

# THE BULK AND THE EXTREMES OF MINIMAL SPANNING ACYCLES AND PERSISTENCE DIAGRAMS OF RANDOM COMPLEXES

NICOLAS FRAIMAN, SAYAN MUKHERJEE, AND GUGAN THOPPE

ABSTRACT. Frieze showed that the expected weight of the minimum spanning tree (MST) of the uniformly weighted graph converges to  $\zeta(3)$ . Recently, this result was extended to a uniformly weighted simplicial complex, where the role of the MST is played by its higher-dimensional analogue—the Minimum Spanning Acycle (MSA). In this work, we provide results on the convergence of the histogram of the weights in random MSAs—both in the bulk and in the extremes. In particular, we focus on the ‘incomplete’ setting, where one has access only to a fraction of the potential face weights. Our first result is that the empirical distribution of the MSA weights asymptotically converges to a measure based on the shadow—the complement of graph components in higher dimensions. As far as we know, this result is the first to explore the connection between the MSA weights and the shadow. Our second result is that the extremal weights converge to an inhomogeneous Poisson point process. A interesting consequence of our two results is that we can also state the distribution of the death times in the persistence diagram corresponding to the above weighted complex, a result of interest in applied topology.

## 1. INTRODUCTION

Random graphs, as models for pairwise relations between vertices, have had a significant impact in discrete mathematics, computer science, engineering, and statistics. Modern-day data analysis also involves looking at higher order relations, resulting in a growing interest to investigate models of random simplicial complexes [10, 8, 11] analogous to random graph models. Our contributions in this paper can be viewed in this light. Specifically, we consider a mean-field model and provide a complete description of the distribution of weights in the minimal spanning tree and also of those in its higher dimensional analogue—the Minimal Spanning Acycle (MSA). As a corollary, we also obtain the distribution of death times in the associated persistence diagram.

In 2014, Robert Adler [1] had summed up the state-of-the art in stochastic topology as follows: while a lot is already known about the asymptotic behaviour of the sums of these weights and death times, respectively, “almost nothing is known” about the distributional properties of the individual values. The work in [12] was the first to partially address this gap. There, the distribution of the extremal MSA face weights and the extremal death times was studied. In this work, not only do we extend those results to a more general scenario but also, and more importantly, describe the behaviour in the bulk as well. In fact, this latter result is new even in the graph case.

**1.1. Summary of Main Results.** Leaving the details to Section 2, we now provide some brief background and a quick summary of our main results along with some visual illustrations.

The Erdős-Rényi graph, denoted by  $G(n, p)$ , is a random graph on  $n$  vertices where each edge is present with probability  $p$  independently. An analogue of this model in higher dimensions is the Linial-Meshulam complex, which we denote here by  $Y_d(n, p)$  or by  $Y(n, p)$  when the dimension  $d$  is clear. This complex has  $n$  vertices and the complete  $(d - 1)$ -dimensional skeleton; further, the  $d$ -dimensional faces are included with probability  $p$  independently. Clearly,  $Y_1(n, p)$  and  $G(n, p)$  are equivalent. Here and henceforth, note that a  $k$ -dimensional face or simply a  $k$ -face refers to a subset of the vertex set with cardinality  $k + 1$ , while the  $k$ -skeleton of a simplicial complex  $\mathcal{K}$  is the collection of all those faces in  $\mathcal{K}$  whose dimension is less than or equal to  $k$ .

Our focus in this paper is on a weighted analogue of  $Y(n, p)$ . Specifically, we look at the case where the weights of the  $d$ -faces in  $Y(n, p)$  come from the uniform  $[0, 1]$  distribution, while the lower dimensional faces have weights all equal to zero. This setup models the scenario where only a fraction of the actual  $d$ -face weights are available to the user. Our main results concern this complex and can be informally stated as follows.

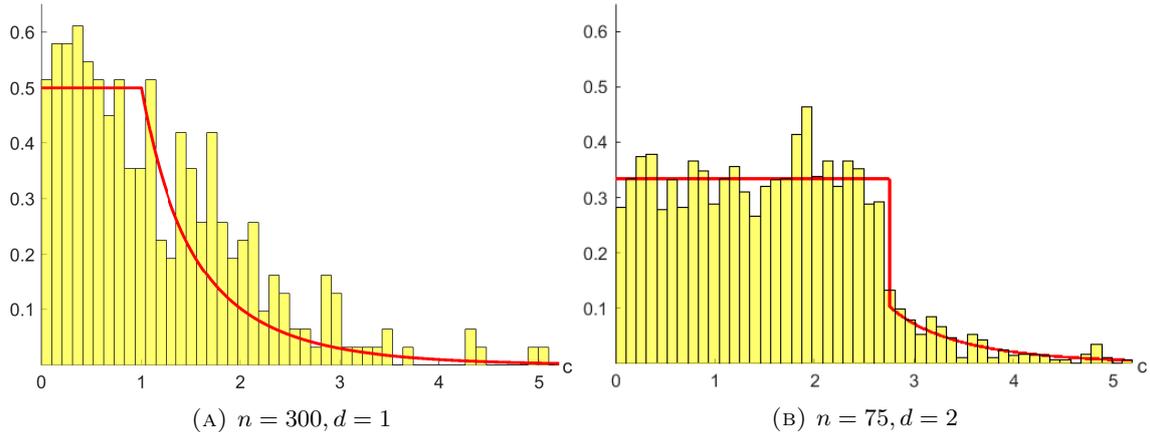


FIGURE 1. Normalized histogram (yellow) of  $\{nw(\sigma) : \sigma \in M_{n,1}\}$  and the density (red) of the shadow based measure given in (8).

**Theorem 1** (Informal summary of results in Sections 2 and 5.3). *Let  $d \geq 1$  and  $p \in (0, 1]$ . In addition, whenever it exists, let  $M_{n,p}$  denote a  $d$ -dimensional MSA (or  $d$ -MSA in short) in the weighted  $Y(n, p)$  complex. Then, as  $n \rightarrow \infty$ , the following claims hold.*

- (1) Bulk: *The empirical measure related to  $\{npw(\sigma) : \sigma \in M_{n,p}\}$  converges to a measure  $\mu$ , whose density is one minus the asymptotic shadow density of  $Y(n, p)$ ; see (8).*
- (2) Extremes:  *$\{npw(\sigma) - d \log(n) + \log(d!) : \sigma \in M_{n,p}\}$  converges to a Poisson point process.*

Some brief comments on this result are in order. First, the shadow of a complex is, loosely, the complement of graph components in higher dimensions. Next, the scaling used in the second claim gives more prominence to the extremal values amongst the  $d$ -MSA weights. This follows from the fact that these extremal values are of the order  $d \log n/n$ ; see Section 1.3 for further details. Finally, the second claim is new for  $p \in (0, 1)$ ; the specific case of  $p = 1$  also appears in [12]. The first claim, in contrast, is new for all  $p$ ; in fact, it is new even in the graph case. Also, we are the first to highlight connection between the MSA weights and the shadow.

A weighted simplicial complex as described above can also be viewed as an evolving complex which, at time  $t \in \mathbb{R}$ , includes all those faces whose weight is less than or equal to  $t$ . There is a natural persistence diagram associated with this process; this diagram is basically a record of the birth and death times of the different topological holes that appear and disappear as the process evolves. In [12, Theorem 3], it was shown that the set of death times in the  $(d - 1)$ -th persistent groups exactly equals the set of weights in the  $d$ -MSA of the original weighted complex. Consequently, our results can also be stated in terms of the death times corresponding to the weighted  $Y(n, p)$  complex; see Theorems 2, 4, and Section 5.3 for the exact statements.

We now use Figs. 1, 2, and 3 to provide an alternative explanation of Theorem 1. The first two figures relate to our result in the  $p = 1$  case, while the third one concerns the  $p \in (0, 1)$  case.

There are two scenarios in Fig. 1, one corresponding to a sample of the weighted  $Y_1(300, 1)$  complex and the other to that of the weighted  $Y_2(75, 1)$  complex. In both the scenarios, the yellow plot shows the normalized histogram of the set  $\{nw(\sigma) : \sigma \in M_{n,1}\}$ , while the red one shows the density of the shadow based limiting measure defined in (8). Observe that the red and the yellow plots more or less resemble each other. This is the crux of our first claim above which, loosely, states that the difference between these two plots decays to zero as  $n \rightarrow \infty$ .

Fig. 2 again shows two distinct scenarios, the first one corresponds to a sample of the weighted  $Y_1(500, 1)$  complex while the second one deals with that of the weighted  $Y_2(100, 1)$  complex. This time, the figure shows a plot of the values in  $\{nw(\sigma) - d \log(n) + \log(d!) : \sigma \in M_{n,1}\}$ . As stated before, this scaling gives more prominence to the extremal weights. Accordingly, observe that the values on the extreme right of the figure have distinctly spread out. Our second claim states that these extremal values converge to a Poisson point process while the rest go to  $-\infty$ .

Finally in Fig. 3, unlike Fig. 1, we discuss our result on the bulk for  $p \in (0, 1)$ . Again, we consider two broad scenarios and the value of  $(n, d)$  in each is  $(500, 1)$  and  $(50, 2)$ , respectively. In each case, we also consider three different values of  $p$ , as indicated, and look at the  $d$ -MSA in one sample

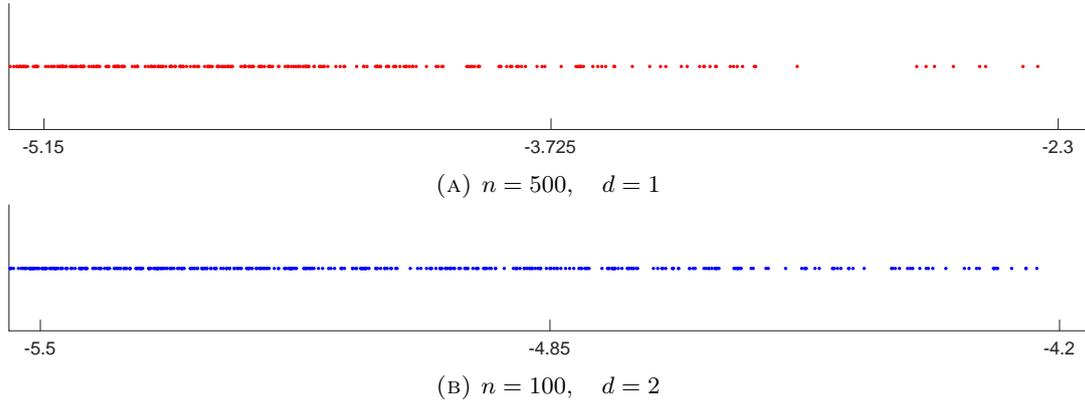


FIGURE 2. Plot of  $\{nw(\sigma) - d \log(n) + \log(d!) : \sigma \in M_{n,1}\}$ .

each of the resultant  $Y_d(n, p)$  complex; we resample if the MSA does not exist. Notice that all the panels have two distinct plots: the blue one is the histogram corresponding to  $\{nw(\sigma) : \sigma \in M_{n,p}\}$ , while the yellow one corresponds to  $\{npw(\sigma) : \sigma \in M_{n,p}\}$ . Clearly, unlike the blue plots, the yellow ones look similar irrespective of the  $p$  values. Our result indeed captures this phenomena. It states that, whatever be the value of  $p$ , the corresponding yellow histogram asymptotically converges to the red curve depicted in Fig. 1. While we haven't shown it, a similar story unfolds in relation to our result on the extremes as well. Therefore, one can conclude that, by uniformly scaling all the weights in the MSA by  $p$ , the limiting behaviour becomes independent of  $p$  itself.

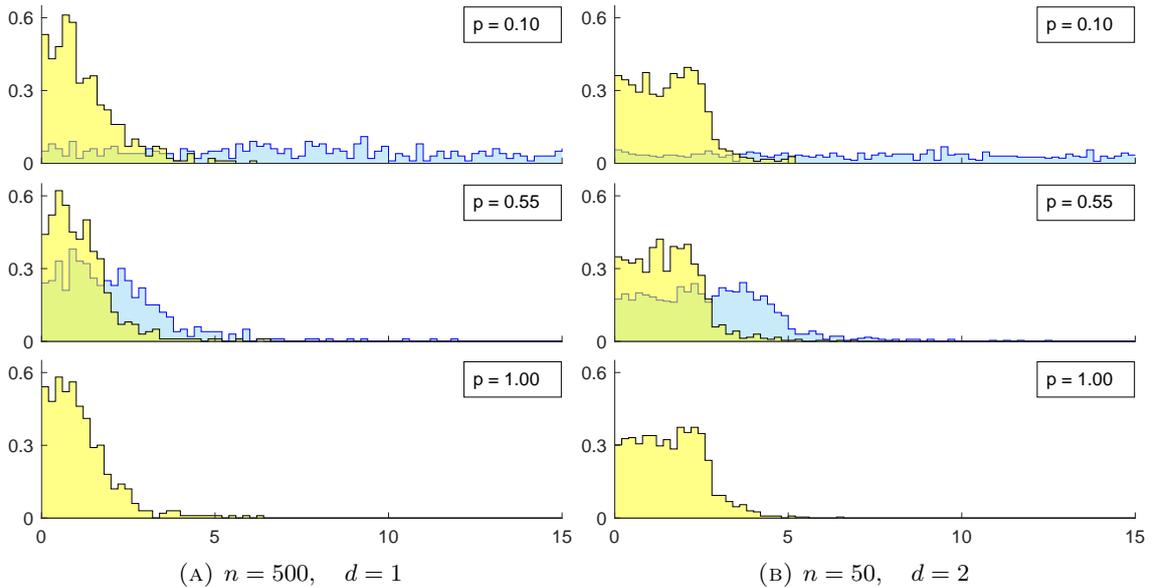


FIGURE 3. Normalized histograms of  $\{npw(\sigma) : \sigma \in M_{n,p}\}$  and  $\{nw(\sigma) : \sigma \in M_{n,p}\}$ , depicted in yellow and blue, respectively. In the last panel, the two histograms coincide since  $p = 1$ .

**1.2. Related Work.** Our work lies at the intersection of two broad strands of research: one concerning component sizes, shadow densities, and homologies of random graphs and random complexes, and the other dealing with the statistics of weights in random minimal spanning trees and acycles. In this section, we look at few of the historical milestones in these two strands.

Erdős and Rényi were the ones who initiated the first strand with their work in [2]. There they showed that  $\log n/n$  is a sharp asymptotic threshold for connectivity in  $G(n, p)$ . And, also that, if  $p_n = (\log n + c)/n$  for  $c \in \mathbb{R}$  then, as  $n \rightarrow \infty$ , almost all the vertices in  $G(n, p_n)$  lie

in one single component; further, the vertices outside this are all isolated and their number has a Poisson distribution with mean  $e^{-c}$ . This result was subsequently refined in [3] and the new statement included the following additional facts. The asymptotic order of the largest component jumps from logarithmic to linear around  $1/n$ . Also, for  $c > 1$ , the size of the largest component in  $G(n, c/n)$ , denoted  $L_n(c)$ , satisfies  $L_n(c)/n \rightarrow 1 - t(c)$ , where  $t(c) \in (0, 1)$  is the unique root of

$$(1) \quad t = e^{-c(1-t)}.$$

A graph is a one dimensional complex, so one can ask if similar phenomena occur in random complexes as well. Since connectivity is associated with the vanishing of the zeroth homology, it is natural to look at the higher order Betti numbers to answer this question. Such a study was first carried out in [8, 11] and it was found that the  $(d-1)$ -th Betti number of  $Y(n, p)$  indeed shows a non-vanishing to vanishing phase transition at  $d \log n/n$ , just as in the graph case. A separate study done in [6] also showed that this Betti number converges in distribution to a Poisson random variable with mean  $e^{-c}$ , when  $p_n = (d \log n - \log(d!) + c)/n$  for some  $c \in \mathbb{R}$ .

The result on component sizes, in contrast, was not so easy to generalize. The challenge was in coming up with a higher dimensional analogue of a component in a simplicial complex. The breakthrough came in [9] with the introduction of the shadow. What motivated the use of the shadow was that, in a sparse graph, a giant component exists if and only if the shadow has a positive density. With this in mind, the behaviour of the  $d$ -dimensional shadow of  $Y(n, p)$  was investigated in [10] and these were the key findings there. One, the size of the  $d$ -shadow changes from  $o(n^{d+1})$  to  $\Omega(n^{d+1})$  at some  $c_d^*/n$ , where  $c_1^* = 1$  but  $c_d^* > 1$  for  $d \geq 2$ . Two, for  $c > c_d^*$ , the  $d$ -shadow of  $Y(n, c/n)$ , denoted  $\text{Sh}(Y(n, c/n))$ , satisfies  $|\text{Sh}(Y(n, c/n))|/\binom{n}{d+1} \rightarrow (1 - t(c))^d$ , where  $t(c) \in (0, 1)$  is the unique root of  $t = \exp[-c(t-t)^d]$ . Note that this latter equation matches the one in (1) when  $d = 1$ . Also,  $|\text{Sh}(Y_1(n, c/n))| = \Theta(n^2)$ , while  $L_1(c) = \Theta(n)$ .

The pioneering work in the second strand was the one in [4] by Frieze. He showed that, given a complete graph on  $n$  vertices with uniform  $[0, 1]$  weights on each edge, the expected sum of weights in the minimum spanning tree converges to  $\zeta(3)$ . This result was recently extended in [5] to the setting of the  $d$ -MSA in the weighted  $Y(n, 1)$  model. Separately, the work in [12] also showed that the extremal weights in this  $d$ -MSA converge to a Poisson point process. As emphasized before, this latter result was the first to deal with the individual values instead of just the sum.

**1.3. Proof Outline.** We now briefly describe how we combine the different facts stated above for proving Theorem 1. Clearly, the weighted  $Y(n, p)$  complex may have some  $d$ -faces missing. The first question then to ask is what is the probability that a  $d$ -spanning acycle, and hence a  $d$ -MSA, exists. As per [12, Lemma 3.6], a sufficient condition to guarantee existence is that the  $(d-1)$ -th Betti number be zero. However, since  $p$  is a constant, this indeed holds for  $Y(n, p)$  asymptotically [8, 11]. Hence, the weighted  $Y(n, p)$  complex does have a  $d$ -MSA a.s. as  $n \rightarrow \infty$ .

For a finite  $n$ , however, there is a non-trivial probability that there may not exist a  $d$ -MSA. Therefore, working directly with the weighted  $Y(n, p)$  complex becomes a bit complicated. Hence, we first derive our results for  $\mathcal{L}(n, p)$ , an intermediary object called the augmented  $d$ -complex, and then extend the same to the original weighted complex. As stated in Definition 1,  $\mathcal{L}(n, p)$  is obtained from the weighted  $Y(n, p)$  complex by adding all the missing  $d$ -faces and then assigning them the largest possible weight, i.e., 1. The  $\mathcal{L}(n, p)$  complex, thus, always has a  $d$ -MSA. Also, asymptotically, its  $d$ -MSA matches the one in the weighted  $Y(n, p)$  complex. Because of this, we now only need to explain how to derive Theorem 1 in the context of  $\mathcal{L}(n, p)$ .

We begin with the first claim in Theorem 1. Let  $\mathcal{L}(n, p)|_{t^-}$  be the simplicial complex obtained from  $\mathcal{L}(n, p)$  by retaining only those faces with weight strictly less than  $t$ . Clearly, the distributions of  $\mathcal{L}(n, p)|_{t^-}$  and  $Y(n, tp)$ , and hence of their shadows, are one and the same. Now, as per [12, Lemma 32], a  $d$ -face with weight  $t$  belongs to the  $d$ -MSA if and only if it does not lie in the shadow of  $\mathcal{L}(n, p)|_{t^-}$ . Given that this face could potentially have been any of those outside of  $\mathcal{L}(n, p)|_{t^-}$ , the probability that it belongs to the  $d$ -MSA must then equal one minus the shadow density of  $\mathcal{L}(n, p)|_{t^-}$ , i.e., the size of shadow divided by the total number of potential  $d$ -faces. Based on this and the shadow density results from [10], the desired result is then easy to see.

Our proof for the second claim builds upon the ideas from [12] and proceeds as follows. As before, note that, for any  $t \in [0, 1]$ , the simplicial complex  $\mathcal{L}(n, p)|_{t^-}$  has the same distribution as  $Y(n, tp)$ . This, in turn, now implies that the distributions of the  $(d-1)$ -th Betti number of

$Y(n, tp)$  and the set of death times exceeding  $t$  in the  $(d - 1)$ -th persistence diagram associated with  $\mathcal{L}(n, p)$  are the same. Separately, due to the equivalence between the MSA weights and the death times [12, Theorem 3], it also follows that the set of  $d$ -MSA weights exceeding  $t$  in  $\mathcal{L}(n, p)$  has the same distribution as the previous two quantities. The phase transition results for the  $(d - 1)$ -th Betti number from [2, 8, 11] then show that the extremal weights in the  $d$ -MSA are  $O(d \log n/n)$ . Separately, the Poisson result from [6] shows that the set of  $d$ -MSA weights that exceed  $(d \log n - \log(d!) + c)/n$  converges in distribution to a Poisson random variable with mean  $e^{-c}$ . The proof of the desired Poisson point process convergence is then a mere technicality.

**1.4. Overview of Contents.** Section 2 provides the necessary terminology as also the formal statements of our main results in the context of  $\mathcal{L}(n, p)$ , both in the bulk and in the extremes. The proof for the result on the bulk is given in Section 3, while the Section 4 deals with that of the extremes. In Section 5, we discuss three distinct extensions of our results for  $\mathcal{L}(n, p)$  including the one for the weighted  $Y(n, p)$  complex. We finally close with a discussion and some future directions. An interesting model discussed here is that of the online setting where the  $d$ -faces are revealed one at a time.

## 2. MAIN RESULTS AND TERMINOLOGY

In this section we state the notation we will use throughout the paper as well the main results.

Throughout,  $d \geq 1$  denotes an arbitrary but fixed dimension of the random simplicial complex. Most of our results and the random objects we consider depend of  $d$ , however as  $d$  is a constant this dependence will not be explicitly highlighted. We use  $\mathcal{K}$  to denote a generic simplicial complex. Further, we use  $\mathcal{F}^d(\mathcal{K})$  or simply  $\mathcal{F}^d$  (when the complex is clear) to denote the  $d$ -dimensional faces in  $\mathcal{K}$ . Similarly,  $\beta_d(\mathcal{K})$  or simply  $\beta_d$  stands for the  $d$ -th reduced Betti number of  $\mathcal{K}$ . For a random variable  $X$ ,  $\|X\|_q = \mathbf{E}(X^q)^{1/q}$  represents its  $L^q$ -th norm. We denote  $|S|$  as the cardinality of a set  $S$ . Lastly, we denote the  $d$ -dimensional Linial-Meshulam complex by  $Y(n, p)$ . Recall this is the random complex on  $n$  vertices with the complete  $(d - 1)$ -dimensional skeleton and where each  $d$ -face is present independently with probability  $p$ .

The following terminology is needed to state our main results. We will assume that the reader is well versed with the basics of simplicial homology and Poisson point processes. In any case, the necessary background can be found in [12, Section 2] and the references therein.

Our first definition concerns a weighted simplicial complex model that we introduce here and refer to as the augmented  $d$ -complex  $\mathcal{L}(n, p)$ . This is obtained from the uniformly weighted Linial-Meshulam complex  $Y(n, p)$  by adding the missing faces and assigning them weight 1.

**Definition 1** (Augmented  $d$ -complex). Let  $\mathcal{K}_n$  denote the complete  $d$ -skeleton on  $n$  vertices and let  $\{U(\sigma) : \sigma \in \mathcal{F}^d\}$  be a collection of i.i.d. uniform  $[0, 1]$  random variables. Then, for  $p \in (0, 1]$  and  $Y$  being a sample of  $Y(n, p)$ , the *augmented  $d$ -complex*  $\mathcal{L}(n, p)$  is the weighted complex  $(\mathcal{K}_n, w)$ , where  $w : \mathcal{K}_n \rightarrow [0, 1]$  is given by:

$$(2) \quad w(\sigma) = \begin{cases} 0 & \text{if } |\sigma| \leq d, \\ U(\sigma) & \text{if } \sigma \in \mathcal{F}^d(Y), \\ 1 & \text{otherwise.} \end{cases}$$

Throughout this work, we mainly work with the  $\mathcal{L}(n, p)$  complex. This is because it is easier to derive our main results first for this model and then extend it to the weighted  $Y(n, p)$  model than to derive it for the latter case directly. The details on the extension can be found in Section 5.

We now describe a filtration or a simplicial process related to  $\mathcal{L}(n, p)$ . Notice that, given any weighted complex  $(\mathcal{K}, w)$ , where the weight function  $w$  is monotone, i.e.,  $w(\sigma) \leq w(\tau)$  whenever  $\sigma \subseteq \tau$ , there is a canonical filtration associated with it. This is given by  $\{\mathcal{K}|_s : s \in \mathbb{R}\}$ , where  $\mathcal{K}|_s := \{\sigma \in \mathcal{K} : w(\sigma) \leq s\}$ . In fact, this process can be seen as an equivalent but alternative description of  $(\mathcal{K}, w)$  itself. The monotonicity of  $w$  is key here. It is only because of this property that  $\mathcal{K}|_s$  is a simplicial complex for each  $s$  and, in turn,  $\{\mathcal{K}|_s : s \in \mathbb{R}\}$  is a well defined filtration. With this in mind, observe that the function  $w$  in Definition 1 is trivially monotone. Therefore,  $\mathcal{L}(n, p)$  also has a filtration associated with it and that is  $\{\mathcal{L}(n, p)|_s : s \in \mathbb{R}\}$ .

Our next aim is to define spanning acycles [7] and minimal spanning acycles, the main objects that we study here. These are higher dimensional generalizations of spanning trees and minimal

spanning trees, respectively. Recall that a spanning tree is a subset of the edges in a connected graph that connects all the vertices together without creating any cycles. In that same spirit, a  $d$ -spanning acycle in a simplicial complex is a subset of  $d$ -faces which when added to  $\mathcal{K}^{d-1}$ , the  $(d-1)$ -skeleton of  $\mathcal{K}$ , kills the existing  $(d-1)$ -th Betti number and also does not create any  $d$ -cycles. Likewise, a minimal spanning acycle in a weighted simplicial complex is simply the spanning acycle with the minimum possible weight. As shown in [12, Lemma 23], a sufficient condition to guarantee the existence of a spanning acycle and, hence, a  $d$ -MSA is  $\beta_{d-1}(\mathcal{K}) = 0$ .

**Definition 2** (Spanning Acycle). Let  $\mathcal{K}$  be a simplicial complex with  $\beta_{d-1}(\mathcal{K}) = 0$ . Then, a  $d$ -spanning acycle of  $\mathcal{K}$  is a set  $S$  of  $d$ -faces in  $\mathcal{K}$  such that  $\beta_{d-1}(\mathcal{K}^{d-1} \sqcup S) = \beta_d(\mathcal{K}^{d-1} \sqcup S) = 0$ .

**Definition 3** (Minimal Spanning Acycle). Let  $(\mathcal{K}, w)$  be a weighted  $d$ -complex with  $\beta_{d-1}(\mathcal{K}) = 0$ . Then, a  $d$ -minimum spanning acycle  $M$  is an element of  $\arg \min_S \{w(S)\}$ , where the minimum is taken over all the  $d$ -spanning acycles and  $w(S)$  represents the sum of weights of the faces in  $S$ .

It is worth noting that minimal spanning acycles are unique when the  $d$ -faces in  $\mathcal{K}$  have unique weights; this follows, e.g., via Kruskal's algorithm [12].

Next, we introduce the concept of a shadow—the main intermediary object we use in order to study the distribution of weights in the  $d$ -MSA of  $\mathcal{L}(n, p)$ . Introduced in [9], a shadow enables one to get around the difficulty of extending the notion of connected components to higher dimensions.

**Definition 4** (Shadow). For a  $d$ -dimensional simplicial complex  $\mathcal{K}$  with vertex set  $V$  and having the complete  $(d-1)$ -skeleton, its  $d$ -shadow is given by

$$(3) \quad \text{Sh}(\mathcal{K}) = \left\{ \sigma \in \binom{V}{d+1} \setminus \mathcal{F}^d : \beta_d(\mathcal{K} \sqcup \sigma) = \beta_d(\mathcal{K}) + 1 \right\},$$

where  $\binom{V}{d+1}$  denotes the set of all  $(d+1)$ -sized subsets of  $V$ .

For  $p \in (0, 1]$ , let  $M(n, p)$  denote the  $d$ -MSA in  $\mathcal{L}(n, p)$ . Separately, let  $D(n, p)$  denote the set of death times in the  $(d-1)$ -th persistence diagram associated with  $\{\mathcal{L}(n, p)|_s : s \in \mathbb{R}\}$ . In analogy with terms for the spectrum of random matrices, we now define the following measures.

**Definition 5** (Bulk Measure). The *bulk measures*  $\mu_{n,p}$  and  $\tilde{\mu}_{n,p}$  are random measures corresponding to the empirical distribution of the re-scaled weights in  $M(n, p)$  and the re-scaled death times in  $D(n, p)$ , respectively. Formally, they are given by

$$(4) \quad \mu_{n,p} := \frac{1}{\binom{n-1}{d}} \sum_{\sigma \in M(n,p)} \delta_{npw(\sigma)} \quad \text{and} \quad \tilde{\mu}_{n,p} := \frac{1}{\binom{n-1}{d}} \sum_{\Delta \in D(n,p)} \delta_{np\Delta}.$$

**Definition 6** (Extremal Measure). The *extremal measures*  $\nu_{n,p}$  and  $\tilde{\nu}_{n,p}$  are random counting measures corresponding to the extremal weights in  $M(n, p)$  and the extremal death times in  $D(n, p)$ , respectively. Their mathematical formulations are

$$(5) \quad \nu_{n,p} := \sum_{\sigma \in M(n,p)} \delta_{npw(\sigma) - d \log(n) + \log(d!)} \quad \text{and} \quad \tilde{\nu}_{n,p} := \sum_{\Delta \in D(n,p)} \delta_{np\Delta - d \log(n) + \log(d!)}.$$

Finally, we define a deterministic measure which will serve as the limiting measure for both  $\mu_{n,p}$  and  $\tilde{\mu}_{n,p}$ . Towards that, for  $d = 1$ , define  $t_* = c_* = 1$ . For  $d \geq 2$ , let  $t_*$  be the unique root in  $(0, 1)$  of the equation

$$(6) \quad (d+1)(1-t) + (1+dt) \log t = 0,$$

and let  $c_* > 0$  be the constant that satisfies  $t_* = e^{-c_*(1-t_*)^d}$ , i.e.,  $c_* = -\log t_*/(1-t_*)^d$ . Further, for both  $d = 1$  and  $d \geq 2$  cases, when  $c \geq c_*$ , let  $t(c)$  denote the smallest root in  $(0, 1)$  of the equation

$$(7) \quad t = e^{-c(1-t)^d}.$$

Now, let  $\mu$  be the measure with density

$$(8) \quad d\mu(x) = \frac{1-s(x)}{d+1} dx,$$

where

$$(9) \quad s(x) = \begin{cases} 0, & x \leq c_*, \\ (1-t(x))^{d+1}, & x > c_*. \end{cases}$$

As shown in [10, Theorem 1.4], value  $s(c)$  above represents the asymptotic density of the shadow of  $Y(n, c/n)$ , i.e.,

$$(10) \quad s(c) = \lim_{n \rightarrow \infty} \frac{|\text{Sh}(Y(n, c/n))|}{\binom{n}{d+1}}, \quad c > 0.$$

With all the ingredients available, we now state our main results. The first one summarizes our findings in Figures 1 and 3. The punchline here is that if you scale the weights of the faces in  $M(n, p)$  by  $p$ , then the corresponding histograms (normalized so that the area under is 1) or equivalently the empirical measures  $\mu_{n,p}$  look similar for any  $p$ . Further, they asymptotically converge to the shadow based measure  $\mu$ . Likewise,  $\tilde{\mu}_{n,p}$  also converges to  $\mu$ .

Let  $K(\rho, \rho')$  denote the Kolmogorov metric between the two measures  $\rho$  and  $\rho'$ , i.e., let

$$K(\rho, \rho') = \sup_{x \in \mathbb{R}} |\rho(-\infty, x] - \rho'(-\infty, x]|.$$

**Theorem 2** (Bulk limit). *Let  $p \in (0, 1]$ . Then, the random measures  $\mu_{n,p}$  and  $\tilde{\mu}_{n,p}$  converge in the Kolmogorov metric to  $\mu$  in  $L^q$ .*

Moreover, if  $f : \mathbb{R} \rightarrow \mathbb{R}$  is a continuous function such that  $\mu|f| < \infty$  and

$$(11) \quad \lim_{b \rightarrow \infty} \sup_{n \geq 0} \|\mu_{n,p}|f| - \mu_{n,p}(|f| \wedge b)\|_q = 0.$$

Then, both  $\mu_{n,p}f$  and  $\tilde{\mu}_{n,p}f$  converge to  $\mu f$  in  $L^q$ .

**Corollary 3.** *Let  $p \in (0, 1]$  and  $f(x) = x^\alpha$  for some  $\alpha > 0$ . Then, both  $\mu_{n,p}f$  and  $\tilde{\mu}_{n,p}f$  converge to  $\mu f$  in  $L^q$ .*

Again, consider the set  $\{pw(\sigma) : \sigma \in M(n, p)\}$  as above. Then, our next result shows that the point process associated with the extremal values in this set, those that exceed  $(d \log n - \log d! + c)/n$  for some  $c \in \mathbb{R}$ , converges to a Poisson point process. The same story is also true for the death times in  $D(n, p)$ . This result for  $p = 1$  was shown in [12, Propositions 35, 36].

**Theorem 4** (Extremes limit). *Let  $p \in (0, 1]$ . Then, both random measures  $\nu_{n,p}$  and  $\tilde{\nu}_{n,p}$  converge vaguely to  $\mathcal{P}$  in distribution, where  $\mathcal{P}$  is the Poisson point process on  $\mathbb{R}$  with intensity  $e^{-x} dx$ .*

*Remark 5.* In Section 5, we discuss three extensions of the above results. The first two relate to cases where the  $d$ -face weights have a generic distribution and/or have additional noisy perturbations. The third one is about the randomly weighted  $Y(n, p)$   $d$ -complex, which is an example of a complex where which not all the potential  $d$ -faces may be present.

*Remark 6.* Note that Theorem 2 applies when  $f$  is additionally bounded as hypothesis (11) is then trivially satisfied. Thus, an immediate consequence of our result is that  $\mu_{n,p}$  converges to  $\mu$  weakly in  $L^q$ . However, and more importantly, our result also applies when  $f$  is unbounded. In particular, Corollary 3 shows that the result holds for the case  $f(x) = x^\alpha$  for  $\alpha > 0$ . This is an important example since Frieze's result [4] concerning the sum of weights in the MST and its recent generalization to higher dimensions by Hino and Kanazawa ([5, Theorem 4.11]) then become special consequences. Moreover, the limiting constant  $I_{d-1}^{(\alpha)}$  obtained in [5] can now be interpreted as the  $\alpha$ -th moment of the measure  $\mu$ .

*Remark 7.* A variant of the second part of Theorem 2, obtained by replacing condition (11) by

$$(12) \quad \lim_{b \rightarrow \infty} \sup_{n \geq 0} \mathbf{P}\{\mu_{n,p}|f| - \mu_{n,p}(|f| \wedge b) \geq \epsilon\} = 0$$

and convergence in  $L^q$  by convergence in probability, also holds. Notice that, while the condition on  $f$  that we require here is weaker the conclusion is also weaker. See Proposition 9 for details.

### 3. BEHAVIOUR IN THE BULK

In this section we prove Theorem 2. The main idea is to connect the weight of the faces in the MSA to the death times in the  $(d-1)$ -th homology in the filtration  $\{\mathcal{L}(n, p)|_s : s \in [0, 1]\}$ . We combine this with the fact that  $\mathcal{L}(n, p)|_s$  is distributed as  $Y(n, ps)$  to obtain the following pointwise (in  $c$ ) convergence result.

**Proposition 8.** *Let  $c \geq 0$ . Then,*

$$(13) \quad \lim_{n \rightarrow \infty} \|\mu_{n,p}[c, \infty) - \mu[c, \infty)\|_q = 0.$$

We postpone the proof of the above result.

*Proof of Theorem 2.* We first show that  $K(\mu_{n,p}, \mu) \rightarrow 0$  in  $L^q$ . Let  $F_{n,p}(x) = \mu_{n,p}(-\infty, x]$  and  $G(x) = \mu(-\infty, x]$ . Fix  $m$  and pick  $c_0, c_1, \dots, c_m$  such that  $c_0 = 0$ ,  $c_m = \infty$  and  $G(c_{i+1}) - G(c_i) = 1/m$ . Then, for  $x \in [c_i, c_{i+1}]$ ,

$$\begin{aligned} |F_{n,p}(x) - G(x)| &\leq \max \{G(c_{i+1}) - F_{n,p}(c_i), F_{n,p}(c_{i+1}) - G(c_i)\} \\ &\leq \max \left\{ G(c_i) + \frac{1}{m} - F_{n,p}(c_i), F_{n,p}(c_{i+1}) - G(c_{i+1}) + \frac{1}{m} \right\} \\ &\leq \sum_{j=1}^m |F_{n,p}(c_j) - G(c_j)| + \frac{1}{m}. \end{aligned}$$

Since the bound is independent of the interval  $[c_i, c_{i+1}]$  we obtain

$$\sup_{x \in \mathbb{R}} |F_{n,p}(x) - G(x)| \leq \sum_{j=1}^m |F_{n,p}(c_j) - G(c_j)| + \frac{1}{m}.$$

Therefore, taking the  $L^q$  norm we get

$$\begin{aligned} \|K(\mu_{n,p}, \mu)\|_q &= \left\| \sup_{x \in \mathbb{R}} |F_{n,p}(x) - G(x)| \right\|_q \leq \left\| \sum_{j=1}^m |F_{n,p}(c_j) - G(c_j)| + \frac{1}{m} \right\|_q \\ &\leq \sum_{j=1}^m \|F_{n,p}(c_j) - G(c_j)\|_q + \frac{1}{m}. \end{aligned}$$

Note that  $G(x) = 1 - \mu(x, \infty)$  and  $\mu(x) = 0$ , likewise  $F_{n,p}(x) = 1 - \mu_{n,p}(x, \infty)$  and  $\mu_{n,p}(x) = 0$  almost surely. Then, by Proposition 8, we have that  $\|F_{n,p}(c_j) - G(c_j)\|_q \rightarrow 0$ . Since  $m$  is arbitrary, it follows that  $\|K(\mu_{n,p}, \mu)\|_q \rightarrow 0$ .

To extend the convergence to unbounded functions, begin by assuming that  $f$  is non-negative. For any  $b > 0$ , triangle inequality gives

$$(14) \quad \|\mu_{n,p}f - \mu f\|_q \leq \|\mu_{n,p}f - \mu_{n,p}(f \wedge b)\|_q + \|\mu_{n,p}(f \wedge b) - \mu(f \wedge b)\|_q + \|\mu f - \mu(f \wedge b)\|_q.$$

Now, pick a large enough  $b > 0$  such that

$$(15) \quad \sup_{n \geq 0} \|\mu_{n,p}f - \mu_{n,p}(f \wedge b)\|_q \leq \epsilon \quad \text{and} \quad \|\mu f - \mu(f \wedge b)\|_q = \mu f - \mu(f \wedge b) \leq \epsilon.$$

Therefore,

$$(16) \quad \|\mu_{n,p}f - \mu f\|_q \leq 2\epsilon + \|\mu_{n,p}(f \wedge b) - \mu(f \wedge b)\|_q.$$

However,  $f \wedge b$  is a bounded continuous function. Hence, since the Kolmogorov metric dominates the Lévy metric which metrizes weak convergence, it follows that

$$(17) \quad \limsup_{n \rightarrow \infty} \|\mu_{n,p}f - \mu f\|_q \leq 2\epsilon.$$

Since  $\epsilon > 0$  is arbitrary, it follows that  $\mu_{n,p}f$  converges  $\mu f$  in  $L^q$ .

Now, consider the case of general  $f$ . Clearly,  $f = f^+ - f^-$ , where  $f^+ = \max\{f, 0\}$  and  $f^- = \max\{-f, 0\}$ . Furthermore, both  $(f^+ - b) \mathbb{1}[f^+ - b]$  and  $(f^- - b) \mathbb{1}[f^- - b]$  are bounded from above by  $(|f| - b) \mathbb{1}[|f| \geq b]$ , whence it follows that  $f^+ - (f^+ \wedge b)$  and  $f^- - (f^- \wedge b)$  are bounded from above by  $|f| - |f| \wedge b$ . Therefore, by repeating the above arguments for both  $f^+$  and  $f^-$ , individually, it is easy to see that the result holds for the case of general  $f$  as well.  $\square$

The following Proposition shows convergence in probability for unbounded functions with a weaker condition as described in Remark 7.

**Proposition 9.** *Let  $p \in (0, 1]$  and  $f : \mathbb{R} \rightarrow \mathbb{R}$  be such that for every  $\epsilon > 0$  equation (12) holds. Then,  $\mu_{n,p}f$  converges to  $\mu f$  in probability.*

*Proof.* Let  $\eta > 0$ . Pick  $b > 0$  so that

$$(18) \quad \sup_{n \geq 0} \mathbf{P}\{|\mu_{n,p}f - \mu(f \wedge b)| > \epsilon\} \leq \eta \quad \text{and} \quad |\mu f - \mu(f \wedge b)| \leq \epsilon.$$

This implies that

$$(19) \quad \{|\mu_{n,p}f - \mu f| > 3\epsilon\} \subseteq \{|\mu_{n,p}f - \mu_{n,p}(f \wedge b)| > \epsilon\} \cup \{|\mu_{n,p}(f \wedge b) - \mu(f \wedge b)| > \epsilon\},$$

whence it follows that

$$(20) \quad \mathbf{P}\{|\mu_{n,p}f - \mu f| > 3\epsilon\} \leq \eta + \mathbf{P}\{|\mu_{n,p}(f \wedge b) - \mu(f \wedge b)| > \epsilon\}.$$

Since  $\mu_{n,p}(f \wedge b)$  converges to  $\mu(f \wedge b)$  in probability and  $\eta$  is arbitrary, the desired result follows.  $\square$

We now show that the function  $f(x) = x^\alpha$  satisfies the hypothesis of Theorem 2.

*Proof of Corollary 3.* We have

$$\begin{aligned} \mu_{n,p}f - \mu_{n,p}(f \wedge b) &= \frac{1}{\binom{n-1}{d}} \sum_{\sigma \in M(n,p)} n^\alpha p^\alpha w(\sigma)^\alpha - n^\alpha p^\alpha w(\sigma)^\alpha \wedge b \\ &= \frac{n^\alpha p^\alpha}{\binom{n-1}{d}} \sum_{\sigma \in M(n,p)} w(\sigma)^\alpha - w(\sigma)^\alpha \wedge \frac{b}{n^\alpha p^\alpha} \\ &= \frac{n^\alpha p^\alpha}{\binom{n-1}{d}} \int_{b/n^\alpha p^\alpha}^1 \beta_{d-1}(t^{1/\alpha}) dt. \end{aligned}$$

Substituting  $s = t^{1/\alpha}np$ , we get

$$\begin{aligned} \mu_{n,p}f - \mu_{n,p}(f \wedge b) &= \frac{\alpha}{\binom{n-1}{d}} \int_{b^{1/\alpha}}^{np} s^{\alpha-1} \beta_{d-1}(s/np) ds \\ &= \frac{\alpha n^d}{\binom{n-1}{d}} \int_{b^{1/\alpha}}^{np} s^{\alpha-1} \frac{\beta_{d-1}(s/np)}{n^d} ds. \end{aligned}$$

By Jensen's inequality we have

$$(21) \quad \|\mu_{n,p}f - \mu_{n,p}(f \wedge b)\|_q \leq \frac{\alpha n^d}{\binom{n-1}{d}} \int_{b^{1/\alpha}}^{np} s^{\alpha-1} \left\| \frac{\beta_{d-1}(s/np)}{n^d} \right\|_q ds.$$

Note that since  $0 \leq \beta_{d-1}(s/n) \leq \binom{n}{d}$  we have

$$0 \leq \frac{\beta_{d-1}(s/n)}{n^d} \leq \frac{1}{d!}.$$

Moreover, (4.14) from [5] shows that there is a  $\ell > 1 \vee q\alpha$  such that

$$\mathbf{E} \left[ \frac{\beta_{d-1}(s/np)}{n^d} \right] \leq 1 \wedge \frac{C p^\ell}{s^\ell}.$$

Combining these two inequalities we get

$$(22) \quad \left\| \frac{\beta_{d-1}(s/n)}{n^d} \right\|_q \leq \left( \frac{1}{d!} \right)^{(q-1)/q} \left( 1 \wedge \frac{C^{1/q} p^{\ell/q}}{s^{\ell/q}} \right)$$

for all  $s \in [0, n]$ . Thus, applying this bound in (21) for all sufficiently large  $b$  and  $n$ , we have

$$(23) \quad \|\mu_{n,p}f - \mu_{n,p}(f \wedge b)\|_q \leq \frac{C}{\ell/q - \alpha} b^{-(\ell/q - \alpha)/\alpha}.$$

This tends to zero as  $b \rightarrow \infty$  uniformly in  $n$ . Thus, the condition in (11) is satisfied.  $\square$

Finally, to prove Proposition 8. Let

$$(24) \quad g(c) = \begin{cases} 0, & c < c_*, \\ ct(c)(1-t(c))^d + \frac{c}{d+1}(1-t(c))^{d+1} - (1-t(c)), & c \geq c_*, \end{cases}$$

and

$$(25) \quad h(c) = 1 - \frac{c}{d+1} + g(c).$$

We next derive three technical results.

**Lemma 10.** *As  $c \rightarrow \infty$ ,  $\lim t(c) = 0$  and  $\lim ct(c) = 0$ .*

*Proof.* We first prove that  $\lim_{c \rightarrow \infty} t(c) = 0$ . For this, it suffices to show that along any sequence of  $c$  values that tends to infinity, there is always a subsequence on which the corresponding  $t(c)$  values have limit zero.

Let  $\{c_k\}$  be an arbitrary sequence of the above kind, i.e.,  $\lim_{k \rightarrow \infty} c_k = \infty$ . Since  $t_* \in (0, 1)$ ,  $t(c) \in (0, t_*)$  for all  $c \geq c_*$ ; hence, one can always find a subsequence  $\{c_\ell\} \equiv \{c_{k_\ell}\}$  such that the sequence  $\{t(c_\ell)\}$  has a limit in  $[0, t_*]$ . However, a limit other than 0 contradicts (7) since  $\lim_{\ell \rightarrow \infty} c_\ell$  would equal infinity then while  $\lim_{\ell \rightarrow \infty} -\log t(c_\ell)/(1-t(c_\ell))^d$  would be a finite number. Therefore,  $\lim_{\ell \rightarrow \infty} t(c_\ell) = 0$ . This shows that  $\lim_{c \rightarrow \infty} t(c) = 0$ .

Because  $\lim_{c \rightarrow \infty} t(c) = 0$ , it follows that  $1 - t(c) \geq 1/2$  (say) for all sufficiently large  $c$  values. Hence, for all such large  $c$  values, it follows from (7) that

$$(26) \quad ct(c) = ce^{-c(1-t(c))^d} \leq ce^{-c/2^d}.$$

This shows that  $\lim_{c \rightarrow \infty} ct(c) = 0$ , as desired.  $\square$

**Lemma 11.** *For any  $c \geq 0$ , we have  $\mu[c, \infty) = h(c)$ .*

*Proof.* We first show that  $\mu[c, \infty)$  is finite. Clearly,

$$\mu[c, \infty) \leq \frac{2}{d+1} \max \left\{ \int_0^{c_*} 1 - s(x) \, dx, \int_{c_*}^{\infty} 1 - s(x) \, dx \right\}.$$

Since  $t(c) \in (0, 1)$  for all  $c \geq c_*$ , it follows from (7) that  $1 - s(x) \leq 2^{d+1}t(x) = 2^{d+1}e^{-x(1-t(x))^d}$ . Lemma 10 shows that  $\lim_{c \rightarrow \infty} t(c) = 0$  which implies that  $1 - s(x) \leq 2^{d+1}e^{-x/2^d}$  for all large values of  $x$ . From this, it is easy to see that  $\mu[c, \infty)$  is finite.

Next, notice that both  $h$  and  $\mu[\cdot, \infty)$  are continuous functions; in particular, they are continuous at  $c_*$ . Hence, to show the desired result, it suffices to establish that  $h$  and  $\mu[\cdot, \infty)$  have the same derivative at any  $x \in \mathbb{R} \setminus c_*$  and that  $\lim_{c \rightarrow \infty} h(c) = 0$ .

From the definition of  $g$  and [10, cLtim 5.3], we have

$$(27) \quad g'(c) = \begin{cases} 0, & c < c_*, \\ \frac{1}{d+1}(1-t(c))^{d+1}, & c > c_*. \end{cases}$$

Hence,

$$(28) \quad h'(c) = \begin{cases} -\frac{1}{d+1}, & c < c_*, \\ -\frac{1}{d+1}[1 - (1-t(c))^{d+1}], & c > c_*. \end{cases}$$

From this, it is easy to see that  $h'(c) = \frac{d}{dc} \mu[c, \infty)$  for all  $c \in \mathbb{R} \setminus c_*$ , as desired.

Clearly,  $\lim_{c \rightarrow \infty} \mu[c, \infty) = 0$ . We now show that  $\lim_{c \rightarrow \infty} h(c) = 0$ . From the definition of  $g$ , we have

$$(29) \quad h(c) = ct(c)(1-t(c))^d + \frac{c}{d+1}(1-t(c))^{d+1} + t(c) - \frac{c}{d+1}.$$

Now, invoking the fact that  $t(c) \in (0, 1)$  for  $c > c_*$ , it follows that  $h(c) = O(ct(c))$ . From Lemma 10, it then follows that  $\lim_{c \rightarrow \infty} h(c) = 0$ .

Together, the above arguments establish the result.  $\square$

**Lemma 12.** *For  $c \geq 0$ ,*

$$(30) \quad \lim_{n \rightarrow \infty} \left\| \frac{\beta_{d-1}(Y(n, c/n))}{n^d} - \frac{h(c)}{d!} \right\|_q = 0.$$

*Proof.* This follows from (4.12) and (4.13) in [5].  $\square$

*Proof of Proposition 8.* Observe that, for all sufficiently large  $n$

$$(31) \quad \binom{n-1}{d} \mu_{n,p}[c, \infty) = \sum_{\sigma \in M(n,p)} \delta_{ntw(\sigma)}[c, \infty) = \left| \left\{ \sigma \in M(n,p) : w(\sigma) \geq \frac{c}{tn} \right\} \right|$$

$$(32) \quad = \left| \left\{ i : D_i \geq \frac{c}{tn} \right\} \right|$$

$$(33) \quad = \beta_{d-1} \left( \mathcal{L}(n,p) \Big|_{c/(tn)} \right)$$

$$(34) \quad \stackrel{d}{=} \beta_{d-1} \left( Y \left( n, \frac{c}{n} \right) \right),$$

where  $\{D_i\}$  is the set of death times in the  $(d-1)$ -th homology in the filtration  $\{\mathcal{L}(n,p)|_s : s \in [0,1]\}$ . However,  $\beta_{d-1} \left( \mathcal{L}(n,p) \Big|_{c/(tn)} \right)$  has the same distribution as that of  $\beta_{d-1} \left( Y(n, c/n) \right)$ , where  $Y(n,p)$  denotes the  $d$ -Linial Meshulam complex on  $n$  vertices. This is because  $\mathcal{L}(n,p)|_s$  is distributed as  $Y(n, ps)$  since in either case each  $d$ -face, independent of everything else, is present with probability  $ps$ . Therefore, from Lemma 12 and the fact that  $\binom{n-1}{d}/n^d \rightarrow 1/d!$ , it follows that

$$(35) \quad \lim_{n \rightarrow \infty} \|\mu_{n,p}[c, \infty) - h(c)\|_q = 0.$$

The desired result now follows from Lemma 11.  $\square$

#### 4. BEHAVIOR IN THE EXTREMES

We derive Theorem 4 here by building upon the proofs for [12, Propositions 33, 35, and 36], which look at the  $p = 1$  case. The reader should, however, note the following distinction between the result in [12] and the one here. In [12], Poisson point process convergence is shown to hold directly for the extremal values amongst the actual  $d$ -MSA face weights in  $\mathcal{L}(n,1)$  and amongst the actual death times in the associated persistence diagram. In contrast, we show that the same result holds for  $\mathcal{L}(n,p)$  when the actual values are additionally scaled by  $p$ , as in Theorem 2.

We begin our analysis by summarizing the key steps in the proofs of [12, Proposition 33, 35, and 36]. There, via the factorial moment method, it is first shown that the extremal values in the set of nearest face distances<sup>1</sup> in  $\mathcal{L}(n,1)$  converges to a Poisson point process. This result is then extended to the death times in the  $(d-1)$ -th persistence diagram of  $\{\mathcal{L}(n,1)|_s : s \in \mathbb{R}\}$ . This is done by exploiting the relation between the nearest face distances and the  $(d-1)$ -th Betti number, cf. the proof of [6, Theorem 1.10]. Finally, the result for the  $d$ -MSA face weights is obtained by using the equivalence between the set of these values and the corresponding death times, as given in [12, Theorem 3]. Notice that this train of thought proceeds in a direction opposite to how the proof for Theorem 2 goes. More specifically, the above result is first shown for the death times and then extended to the  $d$ -MSA face weights instead of the other way around.

As per the above ideas, to derive Theorem 4, we need to sequentially show Poisson point process convergence for three counting measures: the first based on the nearest face distances, the second based on the death times, and the third based on the  $d$ -MSA face weights. The last two measures are precisely  $\tilde{\nu}_{n,p}$  and  $\nu_{n,p}$ , respectively, while the first one is denoted  $\nu'_{n,p}$  and we define it to be

$$(36) \quad \nu'_{n,p} := \sum_{\tau \in \mathcal{F}^{d-1}(\mathcal{L}(n,p))} \delta_{n p C(\tau) - d \log(n) + \log(d!)},$$

where  $C(\tau) := \min\{w(\sigma) : \sigma \in \mathcal{F}^d(\mathcal{L}(n,p)) \text{ and } \sigma \supset \tau\}$ .

*Proof of Theorem 4.* As stated above, our arguments mirror those used in the proof of Propositions 33, 35, and 36 in [12]. Hence, we only highlight the major steps.

*Convergence of  $\nu'_{n,p}$ :* As in the proof of [12, Proposition 33], it suffices to show that the  $\ell$ -th factorial moment of  $\nu'_{n,p}(I)$  satisfies

$$(37) \quad \lim_{n \rightarrow \infty} \mathbf{E}[\nu'_{n,p}{}^{(\ell)}(I)] = \left( \int_I e^{-x} dx \right)^\ell$$

<sup>1</sup>For a  $(d-1)$ -face, the nearest face distance is simply the minimum of the  $d$ -faces incident on it; in the  $d = 1$  case, this can also be seen as the nearest neighbour distance.

for any  $I \subseteq \mathbb{R}$  that is a finite union of disjoint intervals. Furthermore, as shown in [12, p.33], one way to establish the above relation is to let  $\bar{C}(\tau) := npC(\tau) - d \log n + \log d!$  and show that

$$(38) \quad \lim_{n \rightarrow \infty} \ell! \binom{n}{d} \mathbf{E} \left[ \prod_{i=1}^{\ell} \mathbf{1}[\bar{C}(\tau_i) \in (c_i, \infty)] \right] = e^{-\sum_{i=1}^{\ell} c_i},$$

where  $\tau_1, \dots, \tau_{\ell}$  are distinct  $(d-1)$ -faces and  $c_1, \dots, c_{\ell}$  are distinct real numbers. We derive this last equality now.

For any  $\tau \in \mathcal{F}^{d-1}$ , observe that

$$\mathbf{1}[\bar{C}(\tau) \in (c, \infty)] = \mathbf{1}[C(\tau) > c(n)/p],$$

where  $c(n) = (c + d \log n - \log d!)/n$ . Since  $\mathcal{L}(n, p)|_s$  has the same distribution as  $Y(n, ps)$  for  $s \in [0, 1)$ , the event  $[\bar{C}(\tau) \in (c, \infty)]$  is equivalent to the face  $\tau$  being isolated in  $Y(n, c(n))$  whenever  $c(n)/p \in [0, 1)$ . Consequently,

$$\ell! \binom{n}{d} \mathbf{E} \left[ \prod_{i=1}^{\ell} \mathbf{1}[\bar{C}(\tau_i) \in (c_i, \infty)] \right] \sim \frac{n^{d\ell}}{(d!)^{\ell}} \prod_{i=1}^{\ell} \left( 1 - \frac{c_i + d \log(n) - \log(d!)}{n} \right)^{n - \kappa_i}$$

for some arbitrary but fixed (w.r.t.  $n$ ) constants  $\kappa_1, \dots, \kappa_{\ell}$  that depend only on the number of vertices common to the faces  $\tau_1, \dots, \tau_{\ell}$ . This is exactly the same relation that was proven in [12, p.33]. Hence (38) and, in turn, (37) holds.

This shows that  $\nu'_{n,p}$  converges vaguely to  $\mathcal{P}$  in distribution, as desired.

*Convergence of  $\tilde{\nu}_{n,p}$ :* Based on the arguments in the proof of [12, Proposition 35], we only need to establish that, for  $c \in \mathbb{R}$ ,

$$(39) \quad \lim_{n \rightarrow \infty} \mathbf{E} |\tilde{\nu}_{n,p}(c, \infty) - \nu'_{n,p}(c, \infty)| = 0.$$

Let  $c(n)$  be as defined above and  $N_{d-1}(Y(n, p))$  be the number of isolated  $(d-1)$ -faces in  $Y(n, p)$ . Then, note that  $\tilde{\nu}_{n,p}(c, \infty) = \beta_{d-1}(\mathcal{L}(n, p)|_{c(n)/p})$  and  $\nu'_{n,p}(c, \infty) = N_{d-1}(\mathcal{L}(n, p)|_{c(n)/p})$  whenever  $c(n)/p \in [0, 1)$ . Again, since  $\mathcal{L}(n, p)|_s$  has the same distribution as  $Y(n, ps)$  for  $s \in [0, 1)$ , the relation in (39) now follows trivially from [12, Lemma 34].

Thus,  $\tilde{\nu}_{n,p}$  converges vaguely to  $\mathcal{P}$  in distribution.

*Convergence of  $\nu_{n,p}$ :* This is immediate from [12, Theorem 3] and the convergence of  $\tilde{\nu}_{n,p}$ .  $\square$

## 5. EXTENSIONS

Here we discuss three different extensions of our results. The first one relates to the case where the  $d$ -face weights come from some generic distribution. The next one concerns the robustness of our results to noisy perturbations in the  $d$ -face weights. The third and final one is about the randomly weighted  $Y(n, p)$  complex wherein the potential  $d$ -faces may not be all present.

**5.1. Generic distribution for  $d$ -face weights.** With a slight abuse of notation, let  $\mathcal{L}(n, p)$  itself denote the generalization of the complex given in Definition 1 wherein i.)  $\{U(\sigma) : \sigma \in \mathcal{F}^d\}$  are real-valued i.i.d. random variables with some generic distribution  $\mathcal{F}$  and ii.) the function  $w$  is such that  $w(\sigma)$  equals  $-\infty$  when  $|\sigma| \leq d$ , equals  $U(\sigma)$  when  $\sigma \in \mathcal{F}^d(Y)$ , and equals  $\infty$  otherwise. We claim that a version of our results also holds in this setup. However, for this, we need to slightly modify the definition of the measures given in (4) and (5). In particular, we need to replace  $w(\sigma)$  and  $\Delta$  in these definitions with  $\mathcal{F}(w(\sigma))$  and  $\mathcal{F}(\Delta)$ , respectively, i.e., let

$$(40) \quad \mu_{n,p} = \frac{1}{|M(n, p)|} \sum_{\sigma \in M(n, p)} \delta_{np \mathcal{F}(w(\sigma))}$$

and so on.

**Corollary 13.** *Suppose  $\mathcal{F}$  is continuous. Then, Theorems 2 and 4 hold for the modified measures.*

*Proof.* Let  $\mathcal{L}_{\mathcal{F}}(n, p) = (\mathcal{K}_n, w_{\mathcal{F}})$  be a weighted complex that is coupled to  $\mathcal{L}(n, p)$  in the following way. We start with  $\{\mathcal{F}(U(\sigma)) : \sigma \in \mathcal{F}^d(\mathcal{K}_n)\}$ , where the set  $\{U(\sigma)\}$  is the same one that was used in the definition of  $\mathcal{L}(n, p)$ . Since  $\mathcal{F}$  is continuous, the transformed collection is a set of i.i.d.  $U[0, 1]$  random variables. Next, we let  $w_{\mathcal{F}}$  be the weight function that satisfies  $w_{\mathcal{F}}(\sigma) = \mathcal{F}(w(\sigma))$

for all  $\sigma \in \mathcal{K}_n$ . Clearly,  $\mathcal{L}_{\mathcal{F}}(n, p)$  has the same distribution as the augmented  $d$ -complex given in Definition 1. Therefore, Theorems 2 and 4 readily apply to it. Furthermore, the analogues of the measures given in (4) and (5) for  $\mathcal{L}_{\mathcal{F}}(n, p)$  resemble the modified measures described above. The only thing that remains to be checked is the relationship of the  $d$ -MSAs in  $\mathcal{L}(n, p)$  and  $\mathcal{L}_{\mathcal{F}}(n, p)$ . However, since any distribution function is non-decreasing, a  $d$ -MSA in one is also a  $d$ -MSA in the other and vice-versa. The claim is now easy to see.  $\square$

**5.2. Noisy perturbations in  $d$ -face weights.** Here we establish the robustness of our results to additional noisy perturbations in the  $d$ -face weights. Consider the following further generalization of the  $\mathcal{L}(n, p)$  complex described in Section 5.1 above.

**Definition 7.** Let  $\mathcal{K}_n$  be the complete  $d$ -skeleton on  $n$  vertices and let  $\{U(\sigma) : \sigma \in \mathcal{F}^d\}$  be i.i.d. real-valued random variables with a common distribution  $\mathcal{F}$ . Further, let  $\{\epsilon_n(\sigma) : \sigma \in \mathcal{F}^d\}$  be a separate set of random variables denoting noise in the  $d$ -face weights. These latter variables need not be independent nor identically distributed. Then, for  $p \in (0, 1]$  and  $Y$  being a sample of  $Y(n, p)$ , the weighted complex  $\mathcal{L}'(n, p) = (\mathcal{K}_n, w')$ , where  $w' : \mathcal{K}_n \rightarrow [0, 1]$  satisfies

$$w'(\sigma) = \begin{cases} -\infty & \text{if } |\sigma| \leq d, \\ U(\sigma) + \epsilon_n(\sigma) & \text{if } \sigma \in \mathcal{F}^d(Y), \\ \infty & \text{otherwise.} \end{cases}$$

Our aim here is to show that the main results also hold for  $\mathcal{L}'(n, p)$  if  $\|\epsilon_n\|_\infty := \max |\epsilon_n(\sigma)|$  decays to zero sufficiently fast. Towards that, along the lines described in Section 5.1, let  $\mathcal{L}(n, p)$  and  $w$  respectively be the unperturbed  $d$ -complex and the associated weight function that are defined on the same probability space as  $\mathcal{L}'(n, p)$ . Also, let  $M(n, p)$  be a  $d$ -MSA in  $\mathcal{L}(n, p)$  and let  $D(n, p) \equiv \{\Delta_i\}$  be the set of death times in the  $(d-1)$ -th persistence diagram associated with  $\{\mathcal{L}(n, p)|_s : s \in \mathbb{R}\}$ . Similarly, define  $M'(n, p)$  and  $D'(n, p)$  with respect to  $\mathcal{L}'(n, p)$ . Finally, let  $\mu_{n,p}, \tilde{\mu}_{n,p}, \nu_{n,p}$ , and  $\tilde{\nu}_{n,p}$  be the measures as defined in Section 5.1, and let  $\mu'_{n,p}, \tilde{\mu}'_{n,p}, \nu'_{n,p}$ , and  $\tilde{\nu}'_{n,p}$  denote their analogues obtained by replacing  $\mathcal{F}(w(\sigma)), \mathcal{F}(\Delta), M(n, p)$ , and  $D(n, p)$  with  $\mathcal{F}(w'(\sigma)), \mathcal{F}(\Delta'), M'(n, p)$ , and  $D'(n, p)$ , respectively.

**Corollary 14.** *Let  $p \in (0, 1]$ . Suppose  $\mathcal{F}$  is Lipschitz continuous with Lipschitz constant  $\zeta > 0$ . Then the following statements hold.*

- (1) *Bulk: Let the measure  $\mu$  be as in (8). If  $n\|\epsilon_n\|_\infty \rightarrow 0$  in  $L^q$ , then the random measures  $\mu'_{n,p}$  and  $\tilde{\mu}'_{n,p}$  converge in the Kolmogorov metric to  $\mu$  in  $L^q$ . Moreover, if the function  $f$  satisfies the conditions in Theorem 2, then both  $\mu'_{n,p}f$  and  $\tilde{\mu}'_{n,p}f$  converge to  $\mu f$  in  $L^q$ .*
- (2) *Extremes: If  $n\|\epsilon_n\|_\infty \rightarrow 0$  in probability, then both  $\nu'_{n,p}$  and  $\tilde{\nu}'_{n,p}$  converge vaguely in distribution to  $\mathcal{P}$ , where  $\mathcal{P}$  is the Poisson point process on  $\mathbb{R}$  with intensity  $e^{-x} dx$ .*

*Proof.* The  $p = 1$  case in Statement (2) was proved in [12, Theorem 7]. The  $p \in (0, 1)$  case can be derived similarly from the unperturbed version given in Corollary 13 here.

Consider Statement (1). We only discuss the  $\mu'_{n,p}$  result since the  $\tilde{\mu}'_{n,p}$  case can be dealt with similarly. For the  $\mu'_{n,p}$  case, it suffices to show that an analogue of Proposition 8 holds. This is because the arguments from the proof of Theorem 2 can then be again used to get the actual result.

Let  $\delta_n := \zeta n p \|\epsilon_n\|_\infty$ . From<sup>2</sup> [12, Lemma 38], we have

$$\inf_{\gamma} \max_i |w(\sigma) - \gamma(w(\sigma))| \leq \|\epsilon_n\|_\infty,$$

where the infimum is over all the bijections  $\gamma : \{w(\sigma) : \sigma \in M(n, p)\} \rightarrow \{w'(\sigma') : \sigma' \in M'(n, p)\}$ . Hence, for any measurable set  $K \subseteq \mathbb{R}$ , it follows that

$$(41) \quad \mu_{n,p}(K^{-\delta_n}) \subseteq \mu'_{n,p}(K) \subseteq \mu_{n,p}(K^{\delta_n}),$$

where, for  $\delta > 0$ ,

$$K^\delta := \{x \in \mathbb{R} : d(x, K) \leq \delta\} \quad \text{and} \quad K^{-\delta} := \{x \in \mathbb{R} : (x - \delta, x + \delta) \subset K\}.$$

<sup>2</sup>There is a typo in the statement of [12, Lemma 38]. In the second and third displays,  $\gamma(D_i)$  should be  $\gamma(D'_i)$  and  $\gamma(\phi(\sigma_i))$  must be  $\gamma(\phi'(\sigma'_i))$ . This follows from Theorem 4 in *ibid*.

Now, fix an arbitrary  $\delta > 0$ . Then,

$$\begin{aligned}
(42) \quad & |\mu'_{n,p}[c, \infty) - \mu_{n,p}[c, \infty)| \\
& \leq \max \{ |\mu_{n,p}[c - \delta_n, \infty) - \mu_{n,p}[c, \infty)|, |\mu_{n,p}[c, \infty) - \mu_{n,p}[c + \delta_n, \infty)| \} \\
& = \mu_{n,p}[c - \delta_n, c + \delta_n) \\
(43) \quad & \leq \mu_{n,p}[c - \delta, c + \delta) + \mathbf{1}[\delta_n > \delta] \\
& \leq |\mu_{n,p}[c - \delta, c + \delta) - \mu[c - \delta, c + \delta)| + \mu[c - \delta, c + \delta) + \mathbf{1}[\delta_n > \delta].
\end{aligned}$$

where (42) follows from (41), while the second term in (43) is obtained by using the fact that  $\mu_{n,p}[c - \delta_n, c + \delta_n) \leq 1$  which itself holds since  $\mu_{n,p}$  is a probability measure.

A simple application of Proposition 8 along with the given condition on  $\|\epsilon_n\|_\infty$  and the fact that  $\delta$  is arbitrary, then shows that  $|\mu'_{n,p}[c, \infty) - \mu_{n,p}[c, \infty)| \rightarrow 0$  in  $L^q$ , as desired.  $\square$

**5.3. Randomly weighted  $Y(n, p)$  complex.** Here, we have a sample  $Y$  of  $Y(n, p)$  whose  $d$ -faces are assigned i.i.d. weights with a common distribution  $\mathcal{F}$ , while all the lower dimensional faces are assigned weight  $-\infty$ . This clearly differs from the  $\mathcal{L}(n, p)$  complex in Section 5.1 since not all the potential  $d$ -faces may be present here. Despite this, we now show that a version of our results holds for this complex as well. We only talk about the  $\mu_{n,p}$  and  $\nu_{n,p}$  cases, since a similar argument also works for the measures related to the death times.

The first thing one needs to make sure is that there are enough  $d$ -faces to even guarantee the existence of a  $d$ -MSA. This is not that hard. In fact, since  $p$  is a fixed constant, we have from [8, Theorem 1.1] and [11, Theorem 1.1] that  $\lim_{n \rightarrow \infty} \mathbf{P}\{\beta_{d-1}(Y(n, p)) = 0\} = 1$ . It then follows from [12, Lemma 23], for all sufficiently large  $n$ , there indeed exists a  $d$ -MSA with high probability.

Next, let  $\mathcal{L}_Y(n, p)$  be the complex obtained from the given  $Y(n, p)$  complex by adding the missing  $d$ -faces and giving them weight  $+\infty$ . Clearly,  $\mathcal{L}_Y(n, p)$  has the same distribution as the  $\mathcal{L}(n, p)$  complex described in Section 5.1. Further, since the added faces have weight  $+\infty$ , the  $d$ -MSA in  $Y(n, p)$  and  $\mathcal{L}_Y(n, p)$  will be one and the same almost surely on the event  $\mathcal{B}_{n,p} := \{\beta_{d-1}(Y(n, p)) \neq 0\}$ ; the almost sure part is to deal with the case of non-unique weights of the  $d$ -faces in  $Y(n, p)$ . Now, let  $\hat{\mu}_{n,p} = \mathbf{1}_{\mathcal{B}_{n,p}^c} \mu_{n,p} + \mathbf{1}_{\mathcal{B}_{n,p}} \delta_0$ , where  $\delta_0$  is the unit measure at 0, and  $\mu_{n,p}$  is as in (40). Then, for any continuous function  $f$  satisfying (11), we have  $\hat{\mu}_{n,p} f \rightarrow \mu f$  in  $L^q$ , as desired. This can be seen via the following triangle inequality:

$$\begin{aligned}
\|\hat{\mu}_{n,p} f - \mu_{n,p} f\|_q &= \|(\hat{\mu}_{n,p} f - \mu_{n,p} f) \mathbf{1}_{\mathcal{B}_{n,p}}\|_q \\
&\leq \|(\hat{\mu}_{n,p} f) \mathbf{1}_{\mathcal{B}_{n,p}}\|_q + \|(\mu_{n,p} |f| - \mu_{n,p}(|f| \wedge b)) \mathbf{1}_{\mathcal{B}_{n,p}}\|_q \\
&\quad + \|\mu_{n,p}(|f| \wedge b) \mathbf{1}_{\mathcal{B}_{n,p}}\|_q \\
&\leq |f(0)| [\mathbf{P}(\mathcal{B}_{n,p})]^{1/q} + \sup_{n \geq 0} \|\mu_{n,p} |f| - \mu_{n,p}(|f| \wedge b)\|_q + b [\mathbf{P}(\mathcal{B}_{n,p})]^{1/q}.
\end{aligned}$$

The arguments for extending the  $\nu_{n,p}$  result are even more simple. As above, let us define  $\hat{\nu}_{n,p} = \mathbf{1}_{\mathcal{B}_{n,p}^c} \nu_{n,p} + \mathbf{1}_{\mathcal{B}_{n,p}} \delta_0$ . Observe that, for any bounded continuous function  $f : \mathbb{R} \rightarrow \mathbb{R}$  with bounded support, we have

$$\|\hat{\nu}_{n,p} f - \nu_{n,p} f\|_1 \leq \|(\hat{\nu}_{n,p} f - \nu_{n,p} f) \mathbf{1}_{\mathcal{B}_{n,p}}\|_1 \leq |f(0)| \mathbf{P}(\mathcal{B}_{n,p}) + \sqrt{\mathbf{E}[(\nu_{n,p} f)^2]} \sqrt{\mathbf{P}(\mathcal{B}_{n,p})},$$

where the last relation follows from the Cauchy-Schwarz inequality. Therefore,  $\|\hat{\nu}_{n,p} f - \nu_{n,p} f\|_1 \rightarrow 0$  as  $n \rightarrow \infty$ . An application of Slutsky's theorem now shows that  $\hat{\nu}_{n,p} \rightarrow \mathcal{P}$ , as desired.

## 6. SUMMARY AND FUTURE DIRECTIONS

We prove the convergence of the histogram of the weights in random MSA—both in the bulk and in the extremes. We prove the empirical distribution of the MSA weights asymptotically converges to a measure based on the shadow—the complement of graph components in higher dimensions. Our result also is the first to show connections between the MSA weights and the shadow. Our second result states that the extremal weights converge to an inhomogeneous Poisson point process. A interesting consequence of our two results is that we can also state the distribution of the death times in the persistence diagram corresponding to the above weighted complex.

We are interested in an online or streaming version of the results in this paper. Some preliminary simulations suggest that the weight of a MST or MSA generated by an online algorithm, properly scaled, will share similar distributional properties to the weights of the MST and MSA we have

outlined in this paper. In the streaming or online algorithm we observe a sequence of the faces and update our MST or MSA estimate sequentially. Note that unlike the analysis in our paper we cannot assume that the faces will appear in order of the magnitude of the face weights, in general we weights of the faces in the sequence are random. Our goal with the revealing of each face is to update the current estimate of the MST or MSA by either including the face and updating MST/MSA estimate or discarding the face. One can decide to accept or discard the face by testing if a cycle is created and if the face is accepted one can use variants of Kruskal's algorithm to update the MST/MSA. In Fig. 4(A) we plot in blue the weight of the MST after the  $k$ -th face in the sequence is observed,  $w(M_n(k))$ , scaled by  $c_1 = n/\binom{n-1}{d}$ , with  $n = 100$  and  $d = 1$ . The red dashed line in Fig. 4(A) plots our estimate of  $c_1 w(M_n(k))$  as  $c_2/k$  where  $c_2 = \binom{n}{d+1} \mu x$  and the constant  $\mu x$  is  $\zeta(3) \approx 1.20$ . The blue curve in Fig. 4(B) is the plot of the weights of the MSA,  $w(M_n(k))$ , scaled by  $c_1 = n/\binom{n-1}{d}$  as a function of  $k$  with  $n = 30$  and  $d = 2$ . The red dashed line in Fig. 4(B) plots our estimate of  $c_1 w(M_n(k))$  as  $c_2/k$  where  $c_2 = \binom{n}{d+1} \mu x$  where the constant  $\mu x$  is approximately 1.56.

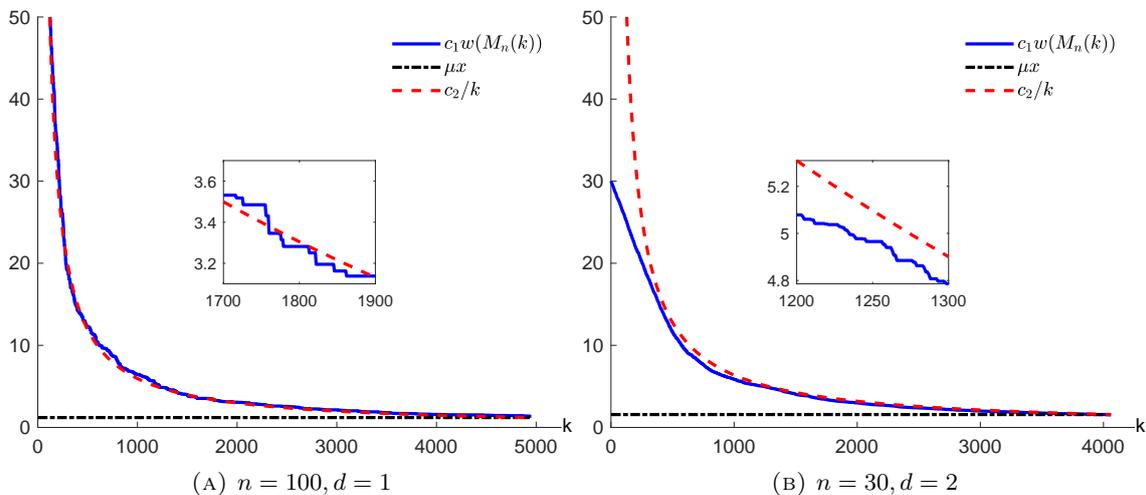


FIGURE 4. Denote the  $d$ -MSA generated by the online algorithm after revealing  $k$  faces as  $M_n(k)$ . The blue curves plot the scaled weights of the  $d$ -MSA,  $c_1 w(M_n(k))$  with  $c_1 = n/\binom{n-1}{d}$ . The red dashed curve is a plot of our estimate of the scaled weights for which we have the analytic expression  $c_2/k$  with  $c_2 = \binom{n}{d+1} \mu x$  where the constant  $\mu x$  is  $\zeta(3) \approx 1.2$  for  $d = 1$  and 1.56 for  $d = 2$ . Note that the limiting value of  $c_2/k$  as  $k \rightarrow \binom{n}{d+1}$  and  $n \rightarrow \infty$  is  $\mu x$ . (A) The plots for  $n = 100, d = 1$ . (B) The plots for  $n = 30, d = 2$ .

There are several future directions of interest:

- (1) Geometric random graphs and complexes: Extending our results from the Erdős-Rényi and Linial-Meshulam models to geometric random graphs and simplicial complexes and understanding differences between the two types of random models.
- (2) Online algorithms: The current model involves unveiling one face at a time. This setup is not an online algorithm, since the initial sorting of the weights requires the observation of all the faces. Providing both an online algorithm as well as the analysis of the algorithm in terms of the weight distribution and relations to the shadow and persistent homology is of interest.
- (3) Large deviation results and central limit theorems: It would be interesting to have general large deviation results for the KS distance  $\mu_{n,p}$  and  $\mu$  as well as central limit theorems characterizing the variance of the weight distribution.
- (4) Rates of convergence: We do not provide rates of convergence to either the bulk or extremal measure. It would be interesting to provide rates of convergence and understand conditions required to provide rates.

## ACKNOWLEDGEMENTS

The authors would like to thank Matthew Kahle, Omer Bobrowski, Primoz Skraba, Ron Rosenthal, Robert Adler, and Christina Goldschmidt for conversations. Sayan Mukherjee would like to acknowledge funding from NSF DEB-1840223, NIH R01 DK116187-01, HFSP RGP0051/2017, NSF DMS 17-13012, and NSF CCF-1934964.

## REFERENCES

- [1] Robert Adler. TOPOS: Pinsky was wrong, Euler was right. <https://imstat.org/2014/11/18/topos-pinsky-was-wrong-euler-was-right/>, 2014. [Online; accessed 14-Nov.-2020].
- [2] Paul Erdős and Alfred Rényi. On random graphs I. *Publicationes Mathematicae Debrecen*, 6(290-297):18, 1959.
- [3] Paul Erdős and Alfred Rényi. On the evolution of random graphs. *Publ. Math. Inst. Hungary. Acad. Sci.*, 5:17–61, 1960.
- [4] Alan M Frieze. On the value of a random minimum spanning tree problem. *Discrete Applied Mathematics*, 10(1):47–56, 1985.
- [5] Masanori Hino and Shu Kanazawa. Asymptotic behavior of lifetime sums for random simplicial complex processes. *Journal of the Mathematical Society of Japan*, 2019.
- [6] Matthew Kahle and Boris Pittel. Inside the critical window for cohomology of random  $k$ -complexes. *Random Structures & Algorithms*, 48(1):102–124, 2016.
- [7] Gil Kalai. Enumeration of  $Q$ -acyclic simplicial complexes. *Israel Journal of Mathematics*, 45(4):337–351, 1983.
- [8] Nathan Linial and Roy Meshulam. Homological connectivity of random 2-complexes. *Combinatorica*, 26(4):475–487, 2006.
- [9] Nathan Linial, Ilan Newman, Yuval Peled, and Yuri Rabinovich. Extremal problems on shadows and hypercuts in simplicial complexes. *arXiv preprint arXiv:1408.0602*, 2014.
- [10] Nathan Linial and Yuval Peled. On the phase transition in random simplicial complexes. *Annals of Mathematics*, pages 745–773, 2016.
- [11] Roy Meshulam and Nathan Wallach. Homological connectivity of random  $k$ -dimensional complexes. *Random Structures & Algorithms*, 34(3):408–417, 2009.
- [12] Primoz Skraba, Gagan Thoppe, and D Yogeshwaran. Randomly Weighted  $d$ - complexes: Minimal Spanning Acycles and Persistence Diagrams. *Electronic Journal of Combinatorics*, 27(2), 2019.

NF: DEPT. OF STATISTICS AND OPERATIONS RESEARCH, UNIV. OF NORTH CAROLINA, CHAPEL HILL, USA  
*Email address:* [fraiman@email.unc.edu](mailto:fraiman@email.unc.edu)

SM: DEPARTMENTS OF STATISTICAL SCIENCE, MATHEMATICS, COMPUTER SCIENCE, BIostatISTICS & BIOINFORMATICS, DUKE UNIVERSITY, DURHAM, NC, USA  
*Email address:* [sayan@stat.duke.edu](mailto:sayan@stat.duke.edu)

GT: DEPT. OF COMPUTER SCIENCE AND AUTOMATION, INDIAN INSTITUTE OF SCIENCE, BENGALURU, INDIA  
*Email address:* [gthoppe@iisc.ac.in](mailto:gthoppe@iisc.ac.in)