SIAM Review
Vol. 62, No. 2, pp. 353–391

Random Walks on Simplicial Complexes and the Normalized Hodge 1-Laplacian*

Michael T. Schaub†
Austin R. Benson‡
Paul Horn§
Gabor Lippner¶
Ali Jadbabaie‖

Abstract. Using graphs to model pairwise relationships between entities is a ubiquitous framework for studying complex systems and data. Simplicial complexes extend this dyadic model of graphs to polyadic relationships and have emerged as a model for multinode relationships occurring in many complex systems. For instance, biological interactions occur between sets of molecules and communication systems include group messages that are not pairwise interactions. While Laplacian dynamics have been intensely studied for graphs, corresponding notions of Laplacian dynamics beyond the node-space have so far remained largely unexplored for simplicial complexes. In particular, diffusion processes such as random walks and their relationship to the graph Laplacian—which underpin many methods of network analysis, including centrality measures, community detection, and contagion models—lack a proper correspondence for general simplicial complexes.

Focusing on coupling between edges, we generalize the relationship between the normalized graph Laplacian and random walks on graphs by devising an appropriate normalization for the Hodge Laplacian—the generalization of the graph Laplacian for simplicial complexes—and relate this to a random walk on edges. Importantly, these random walks are intimately connected to the topology of the simplicial complex, just as random walks on graphs are related to the topology of the graph. This serves as a foundational step toward incorporating Laplacian-based analytics for higher-order interactions. We demonstrate how to use these dynamics for data analytics that extract information about the edge-space of a simplicial complex that complements and extends graph-based analysis. Specifically, we use our normalized Hodge Laplacian to derive spectral embeddings for examining trajectory data of ocean drifters near Madagascar and also develop a generalization of personalized PageRank for the edge-space of simplicial complexes to analyze a book copurchasing dataset.

*Received by the editors July 17, 2018; accepted for publication (in revised form) August 19, 2019; published electronically May 7, 2020.

https://doi.org/10.1137/18M1201019

Funding: The work of the first author was supported by the European Union's Horizon 2020 Research and Innovation Programme under Marie Sklodowska-Curie grant agreement 702410. The work of the second author was partially supported by NSF award DMS-1830274. The work of the third, fourth, and fifth authors was supported by AFOSR grant FA9550-13-1-0097. The work of the fifth author was partially supported by a Vannevar Bush Faculty Fellowship from the Office of the Secretary of Defense. The work of the second and fifth authors was partially supported by ARO award W911NF-19-1-0057.

†Institute for Data, Systems and Society, Massachusetts Institute of Technology, Cambridge, MA 02139, and Department of Engineering Science, University of Oxford, Oxford, OX1 3PJ, United Kingdom (michael.schaub@eng.ox.ac.uk).

‡Department of Computer Science, Cornell University, Ithaca, NY 14853 (arb@cs.cornell.edu).

§Department of Mathematics, University of Denver, Denver, CO 80208 (paul.horn@du.edu).

¶Department of Mathematics, Northeastern University, Boston, MA 02115 (g.lippner@neu.edu).

‖Institute for Data, Systems and Society, Massachusetts Institute of Technology, Cambridge, MA 02139 (jadbabai@mit.edu).

Key words. graph theory, networks, diffusion processes, random walks, simplicial complexes, Hodge Laplacian, Hodge decomposition, spectral embedding, PageRank

AMS subject classifications. 05C82, 05C, 91D30, 55U10, 62-07

DOI. 10.1137/18M1201019

Contents

1. Introduction. Markov chains and diffusion are staples of applied mathematics [19, 26, 70, 87]. Any time-homogeneous finite state Markov chain can be interpreted as a random walk on a graph: the states of the Markov chain are the nodes of the graph, and transitions occur between connected nodes via an appropriately weighted edge. This close relationship between Markov chains and graphs has led to a broad adoption of diffusion-based algorithms in network science [87], with applications to ranking connected objects [59], analyzing disease spreading [111], and respondent-driven sampling [112]. Key to the success of many such algorithms is the link between random walks and (the spectral theory of) the graph Laplacian [37, 38, 87], a matrix that encodes the structure of a graph and has intimate connections to discrete potential theory and harmonic analysis [18]. Indeed, there is well-developed theory relating topological properties of graphs to features of the graph Laplacian and thus to random walks and diffusion processes [18, 37, 38]. For instance, spectral properties of the graph Laplacian are related to expansion, diameter, distance between subsets, and the mixing time of random walks, among others [37, 38]. Thus, analyzing the properties of a random walk on a network (i.e., graph), or alternatively the graph Laplacian, can reveal fundamental properties about the system under investigation.

As network-based system models have become almost ubiquitous across scientific domains [20, 93], graphs and their Laplacians feature prominently in many analysis tasks [9, 12, 40, 64, 87, 89]. However, graphs are in fact special cases of more general mathematical objects called simplicial complexes (SCs),[1] and the graph Laplacian is a special case of the Hodge Laplacian that appears in algebraic topology [85]. We give a formal definition of SCs later, but an intuitive description of an SC is a generalization of a graph whose edge-set can contain subsets of any number of nodes (more specifically, a hypergraph with certain properties). Since SCs can describe richer sets of relationships than a graph, SCs are increasingly used to analyze systems and data (see subsection 1.1).

In this paper, we introduce a certain normalized Hodge Laplacian matrix and show how it relates to random walks on SCs, with an overarching goal of developing data analyses that respect additional aspects of the topology of the data. A primary motivation for this study of diffusion on SCs is to facilitate the translation of the large toolbox of network science for graphs to SCs. To make our results concrete, we present our work in the language of linear algebra and focus on devising random walks on "1-simplices," which may be thought of as edges. Stated differently, our results provide tools for the analysis of signals defined in the *edge-space* of an SC (or graph) that complement the typical node-based analysis conducted for graphs. In particular, we present tools that enable us to extract the relative importance of edges and edge-signals with respect to the higher-order topological properties of the SC. We contrast our methodology with other notions of edge-based random walks such as those based on line graphs [1, 47] and consider how higher-order interactions lead to certain difficulties in formulating a diffusion model that are absent in the theory of random walks on graphs [91, 108].

We describe two applications to illustrate our ideas, which show how our methodology incorporates higher-order topology of the data into the analysis.

In our first application, we develop embeddings of edge flows and trajectory data as a higher-order generalization of diffusion maps [40] and Laplacian eigenmaps [9]. Similar to the embedding of the nodes of a graph into a Euclidean space, this embedding provides us with an effective low-dimensional representation of edges and trajectories

[1] Formally, we use *abstract simplicial complexes*, but we drop "abstract" for easier reading.

in a vector space. This vector space representation can then be used to gain further insights into the observed flows. Here we illustrate how the embedding can be used to differentiate types of trajectories, but other data analysis and machine learning tasks may be addressed with these techniques as well.

Our second application is a variant of (personalized) PageRank [59] for 1-simplices (edges in a graph). Here we demonstrate how our tools enable us to analyze the "role" certain edges play with respect to the global topology of the SC. We point out how these tools may be seen as extensions of ideas from graph signal processing, typically concerned with signals on nodes, to the space of signals defined on edges. Indeed, we demonstrate how our analysis is complementary to node-based analysis and how our tools can highlight, e.g., how much an edge is part of the cycle-space of the SC.

1.1. Additional Background and Motivation.

Networks as Models for Complex Systems. Many complex systems are naturally modeled by graphs [93]. Due to the broad scope of this modeling paradigm, the analysis of systems as networks by means of graph-theoretic tools has been tremendously successful, with applications in biological, technological, and social systems [2, 20, 93, 95, 120]. However, graph-based representations of complex systems have some limitations. Specifically, graphs encode pairwise relationships between entities via edges, but do not account for (simultaneous) interactions between more than two nodes, which can be crucial for the ensuing dynamics of a system [12, 63]. For instance, biochemical reactions often involve more than two species in a reaction [77]; sensors record collections of interactions at a given time [42, 56, 121]; people communicate in small groups with group chat messaging; and individuals form teams or exert peer pressure [25, 75]. Nondyadic interactions have in fact long been an object of study in the social sciences. For instance, structural balance theory implies that three-way relationships in social networks will evolve according to colloquial rules such as "the friend of a friend is my friend" and "the enemy of my enemy is my friend" [30, 86, 124].

Modeling Higher-Order Interactions. There are several modeling frameworks for nondyadic, higher-order interactions, such as SCs [66, 85], set systems [14], hypergraphs [22], and affiliation graphs [48]. Here we focus on SCs, which, in contrast to generic hypergraphs, have special algebraic structure that makes them a core object of study in applied (algebraic) topology [29, 55]. Such algebraic structure is accompanied by analogues of the graph Laplacian for SCs, namely, the Hodge Laplacian, which will be a principal object of our study.

Related Work. Simplicial complex models have been successful in developing new insights into biological data [32, 92], structural and functional connectivity in the brain [57, 58, 104], coverage of sensor networks [42, 56, 90, 121], signal processing [8, 115], mobility analysis [54], and robotics [105]. Simplicial models have further been studied from a geometric perspective [17], in terms of epidemic spreading [69], or in the context of extensions of random graph models [41, 73, 125]. Nevertheless, in contrast to graph-based methods, the analysis of higher-order interaction data using SCs is still nascent, even though the formal use of tools from algebraic topology for the analysis of networks had already been discussed in the 1950s in the context of electrical network and circuit theory [107, 109, 110]. Also, Eckmann's seminal work introduced the ideas underpinning the Hodge Laplacian in 1944 [45]. That being said, little is known about the spectral properties of Hodge Laplacians and how they relate to dynamics on the underlying SCs. Specifically, notions such as random walks and diffusion processes on

SCs have remained barely explored and even then mainly from the perspective of pure mathematics [68, 91, 100, 108]. Preliminary research elucidating spectral connections includes spectral sparsification of SCs [97], embeddedness of edges in the cycle-space of a graph [114], the analysis of flows on graphs and discretized domains [54, 115], and the spectral theory of hypermatrices, tensors, and hypergraphs [11, 12, 33, 53, 106, 126].

1.2. Outline and Notation. We first briefly review SCs and Hodge Laplacians in section 2. Then, in section 3, we discuss our normalized variant of the Hodge Laplacian and how it can be related to models for diffusion processes on SCs in the edge-space, and we analyze its spectral properties. Section 4 describes how SCs can be constructed from data and discusses computational aspects of our formalism. Finally, sections 5 and 6 outline trajectory embeddings and simplicial PageRank as two applications of our random walk model on SCs.

Notation. Matrices and vectors are denoted by bold-faced fonts ($\boldsymbol{A}, \boldsymbol{x}$). Their entries will be denoted by indexed letters (such as $A_{i,j}, x_i$), or for clarity as $(\boldsymbol{A})_{i,j}$. All vectors are assumed to be column vectors. Scalar quantities are denoted by lowercase letters such as a, b. We use $\mathbf{1}$ to denote the vector of all ones, and $\boldsymbol{I}$ to denote the identity matrix. We use $|\boldsymbol{A}|$ to denote the matrix whose elements are given by the absolute values of the entries of $\boldsymbol{A}$. We further denote the positive and negative parts of a real-valued matrix by $\boldsymbol{A}^+ := \max(\boldsymbol{A}, \mathbf{0})$ and $\boldsymbol{A}^- := \max(-\boldsymbol{A}, \mathbf{0})$, where the maximum is applied elementwise. For a vector $\boldsymbol{x}$, the matrix $\operatorname{diag}(\boldsymbol{x})$ denotes the diagonal matrix whose (diagonal) entries are given by the components of the vector $\boldsymbol{x}$. For square matrices $\boldsymbol{A}, \boldsymbol{B}$, we use $\operatorname{diag}(\boldsymbol{A}, \boldsymbol{B})$ to denote a block-diagonal matrix in which the matrices $\boldsymbol{A}, \boldsymbol{B}$ form the diagonal blocks (and the remaining entries are zero). Sets are denoted by calligraphic letters such as $\mathcal{S}$, except for the real numbers, which we denote by $\mathbb{R}$.

2. A Short Review of Graphs, Simplicial Complexes, and Laplacians. We briefly review some ideas from graph theory and algebraic topology. Our exposition is geared toward readers with an understanding of graphs and matrices and is similar to the more detailed exposition by Lim [85].

2.1. Graphs and the Graph Laplacian. An undirected graph $\mathcal{G}$ consists of a set of vertices $\mathcal{V}$ with cardinality $|\mathcal{V}| = n_0$ and a set of edges $\mathcal{E}$, where each edge is an unordered pair of nodes. For convenience, we identify the nodes with the integers $1, \ldots, n_0$. The structure of a graph can be encoded in an adjacency matrix $\boldsymbol{A}$ with entries $A_{i,j} = 1$ if i is connected to j via an edge, and $A_{i,j} = 0$ otherwise. As we consider undirected graphs here, $\boldsymbol{A} = \boldsymbol{A}^\top$. A connected component is a maximal set of nodes $\mathcal{V}_c$ such that there exists a sequence of edges in the graph via which every node in $\mathcal{V}_c$ can be reached from every other node in $\mathcal{V}_c$. The degree of a node i is the number of edges containing i. Accordingly, we can define the matrix of degrees as $\boldsymbol{D} := \operatorname{diag}(\boldsymbol{A}\mathbf{1})$, where $\operatorname{diag}(\boldsymbol{x})$ is the diagonal matrix with the entries of $\boldsymbol{x}$ on the diagonal. The graph Laplacian matrix $\boldsymbol{L}_0 = \boldsymbol{D} - \boldsymbol{A}$ is an algebraic description of a graph, whose spectral properties reveal a number of important topological properties of the graph [23, 37, 38, 89, 118].

2.2. Simplicial Complexes. Let $\mathcal{V}$ be a finite set of vertices. A *k-simplex* $\mathcal{S}^k$ is a subset of $\mathcal{V}$ of cardinality $k+1$ (we do not allow $\mathcal{S}^k$ to be a multiset, i.e., there are no repeated elements in $\mathcal{S}^k$). An SC $\mathcal{X}$ is a set of simplices with the property that if $\mathcal{S} \in \mathcal{X}$, then all subsets of $\mathcal{S}$ are also in $\mathcal{X}$.

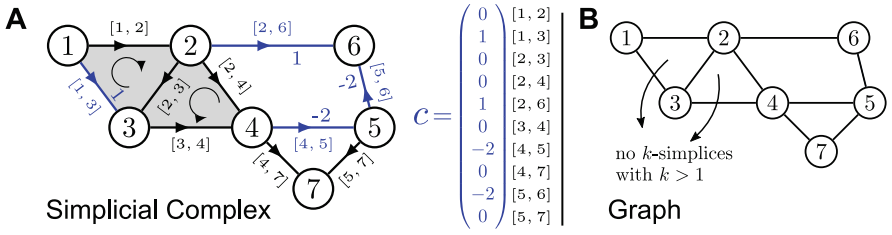

Fig. 2.1 Simplicial complexes and graphs. (A) *Schematic of an SC with a prescribed orientation. This is the running example of an SC in the text. Shaded areas correspond to the* 2*-simplices* $\{1,2,3\}$ *and* $\{2,3,4\}$*. An edge flow* c *along the paths* $2 \to 6 \to 5 \to 4$ *and* $1 \to 3$ *as well as its corresponding vector representation are depicted in blue.* (B) *Schematic of a graph, corresponding to the* 1*-skeleton of the SC in* (A)*. There are no* k*-simplices with* $k > 1$ *in the graph.*

EXAMPLE 1. *Analogous to a graph, the vertices of* $\mathcal{X}$ *in Figure* 2.1A *correspond to "nodes"* $\{1,\dots,7\}$ *and the* 1*-simplices to "edges." The* 2*-simplices* $\{1,2,3\}$ *and* $\{2,3,4\}$ *are depicted by filled triangles.*

A graph, while typically defined via two sets (vertices and edges), may be interpreted as an SC in which all simplices have cardinality at most 2 (Figure 2.1B). An SC can thus be understood as a generalization of a graph encoding higher-order relationships between vertices. To emphasize this connection, we will call the collection of 1-simplices in an SC $\mathcal{X}$ the edges of $\mathcal{X}$.

A *face* of a simplex $\mathcal{S}^k$ is a subset of $\mathcal{S}^k$ with cardinality k, i.e., with one element of $\mathcal{S}^k$ omitted. If $\mathcal{S}^{k-1}$ is a face of simplex $\mathcal{S}^k$, $\mathcal{S}^k$ is called a *coface* of $\mathcal{S}^{k-1}$.

EXAMPLE 1 (continued). *In Figure* 2.1A, $\{1,2\},\{2,3\}$*, and* $\{1,3\}$ *are faces of* $\{1,2,3\}$*. Similarly,* $\{2,3,4\}$ *is a coface of* $\{2,3\}$*,* $\{3,4\}$*, and* $\{2,4\}$*.*

Two k-simplices in an SC $\mathcal{X}$ are *upper adjacent* if they are both faces of the same $(k+1)$-simplex and are *lower adjacent* if both have a common face. For any $\mathcal{S} \in \mathcal{X}$, we define its degree, denoted by $\deg(\mathcal{S})$, to be the number of cofaces of $\mathcal{S}$. We use $\mathcal{X}^k$ to denote the subset of k-simplices in $\mathcal{X}$.

EXAMPLE 1 (continued). *In Figure* 2.1A*, the simplices* $\{3,4\}$ *and* $\{2,4\}$ *are upper adjacent, but* $\{2,6\}$ *and* $\{5,6\}$ *are not. The simplices* $\{3,4\}$ *and* $\{2,4\}$ *are lower adjacent, as are* $\{2,6\}$ *and* $\{5,6\}$*.*

2.3. Oriented Simplicial Complexes and Function Spaces on Simplicial Complexes. While the definition of SCs is based on sets, in order to facilitate computations, we need to define an orientation for each simplex, which we do by fixing an ordering of its vertices. The choice of orientation is a matter for bookkeeping: just as we need to define a node-labeling to represent a graph with an adjacency matrix, we need to define orientations to perform appropriate numerical computations for SCs.[2]

[2]For the expert, there is a small subtlety here. Strictly speaking, orientations are essential because we are concerned with signals defined on SCs represented by vectors with field coefficients in $\mathbb{R}$. If we were to use (binary) field coefficients in $\mathbb{Z}/2$, we would not have to define orientations. However, there is no relevant Hodge theory associated to this case. For applications to data, real-valued coefficients are essential to represent signals on edges and will be our focus.

Formally, an orientation of a k-simplex $\mathcal{S}^k$ ($k > 0$) is an equivalence class of orderings of its vertices, where two orderings are equivalent if they differ by an even permutation. For simplicity, we choose the reference orientation of the simplices induced by the ordering of the vertex labels $\{[i_0, \dots, i_k] : i_0 < \cdots < i_k\}$.

EXAMPLE 1 (continued). *In Figure* 2.1A*, edges and triangles are oriented by arrows on the simplices. In this example, the ordered simplices* $[2, 3, 4]$ *and* $[3, 4, 2]$ *correspond to an equivalent orientation, whereas* $[1, 2]$ *and* $[2, 1]$ *do not.*

A node (0-simplex) can have only one orientation. Hence, issues of orientation do not commonly arise in graph-theoretic settings. An exception is graph-flow problems, in which orientations are defined for edges as above to keep track of the flows appropriately: each flow has a magnitude and a sign to indicate whether the direction of the flow is aligned or antialigned with the chosen reference orientation.

Given a reference orientation for each simplex, for each k we can define the finite-dimensional vector space $\mathcal{C}_k$ with coefficients in $\mathbb{R}$ whose basis elements are the oriented simplices s_i^k. An element $c_k \in \mathcal{C}_k$ is called a *k-chain* and may be thought of as a formal linear combination of these basis elements $c_k = \sum_i \gamma_i s_i^k$. Thus, we can represent each element in $\mathcal{C}_k$ by a vector $\boldsymbol{c} = (\gamma_1, \dots, \gamma_{n_k})^\top$, where $n_k = |\mathcal{X}^k|$ is the number of k-simplices in the SC (Figure 2.1). Note that $\mathcal{C}_k$ is isomorphic to $\mathbb{R}^{n_k}$, so we may think of a k-chain as a vector in $\mathbb{R}^{n_k}$.

EXAMPLE 1 (continued). *In Figure* 2.1A*, the blue vector is the representation of the* 1*-chain* $\boldsymbol{c} = (0, 1, 0, 0, 1, 0, -2, 0, -2, 0)^\top$.

We make one further provision for the construction of $\mathcal{C}_k$—a change of the orientation of the basis element s_i^k is defined to correspond to a change in the sign of the coefficient γ_i. Hence, if we "flip" a basis element s_i^k to its opposite orientation, we have to multiply the corresponding coefficient γ_i by -1. Finally, we endow each space $\mathcal{C}_k$ with the standard ℓ^2 inner product $\langle \boldsymbol{c}_1, \boldsymbol{c}_2 \rangle = \boldsymbol{c}_1^\top \boldsymbol{c}_2$, and thus give $\mathcal{C}_k$ the structure of a finite-dimensional Hilbert space.

An alternative interpretation of the above construction is in terms of the space $\mathcal{C}^k$ of cochains, which is the linear space of all alternating functions $f : \mathcal{C}_k \to \mathbb{R}$ (for a more detailed discussion, see Lim [85]).[3] Since $\mathcal{C}_k$ and $\mathcal{C}^k$ have the same dimension, and there is a canonical isomorphism between the spaces of chains $\mathcal{C}_k$ and cochains $\mathcal{C}^k$, we will treat these two interpretations interchangeably in what follows even though their interpretation can be different.

The reader not familiar with these constructions may simply consider the above spaces in terms of their vector representations. For instance, the space $\mathcal{C}^1$ can be interpreted as the space of *edge flows*, which are commonly encountered in graph theory. Any vector $\boldsymbol{f}$ representing such a flow assigns one scalar value to each edge in the graph, where a negative value indicates that the flow is in the opposite direction with respect to the chosen reference orientation of the edge. To illustrate the abovementioned duality of chains and cochains for the edge-space, think of electrical circuits with unit resistances. In this context, we may think of $\mathcal{C}^1$ as the space of edge-currents, and $\mathcal{C}_1$ as the space of edge-voltages, which encode exactly the same information.

EXAMPLE 1 (continued). *The* 1*-chain* $\boldsymbol{c} = (0, 1, 0, 0, 1, 0, -2, 0, -2, 0)^\top$ *in Figure* 2.1A *can by duality also be thought of as a cochain* $\boldsymbol{f}$*, or simply as the union of an edge flow on* $2 \to 6 \to 5 \to 4$ *and an edge flow on* $1 \to 3$*. Due to the edge orientations,*

[3] Strictly speaking, the space $\mathcal{C}^k$ of cochains corresponds to the dual space of $\mathcal{C}_k$.

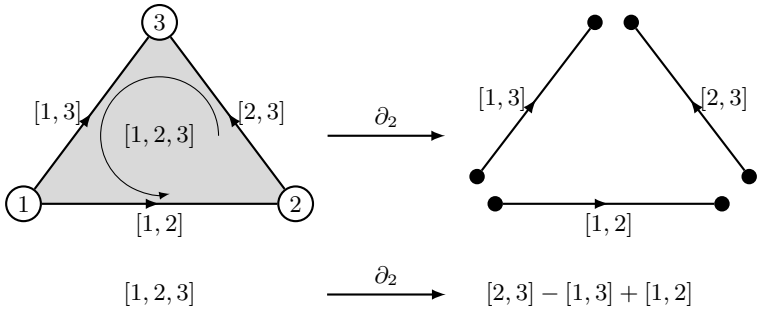

Fig. 2.2 Illustration of the action of a boundary operator on a 2-simplex. *The boundary operator maps the 2-simplex to a linear combination of its faces, respecting the orientation. For simplicity, we choose a basis for $\mathcal{C}_k$ such that $\{s^k = [i_0, \ldots, i_k] : i_0 < \cdots < i_k\}$. Note that within the space of chains (see text), $-[1,3] = [3,1]$, which shows that the above boundary operator gives rise to a cycle $\partial_2([1,2,3]) = [1,2] + [2,3] + [3,1]$ for which $\partial_1(\partial_2([1,2,3])) = 0$.*

some entries of $\boldsymbol{c}$ are negative, corresponding to a flow in the opposite direction to the reference orientation.

An alternative way to think about the space of alternating functions on edges $\mathcal{C}^1$ is to identify it with the set of antisymmetric matrices ($\boldsymbol{A} = -\boldsymbol{A}^\top$), whose sparsity pattern is consistent with the edges present ($A_{i,j} = A_{j,i} = 0$, if $\{i,j\} \notin \mathcal{X}^2$). As $A_{i,j} = -A_{j,i}$, this representation simultaneously encodes both edge orientation $[i,j]$ and $[j,i]$ (with opposite signs, as desired). In what follows, we use the more compact description in terms of vectors, but the reader might find it insightful to keep the above picture in mind.

2.4. Boundary and Coboundary Maps. Given the spaces of chains $\mathcal{C}_k$ defined above, we define the linear boundary maps $\partial_k \colon \mathcal{C}_k \to \mathcal{C}_{k-1}$ by their action on the basis elements as follows:

$$\partial_k([i_0, \ldots, i_k]) = \sum_{j=0}^{k} (-1)^j [i_0, \ldots, i_{j-1}, i_{j+1}, \ldots, i_k].$$

These operators map any chain to a sum of its boundary components, i.e., the simplices lower adjacent to the k-chain considered, with the appropriate orientation (Figure 2.2). We thus call $\mathrm{im}(\partial_k)$ the space of $(k-1)$-boundaries, where $\mathrm{im}(\cdot)$ denotes the image of an operator. It is not difficult to show that if we build a cyclic chain $c_k \in \mathcal{C}_k$ whose start- and end-points are identical, then $\partial_k c_k = 0$; and similarly, any c_k such that $\partial_k c_k = 0$ must correspond to a cycle. Thus, we call $\ker(\partial_k)$ the space of k-cycles.

The boundary operators are linear maps between finite-dimensional vector spaces. After choosing a basis, each of these operators can be represented by a matrix, thereby enabling us to perform computations based on these objects. We will denote the matrix representation of the boundary operators ∂_k by $\boldsymbol{B}_k$ (see [85, 90, 121] for further discussion on how to construct these matrices).

Example 1 (continued). *Consider again the SC in Figure* 2.1. *In this case, the boundary maps $\boldsymbol{B}_1$ (rows indexed by nodes, columns indexed by edges) and $\boldsymbol{B}_2$ (rows*

indexed by edges, columns indexed by 2-simplices) are

$$\boldsymbol{B}_1 = \begin{array}{c|cccccccccc} & [1,2] & [1,3] & [2,3] & [2,4] & [2,6] & [3,4] & [4,5] & [4,7] & [5,6] & [5,7] \\ \hline 1 & -1 & -1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 2 & 1 & 0 & -1 & -1 & -1 & 0 & 0 & 0 & 0 & 0 \\ 3 & 0 & 1 & 1 & 0 & 0 & -1 & 0 & 0 & 0 & 0 \\ 4 & 0 & 0 & 0 & 1 & 0 & 1 & -1 & -1 & 0 & 0 \\ 5 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & -1 & -1 \\ 6 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 1 & 0 \\ 7 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 1 \end{array}, \quad \boldsymbol{B}_2 = \begin{array}{c|cc} & [1,2,3] & [2,3,4] \\ \hline [1,2] & 1 & 0 \\ [1,3] & -1 & 0 \\ [2,3] & 1 & 1 \\ [2,4] & 0 & -1 \\ [2,6] & 0 & 0 \\ [3,4] & 0 & 1 \\ [4,5] & 0 & 0 \\ [4,7] & 0 & 0 \\ [5,6] & 0 & 0 \\ [5,7] & 0 & 0 \end{array}.$$

The matrix $\boldsymbol{B}_1$ is the node-to-edge incidence matrix from algebraic graph theory. Likewise, the higher-order boundary maps induce matrices $\boldsymbol{B}_i$ that can be interpreted as higher-order incidence matrices between simplices and their (co)faces.

Note that for each boundary map there exists a coboundary map $\partial_k^\top : \mathcal{C}_k \to \mathcal{C}_{k+1}$, which is simply the adjoint of the boundary map. The matrix representation of the coboundary operator $\partial_k^\top$ is $\boldsymbol{B}_k^\top$.

2.5. Hodge Laplacians. From the sequences of boundary maps, one can define a hierarchy of Laplacian operators for the SC $\mathcal{X}$. Using our matrix representations discussed above, the kth *combinatorial Hodge Laplacian* is

$$\boldsymbol{L}_k = \boldsymbol{B}_k^\top \boldsymbol{B}_k + \boldsymbol{B}_{k+1}\boldsymbol{B}_{k+1}^\top. \tag{2.1}$$

The standard combinatorial graph Laplacian is a special case of the above and corresponds to $\boldsymbol{L}_0 = \boldsymbol{B}_1\boldsymbol{B}_1^\top$ (as $\boldsymbol{B}_0 = 0$). The matrix $\boldsymbol{L}_1$, which is also referred to simply as the (Hodge) 1-Laplacian, is the primary focus of this paper.

As solutions to the Laplace equation ($\Delta \boldsymbol{x} = 0$) are called harmonic functions, and $\boldsymbol{L}_k$ may be interpreted as a discretized version of the Laplace equation [85], elements $\boldsymbol{h} \in \ker(\boldsymbol{L}_k)$ are called harmonic (functions). These harmonic functions also carry a specific topological meaning. From the definitions of the boundary maps, we can compute that $\partial_{k-1} \circ \partial_k = 0$. Thus, the adjoint of this map is also zero, i.e., $\partial_k^\top \circ \partial_{k-1}^\top = 0$. These equations encapsulate the natural idea that the "boundary of a boundary" is empty. In matrix terms, this annihilation of the boundary maps translates into $\boldsymbol{B}_k\boldsymbol{B}_{k+1} = 0$ and $\boldsymbol{B}_{i+1}^\top \boldsymbol{B}_i^\top = 0$.

Furthermore, since $\partial_k \circ \partial_{k+1} = 0$, $\mathrm{im}(\partial_{k+1})$ is a subspace of $\ker(\partial_k)$. This leads to the definition of the homology vector spaces of $\mathcal{X}$ over $\mathbb{R}$ as those elements in the null space $\ker(\partial_k)$ which are not in the image $\mathrm{im}(\partial_{k+1})$:

$$\mathcal{H}_k := \mathcal{H}(\mathcal{X}, \mathbb{R}) = \ker(\partial_k)/\mathrm{im}(\partial_{k+1}). \tag{2.2}$$

Intuitively, the homology $\mathcal{H}_k$ may be interpreted as accounting for the number of k-dimensional "holes" in the SC $\mathcal{X}$. More precisely, elements of $\mathcal{H}_k$ correspond to k-cycles that are not induced by a k-boundary. The number of k-dimensional holes in the SC is the so-called the kth Betti number. It can be shown that this corresponds precisely to the dimension of the null space of the kth Hodge Laplacian $\ker(\boldsymbol{L}_k)$ [85].

2.6. The Hodge Decomposition. The combinatorial Hodge Laplacian is a sum of two positive semidefinite operators, so any $\boldsymbol{h} \in \ker(\boldsymbol{L}_k)$ fulfills $\boldsymbol{h} \in \ker(\boldsymbol{B}_k)$ and

$\boldsymbol{h} \in \ker(\boldsymbol{B}_{k+1}^\top)$. This implies that the nonzero elements in $\ker(\boldsymbol{L}_k)$ are representatives of the nontrivial equivalence classes in the kth homology.

The space $\mathcal{C}_k$ is isomorphic to $\mathbb{R}^{n_k}$ within our chosen representation, which can be represented by $\ker(\boldsymbol{L}_k) \oplus \operatorname{im}(\boldsymbol{L}_k^T) = \ker(\boldsymbol{L}_k) \oplus \operatorname{im}(\boldsymbol{L}_k)$, where $\oplus$ denotes the union of orthogonal subspaces with respect to the standard inner product. Clearly, $\operatorname{im}(\boldsymbol{L}_k) \subseteq \operatorname{im}(\boldsymbol{B}_k^T) \cup \operatorname{im}(\boldsymbol{B}_{k+1})$. Moreover, as $\operatorname{im}(\boldsymbol{B}_{k+1}) \subset \ker(\boldsymbol{B}_k^T)$ in our case, we have that $\operatorname{im}(\boldsymbol{L}_k) \subseteq \operatorname{im}(\boldsymbol{B}_k^T) \oplus \operatorname{im}(\boldsymbol{B}_{k+1})$. Since the dimension of $\operatorname{im}(\boldsymbol{B}_k^T) \oplus \operatorname{im}(\boldsymbol{B}_{k+1})$ cannot exceed the dimension of $\operatorname{im}(\boldsymbol{L}_k)$, these subspaces must be equal, and we can decompose C_k as follows:

$$\mathcal{C}_k \simeq \mathbb{R}^{n_k} = \operatorname{im}(\boldsymbol{B}_{k+1}) \oplus \operatorname{im}(\boldsymbol{B}_k^\top) \oplus \ker(\boldsymbol{L}_k). \tag{2.3}$$

Equation (2.3) is called the *Hodge decomposition*. Later, we discuss how the Hodge decomposition for $\mathcal{C}_1$ can provide additional insights into the data.

3. Diffusion Processes on Simplicial Complexes. In this section, we outline our model for diffusion processes on SCs that accounts for topological features. For simplicity, we focus on 1-simplices, i.e., diffusion between edges.

Diffusion Processes on Graphs. To understand the complications of defining a diffusion process on an SC, let us revisit a random walk on a graph, a prototypical model for a diffusion process on a graph. A (standard, unbiased) random walk on a graph with adjacency matrix $\boldsymbol{A}$ can be described by the following transition rule:

$$\boldsymbol{p}_{t+1} = \boldsymbol{A}\boldsymbol{D}^{-1}\boldsymbol{p}_t = (\boldsymbol{I} - \boldsymbol{L}_0\boldsymbol{D}^{-1})\boldsymbol{p}_t. \tag{3.1}$$

Here the ith component of the vector $\boldsymbol{p}_t$ denotes the probability of finding a random walker at node i at time t, and $\boldsymbol{p}_0$ corresponds to an initial distribution of the walker.

There are two important features of this formulation. First, the transition matrix of the random walk is directly related to a normalized Hodge Laplacian, namely, $\boldsymbol{\mathcal{L}}_0 = \boldsymbol{L}_0\boldsymbol{D}^{-1}$ is the so-called random walk Laplacian. There is thus a close relationship between the topological features of the graph and the random walk, as the harmonic functions of $\boldsymbol{\mathcal{L}}_0$ are directly related to the connected components of the graph. Second, the state space and the transitions of the random walker are determined by the graph. The nodes are the states of the random walker and transitions occur over the edges.

3.1. Beyond Graphs: Keeping Track of Orientations. When extending the concept of a random walk to SCs, a mismatch between the two features discussed above becomes apparent if we go beyond the node-space. On the one hand, we may define a random walk on the edges, where the edges themselves are defined as the states of the Markov process. To define such a process we could use the line graph [1, 47] or other "dual graph" constructions [97]. However, we would be abandoning the connection to algebraic topology and the Hodge Laplacian, as the Laplacian of the line graph is not directly related to the Hodge Laplacian of the SC. The properties of a random walk on the line graph will therefore not be informative about the topology of the SC.

On the other hand, we face a different issues if we define a random walk based on the $\boldsymbol{L}_1$ Laplacian formally analogous to (3.1). The Laplacian $\boldsymbol{L}_1$ has nontrivial patterns of positive and negative entries that depend on the edge orientations. Hence, the $\boldsymbol{L}_1$ Laplacian is not obviously related to a transition matrix of a Markov chain. We do not face this issue with the $\boldsymbol{L}_0$ Laplacian because orientation is trivial for vertices.

Can a "normalized" variant of the $\boldsymbol{L}_1$ Laplacian be related to a random walk? It turns out that we can indeed construct an edge-based diffusion process that is tied to

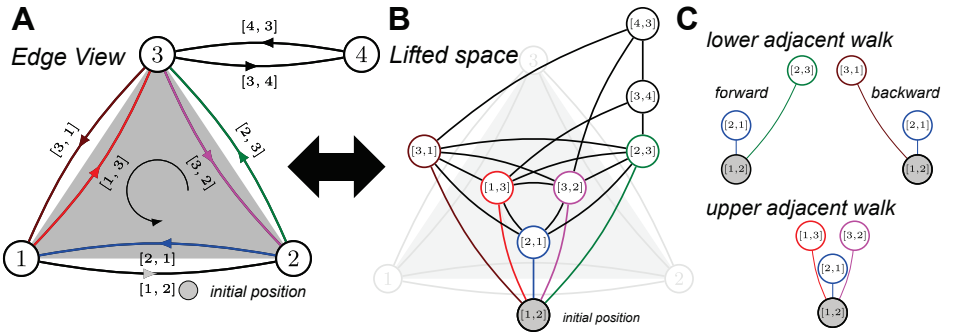

Fig. 3.1 Illustration of lifted simplicial complex. (A) *We may think of the lifted complex as an augmented complex in which each original edge is represented in both possible orientations.* (B) *Alternatively, we may interpret each oriented edge* $[i, j]$ *in the original complex as giving rise to two states* $[i, j]$ *and* $[j, i]$ *on a graph with* $2n_1$ *vertices.* (C) *Starting from* $[1, 2]$*, there are lower adjacent connections "forward" and "backward," as well as upper adjacent connections (see text).*

the topology of the original complex. However, we have to consider a random walk in a higher-dimensional, lifted state space. Our idea is that instead of considering how $\boldsymbol{L}_1$ acts on an edge flow $\boldsymbol{f}$ in the original space, we view its action as an equivalent sequence of three operations: first, we lift $\boldsymbol{f}$ into a higher-dimensional space; second, we act on it via a linear operator; and third, we project the result back down to the original state space. Once we understand these actions, we can normalize the linear transformation in the lifted space such that it corresponds to a diffusion. This leads to the definition of a normalized Hodge Laplacian, to which we can assign a meaning in terms of a random walk in a lifted space.

Decomposing the action of the Hodge Laplacian $\boldsymbol{L}_1$ in the above manner (lift, apply, project) enables us to disentangle the orientation of a flow and the magnitude of the flow. The magnitude of each component indicates the volume of the flow, whereas the sign of the variable indicates the direction of the flow, which can be aligned or antialigned with our chosen reference orientation. As we will see, the magnitude of the flow can be related to a probability, whereas the information about the direction of the flow is a matter of accounting for a reference orientation.

3.2. Lifting of Edge Flows and Matrix Operators. In what follows, we describe how the action of the $\boldsymbol{L}_1$ Laplacian on any cochain vector $\boldsymbol{f}$ (edge flow) can be understood from the point of view of a higher-dimensional, lifted state space. We consider a lifting of an edge flow $\boldsymbol{f} \in \mathcal{C}^1$ into a larger space $\mathcal{D}^1$ in which both possible orientations for each edge are present (see Figure 3.1). Since there are two possible orientations for each edge, $|\mathcal{D}^1| = 2|\mathcal{C}^1|$. As an edge flow in $\mathcal{C}^1$ corresponds to an alternating function, there is a natural inclusion map $V : \mathcal{C}^1 \to \mathcal{D}^1$ which maps any edge flow into $\mathcal{D}^1$ by explicitly representing both edge directions. We choose the basis elements of $\mathcal{D}_1$ such that the matrix representation of V is

$$\boldsymbol{V} = \begin{pmatrix} +\boldsymbol{I}_{n_1} \\ -\boldsymbol{I}_{n_1} \end{pmatrix} \in \mathbb{R}^{2n_1 \times n_1}, \tag{3.2}$$

where $\boldsymbol{I}_{n_1}$ is the identity matrix of dimension $n_1 = |\mathcal{C}_1|$.

EXAMPLE 1 (continued). *Consider the edge flow* $\boldsymbol{f} = (0,1,0,0,1,0,-2,0,-2,0)^\top$ *in Figure* 2.1A. *The lifted edge flow is simply* $\widehat{\boldsymbol{f}} = \boldsymbol{V}\boldsymbol{f} = (\boldsymbol{f}^\top, -\boldsymbol{f}^\top)^\top$. *For instance,* $\widehat{\boldsymbol{f}}$ *has an entry* 1 *for edge* $[2,6]$ *and* -1 *for the (now added) reverse edge* $[6,2]$.

The lifting operator has the property that $\boldsymbol{V}^\top\boldsymbol{V} = 2\boldsymbol{I}_{n_1}$. Thus, the Moore–Penrose pseudoinverse of $\boldsymbol{V}$ is $\boldsymbol{V}^\dagger = \frac{1}{2}\boldsymbol{V}^\top$. Furthermore, it is easy to see that $\boldsymbol{V}\boldsymbol{V}^\top = \boldsymbol{I}_{2n_1} - \boldsymbol{\Sigma}$, where $\boldsymbol{\Sigma}$ is the permutation matrix that maps the original basis simplices to their counterparts with switched orientation:

$$\boldsymbol{\Sigma} = \begin{pmatrix} \boldsymbol{0} & \boldsymbol{I}_{n_1} \\ \boldsymbol{I}_{n_1} & \boldsymbol{0} \end{pmatrix}. \tag{3.3}$$

Having defined a lifting for an edge flow, we now need to define an appropriate notion for a lifting of a matrix operator.

DEFINITION 3.1 (lifting of a matrix). *We say that a matrix* $\boldsymbol{N} \in \mathbb{R}^{2n_k \times 2n_k}$ *is a lifting of a matrix* $\boldsymbol{M} \in \mathbb{R}^{n_k \times n_k}$ *if the following condition holds:*

$$\boldsymbol{V}^\top\boldsymbol{N} = \boldsymbol{M}\boldsymbol{V}^\top. \tag{3.4}$$

This definition implies that if a matrix $\boldsymbol{M}$ has a lifting $\boldsymbol{N}$, then by multiplying from the right with $\boldsymbol{V}$, $\boldsymbol{M} = \frac{1}{2}\boldsymbol{V}^\top\boldsymbol{N}\boldsymbol{V} = \boldsymbol{V}^\dagger\boldsymbol{N}\boldsymbol{V}$. Hence, the action of multiplying by $\boldsymbol{M}$ can be interpreted in terms of a lifting, followed by a linear transformation, and finally a projection into the original lower-dimensional space (we use the term "projection" here to refer to a mapping into a lower-dimensional space).

We now consider a lifting of the combinatorial Hodge Laplacian $\boldsymbol{L}_1$. To state our results compactly we define the following matrices:

$$\widehat{\boldsymbol{B}}_1 := \boldsymbol{B}_1\boldsymbol{V}^\top = \begin{pmatrix} \boldsymbol{B}_1 & -\boldsymbol{B}_1 \end{pmatrix} \quad \text{and} \quad \widehat{\boldsymbol{B}}_2 := \boldsymbol{V}\boldsymbol{B}_2 = \begin{pmatrix} \boldsymbol{B}_2 \\ -\boldsymbol{B}_2 \end{pmatrix}. \tag{3.5}$$

We will, moreover, make use of the positive part $\widehat{\boldsymbol{B}}_i^+$ and the negative part $\widehat{\boldsymbol{B}}_i^-$ of these matrices. In the following we use the notation $\widehat{\cdot}$ to indicate objects that are related to the lifted space. Note that such objects may not always be liftings of matrices (consider, e.g., $\widehat{\boldsymbol{B}}_1$).

LEMMA 3.2 (lifting of the $\boldsymbol{L}_1$ Hodge Laplacian). *The negative of the Hodge Laplacian* $\boldsymbol{L}_1 = \boldsymbol{B}_1^\top\boldsymbol{B}_1 + \boldsymbol{B}_2\boldsymbol{B}_2^\top$ *has a lifting* $\widehat{\boldsymbol{A}} = \widehat{\boldsymbol{A}}_l + \widehat{\boldsymbol{A}}_u$ *with*

$$\widehat{\boldsymbol{A}}_l = (\widehat{\boldsymbol{B}}_1^-)^\top\widehat{\boldsymbol{B}}_1^+ + (\widehat{\boldsymbol{B}}_1^+)^\top\widehat{\boldsymbol{B}}_1^- \text{ and } \widehat{\boldsymbol{A}}_u = \widehat{\boldsymbol{B}}_2^+(\widehat{\boldsymbol{B}}_2^-)^\top + \widehat{\boldsymbol{B}}_2^-(\widehat{\boldsymbol{B}}_2^+)^\top.$$

Let us unpack this result before proving it. We can see immediately that $\widehat{\boldsymbol{A}}$ is symmetric. Next, for any matrix $\boldsymbol{M}$, we have that $(-\boldsymbol{M})^- = \boldsymbol{M}^+$; $(-\boldsymbol{M})^+ = \boldsymbol{M}^-$; and $\boldsymbol{M}^+$ and $\boldsymbol{M}^-$ are nonnegative. Thus, $\widehat{\boldsymbol{A}}$ is also nonnegative and can be interpreted as the (weighted) adjacency matrix of an undirected graph with $2n_1$ nodes. More specifically, $\widehat{\boldsymbol{A}}_l$ describes connections between lower adjacent edges and is composed of

$$((\widehat{\boldsymbol{B}}_1^-)^\top\widehat{\boldsymbol{B}}_1^+)_{[i,j],[k,l]} = \begin{cases} 1 & \text{if } l = i, \\ 0 & \text{otherwise,} \end{cases} \qquad ((\widehat{\boldsymbol{B}}_1^+)^\top\widehat{\boldsymbol{B}}_1^-)_{[i,j],[k,l]} = \begin{cases} 1 & \text{if } k = j, \\ 0 & \text{otherwise.} \end{cases}$$

The first matrix describes a forward walk respecting the edge orientation, where the target node l of the first edge $[k,l]$ has to match with source node i of the second edge

$[i,j]$. The second matrix describes a backward walk in the opposite direction, where the source node k of the first edge $[k,l]$ has to match with the target node j of the second edge $[i,j]$ (see Figure 3.1).

Likewise, $\widehat{\boldsymbol{A}}_u$ describes connections between upper adjacent edges with a joint triangular coface. Using the symbol $\nsim$ to denote that two edges have a different orientation relative to a joint coface, we can write

$$(\widehat{\boldsymbol{A}}_u)_{[i,j],[k,l]} = \begin{cases} 1 & \text{if } [k,l] \nsim [i,j], \\ 0 & \text{otherwise.} \end{cases} \tag{3.6}$$

EXAMPLE 2. *Consider Figure* 3.1. *In the lifted space* $\mathcal{D}^1$*, the lower adjacent forward connections of* $[1,2]$ *are* $[2,1]$ *and* $[2,3]$*; the lower adjacent backward connections are* $[2,1]$ *and* $[3,1]$*. The upper adjacent connections of* $[1,2]$ *with opposite orientation are* $[2,1]$*,* $[3,2]$*, and* $[1,3]$*.*

To conclude this section, we prove Lemma 3.2.

Proof. Since $(-\boldsymbol{M})^- = \boldsymbol{M}^+$ and $(-\boldsymbol{M})^+ = \boldsymbol{M}^-$, for $\widehat{\boldsymbol{A}}_l$ we have that

$$\widehat{\boldsymbol{B}}_1^- \boldsymbol{V} = (\boldsymbol{B}_1, -\boldsymbol{B}_1)^- \boldsymbol{V} = -\boldsymbol{B}_1, \quad \widehat{\boldsymbol{B}}_1^+ \boldsymbol{V} = (\boldsymbol{B}_1, -\boldsymbol{B}_1)^+ \boldsymbol{V} = \boldsymbol{B}_1, \quad \widehat{\boldsymbol{B}}_1^+ \boldsymbol{\Sigma} = \widehat{\boldsymbol{B}}_1^-.$$

Using the transposes of the first two equalities as well as the third equality,

$$\boldsymbol{V}^\top \widehat{\boldsymbol{A}}_l = -\boldsymbol{B}_1^\top \widehat{\boldsymbol{B}}_1^+ + \boldsymbol{B}_1^\top \widehat{\boldsymbol{B}}_1^- = -\boldsymbol{B}_1^\top \widehat{\boldsymbol{B}}_1^+ (\boldsymbol{I} - \boldsymbol{\Sigma}) = -\boldsymbol{B}_1^\top \widehat{\boldsymbol{B}}_1^+ \boldsymbol{V}\boldsymbol{V}^\top = -\boldsymbol{B}_1^\top \boldsymbol{B}_1 \boldsymbol{V}^\top.$$

By analogous arguments for $\widehat{\boldsymbol{A}}_u$, we obtain

$$\boldsymbol{V}^\top \widehat{\boldsymbol{A}}_u = \boldsymbol{B}_2 (\widehat{\boldsymbol{B}}_2^-)^\top - \boldsymbol{B}_2 (\widehat{\boldsymbol{B}}_2^+)^\top = -\boldsymbol{B}_2 (\widehat{\boldsymbol{B}}_2^+)^\top (\boldsymbol{I} - \boldsymbol{\Sigma}) = -\boldsymbol{B}_2 \boldsymbol{B}_2^\top \boldsymbol{V}^\top.$$

The lemma follows by combining these two results. ∎

3.3. The Normalized Hodge 1-Laplacian and Edge-Space Random Walks. Motivated by our lifting result in Lemma 3.2, we now define a normalized Hodge Laplacian for the edge-space and show that its action can be related to a random walk on a lifted complex. There is some flexibility here, however, as multiple operators in the lifted space $\mathcal{D}^1$ will correspond to the same projected matrix. Likewise, there are multiple types of random walks we could define in the lifted space by assigning different weights to the various transitions, leading to different notions of a normalized Laplacian operator. The normalized Hodge Laplacian we consider here is of a "standard form" [62, 68] and admits a (normalized) Hodge decomposition, which is of interest for our applications. A systematic exploration of further normalization schemes and their respective advantages is an interesting avenue for future research.

DEFINITION 3.3 (normalized Hodge 1-Laplacian $\boldsymbol{\mathcal{L}}_1$). *Consider an SC* $\mathcal{X}$ *whose boundary operators can be represented by the matrices* $\boldsymbol{B}_1$ *and* $\boldsymbol{B}_2$*. The normalized Hodge* 1*-Laplacian matrix is then defined by*

$$\boldsymbol{\mathcal{L}}_1 = \boldsymbol{D}_2 \boldsymbol{B}_1^\top \boldsymbol{D}_1^{-1} \boldsymbol{B}_1 + \boldsymbol{B}_2 \boldsymbol{D}_3 \boldsymbol{B}_2^\top \boldsymbol{D}_2^{-1}, \tag{3.7}$$

where $\boldsymbol{D}_2$ *is the diagonal matrix of (adjusted) degrees of each edge,*

$$\boldsymbol{D}_2 = \max(\operatorname{diag}(|\boldsymbol{B}_2|\mathbf{1}), \boldsymbol{I}) \iff (\boldsymbol{D}_2)_{[i,j],[i,j]} = \max\{\deg([i,j]), 1\}, \tag{3.8}$$

$\boldsymbol{D}_1 = 2 \cdot \operatorname{diag}(|\boldsymbol{B}_1| \boldsymbol{D}_2 \mathbf{1})$ *is a diagonal matrix of weighted degrees of the nodes (with the weight of an edge equal to the maximum of* 1 *and the number of cofaces of the edge), and* $\boldsymbol{D}_3 = \frac{1}{3}\boldsymbol{I}$*.*

In the above definition, the matrix $\boldsymbol{D}_2$ defines a weighting of the edges according to their degree, where the elementwise maximum in $\boldsymbol{D}_2$ ensures that the normalized Hodge Laplacian is well-defined (i.e., the edge weight of an existing edge is at least 1). The matrix $\boldsymbol{D}_1$ encodes twice the weighted degree of the nodes according to the weights of the incident edges, and $\boldsymbol{D}_3$ gives a weighting of $1/3$ to each triangular face. Recall that in the standard case of graphs, the Hodge Laplacian is

$$\boldsymbol{L}_0 = \mathbf{0}_1^\top \mathbf{0}_1 + \boldsymbol{B}_1 \boldsymbol{B}_1^\top = \boldsymbol{B}_1 \boldsymbol{B}_1^\top,$$

and the normalized (random walk) version is

$$\boldsymbol{\mathcal{L}}_0 = \boldsymbol{B}_1 \widetilde{\boldsymbol{D}}_3 \boldsymbol{B}_1^\top \widetilde{\boldsymbol{D}}_2^{-1},$$

with $\widetilde{\boldsymbol{D}}_3 = \boldsymbol{I}$ and $\widetilde{\boldsymbol{D}}_2 = \max(\operatorname{diag}(|\boldsymbol{B}_1|\mathbf{1}), \boldsymbol{I})$. Note how $\widetilde{\boldsymbol{D}}_2$ and $\boldsymbol{D}_2$ have exactly the same functional form, but are based on $\boldsymbol{B}_1$ and $\boldsymbol{B}_2$, respectively. Similarly, both $\widetilde{\boldsymbol{D}}_3$ and $\boldsymbol{D}_3$ are simple scalings.

The key difference between the normalized random walk Laplacian and the normalized Hodge 1-Laplacian is that the latter contains both an upper-adjacent and a lower-adjacent term. The following result shows that this normalized Hodge 1-Laplacian has a meaningful connection to random walks. Specifically, the matrix is related to a random walk in the lifted space of edges.

THEOREM 3.4 (stochastic lifting of the normalized Hodge 1-Laplacian). *The matrix $-\boldsymbol{\mathcal{L}}_1/2$ has a stochastic lifting, i.e., there exists a column stochastic matrix $\widehat{\boldsymbol{P}}$ such that $-\frac{1}{2}\boldsymbol{\mathcal{L}}_1 \boldsymbol{V}^\top = \boldsymbol{V}^\top \widehat{\boldsymbol{P}}$. Specifically, $\widehat{\boldsymbol{P}} := \frac{1}{2}\boldsymbol{P}_{\text{lower}} + \frac{1}{2}\boldsymbol{P}_{\text{upper}}$, where $\boldsymbol{P}_{\text{lower}}$ is the transition matrix of a random walk determined by the lower-adjacent connections and $\boldsymbol{P}_{\text{upper}}$ is the transition matrix of a random walk determined by the upper-adjacent connections. The transition matrix $\boldsymbol{P}_{\text{lower}}$ is defined by a "forward walk" and a "backward walk" component moving in the orientation of the edges or against it, respectively:*

$$\boldsymbol{P}_{\text{lower}} := \frac{1}{2}\left(\boldsymbol{P}_{\text{lower,forward}} + \boldsymbol{P}_{\text{lower,backward}}\right), \tag{3.9}$$

$$\boldsymbol{P}_{\text{lower,forward}} = \boldsymbol{M}_f \operatorname{diag}(\boldsymbol{M}_f \mathbf{1})^{-1}, \tag{3.10}$$

$$\boldsymbol{P}_{\text{lower,backward}} = \boldsymbol{M}_b \operatorname{diag}(\boldsymbol{M}_b \mathbf{1})^{-1}, \tag{3.11}$$

where $\boldsymbol{M}_f = \widehat{\boldsymbol{D}}_2 (\widehat{\boldsymbol{B}}_1^-)^\top \widehat{\boldsymbol{B}}_1^+$ and $\boldsymbol{M}_b = \widehat{\boldsymbol{D}}_2 (\widehat{\boldsymbol{B}}_1^+)^\top \widehat{\boldsymbol{B}}_1^-$ are (weighted) lower-adjacency matrices corresponding to forward and backward walks along the edges (see Lemma 3.2) and $\widehat{\boldsymbol{D}}_2 = \operatorname{diag}(\boldsymbol{D}_2, \boldsymbol{D}_2)$. The transition matrix $\boldsymbol{P}_{\text{upper}}$ describes a random walk along upper-adjacent faces as follows:

$$\boldsymbol{P}_{\text{upper}} = \widehat{\boldsymbol{A}}_u \widehat{\boldsymbol{D}}_4^{-1} + \frac{1}{2}\begin{pmatrix} \boldsymbol{I} & \boldsymbol{I} \\ \boldsymbol{I} & \boldsymbol{I} \end{pmatrix} \widehat{\boldsymbol{D}}_5, \tag{3.12}$$

where $\widehat{\boldsymbol{A}}_u = \widehat{\boldsymbol{B}}_2^+ (\widehat{\boldsymbol{B}}_2^-)^\top + \widehat{\boldsymbol{B}}_2^- (\widehat{\boldsymbol{B}}_2^+)^\top$ is the matrix of upper-adjacent connections as defined in Lemma 3.2 and $\widehat{\boldsymbol{D}}_4$ is a diagonal matrix:

$$(\widehat{\boldsymbol{D}}_4)_{[i,j],[i,j]} = \begin{cases} 1 & \text{if} \quad \deg([i,j]) = 0, \\ 3 \cdot \deg([i,j]) & \text{otherwise.} \end{cases} \tag{3.13}$$

Here $\widehat{\boldsymbol{D}}_5$ is the diagonal matrix selecting all edges with no upper-adjacent faces:

$$(\widehat{\boldsymbol{D}}_5)_{[i,j],[i,j]} = \begin{cases} 1 & \text{if} \quad \deg([i,j]) = 0, \\ 0 & \text{otherwise.} \end{cases} \tag{3.14}$$

Proof. The proof closely follows our lifting result above and is in the appendix. ∎

The random walk described by $\widehat{\boldsymbol{P}}$ can be described in words as follows. With a probability of 0.5 each, we take a step via either the upper- or the lower-adjacent connection. If we take a step via the lower-adjacent connections (via the nodes), then with probability of 0.5 each we move either along or against the chosen edge orientation. In either case, the transition probability to a target edge is then proportional to the upper degree of the target edge, which corresponds to the "weight" of that edge. If we take a step via the upper-adjacent connection, there are two cases. If the edge has no upper-adjacent face, then the random walk will stay at the same oriented edge or change orientation with equal probability 0.5. If the edge has an upper-adjacent face, then a walker on edge $[i, j]$ will transition uniformly to an upper-adjacent edge $[k, l] \nsim [i, j]$ with different orientation relative to their shared face. Stated differently, the walker performs an unbiased random walk on the lifted graph with adjacency matrix $\widehat{\boldsymbol{A}}_u$, unless there is no upper-adjacent connection, in which case the walker will either stay put or move to the edge with reverse orientation.

Finally, there exists an interesting link between the construction of the normalized Hodge 1-Laplacian here and previous research that considered certain relationships between Hodge Laplacians and *differences* of random walks on SCs (with absorbing states) [91, 108]. As the projection operator computes a difference between the two possible orientations of an edge in the lifted space, an idea for future work would be to explore this connection in more detail.

Spectral Properties. The ideas underpinning Theorem 3.4 enable us to derive the following results, which have consequences for the spectral properties of $\boldsymbol{\mathcal{L}}_1$.

COROLLARY 3.5. *Define the matrix* $\boldsymbol{Z} = -\boldsymbol{\mathcal{L}}_1/2$ *and the matrix* $\widehat{\boldsymbol{P}}$ *as in Theorem* 3.4. *Then the following identities hold:*

1. $\boldsymbol{Z}\boldsymbol{V}^\top = \boldsymbol{V}^\top \widehat{\boldsymbol{P}}$,
2. $\boldsymbol{Z} = \boldsymbol{V}^\dagger \widehat{\boldsymbol{P}} \boldsymbol{V}$, *and*
3. $\boldsymbol{V}\boldsymbol{Z} = \boldsymbol{V}\boldsymbol{V}^\dagger \widehat{\boldsymbol{P}} \boldsymbol{V} = \widehat{\boldsymbol{P}} \boldsymbol{V}$.

Proof. The first two relations are a simple restatement of Theorem 3.4. The last equality can be shown analogously to Theorem 3.4 and is omitted for brevity. ∎

COROLLARY 3.6. *Consider the (lifted) space* $\mathcal{D}_1$ *of edge flows (cf. subsection* 3.2*). The subspace of alternating functions* $\mathrm{span}(\boldsymbol{V}) \subset \mathcal{D}_1$ *is an invariant subspace of* $\widehat{\boldsymbol{P}}$. *Consequently, the spectrum of* $\boldsymbol{Z}$ *is contained in the spectrum of* $\widehat{\boldsymbol{P}}$, *i.e.,* $\lambda(\boldsymbol{Z}) \subset \lambda(\widehat{\boldsymbol{P}})$. *Furthermore, any eigenvector* $\boldsymbol{x}$ *of* $\boldsymbol{Z}$ *with eigenvalue* λ *corresponds to an eigenvector of* $\widehat{\boldsymbol{P}}$ *of the form* $\boldsymbol{y} = \boldsymbol{V}\boldsymbol{x}$ *with the same eigenvalue.*

Proof. The first claim follows immediately from $\widehat{\boldsymbol{P}}\boldsymbol{V} = \boldsymbol{V}\boldsymbol{Z}$. The latter parts follow by using the identities established in Corollary 3.5 to compute $\widehat{\boldsymbol{P}}\boldsymbol{y} = \widehat{\boldsymbol{P}}\boldsymbol{V}\boldsymbol{x} = \boldsymbol{V}\boldsymbol{Z}\boldsymbol{x} = \lambda \boldsymbol{V}\boldsymbol{x}$, which holds for any eigenvector $\boldsymbol{x}$ of $\boldsymbol{Z}$. ∎

The above result implies that, like the normalized graph Laplacian, the spectrum of $\boldsymbol{\mathcal{L}}_1$ has a bounded support.

3.4. Normalized Hodge Decompositions. Similar to the Hodge 1-Laplacian, the eigenvectors of the normalized Hodge 1-Laplacian $\boldsymbol{\mathcal{L}}_1$ associated to the eigenvalue $\lambda = 0$ of $\boldsymbol{\mathcal{L}}$ and the induced eigenvectors of $\widehat{\boldsymbol{P}}$ are associated with scaled harmonic functions. In fact, we can obtain the following normalized (weighted) Hodge decomposition from our normalized Hodge 1-Laplacian:

$$\mathbb{R}^{n_1} = \mathrm{im}(\boldsymbol{B}_2) \oplus_{\boldsymbol{D}_2^{-1}} \mathrm{im}(\boldsymbol{D}_2 \boldsymbol{B}_1^\top) \oplus_{\boldsymbol{D}_2^{-1}} \ker(\boldsymbol{\mathcal{L}}_1), \tag{3.15}$$

where $\oplus_{\boldsymbol{D}_2^{-1}}$ denotes the union of orthogonal subspaces with respect to the inner product $\langle \boldsymbol{x}, \boldsymbol{y} \rangle_{\boldsymbol{D}_2^{-1}} = \boldsymbol{x}^\top \boldsymbol{D}_2^{-1} \boldsymbol{y}$. Comparing (2.3) with (3.15), it should be apparent that there is an isomorphism between the respective subspaces of the standard and the normalized Hodge Laplacians, and thus a correspondence between the harmonic functions associated with $\boldsymbol{L}_1$ and $\boldsymbol{\mathcal{L}}_1$. If we consider a symmetrized version of $\boldsymbol{\mathcal{L}}_1$ given by $\boldsymbol{\mathcal{L}}_1^{\mathrm{s}} = \boldsymbol{D}_2^{-1/2} \boldsymbol{\mathcal{L}}_1 \boldsymbol{D}_2^{1/2}$, the corresponding Hodge decomposition again holds with respect to the standard inner product:

$$\mathbb{R}^{n_1} = \mathrm{im}(\boldsymbol{D}_2^{-1/2} \boldsymbol{B}_2) \oplus \mathrm{im}(\boldsymbol{D}_2^{1/2} \boldsymbol{B}_1^\top) \oplus \ker(\boldsymbol{\mathcal{L}}_1^{\mathrm{s}}). \tag{3.16}$$

We will use this normalized Hodge decomposition in our application examples (see also subsection 4.3 for further discussion).

In addition to the eigenvectors associated to zero eigenvalues, which have a clear interpretation in terms of harmonic functions (homology), we can provide further insights on the remaining eigenvectors. We state our results in terms of $\boldsymbol{\mathcal{L}}_1^s$. However, these results can be reformulated in terms of the left and right eigenvectors of $\boldsymbol{\mathcal{L}}_1$. In particular, if $\boldsymbol{\mathcal{L}}_1^s \boldsymbol{u} = \lambda \boldsymbol{u}$, then $\boldsymbol{\mathcal{L}}_1 \boldsymbol{u}_R = \lambda \boldsymbol{u}_R$ with $\boldsymbol{u}_R = \boldsymbol{D}_2^{1/2} \boldsymbol{u}$, and similarly $\boldsymbol{u}_L^\top = \boldsymbol{u}^\top \boldsymbol{D}_2^{-1/2}$ for the left eigenvectors of $\boldsymbol{\mathcal{L}}_1$.

THEOREM 3.7. *Consider the matrices*

$$\boldsymbol{G}_1 = \boldsymbol{D}_1^{-1/2} \boldsymbol{B}_1 \boldsymbol{D}_2 \boldsymbol{B}_1^\top \boldsymbol{D}_1^{-1/2} \quad \textit{and} \quad \boldsymbol{G}_2 = \boldsymbol{D}_3^{1/2} \boldsymbol{B}_2^\top \boldsymbol{D}_2^{-1} \boldsymbol{B}_2 \boldsymbol{D}_3^{1/2}. \tag{3.17}$$

Then the following statements hold:

1. *Every eigenvector $\boldsymbol{u}$ of $\boldsymbol{G}_1$ with eigenvalue λ has a corresponding eigenvector of $\boldsymbol{\mathcal{L}}_1^s$ of the form $\boldsymbol{v} = \boldsymbol{D}_2^{1/2} \boldsymbol{B}_1^\top \boldsymbol{D}_1^{-1/2} \boldsymbol{u}$ with eigenvalue λ.*
2. *Every eigenvector $\boldsymbol{u}$ of $\boldsymbol{G}_2$ with eigenvalue λ has a corresponding eigenvector of $\boldsymbol{\mathcal{L}}_1^s$ of the form $\boldsymbol{v} = \boldsymbol{D}_2^{-1/2} \boldsymbol{B}_2 \boldsymbol{D}_3^{1/2} \boldsymbol{u}$ with eigenvalue λ.*

The matrix $\boldsymbol{G}_1 \in \mathbb{R}^{n_0 \times n_0}$ involves only lower-adjacent couplings of the edges and has the form of a weighted graph Laplacian (recall that the the standard graph Laplacian is $\boldsymbol{L}_0 = \boldsymbol{B}_1 \boldsymbol{B}_1^T$). Similarly, $\boldsymbol{G}_2 \in \mathbb{R}^{n_2 \times n_2}$ is completely determined by weighted upper-adjacent couplings of edges. In both cases, the corresponding eigenvectors of $\boldsymbol{\mathcal{L}}_1^s$ can be understood in terms of the Hodge decomposition above (see (3.16)). In particular, the above result shows that eigenvectors of $\boldsymbol{G}_1$ or $\boldsymbol{G}_2$ with near-zero eigenvalues correspond to eigenvectors of $\boldsymbol{\mathcal{L}}_1^s$ with near-zero eigenvalues (i.e., nearly harmonic functions).

Proof. Recall that $\boldsymbol{\mathcal{L}}_1^s$ is defined as

$$\boldsymbol{\mathcal{L}}_1^s = \boldsymbol{D}_2^{1/2} \boldsymbol{B}_1^\top \boldsymbol{D}_1^{-1} \boldsymbol{B}_1 \boldsymbol{D}_2^{1/2} + \boldsymbol{D}_2^{-1/2} \boldsymbol{B}_2 \boldsymbol{D}_3 \boldsymbol{B}_2^\top \boldsymbol{D}_2^{-1/2}. \tag{3.18}$$

Let $\boldsymbol{G}_1 \boldsymbol{u} = \lambda \boldsymbol{u}$ and $\boldsymbol{v} = \boldsymbol{D}_2^{1/2} \boldsymbol{B}_1^\top \boldsymbol{D}_1^{-1/2} \boldsymbol{u}$. Then

$$\begin{aligned} \boldsymbol{\mathcal{L}}_1^s \boldsymbol{v} &= \boldsymbol{D}_2^{1/2} \boldsymbol{B}_1^\top \boldsymbol{D}_1^{-1} \boldsymbol{B}_1 \boldsymbol{D}_2^{1/2} \boldsymbol{v} = \boldsymbol{D}_2^{1/2} \boldsymbol{B}_1^\top \boldsymbol{D}_1^{-1} \boldsymbol{B}_1 \boldsymbol{D}_2 \boldsymbol{B}_1^\top \boldsymbol{D}_1^{-1/2} \boldsymbol{u} \\ &= \boldsymbol{D}_2^{1/2} \boldsymbol{B}_1^\top \boldsymbol{D}_1^{-1/2} \boldsymbol{G}_1 \boldsymbol{u} = \lambda \boldsymbol{D}_2^{1/2} \boldsymbol{B}_1^\top \boldsymbol{D}_1^{-1/2} \boldsymbol{u} = \lambda \boldsymbol{v}. \end{aligned}$$

The proof of the second statement is analogous. ∎

4. Constructing Simplicial Complexes and Computation. Before delving into applications, we first discuss computational aspects of our diffusion framework. Computational costs associated with higher-order interactions are larger than those of

graph-based techniques, but often not prohibitively so. SCs built from data typically induce sparse (co)boundary matrices. Once the SC is constructed, the computations boil down to sparse matrix-vector products, where the sparsity is linear in the number of elements of the SC. This holds when we move to even higher order Hodge Laplacians, although the size of the SC can grow.

4.1. Constructing Simplicial Complexes from Data. We have thus far assumed that we are given an SC. However, an SC $\mathcal{X}$ is typically derived from data in some manner. There are several ways this is done in practice:

1. The original data is a collection of sets $\mathcal{S}$, and we induce $\mathcal{X}$ from elements of $\mathcal{S}$. For example, $\mathcal{X}$ could be induced by all sets in $\mathcal{S}$ with cardinality at most three. Alternatively, we could also include size-3 subsets of sets in $\mathcal{S}$ containing more than three elements. This practice has been used when, for example, the elements of $\mathcal{S}$ are sets of authors on scientific publications [103] or sets of tags annotating questions on Stack Overflow [10].
2. The original data is a point cloud in a metric space, and we use geometric methods to construct, e.g., a Vietoris–Rips complex or a Čech complex. This is standard practice in persistent homology [29], where the data might be a time series of synaptic firings in a brain [57] or a set of images [84].
3. The data is a graph, and $\mathcal{X}$ is the clique complex, where 2-simplices are the 3-cliques in the graph. The 0- and 1-simplices are given by the graph [72, 85].

The first case is a "top-down" construction of the SC, while the second case is a "bottom-up" approach. The third case is somewhere in between—the graph structure imposes the 1-skeleton, and the 1-skeleton contains all of the information for the SC.

The computational complexity of constructing the SC differs in each case. When the data is already a collection of sets, one might only need to process sets one by one. For the case where the data is a point cloud, finding fast algorithms for constructing SCs (or sequences of SCs) is an active area of research [31, 34, 43, 67, 98, 127]. Likewise, enumerating triangles in a graph to construct a clique complex is a well-studied problem [35, 82]. In the worst case, the running time is $O(n_1^{3/2})$, where n_1 is the number of edges (1-simplices). However, practical algorithms are typically faster on real-world graph data exhibiting common structural properties [15, 82].

4.2. Solving $\mathcal{L}_1$ Systems and the Cost of Matrix-Vector Multiplication. Now suppose we have constructed an SC with a maximum simplex size of three. Computationally, our applications will solve systems and compute eigenvectors of matrices that involve the normalized Hodge Laplacian. In typical data applications (including those we study later), the SC induces a sparsely representable normalized Hodge Laplacian. Often, approximate solutions are sufficient for data analysis, so iterative methods are a natural choice in our computations. The driving factor in the running time of the computation (ignoring issues of conditioning) is the cost of a matrix-vector product of the normalized Hodge Laplacian.

Recall that the normalized Hodge 1-Laplacian is

$$\boldsymbol{\mathcal{L}}_1 = \boldsymbol{D}_2\boldsymbol{B}_1^\top\boldsymbol{D}_1^{-1}\boldsymbol{B}_1 + \boldsymbol{B}_2\boldsymbol{D}_3\boldsymbol{B}_2^\top\boldsymbol{D}_2^{-1},$$

where the matrices $\boldsymbol{D}_i$ are simple diagonal matrices that can be directly computed from $\boldsymbol{B}_1$ and $\boldsymbol{B}_2$ with a single matrix vector product. Suppose we have constructed $\boldsymbol{B}_1$ and $\boldsymbol{B}_2$. Then the cost of applying the matrix $\boldsymbol{\mathcal{L}}_1$ to a vector involves a matrix vector product with the diagonal matrices $\boldsymbol{D}_i$ (a simple scaling), and the matrices $\boldsymbol{B}_2^\top$, $\boldsymbol{B}_2$, $\boldsymbol{B}_1$, and $\boldsymbol{B}_1^\top$. These incidence matrices have $O(n_1 + n_2)$, $O(n_1 + n_2)$, $O(n_0 + n_1)$,

and $O(n_0 + n_1)$ nonzeros,[4] respectively, where n_0 is the number of 0-simplices (nodes), n_1 is the number of 1-simplices (edges), and n_2 is the number of 2-simplices (filled triangles). Putting everything together, we can compute a matrix-vector product of the normalized Hodge Laplacian in time $O(n_0 + n_1 + n_2) = O(|\mathcal{X}|)$, i.e., linear time in the size of the data. We note that there is also a growing literature on fast solvers for Laplacian systems [123] with some results for Hodge 1-Laplacians [39].

The other major computational component for applications is the Hodge decomposition of edge flows, which we discuss next.

4.3. Computation and Interpretation of the Hodge Decomposition for Edge Flows. As discussed in subsection 2.6 the Hodge decomposition is an orthogonal decomposition of a vector space. We use the normalized decomposition in (3.16) to provide additional insights into our upcoming applications. The normalized Hodge decomposition of a vector $\boldsymbol{c} \in \mathbb{R}^{n_1}$ (an edge flow) is

$$\boldsymbol{c} = \boldsymbol{g} \oplus \boldsymbol{r} \oplus \boldsymbol{h}, \quad \text{where} \quad \boldsymbol{g} = \boldsymbol{D}_2^{1/2}\boldsymbol{B}_1^\top \boldsymbol{p}, \quad \boldsymbol{r} = \boldsymbol{D}_2^{-1/2}\boldsymbol{B}_2\boldsymbol{w}, \quad \boldsymbol{\mathcal{L}}_1^s \boldsymbol{h} = 0. \tag{4.1}$$

Since the decomposition is orthogonal, computing the decomposition boils down to solving least squares problems:

$$\min_{\boldsymbol{p}} \|\boldsymbol{D}_2^{1/2}\boldsymbol{B}_1^\top \boldsymbol{p} - \boldsymbol{c}\|_2, \qquad \min_{\boldsymbol{w}} \|\boldsymbol{D}_2^{-1/2}\boldsymbol{B}_2\boldsymbol{w} - \boldsymbol{c}\|_2. \tag{4.2}$$

One must exercise a bit of caution here. Although these least squares problems are typically overdetermined, $\boldsymbol{D}_2^{1/2}\boldsymbol{B}_1^\top$ and $\boldsymbol{D}_2^{-1/2}\boldsymbol{B}_2$ are rank-deficient exactly when $\ker(\boldsymbol{\mathcal{L}}_1^s)$ is nontrivial (or, equivalently, $\ker(\boldsymbol{L}_1)$ is nontrivial), i.e., when the SC has a nontrivial first cohomology group $\mathcal{H}_1$. For our purposes, we do not actually need to recover $\boldsymbol{p}$ or $\boldsymbol{w}$; we only need the residuals of the least squares problems. Let $\boldsymbol{e}_p = \boldsymbol{D}_2^{1/2}\boldsymbol{B}_1^\top \boldsymbol{p}^* - \boldsymbol{c}$ and $\boldsymbol{e}_w = \boldsymbol{D}_2^{-1/2}\boldsymbol{B}_2\boldsymbol{w}^* - \boldsymbol{c}$ be the residual error vectors for the least squares problems in (4.2) (with any minimizers $\boldsymbol{p}^*$ and $\boldsymbol{w}^*$). Then the Hodge decomposition is given as follows:

$$\boldsymbol{g} = \boldsymbol{e}_p + \boldsymbol{c} = \boldsymbol{D}_2^{1/2}\boldsymbol{B}_1^\top \boldsymbol{p}^*, \quad \boldsymbol{r} = \boldsymbol{e}_w + \boldsymbol{c} = \boldsymbol{D}_2^{-1/2}\boldsymbol{B}_2\boldsymbol{w}^*, \quad \boldsymbol{h} = \boldsymbol{c} - \boldsymbol{g} - \boldsymbol{r}. \tag{4.3}$$

As discussed above, both $\boldsymbol{B}_1$ and $\boldsymbol{B}_2$ are sparse, and an approximate solution is often satisfactory. Thus, appropriate numerical methods for minimum-length linear least squares problems are iterative solvers such as LSQR [99] and LSMR [49], which produce sequences of residual error vectors. The running time and computational complexity of these algorithms is largely driven by the sparsity of the matrices. If the SC is small enough, then one could first compute the Moore–Penrose pseudoinverses and then compute the projections; in this case, the computational complexity is dominated by the cost of computing the pseudoinverse, which is $O(n_1 n_0^2 + n_2 n_1^2)$.

The components of the Hodge decomposition for edge flows are related to notions from vector calculus [85]. The vector $\boldsymbol{g}$ is the projection of $\boldsymbol{c}$ into $\mathrm{im}(\boldsymbol{D}_2^{1/2}\boldsymbol{B}_1^\top)$, which is a weighted cut-space of the edges [61, 64, 114], i.e., the linear combinations of weighted edge vectors that disconnect the network. Equivalently, $\boldsymbol{g}$ is a weighted gradient flow—the "unweighted" flow $\boldsymbol{D}_2^{-1/2}\boldsymbol{g}$ has no cyclic component, meaning that the sum of its flow along any cyclic path in the complex is zero, taking into

[4] We use $O(\cdot)$ notation here to emphasize that the number of nonzero entries scales linearly in the size of the simplices, even though the exact number of nonzeros could be given. For instance, the matrix $\boldsymbol{B}_1$ has exactly $2n_1$ nonzero entries.

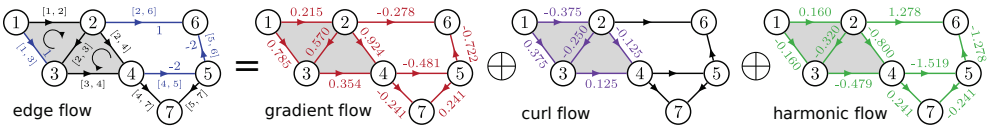

Fig. 4.1 *Standard (unnormalized/unweighted) Hodge decomposition* (2.3) *of the edge flow in the example from Figure* 2.1. *The edge flow can be decomposed into orthogonal gradient, curl, and harmonic flows. Gradient flows around cycles sum to zero; curl flows are comprised of net flows around* 2*-simplices; and harmonic flows sum to zero around* 2*-simplices but are nonzero along longer cycles. Decomposing edge flows with the Hodge decomposition provides additional insights for data analysis applications. While the unnormalized decomposition is illustrative, in practice, we use the normalized Hodge decomposition* (3.16).

account the orientations of the edges (Figure 4.1, second from left). The vector $\boldsymbol{r}$ is the projection of $\boldsymbol{c}$ into $\mathrm{im}(\boldsymbol{D}_2^{-1/2}\boldsymbol{B}_2)$, which consists of all weighted flows that can be composed of local circulations along any 3-node simplex, i.e., weighted circulations around filled triangles (the unweighted version is in Figure 4.1, second from right). Indeed, the operator $\boldsymbol{B}_2$ is a discrete analogue of the familiar notion of a curl in vector calculus [85]. A high projection into the curl subspace thus corresponds to a flow that is mostly composed of local circulations. Finally, the harmonic component $\boldsymbol{h} \in \ker(\boldsymbol{\mathcal{L}}_1^s)$ corresponds to a weighted version of a global circulation that does not sum to zero around every cyclic path but is also inexpressible as a linear combination of curl flows (the unweighted case is in Figure 4.1, far right). A flow with a high projection into the harmonic subspace is thus associated with global cycles within the edge-space that can be directly related to the homology of the SC.

5. Application I: Edge Flow and Trajectory Embeddings. The graph Laplacian with its connection to diffusion processes, harmonic analysis, and algebraic topology has been employed in many learning tasks, including manifold learning, dimensionality reduction, graph clustering, and graph signal processing [9, 40, 87, 96]. Underpinning these methods is the spectral structure of the Laplacian. Eigenvectors associated with zero eigenvalues are associated with the zeroth homology group of the graph, corresponding to connected components. Eigenvectors with eigenvalues close to zero correspond to nearly disconnected components (clusters), as can be quantified by the celebrated Cheeger inequality. By assessing the spectral properties of the graph Laplacian we can thus obtain an approximate notion of the topology, which embodies many spectral embedding and clustering techniques [9, 40]. In the following, we translate these ideas to the context of the Hodge 1-Laplacian by considering spectral embeddings of edges and flows.

5.1. Synthetic Data Example. We introduce our ideas with a synthetic example. Consider a flow $\boldsymbol{f}$ defined on the edges of an SC as depicted in Figure 5.1A. The underlying SC was constructed by (i) drawing 400 random points in the unit square; (ii) generating a triangular lattice via Delaunay triangulation; (iii) eliminating edges inside two predefined regions; and (iv) defining all triangles to be faces. As depicted in Figure 5.1A, the SC $\mathcal{X}$ has two "holes." Accordingly, the (normalized) Hodge Laplacian has exactly two zero eigenvalues. As depicted in Figures 5.1B and 5.1C, these eigenvalues can be associated to two harmonic functions $\boldsymbol{h}_1$ and $\boldsymbol{h}_2$ that encircle the two holes in the SC. To avoid differentiating between left and right eigenvectors, we use the symmetrized Laplacian $\boldsymbol{\mathcal{L}}_1^s$ here and in the remainder of this section, although all results can be translated to $\boldsymbol{\mathcal{L}}$ using the spectral relationships from subsection 3.4.

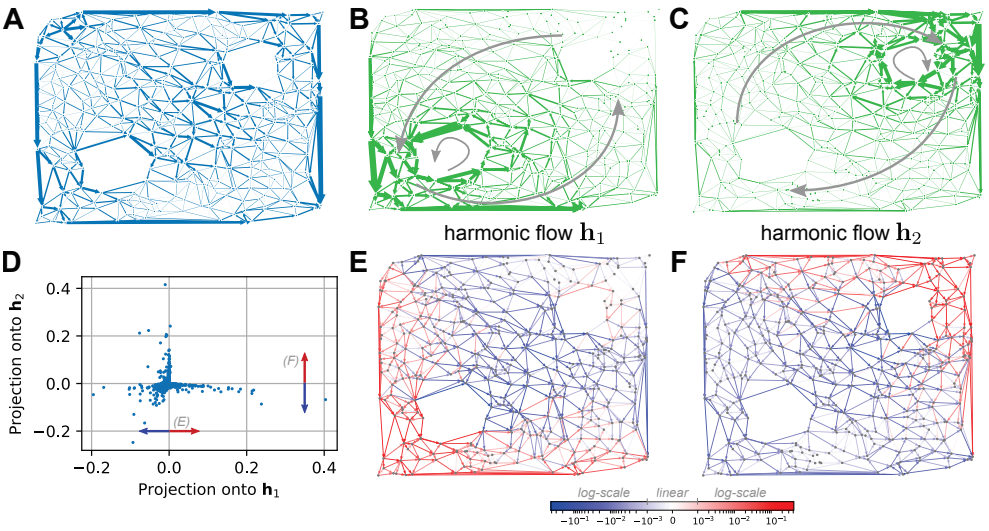

Fig. 5.1 Embedding of edge flows. (A) *A flow on an SC with two "holes," constructed as described in the text. Arrows indicate the direction of the flow on each edge; the magnitude is proportional to the width of the edge.* (B)–(C) *Harmonic functions* $\boldsymbol{h}_1$ *and* $\boldsymbol{h}_2$ *of the symmetric normalized Hodge* 1*-Laplacian* $\boldsymbol{\mathcal{L}}_1^s$ *of the underlying SC. Edge directions correspond to the orientation induced by each harmonic function. Gray arrows indicate how the harmonic flows encircle the two holes.* (D) *Projection of each edge flow* $f_{[i,j]}$ *(depicted in Figure* 5.1A*) onto the harmonic functions.* (E)–(F) *Projection of the edge flow onto the harmonic functions* $\boldsymbol{h}_1$ (E) *and* $\boldsymbol{h}_2$ (F)*. Red indicates a positive projection, blue a negative projection. The arrow direction is the same as in* (A).

Edge Flow Embeddings. Following ideas from Laplacian eigenmaps [9] and diffusion maps [40] for the embedding of nodes of a graph, consider the spectral decomposition of the symmetric normalized Hodge 1-Laplacian $\boldsymbol{\mathcal{L}}_1^s = \boldsymbol{U}\boldsymbol{\Lambda}\boldsymbol{U}^\top$. Here, $\boldsymbol{U} = (\boldsymbol{u}_1, \dots, \boldsymbol{u}_{n_1})$ is the matrix containing the eigenvectors of $\boldsymbol{\mathcal{L}}_1^s$ and $\boldsymbol{\Lambda} = \operatorname{diag}(\lambda_1, \dots, \lambda_{n_1})$ is the diagonal matrix of eigenvalues, where we assume that the eigenvalues have been ordered in increasing magnitude $0 \le \lambda_1 \le \lambda_2 \le \cdots \le \lambda_{n_1}$. For a Laplacian $\boldsymbol{\mathcal{L}}_1^s$ with k zero eigenvalues, we define $\boldsymbol{H} := (\boldsymbol{u}_1, \dots, \boldsymbol{u}_k)$ to be the matrix collecting all the harmonic functions associated to $\boldsymbol{\mathcal{L}}_1^s$.[5]

Let us denote the indicator vector of a positive flow on the positively oriented edge $e = [i, j]$ by $\boldsymbol{e}$, i.e., $\boldsymbol{e}_{[i,j]} = 1$, and 0 otherwise. We now define the *harmonic embedding* of an edge e via the mapping

$$e \mapsto \boldsymbol{l}_e = \boldsymbol{H}^\top \boldsymbol{e} \in \mathbb{R}^k. \tag{5.1}$$

The embedding measures how a unit flow along the oriented edge $[i, j]$ projects into the harmonic subspace, i.e., how much it contributes to the global circular flows represented by the harmonic functions in terms of an inner product. The reference orientation of the edge is important in that a positive projection coordinate indicates that the edge is aligned with the harmonic function; a negative coordinate signifies that the orientations are not aligned.

[5]Note that if $\lambda = 0$ is a degenerate eigenvalue, the eigenvectors are only defined up to a unitary transformation and hence not unique; however, the subspace of harmonic functions is unique.

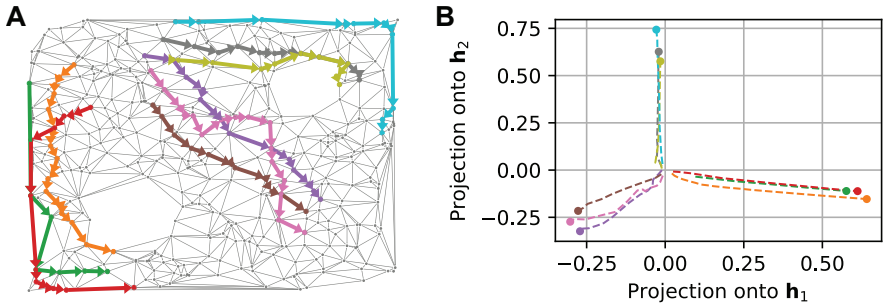

Fig. 5.2 Embedding of trajectories. (A) *A set of nine trajectories, defined on the SC in Figure* 5.1. *Arrows indicate the direction of the flow on each edge.* (B) *Projection of each of the shown trajectories onto the harmonic eigenvectors of* $\boldsymbol{\mathcal{L}}_1^s$.

Figure 5.1A shows a flow $\boldsymbol{f}$ on the SC. For each edge $e = [i, j]$, we construct the weighted indicator vector $\boldsymbol{f}_{[i,j]} = f([i, j])\boldsymbol{e}$ whose value is the amount of flow on the edge relative to the chosen reference orientation. We then compute the projection of this vector into the harmonic subspace $\boldsymbol{l}_f = \boldsymbol{H}^\top \boldsymbol{f}_{[i,j]}$. Figure 5.1D shows the embedding for each edge in the SC with respect to the two harmonic flows of the SC (Figures 5.1B and 5.1C). Edges with a positive projection onto $\boldsymbol{h}_1 = \boldsymbol{u}_1$ are primarily aligned with flows that encircle the lower hole in the complex in the counterclockwise direction (red edges) and edges with a negative projection (blue edges) contribute to clockwise rotations (Figure 5.1E). An analogous argument holds, *mutatis mutandis*, for the second coordinate corresponding to the projection onto $\boldsymbol{h}_2$ (Figure 5.1F). We emphasize that our explanation is geometric, but the extracted features derive solely from the topological information encoded in the SC.

We could use embedding coordinates other than the harmonic ones. For instance, we may choose to project into the curl subspace or the gradient subspace, thereby revealing complementary information, such as how much the edge is aligned with the cut-space. Alternatively, similar to Laplacian eigenmaps [9], we may project onto the first k' eigenvectors of $\boldsymbol{\mathcal{L}}_1^s$, where k' may be different from the size of the harmonic subspace. Such a procedure would also account for contributions of an edge into parts of the gradient and the curl subspaces, namely, those associated with small eigenvalues (i.e., near-harmonic functions [90]).

Trajectory Embeddings. Inspired by Gosh et al. [54], we now consider the embedding of "trajectories" of flow vectors $\boldsymbol{f}$ defined on contiguous edges into the harmonic space. Figure 5.2A displays the same SC as above with a set of nine trajectories. Each of these trajectories is represented by a vector $\boldsymbol{f}$ with entries $f_{[i,j]} = 1$ if $[i, j]$ is part of the trajectory; $f_{[i,j]} = -1$ if $[j, i]$ is part of the trajectory; and $f_{[i,j]} = 0$, otherwise. Figure 5.2B shows the embedding of the complete flow vector along with its "temporal evolution" (dashed lines) in the embedding space, where we update the embedding one edge at a time, leading to a trajectory in embedding space. The embedding differentiates the topological properties of the trajectories. For example, the red, orange, and green trajectories traverse the lower left obstacle aligned with $\boldsymbol{h}_1$. Consequently, they have a similar embedding. Similarly, the brown, pink, and violet trajectories are clustered together in the embedding space, as are the cyan, gray, and olive green trajectories, reflecting their similarity in terms of their projections onto the harmonic subspace.

5.2. Analysis of Ocean Drifter Data. The above idea of trajectory embeddings can be used for a number of tasks in data analysis, similar to node embeddings [9, 40]. This includes clustering trajectories according to their relative positions in the embedding space, definitions of similarity scores between trajectories (even if they have different lengths), and filtering noisy trajectories [115].

The type of data for which these embeddings can be useful includes any type of flow on some discrete (or discretized) domain. For instance, this could be trajectories measured in two-dimensional physical space. To construct an SC from such data we can, e.g., discretize such a flow on a hexagonal grid and choose each hexagonal cell to correspond to a node. An edge is then defined if the number of trajectories crossing from one cell to another exceeds a certain threshold; and a face is defined if sufficiently many trajectories pass through all three neighboring cells. A missing face thus corresponds to an obstacle through which little flow passes.

This style of trajectory analysis was recently used to analyze mobility data with differential forms [54]. In contrast to this work, our formalism depends only on the construction of the SC. Therefore, our methodology is not limited to planar SCs, even though we use planar examples for easy visualization. Indeed, trajectory data collected in applications often has no explicit geometry, but may be understood as a sequence of nodes on an SC (or graph), e.g., moving from one webpage to another [13]. The construction of SCs from such data is a modeling problem, where edges and faces can be chosen as a function of the observed flow pattern. For example, we may create a face from a triangle of nodes if there is enough cyclic flow within this triangle, or we may choose to assign weights to faces to regularize certain aspects of the data [62, Chapter 4]. The success of such methods will be context-dependent and contingent on whether the SC model provides a useful interpretation of the data.

We now apply our harmonic embedding technique for analyzing data from the Global Ocean Drifter Program available at the AOML/NOAA Drifter Data Assembly Center.[6] This data has been analyzed for detecting Lagrangian coherent structures [52], where it was shown that certain flow structures (ocean current) stay coherent over time. While the entire dataset spans several decades of measurements, we focus on data from January 2011–June 2018 and limit ourselves to buoys that were active for at least three months within that time period. We construct trajectories by considering the location information of every buoy every 12 hours. As buoys may fail to record a position, there are trajectories with missing data. In these cases, we split the trajectories into multiple contiguous trajectories. For our analysis, we examine trajectories around Madagascar with a latitude $y_{\text{lat}} \in [-30, -10]$ and longitude $x_{\text{long}} \in [39, 55]$ (Figure 5.3). This results in 400 total trajectories.

To construct an SC, we first transform the data into Euclidean coordinates via an area-preserving (Lambert) projection. We discretize Euclidean space using a hexagonal grid as shown in Figure 5.3A, with the width of the hexagon equal to 1.66° (latitude). Each hexagon corresponds to a node, and we add an edge between two such nodes if there is a nonzero net flow from one hexagon to its adjacent neighbors. We consider all triangles (3-cliques) in this graph to be faces of the SC. The Laplacian $\boldsymbol{\mathcal{L}}_1^s$ of the resulting SC has a two-dimensional harmonic space. Each dimension of this space corresponds to an "obstacle" of the flow.

Finally, we discretize each trajectory by rounding its positional coordinates to the nearest hexagon and consider the resulting sequence of edges that the trajectory traverses in the SC. Figures 5.3B and 5.3C show the results of projecting each trajectory

[6] http://www.aoml.noaa.gov/envids/gld/

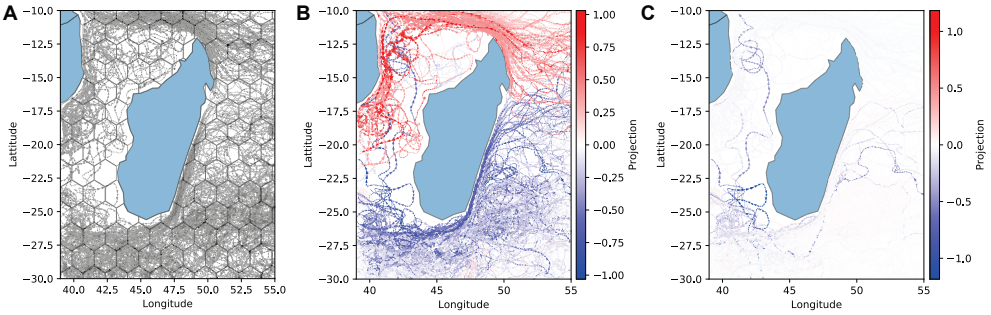

Fig. 5.3 Harmonic embeddings of ocean drifter trajectories. (A) *Visualization of buoy trajectory data around Madagascar. Trajectories are discretized by a hexagonal grid shown in the background.* (B) *Projection of the discretized trajectories onto the first harmonic flow. Trajectories encircling Madagascar in a clockwise direction have a strongly negative projection (blue), and trajectories encircling in a counterclockwise direction have a positive projection (red).* (C) *Projection of the individual trajectories onto the second harmonic flow, which corresponds to an almost local circulation near Southwest Madagascar. Only a small number of trajectories have a large projection onto this flow.*

into the harmonic subspace of $\boldsymbol{\mathcal{L}}_1^s$, where we color the original trajectories according to the projection score. Figure 5.3B shows that the first harmonic function captures the effect of Madagascar as an island. Specifically, the first harmonic function separates the south equatorial current arriving at East of Madagascar into (i) the current flowing toward the north of Madagascar and (ii) the East Madagascar current which flows southwards. The second harmonic function corresponds to a more localized feature (Figure 5.3C) caused by a loopy current near Southwest Madagascar. Accordingly, most trajectories have a small projection onto this space.

6. Application II: PageRank on Simplicial Complexes. Centrality measures, initially conceptualized to quantify the social power of individuals within social networks [21,24,50,51,74], are an important network analysis tool. For instance, centrality measures have been used to identify pivotal elements in infrastructure networks [3, 28] and to target critical nodes in epidemic spreading processes on networks [76,102]. One of the most widely adopted centrality measures is PageRank, which can be interpreted in terms of a random walk on a graph [27]. Initially introduced as a ranking mechanism for hyperlinked webpages, PageRank has been extensively used and studied in other contexts [16,59,81]. For example, (personalized) PageRank has been used for graph clustering, community detection, and semisupervised learning [6,59,60,79,113].

Analogous to the situation of graphs, we would like to assess the importance of certain simplices within an SC, i.e., extend the theory of centrality measures for graphs. One could use simple notions such as the degree of a simplex, but extending other notions of centrality to SCs is nontrivial [46]. Here we leverage the connection of PageRank to random walks to derive a PageRank measure from the $\boldsymbol{\mathcal{L}_1}$ Laplacian that extracts the topological importance of edges in an SC.

6.1. Background: (Personalized) PageRank on Graphs. Following Gleich [59], we adopt the following definition of PageRank.

DEFINITION 6.1 (PageRank on graphs [59]). *Let $\boldsymbol{P}$ be a column-stochastic matrix, $\boldsymbol{\mu}$ a stochastic column vector with $\mathbf{1}^\top \boldsymbol{\mu} = 1$, and $\alpha \in (0,1)$ a teleportation parameter.*

The PageRank vector $\boldsymbol{\pi}$ is the solution to the following linear system:

$$(\boldsymbol{I} - \alpha \boldsymbol{P})\boldsymbol{\pi} = (1-\alpha)\boldsymbol{\mu}. \tag{6.1}$$

The PageRank vector $\boldsymbol{\pi}$ is the stationary distribution of a random walker on a graph who at each step makes transitions according to $\boldsymbol{P}$ with probability α, and with probability $1-\alpha$ teleports to a random node according to the probability distribution $\boldsymbol{\mu}$. The factor $1-\alpha$ facilitates the random walk interpretation but is often omitted as it is simply multiplicative scaling [59].

There are two common types of PageRank [59]. In standard PageRank [27], the teleportation distribution is uniform ($\boldsymbol{\mu} = \mathbf{1}/n_0$) and the PageRank vector $\boldsymbol{\pi}$ is used to rank nodes. In "personalized" PageRank, $\boldsymbol{\mu}$ is an indicator vector on a node i, so with probability $1-\alpha$ we restart our random walk process at i. The resulting PageRank vector $\boldsymbol{\pi}$ can be interpreted as the influence node i exerts on others: the jth entry of $\boldsymbol{\pi}$ is large if i is well connected to j. One can thus find nodes tightly coupled to i, a feature that can be employed for local community detection [6, 79] and graph-based semisupervised learning [60].

6.2. PageRank Vectors on Simplicial Complexes. To generalize PageRank to SCs, we consider the problem on the lifted random walk matrix $\widehat{\boldsymbol{P}}$ from Theorem 3.4:

$$(\boldsymbol{I} - \alpha \widehat{\boldsymbol{P}})\widehat{\boldsymbol{\pi}} = (1-\alpha)\boldsymbol{\mu}. \tag{6.2}$$

As we are interested in the subspace of alternating functions within the lifted space, we project the resulting PageRank vector back into $\mathcal{C}_1$. Specifically, the vector $\boldsymbol{V}^\top \widehat{\boldsymbol{\pi}}$ gives the simplicial PageRank value on oriented edges. We solve for this vector

$$(1-\alpha)\boldsymbol{V}^\top \boldsymbol{\mu} = \boldsymbol{V}^\top (\boldsymbol{I} - \alpha \widehat{\boldsymbol{P}})\widehat{\boldsymbol{\pi}} = \left(\boldsymbol{V}^\top + \frac{\alpha}{2}\boldsymbol{\mathcal{L}}_1 \boldsymbol{V}^\top\right)\widehat{\boldsymbol{\pi}} = \left(\boldsymbol{I} + \frac{\alpha}{2}\boldsymbol{\mathcal{L}}_1\right)\boldsymbol{V}^\top \widehat{\boldsymbol{\pi}}, \tag{6.3}$$

where we used part (1) of Corollary 3.5 in the second equality. Hence, we can compute the projected PageRank vector using $\boldsymbol{\mathcal{L}}_1$ and never have to construct $\widehat{\boldsymbol{P}}$.

Equivalently, for $\beta = 2/\alpha$, we can write the system for $\boldsymbol{V}^\top \widehat{\boldsymbol{\pi}}$ as

$$\boldsymbol{V}^\top \widehat{\boldsymbol{\pi}} = (\beta - 2)(\beta \boldsymbol{I} + \boldsymbol{\mathcal{L}}_1)^{-1} \boldsymbol{V}^\top \boldsymbol{\mu}. \tag{6.4}$$

It is insightful to compare the projected PageRank vector in (6.4) with the graph-based PageRank once more. Note that we can rewrite Definition 6.1 as

$$\boldsymbol{\pi} = (1-\alpha)(\boldsymbol{I} - \alpha \boldsymbol{P})^{-1}\boldsymbol{\mu} = \frac{(1-\alpha)}{\alpha}\left(\frac{1}{\alpha}\boldsymbol{I} - \boldsymbol{P}\right)^{-1}\boldsymbol{\mu} = \beta_0(\beta_0 \boldsymbol{I} + \boldsymbol{\mathcal{L}}_0)^{-1}\boldsymbol{\mu}, \tag{6.5}$$

where $\beta_0 = 1/\alpha - 1$. There is a striking similarity between graph-based PageRank and the projected simplicial PageRank introduced here. Indeed, (6.5) suggests a definition of simplicial PageRank similar to (6.4) on purely notational grounds. However, based on Theorem 3.4, we know that there is a relationship to a random walk, albeit in a lifted state space. While (6.4) is interpretable in terms of a random walk only for $\beta \in (2, \infty)$, the inverse $(\beta \boldsymbol{I} + \boldsymbol{\mathcal{L}}_1)^{-1}$ remains well-defined for smaller positive values of β, as $(\beta \boldsymbol{I} + \boldsymbol{\mathcal{L}}_1)$ is positive definite ($\boldsymbol{\mathcal{L}}_1$ is similar to the positive semidefinite matrix $\boldsymbol{\mathcal{L}}_1^s$, so $\boldsymbol{\mathcal{L}}_1$ is positive semidefinite). We may thus choose to ignore the multiplicative scaling $\beta - 2$, leading to a generalized form of a PageRank vector. These two variants are summarized in the following definition.

DEFINITION 6.2 (PageRank and generalized PageRank vectors for edges in SCs). *Let $\mathcal{X}$ be an SC with normalized Hodge 1-Laplacian $\boldsymbol{\mathcal{L}}_1$, $\boldsymbol{x}$ be a vector of the form $\boldsymbol{x} = \boldsymbol{V}^\top \boldsymbol{\mu}$, where $\boldsymbol{\mu} \in \mathbb{R}^{2n_1}$ is a probability vector, and $\beta \in (2, \infty)$. The PageRank vector $\boldsymbol{\pi}_1$ of the edges is then defined as the solution to the linear system*

$$(\beta \boldsymbol{I} + \boldsymbol{\mathcal{L}}_1)\, \boldsymbol{\pi}_1 = (\beta - 2)\boldsymbol{x}. \tag{6.6}$$

The generalized PageRank vector (for any $\kappa \in \mathbb{R}^+$) is the solution of the linear system

$$(\kappa \boldsymbol{I} + \boldsymbol{\mathcal{L}}_1)\, \boldsymbol{\pi}_1^g = \boldsymbol{x}. \tag{6.7}$$

Since $\boldsymbol{\pi}_1$ corresponds to a projection of a diffusion in a lifted space, the PageRank is effectively a smoothed out version of a distribution $\boldsymbol{\mu}$ in the lifted space $\mathcal{D}^1$ and thus measures how a starting distribution $\boldsymbol{\mu}$ will be shaped by the structure of the SC. Similar to the graph case, certain oriented edges will attract more probability in this process and will in this sense be deemed more important. A key difference is the projection step, which again emphasizes the importance of the orientation. The absolute value of entry $[i, j]$ in $\boldsymbol{\pi}_1$ will be high if there is a large difference in the probability of being at edge $[i, j]$ in the lifted space as compared to $[j, i]$.

Simplicial PageRank Interpretation as a Filter for Edge-Space Signals. To provide further intuition for the above measures, we interpret $\boldsymbol{x}$ as a signal defined in the edge-space, similar to our discussion on edge and trajectory embeddings in section 5. We discuss this issue in terms of the generalized PageRank vector $\boldsymbol{\pi}_1^g$. For any such signal $\boldsymbol{x}$, the PageRank vector corresponds to a transformation of this signal $\boldsymbol{x}$ according to the *generalized PageRank operator* $(\kappa \boldsymbol{I} + \boldsymbol{\mathcal{L}}_1)^{-1}$ for some $\kappa > 0$. To understand how this multiplication acts on $\boldsymbol{x}$, write the spectral decomposition of the Hodge Laplacian as $\boldsymbol{\mathcal{L}}_1 = \boldsymbol{U}_R \boldsymbol{\Lambda} \boldsymbol{U}_L^\top$, where $\boldsymbol{U}_R$ and $\boldsymbol{U}_L$ are the matrices containing the right and left eigenvectors of $\boldsymbol{\mathcal{L}}_1$, respectively. From our discussion at the end of subsection 3.4, we further know that $\boldsymbol{U}_R = \boldsymbol{D}_2 \boldsymbol{U}_L$. The generalized PageRank operator acts as a filter on the signal $\boldsymbol{x}$ by projecting $\boldsymbol{x}$ onto the spectral coordinates of the Laplacian, scaling according to a shifted version of the spectrum (the eigenvalues of $\boldsymbol{\mathcal{L}}_1$ are shifted by κ), and projecting back on the SC:

$$\boldsymbol{\pi}_1 = (\kappa \boldsymbol{I} + \boldsymbol{\mathcal{L}}_1)^{-1} \boldsymbol{x} = \boldsymbol{U}_R\, (\kappa \boldsymbol{I} + \boldsymbol{\Lambda})^{-1}\, \boldsymbol{U}_L^\top \boldsymbol{x} = \boldsymbol{U}_R\, (\kappa \boldsymbol{I} + \boldsymbol{\Lambda})^{-1}\, \boldsymbol{U}_R^\top \boldsymbol{D}_2^{-1} \boldsymbol{x}. \tag{6.8}$$

Thus, there is a close relationship to the embeddings discussed above. However, in contrast to the harmonic embedding, all eigenvectors (modulated by their eigenvalue) are taken into account in the PageRank filtering operation. This type of filtering is analogous to the filtering of signals defined on the nodes of a graph as considered in graph signal processing [96]. The PageRank operator may thus be understood as a filter for *edge signals*. Recent research provides additional filtering perspectives [8, 115] with connections to discrete exterior calculus [62, 71].

Edge Orientations and Personalized Simplicial PageRank. As the projection leading to simplicial PageRank corresponds to a difference of two probabilities, there is no guarantee that a PageRank vector $\boldsymbol{\pi}_1$ has only positive entries. In light of our previous discussions on edge orientation, this again relates to the issue of defining a reference orientation for edges.

We can flip the reference orientation of an edge with a negative entry to obtain a positive PageRank vector entry. The operation of flipping the orientation of an edge can be cast as a gauge transformation or signature-similarity transformation [5] of the

normalized Hodge 1-Laplacian. Let $\boldsymbol{\Theta}_{n_1 \times n_1} = \mathrm{diag}(\theta_1, \ldots, \theta_{n_1})$ be a diagonal matrix with $\theta_i = 1$ for all edges whose orientation is to remain fixed, and $\theta_i = -1$ for those edges whose orientation we would like to reverse. Then $\boldsymbol{\mathcal{L}}_1' = \boldsymbol{\Theta}\boldsymbol{\mathcal{L}}_1\boldsymbol{\Theta}$ describes the normalized Hodge 1-Laplacian corresponding to the SC with the new edge orientations ($\boldsymbol{\Theta}$ defines a similarity transformation, so the spectral properties of $\boldsymbol{\mathcal{L}}_1$ are unaffected; we have merely changed the coordinate system in which flows are measured). Using this transformation, we obtain the following generalized PageRank vector:

$$\boldsymbol{\pi}_1^g = \boldsymbol{\Theta}\left(\kappa \boldsymbol{I} + \boldsymbol{\mathcal{L}}_1\right)^{-1} \boldsymbol{\Theta}\boldsymbol{x}. \tag{6.9}$$

We claim that if $\boldsymbol{x}$ is an indicator vector, then $\boldsymbol{\Theta}$ can be chosen to make $\boldsymbol{\pi}_1^g$ nonnegative. To see this, if $\boldsymbol{x}$ is an indicator vector $\boldsymbol{e}_{[i,j]}$, it will pick out one column of $(\kappa \boldsymbol{I} + \boldsymbol{\mathcal{L}}_1)^{-1}$, which through the action of $\boldsymbol{\Theta}$ can immediately be made nonnegative on all entries except at index $[i, j]$. Since $(\kappa \boldsymbol{I} + \boldsymbol{\mathcal{L}}_1)^{-1}$ is positive definite, the diagonal entry at index $[i, j]$ will be positive. Hence, just as with personalized PageRank on graphs, if $\boldsymbol{x}$ is an indicator vector on edge $[i, j]$, we may interpret the absolute values of the entries of $\boldsymbol{\pi}_1^g$ in terms of the influence edge $[i, j]$ exerts on the edges in the SC. For data that does not induce a natural orientation of the edges (in contrast to our previous examples) we may thus simply use the elementwise absolute value of this "personalized" PageRank vector as an influence measure between the edges.

While this trick of redefining the orientations of the edges also applies if $\boldsymbol{x}$ is not an indicator vector, the situation is a bit more complicated. Since the PageRank operator acts linearly on $\boldsymbol{x}$, we can decompose any vector $\boldsymbol{x}$ into a weighted sum of unit vectors. For each of these unit vectors we can assess the induced PageRank vector in terms of its absolute value. However, due to differences in the sign patterns of the columns of $(\kappa \boldsymbol{I} + \boldsymbol{\mathcal{L}}_1)^{-1}$ the sum of these induced "absolute" (personalized) PageRank vectors is not the same as the absolute value of the PageRank vector associated to $\boldsymbol{x}$, rendering the above interpretation in terms of the influence of the individual edges difficult. In what follows, we will thus concentrate on personalized PageRank vectors with a teleportation vector $\boldsymbol{x}$ localized on a particular edge $[i, j]$. We denote such a personalized PageRank vector by $\boldsymbol{\pi}_1([i, j])$, or $\boldsymbol{\pi}_1^g([i, j])$ for generalized PageRank.

Synthetic Data Example. To illustrate our ideas, we again start with a synthetic example (Figure 6.1A). We construct an SC that consists of four groups of eight nodes (cliques) in a ring configuration, which is connected to a larger clique of 30 nodes. All triangles (3-cliques) in the graph are faces in the SC. Figure 6.1 visualizes three generalized personalized PageRank vectors for $\kappa = 0.001$. For the edge in Figure 6.1B located in the large clique, the generalized PageRank vector is effectively localized on the edge itself. Moreover, the magnitudes of the entries of the vector are small, indicating that its influence within the space of edges is small. The edge in Figure 6.1C corresponds to the "bridge" between the large clique and the ring of small cliques on the right-hand side. While the generalized PageRank vector is also concentrated on a few edges, the magnitudes of its components are substantially larger (since we are using the generalized PageRank, the entries can be substantially larger than 1). Finally, for one of the cycle edges (Figure 6.1D), we see a different picture. In this case, the generalized PageRank vector is also of a large magnitude, but most of the influence of this edge is concentrated on edges around the cycle.

6.3. Decomposing and Aggregating Simplicial PageRank Vectors via the Normalized Hodge Decomposition. By considering again the spectral properties of the normalized Hodge 1-Laplacian $\boldsymbol{\mathcal{L}}_1$, we see that this behavior is a consequence of the

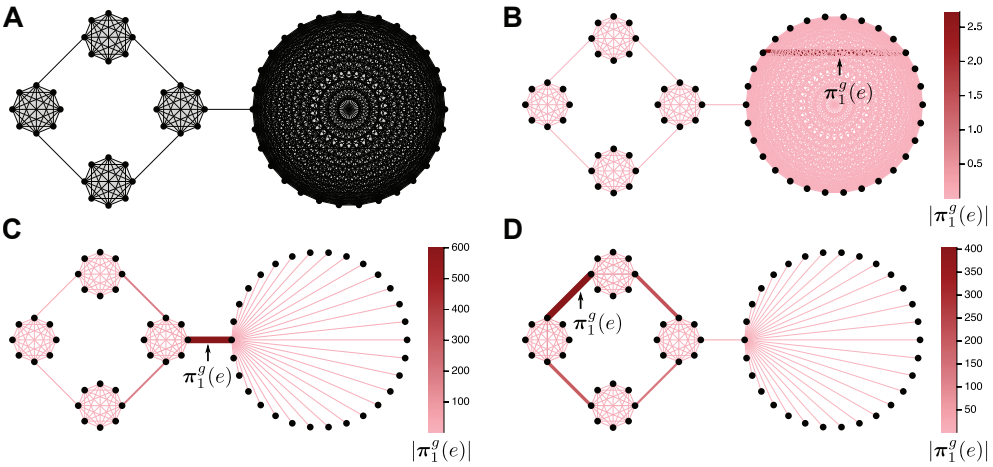

Fig. 6.1 Example analysis of a network using generalized PageRank vectors. (A) *SC constructed from a graph containing five large cliques. Each triangle (3-clique) in the graph corresponds to 2-simplices, as indicated by the gray shading.* (B)–(D) *Generalized personalized PageRank vectors of three edges within different locations of the complex ($\kappa = 0.001$). The color and the edge-width (the thicker the edge, the larger the PageRank) indicate the magnitude of the component of the PageRank vector on the respective edge. The PageRank vector in* (B) *is localized and has small magnitude. The vector in* (C) *is almost localized but has large magnitude. Finally, the vector in* (D) *is not localized and has strong components around the cyclic structure.*

Table 6.1 *Norms of components given by the normalized Hodge decomposition* (4.1) *for the generalized PageRank vectors in Figure* 6.1.

		$\|\boldsymbol{\pi}_1^g\|_{\text{grad}}$	$\|\boldsymbol{\pi}_1^g\|_{\text{curl}}$	$\|\boldsymbol{\pi}_1^g\|_{\text{harm}}$
"Bulk" edge	Figure 6.1B	0.50	2.70	0
"Bridge" edge	Figure 6.1C	707.85	0	0
"Cycle" edge	Figure 6.1D	265.14	0	518.15

topological setup of the SC. As discussed above, the eigenvectors associated to the zero eigenvalues of $\boldsymbol{\mathcal{L}}_1$ correspond to harmonic functions on the complex, which are associated with the cycles in the graph not induced by 2-simplices. As the PageRank vector is computed via a shifted inverse of $\boldsymbol{\mathcal{L}}_1$, edges that have a significant projection into the null space of $\boldsymbol{\mathcal{L}}_1$ will result in a PageRank vector with a strong harmonic (cyclic) component. To better understand the importance of the edges, it is insightful to consider the decomposition of the PageRank vectors in terms of the normalized Hodge decomposition.

We can use the normalized Hodge decomposition from (4.1) to compute the projections of the above-computed PageRank vectors onto the gradient, curl, and harmonic subspaces. This gives a more nuanced picture of the contributions of PageRank vectors (Table 6.1). The edge in Figure 6.1B actually has no harmonic part, and the curl and gradient components largely cancel each other out, apart from the flow on the edge itself. The edge in Figure 6.1C lies in the weighted cut-space. The norm of its corresponding gradient projection is large, and its harmonic and curl projections are zero. Finally, the edge in Figure 6.1D is part of a harmonic cycle in

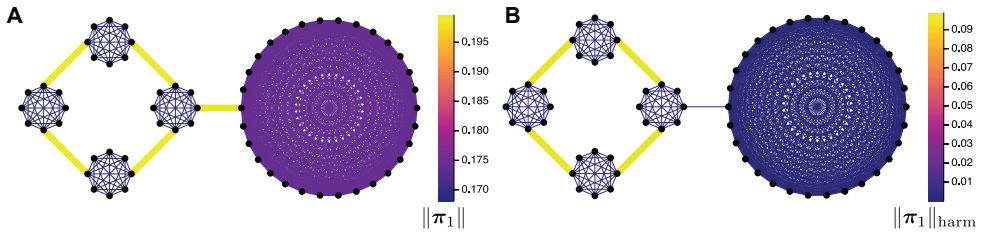

Fig. 6.2 Analyzing edges via norms of personalized PageRank vectors. (A) *The* 2*-norm of the personalized PageRank vector associated to each edge.* (B) *We can alternatively highlight the importance of an edge with respect to the homology of the SC by considering the* 2*-norm of the projection of each PageRank vector into the harmonic space.*

the SC; consequently, the norm of the PageRank vector projected into the harmonic subspace is high.

Instead of decomposing the personalized PageRank vectors associated with each edge, we can assess the importance of an edge within an SC in an aggregated fashion with the 2-norm of the vector, or the 2-norm of any of the projections given by the normalized Hodge decomposition. Figure 6.2 illustrates this procedure ($\beta = 2.5$). In these diagrams, the value along an edge corresponds to the 2-norm of the personalized PageRank vector corresponding to the edge (Figure 6.2A) or to the 2-norm of the harmonic component of the personalized PageRank vector (Figure 6.2B).

Let us remark here that these are clearly not the only ways to extract a "relevance score" of the edges from the computed PageRank vectors, and investigating other functions will be an interesting objective for future work.

6.4. Analysis of Political Book Copurchasing Data. We now analyze a dataset of political book corecommendations. The dataset records the copurchasing of 105 political books—as indicated by the "customers who bought this book also bought these other books" feature on Amazon—around the time of the 2004 presidential election in the USA. The data was collected by Krebs [80][7] and subsequently analyzed by Newman [94]. Newman categorized the books by hand according to their political alignment into three groups: "liberal" (43 books), "conservative" (49 books), and books with bipartisan and centrist views or no clear alignment (13 books), which we refer to as "neutral." Although the data has been analyzed as a network, the edges represent frequent copurchasing of books by the same buyers, and thus the data has an implicit simplicial structure (multiple books are bought together, forming a simplex). To construct an SC, we filled each triangle (3-clique) in the data to form a 2-simplex. Figure 6.3A visualizes this SC, where gray shading indicates the simplices. As already observed by Newman [94], there is a marked community structure, which is commensurate with the political alignment of the books. There are two main clusters corresponding to liberal and conservative books, along with a smaller group of books that act as bridges between these two clusters.

To further analyze the importance of the copurchasing with respect to the homology of the complex, we compute the "harmonic PageRank values" of the edges, which correspond to the 2-norm of the harmonic projection of the personalized PageRank score of each edge ($\beta = 2.5$). Consistent with its importance for the homology, these values highlight edges that act as connectors around "holes," which indicate books

[7] http://vlado.fmf.uni-lj.si/pub/networks/data/cite/polBooks.paj

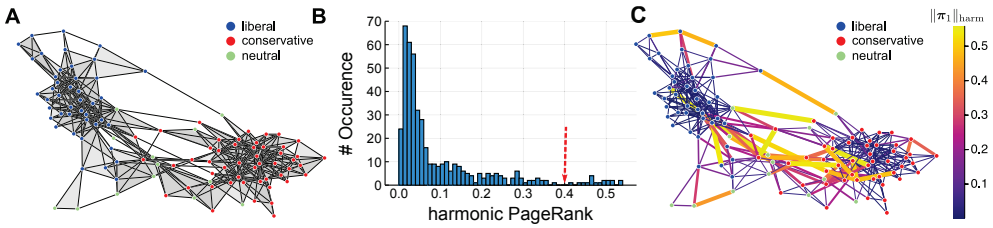

Fig. 6.3 Book copurchasing analysis with simplicial PageRank. (A) *The copurchasing of* 105 *political books—categorized as "liberal," "conservative," or "neutral"—can be abstracted as an SC (gray areas are copurchasing simplices).* (B) *Harmonic PageRank values of edges. We identify a set of* 18 *edges with a high harmonic PageRank value (threshold indicated by dashed red arrow).* (C) *Visualization of all edges with their respective harmonic PageRank values.*

that are never bought together as a set (Figure 6.3). The ranking on edges induced by these harmonic PageRank values is robust with respect to the parameter β: the Spearman rank correlation coefficient ρ between the obtained PageRank vectors for $\beta \in (2.05, 2.67)$ (corresponding to a teleportation parameter $\alpha \in (0.75, 0.975)$) has a mean of 0.78, providing evidence for the consistency of the ranking.

Figure 6.3C shows that even with the separation into political clusters, there are still several edges with high harmonic PageRank within each cluster. Indeed, the edge with the largest aggregated harmonic PageRank corresponds to a connection between two conservative books. To investigate this aspect further, we plotted the histogram of the aggregated harmonic PageRank and identified a tail of 18 copurchases with a harmonic PageRank greater than 0.4 (Figure 6.3B). Out of the 18 edges with the highest harmonic score, 15 are between books of the same category. Edges in the conservative cluster with a high harmonic PageRank are more prevalent: 9 of the 18 highest aggregated harmonic centrality edges connect conservative books. Thus, there appear to be "gaps" in the space of political opinions even within political clusters, indicating a fragmentation of opinions.

Importantly, this information revealed by simplicial PageRank is complementary to standard node-space PageRank (Figure 6.4). We computed the PageRank $\boldsymbol{\pi}$ of the nodes in the graph and assigned each edge (i) the sum or (ii) the difference of the PageRank of its incident nodes. We also compute the PageRank $\boldsymbol{\pi}$ from the line graph corresponding to the copurchasing network. In all cases, we used the standard teleportation parameter $\alpha = 0.85$. As can be seen from Figure 6.4, all of these PageRank scores on edges are essentially uncorrelated to the simplicial (harmonic) PageRank, highlighting that we extract a different kind of topological information from the data than what one could find with existing methodologies.

7. Discussion. The connection between Markov chains, diffusions, and random walks on graphs has led to the successful analysis of complex systems and networks within applied mathematics. However, as we are faced with increasingly complex and diverse datasets, some limitations of traditional graph-based models have come to the fore. Researchers have thus employed richer modeling frameworks—such as multiplex networks, graphons, and simplicial complexes (SCs)—and investigated extensions of graph-based methods such as extensions of centrality measures [7, 46, 117].

SCs and other tools from algebraic and computational topology are promising in this pursuit of extending dyadic network models to account for polyadic interactions

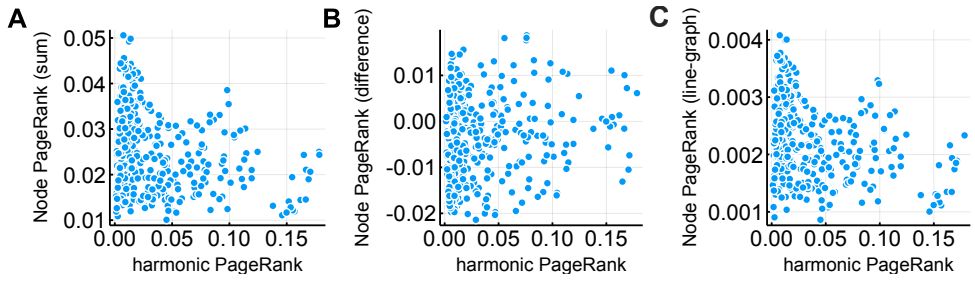

Fig. 6.4 Comparison of simplicial PageRank and graph-based PageRank. *The scatter plots compare the harmonic PageRank (see text) with various edge scores constructed from standard graph-based PageRank.* (A) *Edge scores are the sums of the graph-based PageRank scores of the incident nodes; the Spearman rank correlation is* $\rho = 0.04$. (B) *Edge scores are the differences of the graph-based PageRank of the incident nodes; the Spearman rank correlation is* $\rho = -0.08$. (C) *Edge scores come from graph-based PageRank on the line graph; the Spearman rank correlation is* $\rho = 0.08$.

between groups of nodes, seeing success in a number of applications [32, 42, 56, 58, 92, 104, 121]. However, topological tools have been mostly employed in static contexts, though some works have explored links to dynamical processes on networks [122]. Here we have introduced a normalized Hodge Laplacian operator that enables us to define a diffusion processes on SCs in a principled manner, respecting the topological properties of the complex. In particular, we have focused on diffusion processes in the edge-space with applications in embedding edge trajectory data and simplicial extensions of PageRank for edge importance.

A number of avenues present themselves for future research. There are many variants of random walks we may want to explore and compare to their graph-based counterparts [87]. In particular, a better understanding of how higher-order topological features impact on convergence, mixing, and other random walk properties is of interest in this context. Another direction is to explore the translation of further random walk tools for SCs. We have focused on diffusion-based embeddings and centrality measures. However, random walks have also been employed in many other learning tasks [87]. We expect that enriching such tools through the lens of SCs will be fruitful.

Finally, there is a rich history connecting random walks on graphs to connectivity and near zeroth-order homology, often formalized through ideas from spectral graph theory such as the discrete Cheeger inequality and isoperimetric inequality [4, 36, 78, 83, 88, 116]. For example, personalized PageRank—when viewed as an algorithm for local clustering for graphs—finds low conductance sets, formalized via a local Cheeger inequality [6], which provides credence for our personalized polyadic walks. A major open research direction is a clean generalization of these concepts to SCs. There are a number of connections between the spectra of matrices (such as the Hodge Laplacian) associated with SCs and notions of expansion [44, 65, 97, 101, 119]. However, some types of higher-order Cheeger-like inequalities are impossible due to torsion that may be present in the homology of SCs that is not present in graphs [119]. Properly incorporating near-harmonic components into data analysis remains a challenge, but our framework provides a starting point.

Appendix A. Proof of Theorem 3.4. Recall that the normalized Hodge 1-Laplacian is defined as

$$\mathcal{L}_1 = \boldsymbol{D}_2 \boldsymbol{B}_1^\top \boldsymbol{D}_1^{-1} \boldsymbol{B}_1 + \boldsymbol{B}_2 \boldsymbol{D}_3 \boldsymbol{B}_2^\top \boldsymbol{D}_2^{-1}.$$

Our goal is to show that $-\boldsymbol{\mathcal{L}}_1\boldsymbol{V}^\top = 2\boldsymbol{V}^\top\widehat{\boldsymbol{P}}$, where $\widehat{\boldsymbol{P}}$ is defined in Theorem 3.4. We will prove this result in three steps. First, we show that

$$-(\boldsymbol{D}_2\boldsymbol{B}_1^\top\boldsymbol{D}_1^{-1}\boldsymbol{B}_1)\boldsymbol{V}^\top = \boldsymbol{V}^\top \underbrace{\frac{1}{2}\left(\boldsymbol{M}_f\mathrm{diag}(\boldsymbol{M}_f\mathbf{1})^{-1} + \boldsymbol{M}_b\mathrm{diag}(\boldsymbol{M}_b\mathbf{1})^{-1}\right)}_{\boldsymbol{P}_{lower}}, \tag{A.1}$$

where

$$\boldsymbol{M}_f = \widehat{\boldsymbol{D}}_2(\widehat{\boldsymbol{B}}_1^-)^\top\widehat{\boldsymbol{B}}_1^+ \quad \text{and} \quad \boldsymbol{M}_b = \widehat{\boldsymbol{D}}_2(\widehat{\boldsymbol{B}}_1^+)^\top\widehat{\boldsymbol{B}}_1^- \tag{A.2}$$

are weighted lower-adjacency matrices corresponding to forward and backward walks along the edges (see Theorem 3.4) and $\widehat{\boldsymbol{D}}_2 = \mathrm{diag}(\boldsymbol{D}_2, \boldsymbol{D}_2)$.

Second, we prove that

$$-(\boldsymbol{B}_2\boldsymbol{D}_3\boldsymbol{B}_2^\top\boldsymbol{D}_2^{-1})\boldsymbol{V}^\top = \boldsymbol{V}^\top \underbrace{\left(\widehat{\boldsymbol{A}}_u\widehat{\boldsymbol{D}}_4^{-1} + \frac{1}{2}\begin{pmatrix}\boldsymbol{I} & \boldsymbol{I}\\ \boldsymbol{I} & \boldsymbol{I}\end{pmatrix}\widehat{\boldsymbol{D}}_5\right)}_{\boldsymbol{P}_{upper}}, \tag{A.3}$$

where $\widehat{\boldsymbol{A}}_u = \widehat{\boldsymbol{B}}_2^+(\widehat{\boldsymbol{B}}_2^-)^\top + \widehat{\boldsymbol{B}}_2^-(\widehat{\boldsymbol{B}}_2^+)^\top$ is the matrix of upper-adjacent connections and $\widehat{\boldsymbol{D}}_4$ and $\widehat{\boldsymbol{D}}_5$ are as defined in Theorem 3.4.

Finally, we prove that $\widehat{\boldsymbol{P}} = \frac{1}{2}(\boldsymbol{P}_{lower} + \boldsymbol{P}_{upper})$ is indeed a stochastic matrix.

Preliminary Results. Before embarking on the proof, let us recall some useful facts and develop some lemmas. For any matrix $\boldsymbol{M}$,

$$\boldsymbol{M} = \boldsymbol{M}^+ - \boldsymbol{M}^-, \qquad (-\boldsymbol{M})^- = \boldsymbol{M}^+, \qquad (-\boldsymbol{M})^+ = \boldsymbol{M}^-. \tag{A.4}$$

From some algebra,

$$\boldsymbol{V}\boldsymbol{V}^\top = \boldsymbol{I}_{2n_1} - \boldsymbol{\Sigma}, \tag{A.5}$$

$$\boldsymbol{\Sigma}\boldsymbol{\Sigma} = \boldsymbol{I}_{2n_1}, \tag{A.6}$$

$$\widehat{\boldsymbol{B}}_1 := \boldsymbol{B}_1\boldsymbol{V}^\top = \begin{pmatrix}\boldsymbol{B}_1 & -\boldsymbol{B}_1\end{pmatrix} = \widehat{\boldsymbol{B}}_1^+ - \widehat{\boldsymbol{B}}_1^-, \tag{A.7}$$

$$\widehat{\boldsymbol{B}}_1^-\boldsymbol{\Sigma} = \widehat{\boldsymbol{B}}_1^+ \text{ and } \widehat{\boldsymbol{B}}_1^- = \widehat{\boldsymbol{B}}_1^+\boldsymbol{\Sigma}, \tag{A.8}$$

$$\boldsymbol{B}_1\boldsymbol{V}^\top = \widehat{\boldsymbol{B}}_1^+\boldsymbol{V}\boldsymbol{V}^\top, \tag{A.9}$$

$$\boldsymbol{B}_1 = \widehat{\boldsymbol{B}}_1^+\boldsymbol{V} = -\widehat{\boldsymbol{B}}_1^-\boldsymbol{V}, \tag{A.10}$$

$$\widehat{\boldsymbol{B}}_2 := \boldsymbol{V}\boldsymbol{B}_2 = \begin{pmatrix}\boldsymbol{B}_2 & -\boldsymbol{B}_2\end{pmatrix}^\top = \widehat{\boldsymbol{B}}_2^+ - \widehat{\boldsymbol{B}}_2^-, \tag{A.11}$$

$$\widehat{\boldsymbol{B}}_2^+ = \boldsymbol{\Sigma}\widehat{\boldsymbol{B}}_2^- \text{ and } \boldsymbol{\Sigma}\widehat{\boldsymbol{B}}_2^+ = \widehat{\boldsymbol{B}}_2^-, \tag{A.12}$$

$$\boldsymbol{V}\boldsymbol{B}_2 = \boldsymbol{V}\boldsymbol{V}^\top\widehat{\boldsymbol{B}}_2^+, \tag{A.13}$$

$$\boldsymbol{B}_2 = \boldsymbol{V}^\top\widehat{\boldsymbol{B}}_2^+ = -\boldsymbol{V}^\top\widehat{\boldsymbol{B}}_2^-. \tag{A.14}$$

Now, define the normalizing factors of the forward and backward walks using lower-adjacency relationships as follows:

$$\boldsymbol{Q}_f = \mathrm{diag}(\mathbf{1}^\top\boldsymbol{M}_f) \text{ and } \boldsymbol{Q}_b = \mathrm{diag}(\mathbf{1}^\top\boldsymbol{M}_b).$$

LEMMA A.1. $\boldsymbol{Q}_f\boldsymbol{\Sigma} = \boldsymbol{\Sigma}\boldsymbol{Q}_b$.

Proof. The diagonal entries of $\boldsymbol{Q}_f$ are given by the vector

$$\begin{aligned}
\boldsymbol{q}_f^\top := \mathbf{1}^\top \boldsymbol{M}_f &= \mathbf{1}^\top \widehat{\boldsymbol{D}}_2 (\widehat{\boldsymbol{B}}_1^-)^\top \widehat{\boldsymbol{B}}_1^+ \\
&= \mathbf{1}^\top \widehat{\boldsymbol{D}}_2 \boldsymbol{\Sigma}\boldsymbol{\Sigma} (\widehat{\boldsymbol{B}}_1^-)^\top \widehat{\boldsymbol{B}}_1^+ \\
&= \mathbf{1}^\top \widehat{\boldsymbol{D}}_2 \boldsymbol{\Sigma} (\widehat{\boldsymbol{B}}_1^-)^\top \widehat{\boldsymbol{B}}_1^+ \\
&= \mathbf{1}^\top \widehat{\boldsymbol{D}}_2 (\widehat{\boldsymbol{B}}_1^+)^\top \widehat{\boldsymbol{B}}_1^+ \\
&= \mathbf{1}^\top \begin{pmatrix} \boldsymbol{D}_2 & \mathbf{0} \\ \mathbf{0} & \boldsymbol{D}_2 \end{pmatrix} \begin{pmatrix} (\boldsymbol{B}_1^+)^\top \boldsymbol{B}_1^+ & (\boldsymbol{B}_1^+)^\top \boldsymbol{B}_1^- \\ (\boldsymbol{B}_1^-)^\top \boldsymbol{B}_1^+ & (\boldsymbol{B}_1^-)^\top \boldsymbol{B}_1^- \end{pmatrix} \\
&= \mathbf{1}^\top \begin{pmatrix} \boldsymbol{D}_2 |\boldsymbol{B}_1|^\top \boldsymbol{B}_1^+ & \boldsymbol{D}_2 |\boldsymbol{B}_1|^\top (\boldsymbol{B}_1^-) \end{pmatrix}.
\end{aligned}$$

The first line follows the definition of $\boldsymbol{Q}_f$; the second line uses (A.6); the third line follows from $\mathbf{1}^\top \widehat{\boldsymbol{D}}_2 \boldsymbol{\Sigma} = \mathbf{1}^\top \widehat{\boldsymbol{D}}_2$; the fourth line uses (A.8); and the fifth and sixth lines follows from expanding notation. Similarly, the diagonal entries of $\boldsymbol{Q}_b$ are given by

$$\begin{aligned}
\boldsymbol{q}_b^\top := \mathbf{1}^\top \boldsymbol{M}_b &= \mathbf{1}^\top \widehat{\boldsymbol{D}}_2 (\widehat{\boldsymbol{B}}_1^+)^\top \widehat{\boldsymbol{B}}_1^- \\
&= \mathbf{1}^\top \widehat{\boldsymbol{D}}_2 \boldsymbol{\Sigma}\boldsymbol{\Sigma} (\widehat{\boldsymbol{B}}_1^+)^\top \widehat{\boldsymbol{B}}_1^- \\
&= \mathbf{1}^\top \widehat{\boldsymbol{D}}_2 \boldsymbol{\Sigma} (\widehat{\boldsymbol{B}}_1^+)^\top \widehat{\boldsymbol{B}}_1^- \\
&= \mathbf{1}^\top \widehat{\boldsymbol{D}}_2 (\widehat{\boldsymbol{B}}_1^-)^\top \widehat{\boldsymbol{B}}_1^- \\
&= \mathbf{1}^\top \begin{pmatrix} \boldsymbol{D}_2 & \mathbf{0} \\ \mathbf{0} & \boldsymbol{D}_2 \end{pmatrix} \begin{pmatrix} (\boldsymbol{B}_1^-)^\top \boldsymbol{B}_1^- & (\boldsymbol{B}_1^-)^\top \boldsymbol{B}_1^+ \\ (\boldsymbol{B}_1^+)^\top \boldsymbol{B}_1^- & (\boldsymbol{B}_1^+)^\top \boldsymbol{B}_1^- \end{pmatrix} \\
&= \mathbf{1}^\top \begin{pmatrix} \boldsymbol{D}_2 |\boldsymbol{B}_1|^\top \boldsymbol{B}_1^- & \boldsymbol{D}_2 |\boldsymbol{B}_1|^\top \boldsymbol{B}_1^+ \end{pmatrix} = \boldsymbol{q}_f^\top \boldsymbol{\Sigma}.
\end{aligned}$$

The lemma then follows by considering how the permutation $\boldsymbol{\Sigma}$ acts on the diagonal matrices $\boldsymbol{Q}_f$ and $\boldsymbol{Q}_b$. ∎

LEMMA A.2. $\boldsymbol{D}_1^{-1} \widehat{\boldsymbol{B}}_1^+ = \frac{1}{2} \widehat{\boldsymbol{B}}_1^+ \boldsymbol{Q}_f^{-1}$.

Proof. By definition, $\widehat{\boldsymbol{B}}_1^+ = \begin{pmatrix} \boldsymbol{B}_1^+ & \boldsymbol{B}_1^- \end{pmatrix}$ is an indicator matrix with a single entry equal to 1 per column. More specifically, $\boldsymbol{B}_1^+$ picks out the target nodes of the oriented edges and $\boldsymbol{B}_1^-$ picks out the source nodes:

$$(\boldsymbol{B}_1^+)_{[i],[j,k]} = \begin{cases} 1 & \text{if } i = k, \\ 0 & \text{otherwise,} \end{cases} \qquad (\boldsymbol{B}_1^-)_{[i],[j,k]} = \begin{cases} 1 & \text{if } i = j, \\ 0 & \text{otherwise.} \end{cases}$$

Recall that $\boldsymbol{D}_1 = 2 \cdot \operatorname{diag}(\boldsymbol{a})$, where $\boldsymbol{a} = |\boldsymbol{B}_1| \boldsymbol{D}_2 \mathbf{1}$. Thus,

$$\left[\boldsymbol{D}_1^{-1} \boldsymbol{B}_1^+\right]_{[i],[j,k]} = \begin{cases} \frac{1}{2a_{ii}} & \text{if } i = k, \\ 0 & \text{otherwise,} \end{cases} \qquad \left[\boldsymbol{D}_1^{-1} \boldsymbol{B}_1^-\right]_{[i],[j,k]} = \begin{cases} \frac{1}{2a_{ii}} & \text{if } i = j, \\ 0 & \text{otherwise.} \end{cases}$$

From the proof of Lemma A.1,

$$\boldsymbol{Q}_f = \operatorname{diag} \begin{pmatrix} \boldsymbol{Q}_{f,1} \\ \boldsymbol{Q}_{f,2} \end{pmatrix} = \operatorname{diag} \begin{pmatrix} (\boldsymbol{B}_1^+)^\top \boldsymbol{a} \\ (\boldsymbol{B}_1^-)^\top \boldsymbol{a} \end{pmatrix}.$$

Row $[j,k]$ of $(\boldsymbol{B}_1^+)^\top$ equals $\boldsymbol{e}_k^\top$ and row $[j,k]$ of $(\boldsymbol{B}_1^-)^\top$ equals $\boldsymbol{e}_j^\top$. Thus,

$$\left[\frac{1}{2} \boldsymbol{B}_1^+ \boldsymbol{Q}_{f,1}^{-1}\right]_{[i],[j,k]} = \begin{cases} \frac{1}{2a_{kk}} & \text{if } i = k, \\ 0 & \text{otherwise,} \end{cases} \qquad \left[\frac{1}{2} \boldsymbol{B}_1^- \boldsymbol{Q}_{f,2}^{-1}\right]_{[i],[j,k]} = \begin{cases} \frac{1}{2a_{jj}} & \text{if } i = j, \\ 0 & \text{otherwise.} \end{cases}$$

Since the case statements hold when $i = k$ (first block) or $i = j$ (second block), we get the desired equality. ∎

Proof of (A.1). We now prove the first part of Theorem 3.4, i.e., (A.1).

Proof.

$$\begin{aligned}
-2[\boldsymbol{D}_2\boldsymbol{B}_1^\top\boldsymbol{D}_1^{-1}\boldsymbol{B}_1]\boldsymbol{V}^\top &= -2\boldsymbol{D}_2\boldsymbol{B}_1^\top\boldsymbol{D}_1^{-1}\widehat{\boldsymbol{B}}_1^+\boldsymbol{V}\boldsymbol{V}^\top \quad \text{(A.9)}\\
&= -\boldsymbol{D}_2\boldsymbol{B}_1^\top\widehat{\boldsymbol{B}}_1^+\boldsymbol{Q}_f^{-1}\boldsymbol{V}\boldsymbol{V}^\top \quad \text{(Lemma A.2)}\\
&= -\boldsymbol{D}_2\boldsymbol{B}_1^\top\widehat{\boldsymbol{B}}_1^+\boldsymbol{Q}_f^{-1}(\boldsymbol{I}-\boldsymbol{\Sigma}) \quad \text{(A.5)}\\
&= -\boldsymbol{D}_2\boldsymbol{B}_1^\top\widehat{\boldsymbol{B}}_1^+\boldsymbol{Q}_f^{-1} + \boldsymbol{D}_2\boldsymbol{B}_1^\top\widehat{\boldsymbol{B}}_1^+\boldsymbol{Q}_f^{-1}\boldsymbol{\Sigma} \quad \text{(expansion)}\\
&= -\boldsymbol{D}_2\boldsymbol{B}_1^\top\widehat{\boldsymbol{B}}_1^+\boldsymbol{Q}_f^{-1} + \boldsymbol{D}_2\boldsymbol{B}_1^\top\widehat{\boldsymbol{B}}_1^-\boldsymbol{\Sigma}\boldsymbol{Q}_f^{-1}\boldsymbol{\Sigma} \quad \text{(A.8)}\\
&= -\boldsymbol{D}_2\boldsymbol{B}_1^\top\widehat{\boldsymbol{B}}_1^+\boldsymbol{Q}_f^{-1} + \boldsymbol{D}_2\boldsymbol{B}_1^\top\widehat{\boldsymbol{B}}_1^-\boldsymbol{Q}_b^{-1}\boldsymbol{\Sigma}\boldsymbol{\Sigma} \quad \text{(Lemma A.1)}\\
&= -\boldsymbol{D}_2\boldsymbol{B}_1^\top\widehat{\boldsymbol{B}}_1^+\boldsymbol{Q}_f^{-1} + \boldsymbol{D}_2\boldsymbol{B}_1^\top\widehat{\boldsymbol{B}}_1^-\boldsymbol{Q}_b^{-1} \quad \text{(A.6)}\\
&= \boldsymbol{D}_2\boldsymbol{V}^\top(\widehat{\boldsymbol{B}}_1^-)^\top\widehat{\boldsymbol{B}}_1^+\boldsymbol{Q}_f^{-1} + \boldsymbol{D}_2\boldsymbol{V}^\top(\widehat{\boldsymbol{B}}_1^+)^\top\widehat{\boldsymbol{B}}_1^-\boldsymbol{Q}_b^{-1} \quad \text{(A.10)}\\
&= \boldsymbol{V}^\top\widehat{\boldsymbol{D}}_2(\widehat{\boldsymbol{B}}_1^-)^\top\widehat{\boldsymbol{B}}_1^+\boldsymbol{Q}_f^{-1} + \boldsymbol{V}^\top\widehat{\boldsymbol{D}}_2(\widehat{\boldsymbol{B}}_1^+)^\top\widehat{\boldsymbol{B}}_1^-\boldsymbol{Q}_b^{-1}\\
&= \boldsymbol{V}^\top\left(\boldsymbol{M}_f\boldsymbol{Q}_f^{-1} + \boldsymbol{M}_b\boldsymbol{Q}_b^{-1}\right) \quad \text{(by definition).}
\end{aligned}$$

Note again that $\boldsymbol{M}_f\boldsymbol{Q}_f^{-1}$ and $\boldsymbol{M}_b\boldsymbol{Q}_b^{-1}$ are simply transition matrices of random walks on graphs with weighted adjacency matrices $\boldsymbol{M}_f$ and $\boldsymbol{M}_b$; accordingly, any convex combination is also a valid transition matrix. ∎

Proof of (A.3). Finally, we prove the second part of Theorem 3.4, i.e., (A.3).

Proof. First, observe that

$$\boldsymbol{V}^\top\begin{pmatrix}\boldsymbol{I} & \boldsymbol{I}\\ \boldsymbol{I} & \boldsymbol{I}\end{pmatrix}\widehat{\boldsymbol{D}}_5 = \boldsymbol{0}\widehat{\boldsymbol{D}}_5 = \boldsymbol{0},$$

which implies that the contribution to our projection of the walks that correspond to edges without an upper-adjacent face is zero. Hence, it suffices to show that

$$-(\boldsymbol{B}_2\boldsymbol{D}_3\boldsymbol{B}_2^\top\boldsymbol{D}_2^{-1})\boldsymbol{V}^\top = \boldsymbol{V}^\top\widehat{\boldsymbol{A}}_u\widehat{\boldsymbol{D}}_4^{-1}.$$

We then have that

$$\begin{aligned}
\boldsymbol{V}^\top\widehat{\boldsymbol{A}}_u &= \boldsymbol{V}^\top\left[\widehat{\boldsymbol{B}}_2^+(\widehat{\boldsymbol{B}}_2^-)^\top + \widehat{\boldsymbol{B}}_2^-(\widehat{\boldsymbol{B}}_2^+)^\top\right] \quad \text{(definition of } \widehat{\boldsymbol{A}}_u\text{)}\\
&= \boldsymbol{B}_2(\widehat{\boldsymbol{B}}_2^-)^\top - \boldsymbol{B}_2(\widehat{\boldsymbol{B}}_2^+)^\top \quad \text{(A.14)}\\
&= -\boldsymbol{B}_2\left[(\widehat{\boldsymbol{B}}_2^+)^\top - (\widehat{\boldsymbol{B}}_2^-)^\top\right]\\
&= -\boldsymbol{B}_2(\widehat{\boldsymbol{B}}_2^+)^\top(\boldsymbol{I}-\boldsymbol{\Sigma}) \quad \text{(A.12)}\\
&= -\boldsymbol{B}_2(\widehat{\boldsymbol{B}}_2^+)^\top\boldsymbol{V}\boldsymbol{V}^\top \quad \text{(A.5)}\\
&= -\boldsymbol{B}_2\boldsymbol{B}_2^\top\boldsymbol{V}^\top \quad \text{(A.13).}
\end{aligned}$$

Since $\boldsymbol{D}_3$ is simply a scaling by $1/3$,

$$\begin{aligned}
\boldsymbol{V}^\top\widehat{\boldsymbol{A}}_u\widehat{\boldsymbol{D}}_4^{-1} &= -\boldsymbol{B}_2\boldsymbol{D}_3\boldsymbol{B}_2^\top\boldsymbol{V}^\top(3\cdot\widehat{\boldsymbol{D}}_4^{-1})\\
&= -\boldsymbol{B}_2\boldsymbol{D}_3\boldsymbol{B}_2^\top\boldsymbol{V}^\top\widehat{\boldsymbol{D}}_2^{-1}\\
&= -\boldsymbol{B}_2\boldsymbol{D}_3\boldsymbol{B}_2^\top\boldsymbol{D}_2^{-1}\boldsymbol{V}^\top.
\end{aligned}$$

In the second equality, we used the fact that if $\deg([i,j]) = 0$, then the corresponding column of $\boldsymbol{B}_2\boldsymbol{D}_3\boldsymbol{B}_2^\top$ will be zero. ∎

Stochasticity of $\widehat{\boldsymbol{P}}$. The matrix $\boldsymbol{P}_{\text{lower}}$ is column stochastic by construction. The last item we need to show is that $\boldsymbol{P}_{\text{upper}}$ is also column stochastic, so that we indeed have a stochastic lifting of the normalized Hodge 1-Laplacian. Recall that

$$\begin{aligned}\boldsymbol{P}_{\text{upper}} &= \widehat{\boldsymbol{A}}_u\widehat{\boldsymbol{D}}_4^{-1} + \frac{1}{2}\begin{pmatrix}\boldsymbol{I} & \boldsymbol{I}\\ \boldsymbol{I} & \boldsymbol{I}\end{pmatrix}\widehat{\boldsymbol{D}}_5\\ &= \left[\widehat{\boldsymbol{B}}_2^+(\widehat{\boldsymbol{B}}_2^-)^\top + \widehat{\boldsymbol{B}}_2^-(\widehat{\boldsymbol{B}}_2^+)^\top\right]\widehat{\boldsymbol{D}}_4^{-1} + \frac{1}{2}\begin{pmatrix}\boldsymbol{I} & \boldsymbol{I}\\ \boldsymbol{I} & \boldsymbol{I}\end{pmatrix}\widehat{\boldsymbol{D}}_5.\end{aligned}$$

When $\deg([i,j]) = 0$, the corresponding column of $\widehat{\boldsymbol{A}}_u$ is zero and $\widehat{\boldsymbol{D}}_5$ picks out a column of the stacked identity matrices, which is stochastic when multiplied by $1/2$. When $\deg([i,j]) > 0$, the corresponding column of $\widehat{\boldsymbol{A}}_u$ has nonzero entries for exactly those edges that have a different orientation relative to each coface of $[i,j]$ (each coface results in three entries; see (3.6)). Thus, the column sum is exactly $3 \cdot \deg([i,j])$, so scaling by $\widehat{\boldsymbol{D}}_4^{-1}$ makes the matrix stochastic.

Acknowledgment. We thank Jean-Baptiste Seby for a careful reading of the manuscript and useful comments.

REFERENCES

[1] Y.-Y. Ahn, J. P. Bagrow, and S. Lehmann, *Link communities reveal multiscale complexity in networks*, Nature, 466 (2010), p. 761. (Cited on pp. 355, 362)

[2] R. Albert and A.-L. Barabási, *Statistical mechanics of complex networks*, Rev. Modern Phys., 74 (2002), p. 47. (Cited on p. 356)

[3] R. Albert, H. Jeong, and A.-L. Barabási, *Error and attack tolerance of complex networks*, Nature, 406 (2000), pp. 378–382. (Cited on p. 375)

[4] N. Alon and V. Milman, λ_1*, isoperimetric inequalities for graphs, and superconcentrators*, J. Combin. Theory Ser. B, 38 (1985), pp. 73–88, https://doi.org/10.1016/0095-8956(85)90092-9. (Cited on p. 382)

[5] C. Altafini, *Stability analysis of diagonally equipotent matrices*, Automatica, 49 (2013), pp. 2780–2785. (Cited on p. 377)

[6] R. Andersen, F. Chung, and K. Lang, *Local graph partitioning using PageRank vectors*, in 47th Annual IEEE Symposium on Foundations of Computer Science, IEEE, 2006, pp. 475–486. (Cited on pp. 375, 376, 382)

[7] M. Avella-Medina, F. Parise, M. T. Schaub, and S. Segarra, *Centrality measures for graphons: Accounting for uncertainty in networks*, IEEE Trans. Network Sci. Engrg., (2018), https://doi.org/10.1109/TNSE.2018.2884235. (Cited on p. 381)

[8] S. Barbarossa and M. Tsitsvero, *An introduction to hypergraph signal processing*, in the 2016 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), IEEE, 2016, pp. 6425–6429. (Cited on pp. 356, 377)

[9] M. Belkin and P. Niyogi, *Laplacian eigenmaps and spectral techniques for embedding and clustering*, in Advances in Neural Information Processing Systems, 2002, pp. 585–591. (Cited on pp. 355, 371, 372, 373, 374)

[10] A. R. Benson, R. Abebe, M. T. Schaub, A. Jadbabaie, and J. Kleinberg, *Simplicial closure and higher-order link prediction*, Proc. Natl. Acad. Sci. USA, 115 (2018), pp. E11221–E11230, https://doi.org/10.1073/pnas.1800683115. (Cited on p. 369)

[11] A. R. Benson, D. F. Gleich, and J. Leskovec, *Tensor spectral clustering for partitioning higher-order network structures*, in Proceedings of the 2015 SIAM International Conference on Data Mining, SIAM, 2015, pp. 118–126, https://doi.org/10.1137/1.9781611974010.14. (Cited on p. 357)

[12] A. R. Benson, D. F. Gleich, and J. Leskovec, *Higher-order organization of complex networks*, Science, 353 (2016), pp. 163–166. (Cited on pp. 355, 356, 357)

[13] A. R. Benson, R. Kumar, and A. Tomkins, *Modeling user consumption sequences*, in Proceedings of the 25th International Conference on World Wide Web, ACM Press, 2016, pp. 519–529, https://doi.org/10.1145/2872427.2883024. (Cited on p. 374)
[14] C. Berge and E. Minieka, *Graphs and Hypergraphs*, North-Holland, 1973. (Cited on p. 356)
[15] J. W. Berry, L. K. Fostvedt, D. J. Nordman, C. A. Phillips, C. Seshadhri, and A. G. Wilson, *Why do simple algorithms for triangle enumeration work in the real world?*, in Proceedings of the 5th Conference on Innovations in Theoretical Computer Science, ACM Press, 2014, pp. 225–234, https://doi.org/10.1145/2554797.2554819. (Cited on p. 369)
[16] M. Bianchini, M. Gori, and F. Scarselli, *Inside PageRank*, ACM Trans. Internet Technology, 5 (2005), pp. 92–128. (Cited on p. 375)
[17] G. Bianconi and C. Rahmede, *Emergent hyperbolic network geometry*, Sci. Rep., 7 (2017), art. 41974. (Cited on p. 356)
[18] N. Biggs, *Algebraic potential theory on graphs*, Bull. London Math. Soc., 29 (1997), pp. 641–682. (Cited on p. 355)
[19] P. Billingsley, *Statistical methods in Markov chains*, Ann. Math. Statist., 32 (1961), pp. 12–40. (Cited on p. 355)
[20] S. Boccaletti, V. Latora, Y. Moreno, M. Chavez, and D.-U. Hwang, *Complex networks: Structure and dynamics*, Phys. Rep., 424 (2006), pp. 175–308. (Cited on pp. 355, 356)
[21] P. Boldi and S. Vigna, *Axioms for centrality*, Internet Math., 10 (2014), pp. 222–262. (Cited on p. 375)
[22] B. Bollobás, *Combinatorics: Set Systems, Hypergraphs, Families of Vectors, and Combinatorial Probability*, Cambridge University Press, 1986. (Cited on p. 356)
[23] B. Bollobás, *Modern Graph Theory*, Grad. Texts in Math. 184, Springer Science & Business Media, 2013. (Cited on p. 357)
[24] P. Bonacich, *Power and centrality: A family of measures*, Amer. J. Soc., 92 (1987), pp. 1170–1182. (Cited on p. 375)
[25] P. Bonacich, A. C. Holdren, and M. Johnston, *Hyper-edges and multidimensional centrality*, Social Networks, 26 (2004), pp. 189–203. (Cited on p. 356)
[26] S. Boyd, P. Diaconis, and L. Xiao, *Fastest mixing Markov chain on a graph*, SIAM Rev., 46 (2004), pp. 667–689, https://doi.org/10.1137/S0036144503423264. (Cited on p. 355)
[27] S. Brin and L. Page, *Reprint of: The anatomy of a large-scale hypertextual web search engine*, Comput. Networks, 56 (2012), pp. 3825–3833. (Cited on pp. 375, 376)
[28] D. S. Callaway, M. E. Newman, S. H. Strogatz, and D. J. Watts, *Network robustness and fragility: Percolation on random graphs*, Phys. Rev. Lett., 85 (2000), pp. 5468–5471. (Cited on p. 375)
[29] G. Carlsson, *Topology and data*, Bull. Amer. Math. Soc., 46 (2009), pp. 255–308. (Cited on pp. 356, 369)
[30] D. Cartwright and F. Harary, *Structural balance: A generalization of Heider's theory*, Psych. Rev., 63 (1956), pp. 277–293. (Cited on p. 356)
[31] E. W. Chambers, J. Erickson, and P. Worah, *Testing contractibility in planar rips complexes*, in Proceedings of the Twenty-Fourth Annual Symposium on Computational Geometry, ACM Press, 2008, pp. 251–259, https://doi.org/10.1145/1377676.1377721. (Cited on p. 369)
[32] J. M. Chan, G. Carlsson, and R. Rabadan, *Topology of viral evolution*, Proc. Natl. Acad. Sci. USA, 110 (2013), pp. 18566–18571.(Cited on pp. 356, 382)
[33] T.-H. H. Chan, A. Louis, Z. G. Tang, and C. Zhang, *Spectral properties of hypergraph Laplacian and approximation algorithms*, J. ACM, 65 (2018), art. 15, https://doi.org/10.1145/3178123. (Cited on p. 357)
[34] C. Chen and M. Kerber, *Persistent homology computation with a twist*, in Proceedings 27th European Workshop on Computational Geometry, 2011, pp. 197–200. (Cited on p. 369)
[35] N. Chiba and T. Nishizeki, *Arboricity and subgraph listing algorithms*, SIAM J. Comput., 14 (1985), pp. 210–223, https://doi.org/10.1137/0214017. (Cited on p. 369)
[36] F. Chung, *The heat kernel as the PageRank of a graph*, Proc. Natl. Acad. Sci. USA, 104 (2007), pp. 19735–19740, https://doi.org/10.1073/pnas.0708838104. (Cited on p. 382)
[37] F. R. Chung, *Spectral Graph Theory*, CBMS Regional Conf. Ser. in Math. 92, AMS, 1997. (Cited on pp. 355, 357)
[38] F. R. Chung and L. Lu, *Complex Graphs and Networks*, CBMS Reg. Conf. Ser. Math. 107, AMS, 2006. (Cited on pp. 355, 357)
[39] M. B. Cohen, B. T. Fasy, G. L. Miller, A. Nayyeri, R. Peng, and N. Walkington, *Solving* 1*-Laplacians in nearly linear time: Collapsing and expanding a topological ball*, in Proceedings of the Twenty-Fifth Annual ACM-SIAM Symposium on Discrete Algorithms, SIAM, 2014, pp. 204–216, https://doi.org/10.1137/1.9781611973402.15. (Cited on p. 370)

[40] R. R. Coifman and S. Lafon, *Diffusion maps*, Appl. Comput. Harmon. Anal., 21 (2006), pp. 5–30. (Cited on pp. 355, 371, 372, 374)
[41] O. T. Courtney and G. Bianconi, *Generalized network structures: The configuration model and the canonical ensemble of simplicial complexes*, Phys. Rev. E, 93 (2016), art. 062311. (Cited on p. 356)
[42] V. de Silva and R. Ghrist, *Coverage in sensor networks via persistent homology*, Algebr. Geom. Topol., 7 (2007), pp. 339–358. (Cited on pp. 356, 382)
[43] V. de Silva, D. Morozov, and M. Vejdemo-Johansson, *Dualities in persistent (co)homology*, Inverse Problems, 27 (2011), art. 124003, https://doi.org/10.1088/0266-5611/27/12/124003. (Cited on p. 369)
[44] D. Dotterrer and M. Kahle, *Coboundary expanders*, J. Topology Anal., 4 (2012), pp. 499–514, https://doi.org/10.1142/s1793525312500197. (Cited on p. 382)
[45] B. Eckmann, *Harmonische Funktionen und Randwertaufgaben in einem Komplex*, Comment. Math. Helv., 17 (1944), pp. 240–255. (Cited on p. 356)
[46] E. Estrada and G. J. Ross, *Centralities in simplicial complexes. Applications to protein interaction networks*, J. Theoret. Biol., 438 (2018), pp. 46–60. (Cited on pp. 375, 381)
[47] T. Evans and R. Lambiotte, *Line graphs, link partitions, and overlapping communities*, Phys. Rev. E, 80 (2009), art. 016105. (Cited on pp. 355, 362)
[48] S. L. Feld, *The focused organization of social ties*, Amer. J. Soc., 86 (1981), pp. 1015–1035. (Cited on p. 356)
[49] D. C.-L. Fong and M. Saunders, *LSMR: An iterative algorithm for sparse least-squares problems*, SIAM J. Sci. Comput., 33 (2011), pp. 2950–2971, https://doi.org/10.1137/10079687x. (Cited on p. 370)
[50] L. C. Freeman, *A set of measures of centrality based on betweenness*, Sociometry, 40 (1977), pp. 35–41. (Cited on p. 375)
[51] L. C. Freeman, *Centrality in social networks conceptual clarification*, Social Networks, 1 (1978), pp. 215–239. (Cited on p. 375)
[52] G. Froyland and K. Padberg-Gehle, *A rough-and-ready cluster-based approach for extracting finite-time coherent sets from sparse and incomplete trajectory data*, Chaos, 25 (2015), art. 087406. (Cited on p. 374)
[53] A. Gautier, F. Tudisco, and M. Hein, *A unifying Perron–Frobenius theorem for nonnegative tensors via multihomogeneous maps*, SIAM J. Matrix Anal. Appl., 40 (2019), pp. 1206–1231, https://doi.org/10.1137/18M1165049. (Cited on p. 357)
[54] A. Ghosh, B. Rozemberczki, S. Ramamoorthy, and R. Sarkar, *Topological signatures for fast mobility analysis*, in Proceedings of the 26th ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems, ACM, 2018, pp. 159–168. (Cited on pp. 356, 357, 373, 374)
[55] R. Ghrist, *Barcodes: The Persistent Topology of Data*, Bull. Amer. Math. Soc., 45 (2008), pp. 61–75. (Cited on p. 356)
[56] R. Ghrist and A. Muhammad, *Coverage and hole-detection in sensor networks via homology*, in the Fourth International IEEE Symposium on Information Processing in Sensor Networks, IPSN 2005, IEEE, 2005, pp. 254–260. (Cited on pp. 356, 382)
[57] C. Giusti, R. Ghrist, and D. S. Bassett, *Two's company, three (or more) is a simplex: Algebraic-topological tools for understanding higher-order structure in neural data*, J. Comput. Neurosci., 41 (2016), pp. 1–14, https://10.1007/s10827-016-0608-6. (Cited on pp. 356, 369)
[58] C. Giusti, E. Pastalkova, C. Curto, and V. Itskov, *Clique topology reveals intrinsic geometric structure in neural correlations*, Proc. Natl. Acad. Sci. USA, 112 (2015), pp. 13455–13460. (Cited on pp. 356, 382)
[59] D. F. Gleich, *PageRank beyond the Web*, SIAM Rev., 57 (2015), pp. 321–363, https://doi.org/10.1137/140976649. (Cited on pp. 355, 356, 375, 376)
[60] D. F. Gleich and M. W. Mahoney, *Using local spectral methods to robustify graph-based learning algorithms*, in Proceedings of the 21th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, ACM Press, 2015, pp. 359–368, https://doi.org/10.1145/2783258.2783376. (Cited on pp. 375, 376)
[61] C. Godsil and G. F. Royle, *Algebraic Graph Theory*, Springer Science & Business Media, 2013. (Cited on p. 370)
[62] L. J. Grady and J. R. Polimeni, *Discrete Calculus: Applied Analysis on Graphs for Computational Science*, Springer Science & Business Media, 2010. (Cited on pp. 365, 374, 377)
[63] J. Grilli, G. Barabás, M. J. Michalska-Smith, and S. Allesina, *Higher-order interactions stabilize dynamics in competitive network models*, Nature, 548 (2017), p. 210. (Cited on p. 356)

[64] S. Guattery, *Graph Embeddings, Symmetric Real Matrices, and Generalized Inverses*, Tech. Report NASA/CR-1998-208462, Institute for Computer Applications in Science and Engineering, NASA Langley Research Center, 1998, http://handle.dtic.mil/100.2/ADA351054. (Cited on pp. 355, 370)
[65] A. Gundert and U. Wagner, *On Laplacians of random complexes*, in Proceedings of the Twenty-Eighth Annual Symposium on Computational Geometry, ACM Press, 2012, pp. 151–160, https://doi.org/10.1145/2261250.2261272. (Cited on p. 382)
[66] A. Hatcher, *Algebraic Topology*, Cambridge University Press, 2002. (Cited on p. 356)
[67] G. Henselman and R. Ghrist, *Matroid Filtrations and Computational Persistent Homology*, preprint, https://arxiv.org/abs/1606.00199, 2016. (Cited on p. 369)
[68] D. Horak and J. Jost, *Spectra of combinatorial Laplace operators on simplicial complexes*, Adv. Math., 244 (2013), pp. 303–336, https://doi.org/10.1016/j.aim.2013.05.007. (Cited on pp. 357, 365)
[69] I. Iacopini, G. Petri, A. Barrat, and V. Latora, *Simplicial models of social contagion*, Nature Commun., 10 (2019), pp. 1–9. (Cited on p. 356)
[70] N. Ikeda and S. Watanabe, *Stochastic Differential Equations and Diffusion Processes*, North-Holland Math. Lib. 24, Elsevier, 2014. (Cited on p. 355)
[71] J. Jia, M. T. Schaub, S. Segarra, and A. R. Benson, *Graph-based semi-supervised & active learning for edge flows*, in Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, ACM Press, 2019, pp. 761–771. (Cited on p. 377)
[72] X. Jiang, L.-H. Lim, Y. Yao, and Y. Ye, *Statistical ranking and combinatorial Hodge theory*, Math. Program., 127 (2011), pp. 203–244, https://doi.org/10.1007/s10107-010-0419-x. (Cited on p. 369)
[73] M. Kahle, *Topology of random simplicial complexes: A survey*, AMS Contemp. Math, 620 (2014), pp. 201–222. (Cited on p. 356)
[74] L. Katz, *A new status index derived from sociometric analysis*, Psychometrika, 18 (1953), pp. 39–43. (Cited on p. 375)
[75] K. F. Kee, L. Sparks, D. C. Struppa, and M. Mannucci, *Social groups, social media, and higher dimensional social structures: A simplicial model of social aggregation for computational communication research*, Commun. Quart., 61 (2013), pp. 35–58. (Cited on p. 356)
[76] M. Kitsak, L. K. Gallos, S. Havlin, F. Liljeros, L. Muchnik, H. E. Stanley, and H. A. Makse, *Identification of influential spreaders in complex networks*, Nature Phys., 6 (2010), pp. 888–893. (Cited on p. 375)
[77] S. Klamt, U.-U. Haus, and F. Theis, *Hypergraphs and cellular networks*, PLoS Comput. Biol., 5 (2009), art. e1000385. (Cited on p. 356)
[78] K. Kloster and D. F. Gleich, *Heat kernel based community detection*, in Proceedings of the 20th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, ACM Press, 2014, pp. 1386–1395, https://doi.org/10.1145/2623330.2623706. (Cited on p. 382)
[79] I. M. Kloumann, J. Ugander, and J. Kleinberg, *Block models and personalized PageRank*, Proc. Natl. Acad. Sci. USA, 114 (2017), pp. 33–38. (Cited on pp. 375, 376)
[80] V. Krebs, Orgnet LLC, http://www.orgnet.com/, unpublished data. (Cited on p. 380)
[81] A. N. Langville and C. D. Meyer, *Deeper inside PageRank*, Internet Math., 1 (2004), pp. 335–380. (Cited on p. 375)
[82] M. Latapy, *Main-memory triangle computations for very large (sparse (power-law)) graphs*, Theoret. Comput. Sci., 407 (2008), pp. 458–473, https://doi.org/10.1016/j.tcs.2008.07.017. (Cited on p. 369)
[83] G. F. Lawler and A. D. Sokal, *Bounds on the L^2 spectrum for Markov chains and Markov processes: A generalization of Cheeger's inequality*, Trans. Amer. Math. Soc., 309 (1988), pp. 557–580, https://doi.org/10.2307/2000925. (Cited on p. 382)
[84] A. B. Lee, K. S. Pedersen, and D. Mumford, *The nonlinear statistics of high-contrast patches in natural images*, Internat. J. Comput. Vision, 54 (2003), pp. 83–103, https://doi.org/10.1023/a:1023705401078. (Cited on p. 369)
[85] L.-H. Lim, *Hodge Laplacians on graphs*, SIAM Rev., 62 (3) (2020), to appear. (Cited on pp. 355, 356, 357, 359, 360, 361, 369, 370, 371)
[86] S. A. Marvel, J. Kleinberg, R. D. Kleinberg, and S. H. Strogatz, *Continuous-time model of structural balance*, Proc. Natl. Acad. Sci. USA, 108 (2011), pp. 1771–1776. (Cited on p. 356)
[87] N. Masuda, M. A. Porter, and R. Lambiotte, *Random walks and diffusion on networks*, Phys. Rep., 716/717 (2017), pp. 1–58. (Cited on pp. 355, 371, 382)

[88] M. Mihail, *Conductance and convergence of Markov chains—a combinatorial treatment of expanders*, in 30th Annual Symposium on Foundations of Computer Science, IEEE, 1989, pp. 526–531, https://doi.org/10.1109/sfcs.1989.63529. (Cited on p. 382)
[89] B. Mohar, Y. Alavi, G. Chartrand, and O. Oellermann, *The Laplacian spectrum of graphs*, in Graph Theory, Combinatorics, and Applications, Vol. 2, Wiley, 1991, pp. 871–898. (Cited on pp. 355, 357)
[90] A. Muhammad and M. Egerstedt, *Control using higher order Laplacians in network topologies*, in Proc. 17th International Symposium on Mathematical Theory of Networks and Systems, 2006, pp. 1024–1038. (Cited on pp. 356, 360, 373)
[91] S. Mukherjee and J. Steenbergen, *Random walks on simplicial complexes and harmonics*, Random Structures Algorithms, 49 (2016), pp. 379–405, https://doi.org/10.1002/rsa.20645. (Cited on pp. 355, 357, 367)
[92] V. Nanda and R. Sazdanović, *Simplicial models and topological inference in biological systems*, in Discrete and Topological Models in Molecular Biology, Springer, 2014, pp. 109–141. (Cited on pp. 356, 382)
[93] M. E. J. Newman, *The structure and function of complex networks*, SIAM Rev., 45 (2003), pp. 167–256, https://doi.org/10.1137/S003614450342480. (Cited on pp. 355, 356)
[94] M. E. J. Newman, *Modularity and community structure in networks*, Proc. Natl. Acad. Sci. USA, 103 (2006), pp. 8577–8582, https://doi.org/10.1073/pnas.0601602103. (Cited on p. 380)
[95] M. E. J. Newman, *Networks: An Introduction*, Oxford University Press, 2010. (Cited on p. 356)
[96] A. Ortega, P. Frossard, J. Kovacevic, J. M. F. Moura, and P. Vandergheynst, *Graph Signal Processing: Overview, Challenges, and Applications*, Proc. IEEE, 106 (2018), pp. 808–828, https://doi.org/10.1109/JPROC.2018.2820126. (Cited on pp. 371, 377)
[97] B. Osting, S. Palande, and B. Wang, *Towards Spectral Sparsification of Simplicial Complexes Based on Generalized Effective Resistance*, preprint, https://arxiv.org/abs/1708.08436, 2017. (Cited on pp. 357, 362, 382)
[98] N. Otter, M. A. Porter, U. Tillmann, P. Grindrod, and H. A. Harrington, *A roadmap for the computation of persistent homology*, EPJ Data Sci., 6 (2017), art. 17, https://doi.org/10.1140/epjds/s13688-017-0109-5. (Cited on p. 369)
[99] C. C. Paige and M. A. Saunders, *LSQR: An algorithm for sparse linear equations and sparse least squares*, ACM Trans. Math. Software, 8 (1982), pp. 43–71, https://doi.org/10.1145/355984.355989. (Cited on p. 370)
[100] O. Parzanchevski and R. Rosenthal, *Simplicial complexes: Spectrum, homology and random walks*, Random Structures Algorithms, 50 (2017), pp. 225–261. (Cited on p. 357)
[101] O. Parzanchevski, R. Rosenthal, and R. J. Tessler, *Isoperimetric inequalities in simplicial complexes*, Combinatorica, 36 (2015), pp. 195–227, https://doi.org/10.1007/s00493-014-3002-x. (Cited on p. 382)
[102] R. Pastor-Satorras, C. Castellano, P. Van Mieghem, and A. Vespignani, *Epidemic processes in complex networks*, Rev. Modern Phys., 87 (2015), pp. 925–979. (Cited on p. 375)
[103] A. Patania, G. Petri, and F. Vaccarino, *The shape of collaborations*, EPJ Data Sci., 6 (2017), art. 18, https://doi.org/10.1140/epjds/s13688-017-0114-8. (Cited on p. 369)
[104] G. Petri, P. Expert, F. Turkheimer, R. Carhart-Harris, D. Nutt, P. J. Hellyer, and F. Vaccarino, *Homological scaffolds of brain functional networks*, J. Roy. Soc. Interface, 11 (2014), art. 20140873. (Cited on pp. 356, 382)
[105] F. T. Pokorny, M. Hawasly, and S. Ramamoorthy, *Topological trajectory classification with filtrations of simplicial complexes and persistent homology*, Internat. J. Robotics Res., 35 (2016), pp. 204–223, https://doi.org/10.1177/0278364915586713. (Cited on p. 356)
[106] L. Qi and Z. Luo, *Tensor Analysis: Spectral Theory and Special Tensors*, SIAM, 2017, https://doi.org/10.1137/1.9781611974751. (Cited on p. 357)
[107] F. Reza, *Some topological considerations in network theory*, IRE Trans. Circuit Theory, 5 (1958), pp. 30–42. (Cited on p. 356)
[108] R. Rosenthal, *Simplicial branching random walks and their applications*, preprint, https://arxiv.org/abs/1412.5406, 2014. (Cited on pp. 355, 357, 367)
[109] J. P. Roth, *An application of algebraic topology to numerical analysis: On the existence of a solution to the network problem*, Proc. Natl. Acad. Sci. USA, 41 (1955), pp. 518–521. (Cited on p. 356)
[110] J. P. Roth, *An application of algebraic topology: Kron's method of tearing*, Quart. Appl. Math., 17 (1959), pp. 1–24. (Cited on p. 356)

[111] M. Salathé and J. H. Jones, *Dynamics and control of diseases in networks with community structure*, PLoS Comput. Biol., 6 (2010), art. e1000736, https://doi.org/10.1371/journal.pcbi.1000736. (Cited on p. 355)

[112] M. J. Salganik and D. D. Heckathorn, *Sampling and estimation in hidden populations using respondent-driven sampling*, Sociological Method., 34 (2004), pp. 193–240, https://doi.org/10.1111/j.0081-1750.2004.00152.x. (Cited on p. 355)

[113] M. T. Schaub, J.-C. Delvenne, M. Rosvall, and R. Lambiotte, *The many facets of community detection in complex networks*, Appl. Network Sci., 2 (2017), art. 4, https://doi.org/10.1007/s41109-017-0023-6. (Cited on p. 375)

[114] M. T. Schaub, J. Lehmann, S. N. Yaliraki, and M. Barahona, *Structure of complex networks: Quantifying edge-to-edge relations by failure-induced flow redistribution*, Network Sci., 2 (2014), pp. 66–89. (Cited on pp. 357, 370)

[115] M. T. Schaub and S. Segarra, *Flow smoothing and denoising: Graph signal processing in the edge-space*, in 2018 IEEE Global Conference on Signal and Information Processing (GlobalSIP), IEEE, 2018, pp. 735–739. (Cited on pp. 356, 357, 374, 377)

[116] A. Sinclair and M. Jerrum, *Approximate counting, uniform generation and rapidly mixing Markov chains*, Inform. and Comput., 82 (1989), pp. 93–133, https://doi.org/10.1016/0890-5401(89)90067-9. (Cited on p. 382)

[117] A. Solé-Ribalta, M. De Domenico, S. Gómez, and A. Arenas, *Centrality rankings in multiplex networks*, in Proceedings of the 2014 ACM Conference on Web Science, ACM, 2014, pp. 149–155. (Cited on p. 381)

[118] D. A. Spielman, *Spectral graph theory and its applications*, in the 48th Annual IEEE Symposium on Foundations of Computer Science, FOCS'07, IEEE, 2007, pp. 29–38. (Cited on p. 357)

[119] J. Steenbergen, C. Klivans, and S. Mukherjee, *A Cheeger-type inequality on simplicial complexes*, Adv. Appl. Math., 56 (2014), pp. 56–77, https://doi.org/10.1016/j.aam.2014.01.002. (Cited on p. 382)

[120] S. H. Strogatz, *Exploring complex networks*, Nature, 410 (2001), pp. 268–276, https://doi.org/10.1038/35065725. (Cited on p. 356)

[121] A. Tahbaz-Salehi and A. Jadbabaie, *Distributed coverage verification in sensor networks without location information*, IEEE Trans. Automat. Control, 55 (2010), pp. 1837–1849, https://doi.org/10.1109/TAC.2010.2047541. (Cited on pp. 356, 360, 382)

[122] D. Taylor, F. Klimm, H. A. Harrington, M. Kramár, K. Mischaikow, M. A. Porter, and P. J. Mucha, *Topological data analysis of contagion maps for examining spreading processes on networks*, Nature Commun., 6 (2015), art. 7723. (Cited on p. 382)

[123] N. K. Vishnoi, *$Lx = b$. Laplacian solvers and their algorithmic applications*, Found. Trends Theoret. Comput. Sci., 8 (2012), pp. 1–141. (Cited on p. 370)

[124] S. Wasserman and K. Faust, *Social Network Analysis: Methods and Applications*, Cambridge University Press, 1994. (Cited on p. 356)

[125] J.-G. Young, G. Petri, F. Vaccarino, and A. Patania, *Construction of and efficient sampling from the simplicial configuration model*, Phys. Rev. E, 96 (2017), art. 032312. (Cited on p. 356)

[126] D. Zhou, J. Huang, and B. Schölkopf, *Learning with hypergraphs: Clustering, classification, and embedding*, in Advances in Neural Information Processing Systems, 2007, pp. 1601–1608. (Cited on p. 357)

[127] A. Zomorodian, *Fast construction of the Vietoris–Rips complex*, Comput. Graphics, 34 (2010), pp. 263–271, https://doi.org/10.1016/j.cag.2010.03.007. (Cited on p. 369)