Asymptotics of integrals of Betti numbers for random simplicial complex processes

Masanori Hino

Abstract We discuss a higher-dimensional analogue of Frieze's $\zeta(3)$-limit theorem for the Erdős–Rényi graph process applied to a family of increasing random simplicial complexes. In particular, we consider the time integrals of Betti numbers, which are interpreted as lifetime sums in the context of persistent homologies. We survey some recent results regarding their asymptotic behavior that answer some questions posed in an earlier study by Hiraoka and Shirai.

Key words: random simplicial complex, Betti number, persistent homology, lifetime sum

Mathematics Subject Classifications (2010). Primary: 60D05; Secondary: 05C80, 55U10, 05E45, 60C05

1 Introduction

Extensive studies on limit behavior of random graphs have their origins in the work of Erdős and Rényi [4, 5]. Graph characteristics such as the threshold probability of connectivity and the limit behavior around the critical probability provide good descriptions of such complicated random discrete objects. In recent studies, the scaling limits of random graphs themselves have attracted attention in pursuit of a more comprehensive understanding; typical limit objects are continuum random trees, which have fractal structures (e.g., see [1, 21] and the references therein). The importance of fractal analysis in the study of random graphs will be emphasized more in future work.

Meanwhile, the homological structures of random simplicial complexes have also attracted interest recently as higher-dimensional counterparts of random graphs; see

Masanori Hino
Department of Mathematics, Kyoto University, Kyoto 606–8502, Japan, e-mail: hino@math.kyoto-u.ac.jp

Kahle [16] for a survey of recent studies. In this connection, Hiraoka and Shirai [11] studied the asymptotic behavior of persistent homologies of random simplicial complex processes, and Hino and Kanazawa [10] advanced their research by solving some of the problems that they had posed. A natural question to consider next is to characterize suitable scaling limits of random simplicial complexes, which are certain to have fractal structures. However, unlike the case with random graphs, there are as yet no concrete results about this question because the theory and techniques are yet to be developed fully.

In this article, we follow [11, 10] and survey some recent results and new ideas in the study of the homologies of random simplicial complexes. We hope that this survey will serve as a preliminary to studying such objects from the perspective of fractal analysis.

The rest of this article is organized as follows. In Section 2, we introduce various concepts regarding random graphs, random graph processes, and their higher-dimensional analogues, and we state some results regarding their asymptotic behavior. In Section 3, we provide basic ideas for proving the main theorems. In Section 4, we enumerate several problems for future research.

2 Frameworks and theorems

A typical random graph model is the Erdős–Rényi model $G(n,p)$ [9, 4, 5], which is defined as the distribution of a random graph consisting of n vertices with the edges between each pair of vertices included with probability p independently.[1] In one of the earliest results of random graph theory, Erdős and Rényi proved the following.

Theorem 2.1 ([5]). *Let $\varepsilon > 0$ and $p = p(n)$ depend on n.*

- *If $p < (1-\varepsilon)(\log n)/n$ for sufficiently large n, then*

$$\mathbb{P}(\textit{the graph is disconnected}) \to 1 \quad \textit{as } n \to \infty.$$

- *If $p > (1+\varepsilon)(\log n)/n$ for sufficiently large n, then*

$$\mathbb{P}(\textit{the graph is connected}) \to 1 \quad \textit{as } n \to \infty.$$

This theorem shows that the connectivity changes drastically around $p = (\log n)/n$. Since then, there have been many studies of the behavior around the threshold probability, which is one of the central topics of random graph theory.

Meanwhile, there have been other types of studies on the limit behavior of the Erdős–Rényi model. To explain one such type, we introduce a canonical realization of the family of Erdős–Rényi models $\{G(n,p)\}_{p\in[0,1]}$ for fixed n. Let $K_n = V_n \sqcup E_n$ be the complete graph with n vertices, where V_n and E_n are the vertex set and the

[1] This definition is due to Gilbert [9]. The model that Erdős and Rényi introduced in [4, 5] is slightly different, but the two models behave similarly as the number of vertices tends to infinity.

edge set, respectively. We assign independent and identically distributed random variables $\{u_e\}_{e\in E_n}$ that are uniformly distributed on $[0,1]$. We construct a family of random graphs $\mathcal{X}_n = \{X_n(t)\}_{t\in[0,1]}$ so that each u_e is the birth time of the edge $e \in E_n$. More precisely, for each $t \in [0,1]$, the random graph $X_n(t)$ is defined as

$$X_n(t) = V_n \sqcup \{e \in E_n \mid u_e \le t\}.$$

By construction, $X_n(t)$ is nondecreasing with respect to t almost surely, and the law of $X_n(t)$ is equal to $G(n,t)$ for every $t \in [0,1]$.

Let $L_0(\mathcal{X}_n)$ be the minimal weight of spanning trees[2] of K_n, that is,

$$L_0(\mathcal{X}_n) = \inf\left\{\sum_{e\in T} u_e \;\middle|\; T\text{: a spanning tree of } K_n\right\}.$$

This quantity has several interpretations: By Kruskal's algorithm [18], the identities

$$L_0(\mathcal{X}_n) = \int_0^1 \beta_0(X_n(t))\,dt = \sum_{i=1}^{n-1} t_i \tag{2.1}$$

hold, where $\beta_0(G)$ denotes the number of connected components of the graph G minus one, and t_i denotes the ith random time when the number of connected components of $X_n(t)$ decreases. Frieze [7] proved the asymptotic behavior of $L_0(\mathcal{X}_n)$ as follows.

Theorem 2.2 ([7]). *It holds that*

$$\lim_{n\to\infty} \mathbb{E}[L_0(\mathcal{X}_n)] = \zeta(3)\left(= \sum_{k=1}^{\infty} k^{-3}\right).$$

Moreover, for any $\varepsilon > 0$,

$$\lim_{n\to\infty} \mathbb{P}(|L_0(\mathcal{X}_n) - \zeta(3)| > \varepsilon) = 0.$$

Recently, Hiraoka and Shirai [11] studied a higher-dimensional analogue of (2.1) and Theorem 2.2, with random graphs and the number of connected components replaced by *random simplicial complexes* and the *(reduced) Betti number*, respectively. Let us briefly review the concepts of simplicial complexes and their homologies.

Let V be a nonempty finite set. A collection X of nonempty subsets of V is called an (abstract) *simplicial complex* over V if the following conditions are satisfied.

- For every $v \in V$, $\{v\}$ belongs to X.
- For any $\sigma \in X$, every nonempty subset of σ belongs to X.

For $\sigma \in X$, $k := \#\sigma - 1$ is called the dimension of σ and is denoted by $\dim\sigma$. We call σ a k-dimensional simplex or, equivalently, a k-simplex, and we call the maximum

[2] A spanning tree of a graph G is, by definition, a tree that includes all the vertices of G.

of $\dim \sigma$ the dimension of X. Any finite graph can be regarded as either a zero- or one-dimensional simplicial complex. If two simplices σ and τ satisfy $\sigma \subset \tau$, then σ is called a face of τ.

For $k \geq 0$, $\sigma = (v_0, v_1, \ldots, v_k) \in V^{k+1}$ is called an ordered k-simplex of X if $\{v_0, v_1, \ldots, v_k\}$ is a k-simplex of X. Two ordered simplices are called equivalent if one is an even permutation of the other. The equivalence class of an ordered k-simplex σ is denoted by $\langle \sigma \rangle$ or $\langle v_0, v_1, \ldots, v_k \rangle$ and is called an oriented k-simplex of X. The space $C_k(X)$ of k-chains on X is defined as the real vector space consisting of all linear combinations of oriented k-simplices under the relation that $\langle v_0, v_1, \ldots, v_k \rangle = -\langle v_1, v_0, \ldots, v_k \rangle$ for any oriented k-simplices.

For $k \geq 1$, the kth boundary operator $\partial_k : C_k(X) \to C_{k-1}(X)$ is defined as a linear map such that for any $\langle \sigma \rangle = \langle v_0, v_1, \ldots, v_k \rangle \in C_k(X)$,

$$\partial_k \langle \sigma \rangle = \sum_{i=0}^{k} (-1)^i \langle v_0, \ldots, v_{i-1}, v_{i+1}, \ldots, v_k \rangle.$$

By convention, we define $C_{-1}(X) = \mathbb{R}$, and $\partial_0 \colon C_0(X) \to C_{-1}(X)$ is defined as a linear map such that $\partial_0 \langle v \rangle = 1$ for $v \in V$. Then, it holds that $\partial_k \circ \partial_{k+1} = 0$ for all $k \geq 0$. The *kth homology group* of X over $\mathbb{R}$ and the *kth (reduced) Betti number* are defined as $H_k(X) := \ker \partial_k / \operatorname{Im} \partial_{k+1}$ and $\beta_k(X) := \dim H_k(X)$, respectively.[3] Intuitively, $\beta_k(X)$ is interpreted as the number of k-dimensional holes in X. In particular, $\beta_0(X)$ is equal to the number of connected components of X minus one. In the standard definition, we note that ∂_0 would be defined as the zero operator, which makes our zeroth Betti number defined above equal to the conventional zeroth Betti number minus one.

Research interest has been growing in the higher-dimensional analogue of Theorem 2.1 and related topics; see the survey by Kahle [16] for recent studies. In general, the homological structures of large simplicial complexes are expected to be very complicated. Indeed, as the number of simplices increases, so the effect of creating holes competes against that of filling holes, thereby making the situation more problematic than simply analyzing graphs. A distant goal is to extract nice fractal structures from these simplicial complexes, but initially it would be meaningful to develop effective tools with which to study the limit behavior as the number of vertices tends to infinity.

We now consider a family $\mathcal{X} = \{X(t)\}_{t \geq 0}$ of subcomplexes of X, and we call it a *right-continuous filtration* of X if $X(s) \subset X(t)$ for $0 \leq s \leq t$ and $X(t) = \bigcap_{t' > t} X(t')$ for $t \geq 0$. Here, $X(t)$ can be an empty set, which is regarded as a (-1)-dimensional simplicial complex. Let $\mathbb{R}[\mathbb{R}_{\geq 0}]$ be a real vector space of formal linear combinations of finite elements of $\mathbb{R}_{\geq 0}$. We describe each element of $\mathbb{R}_{\geq 0}$ as z^t ($t \in \mathbb{R}_{\geq 0}$), where z is indeterminate. The product of two elements of $\mathbb{R}[\mathbb{R}_{\geq 0}]$ is defined so as to be consistent with $az^s \cdot bz^t = abz^{s+t}$ ($a, b \in \mathbb{R}$ and $s, t \in \mathbb{R}_{\geq 0}$). This operation equips $\mathbb{R}[\mathbb{R}_{\geq 0}]$ with a ring structure. For $k \geq 0$, the kth persistent homology $\mathrm{PH}_k(\mathcal{X})$ of

[3] In general, we can define the spaces $C_k(X, R)$ and $H_k(X, R)$ as R-modules for a commutative ring R. In this paper, we consider only the case $R = \mathbb{R}$.

$\mathcal{X} = \{X(t)\}_{t\geq 0}$ is defined as

$$\mathrm{PH}_k(\mathcal{X}) = \bigoplus_{t\geq 0} H_k(X(t)),$$

which is regarded as a graded module over $\mathbb{R}[\mathbb{R}_{\geq 0}]$. Here, $H_k(X(s))$ is considered as a subset of $H_k(X(t))$ for $0 \leq s \leq t$ by a natural inclusion from $X(s)$ to $X(t)$. The structure theorem of the persistent homology is stated as follows.

Theorem 2.3 (e.g., see [22, 11]). *For each $k \geq 0$, there exist unique indices $p, q \in \mathbb{Z}_{\geq 0}$ and $\{b_i\}_{i=1}^{p+q}$, $\{d_i\}_{i=1}^{p} \subset \mathbb{R}_{\geq 0}$ such that $b_i < d_i$ for all $i = 1, \dots, p$, and the following graded module isomorphism holds:*

$$\mathrm{PH}_k(\mathcal{X}) \simeq \bigoplus_{i=1}^{p} \big((z^{b_i})/(z^{d_i})\big) \oplus \bigoplus_{i=p+1}^{p+q} (z^{b_i}),$$

where (z^a) denotes the ideal in $\mathbb{R}[\mathbb{R}_{\geq 0}]$ that is generated by the monomial z^a.

In Theorem 2.3, we call b_i and d_i the kth birth and death times, respectively, which indicate the appearance and disappearance of each k-dimensional "hole" in $\mathcal{X}$. The corresponding lifetime is defined as $l_i := d_i - b_i$. We set $d_i = l_i = \infty$ for $i = p+1, \dots, p+q$, and we define the lifetime sum $L_k(\mathcal{X})$ as

$$L_k(\mathcal{X}) = \sum_{i=1}^{p+q} (d_i - b_i).$$

The following is a generalization of the second identity of (2.1) to filtrations.

Theorem 2.4 (Lifetime formula [11, Proposition 2.2]). *It holds that*

$$L_k(\mathcal{X}) = \int_0^\infty \beta_k(X(t))\, dt.$$

Analogously, by defining

$$(L_k(\mathcal{X}))_T = \sum_{i=1}^{p+q} \big((d_i \wedge T) - (b_i \wedge T)\big)$$

for $T > 0$, we have

$$(L_k(\mathcal{X}))_T = \int_0^T \beta_k(X(t))\, dt.$$

An analogue of the first identity of (2.1) has also been obtained by introducing the concept of spanning acycles; see [11] for further details.

Now, we are interested in the asymptotic behavior of $L_k(\mathcal{X})$ for *random* filtrations as the number of vertices tends to infinity. The random models are introduced as follows.

For each $i \in \mathbb{Z}_{\geq 0}$, we take a probability distribution function p_i on $[0, +\infty]$. Let $n \in \mathbb{N}$ and let $K(n)$ denote the complete $(n-1)$-dimensional simplicial complex, namely the family of all nonempty subsets of an n-point set. We take a family of independent random variables $\{u_\tau\}_{\tau \in K(n)}$ such that each u_τ obeys the distribution function $p_{\dim \tau}$. We then define a random simplicial complex process $\mathcal{X}_n = \{X_n(t)\}_{t \geq 0}$ over n vertices by

$$X_n(t) := \{\sigma \in K(n) \mid u_\tau \leq t \text{ for every simplex } \tau (\neq \emptyset) \text{ with } \tau \subset \sigma\}. \tag{2.2}$$

We call this process a multi-parameter random complex process. We can also consider $\mathcal{X}_n = \{X_n(t)\}_{t \in [0,T]}$ for fixed $T > 0$ in an obvious manner. In this case, we write $L_k(\mathcal{X}_n)$ for $(L_k(\mathcal{X}_n))_T$.

We have the following typical examples in mind.

Example 2.5 (cf. [19]). Let $d \in \mathbb{N}$ be fixed. For each $i \in \mathbb{Z}_{\geq 0}$, define

$$p_i(t) = \begin{cases} 1 & (i < d) \\ t \wedge 1 & (i = d) \\ 0 & (i > d) \end{cases} \quad \text{for } t \geq 0.$$

In [10], the corresponding process $\mathcal{K}_n^{(d)} = \{K_n^{(d)}(t)\}_{t \in [0,1]}$ for $n > d$ and $T = 1$ is called the d-Linial–Meshulam complex process. By definition, for each $t \in [0,1]$, the random simplicial complex $K_n^{(d)}(t)\,(\subset K(n))$ is described as follows:

- $K_n^{(d)}(t)$ includes every simplex of $K(n)$ whose dimension is less than d.
- $K_n^{(d)}(t)$ includes each d-dimensional simplex of $K(n)$ with probability t independently.
- $K_n^{(d)}(t)$ includes no simplex of $K(n)$ whose dimension is greater than d.

The Erdős–Rényi graph process is identified with $\mathcal{K}_n^{(1)}$.

Example 2.6 (cf. [14]). Let $d \in \mathbb{N}$ be fixed. For each $i \in \mathbb{Z}_{\geq 0}$, define

$$p_i(t) = \begin{cases} 1 & (i < d) \\ t \wedge 1 & (i = d) \\ 1 & (i > d) \end{cases} \quad \text{for } t \geq 0.$$

In [10], the corresponding process $\mathcal{C}_n^{(d)} = \{C_n^{(d)}(t)\}_{t \in [0,1]}$ for $n > d$ and $T = 1$ is called the d-flag complex process. By definition, for each $t \in [0,1]$, the random simplicial complex $C_n^{(d)}(t)\,(\subset K(n))$ is described as follows:

- $C_n^{(d)}(t)$ includes every simplex of $K(n)$ whose dimension is less than d.
- $C_n^{(d)}(t)$ includes each d-dimensional simplex of $K(n)$ with probability t independently.
- $C_n^{(d)}(t)$ includes each simplex σ of $K(n)$ whose dimension is greater than d if and only if every d-dimensional face of σ belongs to $C_n^{(d)}(t)$.

$C_n^{(1)}$ is also called the random clique complex process.

Our main concern is the asymptotic behavior of $\mathbb{E}[L_k(\mathcal{X}_n)]$ as $n \to \infty$. To state the results, we introduce the following functions:

$$
\begin{aligned}
q_{-1}(t) &:= 1, \quad q_k(t) := \prod_{i=0}^{k} \{p_i(t)\}^{\binom{k+1}{i+1}} \quad (k \geq 0), \\
r_k(t) &:= \frac{q_{k+1}(t)}{q_k(t)} = \prod_{i=0}^{k+1} \{p_i(t)\}^{\binom{k+1}{i}} \quad (k \geq -1).
\end{aligned}
\tag{2.3}
$$

Note that $q_k(t)$ denotes the probability of a fixed k-simplex appearing at time t. For a k-simplex σ and a $(k+1)$-simplex τ with $\sigma \subset \tau$, $r_k(t)$ represents the conditional probability of τ appearing at time t given σ appearing.

Let $\check{r}_k$ denote the generalized inverse function of r_k, namely

$$
\check{r}_k(u) = \inf\{t \geq 0 \mid r_k(t) > u\} \quad \text{for } u < 1,
$$

and $\check{r}_k(1) = \infty$. We further define

$$
\begin{aligned}
&Q_k(t) = \int_0^t q_k(s)\, ds \quad \text{for } t \geq 0, \\
&\Phi_k(u) = Q_k(\check{r}_k(u)) \text{ and } \Psi_k(u) = Q_k(\check{r}_{k-1}(u)) \quad \text{for } u \in [0,1).
\end{aligned}
$$

In what follows, we use the standard notations big-O and little-o, and

- $f(u) = \Theta(g(u))$ means that $f(u) = O(g(u))$ and $g(u) = O(f(u))$ as $u \to 0$;
- $a_n \asymp b_n$ means that $a_n = O(b_n)$ and $b_n = O(a_n)$ as $n \to \infty$.

Below, k is a fixed number. The following result is a special case of more-general estimates [10, Theorems 4.3 and 4.4].

Theorem 2.7 ([10, Corollary 4.5]). *Suppose that $\Phi_k(u) = \Theta(u^a)$ for some $a \in [0, \infty)$ and $\Psi_k(u) = o(\Phi_k(u))$ as $u \to 0$. Then, for each $T > 0$,*

$$
\mathbb{E}[(L_k(\mathcal{X}_n))_T] \asymp n^{k+1-a}. \tag{2.4}
$$

Moreover, if $\int_0^\infty t^{1+\delta}\, dq_{k+1}(t) < \infty$ for some $\delta > 0$, then

$$
\mathbb{E}[L_k(\mathcal{X}_n)] \asymp n^{k+1-a}. \tag{2.5}
$$

The following is a rather simple case but is not treated in Theorem 2.7.

Theorem 2.8 ([10, Theorem 4.6]). *If $\Phi_k(u) = \Psi_k(u)$ for all $u \in [0,1)$, then $L_k(\mathcal{X}_n) = 0$ almost surely for all $n \in \mathbb{N}$.*

We apply these results to Examples 2.5 and 2.6.

Example 2.9 ([10, Example 4.8]). We consider the d-flag complex process $C_n^{(d)} = \{C_n^{(d)}(t)\}_{t\in[0,1]}$ as in Example 2.6. From straightforward computation, we obtain

$$(\Phi_k(u), \Psi_k(u)) = \begin{cases} (0,0) & (k < d-1), \\ (u,0) & (k = d-1), \\ \left(\Theta\left(u^{\frac{k+1-d}{d+1}+\binom{k+1}{d}^{-1}}\right), \Theta\left(u^{\frac{k+1}{d+1}+\binom{k}{d}^{-1}}\right)\right) & (k \geq d). \end{cases}$$

From Theorems 2.7 and 2.8, we have

$$\mathbb{E}[L_k(C_n^{(d)})] \asymp \begin{cases} 0 & (k < d-1), \\ n^{\frac{(k+2)d}{d+1}-\binom{k+1}{d}^{-1}} & (k \geq d-1). \end{cases}$$

In particular,

$$\mathbb{E}[L_k(C_n^{(1)})] \asymp n^{k/2+1-1/(k+1)}.$$

This estimate improves Theorem 6.10 in [11] and determines the growth order, thereby answering the question posed in [11, Section 7.4].

Example 2.10 ([10, Example 4.7]). We consider the d-Linial–Meshulam complex process $\mathcal{K}_n^{(d)} = \{K_n^{(d)}(t)\}_{t\in[0,1]}$ as in Example 2.5. It is straightforward to see that

$$(\Phi_k(u), \Psi_k(u)) = \begin{cases} (0,0) & (k < d-1), \\ (u,0) & (k = d-1), \\ (1/2, u^2/2) & (k = d), \\ (0,0) & (k > d). \end{cases}$$

From Theorems 2.7 and 2.8, we have

$$\mathbb{E}[L_k(\mathcal{K}_n^{(d)})] \asymp \begin{cases} 0 & (k \neq d-1, d), \\ n^{d-1} & (k = d-1), \\ n^{d+1} & (k = d). \end{cases}$$

The case $k = d-1$ corresponds to [11, Theorem 1.2].

In fact, we have more-precise asymptotics for $L_{d-1}(\mathcal{K}_n^{(d)})$. Following [20, 11], we introduce the limit constant. Let $t_1^* = c_1^* = 1$. For $d \geq 2$, let t_d^* be the unique root in $(0,1)$ of

$$(d+1)(1-t) + (1+dt)\log t = 0, \tag{2.6}$$

and define $c_d^* = (-\log t_d^*)/(1-t_d^*)^d > 0$. For $c \geq c_d^*$, let t_c denote the smallest positive root of $(-\log t)/(1-t)^d = c$. Define functions g_d and h_d on $[0,\infty)$ as

$$g_d(c) = \begin{cases} 0 & (c < c_d^*), \\ ct_c(1-t_c)^d + \dfrac{c}{d+1}(1-t_c)^{d+1} - (1-t_c) & (c \geq c_d^*), \end{cases}$$

and

$$h_d(c) = 1 - \frac{c}{d+1} + g_d(c).$$

We also define

$$I_{d-1} := \frac{1}{d!} \int_0^\infty h_d(s)\, ds.$$

Then, the limit behavior of $L_{d-1}(\mathcal{K}_n^{(d)})$ is described as follows.

Theorem 2.11 (part of [10, Theorem 4.11]). *Let $d \geq 1$. The constant I_{d-1} is finite, and for any $r \in [1, \infty)$,*

$$\lim_{n\to\infty} \mathbb{E}\left[\left|\frac{L_{d-1}(\mathcal{K}_n^{(d)})}{n^{d-1}} - I_{d-1}\right|^r\right] = 0;$$

in particular, $\mathbb{E}[L_{d-1}(\mathcal{K}_n^{(d)})]/n^{d-1}$ converges to I_{d-1} as $n \to \infty$.

This claim is a justification of an informal discussion in [11, Section 7.1]. Note that $I_0 = \zeta(3)$, and Theorem 2.11 with $d = 1$ is consistent with Theorem 2.2. See [10, Section 4.4] for explicit expressions for general I_{d-1} and further information. In particular, we have

$$\begin{aligned} I_1 &= \frac{1}{2}\left[\mathrm{Li}_2(t_2^*) + (\log t_2^*)\log(1-t_2^*) + \frac{t_2^*(\log t_2^*)^2}{2(1-t_2^*)} + \frac{(\log t_2^*)\{\log t_2^* + (1-t_2^*)\}}{4(1-t_2^*)^2}\right] \\ &= \frac{1}{2}\left[\mathrm{Li}_2(t_2^*) + (\log t_2^*)\log(1-t_2^*) + \frac{3(1-t_2^*)(1+3t_2^*)}{2(1+2t_2^*)^2}\right] \end{aligned} \tag{2.7}$$

and

$$\begin{aligned} I_2 &= \frac{1}{12}\left[\mathrm{Li}_2(t_3^*) + (\log t_3^* - 1)\log(1-t_3^*) + \frac{t_3^*(\log t_3^*)(\log t_3^* - 2)}{2(1-t_3^*)}\right. \\ &\qquad \left. + \frac{t_3^*(\log t_3^*)^2}{2(1-t_3^*)^2} + \frac{(\log t_3^*)\{\log t_3^* + (1-t_3^*)\}}{3(1-t_3^*)^3}\right] \\ &= \frac{1}{12}\left[\mathrm{Li}_2(t_3^*) + (\log t_3^* - 1)\log(1-t_3^*) + \frac{4((t_3^*)^2 + 5t_3^* + 1)}{(1+3t_3^*)^2}\right], \end{aligned} \tag{2.8}$$

where $\mathrm{Li}_2(x)$ denotes the dilogarithm

$$\mathrm{Li}_2(x) = \sum_{k=1}^\infty \frac{x^k}{k^2} \quad (-1 \leq x \leq 1).$$

We remark that the second identities of (2.7) and (2.8) follow from the fact that t_d^* is a root of (2.6).

3 Ideas for proving the theorems

In this section, we explain some basic ideas for proving the main results (Theorems 2.7 and 2.11) in the previous section, following [10]. Because

$$\mathbb{E}[(L_k(\mathcal{X}))_T] = \int_0^T \mathbb{E}[\beta_k(X(t))]\,dt \quad \text{and} \quad \mathbb{E}[L_k(\mathcal{X})] = \int_0^\infty \mathbb{E}[\beta_k(X(t))]\,dt$$

from Theorem 2.4, it suffices to obtain a sufficiently sharp estimate of $\mathbb{E}[\beta_k(X(t))]$ for each random simplicial complex $X(t)$. In general, it is a difficult problem to obtain a good estimate of a Betti number for a large simplicial complex X. The following is a basic estimate.

Lemma 3.1. *For every $k \geq 0$,*

$$f_k(X) - f_{k+1}(X) - f_{k-1}(X) \leq \beta_k(X) \leq f_k(X), \tag{3.1}$$

where $f_k(X)$ denotes the number of all k-simplices of X, and $f_{-1}(X) = 1$ by convention.

This is a version of the Morse inequality and is proved by simple application of linear algebra. Lemma 3.1 provides good upper and lower estimates of $\mathbb{E}[\beta_k(X(t))]$ if t is sufficiently small. In fact, as crucially noticed in [11], replacing X in the first inequality of (3.1) by $X(t)$ and integrating with respect to t on a small interval $[0, t_0]$ gives a lower estimate in (2.4) with the correct growth order. Thus, the main difficulty in the proof of Theorem 2.7 is the upper estimate in (2.4) and (2.5).

For general t, we require another strategy for estimating $\beta_k(X(t))$. To explain this strategy, we introduce several concepts from graph theory and topology. Let G be a finite undirected graph with a vertex set V, an edge set E, and with no loops or multiple edges. The degree $\deg(v)$ of a vertex $v \in V$ is defined as the number of $w \in V$ such that $\{v, w\} \in E$. The averaging matrix $A[G] = \{a_{vw}\}_{v,w\in V}$ of G is defined as

$$a_{vw} := \begin{cases} 1/\deg(v) & \text{if } \{v, w\} \in E, \\ 1 & \text{if } \deg(v) = 0 \text{ and } v = w, \\ 0 & \text{otherwise.} \end{cases}$$

This is interpreted as the transition probability of a simple random walk on G. The Laplacian $\mathcal{L}[G]$ of G is defined as $\mathcal{L}[G] = I_V - A[G]$, where I_V is the matrix that acts as the identity operator on V. Let $\{\lambda_i\}_{i=1}^{\#V}$ be all the (not necessarily distinct) eigenvalues of $\mathcal{L}[G]$. Note that $\lambda_i \in [0, 2]$ for all i and at least one λ_i is zero. Define

$$\gamma(G; \alpha) := \#\{i \mid \lambda_i \leq \alpha\} - 1 \,(\geq 0)$$

for $\alpha \geq 0$. By convention, $\gamma(\emptyset; \alpha) := 0$.

Given a D-dimensional simplicial complex X and a j-simplex τ in X with $-1 \leq j \leq D$, the *link* $\mathrm{lk}_X(\tau)$ of τ in X is defined as

$$\mathrm{lk}_X(\tau) := \{\sigma \in X \mid \tau \cap \sigma = \emptyset \text{ and } \tau \cup \sigma \in X\}.$$

This is either an empty set or a simplicial complex whose dimension is at most $D - j - 1$. Let $\mathrm{lk}_X(\tau)^{(1)}$ denote the 1-skeleton of $\mathrm{lk}_X(\tau)$, that is, the totality of the simplices of $\mathrm{lk}_X(\tau)$ whose dimensions are at most 1. This is either an empty set or a graph.

A key estimate is described as follows.

Theorem 3.2 ([10, Theorem 2.5]). *Suppose that the dimension D of X is greater than or equal to* 1. *Then*

$$\beta_{D-1}(X) \le \sum_{\tau} \gamma\big(\mathrm{lk}_X(\tau)^{(1)}; 1 - D^{-1}\big), \tag{3.2}$$

where τ in the summation is taken to be all $(D-2)$-simplices of X.

Informally speaking, this claim says that the Betti number is dominated by the sum of the number of small eigenvalues of the Laplacian on the 1-skeleton of each link of X. In particular, if the right-hand side of (3.2) is zero, then $H_{D-1}(X) = \{0\}$. In this sense, Theorem 3.2 is regarded as a quantitative generalization of the *cohomology vanishing theorem*[4] [8, 2]. The proof of Theorem 3.2 is based on a careful modification of that of [2, Theorem 2.1] and some additional arguments to remove extra assumptions.

From Theorem 3.2, under the assumption that (3.2) provides a sufficiently sharp estimate, the upper estimate of the Betti number is reduced to counting small eigenvalues of Laplacians on graphs. If X is a random simplicial complex, then this is closely related to the study of the eigenvalues of random matrices.

We apply this estimate to the following multi-parameter random simplicial complexes that were introduced in [3, 6]. Let $\{p_i\}_{i=0}^{\infty}$ be fixed parameters with $0 \le p_i \le 1$ for all i. We define a sequence of random simplicial complexes $\{X_n\}_{n\in\mathbb{N}}$ as follows. For each $n \in \mathbb{N}$, we start with a set V of n vertices and retain each vertex with independent probability p_0. Each edge with both ends retained is added with probability p_1, independently. Iteratively, for $i = 1, 2, \ldots, n-1$, each i-simplex for which all faces were added by the previous procedures is added with probability p_i, independently. The resulting random simplicial complex is X_n. From the definition, $\{X_n(t)\}_{n\in\mathbb{N}}$ defined in (2.2) for fixed t is nothing but $\{X_n\}_{n\in\mathbb{N}}$ with parameters $\{p_i(t)\}_{i=0}^{\infty}$.

Just as in (2.3), we define

$$q_{-1} := 1, \quad q_k := \prod_{i=0}^{k} p_i^{\binom{k+1}{i+1}} \quad (k \ge 0),$$
$$r_k := \frac{q_{k+1}}{q_k} = \prod_{i=0}^{k+1} p_i^{\binom{k+1}{i}} \quad (k \ge -1).$$

Then, a crucial estimate is described as follows.

---

[4] The proof is based on the discussion of the cohomology, not the homology. However, they are isomorphic.

Theorem 3.3 ([10, Theorem 3.6]). *Let $k \geq 0$ and $l \in \mathbb{N}$. Then, there exists a positive constant C depending only on k and l such that, for all $n \in \mathbb{N}$,*

$$\mathbb{E}[\beta_k(X_n)] \leq n^{k+1} q_k \left\{1 \wedge C(nr_k)^{-l}\right\}. \tag{3.3}$$

We give a brief outline of the proof of Theorem 3.3. Lemma 3.1 immediately implies the inequality

$$\mathbb{E}[\beta_k(X_n)] \leq n^{k+1} q_k. \tag{3.4}$$

Therefore, it suffices to prove the inequality

$$\mathbb{E}[\beta_k(X_n)] \leq C n^{k+1} q_k (nr_k)^{-l} \tag{3.5}$$

for some C. The proof is decomposed into the following three cases. The constants $K_1 \leq K_2$ below should be taken appropriately.

Case 1 If $r_k \geq \dfrac{K_1}{n} \vee \dfrac{(nr_{k-1})^{1/l}}{n}$, then the effect of "filling k-dimensional holes" is strong; (3.5) follows from a variant of the cohomology vanishing theorem of random simplicial complexes (e.g., [15, Theorem 1.1 (1)] and [6, Theorem 1.1]) that is based on insightful results regarding spectral gaps on random graphs by Hoffman, Kahle, and Paquette [12].

Case 2 If $\dfrac{K_2}{n} \leq r_k \leq \dfrac{(nr_{k-1})^{1/l}}{n}$, then we use a general inequality

$$\begin{aligned}
&\#\{\text{eigenvalues of } L \text{ (counting multiplicities) greater than } \alpha\} \\
&= \#\{\text{eigenvalues of } (L/\alpha)^l \text{ (counting multiplicities) greater than unity}\} \\
&\leq \operatorname{tr}((L/\alpha)^l) = \alpha^{-l} \operatorname{tr}(L^l)
\end{aligned}$$

for nonnegative-definite symmetric matrices L and $\alpha > 0$. Applying this by letting $L = \mathcal{L}[\operatorname{lk}_{X_n}(\tau)]$ with $\tau \in X_n$ and $\alpha = 1 - 1/(k+1)$, and using some combinatorial arguments for estimating $\operatorname{tr}(L^l)$, we can prove (3.5) via Theorem 3.2.

Case 3 If $r_k \leq \dfrac{K_2}{n}$, then (3.4) implies (3.5) for a suitable C.

Remark 3.4. As seen from the above explanation, the novel Betti-number estimate is that in the intermediate range (Case 2). We remark that combinatorial arguments that are similar in spirit are also found in the classical proof of Wigner's semicircle law of random matrices, albeit in a slightly different situation.

Now, we obtain

$$\begin{aligned}
\mathbb{E}[L_k(\mathcal{X}_n)] &= \int_0^\infty \mathbb{E}[\beta_k(X_n(t))]\, dt \quad \text{(from Theorem 2.4)} \\
&\leq \int_0^\infty n^{k+1} q_k(t) \left\{1 \wedge C(nr_k(t))^{-l}\right\} dt \quad \text{(from Theorem 3.3).}
\end{aligned}$$

Taking l to be sufficiently large and performing some elementary calculations, we reach an estimate $\mathbb{E}[L_k(\mathcal{X}_n)] = O(n^{k+1-a})$ as $n \to \infty$. The estimate of $\mathbb{E}[(L_k(\mathcal{X}_n))_T]$ is similarly proved, which completes the proof of Theorem 2.7.

In proving Theorem 2.11, the following is the key fact and follows from the results by Linial and Peled [20] that come from the convergence of a sequence of random graphs induced by $\{K_n^{(d)}(s/n)\}_{n\in\mathbb{N}}$ for fixed $s \geq 0$.

Theorem 3.5. *For any $s \geq 0$ and $\varepsilon > 0$,*

$$\lim_{n\to\infty} \mathbb{P}\left(\left|\frac{\beta_d(K_n^{(d)}(s/n))}{\binom{n}{d}} - g_d(s)\right| > \varepsilon\right) = 0.$$

With the help of the Euler–Poincaré formula, we can prove that, for each $s \geq 0$ and $\varepsilon > 0$,

$$\lim_{n\to\infty} \mathbb{P}\left(\left|\frac{\beta_{d-1}(K_n^{(d)}(s/n))}{\binom{n}{d}} - h_d(s)\right| > \varepsilon\right) = 0. \tag{3.6}$$

We note that

$$\begin{aligned}\left\|\frac{L_{d-1}(\mathcal{K}_n^{(d)})}{n^{d-1}} - I_{d-1}\right\|_{L^r} &= \left\|\int_0^\infty \left(\frac{\beta_{d-1}(K_n^{(d)}(s/n))}{n^d} 1_{[0,n]}(s) - \frac{1}{d!} h_d(s)\right) ds\right\|_{L^r} \\ &\leq \int_0^\infty U_n(s)\, ds,\end{aligned}$$

where

$$U_n(s) = \left\|\frac{\beta_{d-1}(K_n^{(d)}(s/n))}{n^d} 1_{[0,n]}(s) - \frac{1}{d!} h_d(s)\right\|_{L^r}.$$

Combining Theorem 3.3 and (3.6), we obtain $\lim_{n\to\infty} U_n(s) = 0$ for each $s \geq 0$ and $\sup_{n\in\mathbb{N}} U_n(s)$ is Lebesgue integrable over $[0,\infty)$. The dominated convergence theorem implies that $\int_0^\infty U_n(s)\, ds$ converges to zero as $n \to \infty$, which finishes the proof of Theorem 2.11.

A similar outline was discussed informally in [11, Section 7.1]. However, because we now have the uniform estimate (3.3), we can provide a rigorous proof.

4 Concluding remarks

Theorems 2.7 and 2.11 remain at the beginning of the homological study of families of random simplicial complexes. We will describe some potential directions for future research.

1. In [11], discrete Morse theory was used for estimating $L_k(\mathcal{X}_n)$. Although the argument therein did not provide the optimal asymptotics, it may be interesting to investigate that approach further.

2. Work is in progress [17] to prove the existence and identify the limit of scaled expectations of $L_k(\mathcal{X}_n)$ (Theorem 2.11) for general models other than d-Linial–Meshulam complex processes.
3. The limit constants [e.g., (2.7) and (2.8)] for d-Linial–Meshulam complex processes are regarded as "higher-dimensional" analogues of $\zeta(3)$, but the question remains as to whether they have simpler expressions.
4. As already mentioned in [11], the next problem to be considered is proving the central limit theorem for $L_k(\mathcal{X}_n)$. In the case of the Erdős–Rényi process, this has been proved by Janson [13].
5. The sum of the αth power ($\alpha > 0$) of lifetimes was studied in [10, Theorem 4.11] for d-Linial–Meshulam complex processes. In any further investigation, it would not be sufficient to study only the homologies of the simplicial complexes $X_n(t)$ for fixed t: we require the homological structure of the filtration $\{X_n(t)\}_{t\geq 0}$ itself.
6. Regarding item 5 in this list, the scaling limit of graphs in the Gromov–Hausdorff–Prokhorov topology has also been studied extensively (see [1, 21] and the references therein for recent studies). The limit objects in that case would have fractal structures and should provide detailed information about random graphs. Studying the counterpart of random simplicial complexes or their filtrations would be required for more-comprehensive understanding.

Acknowledgements This study was supported by JSPS KAKENHI Grant Numbers JP19H00643 and JP19K21833.

References

1. L. Addario-Berry, N. Broutin, C. Goldschmidt, and G. Miermont, The scaling limit of the minimum spanning tree of the complete graph, Ann. Probab. **45** (2017), 3075–3144.
2. W. Ballmann and J. Świątkowski, On L^2-cohomology and property (T) for automorphism groups of polyhedral cell complexes, Geom. Funct. Anal. **7** (1997), 615–645.
3. A. Costa and M. Farber, Random simplicial complexes, in: *Configuration spaces*, Springer INdAM Ser. **14**, 129–153, Springer, 2016.
4. P. Erdős and A. Rényi, On random graphs, Publ. Math. Debrecen **6** (1959), 290–297.
5. P. Erdős and A. Rényi, On the evolution of random graphs, Publ. Math. Inst. Hungarian Acad. Sci. **5A** (1960), 17–61.
6. C. F. Fowler, Homology of multi-parameter random simplicial complexes, Discrete Comput. Geom. **62** (2019), 87–127.
7. A. M. Frieze, On the value of a random minimum spanning tree problem, Discrete Applied Math. **10** (1985), 47–56.
8. H. Garland, p-adic curvature and the cohomology of discrete subgroups of p-adic groups, Ann. of Math. (2) **97** (1973), 375–423.
9. E. N. Gilbert, Random graphs, Ann. Math. Statist. **30** (1959), 1141–1144.
10. M. Hino and S. Kanazawa, Asymptotic behavior of lifetime sums for random simplicial complex processes, J. Math. Soc. Japan **71** (2019), 765–804.
11. Y. Hiraoka and T. Shirai, Minimum spanning acycle and lifetime of persistent homology in the Linial–Meshulam process, Random Structures Algorithms **51** (2017), 315–340.
12. C. Hoffman, M. Kahle, and E. Paquette, Spectral gaps of random graphs and applications, to appear in Int. Math. Res. Not. IMRN. doi: 10.1093/imrn/rnz077

13. S. Janson, The minimal spanning tree in a complete graph and a functional limit theorem for trees in a random graph, Random Structures Algorithm **7** (1995), 337–355.
14. M. Kahle, Topology of random clique complexes, Discrete Math. **309** (2009), 1658–1671.
15. M. Kahle, Sharp vanishing thresholds for cohomology of random flag complexes, Ann. of Math. (2) **179** (2014), 1085–1107.
16. M. Kahle, Topology of random simplicial complexes: a survey, in: *Algebraic topology: applications and new directions*, 201–221, Contemp. Math. **620**, Amer. Math. Soc., Providence, RI, 2014.
17. S. Kanazawa, in preparation.
18. J. B. Kruskal, Jr., On the shortest spanning subtree of a graph and the traveling salesman problem, Proc. Amer. Math. Soc. **7** (1956), 48–50.
19. N. Linial and R. Meshulam, Homological connectivity of random 2-complexes, Combinatorica **26** (2006), 475–487.
20. N. Linial and Y. Peled, On the phase transition in random simplicial complexes, Ann. of Math. (2) **184** (2016), 745–773.
21. K. Panagiotou, B. Stufler, and K. Weller, Scaling limits of random graphs from subcritical classes, Ann. Probab. **44** (2016), 3291–3334.
22. A. Zomorodian and G. Carlsson, Computing persistent homology, Discrete Comput. Geom. **33** (2005), 249–274.