Towards Spectral Sparsification of Simplicial Complexes Based on Generalized Effective Resistance

Braxton Osting* University of Utah
Sourabh Palande† University of Utah
Bei Wang‡ University of Utah

Abstract

As a generalization of the use of graphs to describe pairwise interactions, simplicial complexes can be used to model higher-order interactions between three or more objects in complex systems. There has been a recent surge in activity for the development of data analysis methods applicable to simplicial complexes, including techniques based on computational topology, higher-order random processes, generalized Cheeger inequalities, isoperimetric inequalities, and spectral methods. In particular, spectral learning methods (e.g. label propagation and clustering) that directly operate on simplicial complexes represent a new direction emerging from the confluence of computational topology and machine learning. Similar to the challenges faced by massive graphs, computational methods that process simplicial complexes are severely limited by computational costs associated with massive datasets.

To apply spectral methods in learning to massive datasets modeled as simplicial complexes, we work towards the sparsification of simplicial complexes based on preserving the spectrum of the associated Laplacian operators. We show that the theory of Spielman and Srivastava for the sparsification of graphs extends to simplicial complexes via the up Laplacian. In particular, we introduce a generalized effective resistance for simplexes; provide an algorithm for sparsifying simplicial complexes at a fixed dimension; and give a specific version of the generalized Cheeger inequality for weighted simplicial complexes. Finally, we demonstrate via experiments the preservation of the up Laplacian during sparsification, as well as the utility of sparsification with respect to spectral clustering.

*E-mail: osting@math.utah.edu.
†E-mail: sourabh@sci.utah.edu.
‡E-mail: beiwang@sci.utah.edu.

1 Motivation

Our work towards spectral sparsification of simplicial complexes is primarily motivated by learning based on simplicial complexes. Simplicial complexes capture higher-order interactions in complex systems; and there has been much recent activity in developing spectral theory for higher-order Laplacians as well as learning algorithms that operate on these Laplacians. We are interested in understanding the behavior of spectral learning algorithms on compact representations that preserve the spectral structure of data.

Simplicial complexes in data analysis. Understanding massive systems with complex interactions and multi-scale dynamics is important in a variety of social, biological, and technological settings. One approach to understanding such systems is to represent them as graphs where vertices represent objects and (weighted) edges represent *pairwise interactions* between the objects. A large arsenal of methods has now been developed to analyze properties of graphs, which can then be combined with domain-specific knowledge to infer properties of the system being studied. These tools include graph partitioning and clustering [42, 56, 57], random processes on graphs [25], graph distances, various measures of graph connectivity [41], combinatorial graph invariants [16], and spectral graph theory [12].

While graphs have been used with great success to describe pairwise interactions between objects in datasets, they fail to capture *higher-order interactions* that occur between three or more objects. These structures in data can be described using *simplicial complexes* [27, 38]. There has recently been a surge in activity for the development of data analysis methods that focus on simplicial complexes, including methods based on computational topology [8, 20, 23, 27], higher-order random processes [5, 26], generalized Cheeger inequalities [28, 54], isoperimetric inequalities [45], high-dimensional expanders [17, 35, 44], and spectral methods [29]. In particular, topological data analysis methods using simplicial complexes as the underlying combinatorial structures have been successfully employed for applications as diverse as the discovery of a new subtype of breast cancer [40], describing high-contrast patches in images [34], time series analysis [46], multi-channel communication [47] and sensor networks [13], statistical ranking [31, 43], and visualization [58]. Simplicial complexes have also been used to generalize graphical models in machine learning, where faces of dimension two or higher represent higher-order conditional dependence relations between random variables [19].

Sparsification of simplicial complexes. For unstructured graphs representing massive datasets, the computational costs associated with naïve implementations of many graph-based tools is prohibitive. In this scenario, it is useful to approximate the original graph with one having fewer edges or vertices while preserving some properties of interest in some appropriate metric, known as *graph sparsification*. A variety of graph sparsification methods have been developed, allowing for their efficient storage and computation [3, 50, 51]. Our work is inspired by and based on the seminal work by Spielman and Srivastava [50].

Similar to the challenges faced by massive graphs, computational methods that operate on simplicial complexes are severely limited by computational costs associated with massive datasets. There have been recent approaches from computational topology to construct sparse simplicial complexes that give good approximation results for computing persistent homology [6, 7, 9, 11, 14, 15, 33, 49, 55]. Persistence homology [21] turns the algebraic concept of homology into a multi-scale notion. It typically operates on a sequence of simplicial complexes (referred to as a *filtration*), constructs a series of homology groups and measures their relevant scales in the filtration. Common simplicial filtrations arise from Čech or (Vietoris-)Rips complexes, and most of the aforementioned

techniques produce sparsified Rips complexes that give guaranteed approximations to the persistent homology of the unsparsified filtration. The sparsification processes involve either the removal or subsampling of vertices, or edge contractions from the sparse filtration. While these techniques focus on the preservation of homological features in the data (referred to as *homological sparsification*), to the best of our knowledge, there are no known results regarding the higher-order spectral properties of the sparsified simplicial complexes. Our approach significantly differs from these techniques in the sense that it focuses on *spectral sparsification* which preserves higher-order spectral properties of the simplicial complexes, instead of homological ones.

Spectral sparsification and machine learning. A recurring theme in machine learning focuses on graph-based learning, where the data assume an underlying graph structure and one would like to infer information about the nodes of the graph. Spectral methods for graph-based learning such as spectral clustering (e.g. [1, 52]) are essential to many seminal papers in the field, which typically have good theoretical guarantees and efficient solutions for problems ranging from image segmentation [36] to community detection [2]. Some important instances of semi-supervised graph-based learning algorithms are often referred to as label propagation methods [30, 60] where labels on the nodes are propagated along the edges of the graph.

Inspired by graph-based learning, a setting which can be used to describe complex data, learning (indirectly or directly) based on simplicial complexes represents a new direction recently emerging from the confluence of computational topology and machine learning. On one hand, topological features derived from simplicial complexes, used as input to machine learning algorithms, have shown to increase the strength of prediction or classification compared to graph-theoretic features [4, 59]. On the other hand, we would like to develop learning algorithms that directly operate on simplicial complexes. For example, researchers have begun to develop mathematical intuition behind higher-dimensional notions of spectral clustering and label propagation [37, 54, 57].

Our work on spectral sparsification of simplicial complexes would play an important role in applying spectral learning algorithms to massive datasets modeled as simplicial complexes.

Overview. In this paper, we work towards developing computational methods for the spectral sparsification of simplicial complexes, in particular:

- We introduce a *generalized effective resistance* of simplexes by extending the notion of *effective resistance* of edges (e.g. [10, 18, 22]); see Section 3.
- We show that the algorithm in [50] for sparsifying graphs can be generalized to the simplification of simiplicial complexes at a fixed dimension, and prove that the spectrum of the *up Laplacian* is preserved under sparsification in the sense that the spectrum of the up Laplacian for the sparsified simplicial complex can be bounded in terms of the spectrum of the up Laplacian for the original simplicial complex; see Theorem 3.1.
- We generalize the Cheeger constant of Gundert and Szedlák for *unweighted* simplicial complexes [28] to *weighted* simplicial complexes and verify that the Cheeger inequality involving the first non-trivial eigenvalue of the *weighted* up Laplacian holds in the sparsfied setting; see Proposition 4.1.
- Our theoretical results are supported by numerical experiments in Section 5. These experiments illustrate the inequalities bounding the spectrum of the up Laplacian of the sparsified simplicial complex, proven in Theorem 3.1. Further experiments demonstrate that simplicial complex clusters, obtained via a method extending spectral clustering to simplicial complexes,

are preserved by spectral sparsification. This application exemplifies the utility of the spectral sparsification methods developed here.

We proceed by reviewing background results and introducing notation in Section 2 that gives a brief description of relevant algebraic concepts, effective resistance, and spectral sparsification of graphs. The theory and algorithm for sparsifying simplicial complexes are provided in Section 3. We state the implications of the algorithm for a generalized Cheeger cut for the simplicial complex in Section 4. We showcase some experimental results validating our algorithms in Section 5. We conclude with some open questions in Section 6.

2 Background

Simplicial complexes. A *simplicial complex* K is a finite collection of simplices such that every face of a simplex of K is in K and the intersection of any two simplexes of K is a face of each of them [38]. 0-, 1- and 2-simplices correspond to vertices, edges and triangles. An *oriented simplex* is a simplex with a chosen ordering of its vertices. Let $S_p(K)$ denote the collection of all oriented p-simplices of K and $n_p = |S_p(K)|$. The *p-skeleton* of K is denoted as $K^{(p)} := \cup_{0\le i\le p} S_i(K)$. For the remainder of this paper, let K be an oriented simplicial complex on a vertex set $[n] = \{1, 2, \ldots, n\}$. Let $\dim K$ denote the dimension of K. For a review of simplicial complexes, see [24, 27, 38].

Laplace operators on simplicial complexes. The i-th *chain group* $C_i(K) = C_i(K, \mathbb{R})$ of a complex K with coefficient $\mathbb{R}$ is a vector space over the field $\mathbb{R}$ with basis in $S_i(K)$. The i-th *cochain group* $C^i(K) = C^i(K, \mathbb{R})$ is the dual of the chain group, defined by $C^i(K) := \mathrm{Hom}(C_i(K), \mathbb{R})$, where $\mathrm{Hom}(C_i, \mathbb{R})$ denotes all homomorphisms of C_i into $\mathbb{R}$. The coboundary operator, $\delta_i \colon C^i(K) \to C^{i+1}(K)$, is defined as

$$(\delta_i f)([v_0, \ldots, v_{i+1}]) = \sum_{j=1}^{i+1} (-1)^j f([v_0, \ldots, \hat{v}_j, \ldots, v_{i+1}]),$$

where $\hat{v}_j$ denotes that the vertex v_j has been omitted. It satisfies the property $\delta_i \delta_{i-1} = 0$ which implies that $\mathrm{im}(\delta_{i-1}) \subset \ker(\delta_i)$. The boundary operators, δ_i^*, are the adjoints of the coboundary operators,

$$\cdots \quad C^{i+1}(K) \underset{\delta_i^*}{\overset{\delta_i}{\leftrightarrows}} C^i(K) \underset{\delta_{i-1}^*}{\overset{\delta_{i-1}}{\leftrightarrows}} C^{i-1}(K) \quad \cdots$$

satisfying $(\delta_i a, b)_{C^{i+1}} = (a, \delta_i^* b)_{C^i}$ for every $a \in C^i(K)$ and $b \in C^{i+1}(K)$, where $(\cdot, \cdot)_{C^i}$ denote the scalar product on the cochain group.

Following [29], we define three combinatorial Laplace operators that operate on $C^i(K)$ (for the i-th dimension). Namely, the *up Laplacian*,

$$\mathcal{L}_i^{\mathrm{up}}(K) = \delta_i^* \delta_i,$$

the *down Laplacian*, $\mathcal{L}_i^{\mathrm{down}}(K) = \delta_{i-1} \delta_{i-1}^*$, and the *Laplacian*, $\mathcal{L}_i(K) = \mathcal{L}_i^{\mathrm{up}}(K) + \mathcal{L}_i^{\mathrm{down}}(K)$. All three operators are self-adjoint, non-negative, compact and enjoy a collection of spectral properties, as detailed in [29]. We restrict our attention to the up Laplacians.

Explicit expression for the up Laplacian. To make the expression of up Laplacian explicit, we need to choose a scalar product on the coboundary vector spaces, which can be viewed in terms of weight functions [29]. In particular, the weight function w is evaluated on the set of all simplices of K,

$$w : \bigcup_{i=0}^{\dim K} S_i(K) \to \mathbb{R}^+,$$

where the weight of a simplex f is $w(f)$. Let $w_i \colon S_i(K) \to \mathbb{R}^+$. Then $C^i(K)$ is the space of real-valued functions on $S_i(K)$, with inner product

$$(a, b)_{C^i} := \sum_{f \in S_i(K)} w_i(f) a(f) b(f),$$

for every $a, b \in C^i(K)$.

Choosing the natural bases, we identify each coboundary operator δ_p with an incidence matrix D_p. The *incidence matrix* $D_p \in \mathbb{R}^{n_{p+1}} \times \mathbb{R}^{n_p}$ encodes which p-simplices are incident to which $(p+1)$ simplices in the complex, and is defined as

$$D_p(i,j) = \begin{cases} 0 & \text{if } \sigma_j^p \text{ is not on the boundary of } \sigma_i^{p+1} \\ 1 & \text{if } \sigma_j^p \text{ is coherent with the induced orientation of } \sigma_i^{p+1} \\ -1 & \text{if } \sigma_j^p \text{ is not coherent with the induced orientation of } \sigma_i^{p+1} \end{cases}$$

Let D_p^T be the transpose of D_p. Let W_i be the diagonal matrix representing the scalar product on $C^i(K)$. The i-dimensional up Laplacian can then be expressed in the chosen bases, as the matrix

$$\mathcal{L}_{K,i} := \mathcal{L}_i^{\text{up}}(K) = W_i^{-1} D_i^T W_{i+1} D_i.$$

With this notation, $L = \mathcal{L}_{K,0}$ is the graph Laplacian.

Effective resistance. We quickly review the notation in [50] regarding effective resistance. Let $G = (V, E, w)$ be a connected weighted undirected graph with n vertices and m edges and edge weights $w_e \in \mathbb{R}^+$. W is an $m \times m$ diagonal matrix with $W(e,e) = w_e$. Suppose the edges are oriented arbitrarily. Its graph Laplacian $L \in \mathbb{R}^{n \times n}$ can be written as

$$L = B^T W B,$$

where $B \in \mathbb{R}^{m \times n}$ is the signed edge-vertex incidence matrix, that is,

$$B(i,j) = \begin{cases} 0 & \text{if vertex } j \text{ is not on the boundary of edge } i \\ 1 & \text{if } j \text{ is } i\text{'s head} \\ -1 & \text{if } j \text{ is } i\text{'s tail.} \end{cases}$$

The effective resistance R_e at an edge e is the energy dissipation (potential difference) when a unit current is injected at one end and removed at the other end of e [50]. Define the matrix

$$R := B(L)^+ B^T = B(B^T W B)^+ B^T,$$

where L^+ denote the Moore-Penrose pseudoinverse of L. The diagonal entry $R(e,e)$ of R, is the effective resistance R_e across e. That is, $R_e = R(e,e)$.

The above expression for L is consistent with previous notation of up Laplacian, by setting $B = D_0$, $W = W_1$ for $L = \mathcal{L}_{K,0} = W_0^{-1} D_0^T W_1 D_0$. Suppose $W_0 = I$ (identity matrix), then R could be expressed as

$$R = D_0 (L)^+ D_0^T = D_0 (D_0^T W_1 D_0)^+ D_0^T.$$

Graph sparsification. There are several different notions of approximation for graph sparsification, including the following based on spectral properties of the associated graph Laplacian. We say $H = (V, F, u)$ is an *ϵ-approximate sparse graph* of $G = (V, E, w)$ if

$$(1-\epsilon)L_G \preceq L_H \preceq (1+\epsilon)L_G, \tag{1}$$

where L_G and L_H are the graph Laplacians of G and H respectively and the inequalities are to be understood in the sense of the semi-definite matrix ordering. That is, $\forall x \in \mathbb{R}^n$,

$$(1-\epsilon)x^T L_G x \leq x^T L_H x \leq (1+\epsilon)x^T L_G x. \tag{2}$$

It is well-known from spectral graph theory that the spectrum of the graph Laplacian controls a variety of properties of interest including the size of cuts (i.e. bottlenecks), clusters (i.e. communities), distances, various random processes (i.e. PageRank), and combinatorial properties (e.g. coloring, spanning trees, etc.). It follows that a sparsified graph, H, with graph Laplacian satisfying Equation (1), yields a great deal of information about the original graph, G, at a reduced computational cost. Recently, methods have been developed to efficiently compute sparsified graphs. In particular, it has been proven that every weighted graph G with n vertices and m edges has as an *ϵ-approximate sparse graph* H with at most $O(n \cdot \log(n)/\epsilon^2)$ edges and moreover, by subsampling the original graph with probabilities based on effective resistance, this graph can be found efficiently in $O(m/\epsilon^2)$ time [50].

3 Sparsification of simplicial complexes

To prove the existence of ϵ-approximate sparse simplicial complex, we will follow the approach of [51] for the analogus problem for graphs.

Generalized effective resistance for simplicial complexes. To generalize effective resistance for simplices beyond dimension 1 (i.e. edges), we consider the operator $R_i \colon C^i \to C^i$, defined by

$$R_i = D_{i-1}(\mathcal{L}_{K,i-1})^+ D_{i-1}^T = D_{i-1}\left(W_{i-1}^{-1} D_{i-1}^T W_i D_{i-1}\right)^+ D_{i-1}^T.$$

Specifically, setting $W_{i-1} = I$, we have

$$R_i = D_{i-1}(\mathcal{L}_{K,i-1})^+ D_{i-1}^T = D_{i-1}\left(D_{i-1}^T W_i D_{i-1}\right)^+ D_{i-1}^T,$$

which is the projection onto the image of D_{i-1}[1]. The *generalized effective resistance* on the i-dimensional simplex f, is defined to be the diagonal entry, $R_i(f, f)$.

For $i = 1$, the generalized effective resistance reduces to the effective resistance on the graph [22]. That is, recall in the notation from Section 2, let $B = D_0$ and assume $W_0 = I$, we have $R = R_1 = D_0(D_0^T W_1 D_0)^+ D_0^T$.

Sparsification algorithm. Algorithm 1, is a natural generalization of the **Sparsify** Algorithm given in [50]. The algorithm sparsifies a given simplicial complex K at a fixed dimension i (while ignoring all dimensions larger than i). The main idea is to include each i-simplex f of K in the sparsifier J with probability proportional to its generalized effective resistance. Specifically, for a

[1]For the rest of this section, we always consider R_i by assuming $W_{i-1} = I$ in our simplicial complex K. However our result will hold for any other choice of weights W_{i-1} in dimension $i-1$ since for symmetric matrices A, B, $A \succeq B$ if and only if $DA \succeq DB$ for any positive definite diagonal matrix D.

Algorithm 1: $J = \textbf{Sparsify}(K, i, q)$, a sparsification algorithm for simplicial complexes.

Data: A weighted, oriented simplicial complex K, a dimension i (where $1 \le i \le \dim K$), and an integer q.

Result: A weighted, oriented simplicial complex J which is sparsified at dimension i, with equivalent $(i-1)$-skeleton to K and $\dim J = i$.

$J := K^{(i-1)}$

Sample q i-dimensional simplices independently with replacement according to the probability

$$p_f = \frac{w(f) R_i(f,f)}{\sum_f w(f) R_i(f,f)},$$

and add sampled simplexes to J with weight $w(f)/qp_f$. If a simplex is chosen more than once, the weights are summed.

fixed dimension i, the algorithm chooses a random i-simplex f of K with probability p_f (proportional to $w_f R_f$), and adds f to J with weight w_f/qp_f; then q samples are taken independently with replacement, while summing weights if a simplex is chosen more than once. The following theorem (Theorem 3.1) shows that if q is sufficiently large, the i-dimensional up Laplacians of K and J are close.

Theorem 3.1. *Let K be a weighted, oriented simpicial complex, and $J = \textbf{Sparsify}(K, i, q)$ for some fixed i (where $1 \le i \le \dim K$). Suppose K and J have $(i-1)$-th up Laplacians $\mathcal{L}_K := \mathcal{L}_{K,i-1}$ and $\mathcal{L}_J := \mathcal{L}_{J,i-1}$ respectively. Let n_{i-1} denote the number of $(i-1)$-simplexes in K. Fix $\epsilon > 0$ (where $1/\sqrt{n_{i-1}} < \epsilon \le 1$), and let $q = 9C^2 n_{i-1} \log n_{i-1}/\epsilon^2$, where C is an absolute constant. If n_{i-1} is sufficently large, then with probability at least $1/2$,*

$$(1-\epsilon)\mathcal{L}_K \preceq \mathcal{L}_J \preceq (1+\epsilon)\mathcal{L}_K, \tag{3}$$

where the inequalities are to be understood in the sense of the semi-definite matrix ordering. Equivalently, this means, $\forall x \in \mathbb{R}^{n_{i-1}}$,

$$(1-\epsilon)x^T\mathcal{L}_K x \le x^T\mathcal{L}_J x \le (1+\epsilon)x^T\mathcal{L}_K x. \tag{4}$$

Proof. For simplicity in notation, let $\mathcal{L} = \mathcal{L}_K := \mathcal{L}_{K,i-1}$ and $\tilde{\mathcal{L}} = \mathcal{L}_J := \mathcal{L}_{J,i-1}$, with corresponding weight matrices denoted as W_i and $\tilde{W}_i$ respectively.

Our proof follows the proof of [50, Theorem 1]. We consider the projection matrix $\Pi = W_i^{1/2} R_i W_i^{1/2}$. We also define the $n_i \times n_i$ nonnegative, diagonal matrix S_i with entries

$$S_i(f,f) = \frac{\tilde{w}_f}{w_f} = \frac{\#\text{ times } f \text{ is sampled}}{qp_f},$$

where the random entry $S_i(f,f)$ captures the amount of i-simplexes f included in J by **Sparsify**. The weight of an i-simplex f in J is $\tilde{w}_f = S_i(f,f)w_f$. Since $\tilde{W}_{i-1} = W_{i-1}S_{i-1} = W_{i-1}^{1/2}S_{i-1}W_{i-1}^{1/2}$ and $\tilde{W}_i = W_i S_i = W_i^{1/2} S_i W_i^{1/2}$, the $(i-1)$-dimensional up Laplacian of J is therefore

$$\tilde{\mathcal{L}} = \mathcal{L}_{J,i-1} = \tilde{W}_{i-1} D_{i-1}^T \tilde{W}_i D_{i-1} = (W_{i-1}^{1/2} S_{i-1} W_{i-1}^{1/2}) D_{i-1}^T (W_i^{1/2} S_i W_i^{1/2}) D_{i-1}.$$

Since $\mathbb{E}S_i = I$, $\mathbb{E}S_{i-1} = I$ and suppose $W_{i-1} = I$, therefore $\mathbb{E}\tilde{\mathcal{L}} = \mathcal{L}$. It is not difficult to show that if S is a non-negative diagonal matrix such that

$$\|\Pi S \Pi - \Pi\Pi\|_2 \le \epsilon \tag{5}$$

then (3) holds. But, $\Pi S \Pi$ can be expressed as the average of symmetric, rank-one matrices. Now applying a result of Rudelson and Vershynin [48, Theorem 3.1], we have that $\mathbb{E}\|\Pi S \Pi - \Pi\Pi\|_2 \leq \frac{\epsilon}{2}$. By Markov's inequality this implies that (5) holds with probability at least $\frac{1}{2}$. □

Computing the generalized effective resistance. Recall the work in [50] uses the linear system solver of Spielman and Teng [53], which applies to all symmetrical positive semidefinite matrices, for graph sparsification. Recent methods for solving symmetric diagonally dominant (SDD) linear systems based on low-stretch spanning trees improves the running time further [32]. In particular, these solvers are applicable for approximating the generalized effective resistance here.

4 Generalized Cheeger inequalities for simplicial complexes

Cheeger constant and inequality for graphs. The Cheeger constant for an unweighted graph $G = (V, E)$ is given by

$$h(G) := \min_{\varnothing \subsetneq A \subsetneq V} \frac{|V|\ |E(A, V \setminus A)|}{|A|\ |V \setminus A|}, \tag{6}$$

where $E(A, B)$ is the set of edges that connect $A \subset V$ to $B \subset V$. For a weighted graph, $G = (V, E, w)$, the Cheeger constant is typically generalized to

$$h(G) := \min_{\varnothing \subsetneq A \subsetneq V} \frac{|V|}{|A|\ |V \setminus A|} \sum_{(i,j) \in E(A, V \setminus A)} w_{ij}. \tag{7}$$

Using the variational formulation for eigenvalues and a suitable test function, it is not difficult to prove the Cheeger inequality,

$$\lambda_1(L_G) \leq 2h(G),$$

where $\lambda_1(L_G)$ is the first non-trivial eigenvalue of the weighted graph Laplaican [12, Lemma 2.1].

Generalized Cheeger inequality for simplicial complexes of Gundert and Szedlák. We first recall the generalized Cheeger inequality for simplicial complexes of Gundert and Szedlák [28]. For a k-dimensional simplicial complex K, its *k-dimensional completion* is defined to be

$$\bar{K} := K \bigcup \{\tau^* \in \binom{V}{k+1} \mid (\tau^* \setminus v) \in X, \forall v \in \tau^*\}.$$

When K has a complete $(k-1)$-skeleton, $\bar{K}$ is the complete k-dimensional complex. The generalized Cheeger constant for *unweighted* simplicial complexes is defined to be

$$h(K) := \min_{V = \bigsqcup_{i=0}^{k} A_i \atop A_i \neq \varnothing} \frac{|V||F(A_0, A_1, \ldots, A_k)|}{|F^*(A_0, A_1, \ldots, A_k)|}, \tag{8}$$

where $F(A_0, A_1, \ldots, A_k)$ and $F^*(A_0, A_1, \ldots, A_k)$ are the sets of all k-simplices of K and $\overline{K}$, respectively, with one node in A_i for all $0 \leq i \leq k$.

Theorem 4.1 ([28, Theorem 2])**.** *If $\lambda_1(\mathcal{L}_K)$ is the first non-trivial eigenvalue of the k-th up-Laplacian and C^* is a the maximum number of $(k-1)$-simplices contained in a k-simplex of K, then*

$$\lambda_1(\mathcal{L}_K) \leq \frac{(k+1)\ C^*}{|V|}\ h(K).$$

We remark that an alternative Cheeger inequality is given in [45].

A generalized Cheeger constant for weighted simplicial complexes. In analogy to the generalization of the unweighted Cheeger constant in Equation (6) to the weighted Cheeger constant in Equation (7), we define the generalized Cheeger constant for weighted simplicial complexes by

$$h(K) := \min_{\substack{V=\bigsqcup_{i=0}^{k} A_i \\ A_i \neq \varnothing}} \frac{|V|}{|F^*(A_0, A_1, \ldots, A_k)|} \sum_{X \in F(A_0, A_1, \ldots, A_k)} w_k(X). \tag{9}$$

Observe that Equation (9) agrees with Equation (8) in the case when all weights are unity. The following result can be proved analogously to Theorem 4.1.

Proposition 4.1. *If $\lambda_1(\mathcal{L}_K)$ is the first non-trivial eigenvalue of the k-th weighted up-Laplacian and C^* is a the maximum number of $(k-1)$-simplices contained in a k-simplex of K, then*

$$\lambda_1(\mathcal{L}_K) \leq \frac{(k+1)\ C^*}{|V|}\ h(K).$$

The proof of Proposition 4.1 involves only a slight modification of the arguments in [28] by adapting weights in the definition of the generalized Cheeger constant (Equation 9), we omit its proof here.

The following result now follows from combining Theorem 3.1 and Proposition 4.1.

Corollary 4.1. *In the setting as Theorem 3.1 and Proposition 4.1, we have with probability $\frac{1}{2}$,*

$$h(J) \geq \frac{|V|}{(k+1)\ C^*}\ \lambda_1(\mathcal{L}_J) \geq \frac{|V|}{(k+1)\ C^*}\ (1-\epsilon)\ \lambda_1(\mathcal{L}_K).$$

Thus, the Cheeger constant of the sparsified simplicial complex, J, is bounded below by a multiplicative factor of the first nontrivial eigenvalue of the up Laplacian for the original complex, K.

5 Numerical Experiments

In Section 5.1, we conduct numerical experiments to illustrate the inequalities bounding the spectrum of the up Laplacian of the sparsified simplicial complex, proven in Theorem 3.1. In Section 5.2 we extend spectral clustering methods to simplicial complexes. We show that the clusters obtained for sparsified simplicial complexes are similar to those of the original simplicial complex. In both sections, we also present the analogous results for graph sparsification to serve as a comparison.

5.1 Preservation of the spectrum of the up Laplacian

Experimental set up. In the setting of graph sparsification [50], we recall that if a graph H is an ε-approximation of a graph G, n is the number of vertices in H and G, then, as in (2), we have the following inequality,

$$(1-\varepsilon)x^T L_G x \ \leq\ x^T L_H x \ \leq\ (1+\varepsilon)x^T L_G x, \qquad \forall x \in \mathbb{R}^n.$$

Subtracting $x^T L_G x$ from all terms in this inequality, we obtain

$$-\varepsilon x^T L_G x \ \leq\ x^T (L_H - L_G) x \ \leq\ \varepsilon x^T (L_G) x, \qquad \forall x \in \mathbb{R}^n.$$

Let $\lambda_{max}(L_G)$, $\lambda_{max}(L_H)$ and $\lambda_{max}(L_H - L_G)$ be the maximum eigenvalues of L_G and L_H and $L_H - L_G$ respectively. Also, let $\lambda_{min}(L_G)$ be the minimum eigenvalue of L_G. With some algebraic manipulations, we obtain on the right hand side,

$$\lambda_{max}(L_H - L_G) = \max_{||x||=1} x^T(L_H - L_G)x \leq \varepsilon \max_{||x||=1} x^T(L_G)x = \varepsilon\lambda_{max}(L_G).$$

Similarly, on the left hand side, we obtain

$$\begin{aligned} 0 = -\varepsilon\lambda_{min}(L_G) &= -\varepsilon \min_{||x||=1} x^T L_G x \\ &= \max_{||x||=1} -\varepsilon x^T L_G x \\ &\leq \max_{||x||=1} x^T(L_H - L_G)x = \lambda_{max}(L_H - L_G). \end{aligned}$$

Together we have the inequality

$$0 \;\leq\; \lambda_{max}(L_H - L_G) \;\leq\; \varepsilon\lambda_{max}(L_G). \tag{10}$$

We can obtain the analogous inequality in the setting of simplicial complex sparsification. Let J be a sparsified version of K following the setting of Theorem 3.1. Suppose for a fixed dimension i (where $1 \leq i \leq \dim K$), K and J have $(i-1)$-th up Laplacians $\mathcal{L}_K := \mathcal{L}_{K,i-1}$ and $\mathcal{L}_J := \mathcal{L}_{J,i-1}$ respectively, we have,

$$(1-\varepsilon)x^T\mathcal{L}_K x \leq x^T\mathcal{L}_J x \leq (1+\varepsilon)x^T\mathcal{L}_K x, \qquad \forall x \in \mathbb{R}^{n_{i-1}}. \tag{11}$$

A similar argument leads to the following inequality,

$$0 \;\leq\; \lambda_{max}(\mathcal{L}_J - \mathcal{L}_K) \;\leq\; \varepsilon\lambda_{max}(\mathcal{L}_K). \tag{12}$$

Notice that inequality (10) is a special case of the inequality (12).

Preservation of the spectrum of the sparsified graph Laplacian. To demonstrate how the spectrum of the graph Laplacian is preserved during graph sparsification, we set up the following experiment. Consider a complete graph G with $n_0 = 40$ vertices and $n_1 = 780$ edges. We run multiple sparsification processes on this graph G and study the convergence behavior based on the inequality in (2). For each sparsification process, we use a sequence of sample sizes, ranging between 10 and $2n_1$. For each sample size q, we set $\varepsilon = \sqrt{n_0 \log n_0 / q}$ by assuming that $9C^2 = 1$ in the hypothesis of Theorem 3.1. As q varies, we correspondingly obtain a sequence of varying ε values.

In particular, we run 25 simulations on G. For each simulation, we fix a unit vector x uniformly randomly sampled from $\mathbb{S}^{n_0}$, and perform 25 instances of experiments. For each instance, we apply our sparsification procedure to generate the convergence plot using the list of fixed sample sizes q and their corresponding ε's. Specifically, for each sample size, we obtain a sparse graph H and compute $x^T L_H x$ and $\lambda_{max}(L_H - L_G)$; and we observe the convergence behavior of these quantities as the sample size increases.

In Figure 1(a), we show the convergence behavior based on the inequality in (2). For a single simulation, we compute the point-wise average of $x^T L_H x$ across the 25 instances, and plot these values as function of the sample size q, which gives rise to a single convergence curve in aqua. Then we compute the point-wise average of the aqua curves across all simulations, producing the red curve. Since each simulation (for a fixed x) has a different upper bound curve $(1-\varepsilon)x^T L_G x$

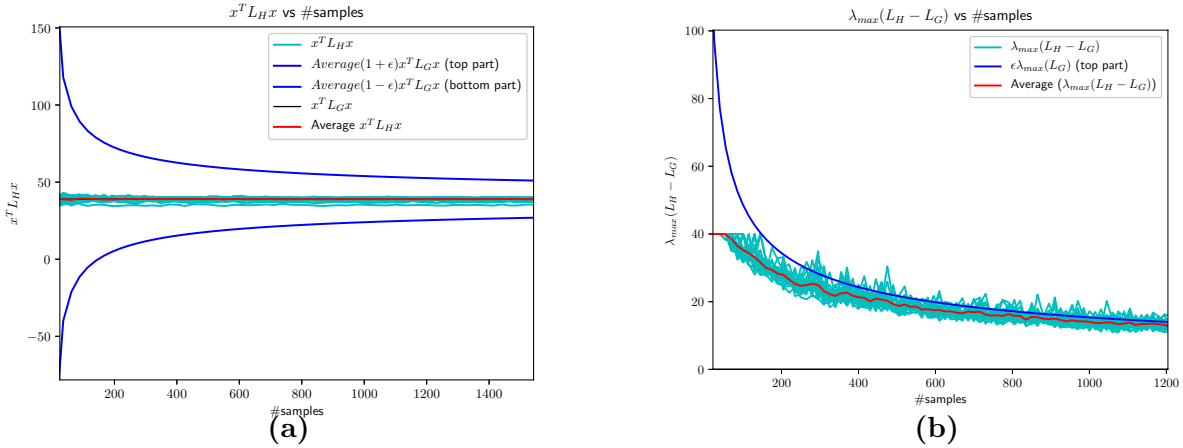


Figure 1: The results of a numerical experiment illustrating inequalities which control the spectrum of sparsified graph Laplaicans; see Section 5.1 for details. **(a)** For an ensemble of vectors, $x \in \mathbb{S}^{n_0}$, and sparsified graphs, H, we plot the terms in inequality (2). **(b)** For an ensemble of sparsified graphs, H, we plot the terms in the inequality (10).

and lower bound curve $(1+\varepsilon)x^T L_G x$ respectively (not shown here), The point-wise average of the upper and lower bound curves across all simulations is plotted in blue. We observe that on average, these curves respect the inequality (2), that is, the red curve is nested within its approximated theoretical upper and lower bounds in blue.

In Figure 1(b), we illustrate the theoretical upper and lower bounds for $\lambda_{max}(L_H - L_G)$ given in inequality (10) as the sample size q increases. In particular, we run a single simulation with 25 instances, computing $\lambda_{max}(L_H - L_G)$. Each instance gives us a convergence curve shown in aqua. We compare the point-wise average of $\lambda_{max}(L_H - L_G)$ (in red) with its (approximated) theoretical upper bound in blue and lower bound (*i.e.*, 0, the x-axis). On average, the experimental results respect the inequality (10). Figure 3(a) illustrates how the number of edges scale with the increasing number of samples across all instances.

Preservation of the spectrum of the up Laplacian for a sparsified simplicial complex. To demonstrate that the spectrum of the up Laplacian is preserved during the sparsification of a simplicial complex, we set up a similar experiment. We start with a 2-dimensional simplicial complex, K, that contains all edges and triangles on $n_0 = 40$ vertices (with $n_1 = 780$ edges and $n_2 = 9880$ faces.) and a sequence of fixed sample sizes q. For each sample size q, we solve for $\varepsilon = \sqrt{n_1 \log n_1 / q}$ assuming that $9C^2 = 1$ in the hypothesis of Theorem 3.1, to get the corresponding sequence of ε values. With the simplicial complex K and the sequence of sample sizes fixed, we run 25 simulations, each simulation consisting 25 instances and a fixed randomly sampled unit vector x as described previously; only this time, we sparsify the faces of the simplicial complex by applying Algorithm 1 with $i = 2$. In Figure 2, we plot the terms in inequalities describing the spectrum for these sparsified simplicial complexes.

In Figure 2(a), following the same procedure as for graph sparsification, we obtain a plot that respects the inequality (11). The curves in aqua show the point-wise averages of $x^T \mathcal{L}_J x$ across all instances in a single simulation, whereas the red curve represents point-wise average across all instances and all simulations. Since the random vector x is resampled for each simulation, the upper

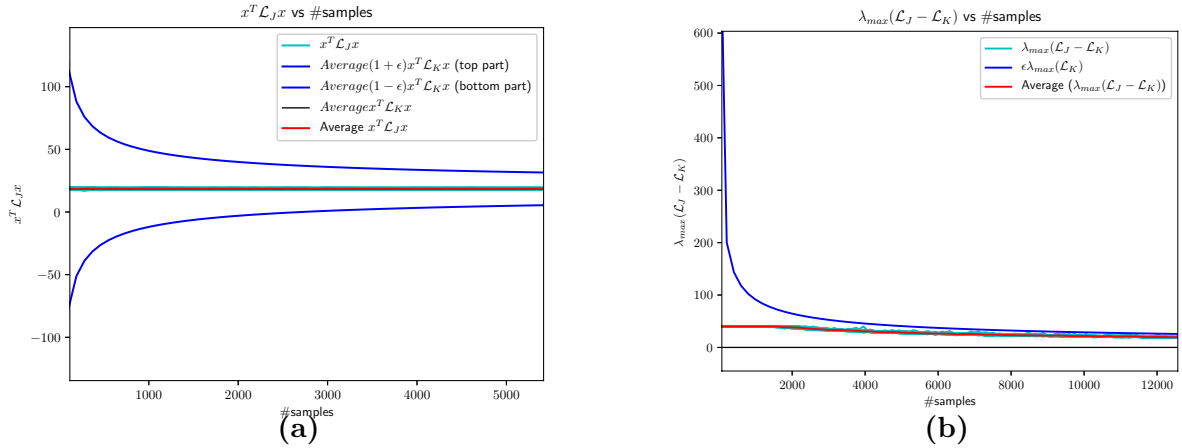


Figure 2: The results of a numerical experiment illustrating inequalities which control the spectrum of the up Laplacian for sparsified simplicial complexes; see Section 5.1 for details. **(a)** For an ensemble of vectors, $x \in \mathbb{S}^{n_1}$, and sparsified simplicial complexes, J, we plot the terms in inequality (11). **(b)** For an ensemble of sparsified simplicial complexes, J, we plot the terms in the inequality (12).

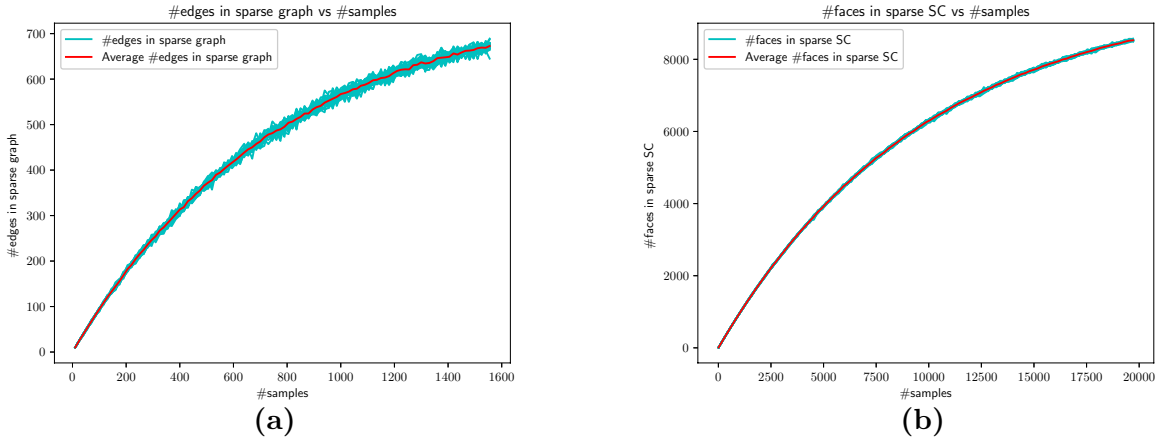


Figure 3: Figures illustrating how **(a)** the number of edges in the case of graph sparsification and **(b)** the number of faces/triangles in the case of simplicial complex sparsification vary with increasing sample size.

and lower bound curves are different for every simulation. In Figure 2(a) we plot their point-wise average across all simulations as the upper and lower bound curves in blue.

In Figure 2(b), to illustrate inequality (12), we run a single simulation with 25 instances. Each instance gives us a sequence of $\lambda_{max}(\mathcal{L}_J - \mathcal{L}_K)$ values as function of sample size. We plot them as curves in aqua. We compare the point-wise averages of $\lambda_{max}(\mathcal{L}_J - \mathcal{L}_K)$ (in red) with its (approximated) theoretical upper and lower bounds in blue. Figure 3(b) shows how the number of faces scales with increasing number of samples across all instances.

5.2 Spectral clustering

We demonstrate via numerical experiments, that preserving the structure of the up Laplacian via sparsification also preserves the results of spectral clustering on simplicial complexes. Spectral clustering can be considered as a class of algorithms with many variations. Here, we apply spectral clustering to simplicial complexes before and after sparsification, by extending a rather standard implementation for graphs.

Datasets. We consider a graph that contains two complete subgraphs with 20 vertices (and 190 edges) each, which are connected by $64 = 8 \times 8$ edges spanning across the two subgraphs. We refer to this graph, G, as the *dumbbell graph*; it has $n_0 = 40$ vertices and $n_1 = 444$ edges. All edge weights are set to be 1. To compute the sparsified graph, the number of samples, q, is set to be $0.5n_1$.

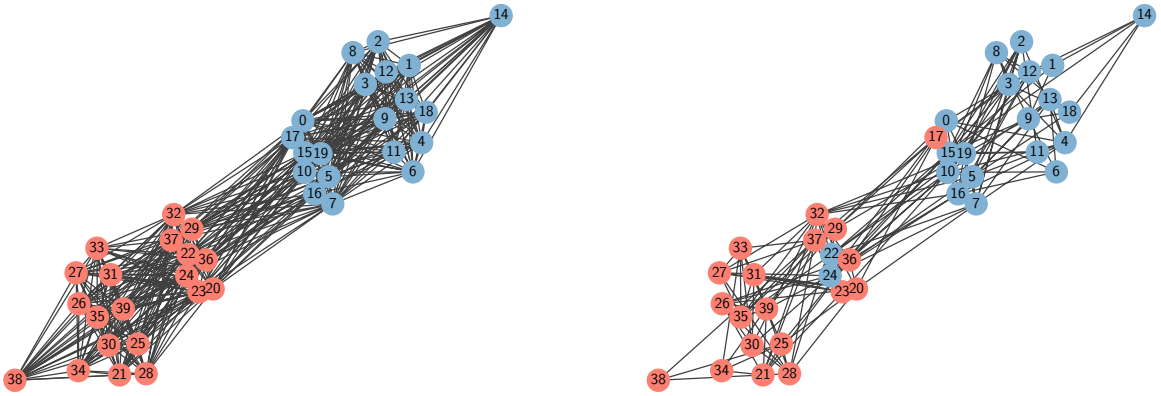
Similarly, we consider a simplicial complex that contains two complete sub-complexes with 10 vertices, 45 edges and 120 triangles each. The two sub-complexes are connected by 16 cross edges and 48 cross triangles so that the simplicial complex is made up of $n_0 = 20$ vertices, $n_1 = 106$ edges and $n_2 = 288$ triangles. We refer to this simplicial complex, K, as the *dumbbell complex*. The weights on all edges and triangles are set to be 1. To compute the sparsified simplicial complex, the number of samples, q, is set to be $0.75n_2$.

Spectral clustering algorithm for graphs. We use the Ng-Jordan-Weiss algorithm [39] to perform spectral clustering of graphs. Let n_0 be the number of vertices in a graph. The algorithm can be summarized:

1. Compute the *vertex-vertex adjacency matrix* $A \in \mathbb{R}^{n_0 \times n_0}$ such that $A_{ij} = 1$ if vertices v_i and v_j are connected by an edge, otherwise $A_{ij} = 0$.
2. Compute the *degree matrix*, $\Delta \in \mathbb{R}^{n_0 \times n_0}$, a diagonal matrix with diagonal elements Δ_{ii} being the number of edges incident to vertex v_i, that is, $\Delta_{ii} = \sum_j A_{ij}$.
3. Construct the matrix $M = \Delta^{-1/2} A \Delta^{-1/2}$.
4. Find $u_1, u_2, \cdots, u_k$, the eigenvectors of M corresponding to the k largest eigenvalues (chosen to be orthogonal to each other in the case of repeated eigenvalues), and form the matrix $X = [u_1 u_2 \cdots u_k] \in \mathbb{R}^{n_0 \times k}$ by stacking the eigenvectors in columns.
5. Form the matrix Y from X by re-normalizing each of X's rows to have unit length, that is, $Y_{ij} = X_{ij} / \left(\sum_j X_{ij}^2 \right)^{1/2}$.
6. Treating each row of Y as a point in $\mathbb{R}^k$, cluster them into k clusters via the k-means algorithm.
7. Finally, assign the original vertex v_i to cluster j if and only if row i of the matrix Y is assigned to cluster j.

Recall the graph Laplacian can be written as $L = \Delta - A$. Furthermore $M = I - L_N$, where $L_N = \Delta^{-1/2} L \Delta^{-1/2}$ is referred to as the normalized graph Laplacian.

To demonstrate the utility of the sparsification, we illustrate the spectral clustering results before and after graph sparsification in Figure 4. Since graph sparsification preserves the spectral properties of graph Laplacian, we expect it to also preserve (to some extent) the results of spectral methods, such as spectral clustering.



(a) **(b)**

Figure 4: Spectral clustering of graphs before **(a)** and after **(b)** sparsification. We observe that the clusters are very similar. See Section 5.2 for details.

Spectral clustering algorithm for simplicial complex. We seek to extend the Ng-Jordan-Weiss algorithm [39] to simplicial complexes, which, as far as we are aware, has not yet been studied. We seek the simplest generalization by replacing the vertex-vertex adjacency matrix with an edge-edge adjacency matrix A, where two edges are considered to be adjacent if they are faces of the same triangle. This definition is a straightforward extension of the adjacency among vertices in graphs, however it does not account for the orientation of edges or triangles.

Formally, let n_1 be the number of edges. We define the *edge-edge adjacency matrix* $A \in \mathbb{R}^{n_1 \times n_1}$, where

$$A_{i,j} = \begin{cases} 1 & \text{if edges } e_i \text{ and } e_j \text{ are faces of the same triangle} \\ 0 & \text{otherwise} \end{cases}.$$

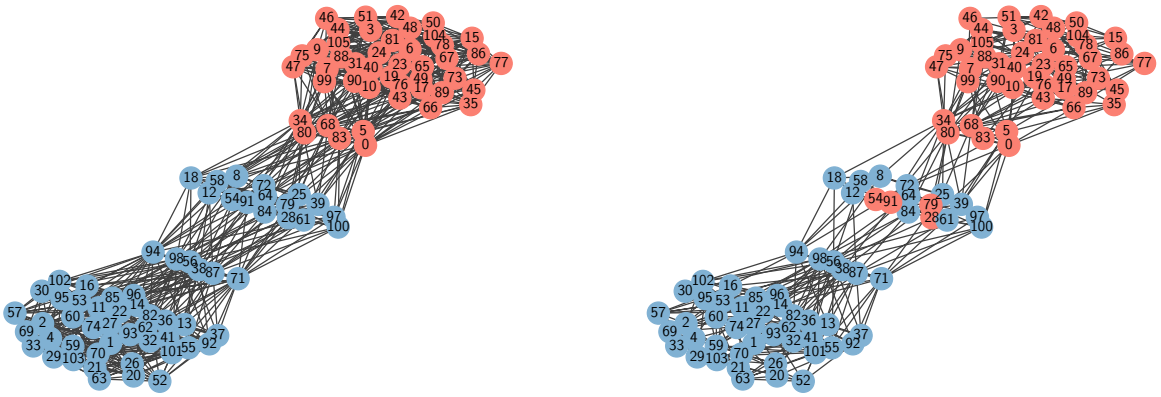
We define the *degree matrix* $\Delta \in \mathbb{R}^{n_1 \times n_1}$ to be the diagonal matrix with element $\Delta_{i,i}$ is the number of triangles incident to edge e_i. With A and Δ defined this way, we can apply the Ng-Jordan-Weiss algorithm to cluster the edges of the simplicial complex K.

This is the same as applying spectral clustering to the *dual graph* of K. A *dual graph* G of a given simplicial complex K is created as follows: each edge in K becomes a vertex in the dual graph G, and there is an edge between two vertices in G if their corresponding edges in K share the same triangle. We then apply spectral clustering to the dual graph G as usual and obtain the resulting clustering of vertices in G (which correspond to the clustering of edges in K). To better illustrate our edge clustering results, we visualize the resulting clusters based upon the dual graph. The results are plotted in Figure 5 for two clusters and Figure 6 for three clusters. Applying the spectral algorithm with this new adjacency definition results in clusters that agree reasonably well before and after sparsification.

The adjacency matrix, A, does not take into consideration the orientation of the edges, so the above clustering algorithm effectively relies on computing the eigenvectors of a normalized up Laplacian. To wit, one can verify that the dimension 1 up Laplacian can be written as

$$\mathcal{L}_{K,1} = \Delta - A^*,$$

where Δ is the diagonal degree matrix defined previously and the *oriented edge-edge adjacency*



(a) **(b)**

Figure 5: Spectral clustering of simplicial complexes into two clusters, before **(a)** and after **(b)** sparsification. See Section 5.2 for details.

matrix, $A^* \in \mathbb{R}^{n_1 \times n_1}$, is given by

$$A^*_{i,j} = \begin{cases} -1 & \text{edges } e_i \text{ and } e_j \text{ are adjacent and either both agree or disagree with the} \\ & \text{orientation of their shared triangle} \\ 1 & \text{if either } e_i \text{ or } e_j \text{ (but not both) agree with the orientation of their shared triangle} \\ 0 & \text{if } e_i \text{ and } e_j \text{ are not adjacent} \end{cases}.$$

It follows that the edge-edge adjacency matrix $A = |A^*|$ where the absolute value operation is applied element-wise. The relation between the normalized up Laplacian $\Delta - A$ and the up Laplacian, $\mathcal{L}_{K,1}$, we used for sparsification, remains unclear.

6 Discussion

The results in this paper constitute a first step towards spectral sparsification of simplicial complexes. Our work is strongly motivated by the study of an emerging class of learning algorithms based on simplicial complexes and, in particular, those spectral algorithms that operate with higher-order Laplacians. We would like to understand how such learning algorithms can be applicable to more compact representations of the data. Several on-going and future directions are described below.

Label prorogation on simplicial complexes. A very good example of spectral methods in learning arises from extending label propagation algorithms on graphs to simplicial complexes, in particular, the work by Mukherjee and Steenbergen [37]. Specifically, they adapt the label propagation algorithm to higher dimensional walks on oriented edges, and give visual examples of applying label propagation with the 1-dimensional up Laplacian $\mathcal{L}_1^{up}$, down Laplacian $\mathcal{L}_1^{down}$, and Laplacian $\mathcal{L}_1$. We could envision label propagation to be generalized to random walks on even higher-dimensional simplexes, such as triangles. A direct application of our work is to sparsify the top-dimensional simplexes (e.g. triangles in a 2-dimensional simplicial complex) and examine how label propagation behaves on these top-dimensional simplexes of the sparsified representation.

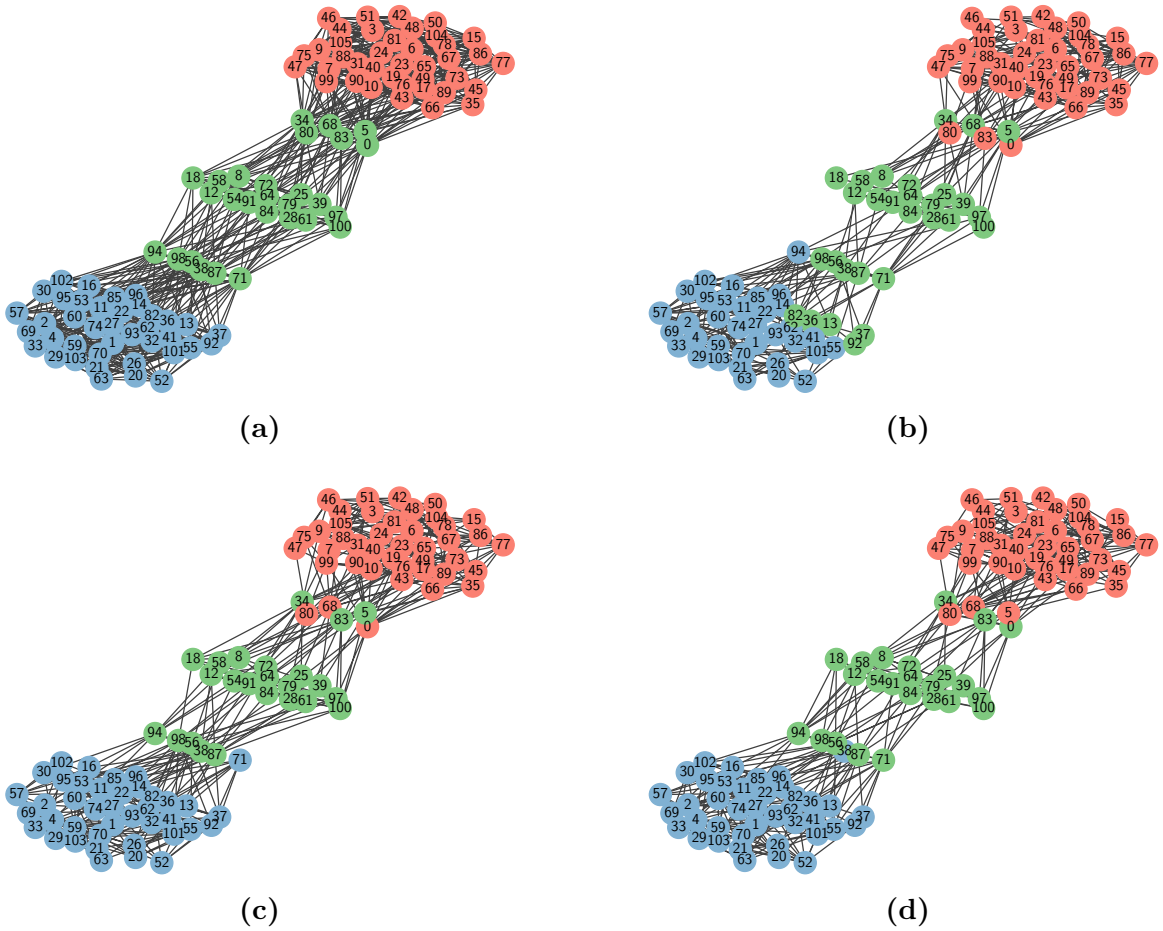


Figure 6: Spectral clustering of simplicial complexes into three clusters, before **(a)** and after **(b)-(d),** sparsification. See Section 5.2 for details.

Physical meaning of generalized effective resistance. We believe the generalization of effective resistance to simplicial complexes has interesting implications to our understanding of spectral clustering of simplicial complexes. Such a generalization is algebraically straightforward, but does not admit an obvious physical interpretation. As part of the future work, we will seek an interpretation in terms of a random process, such as an effective commute time as in the case of a graph [22]. This could also be related to properties of minimum spanning objects in the simplicial complex.

Multilevel and Hodge sparsification. We are also interested in performing multilevel sparsification of simplicial complexes; for example, we would like to sparsify triangles and edges simultaneously while preserving spectral properties of the dimension 0 and dimension 1 up Laplacians. This is very challenging if we would like to maintain structures of simplicial complexes; it may be possible if we could relax our structural constraints to work with hyper-graphs instead. In addition, multilevel sparsification is also related to preserving the spectral properties of the (Hodge) Laplacian. Finally, we are also interested in deriving formal connections between homological sparsification and spectral sparsification of simplicial complexes.

Acknowledgements

This work was partially supported by NSF DMS-1461138, NSF IIS-1513616, and the University of Utah Seed Grant 10041533. We would like to thank Todd H. Reeb for contributing to early discussions regarding this line of work.

References

[1] R. Andersen, F. Chung, and K. Lang. Local graph partitioning using pagerank vectors. *Proceedings 47th Annual IEEE Symposium on Foundations of Computer Science*, 2006.

[2] R. Andersen and K. J. Lang. Communities from seed sets. *Proceedings 15th International Conference on the World Wide Web*, pages 223–232, 2006.

[3] A. A. Benczúr and D. R. Karger. Approximating s-t minimum cuts in $\tilde{O}(n^2)$ time. In *Proceedings ACM Symposium on Theory of Computing*, pages 47–55. ACM, 1996.

[4] P. Bendich, E. Gasparovic, J. Harer, R. Izmailov, and L. Ness. Multi-scale local shape analysis for feature selection in machine learning applications. *IJCNN*, 2015.

[5] A. R. Benson, D. F. Gleich, and J. Leskovec. Tensor spectral clustering for partitioning higher-order network structures. In *Proceedings of the 2015 SIAM ICDM*, pages 118–126, 2015.

[6] M. B. Botnan and G. Spreemann. Approximating persistent homology in euclidean space through collapses. *Applicable Algebra in Engineering, Communication and Computing*, pages 1–29, 2015.

[7] M. Buchet, F. Chazal, S. Y. Oudot, and D. R. Sheehy. Efficient and robust persistent homology for measures. *ACM-SIAM Symposium on Discrete Algorithms*, pages 168–180, 2015.

[8] G. Carlsson. Topology and data. *Bull. Amer. Math. Soc.*, 46(2):255–308, 2009.

[9] N. J. Cavanna, M. Jahanseir, and D. R. Sheehy. A geometric perspective on sparse filtrations. In *Proceedings Canadian Conference on Computational Geometry*, 2015.

[10] A. K. Chandra, P. Raghavan, W. L. Ruzzo, R. Smolensky, and P. Tiwari. The electrical resistance of a graph captures its commute and cover times. *Computational Complexity*, 6:312–340, 1996.

[11] A. Choudhary, M. Kerber, and S. Raghvendra. Polynomial-sized topological approximations using the permutahedron. arXiv:1601.02732, 2016.

[12] F. R. K. Chung. *Spectral Graph Theory*, volume 92. American Mathematical Society, 1997.

[13] V. de Silva and R. Ghrist. Coverage in sensor networks via persistent homology. *Algebraic and Geometric Topology*, 7:339–358, 2007.

[14] T. K. Dey, F. Fan, and Y. Wang. Computing topological persistence for simplicial maps. *Proceedings 30th Annual Symposium on Computational Geometry*, pages 345–354, 2014.

[15] T. K. Dey, F. Fan, and Y. Wang. Graph induced complex on point data. *Computational Geometry*, 48(8):575–588, 2015.

[16] R. Diestel. *Graph Theory*. Springer Graduate Texts in Mathematics, 2000.

[17] D. Dotterrer and M. Kahle. Coboundary expanders. *Journal of Topology and Analysis*, 4:499–514, 2012.

[18] P. G. Doyle and J. L. Snell. *Random walks and electric networks*. Mathematical Association of America, 1984.

[19] M. Drton, B. Sturmfels, and S. Sullivant. *Lectures on algebraic statistics*, volume 39. Springer Science & Business Media, 2008.

[20] H. Edelsbrunner and J. Harer. Persistent homology-a survey. *Contemporary mathematics*, 453:257–282, 2008.

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

[22] A. Ghosh, S. Boyd, and A. Saberi. Minimizing effective resistance of a graph. *SIAM Review*, 50(1):37–66, 2008.

[23] R. Ghrist. Barcodes: the persistent topology of data. *Bull. Amer. Math. Soc.*, 45(1):61–75, 2008.

[24] R. Ghrist. *Elementary Applied Topology*. Createspace, 2014.

[25] D. F. Gleich. Pagerank beyond the web. *SIAM Review*, 57(3):321–363, 2015.

[26] D. F. Gleich, L.-H. Lim, and Y. Yu. Multilinear pagerank. *SIAM Journal on Matrix Analysis and Applications*, 36(4):1507–1541, 2015.

[27] L. J. Grady and J. Polimeni. *Discrete Calculus: Applied Analysis on Graphs for Computational Science*. Springer, 2010.

[28] A. Gundert and M. Szedlák. Higher dimensional discrete Cheeger inequalities. *Journal of Computational Geometry*, 6(2):54–71, 2015.

[29] D. Horak and J. Jost. Spectra of combinatorial Laplace operators on simplicial complexes. *Advances in Mathematics*, 244:303–336, 2013.

[30] M. S. T. Jaakkola and M. Szummer. Partially labeled classification with markov random walks. *Advances in neural information processing systems (NIPS)*, 14:945–952, 2002.

[31] X. Jiang, L.-H. Lim, Y. Yao, and Y. Ye. Statistical ranking and combinatorial hodge theory. *Mathematical Programming*, 127(1):203–244, 2011.

[32] J. A. Kelner, L. Orecchia, A. Sidford, and Z. A. Zhu. A simple, combinatorial algorithm for solving sdd systems in nearly-linear time. *Proceedings 45th annual ACM symposium on Theory of computing*, pages 911–920, 2013.

[33] M. Kerber and R. Sharathkumar. Approximate Cěch complexes in low and high dimensions. *Proceedings 24th International Symposium on Algorithms and Computation, LNCS*, 8283:666–676, 2013.

[34] A. B. Lee, K. S. Pedersen, and D. Mumford. The nonlinear statistics of high-contrast patches in natural images. *International Journal of Computer Vision*, 54(1-3):83–103, 2003.

[35] A. Lubotzky. Ramanujan complexes and high dimensional expanders. *Japanese Journal of Mathematics*, 9:137–169, 2014.

[36] M. W. Mahoney, L. Orecchia, and N. K. Vishnoi. A local spectral method for graphs: With applications to improving graph partitions and exploring data graphs locally. *Journal of Machine Learning Research*, 13:2339–2365, 2012.

[37] S. Mukherjee and J. Steenbergen. Random walks on simplicial complexes and harmonics. *Random Structures & Algorithms*, 2016.

[38] J. R. Munkres. *Elements of algebraic topology*. Addison-Wesley, Redwood City, CA, USA, 1984.

[39] A. Y. Ng, M. I. Jordan, and Y. Weiss. On spectral clustering: Analysis and an algorithm. *Advances In Neural Information Processing Systems*, 2001.

[40] M. Nicolau, A. J. Levine, and G. Carlsson. Topology based data analysis identifies a subgroup of breast cancers with a unique mutational profile and excellent survival. *Proceedings of the National Academy of Sciences*, 108(17):7265–7270, 2011.

[41] B. Osting, C. Brune, and S. J. Osher. Optimal data collection for informative rankings expose well-connected graphs. *Journal of Machine Learning Research*, 15:2981–3012, 2014.

[42] B. Osting, C. D. White, and E. Oudet. Minimal Dirichlet energy partitions for graphs. *SIAM J. Scientific Computing*, 36(4):A1635–A1651, 2014.

[43] B. Osting, Y. Yao, J. Xiong, and Q. Xu. Analysis of crowdsourced sampling strategies for hodgerank with sparse random graphs. *Applied and Computational Harmonic Analysis*, 41(2):540–560, 2016.

[44] O. Parzanchevski and R. Rosenthal. Simplicial complexes: Spectrum, homology and random walks. *Random Structures & Algorithms*, 2016.

[45] O. Parzanchevski, R. Rosenthal, and R. J. Tessler. Isoperimetric inequalities in simplicial complexes. *Combinatorica*, pages 1–33, 2015.

[46] J. A. Perea and J. Harer. Sliding windows and persistence: An application of topological methods to signal analysis. *Foundations of Computational Mathematics*, 15(3):799–838, 2015.

[47] W. Ren, Q. Zhao, R. Ramanathan, J. Gao, A. Swami, A. Bar-Noy, M. P. Johnson, and P. Basu. Broadcasting in multi-radio multi-channel wireless networks using simplicial complexes. *Wireless networks*, 19(6):1121–1133, 2013.

[48] M. Rudelson and R. Vershynin. Sampling from large matrices: An approach through geometric functional analysis. *Journal of the ACM*, 54(4):21, 2007.

[49] D. Sheehy. Linear-size approximations to the Vietoris-Rips filtration. *Discrete & Computational Geometry*, 49(4):778–796, 2013.

[50] D. A. Spielman and N. Srivastava. Graph sparsification by effective resistances. *SIAM Journal on Computing*, 40(6):1913–1926, 2011.

[51] D. A. Spielman and S.-H. Teng. Spectral sparsification of graphs. *SIAM Journal on Computing*, 40(4):981–1025, 2011.

[52] D. A. Spielman and S.-H. Teng. A local clustering algorithm for massive graphs and its application to nearly-linear time graph partitioning. *SIAM Journal on Computing*, 42(1):1–26, 2013.

[53] D. A. Spielman and S.-H. Teng. Nearly linear time algorithms for preconditioning and solving symmetric, diagonally dominant linear systems. *SIAM Journal on Matrix Analysis and Applications*, 35(3):835–885, 2014.

[54] J. Steenbergen, C. Klivans, and S. Mukherjee. A Cheeger-type inequality on simplicial complexes. *Advances in Applied Mathematics*, 56:56–77, 2014.

[55] A. Tausz and G. Carlsson. Applications of zigzag persistence to topological data analysis. *arxiv:1108.3545*, 2011.

[56] Y. van Gennip, N. Guillen, B. Osting, and A. Bertozzi. Mean curvature, threshold dynamics, and phase field theory on finite graphs. *Milan Journal of Mathematics*, 82(1):3–65, 2014.

[57] U. von Luxburg. A tutorial on spectral clustering. *Statistics and Computing*, 17(4):395–416, 2007.

[58] B. Wang, B. Summa, V. Pascucci, and M. Vejdemo-Johansson. Branching and circular features in high dimensional data. *IEEE Transactions on Visualization and Computer Graphics (TVCG)*, 17(12):1902–1911, 2011.

[59] E. Wong, S. Palande, B. Wang, B. Zielinski, J. Anderson, and P. T. Fletcher. Kernel partial least squares regression for relating functional brain network topology to clinical measures of behavior. *International Symposium on Biomedical Imaging (ISBI)*, 2016.

[60] X. Zhu, Z. Ghahramani, and J. Lafferty. Semi-supervised learning using gaussian fields and harmonic functions. *Proceedings 20th International Conference on Machine Learning*, pages 912–919, 2003.