $\sigma \in S$. Then there is a $d-$cycle $\mathscr{C} \subset M \cup \sigma$ such that $\sigma \in \mathscr{C}$. Since $\mathscr{C} \not\subset S$, there exists a d-face $\sigma_1 \in M \setminus S$ that is part of a d-cycle containing σ. By the choice of σ, $w(\sigma_1) > w(\sigma)$. From Lemma 3.7, $M \cup \sigma \setminus \sigma_1$ is a spanning cycle. But $w(M \cup \sigma \setminus \sigma_1) < w(M)$, a contradiction. $\qquad\square$

Remark 3.9. *Suppose the weight function w is not injective on $\mathcal{F}^d$ but nevertheless monotonic on $\mathcal{K}$. Then as discussed in Remark 2.5, this weight function shall yield a total order on $\mathcal{K}$ and so on $\mathcal{F}^d$ as well. In such a case, the above theorem guarantees that the minimal spanning acycle is unique with respect to the chosen total order.*

3.2. **Kruskal's algorithm.** The classical Kruskal's [Kru56] algorithm helps find minimal spanning trees. We now discuss its generalization that will be useful for finding minimal spanning acycles. Generally, greedy algorithms exist to output a minimal basis for matroids [Wel76, Chapter 19] and the following result can be considered folklore. However, we make use of this repeatedly throughout the remainder of the paper and so we provide a self-contained proof.

Let $\mathcal{K}$ be a simplicial complex weighted by $w : \mathcal{K} \to \mathbb{R}$. By Lemma 2.3, every $\sigma \in \mathcal{F}^d$ is either positive or negative, but not both, with respect to a subcomplex $\mathcal{K}_1$ such that $\mathcal{K}_1^{d-1} = \mathcal{K}^{d-1}$ and $\sigma \notin \mathcal{K}_1$. Using this, we give the simplicial Kruskal's algorithm below.

Algorithm 2 Simplicial Kruskal's Algorithm

Input: $d \geq 1$, $\mathcal{K}, w$
Main Procedure:
set $S = \emptyset$, $F = \mathcal{F}^d$
while $F \neq \emptyset$ and $\beta_{d-1}(\mathcal{K}^{d-1} \cup S) \neq 0$
- remove a face σ with minimum weight from F (w.r.t. $<_l$).
- if σ is negative w.r.t. $\mathcal{K}^{d-1} \cup S$, then add σ to S.

Output $M = S$.

Lemma 3.10. *Let $\mathcal{K}$ be a weighted simplicial complex with $\beta_{d-1}(\mathcal{K}) = 0$ and let M be the output of the simplicial Kruskal's algorithm. Then, M is a minimal spanning acycle.*

Proof. We shall assume that the weight function w is injective. For the general case, similar arguments can be carried out by using Remarks 2.5 and 3.9. From Lemmas 3.6 and 3.8, it follows that there is a unique minimal spanning acycle which we denote by M_1.

We now show that M is a spanning acycle. Clearly, $\beta_d(\mathcal{K}^{d-1}) = 0$ and by our algorithm and Lemma 2.3, it remains the same at every stage of the algorithm and so $\beta_d(\mathcal{K}^{d-1} \cup M) = 0$, proving that M is an acycle. Clearly, each face in $\mathcal{F}^d \setminus M$ is positive with respect to $\mathcal{K}^{d-1} \cup M$. Hence, M is spanning as using Lemma 3.2 we have that

$$\beta_{d-1}(\mathcal{K}^{d-1} \cup M) = \beta_{d-1}(\mathcal{K}^{d-1} \cup M \cup \mathcal{F}^d \setminus M) = \beta_{d-1}(\mathcal{K}) = 0.$$

For the proof of minimality, we argue as in the Kruskal's algorithm for minimal spanning tree. We prove that at any stage of the algorithm, $S \subseteq M_1$. Assuming that the above claim is true, $M \subseteq M_1$. Since M and M_1 are both spanning acycles, $M_1 = M$ as desired.

It remains to prove that $S \subseteq M_1$ at any stage. We use induction for the same. Trivially, this is true for $S = \emptyset$. Suppose that the claim holds for S at some stage of the algorithm, i.e., $S \subset M_1$ but $S \neq M$. This implies that there does exist a $d-$face in $\mathcal{F}^d \setminus S$ which is negative w.r.t. $\mathcal{K}^{d-1} \cup S$ and hence, from Lemma 3.3, $S \neq M_1$. Let σ be the next face that is added to S and suppose that $\sigma \notin M_1$. Clearly, $\beta_d(\mathcal{K}^{d-1} \cup M_1 \cup \sigma) = 1$. Hence, there exists a $d-$cycle in $\mathcal{K}^{d-1} \cup M_1 \cup \sigma$ whose support[6] $\mathscr{C}$ contains σ. Since $\beta_d(\mathcal{K}^{d-1} \cup S \cup \sigma) = 0$, $\mathscr{C} \not\subseteq S \cup \sigma$ and so there exists $\sigma_1 \in \mathscr{C} \cap M_1 \setminus S$. Clearly, either $w(\sigma) < w(\sigma_1)$ or $w(\sigma) > w(\sigma_1)$ as w is injective. Suppose that $w(\sigma) < w(\sigma_1)$. By the exchange property of matroids, it follows that $M_1 \cup \sigma \setminus \sigma_1$ is spanning acycle with $w(M_1 \cup \sigma \setminus \sigma_1) < w(M_1)$, a contradiction. Suppose that $w(\sigma) > w(\sigma_1)$. Since $S \subsetneq M_1$, $\sigma_1 \in M_1 \setminus S$, and M_1 is a spanning acycle, it follows from Lemma 3.2 that σ_1 is negative w.r.t. $\mathcal{K}^{d-1} \cup S$. Thus, it follows that the algorithm would have chosen σ_1 before σ, a contradiction. The desired claim now follows. $\qquad\square$

As with minimal spanning trees, the Kruskal's algorithm has a number of useful consequences. We conclude this section with a definition of a (topological) notion of a *cut* for a simplicial complex and show that it has the desired properties which will prove useful in Section 4.1.

Definition 3.11 (Cut). *Let $d \geq 1$. Given a simplicial complex $\mathcal{K}$ with $\beta_{d-1}(\mathcal{K}) = 0$, a subset $\mathscr{C} \subseteq \mathcal{F}^d$ is a cut if $\beta_{d-1}(\mathcal{K} - \mathscr{C}) > 0$ and for any $\mathscr{C}_1 \subsetneq \mathscr{C}$, $\beta_{d-1}(\mathcal{K} - \mathscr{C}_1) < \beta_{d-1}(\mathcal{K} - \mathscr{C})$.*

As expected, this definition yields a corresponding *cut property*.

Lemma 3.12 (Cut Property). *Let $\mathcal{K}$ be a weighted simplicial complex with $\beta_{d-1}(\mathcal{K}) = 0$. Let $\mathscr{C} \subseteq \mathcal{F}^d$ be a cut. Then $\mathscr{C} \cap S \neq \emptyset$ for any spanning acycle S and every minimum weight face in $\mathscr{C}$ belongs to some minimal spanning acycle.*

Proof. Let S be a spanning acycle and suppose that $\mathscr{C} \cap S = \emptyset$. On one hand, because S is spanning, $\beta_{d-1}(\mathcal{K}^{d-1} \cup S) = 0$. On the other hand, since $\mathscr{C}$ is a cut, we have $\beta_{d-1}(\mathcal{K} - \mathscr{C}) > 0$. The latter, when combined with the second inequality in Corollary 3.1 and the fact that

---

[6]For a $d-$chain $\sum_i a_i \sigma_i$, its support is $\{\sigma_i \in \mathcal{F}^d : a_i \neq 0\}$

$S \subseteq \mathcal{F}^d(\mathcal{K} - \mathscr{C})$, implies $\beta_{d-1}(\mathcal{K}^{d-1} \cup S) > 0$. This leads to a contradiction and, thus, the first conclusion holds.

Now for the second part. Let σ_1 be a minimum weight face in the cut $\mathscr{C}$ and let $<_l$ be a total order in which this is the unique minimum weight face in the cut $\mathscr{C}$. Consider the simplicial Kruskal's algorithm under this $<_l$ and let S_1 be the acycle constructed when σ_1 is the minimum weight face in F. Clearly $\mathcal{K}^{d-1} \cup S_1 \subseteq \mathcal{K} - \mathscr{C}$. Setting $\mathscr{C}_1 = \mathscr{C} \setminus \sigma_1$, the cut property implies that

$$\beta_{d-1}(\mathcal{K} - \mathscr{C}_1) < \beta_{d-1}(\mathcal{K} - \mathscr{C}).$$

Thus σ_1 is negative w.r.t. $\mathcal{K} - \mathscr{C}$ and, by Lemma 3.2, is also negative w.r.t. $\mathcal{K}^{d-1} \cup S_1$. Hence σ_1 will be added to the minimal spanning acycle by the simplicial Kruskal's algorithm . $\square$

We note that this agrees with the graph notion of a cut. This will prove useful when considering extremal faces. We conclude with the following consequence. Let $\mathcal{K}$ be a simplicial complex and let $\tau \in \mathcal{F}^{d-1}$. Then $\sigma \in \mathcal{F}^d$ is said to be a coface of τ, if $\tau \subset \sigma$. Since the set of all cofaces of a $(d-1)$-face forms a cut, the below result is immediate.

Corollary 3.13. *Let $\mathcal{K}$ be a weighted simplicial complex with $\beta_{d-1}(\mathcal{K}) = 0$. Let $\tau \in \mathcal{F}^{d-1}(\mathcal{K})$ and $\sigma_0 := \arg\min\{w(\sigma) : \sigma \in \mathcal{F}^d(\mathcal{K}), \tau \subset \sigma.\}$. Then, $\sigma_0 \in M$ for some minimal spanning acycle M.*

3.3. **Persistence diagrams and minimal spanning acycles.** In this section, we prove the connection between persistence diagrams and minimal spanning acycles (Theorem 1.3) and some consequences. Though this correspondance is striking in its simplicity and completely consistent with the minimal spanning tree case, we will see that this has some nontrivial consequences in the study of weighted complexes.

The minimal spanning acycle represents the persistence boundary basis w.r.t. the sublevel set filtration induced by weights on the simplices. This is explicit from the incremental algorithm (Algorithm 1). From the decomposition of a filtration into a persistence diagram, it follows that a positive simplex generates a new homology class and hence forms a new cycle, while a negative simplex bounds an existing non-trivial homology class and hence is a boundary. Our proof will make this idea precise.

Proof of Theorem 1.3. We only prove the result for death times $\mathcal{D}$ since the result for birth times $\mathcal{B}$ is then immediate. This is because, on one hand, every $d-$simplex is either positive or negative with respect to $\mathcal{K}(\sigma^-)$ (see (2.6) and (2.7)). On the other hand, by the incremental algorithm (Algorithm 1), negative simplices correspond to death times and positive simplices correspond to birth times (2.8).

We again only consider the case when the filtration values are unique and appeal to Remark 2.5 to complete the proof in the general case. Note that, in the general case, we use the same total ordering for the incremental algorithm (Algorithm 1) generating death and birth times as well as the simplicial Kruskal's algorithm (Algorithm 2).

By uniqueness of weights on $\mathcal{F}^d$, the Kruskal's algorithm gives us the minimal spanning acycle M. Firstly, by the relation (2.4), the condition to add σ to S in Kruskal's algorithm is equivalent to $\partial(C_d(S)) \subsetneq \partial(C_d(S \cup \sigma))$. Similarly, the incremental algorithm adds $c = w(\sigma)$

to $\mathcal{D}$ if $\partial(C_d(\mathcal{K}(c-))) \subsetneq \partial(C_d(\mathcal{K}(c-) \cup \sigma))$, where, for $c \in \mathbb{R}$, $\mathcal{K}(c) := \{\sigma \in K : w(\sigma) \leq c\}$ and $\mathcal{K}(c-) := \{\sigma \in K : w(\sigma) < c\}$.

Let c be a non-trivial value in the filtration, i.e., there exists $\sigma \in \mathcal{K}$ such that $w(\sigma) = c$. Let $M(c)$ denote the acycle generated by Kruskal's algorithm on $\mathcal{K}(c)$, i.e., $M(c) := M \cap \mathcal{K}(c)$; similarly, define the notation $M(c-)$. By the above discussion on Kruskal's algorithm and incremental algorithm, our proof is complete if we show that $\partial(C_d(M(c))) = \partial(C_d(\mathcal{K}(c)))$. Trivially, $\partial(C_d(M(c))) \subseteq \partial(C_d(\mathcal{K}(c)))$ and we shall now show the other inclusion.

Suppose the other inclusion does not hold, then there exists a $\tau \in \mathcal{F}^d(\mathcal{K}(c)) \setminus M(c)$ such that $\partial\tau \notin \partial(C_d(M(c)))$. Let $w(\tau) = b \leq c$. Then, clearly $\partial\tau \notin \partial(C_d(M(b-)))$; hence, by (2.4), τ will be a negative face with respect to $\mathcal{K}^{d-1} \cup M(b-)$. Therefore, Kruskal's algorithm would have added σ to the acycle $M(b-)$ contradicting the assumption that $\tau \notin M(c)$. Thus, we have $\partial(C_d(M(c))) = \partial(C_d(\mathcal{K}(c)))$ and the proof is complete. $\qquad\square$

The above result has powerful applications for random complexes as will be seen in the next section but we will now mention few applications in the deterministic setting as well. As already mentioned in the introduction, we obtain [HS17, Theorem 1.1] (see (1.1)) as an easy corollary of our previous theorem. Further, we can easily prove a fundamental uniqueness result for minimal spanning acycles relying upon this correspondence and the uniqueness of persistence diagrams [ZC05, Theorem 2.1],[CDSGO16, Theorem 1.3], [CB15, Theorem 1.1][7].

Lemma 3.14. *Let $\mathcal{K}$ be a weighted $d-$complex such that $\beta_{d-1}(\mathcal{K}^d) = 0$ and M_1, M_2 be two $d-$minimal spanning acycles in $\mathcal{K}$. Let $c \in \mathbb{R}$. Then we have that*

$$|\{\sigma \in M_1 : w(\sigma) = c\}| = |\{\sigma \in M_2 : w(\sigma) = c\}|.$$

In the case of unique weights, the minimal spanning acycle is unique making the above lemma trivially true. In the case of non-unique weights, the minimal spanning acycle we obtain will depend on our choice of extension to a total order. However, the above theorem states that the weights of a minimal spanning acycle will be independent of this choice.

We now give an alternative characterization of a minimal spanning acycle that follows from the proof of Theorem 1.3. Such a characterization of a minimal spanning tree has been very useful in the study of minimal spanning trees on infinite graphs ([LP17, Chapter 11], [Ale95, Proposition 2.1]). A similar characterization for minimal spanning tree is known as *the creek-crossing criterion* in [Ale95]. However, we wish to point out now that these different characterizations do not coincide even in the infinite graph case ([Ale95, Proposition 2.1]).

Lemma 3.15. *Let $\mathcal{K}$ be a weighted simplicial complex with $\beta_{d-1}(\mathcal{K}) = 0$. Let $\sigma \in \mathcal{F}^d$ and M be the minimal spanning acycle with respect to a total order $<_l$ extending the partial order induced by w. Then $\sigma \in M$ iff $\partial\sigma \notin \partial(C_d(\mathcal{K}(\sigma^-)))$.*

Proof. From the proof of Theorem 1.3, we know that $\partial(C_d(M \cap \mathcal{K}(\sigma^-)) = \partial(C_d(\mathcal{K}(\sigma^-))$. Thus, by Kruskal's algorithm and (2.4), we have that $\sigma \in M$ iff $\partial\sigma \notin \partial(C_d(\mathcal{K}(\sigma^-)))$. $\qquad\square$

---

[7]Uniqueness follows from certain assumptions on finiteness and the Krull-Remak-Schmidt theorem of isomorphisms of indecomposable subgroups, which always hold in the setting of finite simplicial complexes.

3.4. **Stability result.** Here, we provide a proof for Theorem 1.4.

Proof of Theorem 1.4. Again, it suffices to prove the theorem for death times and the proof for birth times is quite identical. Secondly, due to Theorem 1.3, we shall prove the stability result for weights of a minimal spanning acycle. We shall also assume $0 \leq p < \infty$ and the extension to $p = \infty$ follows by a standard limiting argument.

Let M, M' be the two minimal spanning acycles corresponding to f, f'. We begin with the following case: where f, f' differ precisely on one simplex σ and $f(\sigma) = a, f'(\sigma) = a', |a-a'| = c$. In this case, as we shall show later, $|M \triangle M'| \leq 2$, where $\triangle$ denotes the symmetric difference between the two sets. Since M, M' have equal cardinalities, $|M \triangle M'| \in \{0, 2\}$. If $M \triangle M' = \emptyset$, we are done since the identity map between the simplices in M, M' gives that

$$\inf_{\pi} \sum_{\sigma \in M} |f(\sigma) - f'(\pi(\sigma))|^p \leq c^p = \sum_{\sigma \in \mathcal{F}^d} |f(\sigma) - f'(\sigma)|^p.$$

In the other case, $M \triangle M' = \{\sigma_1, \sigma_2\}$ with $\sigma_1 \in M, \sigma_2 \in M'$ and one of the σ_i's is σ. Below, we shall also show that $|f(\sigma_1) - f'(\sigma_2)| \leq c$. This again shows that

$$\inf_{\pi} \sum_{\sigma \in M} |f(\sigma) - f'(\pi(\sigma))|^p \leq c^p = \sum_{\sigma \in \mathcal{F}^d} |f(\sigma) - f'(\sigma)|^p.$$

By a recursive application of the above case, we can prove the theorem for the general case of f, f' differing in many simplices.

For the rest of the proof, we shall focus only on the case of f, f' differing on exactly one simplex, say $\sigma \in \mathcal{F}^d$, and derive the claims made above. Without loss of generality, assume that f, f' assign distinct weights to distinct faces; the case of non-distinct weights can be proved by appealing again to Remarks 2.5 and 3.9. Given a set $I \subset \mathbb{R}$, $M(I) := \{\sigma \in M : f(\sigma) \in I\}$. Also, as before, let $M(a) = M((-\infty, a])$ and $M(a-) = M((-\infty, a))$. Define these notions, similarly, for notions for M'.

We shall break the proof into four cases where the first two take care of the trivial cases, i.e., when $M \triangle M' = \emptyset$. We shall assume that both M and M' are generated by simplicial Kruskal's algorithm (Algorithm 2).

Case 1: Suppose $\sigma \in M$ and $a > a'$, i.e., $f(\sigma) > f'(\sigma)$. In this case, since $M(a)$ and $M'(a)$ are both maximal acycles in $\mathcal{K}(a)$, we have that $|M(a)| = |M'(a)|$. Further, by Kruskal's algorithm, $M(a'-) = M'(a'-)$. Since $\sigma \in M$, σ is negative w.r.t. $M(a-)$. Now, because $M'(a'-) \subset M(a-)$, it follows from Lemma 3.2 that σ is negative w.r.t. $M'(a'-)$ and, therefore, $\sigma \in M'$. Similarly, by Lemma 3.2, it is also easy to see that $M'((a', a)) \subset M((a', a))$. Consequently, it follows that $M(a) = M'(a)$ since $M(a)$ and $M'(a)$ have equal cardinalities. Continuing with Kruskal's algorithm from a onwards gives $M = M'$.

Case 2: Suppose $\sigma \notin M$ and $a' > a$. This case is similar to Case 1 above. The main differences are as follows. First, we note that $M(a-) = M'(a-)$. Second, since $\sigma \notin M$, $M((a, a')) \subset M'((a, a'))$. Arguing as before, it then follows that $M = M'$.

25

Case 3: Suppose $\sigma \in M$ and $a < a'$. If $\sigma \in M'$, then arguing as in Case 1 gives $M = M'$. Thus, let $\sigma \notin M'$. We show that, for some $d-$face τ, $M'((a, a')) \setminus M((a, a')) = \{\tau\}$, $M \triangle M' = \{\sigma, \tau\}$, and $f'(\tau) - f(\sigma) \le a' - a = c$, as needed.

To show the same, note that by Kruskal's algorithm $M(a-) = M'(a-)$ and $|M(a')| = |M'(a')|$. Further, if τ' with $f(\tau') \in (a, a')$ is negative w.r.t. $\mathcal{K}(\tau'-)$, then Lemma 3.2 shows that τ' is also negative w.r.t. $\mathcal{K}'(\tau'-)$ as well. Hence, from Lemma 3.15 and (2.4), it follows that $M((a, a')) \subset M'((a, a'))$. Now, because of the equality of cardinalities, there exists a $\tau \notin M$ with $f'(\tau) \in (a, a')$ such that $M'(a') = M(a) \cup M((a, a')) \cup \tau$. The desired results are then easy to see.

Case 4: Suppose $\sigma \notin M$ and $a' < a$. Then either $\sigma \notin M'$ or $\sigma \in M'$. If $\sigma \notin M'$, then $M = M'$ as in Case 2. If $\sigma \in M'$, arguing as in Case 3, we have that $M((a', a)) \setminus M'((a', a)) = \{\tau\}$, $M(a, \infty) = M'(a, \infty)$ and hence $M \triangle M' = \{\sigma, \tau\}$ with $f(\tau) - f'(\sigma) \le a - a' = c$. $\qquad \square$

4. Weighted random complexes

Our first aim here is to look at weighted random complexes (Definition 1.6) and derive our point process convergence result (Theorem 1.7). Our second aim is to show the other important consequence of our stability result (Corollary 1.8).

Towards proving Theorem 1.7, we first consider a special case where the weights are i.i.d. uniform on all possible $d-$faces and 0 elsewhere.

4.1. Random $d-$ complex : I.I.D. uniform weights.

The *uniformly weighted $d-$complex* $\mathcal{U}_{n,d}$ is the randomly weighted $d-$ complex $\mathcal{L}'_{n,d}$ with $\|\epsilon_n\|_\infty = 0$ and $\mathscr{F}$ being the uniform distribution on $[0, 1]$ (see Definition 1.6); hence, $\phi = \phi'$ in this case. The canonical filtration associated with $\mathcal{U}_{n,d}$ is $\{\mathcal{U}_{n,d}(t) : t \in [0, 1]\}$. Trivially, the well-known random $d-$complex $Y_{n,d}(t)$ defined before Lemma 1.5 is the same as $\mathcal{U}_{n,d}(t)$ in distribution.

Fix $d \ge 1$. In this section, we show that the three point sets - nearest neighbour distances, death times, weights in the minimal spanning acycle - corresponding to $\mathcal{U}_{n,d}$ (see below Definition 1.6), under appropriate scaling converge to a Poisson point process as $n \to \infty$.

4.1.1. *Extremal nearest neighbour distances.* Fix $\sigma \in \mathcal{F}^{d-1}(\mathcal{U}_{n,d})$. Then, $C(\sigma)$ defined w.r.t. ϕ, as in (1.3), denotes the nearest neighbour distance of σ. By considering the filtration $\{\mathcal{U}_{n,d}(t) : t \in [0, 1]\}$, note that σ is isolated (not part of any $d-$face) exactly between times 0 and $C(\sigma)$ in $\{\mathcal{U}_{n,d}(t) : t \in [0, 1]\}$. That is, the first coface of σ appears at $t = C(\sigma)$.

For each $\sigma \in \mathcal{F}^{d-1}(\mathcal{U}_{n,d})$, let $\bar{C}(\sigma) := nC(\sigma) - d \log n + \log(d!)$ and let $\mathscr{P}^C_{n,d}$ be the scaled point set given by

$$(4.1) \qquad \mathscr{P}^C_{n,d} := \{\bar{C}(\sigma) : \sigma \in \mathcal{F}^{d-1}(\mathcal{U}_{n,d})\}.$$

Viewing the latter as a point process, for any $R \subseteq \mathbb{R}$, we set

$$(4.2) \qquad \mathscr{P}^C_{n,d}(R) := |\{\sigma \in \mathcal{F}^{d-1}(\mathcal{U}_{n,d}) : \bar{C}(\sigma) \in R\}|.$$

For any $c \in \mathbb{R}$, let $\mathscr{P}^C_{n,d}(c, \infty) \equiv \mathscr{P}^C_{n,d}((c, \infty))$. Separately, let $N_{n,d-1}(p)$ denote the number of isolated $(d-1)-$faces in $Y_{n,d}(p)$.

Since $\mathcal{U}_{n,d}(p)$ has the same distribution as $Y_{n,d}(p)$, it follows that $\mathscr{P}^C_{n,d}(np - d\log n + \log(d!), \infty)$ has the same distribution as $N_{n,d-1}(p)$. Also, whenever p_n is of the form as in (1.2), then we know from Lemma 1.5 that, as $n \to \infty$, $N_{n,d-1}(p_n)$ converges to $\mathrm{Poi}(e^{-c})$, the poisson random variable with mean e^{-c}. From this, we have $\mathscr{P}^C_{n,d}(c, \infty) \Rightarrow \mathrm{Poi}(e^{-c})$ as $n \to \infty$. We now extend this to a multivariate convergence, thereby proving convergence of point processes $\mathscr{P}^C_{n,d}$. Recall that $\mathscr{P}_{\mathrm{poi}}$ is the Poisson point process as in Theorem 1.7.

Proposition 4.1. *As $n \to \infty$, $\mathscr{P}^C_{n,d}$ converges in distribution to $\mathscr{P}_{\mathrm{poi}}$.*

Proof. Let $I := \cup_{j=1}^m (a_{2j-1}, a_{2j}] \subseteq \mathbb{R}$ be an arbitrary but fixed union of finite number of disjoint intervals. Since $\mathscr{P}_{\mathrm{poi}}$ is simple and does not contain atoms, as per Lemma A.1, it suffices to prove the following two statements in order to prove weak convergence of the point process $\mathscr{P}^C_{n,d}$:

$$(i) \lim_{n\to\infty} \mathbb{E}[\mathscr{P}^C_{n,d}(I)] = \mathbb{E}[\mathscr{P}_{\mathrm{poi}}(I)] \text{ and } (ii) \mathscr{P}^C_{n,d}(I) \xrightarrow{d} \mathscr{P}_{\mathrm{poi}}(I) \text{ as } n \to \infty.$$

In turn, to establish these two statements, we make use of the method of factorial moments, i.e., show that

$$(4.3) \qquad \mathbb{E}[(\mathscr{P}^C_{n,d}(I))^{(\ell)}] \to \left(\int_I e^{-x} dx\right)^\ell = \mathbb{E}[(\mathscr{P}_{\mathrm{poi}}(I))^{(\ell)}], \quad \forall \ell \geq 1,$$

where, for $m \in \mathbb{N}$, the notation $m^{(\ell)} = m(m-1)\cdots(m-\ell+1)$ so that $\mathbb{E}[(\mathscr{P}^C_{n,d}(I))^{(\ell)}]$ represents the $\ell-$th factorial moment of the random variable $\mathscr{P}^C_{n,d}(I)$. This suffices since Statement (i) above is precisely the $\ell = 1$ case, while Statement (ii) follows due to [vdH16, Theorem 2.4]. For a brief motivation on the method of factorial moments, see Appendix B.

The rest of the proof concerns proving (4.3). Let $\ell \geq 1$ be fixed. Denote $\ell-$th factorial moment of $\mathscr{P}^C_{n,d}(I)$ by $M^{(\ell)}_{n,d}$. For $\sigma \in \mathcal{F}^{d-1}(\mathcal{U}_{n,d})$ and $R \subseteq \mathbb{R}$, let $1(\sigma; R) \equiv \mathbf{1}[\bar{C}(\sigma) \in R]$, where $\mathbf{1}$ denotes the indicator function. Then, clearly,

$$\mathscr{P}^C_{n,d}(I) = \sum_{\sigma \in \mathcal{F}^{d-1}(\mathcal{U}_{n,d})} 1(\sigma; I).$$

Note that if $X = 1_a + 1_b$, i.e., it is a sum of two indicators, then $X^{(2)} = 1_a 1_b + 1_b 1_a$, while $X^{(\ell)} = 0$ for all $\ell \geq 3$. On the other hand, if $X = 1_a + 1_b + 1_c$, then $X^{(2)} = 2 \times 1_a 1_b + 2 \times 1_a 1_c + 2 \times 1_b 1_c$, $X^{(3)} = 6 \times 1_a 1_b 1_c$, while $X^{(\ell)} = 0$ for all $\ell \geq 4$. Proceeding along these lines, it follows using induction on ℓ and linearity of expectation that

$$M^{(\ell)}_{n,d} = \sum_{\boldsymbol{\sigma} \in \mathscr{I}^{(\ell)}_{n,d}} \mathbb{E}\left[\prod_{i=1}^\ell 1(\sigma_i; I)\right],$$

where

$$\mathscr{I}^{(\ell)}_{n,d} := \{\boldsymbol{\sigma} \equiv (\sigma_1, \ldots, \sigma_\ell) : \sigma_i \in \mathcal{F}^{d-1}(\mathcal{U}_{n,d}) \text{ and no two of } \sigma_1, \ldots, \sigma_\ell \text{ are same}\}.$$

27

To simplify the computation of $M_{n,d}^{(\ell)}$, we group the faces $\boldsymbol{\sigma} \in \mathscr{I}_{n,d}^{(\ell)}$ which give the same value for $\mathbb{E}\left[\prod_{i=1}^{\ell} 1(\sigma_i; I)\right]$. We do this as follows. For $\boldsymbol{\sigma} \in \mathscr{I}_{n,d}^{(\ell)}$, let

$$\gamma(\boldsymbol{\sigma}) \equiv (|\cap_{i \in S} \sigma_i| : S \subseteq \{1, \ldots, \ell\}, |S| \geq 2)$$

denote its intersection type. For $\boldsymbol{\sigma}, \boldsymbol{\sigma'} \in \mathscr{I}_{n,d}^{(\ell)}$, we will say that both have similar intersection type, denoted by $\boldsymbol{\sigma} \sim \boldsymbol{\sigma'}$, if there exists a permutation π of the faces in $\boldsymbol{\sigma'}$ such that $\gamma(\boldsymbol{\sigma}) = \gamma(\pi(\boldsymbol{\sigma'}))$. It is easy to see that $\sim$ is an equivalence relation. Let $\Gamma := \{[\boldsymbol{\sigma}]\}$ denote the quotient of $\mathscr{I}_{n,d}^{(\ell)}$ under $\sim$ with $[\boldsymbol{\sigma}]$ denoting the equivalence class of $\boldsymbol{\sigma}$. Since the number of ways in which ℓ distinct $(d-1)-$faces can intersect each other is finite, we have that the number of equivalence classes in Γ, i.e., $|\Gamma|$, is upper bounded by some constant (w.r.t. n). Indeed $|\Gamma|$ depends on d and ℓ, but these are fixed a priori in our setup. Lastly, note that for $\boldsymbol{\sigma} \in \mathscr{I}_{n,d}^{(\ell)}$, the cardinality of its equivalence class $|[\boldsymbol{\sigma}]|$ indeed depends on n.

Fix $\boldsymbol{\sigma} \equiv (\sigma_1, \ldots, \sigma_\ell)$ and $\boldsymbol{\sigma'} \equiv (\sigma'_1, \ldots, \sigma'_\ell)$ in $\mathscr{I}_{n,d}^{(\ell)}$ such that $\boldsymbol{\sigma} \sim \boldsymbol{\sigma'}$. Then

$$\mathbb{E}\left[\prod_{i=1}^{\ell} 1(\sigma_i; I)\right] = \mathbb{E}\left[\prod_{i=1}^{\ell} 1(\sigma'_i; I)\right].$$

Hence, $M_{n,d}^{(\ell)}$ can be rewritten as

$$M_{n,d}^{(\ell)} = \sum_{[\boldsymbol{\sigma}] \in \Gamma} \sum_{\boldsymbol{\sigma'} \in \mathscr{I}_{n,d}^{(\ell)} : \boldsymbol{\sigma'} \sim \boldsymbol{\sigma}} \mathbb{E}\left[\prod_{i=1}^{\ell} 1(\sigma'_i; I)\right] = \sum_{[\boldsymbol{\sigma}] \in \Gamma} |[\boldsymbol{\sigma}]| \mathbb{E}\left[\prod_{i=1}^{\ell} 1(\sigma_i; I)\right].$$

Counting the number of ways in which ℓ distinct $(d-1)-$faces from a total of $\binom{n}{d}$ can be arranged, we have $|\mathscr{I}_{n,d}^{(\ell)}| = \ell!\binom{\binom{n}{d}}{\ell}$. For each $[\boldsymbol{\sigma}] \in \Gamma$, we have $[\boldsymbol{\sigma}] \subseteq \mathscr{I}_{n,d}^{(\ell)}$; hence, $|[\boldsymbol{\sigma}]| = c_n([\boldsymbol{\sigma}])|\mathscr{I}_{n,d}^{(\ell)}|$, for some number $c_n([\boldsymbol{\sigma}]) \in [0, 1]$. Clearly

(4.4)
$$\sum_{[\boldsymbol{\sigma}] \in \Gamma} c_n([\boldsymbol{\sigma}]) = 1.$$

Hence, it follows that

$$M_{n,d}^{(\ell)} = \sum_{[\boldsymbol{\sigma}] \in \Gamma} c_n([\boldsymbol{\sigma}])\ell!\binom{\binom{n}{d}}{\ell} \mathbb{E}\left[\prod_{i=1}^{\ell} 1(\sigma_i; I)\right].$$

For every $\sigma \in \mathcal{F}^{d-1}(\mathcal{U}_{n,d})$, we have

$$1(\sigma; I) = \sum_{j=1}^{m} 1(\sigma; (a_{2j-1}, a_{2j}]) = \sum_{j=1}^{m} (1(\sigma; (a_{2j-1}, \infty)) - 1(\sigma; (a_{2j}, \infty))).$$

Therefore, it follows that for any $\boldsymbol{\sigma} \equiv (\sigma_1, \ldots, \sigma_\ell) \in \mathscr{I}_{n,d}^{(\ell)}$,

$$\prod_{i=1}^{\ell} 1(\sigma_i; I) = \sum_{(\alpha_1, \ldots, \alpha_\ell) \in \{1, \ldots, 2m\}^{\ell}} \prod_{i=1}^{\ell} (-1)^{\alpha_i+1} 1(\sigma_i; (a_{\alpha_i}, \infty)),$$

28

where $\{1, \ldots, 2m\}^\ell$ is the the $\ell-$ary cartesian power of $\{1, \ldots, 2m\}$. Hence,

$$(4.5) \qquad M_{n,d}^{(\ell)} = \sum_{[\boldsymbol{\sigma}] \in \Gamma} c_n([\boldsymbol{\sigma}]) \sum_{(\alpha_1, \ldots, \alpha_\ell) \in \{1, \ldots, 2m\}^\ell} \ell! \binom{\binom{n}{d}}{\ell} (-1)^{\sum_i \alpha_i + \ell} \mathbb{E}\left[\prod_{i=1}^{\ell} \mathbf{1}(\sigma_i; (a_{\alpha_i}, \infty))\right].$$

From the scaling of $C(\sigma)$, for any $\sigma \in \mathcal{F}^{d-1}(\mathcal{U}_{n,d})$ and any $a \in \mathbb{R}$,

$$\mathbf{1}(\sigma; (a, \infty)) = \mathbf{1}\left[C(\sigma) > \frac{a + d\log n - \log(d!)}{n}\right].$$

Combining this with the definitions of $C(\sigma)$ and $\mathcal{U}_{n,d}$, observe that

$$\ell! \binom{\binom{n}{d}}{\ell} \mathbb{E}\left[\prod_{i=1}^{\ell} \mathbf{1}(\sigma_i; (a_{\alpha_i}, \infty))\right] \sim \frac{n^{d\ell}}{(d!)^\ell} \prod_{i=1}^{\ell} \left(1 - \frac{a_{\alpha_i} + d\log n - \log(d!)}{n}\right)^{n - \kappa_i}.$$

Here $\kappa_1, \ldots, \kappa_\ell \geq 0$ are some constants depending on how many vertices are common between the faces $\sigma_1, \ldots, \sigma_\ell$. From this, irrespective of $\kappa_1, \ldots, \kappa_\ell$ (as these are constants independent of n), we have

$$\lim_{n \to \infty} \ell! \binom{\binom{n}{d}}{\ell} \mathbb{E}\left[\prod_{i=1}^{\ell} \mathbf{1}(\sigma_i; (a_{\alpha_i}, \infty))\right] = e^{-\sum_{i=1}^{\ell} a_{\alpha_i}}.$$

Therefore, the inner sum in (4.5) converges to $\left(\sum_{j=1}^{m} [e^{-a_{2j-1}} - e^{-a_{2j}}]\right)^\ell = \left(\int_I e^{-x} dx\right)^\ell$ for every $[\boldsymbol{\sigma}] \in \Gamma$. Now, using (4.4), it follows that

$$\lim_{n \to \infty} M_{n,d}^{(\ell)} = \left(\int_I e^{-x} dx\right)^\ell,$$

as desired in (4.3). $\qquad \square$

4.1.2. *Extremal death times.* We now discuss death times in the persistence diagram. First, we state a lemma explaining why nearest neighbour distances approximate death times.

Lemma 4.2. *Fix $d \geq 1$. Let $N_{d-1}(Y_{n,d}(p_n))$ be the number of isolated $(d-1)-$faces in $Y_{n,d}(p_n)$ with p_n as in (1.2). Then*

$$\lim_{n \to \infty} \mathbb{E}|\beta_{d-1}(Y_{n,d}(p_n)) - N_{d-1}(Y_{n,d}(p_n))| = 0.$$

This lemma essentially follows from ideas in the proofs in [KP14, Theorem 1.10]. But, to the best of our knowledge, it has not been explicitly mentioned anywhere. The proof for the case $d \geq 2$ requires cohomological arguments and hence the entire proof along with more details on cohomology theory has been provided in Section C in the Appendix.

Let $\mathscr{P}_{n,d}^{D}$ denote the set of scaled death times in $H_{d-1}(\mathcal{U}_{n,d})$ as in the second item listed below (1.3). Let $c \in \mathbb{R}$ be some arbitrary but fixed constant and let p_n be as defined in (1.2). Then, for n large enough, we have

$$\mathscr{P}_{n,d}^{D}(c, \infty) = \beta_{d-1}(\mathcal{U}_{n,d}(p_n)) \quad \text{and} \quad \mathscr{P}_{n,d}^{C}(c, \infty) = N_{d-1}(\mathcal{U}_{n,d}(p_n)).$$

From Lemma 4.2, it then immediately follows that

$$(4.6) \qquad\qquad \lim_{n \to \infty} \mathbb{E}|\mathscr{P}_{n,d}^{D}(c, \infty) - \mathscr{P}_{n,d}^{C}(c, \infty)| = 0.$$

Now we are ready to prove the convergence result for scaled death times.

Proposition 4.3. *As $n \to \infty$, $\mathscr{P}_{n,d}^D$ converges in distribution to the Poisson point process $\mathscr{P}_{poi}$.*

Proof. Let $I := \cup_{j=1}^m (a_{2j-1}, a_{2j}] \subseteq \mathbb{R}$ be some finite union of disjoint intervals. Since $\mathscr{P}_{poi}$ is simple and does not contain atoms, again as per Lemma A.1, to prove the desired result, it suffices to show that:

$$(i) \lim_{n\to\infty} \mathbb{E}[\mathscr{P}_{n,d}^D(I)] = \mathbb{E}[\mathscr{P}_{poi}(I)] \text{ and } (ii) \mathscr{P}_{n,d}^D(I) \overset{d}{\Rightarrow} \mathscr{P}_{poi}(I) \text{ as } n \to \infty.$$

From triangle inequality,

$$(4.7) \qquad |\mathscr{P}_{n,d}^D(I) - \mathscr{P}_{n,d}^C(I)| \le \sum_{j=1}^{2m} |\mathscr{P}_{n,d}^D(a_j, \infty) - \mathscr{P}_{n,d}^C(a_j, \infty)|.$$

By combining this with (4.6) and Statement (i) from above (4.3), we get (i).

The same argument also shows that $|\mathscr{P}_{n,d}^D(I) - \mathscr{P}_{n,d}^C(I)| \to 0$ in probability as $n \to \infty$. Combining this with Slutsky's theorem[8] [EK09, Chapter 3, Corollary 3.3] and Statement (ii) from above (4.3), we obtain (ii) as desired. $\square$

4.1.3. *Extremal weights in the $d-$minimal spanning acycle.* Again fix $d \ge 1$. Viewing $\mathcal{U}_{n,d}$ as a weighted simplicial complex, let M denote its $d-$minimal spanning acycle. And let

$$(4.8) \qquad \mathscr{P}_{n,d}^M := \{nw(\sigma) - d\log n + \log(d!) : \sigma \in M\}$$

denote the set of scaled weights of the faces in the $d-$minimal spanning acycle of $\mathcal{U}_{n,d}$. Using Theorem 1.3 and Proposition 4.3, we get the following result immediately.

Proposition 4.4. *As $n \to \infty$, $\mathscr{P}_{n,d}^M$ converges in distribution to the Poisson point process $\mathscr{P}_{poi}$.*

4.2. **Random $d-$ complexes : I.I.D. generic weights with perturbation.** We shall now prove our most general point process convergence result (Theorem 1.7) and then describe corollaries which give simpler bounds to verify the assumptions of this result. For the proof, we shall first consider the simplicial complex $\mathcal{K}_n^d$ weighted by ϕ alone, which we shall refer to as $\mathcal{L}_{n,d}$. With respect to this $\mathcal{L}_{n,d}$, define $C(\sigma), D_i, M, \mathscr{P}_{n,d}^C, \mathscr{P}_{n,d}^D$, and $\mathscr{P}_{n,d}^M$, exactly as below Definition 1.6.

Proposition 4.5. *Suppose that $\mathscr{F}$ is continuous. Then, the point processes $\mathscr{P}_{n,d}^C, \mathscr{P}_{n,d}^D$, and $\mathscr{P}_{n,d}^M$, converge in distribution to $\mathscr{P}_{poi}$ as $n \to \infty$.*

Proof. Clearly, $\{\mathscr{F}(\phi(\sigma))\}_{\sigma \in \mathcal{F}^d(\mathcal{L}_{n,d})}$ are i.i.d. uniform $[0,1]$ random variables. The desired result is now immediate from Propositions 4.1, 4.3, and 4.4 $\square$

We need a comparison lemma to prove the main point process convergence result. The first inequality is obvious and the next two follow from Theorem 1.4 for $p = \infty$ and Theorem 1.3.

---

[8]The relevant version of Slutsky's theorem that we use is the following: If the random variables X, $X_1, X_2, \ldots$, and $Y_1, Y_2, \ldots$ is such that $X_n \Rightarrow X$ and $|X_n - Y_n| \to 0$ in probability, then $Y_n \Rightarrow X$.

Lemma 4.6. *For fixed $n, d \geq 1$, we have the following inequalities:*

$$\max_{\sigma \in \mathcal{F}^{d-1}(\mathcal{L}'_{n,d})} |C'(\sigma) - C(\sigma)| \leq \|\epsilon_n\|_\infty,$$

$$\inf_\gamma \max_i |D'_i - \gamma(D_i)| \leq \|\phi' - \phi\|_\infty \leq \|\epsilon_n\|_\infty,$$

where the infimum is over all possible bijections $\gamma : \{D'_i\} \to \{D_i\}$, and

$$\inf_\gamma \max_i |\phi'(\sigma'_i) - \gamma(\phi(\sigma_i))| \leq \|\phi' - \phi\|_\infty = \|\epsilon_n\|_\infty,$$

where the infimum is over all possible bijections $\gamma : \{\phi'(\sigma') : \sigma' \in M'\} \to \{\phi(\sigma) : \sigma \in M\}$.

Proof of Theorem 1.7. We only show that $\mathscr{P}^{D'}_{n,d} \Rightarrow \mathscr{P}_{\text{poi}}$ as $n \to \infty$ using Lemma 4.6, as the other results follow similarly. Let d_v be the vague metric given in (A.1). Suppose we show that $d_v(\mathscr{P}^{D'}_{n,d}, \mathscr{P}^{D}_{n,d}) \to 0$ in probability as $n \to \infty$. Then, since $(M_p(\mathbb{R}), d_v)$ is a Polish space (see Appendix 4.2), we can apply Slutsky's theorem ([EK09, Chapter 3, Corollary 3.3]) and Proposition 4.5 to derive that $\mathscr{P}^{D'}_{n,d} \Rightarrow \mathscr{P}_{\text{poi}}$ as desired. It thus suffices to prove that $d_v(\mathscr{P}^{D'}_{n,d}, \mathscr{P}^{D}_{n,d}) \to 0$ in probability as $n \to \infty$.

Let d_B be as in Definition 2.9. Then by Lemma 4.6, we have that

$$d_B(\mathscr{P}^{D'}_{n,d}, \mathscr{P}^{D}_{n,d}) \leq \zeta n \|\epsilon_n\|_\infty,$$

where we have assumed that the Lipschitz constant associated with $\mathscr{F}$ is ζ. Now, by assumption, $d_B(\mathscr{P}^{D'}_{n,d}, \mathscr{P}^{D}_{n,d}) \to 0$ in probability. Fix $\epsilon \in (0,1)$ and choose $\delta = \frac{\epsilon}{k}$ for some $k \geq 1$. Then, we have

$$
\begin{aligned}
\mathbb{P}\{d_v(\mathscr{P}^{D'}_{n,d}, \mathscr{P}^{D}_{n,d}) > \epsilon\} &\leq \mathbb{P}\{d_v(\mathscr{P}^{D'}_{n,d}, \mathscr{P}^{D}_{n,d}) > \epsilon, d_B(\mathscr{P}^{D'}_{n,d}, \mathscr{P}^{D}_{n,d}) \leq \delta\} \\
&\quad + \mathbb{P}\{d_v(\mathscr{P}^{D'}_{n,d}, \mathscr{P}^{D}_{n,d}) > \epsilon, d_B(\mathscr{P}^{D'}_{n,d}, \mathscr{P}^{D}_{n,d}) > \delta\} \\
&\leq \mathbb{P}\{2\lambda_\epsilon \mathscr{P}^{D}_{n,d}(K_\epsilon)\delta > \epsilon/2\} + \mathbb{P}\{d_B(\mathscr{P}^{D'}_{n,d}, \mathscr{P}^{D}_{n,d}) > \delta\}, \\
&\leq \mathbb{P}\left\{\mathscr{P}^{D}_{n,d}(K_\epsilon) > \frac{k}{4\lambda_\epsilon}\right\} + \mathbb{P}\{d_B(\mathscr{P}^{D'}_{n,d}, \mathscr{P}^{D}_{n,d}) > \delta\}.
\end{aligned}
$$

(4.9)

For the second inequality, we have used (A.2) with λ_ϵ and the compact set K_ϵ as given there. Since the second term in (4.9) converges to 0 as $n \to \infty$, it follows using Proposition 4.5 that

$$\limsup_{n \to \infty} \mathbb{P}\{d_v(\mathscr{P}^{D'}_{n,d}, \mathscr{P}^{D}_{n,d}) > \epsilon\} \leq \lim_{n \to \infty} \mathbb{P}\left\{\mathscr{P}^{D}_{n,d}(K_\epsilon) > \frac{k}{4\lambda_\epsilon}\right\} = \mathbb{P}\left\{\mathscr{P}_{\text{poi}}(K_\epsilon) > \frac{k}{4\lambda_\epsilon}\right\}.$$

Now letting $k \to \infty$, we have that $d_v(\mathscr{P}^{D'}_{n,d}, \mathscr{P}^{D}_{n,d}) \to 0$ in probability as desired. $\square$

Except for a few trivial cases, determining the distribution of the maximum $\|\epsilon_n\|_\infty$ is not easy and hence we give two simple corollaries to verify the bounds.

Corollary 4.7. *For each n, let $\{\psi(\sigma) : \sigma \in \mathcal{F}^d(\mathcal{L}'_{n,d})\}$ have the same distribution as the real valued random variable ψ which, for some $s > 0$, satisfies $\mathbb{E}[e^{s|\psi|}] < \infty$. Define $\epsilon_n(\sigma) = a_n^{-1}\psi(\sigma)$ where a_n is a sequence such that[9] $a_n = \omega(n \log n)$. If $\mathscr{F}$ is Lipschitz continuous, then, each of $\mathscr{P}^{C'}_{n,d}$, $\mathscr{P}^{D'}_{n,d}$, and $\mathscr{P}^{M'}_{n,d}$ converges in distribution to $\mathscr{P}_{\text{poi}}$.*

---

[9] Here w is the small omega notation.

Proof. Using Jensen's inequality for the second inequality below and since $|\mathcal{F}^d(\mathcal{L}'_{n,d})| \leq n^{d+1}$,

$$s\mathbb{E}[\|\epsilon_n\|_\infty] = \frac{1}{a_n} \log e^{s\mathbb{E}\left[\max_{\sigma \in \mathcal{F}^d(\mathcal{L}'_{n,d})} |\psi(\sigma)|\right]} \leq \frac{1}{a_n} \log \mathbb{E}\left[e^{s\max_{\sigma \in \mathcal{F}^d(\mathcal{L}'_{n,d})} |\psi(\sigma)|}\right]$$

$$= \frac{1}{a_n} \log \mathbb{E}\left[\max_{\sigma \in \mathcal{F}^d(\mathcal{L}'_{n,d})} e^{s|\psi(\sigma)|}\right] \leq \frac{1}{a_n} \log(n^{d+1}\mathbb{E}[e^{s|\psi|}]).$$

Hence, $n\|\epsilon_n\|_\infty \to 0$ in probability as $n \to \infty$. The result now follows from Theorem 1.7. $\quad\square$

The following corollary follows from Theorem 1.7 using Markov's inequality and $\|\epsilon_n\|_\infty \leq \|\epsilon_n\|_1$.

Corollary 4.8. *For each n, let $\{\epsilon_n(\sigma) : \sigma \in \mathcal{F}^d(\mathcal{L}'_{n,d})\}$ be identically distributed random variables with $\mathbb{E}|\epsilon_n(\sigma)| = o(n^{-d-2})$ for each σ. If $\mathscr{F}$ is Lipschitz continuous, then each of $\mathscr{P}^{C'}_{n,d}, \mathscr{P}^{D'}_{n,d},$ and $\mathscr{P}^{M'}_{n,d}$ converges in distribution to $\mathscr{P}_{poi}$.*

In relation to $\epsilon_n(\sigma)$'s from Definition 1.6, let $\|\epsilon_n\|_p := (\sum_{\sigma \in \mathcal{F}^d(\mathcal{L}'_{n,d})} |\epsilon_n(\sigma)|^p)^{1/p}$ for $p \in \{1, 2, \ldots\}$.

Proof of Corollary 1.8. Fix a $p \in \{1, 2, \ldots, \}$. Let π be a bijection from $\{D_i\}$ to $\{D'_i\}$ achieving the infimum in Theorem 1.4. Due to the finiteness of the complex, such a bijection exists. Now, we derive from mean-value theorem, Hölder's inequality and our stability result (Theorem 1.4) that

$$|L^p_{n,d-1} - (L')^p_{n,d-1}| \leq p\sum_i (D_i^{p-1} + \pi(D_i)^{p-1})|D_i - \pi(D_i)|$$

$$\leq p(\sum_i |D_i - \pi(D_i)|^p)^{1/p}\left[(\sum_i (D'_i)^p)^{(p-1)/p} + (\sum_i D_i^p)^{(p-1)/p}\right]$$

$$\leq p\|\epsilon_n\|_p\left[(\sum_i (D'_i)^p)^{(p-1)/p} + (\sum_i D_i^p)^{(p-1)/p}\right]$$

Now taking expectations and again using Hölder's inequality, we obtain that

$$\mathbb{E}[|L^p_{n,d-1} - (L')^p_{n,d-1}|] \leq p\mathbb{E}[\|\epsilon_n\|_p^p]^{1/p}\left(\mathbb{E}[\sum_i (D'_i)^p)]^{(p-1)/p} + \mathbb{E}[\sum_i D_i^p]^{(p-1)/p}\right)$$

(4.10)
$$\leq p\mathbb{E}[\|\epsilon_n\|_p^p]^{1/p}\left(\mathbb{E}[L^p_{n,d-1}]^{(p-1)/p} + \mathbb{E}[(L')^p_{n,d-1}]^{(p-1)/p}\right)$$

By the decay bounds on $\epsilon_n(\sigma)$'s, we can derive that $n^{-(d-p)}\mathbb{E}[\|\epsilon_n\|_p^p] \to 0$. From this convergence and (1.5), our proof is complete if we show boundedness of $n^{-(d-p)}\mathbb{E}[(L')^p_{n,d-1}]$. Now, using our stability result (Theorem 1.4), we obtain that

$$\mathbb{E}[(L')^p_{n,d-1}] \leq 2^{p-1}\left(\mathbb{E}[L^p_{n,d-1}] + \mathbb{E}[\|\epsilon_n\|_p^p]\right),$$

and thus the required boundedness follows from convergence of $n^{-(d-p)}\mathbb{E}[\|\epsilon_n\|_p^p]$ and (1.5). $\quad\square$

APPENDIX A. CONVERGENCE OF POINT PROCESSES

We discuss here convergence of point processes under vague topology. For notations and definitions, see Subsection 2.1. Firstly, it is known that $(M_p(\mathbb{R}), \mathcal{M}_p(\mathbb{R}))$ is metrizable as a complete, separable metric space ([Res13, Chapter 3, Proposition 3.17]), i.e., it is a Polish space. In the proof of this proposition, the vague metric d_v has been used which we describe next.

Let $\{G_i\}$ be the collection of open intervals in $\mathbb{R}$ with rational end points and let $\{h_j\}$ be suitable piece-wise linear approximations to the indicator function of these sets. The functions h_j are chosen so that they lie in $C_c^+(\mathbb{R})$ and are Lipschitz continuous (while it is not explicitly highlighted, from the definition of h_j in the proofs of Propositions 3.11, 3.17 from [Res13, Chapter 3], one can check that it is Lipschitz). Then for $m_1, m_2 \in M_p(\mathbb{R})$,

$$(A.1) \quad d_v(m_1, m_2) := \sum_{j=1}^{\infty} \frac{1 - \exp\{-|m_1(h_j) - m_2(h_j)|\}}{2^j} \leq \sum_{j=1}^{\infty} \frac{\min\{|m_1(h_j) - m_2(h_j)|, 1\}}{2^j}.$$

The following is an oft-used result to prove weak convergence of point processes.

Lemma A.1. [Res13, Proposition 3.22] *Let $\{\mathscr{P}_n\}, \mathscr{P}$ be point processes on $\mathbb{R}$ with $\mathscr{P}$ being simple. Let $\mathcal{I}(\mathbb{R})$ be the collection of all finite union of intervals in $\mathbb{R}$. Suppose that for each $I \in \mathcal{I}(\mathbb{R})$, with $\mathbb{P}\{\mathscr{P}(\partial I) = 0\} = 1$, we have*

$$\lim_{n \to \infty} \mathbb{P}\{\mathscr{P}_n(I) = 0\} = \mathbb{P}\{\mathscr{P}(I) = 0\} \text{ and } \lim_{n \to \infty} \mathbb{E}[\mathscr{P}_n(I)] = \mathbb{E}[\mathscr{P}(I)] < \infty.$$

Then $\mathscr{P}_n \Rightarrow \mathscr{P}$ in $M_p(\mathbb{R})$.

We now prove a lemma that will be useful when combining results from computational topology (which uses bottleneck distance) and point process theory (vague topology).

Lemma A.2. *The topology of bottleneck distance is stronger than that of vague topology on $M_p(\mathbb{R})$. In particular, for every $\epsilon > 0$, there exists a constant $\lambda_\epsilon > 0$ and a compact set K_ϵ such that, whenever $d_B(m_1, m) \leq 1/2$, we have*

$$(A.2) \qquad\qquad d_v(m_1, m) \leq 2\lambda_\epsilon m(K_\epsilon) d_B(m_1, m) + \frac{\epsilon}{2}.$$

Proof. We first establish (A.2). Let $\epsilon > 0$ be arbitrary. For any $m, m_1 \in M_p(\mathbb{R})$, it follows from (A.1) that we can choose k (independent of m, m_1) such that:

$$(A.3) \qquad\qquad d_v(m_1, m) \leq \sum_{j=1}^{k} |m_1(h_j) - m(h_j)| + \frac{\epsilon}{2}.$$

Let K_j be the compact support of h_j and λ_j, the associated Lipschitz constant. Set $\lambda_\epsilon = \sum_{i=1}^{k} \lambda_j$ and $K_\epsilon = \cup_{j=1}^{k} K_j^1$, where $K^\rho := \{x \in \mathbb{R} : \exists y \in K \text{ s.t. } |x - y| \leq \rho\}$.

Let m, m_1 be such that $\delta := 2d_B(m_1, m) \leq 1$. Let $\gamma : supp(m) \to supp(m_1)$ be the bijection such that $\max_{x \in \text{supp}(m)} |x - \gamma(x)| \leq \delta$. Also, let $M = supp(m), M_1 = supp(m_1)$.

By the definition of Bottleneck distance, we have that, for any compact set K,

$$(A.4) \qquad\qquad m_1(K) \leq m(K^\delta) \leq m_1(K^{2\delta}).$$

Fix a $j \in \{1, \ldots, k\}$. By the definition of $m(h_j)$,

$$
\begin{aligned}
|m_1(h_j) - m(h_j)| &= |\sum_{x \in M} h_j(x) - \sum_{x \in M_1} h_j(x)| = |\sum_{x \in M} [h_j(x) - h_j(\gamma(x))]| \\
&\leq \lambda_j[m(K_j) + m_1(K_j)]\delta \leq 2\lambda_j m(K_j^1)\delta \leq 2\lambda_j m(K_\epsilon)\delta,
\end{aligned}
$$

where in the last inequality we have used (A.4) and the fact that $\delta \leq 1$. Substituting the above relation in (A.3), we get

$$d_v(m_1, m) \leq 2\lambda_\epsilon m(K_\epsilon)\delta + \frac{\epsilon}{2},$$

as desired.

From this, it follows that for every $m \in M_p(\mathbb{R})$ and $\epsilon > 0$, there exists ρ (depending on m and ϵ) such that $d_B(m_1, m) \leq \rho$ implies $d_v(m_1, m) \leq \epsilon$, which completes the proof. $\qquad\square$

Appendix B. Method of Factorial Moments

Here, we provide a brief motivation for the method of factorial moments. First, this is very closely related to the method of moments and both these methods are useful when the goal is to establish convergence in distribution. Formally, suppose a random variable X is such that its distribution is completely characterised by its moments $\{\mathbb{E}[X^k] : k \geq 1\}$. Since polynomials are dense in the class of continuous functions, note that the above statement is true for a broad class of random variables, including the Poisson random variable. A standard result in probability theory then states that if all the moments of a sequence of random variables converge to that of X, then the sequence itself converges in distribution to X. Keeping this in mind, the method of moments idea to verify if $X_n \Rightarrow X$ is to check if $\mathbb{E}[X_n^k] \to \mathbb{E}[X^k]$ or not for all $k \geq 1$. Now, since moments of a random variable are linear combinations of its factorial moments and vice versa, we can alternatively also work with factorial moments. In case of a Poisson random variable, working with latter makes a lot of sense since the resulting expressions are much simpler than those for the corresponding moments. In particular, if $X \sim \text{Poi}(\lambda)$, then $\mathbb{E}[X^{(\ell)}] = \lambda^\ell$.

Appendix C. Betti numbers and Isolated faces in $Y_{n,d}(p)$

Lemma 4.2 is proved here. The cases $d = 1$ and $d \geq 2$ are dealt with separately, with the latter requiring cohomological arguments.

Proof of Lemma 4.2 for $d = 1$. Let V_n be the vertex set of $Y_{n,1}(p_n)$. Since there can be at most one component of size bigger than $n/2$, for reduced β_0,

$$|\beta_0(Y_{n,1}(p_n)) - N_0(Y_{n,1}(p_n))| \leq \sum_{V \subset V_n, 2 \leq |V| \leq n/2} 1_V + 1[N_0(Y_{n,1}(p_n)) = n],$$

where $1_V = 1$ whenever V forms a connected component in $Y_{n,1}(p_n)$ and there is no edge between a vertex in V and a vertex in V^c. If $|V| = k$, then for all sufficiently large n,

$$\mathbb{E}[1_V] \leq k^{k-2} p_n^{k-1}(1 - p_n)^{k(n-k)}.$$

This is because, when $|V| = k$, there are k^{k-2} possible spanning trees in V, the probability of getting a particular spanning tree in V is p_n^{k-1}, and the probability of having no edge between V and V^c is $(1 - p_n)^{k(n-k)}$. We say sufficiently large because p_n may be negative for small n if c is negative. Hence, for all sufficiently large n,

$$\mathbb{E}|\beta_0(Y_{n,1}(p_n)) - N_0(Y_{n,1}(p_n))| \leq \sum_{k=2}^{n/2} \binom{n}{k} k^{k-2} p_n^{k-1}(1 - p_n)^{k(n-k)} + (1 - p_n)^{\binom{n}{2}}.$$

Since $(1 - p_n) \leq e^{-p_n}$ and $\binom{n}{k} \leq e^k n^k / k^k$, for all sufficiently large n,

$$\mathbb{E}|\beta_0(Y_{n,1}(p_n)) - N_0(Y_{n,1}(p_n))| \leq \sum_{k=2}^{n/2} e^k n^k k^{-2} p_n^{k-1} e^{-p_n k(n-k)} + e^{-p_n \binom{n}{2}}.$$

As $p_n = (\log n + c)/n$, the second term decays to 0 with n.

With regards to the first term,

$$\sum_{k=2}^{n/2} e^k n^k k^{-2} p_n^{k-1} e^{-p_n k(n-k)} = \sum_{k=2}^{n/2} e^k k^{-2} p_n^{k-1} e^{k^2 p_n - kc} \leq \sum_{k=2}^{n/2} e^{T_k},$$

where

$$T_k = k - 2\log(k) + (k-1)\log(\log n + |c|) - (k-1)\log n + k|c| + k^2 \left[\frac{\log n + |c|}{n} \right].$$

Fix n. Treating k as a continuous variable, observe that the second derivative of T_k w.r.t. k is strictly positive for $k \in (3, n/2)$. This shows that T_k is convex in $(3, n/2)$ and hence $T_k \leq \max\{T_3, T_{n/2}\}$ for $k \in \{3, \ldots, n/2\}$. But $T_3 > T_{n/2}$ for all sufficiently large n. Hence, for all sufficiently large n,

$$\sum_{k=2}^{n/2} e^k n^k k^{-2} p_n^{k-1} e^{-p_n k(n-k)} \leq e^{T_2} + \frac{n}{2} e^{T_3}.$$

But the RHS converges to 0 with n. The desired result now follows. $\qquad \square$

We now give a brief exposition about reduced cohomology (w.r.t. $\mathbb{Z}_2$ for simplicity) here which is necessary for proving Lemma 4.2 for the case $d \geq 2$. In one line, it can be said that cohomology is the dual theory of homology and can be derived by considering the dual of the boundary operator ∂.

Consider a simplicial complex $\mathcal{K}$. For $d \geq 0$, a $d-$cochain of $\mathcal{K}$ is a map $g : \mathcal{F}^d \to \mathbb{Z}_2$. Its support[10] is given by $\mathrm{supp}(g) := \{\sigma \in \mathcal{F}^d : g(\sigma) = 1\}$. Let $C^d := \{g : \mathcal{F}^d \to \mathbb{Z}_2\}$ denote the set of all $d-$cochains and it is $\mathbb{Z}_2$-vector space under natural addition and scalar

---

[10]This notion of support is different than that given for chains in Section 2.1.1

multiplication operations on C^d. The $d-$th coboundary operator $\delta_d : C^d \to C^{d+1}$ is defined as follows :

$$(C.1) \qquad \delta_d(g)(\sigma) := \sum_{\tau \in \mathcal{F}^d : \tau \subset \sigma} g(\tau), \quad g \in C^d, \sigma \in \mathcal{F}^{d+1}.$$

For $d \geq 1$, let $B^d := \text{im } (\delta_{d-1})$; and, for $d \geq 0$, let $Z^d := \ker(\delta_d)$. Let $B^0 := \{0,1\}$, where 0 and 1 are respectively the $0-$cochains that assign 0 and 1 to all vertices. The elements of B^d are called *coboundaries* while the those of Z^d are called *cocycles*. As in homology, we have that $\delta_d \circ \delta_{d-1} = 0$ and hence we define the $d-$th cohomology group

$$H^d := \frac{Z^d}{B^d}.$$

It is well known that the $d-$th homology group H_d is isomorphic to the $d-$th cohomology group H^d and so we have that $\beta_d(\mathcal{K}) := \text{rank}(H^d(\mathcal{K}))$.

We first describe upper and lower bounds for Betti numbers, which to the best of our knowledge, have not been explicitly mentioned anywhere. But they follow from the proofs in [LM06, MW09, KP14].

We first discuss upper bounds for Betti numbers. Fix an arbitrary $d \geq 0$. For $g \in C^d$, let $[g] = g + B^d$ and

$$(C.2) \qquad w(g) = \min\{|\text{supp}(g')| : g' \in [g]\},$$

where $|\cdot|$ denotes cardinality. Then it follows that

$$H^d = \{[g] : g \in Z^d\} = \{[0]\} \cup \{[g] : g \in Z^d, |\text{supp}(g)| \geq 1, w(g) = |\text{supp}(g)|\}$$

and hence

$$(C.3) \qquad \beta_d(\mathcal{K}) = \text{rank}(\{[g] : g \in Z^d, |\text{supp}(g)| \geq 1, w(g) = |\text{supp}(g)|\}).$$

For $d = 0$, call every $\mathcal{A} \subseteq \mathcal{F}^0$ connected. For $d \geq 1$, call $\mathcal{A} \subseteq \mathcal{F}^d$ connected, if for every $\sigma_1, \sigma_2 \in \mathcal{A}$, there exists a sequence $\tau_1, \ldots, \tau_i \in \mathcal{A}$ with $\tau_1 = \sigma_1$ and $\tau_i = \sigma_2$ such that, for each j, τ_j and τ_{j+1} share a common $(d-1)-$face. Now fix $d \geq 1$ and consider $g \in Z^d$ such that $\text{supp}(g)$ is not connected. Then clearly there exists $\{\mathcal{A}_i : \mathcal{A}_i \subseteq \mathcal{F}^d\}$ such that each $\mathcal{A}_i$ is non-empty and connected; $\mathcal{A}_i \cap \mathcal{A}_j = \emptyset$; for all $\sigma_i \in \mathcal{A}_i$ and $\sigma_j \in \mathcal{A}_j$, σ_i and σ_j do not have a common $(d-1)-$face; and $\text{supp}(g) = \cup_i \mathcal{A}_i$. Let $g_{\mathcal{A}_i} \in C^d$ be such that $\text{supp}(g_{\mathcal{A}_i}) = \mathcal{A}_i$. It is then easy to see that

$$g = \sum_i g_{\mathcal{A}_i}.$$

From the above relation and our assumption that $g \in Z^d$, it necessarily follows that each $g_{\mathcal{A}_i} \in Z^d$. Suppose not. Then there exists i and $\sigma \in \mathcal{F}^{d+1}$ such that

$$\delta_d(g_{\mathcal{A}_i})(\sigma) = \sum_{\tau \in \mathcal{F}^d, \tau \subset \sigma} g_{\mathcal{A}_i}(\tau) = 1.$$

Since no $\sigma_i \in \mathcal{A}_i$ shares a $(d-1)-$face with any $d-$face in $\cup_{j \neq i} \mathcal{A}_j$, the above necessarily implies that $\delta_d(g)(\sigma) = 1$; which is a contradiction.

36

From the above discussion and that the fact that the rank only depends upon independent elements, we have, for each $d \geq 0$,

$$\beta_d(\mathcal{K}) = \text{rank}(\{[g] : g \in Z^d, |\text{supp}(g)| \geq 1, w(g) = |\text{supp}(g)|, \text{supp}(g) \text{ is connected}\}).$$

If we define $1_g = \mathbf{1}[g \in Z^d]$ and

(C.4) $$\mathcal{G}_d(\mathcal{K}) = \{g \in C^d : |\text{supp}(g)| \geq 1, w(g) = |\text{supp}(g)|, \text{supp}(g) \text{ is connected}\},$$

then the above discussion yields the upper bound

(C.5) $$\beta_d(\mathcal{K}) \leq \sum_{g \in \mathcal{G}_d(\mathcal{K})} 1_g.$$

We now obtain a lower bound for the different Betti numbers. As usual, let $N_d(\mathcal{K})$ denote the number of isolated $d-$faces in the given simplicial complex $\mathcal{K}$. When $d = 0$, we have

$$N_0(\mathcal{K}) - 1[N_0(\mathcal{K}) = |\mathcal{F}^0|] \leq \beta_0(\mathcal{K}).$$

This shows that the number of isolated vertices is a lower bound for $\beta_0(\mathcal{K})$ except in one particular case when all vertices in $\mathcal{K}$ are isolated. For $d \geq 1$, however, one can easily come up with several examples when the number of isolated $d-$faces exceeds $\beta_d(\mathcal{K})$. From this, it follows that $\beta_d(\mathcal{K})$ for $d \geq 1$ needs to treated a little differently.

Fix $d \geq 1$. In contrast to the setup used for the upper bound, we will assume here that the $d-$skeleton $\mathcal{K}^d$ of the given simplicial complex $\mathcal{K}$ is complete. Consider $\tilde{N}_d(\mathcal{K})$, which we define to be the number of disjoint isolated $d-$faces in $\mathcal{K}$. We call a $d-$face disjoint isolated if it is isolated in $\mathcal{K}$ and none of its neighbouring $d-$faces (i.e.,$\sigma' \in \mathcal{F}^d$ which share a $(d-1)-$face with σ) are isolated. Let $\sigma_1, \ldots, \sigma_{\tilde{N}_d}$ be all the disjoint isolated $d-$faces in $\mathcal{K}$ and let $g_{\sigma_1}, \ldots, g_{\sigma_{\tilde{N}_d}}$ be their associated indicator $d-$cochains. We claim that

(C.6) $$\text{rank}(\{[g_{\sigma_1}], \ldots, [g_{\sigma_{\tilde{N}_d}}]\}) = \tilde{N}_d.$$

If $\tilde{N}_d = 1$, then the above is obviously true. We need to verify it for $\tilde{N}_d > 1$. Suppose not. Then there exists $I \subseteq \{1, \ldots, \tilde{N}_d\}$ such and a $f \in C^{d-1}$ such that

(C.7) $$\sum_{i \in I} g_{\sigma_i} = \delta_{d-1}(f) \text{ i.e., } \sum_{i \in I} g_{\sigma_i} \in [0].$$

Fix an arbitrary $i \in I$ and consider σ_i. Suppose $\sigma_i = \{v_0, \ldots, v_d\}$. Let $v_{d+1} \in \mathcal{F}^0$ be such that $v_{d+1} \notin \sigma_i$. For $j = 0$ to d, let $\tau_{ij} = \{v_0, \ldots, v_{d+1}\} \backslash \{v_j\}$. Note that each τ_{ij} is a $d-$face and it belongs to $\mathcal{F}^d$ because of our assumption that $\mathcal{K}^d$ is complete. Clearly none of the τ_{ij}'s are σ_k's for any i, j, k. Thus, from (C.7), we have that for $i, j \in \{0, \ldots, d\}$

$$\delta_{d-1}(f)(\tau_{ij}) = 0 \text{ and } \delta_{d-1}(f)(\sigma_i) = 1.$$

But by the property of the coboundary operator, we derive a the contradiction that

$$\delta_d(\delta_{d-1}(f))[v_0, \ldots, v_{d+1}] = \sum_{j=0}^{d} \delta_{d-1}(f)(\tau_{ij}) + \delta_{d-1}(f)(\sigma_i) = 0.$$

37

Thus (C.6) holds even when $\tilde{N}_d > 1$. From (C.6) and (C.3), we now have the lower bound

$$(C.8) \qquad\qquad \tilde{N}_d(\mathcal{K}) \leq \beta_d(\mathcal{K}).$$

Some preliminary results are required before we prove Lemma 4.2 for $d \geq 2$. The below result can be proved using above arguments. But we give a slightly more general proof.

Lemma C.1. *Fix $d \geq 0$. Let $\mathcal{K}$ be a simplicial complex with complete $d-$skeleton such that $|\mathcal{F}^0| \geq d + 2$. Let $\sigma \in \mathcal{F}^d$ and g_σ be the associated indicator $d-$cochain. Then $g_\sigma \in \mathcal{G}_d(\mathcal{K})$. Also σ is isolated in $\mathcal{K}$ if and only if $g_\sigma \in Z^d$.*

Proof. It is easy to see that if $g_\sigma \notin B^d$, then the desired result follows. For $d = 0$, since $|\mathcal{F}^0| \geq 2$ and $B^0 = \{0, 1\}$, it is immediate that $g_\sigma \notin B^0$. Let $d \geq 1$ and suppose that $g_\sigma \in B^d(\mathcal{K})$, i.e., there exists $f \in C^{d-1}(\mathcal{K})$ such that $\delta_{d-1}(f) = g_\sigma$. Since $|\mathcal{F}^0(\mathcal{K})| \geq d + 2$, it follows that there exists $v \notin \sigma$. Construct a new simplicial complex $\mathcal{K}_*$ such that $\mathcal{K}_*^d = \mathcal{K}^d$ and $\{v\} \cup \sigma$ is the only $(d+1)-$face in $\mathcal{K}_*$. Clearly $C^{d-1}(\mathcal{K}_*) = C^{d-1}(\mathcal{K})$, $C^d(\mathcal{K}_*) = C^d(\mathcal{K})$, and $B^d(\mathcal{K}_*) = B^d(\mathcal{K})$, and hence $g_\sigma \in B^d(\mathcal{K}_*)$ and $f \in C^{d-1}(\mathcal{K}_*)$ with $g_\sigma = \delta_{d-1}(f)$. For $\delta_d : C^d(\mathcal{K}_*) \to C^{d+1}(\mathcal{K}_*)$, we have $\delta_d(g_\sigma)(\{v\} \cup \sigma) = 1$. But this is a contradiction, since by definition of boundary maps, $\delta_d(\delta_{d-1}(f))(\{v\} \cup \sigma) = 0$. Hence $g_\sigma \notin B^d(\mathcal{K}_*)$, which implies that $g_\sigma \notin B^d(\mathcal{K})$. The desired result now follows. $\qquad\square$

Denoting by $N_d(\mathcal{K})$, the number of isolated $d-$faces in $\mathcal{K}$, we have the following consequence of the above result :

$$(C.9) \qquad N_d(\mathcal{K}) + \sum_{g \in \mathcal{G}_d(\mathcal{K}), |supp(g)| > 1} 1_g = \sum_{g \in \mathcal{G}_d(\mathcal{K})} 1_g.$$

Lemma C.2. *Fix $d \geq 1$. Let $\mathcal{K}$ be a simplicial complex with complete $d-$skeleton such that $|\mathcal{F}^0| \geq 2d + 4$. Let $\sigma, \sigma' \in \mathcal{F}^d$ be such that $\sigma \neq \sigma'$ and σ and σ' share a common $(d-1)-$face. If $g_{\sigma,\sigma'}$ is that $g \in C^d$ which has $\text{supp}(g) = \{\sigma, \sigma'\}$, then $g_{\sigma,\sigma'} \in \mathcal{G}_d(\mathcal{K})$.*

Proof. Since $|\text{supp}(g_{\sigma,\sigma'})| = 2$ and $\text{supp}(g_{\sigma,\sigma'})$ is connected, it suffices to show that $w(g_{\sigma,\sigma'}) = 2$, where $w(\cdot)$ is as in (C.2). By repeating the argument below (C.7), we can derive a contradiction to $[g_{\sigma,\sigma'}] = [0]$ and so $w(g_{\sigma,\sigma'}) \neq 0$. Suppose $w(g_{\sigma,\sigma'}) = 1$. Then there exists some indicator $d-$co-chain $g_{\hat{\sigma}}$ associated with the $d-$face $\hat{\sigma}$ such that

$$g_{\sigma,\sigma'} + g_{\hat{\sigma}} = \delta_{d-1}(f).$$

Now either $\hat{\sigma}$ belongs to $\{\sigma, \sigma'\}$ or not. In the former case, without loss of generality, assume that $\hat{\sigma} = \sigma'$. Then it follows that $g_\sigma = g_{\sigma,\sigma'} + g_{\hat{\sigma}} = \delta_{d-1}(f)$. Repeating the argument below (C.7) (with σ_i there replaced by σ), we get a contradiction. Now suppose that $\hat{\sigma} \notin \{\sigma, \sigma'\}$. Since $|\mathcal{F}^0| \geq 2d + 4$, there exists $v_{d+1} \in \mathcal{F}^0$ such that $v_{d+1} \notin \sigma \cup \sigma' \cup \hat{\sigma}$. again repeating the argument below (C.7), we again get a contradiction and hence $w(g_{\sigma,\sigma'}) = 2$ as required. $\quad\square$

Theorem C.3. *Fix $d \geq 1$. Let $\mathcal{K}$ be a simplicial complex with complete $d-$skeleton such that $|\mathcal{F}^0| \geq 2d + 4$. Let $N_d(\mathcal{K})$ denote the number of isolated $d-$faces in $\mathcal{K}$. Then*

$$|\beta_d(\mathcal{K}) - N_d(\mathcal{K})| \leq 3 \sum_{g \in \mathcal{G}_d(\mathcal{K}), |\text{supp}(g)| > 1} 1_g.$$

Proof. Let $\tilde{N}_d(\mathcal{K})$ denote the number of disjoint isolated $d-$faces in $\mathcal{K}$. Then from (C.8), (C.5) and (C.9), we have

$$\tilde{N}_d(\mathcal{K}) \leq \beta_d(\mathcal{K}) \leq \sum_{g \in \mathcal{G}_d(\mathcal{K})} 1_g \; ; \; \tilde{N}_d(\mathcal{K}) \leq N_d(\mathcal{K}) \leq \sum_{g \in \mathcal{G}_d(\mathcal{K})} 1_g.$$

Combining the above two relations, we get

$$(\text{C.10}) \quad |\beta_d(\mathcal{K}) - N_d(\mathcal{K})| \leq \sum_{g \in \mathcal{G}_d(\mathcal{K})} 1_g - \tilde{N}_d(\mathcal{K}) = \sum_{g \in \mathcal{G}_d(\mathcal{K})} 1_g - N_d(\mathcal{K}) + N_d(\mathcal{K}) - \tilde{N}_d(\mathcal{K}).$$

Let $1(\sigma) = \mathbf{1}[\sigma \text{ is an isolated } d-\text{face in } \mathcal{K}]$ and $\hat{1}(\sigma) = \mathbf{1}[\sigma \text{ is a disjoint isolated } d-\text{face in } \mathcal{K}]$. Then, we have that

$$N_d(\mathcal{K}) - \tilde{N}_d(\mathcal{K}) = \sum_{\sigma \in \mathcal{F}^d} [1(\sigma) - \hat{1}(\sigma)] \leq \sum_{\sigma \in \mathcal{F}^d} \sum_{\sigma'} 1(\sigma) 1(\sigma'),$$

where the second sum is over all neighbouring $d-$faces of σ. For a neighbouring $d-$face σ',

$$1(\sigma) 1(\sigma') \leq 1_{g_{\sigma,\sigma'}},$$

where $g_{\sigma,\sigma'}$ is that $g \in C^d$ and $\text{supp}(g) = \{\sigma, \sigma'\}$. From the above discussion, we have

$$N_d(\mathcal{K}) - \tilde{N}_d(\mathcal{K}) \leq 2 \sum_{g \in \mathcal{G}_d(\mathcal{K}), |\text{supp}(g)|=2} 1_g,$$

The factor 2 comes because $g_{\sigma,\sigma'} = g_{\sigma',\sigma}$. Now using (C.9) in (C.10), the proof is complete. $\square$

Lemma C.4. [KP14, (3.5), (5.1)] *Let $d \geq 2$. Consider the random $d-$complex $Y_{n,d}(p_n)$ with*

$$p_n = \frac{d \log n + c - \log(d!)}{n}$$

for some fixed $c \in \mathbb{R}$. Let $N_{d-1}(Y_{n,d}(p_n))$ be the number of isolated $(d-1)-$faces in $Y_{n,d}(p_n)$. Also let $\mathcal{G}_{d-1}(Y_{n,d}(p_n))$ be as in (C.4). Then

$$\lim_{n \to \infty} \mathbb{E}\left[\sum_{g \in \mathcal{G}_{d-1}(Y_{n,d}(p_n))} 1_g - N_{d-1}(Y_{n,d}(p_n)) \right] = \lim_{n \to \infty} \mathbb{E}\left[\sum_{g \in \mathcal{G}_{d-1}(Y_{n,d}(p_n)), |\text{supp}(g)|>1} 1_g \right] = 0.$$

In [KP14] (see in particular Section 5 there), each $g \in \mathcal{G}_{d-1}(Y_{n,d}(p_n))$ is identified by an appropriate hypergraph H. $X(H)$ is the number of $d-$faces in $Y_{n,d}(p_n)$ that contain an odd number of faces of $\text{supp}(g)$. Hence the result follows from the following inequality :

$$\mathbb{E}[1_g] = (1 - p_n)^{X(H)} \leq e^{-p_n X(H)}.$$

Proof of Lemma 4.2 for $d \geq 2$. This is easy to see from Theorem C.3 and Lemma C.4. $\square$

References

[Ale95]   K. S. Alexander. Percolation and minimal spanning forests in infinite graphs. *Ann. Probab.*, 23(1):87–104, 1995.

[AR02]   M.J.B. Appel and R.P. Russo. The connectivity of a graph on uniform points on $[0, 1]^d$. *Statistical Probability Letters*, 60:351–357, 2002.

[BBC14]   C. Bajo, B. Burdick, and S. Chmutov. On the Tutte–Krushkal–Renardy polynomial for cell complexes. *Journal of Combinatorial Theory, Series A*, 123(1):186 – 201, 2014.

[BK14]   O. Bobrowski and M. Kahle. Topology of random geometric complexes: a survey. *Journal of Applied and Computational Topology*, pages 1–34, 2014.

[BM19]   C. A. N. Biscio and J. Møller. The accumulated persistence function, a new useful functional summary statistic for topological data analysis, with a view to brain artery trees and spatial point process applications. *Journal of Computational and Graphical Statistics*, 28(3):671–681, 2019.

[Bol01]   B. Bollobàs. *Random graphs*, volume 73 of *Cambridge Studies in Advanced Mathematics*. Cambridge University Press, Cambridge, 2001.

[Bub15]   P. Bubenik. Statistical topological data analysis using persistence landscapes. *Journal of Machine Learning Research*, 16(1):77–102, 2015.

[Car09]   G. Carlsson. Topology and data. *Bulletin of the Ameriancan Mathematical Society*, 46(2):255–308, 2009.

[Car14]   G. Carlsson. Topological pattern recognition for point cloud data. *Acta Numerica*, 23:289–368, 2014.

[CB15]   W. Crawley-Boevey. Decomposition of pointwise finite-dimensional persistence modules. *Journal of Algebra and Its Applications*, 14(05):1550066, 2015.

[CCFK12]   D. Cohen, A. Costa, M. Farber, and T. Kappeler. Topology of random 2-complexes. *Discrete and Computational Geometry*, 47(1):117–149, 2012.

[CCK15]   M. J Catanzaro, V. Y. Chernyak, and J. R. Klein. Kirchhoff's theorems in higher dimensions and reidemeister torsion. *Hom. Hom. Applns.*, 17(1):165–189, 2015.

[CCSG⁺09]   F. Chazal, D. Cohen-Steiner, M. Glisse, L. J. Guibas, and S. Y. Oudot. Proximity of persistence modules and their diagrams. In *Proceedings of the twenty-fifth annual Symposium on Computational geometry*, pages 237–246. ACM, 2009.

[CDSGO16]   F. Chazal, V. De Silva, M. Glisse, and S. Oudot. *The structure and stability of persistence modules*. Springer, 2016.

[CDSO14]   F. Chazal, V. De Silva, and S. Y. Oudot. Persistence stability for geometric complexes. *Geometriae Dedicata*, 173(1):193–214, 2014.

[CF15a]   A. E. Costa and M. Farber. The asphericity of random 2-dimensional complexes. *Random Structures and Algorithms*, 46(2):261–273, 2015.

[CF15b]   A. E. Costa and M. Farber. Geometry and topology of random 2-complexes. *Israel Journal of Mathematics*, 209(2):883–927, 2015.

[CFK12]   A.E. Costa, M. Farber, and T. Kappeler. Topics of stochastic algebraic topology. *Electronic Notes in Theoretical Computer Science*, 283(0):53 – 70, 2012. Proceedings of the workshop on Geometric and Topological Methods in Computer Science (GETCO).

[CLR09]   T. H. Cormen, C.E. Leiserson, and R.L. Rivest. *Introduction to Algorithms*. MIT Press and McGraw-Hill., 2009.

[CSEH07]   D. Cohen-Steiner, H. Edelsbrunner, and J. Harer. Stability of persistence diagrams. *Discrete and Computational Geometry*, 37(1):103–120, 2007.

[CSEHM10]   D. Cohen-Steiner, H. Edelsbrunner, J. Harer, and Y. Mileyko. Lipschitz functions have l p-stable persistence. *Foundations of computational mathematics*, 10(2):127–139, 2010.

[DE93]   C J A Delfinado and H Edelsbrunner. An incremental algorithm for Betti numbers of simplicial complexes. In *Proceedings of the ninth annual Symposium on Computational geometry*, pages 232–239. ACM, 1993.

[DKM09]   A M. Duval, C J. Klivans, and J L. Martin. Simplicial matrix-tree theorems. *Transactions of the American Mathematical Society*, 361(11):6073–6114, 2009.

[DKM16]   A. M. Duval, C. J. Klivans, and J. L. Martin. Simplicial and cellular trees. In A. Beveridge, J. R. Griggs, L. Hogben, G. Musiker, and P. Tetali, editors, *Recent Trends in Combinatorics*, pages 713–752. Springer International Publishing, Cham, 2016.

[EH10]   H. Edelsbrunner and J.L. Harer. *Computational Topology, An Introduction*. American Mathematical Society, Providence, RI, 2010.

[EK09]   S. N. Ethier and T. G. Kurtz. *Markov processes: characterization and convergence*, volume 282. John Wiley & Sons, 2009.

[ELZ02]   H. Edelsbrunner, D. Letscher, and A. Zomorodian. Topological persistence and simplification. *Discrete and Computational Geometry*, 28(4):511–533, 2002.

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

[FK16]   A. Frieze and M. Karoǹski. *Introduction to Random Graphs*. Cambridge University Press, 2016.

[Fri85]   A. Frieze. On the value of a random minimum spanning tree problem. *Discrete Applied Mathematics*, 10(1):47 – 56, 1985.

[Hat02]   A. Hatcher. *Algebraic Topology*. Cambridge University Press, Cambridge, New York, 2002.

[Hen82]   N. Henze. The limit distribution for maxima of "weighted" rth-nearest-neighbour distances. *Journal of Applied Probability*, 19(2):344–354, 1982.

[HK19]   M. Hino and S. Kanazawa. Asymptotic behavior of lifetime sums for random simplicial complex processes. *Journal of the Mathematical Society of Japan*, 71(3):765–804, 2019.

[HR05]   T. Hsing and H. Rootzen. Extremes on trees. *Annals of Probability*, 33(1):413–444, 2005.

[HS16]   Y. Hiraoka and T. Shirai. Tutte polynomials and random-cluster models in bernoulli cell complexes (stochastic analysis on large scale interacting systems). *RIMS Kokyuroku Bessatsu*, (59):289–304, 2016.

[HS17]   Y. Hiraoka and T. Shirai. Minimum spanning acycle and lifetime of persistent homology in the linial–meshulam process. *Random Structures & Algorithms*, 51(2):315–340, 2017.

[IKR+11] M. E.H. Ismail, E. Koelink, V. Reiner, A. M. Duval, C. J. Klivans, and J. L. Martin. Cellular spanning trees and laplacians of cubical complexes. *Advances in Applied Mathematics*, 46(1):247 – 274, 2011.

[JLR00]   S. Janson, T. Łuczak, and A. Rucinski. *Random Graphs*. Wiley-Interscience Series in Discrete Mathematics and Optimization. Wiley, New York, 2000.

[Kah14a] M. Kahle. Sharp vanishing thresholds for cohomology of random flag complexes. *Annals of Mathematics*, pages 1085–1107, 2014.

[Kah14b] M. Kahle. Topology of random simplicial complexes: a survey. *AMS Contemporary Mathematics*, 620:201–222, 2014.

[Kah17]   M. Kahle. Random simplicial complexes. pages 581–603, 2017.

[Kal83]   G. Kalai. Enumeration of $\mathbb{Q}$-acyclic simplicial complexes. *Israel Journal of Mathematics*, 45(4):337–351, 1983.

[KKL19]   S. Kališnik, V. Kurlin, and D. Lešnik. A higher-dimensional homologically persistent skeleton. *Advances in Applied Mathematics*, 102:113–142, 2019.

[KP14]   M. Kahle and B. Pittel. Inside the critical window for cohomology of random k-complexes. *Random Structures and Algorithms*, 2014.

[KR14]   V. Krushkal and D. Renardy. A polynomial invariant and duality for triangulations. *Electronic Journal of Combinatorics*, 21(3), 2014.

[Kru56]   J. B. Kruskal. On the shortest spanning subtree of a graph and the traveling salesman problem. *Proceedings of the American Mathematical Society*, (7):48—50, 1956.

[LM06]   N. Linial and R. Meshulam. Homological connectivity of random 2-complexes. *Combinatorica*, 26(4):475–487, 2006.

[LNPR19]   N. Linial, I. Newman, Y. Peled, and Y. Rabinovich. Extremal hypercuts and shadows of simplicial complexes. *Israel Journal of Mathematics*, 229(1):133–163, 2019.

[LP16]   N. Linial and Y. Peled. On the phase transition in random simplicial complexes. *Annals of Mathematics*, 184(3):745–773, 2016.

[LP17]   R. Lyons and Y. Peres. *Probability on trees and networks*, volume 42. Cambridge University Press, 2017.

[Lyo09]   R. Lyons. Random complexes and ℓ^2-betti numbers. *Journal of Topology and Analysis*, 01(02):153–175, 2009.

[MNRR15]   R. Mathew, I. Newman, Y. Rabinovich, and D. Rajendraprasad. Boundaries of Hypertrees and Hamiltonian Cycles in Simplicial Complexes. arXiv:1507.04471, 2015.

[Mun84]   J.R. Munkres. *Elements of Algebraic Topology*. Addison-Wesley, 1984.

[MW09]   R Meshulam and N Wallach. Homological connectivity of random k-dimensional complexes. *Random Structures and Algorithms*, 34(3):408–417, 2009.

[Oxl03]   J. Oxley. What is a matroid? *Cubo Matemática Educacional*, 5(3):179–218, 2003.

[Pen97]   M. D Penrose. The longest edge of the random minimal spanning tree. *Annals of Applied Probability*, pages 340–361, 1997.

[Pen03]   M. D. Penrose. *Random geometric graphs*, volume 5 of *Oxford Studies in Probability*. Oxford University Press, Oxford, 2003.

[PET⁺14]   G. Petri, P. Expert, F. Turkheimer, R. Carhart-Harris, D. Nutt, P.J. Hellyer, and F. Vaccarino. Homological scaffolds of brain functional networks. *Journal of The Royal Society Interface*, 11(101), 2014.

[Res13]   S. I. Resnick. *Extreme values, regular variation and point processes*. Springer, 2013.

[ST86]   J. M. Steele and L. Tierney. Boundary domination and the distribution of the largest nearest-neighbor link in higher dimensions. *Journal of Applied Probability*, 23(2):524–528, 1986.

[Ste69]   V. E. Stepanov. Combinatorial algebra and random graphs. *Theory of Probability and Its Applications*, 14(3):373–399, 1969.

[STY17]   P. Skraba, G. Thoppe, and D. Yogeshwaran. Randomly Weighted $d-$complexes: Minimal Spanning Acycles and Persistence Diagrams, 2017. arXiv:1701.00239v1.

[SVJ13]   P. Skraba and M. Vejdemo-Johansson. Persistence modules: algebra and algorithms. *arXiv:1302.2015*, 2013.

[Szp30]   E. Szpilrajn. Sur l'extension de l'ordre partiel. *Fundamenta mathematicae*, 16(1):386–389, 1930.

[vdH16]   R. van der Hofstad. *Random Graphs and Complex Networks, Volume 1. Cambridge Series in Statistical and Probabilistic Mathematics*. Cambridge university press, 2016.

[Wel76]   D. J. A. Welsh. *Matroid theory*. Academic Press [Harcourt Brace Jovanovich, Publishers], London-New York, 1976. L. M. S. Monographs, No. 8.

[Wik16]   Wikipedia. Minimum spanning tree — wikipedia, the free encyclopedia, 2016. [Online; accessed 10-February-2016].

[ZC05]   A. Zomorodian and G. Carlsson. Computing persistent homology. *Discrete and Computational Geometry*, 33(2):249–274, 2005.