Persistent homology of convection cycles in network flows

Minh Quang Le and Dane Taylor

Phys. Rev. E **105**, 044311 — Published 18 April 2022

DOI: 10.1103/PhysRevE.105.044311

Persistent Homology of Convection Cycles in Network Flows

Minh Quang Le[1,*] and Dane Taylor[1,†]

[1]*Department of Mathematics, University at Buffalo, State University of New York, Buffalo, NY 14260, USA*

(Dated: March 31, 2022)

Convection is a well-studied topic in fluid dynamics, yet it is less understood in the context of networks flows. Here, we incorporate techniques from topological data analysis (namely, persistent homology) to automate the detection and characterization of convective/cyclic/chiral flows over networks, particularly those that arise for irreversible Markov chains (MCs). As two applications, we study convection cycles arising under the PageRank algorithm, and we investigate chiral edges flows for a stochastic model of a bi-monomer's configuration dynamics. Our experiments highlight how system parameters—e.g., the teleportation rate for PageRank and the transition rates of external and internal state changes for a monomer—can act as *homology regularizers* of convection, which we summarize with *persistence barcodes* and *homological bifurcation diagrams.* Our approach establishes a new connection between the study of convection cycles and homology, the branch of mathematics that formally studies cycles, which has diverse potential applications throughout the sciences and engineering.

PACS numbers: 89.75.Hc, 02.50.Ga, 87.15.hj,84.35.+i

I. INTRODUCTION

One of the main goals of topological data analysis (TDA) is to characterize the structure of an object—usually a point cloud—through its topological features. In particular, persistent homology [1, 2] is a family of techniques that detect and summarize multiscale topological features and has been applied to a wide variety of applications including time-series data [3, 4], image processing [5], machine learning [6], and artificial intelligence [7, 8]. Complementing the study of point-cloud data, another line of research involves utilizing the TDA toolset to study complex systems, for which applications include the analysis of spreading processes over social networks [9], network neuroscience [10–12], mechanical-force networks [13], jamming in granular material [14], molecular structure [15], and DNA folding [16]. In this paper, we employ techniques from TDA to study Markov chains (MCs), which provide a foundation to numerous areas of science and engineering including queuing theory [17], population dynamics [18], as well as statistical (and machine learning) models that rely on Markov chain Monte Carlo [19], hidden Markov models [20], and Markov decision process [21].

We utilize the mathematical framework of persistent homology to automate the detection (and summarize the multiscale properties) of convection cycles that arise for the stationary flows of irreversible MCs. Notably, while convection cycles have been extensively studied in fluid dynamics, they are less understood in the context of flows over networks. For example, it was recently observed that the coupling together of reversible MCs can give rise to an irreversible MC with convection cycles that are an emergent property [9]. Emergent convection cycles have also been recently found to describe the phenomenon of "chiral edge flows" [22], providing new insights into the quantum Hall effect, configurational dynamics of monomers, and biological (e.g., circadian) rhythms. Given the inherent prevalence of convection cycles in MCs and other network flows, it is important that we place their study on a stronger mathematical, computational, and theoretical footing.

We study convection using a branch of mathematics called *homology* and the related field *computational homology* [23]. Both are concerned with studying the absence/presence of k-dimensional "holes" (and their connectivity) within a topological space such as a simplicial complex. Importantly, cycles on a graph are 1-dimensional (1D) holes, and so persistent homology is a natural fit to analyze convection cycles. We construct filtrations of graphs by including edges according to the stationary flows across them (which is done in descending order so that the last edges to be included are those with the smallest stationary flows), and we summarize the persistent homology of the filtered graphs' associated clique complexes. See Fig. 1 for a graph and its associated clique complex. Computationally, we implement these techniques by building on a popular TDA framework called Gudhi [24], which we adapt to implement *edge-value clique (EVC) filtrations* of scalar-functions that are defined over the the edges of a graph.

We apply this technique to two applications. First, we study convection cycles arising under the PageRank algorithm [25], examining the role of the *teleportation parameter*. Second, we study chiral edge flows that emerge for a 4-state model that describes the configurational changes of a bi-monomer [22], examining the roles of the external and internal transition rates. These parameters significantly affect convection cycles arising for these respective applications, and we show that they act as "homology regularizers" of convection. We introduce "homological bifurcation diagrams" to summarize these effects. Our methods provide mathematically principled (and automated) tools to gain a deeper understanding of the structural patterns of convection on networks, and they are expected to be useful to myriad applications across the physical, social, biological and computational sciences.

The remainder of this paper is organized as follows: We present background information in Sec. II, our methodology in Sec. III, applications in Sec. IV, and a discussion in Sec. V.

* minhquan@buffalo.edu
† danet@buffalo.edu

II. BACKGROUND INFORMATION

Here, we present introductory material about simplicial complexes and homology (Sec. II A), persistent homology of graphs (Sec. II B), and discrete-time MCs (Sec. II C).

A. Simplicial complexes (SCs) and their homology

We first define an *undirected graph* $G = \{\mathcal{V}, \mathcal{E}\}$, where $\mathcal{V} = \{1, \ldots, N_0\}$ is a set of N_0 vertices and $\mathcal{E} \subset \mathcal{V} \times \mathcal{V}$ is a set of edges. Note that each vertex is specified by a single index $i \in V$, and each edge is specified by an unordered pair $(i, j) \in \mathcal{V} \times \mathcal{V}$. More generally, we define a $(k+1)$-tuple of vertices $\sigma = (i_0, i_1, \ldots, i_k)$ as a k-dimensional simplex, or *k-simplex* [26]. Vertices and edges are equivalent to 0-simplices and 1-simplices, respectively. An abstract SC is a set of simplices of possibly different dimensions, and it is a generalization of an undirected graph. It is also a type of *hypergraph* with a constraint on which simplices can exist. That is, for any k-simplex $(i_0, i_1, \ldots, i_k)$ in an SC, its *faces* are the $(k-1)$ simplices in which one of the of indices is omitted (e.g., i_1 is omitted to yield $(i_0, i_2, \ldots, i_k)$). The *cofaces* of a $(k-1)$-simplex are the k-simplices for which it is a face. Note that the faces of an edge (i, j) are the vertices i and j, and likewise, (i, j) is a coface of each of these vertices.

With these definitions, we state the two restrictions that are required for a SC: (i) for any face, its faces must be included in the SC; and (ii) the intersection of any two faces is either a face of both, or it is an empty set. The dimension of an SC is the maximum dimension of its simplices, and an undirected graph is a 1-dimensional SC—it contains 0-simplices and 1-simplices, and for any edge (i, j), the vertices i and j must exist. We will focus on a particular type of SC that can be generated from a graph and is called a *clique complex*. A clique complex $K(G)$ of a graph G is the SC in which there is a 1-to-1 correspondence between the $(k+1)$-cliques in the graph and the k-simplices in the SC. (Recall that an *n-clique* is a complete subgraph on $n+1$ vertices of a graph.) Given this 1-to-1 correspondence, the map from G to $K(G)$ is invertible, and G can be recovered as the 1-skeleton of $K(G)$. (A *k-skeleton* of a SC is the SC that is obtained after removing all simplices having dimensions that are greater than k.)

We next discuss *simplicial homology*, which will lead to a formal definition of "homological" cycles. To this end, we consider vector spaces defined over the k-simplices in a SC. A *k-chain*, $\sum_{n=1}^{N_k} \alpha_n \sigma_n$, is a linear combination of k-simplices $\{\sigma_n\}$ with weights $\{\alpha_k\}$. (Note that N_0 and N_1 are the numbers of vertices and edges, respectively.) If a SC contains N_k different k-simplices, then the vector space of k-chains is N_k-dimensional, and it is isomorphic to $\mathbb{R}^{N_k}$ if one assumes $\alpha_k \in \mathbb{R}$. We now consider a simplicial map $f : X_k \to X_{k-1}$ between X_k, which is a SC of dimension k, and X_{k-1}, which is a SC of dimension $k-1$ that contains the faces of simplices in X_k. Considering the vector space C_k of k-chains defined over k-simplices in X_k and vector space C_{k-1} of $(k-1)$-chains defined over their cofaces in X_{k-1}, we define the linear *boundary map* $\partial_k : C_k \longrightarrow C_{k-1}$, where the action of ∂_k on any k-simplex is given by

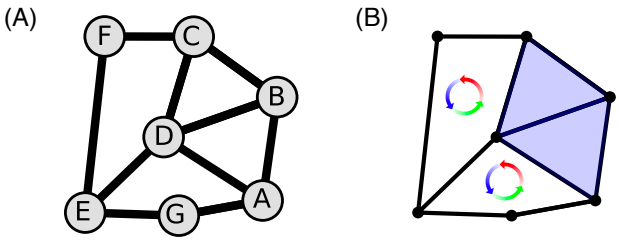

FIG. 1. **A graph and an associated simplicial complex (SC).** (A) Graph $\mathcal{G}$ with $N = 7$ vertices and $M = 10$ undirected, unweighted edges. (B) The corresponding *clique complex* $\mathcal{S}$, where each k-clique gives rise to a $(k-1)$-simplex. Each triangle (i.e., 3-clique) gives rise to a 2-simplex (see shaded triangles). The SC is a topological space and has associated vector spaces. Consider the space $\mathbb{R}^7$ of real-valued functions defined over the vertices. Since there is just 1 connected component, the SC's 0-dimensional (0D) homology is a 1D subspace of $\mathbb{R}^7$. Similarly, there are two "homological" 1-cycles that are not a boundary of a 2-simplex, and so the 1D homology is a 2D subspace of $\mathbb{R}^{10}$ (i.e., the space of real-valued functions defined over the ten edges).

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

The boundary map allows one to relate vectors in C_k to those in C_{k-1}. For example, the *boundary* of a 2-simplex (i.e., triangle) (i, j, k) is the signed combination if the associated edges, $\partial_2(i, j, k) = (j, k) - (i, k) + (i, j)$. Notably, the boundary of any closed path is zero, which yields an algebraic definition of a k-cycle: any k-chain that lies within the subspace Z_k, where $Z_k = \ker(\partial_k) \subseteq C_k$ is the *vector space of k-cycles*.

Notably, k-cycles can arise for different reasons, and we distinguish two types. The boundary map satisfies the property $\partial_k \circ \partial_{k+1} = 0$, which essentially states that the boundary of a boundary is zero. [For the triangle, $\partial_1 \circ \partial_2(i, j, k) = \partial_1(j, k) - \partial_1(i, k) + \partial_1(i, j) = (k - j) - (k - i) + (j - i) = 0$.] Thus we define $B_k = \text{image}(\partial_{k+1})$ as the *subspace of $(k+1)$-boundaries*, and it follows that $B_k \subseteq Z_k$. In other words, some cycles arise simply because they are boundaries of $(k+1)$-simplices For example, observe in Fig. 1(B) that there are two "triangular" cycles that exist around the two 2-simplices, but that there are two other cycles that also exist. The *k-th simplicial homology* is defined as the quotient space $H_k = Z_k / B_k$, and it represents the subspace of k-dimensional cycles (i.e., k-cycles) that do not arise simply as the boundary of a $(k+1)$-simplex.

The k-th simplicical homology can be represented by the span of *homology generators*, which are a linearly independent set of k-chains that span H_n and represent the associated k-cycles. The number of linearly independent homology generators is called a *Betti number*

$$\beta_k = \dim H_k = \dim(Z_k) - \dim(B_k). \quad (2)$$

Informally, β_0 is the number of connected components; β_1 is

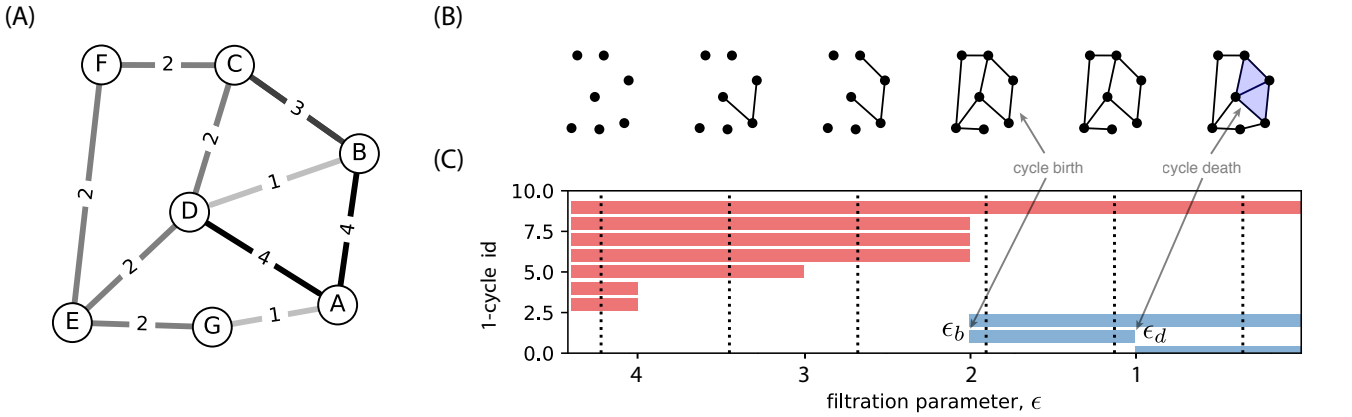

FIG. 2. **Persistent homology of a graph according to an *edge-value clique (EVC) filtration*.** (A) An undirected graph $G(\mathcal{V}, \mathcal{E})$ with a scalar function $f : \mathcal{E} \to \mathbb{R}$ defined over the edges $\mathcal{E}$. We apply an EVC filtration to the graph by considering a monotonically decreasing *filtration parameter* $\epsilon = (0, 5]$, and by considering a filtered sequence of graphs $G(\mathcal{V}, \mathcal{E}_\epsilon)$, where $\mathcal{E}_\epsilon = \{(i,j)|f(i,j) > \epsilon\}$ is the subset of edges for which $f(i,j)$ are larger than a threshold ϵ. (B) Visualization of the graphs' associated clique complexes $K_\epsilon \equiv K(G(\mathcal{V}, \mathcal{E}_\epsilon))$ for several ϵ. (C) A *persistence barcode* summarizes how the 0-dimensional (red) and 1-dimensional (blue) homological k-cycles of K_ϵ change with decreasing ϵ. The arrows highlight two events: at $\epsilon = \epsilon_b$, a homological 1-cycle involving four edges is *born*; at $\epsilon = \epsilon_d$, the 1-cycle *dies*, since it is "filled in" by a 1-simplex and two 2-simplices.

the number of 1-dimensional cycles or "loops" (that is, not including the triangular boundaries of 2-simplices); and β_2 is the number of 2-dimensional holes or "voids" (e.g., the interior of a triangulated sphere). For the SC shown in Fig. 1(B), $\beta_0 = 1$ since there's one connected component, and $\beta_1 = 2$ since there are two cycles that are not simply the boundaries of 2-simplices.

By formulating k-cycles algebraically, one can consider the linear dependence and independence of k-cycles. As such, one can not only identify cycles, but also investigate the relations/connectivity between cycles, which we find to be instrumental for understanding pattern formation for cycle. We also highlight that a given homological k-cycle can potentially have more than one homological generator. Such generators are said to be *homologically equivalent*, and they can be obtained by considering linear combinations of k-cycles (including both homological k-cycles and boundaries). We will later show that this complicates the investigation of convection cycles through the lens of homological k-cycles.

B. Persistent homology of scalar functions defined over edges

One of the greatest tools of topological data analysis is the study of *persistent homology* [1, 2]. Here, we examine how the homology of a topological object changes as it undergoes a *filtration* to yield a monotonically increasing sequence $X_0 \subseteq X_1 \subseteq X_1 \subseteq ...$ (e.g., of simplicial complexes $\{X_t\}$). We consider filtrations in which one has a scalar function $f : \mathcal{E} \longrightarrow \mathbb{R}$ over the edges, and each edge $(i,j) \in \mathcal{E}$ is retained/removed according to $f(i,j)$. The values $f(i,j)$ could be edge weights for a weighted graph, but in general they can encode any scalar property. We visualize such a graph and the values $f(i,j)$ in Fig. 2(A).

We call the process an *edge-value clique (EVC) filtration*, and we construct it as follows. Given a graph $G(\mathcal{V}, \mathcal{E})$ and a *filtration function* f, we define the subsets $\mathcal{E}_\epsilon = \{(i,j)|f(i,j) > \epsilon\}$ in which one retains edges only for which $f(i,j)$ is sufficiently large. Note that the subsets $\{\mathcal{E}_\epsilon\}$ are non-decreasing as ϵ decreases (i.e., $\mathcal{E}_\epsilon \subseteq \mathcal{E}_{\epsilon'}$ for any $\epsilon' < \epsilon$). We must specify a range over which to decrease $\epsilon \in (\epsilon_B, \epsilon_A]$, and in practice we assume $\epsilon_A > \max_{(i,j)\in\mathcal{E}} f(i,j)$ and $\epsilon_B < \min_{(i,j)\in\mathcal{E}} f(i,j)$. It then follows that $\mathcal{E}_\epsilon = \emptyset$ is an empty set of edges when $\epsilon \geq \epsilon_A$, and $\mathcal{E}_\epsilon = \mathcal{E}$ (i.e., all edges are retained) when $\epsilon \leq \epsilon_B$. See [27] for our codebase that implements EVC filtrations by adapting the TDA framework called Gudhi [24], and which reproduces the results of this paper.

In Fig. 2(B), we visualize a sequence of filtered clique complexes $\{K_\epsilon\}$ that are associated with the filtered graphs $\{G_\epsilon\}$ that are defined with the edge sets $\{\mathcal{E}_\epsilon\}$. In Fig. 2(C), we summarize the persistent homology of $\{K_\epsilon\}$ in a *persistence barcode*, which reveals how homology changes with ϵ. Observe that when ϵ is sufficiently large, K_ϵ contains vertices but no edges. On the other hand, when ϵ decreases to be sufficiently small, then K_ϵ recovers the original clique complex [recall Fig. 1(B)]. The values of ϵ that were used to create Fig. 2(B) are indicated by the vertical dotted lines in Fig. 2(C).

Each horizontal bar in the persistence barcode shown in Fig. 2(C) indicates the *lifetime* of a homological 1-cycle—that is, the values of ϵ for which it exists. The red and blue bars reflect 0-homology and 1-homology respectively. The dimensions of the homology spaces (i.e., Betti numbers) can be found by counting the number of homological 1-cycles at any particular ϵ. For example, one can observe that $\beta_1 = 0$ when $\epsilon = 3.5$, $\beta_1 = 1$ when $\epsilon = 2.5$, and $\beta_1 = 2$ when $\epsilon = 1.5$. Clearly, the homological 1-cycles are undergoing bifurcations as ϵ varies. A persistence barcode is convenient to identify for each generator: the value ϵ_b of ϵ when it is "born" (i.e., the homological k-cycle does not exist when $\epsilon > \epsilon_b$); the value ϵ_d of ϵ when it "dies" (i.e., the homological k-cycle

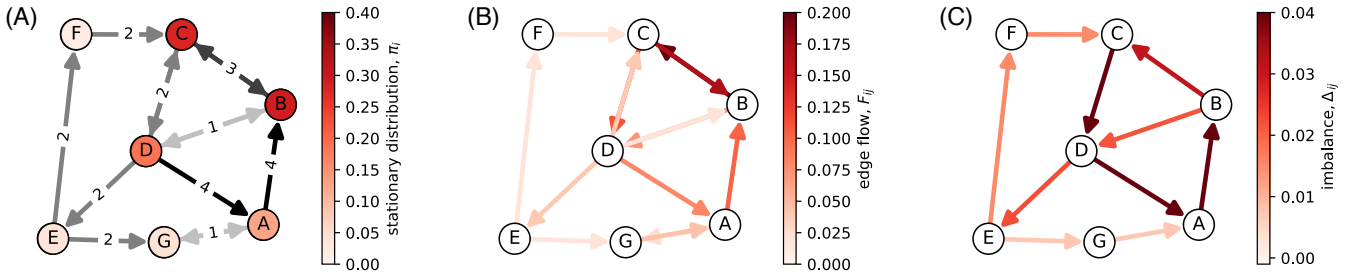

FIG. 3. **Stationary distribution, edge flows, and convection cycles for an irreversible MC.** We study a discrete-time random walk over a directed, weighted graph that resembles the undirected graph in Fig. 2, except that the edges are now either directed or bidirectional. (Recall that undirected graphs give rise to reversible MCs that lack convection cycles.) (A) The color of each vertex indicates the *stationary distribution* π_i of random walkers at each vertex i. (B) Edge colors indicate the *stationary flows* $F_{ij} = \pi_i P_{ij}$ across edges, i.e., the stationary fraction of random walkers that traverse each directed edge. (C) *Flow imbalances* $\Delta_{ij} = (F_{ij} - F_{ji})$ manifest as a pattern of *convection cycles*. By construction, $\Delta_{ij} = -\Delta_{ji}$, and we use arrows to indicate the directions of imbalance, e.g., $i \to j$ if $\Delta_{ij} > 0$.

does not exist when $\epsilon < \epsilon_d$); its *lifetime* $(\epsilon_d, \epsilon_b]$; and *lifespan* $|\epsilon_d - \epsilon_b|$. A cycle's lifespan quantifies its persistence under the filtration, and it is often interpreted as a proxy for the cycle's significance (although short-lifetime cycles can also be important in certain contexts).

C. Convection cycles for irreversible Markov chains (MCs)

We will apply persistence homology to study convection cycles in irreversible MCs [28], which we now briefly summarize. A discrete-time MC is a "memoryless" random process in which for time steps $t = 0, 1, 2, \ldots$, the system state $S_t \in \mathcal{V}$ satisfies the Markov property $\mathrm{P}[S_{t+1} = i | S_0 = i_0, ..., S_t = i_t] = \mathrm{P}[S_{t+1} = i | S_t = i_t]$, which implies that the probability of a state occurring at the next time step only depends on the current state and not earlier states. In our case, we consider MCs that correspond to a random walk on a (possibly) weighted and directed graph having an *adjacency matrix* $\mathbf{A}$ in which $A_{ij} \in \mathbb{R}$ is nonzero if (i, j) is an edge, $(i, j) \in \mathcal{E}$, and $A_{ij} = 0$ otherwise. We similarly define a *transition matrix*, $\mathbf{P} = \mathbf{D}^{-1}\mathbf{A}$, where $\mathbf{D}$ is a diagonal matrix with entries that encode the (possibly) weighted vertex degrees $\mathbf{D}_{ii} = \sum_j A_{ij}$. For directed graphs, each (i, j) is considered to be an ordered pair, and each $\mathbf{D}_{ii}$ encodes the out-degree of vertex i. Each matrix element P_{ij} gives the probability for a random walk to transition from vertex i to j. Letting $x_i(t)$ denote the probability that the system is in state i (or equivalently, the probability that a random walker is at vertex i) at time t, one can utilize the Markov property to obtain the linear discrete-time system $x_j(t + 1) = \sum_i x_i(t) P_{ij}$. By defining $\mathbf{x}(t) = [x_1(t), \ldots, x_{N_0}]^T$, one equivalently has

$$\mathbf{x}(t + 1)^T = \mathbf{x}(t)^T \mathbf{P}. \tag{3}$$

Since $\mathbf{x}(t)$ is a vector of probabilities, we assume that it is normalized in 1-norm, $\sum_i x_i(t) = 1$.

Herein, we focus on network flows after a system reaches a *stationary state*, in which case $\mathbf{x}(t)$ converges to a limiting vector $\pi = \lim_{t\to\infty} \mathbf{x}(t)$ that satisfies the eigenvalue equation $\pi^T = \pi^T \mathbf{P}$. By construction, π is a vector of probabilities and contains nonnegative entries. Furthermore, as a row-stochastic matrix, $\mathbf{P}$ has an eigenvalue equal to one (i.e., the largest eigenvalue) and its right dominant eigenvector is the vector containing 1's as entries. Our assumption of convergence requires that matrix $\mathbf{P}$ is an irreducible and aperiodic [29] or that the initial condition $\mathbf{x}(0)$ lies in a converging subspace. In the stationary state, the *stationary flow* across each edge (i, j) per time step is given by

$$F_{ij} = \pi_i P_{ij}. \tag{4}$$

We study *convection cycles* using an approach that was developed in [30]. Specifically, for each edge we define the stationary *flow imbalance*

$$\Delta_{ij} = F_{ij} - F_{ji}. \tag{5}$$

By construction, $\Delta_{ji} = -\Delta_{ij}$, and we say that the *imbalance direction* is from i to j when $\Delta_{ij} > 0$. Importantly, the defining feature of a *reversible MC* is that $\Delta_{ij} = 0$ for all i and j. That is, the directional flows match $\pi_i P_{ij} = \pi_j P_{ji}$ for any edge (i, j). This is the case for any undirected graph, since in this case $\mathbf{A} = \mathbf{A}^T$, and it follows that $\pi_i = D_{ii} / \sum_j D_{jj}$. In contrast, an *irreversible MC* yields asymmetric stationary flows and Δ_{ij} is nonzero for some edges. To formally define convection cycles, we consider a new graph $G_\Delta(\mathcal{V}, \mathcal{E}_\Delta)$ such that each positive value Δ_{ij} gives rise to a directed edge $(i, j, \Delta_{ij}) \in \mathcal{E}_\Delta$ having weight Δ_{ij}. We then define a *convection cycle* to be any non-intersecting closed path in $G_\Delta(\mathcal{V}, \mathcal{E}_\Delta)$.

In Fig. 3, we illustrate for an example MC how flow imbalances manifest as a pattern of convection cycles. In Figs. 3(A), 3(B), and Fig. 3(C), we use edge colors to indicate the stationary distribution π, stationary edge flows F_{ij}, and flow imbalances Δ_{ij}, respectively. Observe that some of the arrows in Figs. 3(A)–(B) are bidirectional, since some of the graph's edges are bidirectional. In contrast, the arrows in Fig. 3(C) are exclusively directed since they now indicate the directions of flow imbalances. There exists an edge $i \to j$ only if $\Delta_{ij} > 0$, which also implies $j \to i$ is not an edge since $\Delta_{ji} = -\Delta_{ij}$. Observe in Fig. 3(C) that this yields five convection cycles. In

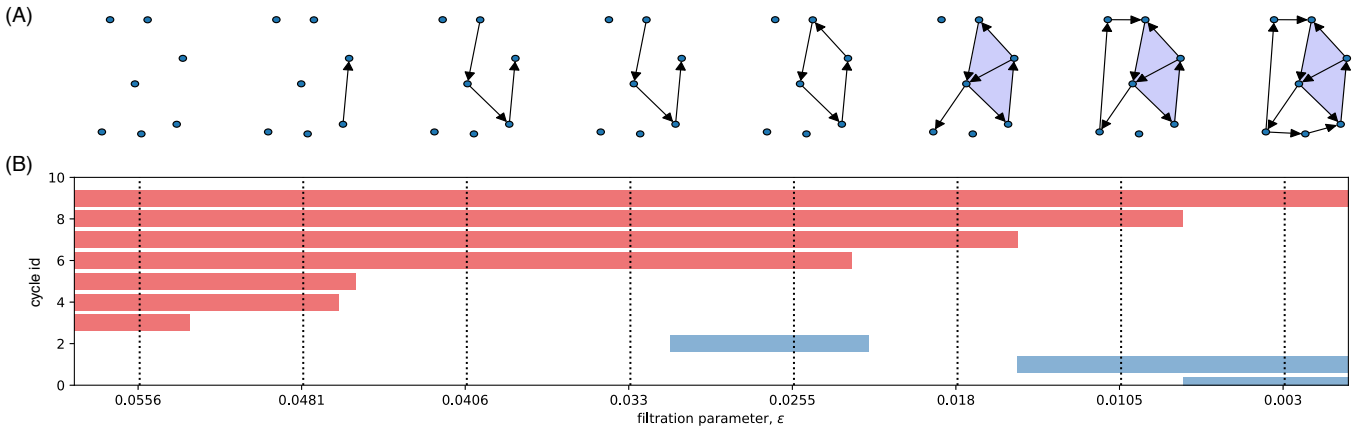

FIG. 4. **Persistent homology of convection cycles.** (A) Visualization of an EVC filtration applied to flow imbalances arising for the irreversible MC shown in Fig. 3, and we use the magnitude $|\Delta_{ij}|$ of flow imbalance as the filtration function $f : \mathcal{E} \to \mathbb{R}$. We indicate flow imbalances' directions with arrows, noting that the clique complexes that are constructed by the filtration are undirected, since the filtration does not incorporate information about edge directions. (B) Persistence barcodes for homological 1-cycles. Observe that the 1-cycle that first appears dies before the other 1-cycles are born.

Sec. III B, we will further discuss these convection cycles and their relation to homological 1-cycles.

Before continuing, we highlight that convection cycles revealed through flow imbalances [30] do not take into account the probability of transitioning to or away from a convection cycle, and so they are not necessarily "cyclic traps." That is, the presence of a convection cycle does not imply that it is unlikely for a random walker to leave (or move in an opposite direction as) the cycle. For example, observe in Fig. 3 that the counter-clockwise flow around convection cycle $A \to B \to C \to D$ is approximately 0.025, yet there is a flow of approximately 0.02 that leaves the cycle at node D, and a flow of approximately 0.15 moves in the opposite direction from node C to B. Future research will likely uncover complementary notions of convection with different advantages/disadvantages, and our proposed techniques using persistent homology can likely be similarly extended.

III. HOMOLOGICAL ANALYSES OF CONVECTION

We now employ persistent homology to automate the detection, characterization, and summarization of the homological patterns of convection cycles. In Sec. III A, we study the MC that was presented in Fig. 3. In Sec. III B, we discuss the relation between convection cycles and homological 1-cycles.

A. Persistent homology of convection cycles

Recall from Sec. II B that EVC filtrations were defined for an undirected graph with a scalar function defined on the edges. Therefore, given an MC corresponding to a (potentially) directed and weighted graph, we first consider the associated undirected graph. Then we study homology under an EVC filtration in which the filtration function $f : \mathcal{E} \longrightarrow \mathbb{R}$ is given by the magnitudes of the flow imbalances

$$f(i,j) = |\Delta_{ij}|. \tag{6}$$

In this way, the persistent homology that is revealed corresponds to the convection cycles that arise under flow imbalances.

In Fig. 4, we visualize persistence barcodes for an EVC filtration associated with the convection cycles shown in Fig. 3(C). Note that this figure is analogous to Fig. 2(C), where we had previously chosen the filtration function $f(i,j)$ to be the edge weights. Since we now use a different function f, the cycles now have different births, deaths, lifetimes and lifespans. Interestingly, the 1-cycle involving vertices $\{A, B, C, D\}$ is now born and dies before the other two 1-cycles are born. While there is an obvious connection between the EVC homology of a graph induced by edge weights and that which is induced by convective flows, this relation remains unclear and should be explored in future work.

We note that one could also construct EVC filtrations by increasing ϵ and retaining edges (i,j) for which $|\Delta_{ij}|$ is smaller than ϵ. In Appendix A, we provide an example illustrating why EVC filtrations with decreasing ϵ are superior to those with increasing ϵ for the goal of studying convection cycles. In particular, EVC filtrations that decrease ϵ focus on 1-cycles that are associated with large-flow convection cycles (i.e., large values of $|\Delta_{ij}|$), which we consider to be the ones that are more significant. In contrast, EVC filtrations that increase ϵ focus on 1-cycles that are associated with small-flow convection cycles (i.e., small values of $|\Delta_{ij}|$), which we consider to be less significant.

B. Comparing convection cycles and homological 1-cycles

We propose to study pattern formation for convection cycles using persistent homology techniques for homological

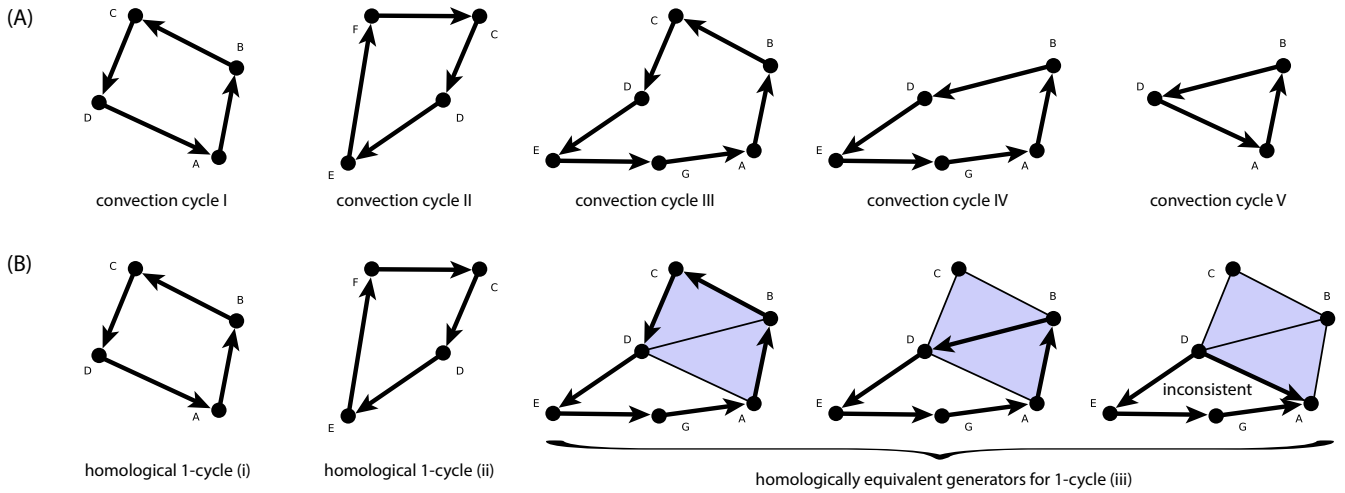

FIG. 5. **Relation between convection cycles and homological 1-cycles.** (A) The flow imbalances shown in Fig. 3(C) give rise to five convection cycles, which we label I–V. (B) Persistent homology using EVC filtrations applied to a network of flow imbalances reveals three *homological 1-cycles*, which we label (i)–(iii). Each homological 1-cycle represents a "1-dimensional hole" and can be represented by one or more *homological generator* (recall Sec. II A). Observe that there is a one-to-one correspondence between convection cycles I and II and homological 1-cycles (i) and (ii). In contrast, there are three homologically equivalent generators for 1-cycle (iii) as shown. Two of the generators correespond to convection cycles III and IV. The third generator does not correspond to a convection cycle, because the edge directions are not consistently in the same orientation (i.e., always clockwise or counter-clockwise).

1-cycles; however, one should keep in mind that these are two different notions for cycles. Homological 1-cycles are 1-dimensional holes for a topological space, and k-cycles generalize to higher dimensional by representing higher-dimensional holes (Sec. II A). In contrast, we define convection cycles to be closed non-backtracking paths in a directed graph that encodes flow imbalances (Sec. II C). In this section, we will clarify the relationship between homological k-cycles and convection cycles, thereby revealing the capabilities and disadvantages of existing persistent homology techniques for studying convection cycles. Continuing with the previous example [see Figs. 3-4], we find that flow imbalances give rise to five convection cycles, which we enumerate I–V and visualize in Fig. 5(A). In contrast, we identify three homological 1-cycles using persistent homology with EVC filtrations, which we enumerate (i)–(iii) and visualize in Fig. 5(B).

Observe that there is a one-to-one correspondence between convection cycles I and II and homological 1-cycles (i) and (ii). Also observe that homological 1-cycle (iii) has three homologically equivalent generators, and any of them can be used to represent the 1-cycle (which again, is defined as a 1-dimensional hole). Each subsequent generator can be obtained via a *topological retraction* in which a 2-simplex is collapsed down onto one of its edges. Interestingly, the first two homological generators for 1-cycle (iii) correspond to convection cycles III and IV. In contrast, the third generator corresponds to a loop that is not a convection cycle, since the flow-imbalances' directions do not point in a consistent direction along the cycle (i.e., clockwise or counter-clockwise). Finally, observe that convection cycle V is a boundary of a 2-simplex, and it therefore does not contribute to the 1-dimensional simplicial homology.

Thus, it is important to not misinterpret one notion of cycle for the other. At the same time, our findings in Fig. 5 also highlight that there is a need for new persistent homology techniques that cater specifically to convection cycles and directed graphs. For example, if one were to omit the 2-simplex that involves vertices A, B and D from the clique complexes that arise under an EVC filtration, then convection cycle V would coincide with a homological 1-cycle. However, the aim of this paper is not to develop new methods for persistent homology. Instead, we proposed to begin this pursuit by studying convection cycles using existing methods for persistent homology. Even though there is not an exact one-to-one match between convection cycles and homological 1-cycles, because they are closely related, we find that persistent homology can effectively detect and summarize convection cycles' patterns.

IV. APPLICATIONS

In this section, we apply our approach to two applications. In Sec. IV A, we study MCs arising for the Google PageRank algorithm, exploring how convection cycles are effected by the teleportation parameter α. In Sec. IV B, we study a type of emergent convection cycle called a chiral edge flow.

A. Teleportation is a homology regularizer for PageRank

We now study the persistent homology of convection cycles arising for the PageRank algorithm [25, 31], which is a popular technique to rank the importance of vertices in graphs. It has been applied to numerous applications (see survey [32]),

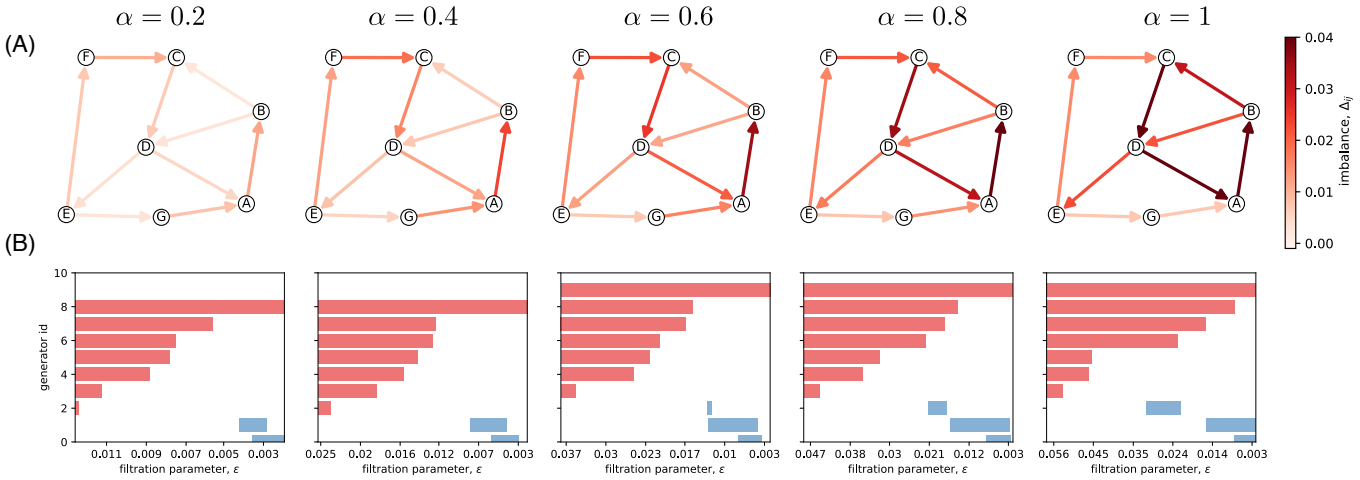

FIG. 6. **Persistent homology of convection cycles arising under PageRank.** We study the MC associated with PageRank for the directed graph from Fig. 3(A) under several choices of α. (A) Flow imbalances $\Delta_{ij}(\alpha)$ give rise to convection cycles. For clarity, we do not visualize flow imbalances for transitions due to teleportation. (B) Their homology changes with α, which is summarized by persistence barcodes.

but most notably, for many years it was utilized by Google to rank website and facilitate web search. The PageRank of a vertex i is given by the stationary density $\pi_i(\alpha)$ of MC having a transition matrix of the form

$$\mathbf{P}(\alpha) \equiv \alpha \mathbf{P} + (1-\alpha) N^{-1} \mathbf{1}\mathbf{1}^T, \tag{7}$$

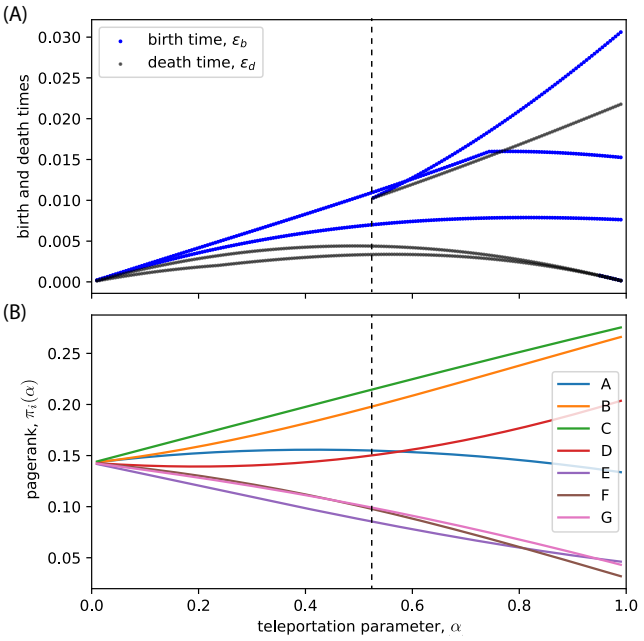

FIG. 7. **Bifurcation diagram summarizes homological changes onset by α.** (A) Birth and death times $\epsilon_{b,d}$ of 1-cycles arising under PageRank versus α. (B) For comparison, we depict the vertices' PageRanks $\pi_i(\alpha)$. Vertical dashed lines near $\alpha^* = 0.54$ highlight that there are three cycles when $\alpha > \alpha^*$, but only two when $\alpha < \alpha^*$.

where $\mathbf{P}$ is the transition matrix described in Sec. II C and $\alpha \in (0,1)$ is the *teleportation parameter*. As $\alpha \to 1$, the second term vanishes and $\mathbf{P}(\alpha) \to \mathbf{P}$. Usually, α is chosen to be near 1 (often 0.85) so that the second term can be considered as a small perturbation that improves the mathematical characteristics of $\mathbf{P}$—or more formally, it is a "'regularization" of matrix $\mathbf{P}$. In particular, when $\alpha \in (0,1)$ the matrix $\mathbf{P}(\alpha)$ is guaranteed to be irreducible, aperiodic and positive, and the Perron-Frobenius theorem [29] ensures that its dominant left eigenvector π is unique and has positive entries (i.e., $\pi_i(\alpha) > 0$ for all i). In other words, the PageRanks are well-defined for all vertices.

We now show that the introduction of teleportation also regularizes the homology of convection cycles. In this experiment, we construct EVC filtrations with the filtration function $f(i,j) = |\Delta_{ij}(\alpha)|$, which now depends on α. In Fig. 6(A), we illustrate for several choices of α the flow imbalances that arise under PageRank, which we apply to the graph from Fig. 3(A). In Fig. 6(B), we visualize their associated persistence barcodes, which we create using EVC filtrations. Note that the choice $\alpha = 1$ recovers the transition matrix, stationary distribution, flow imbalances, and persistence barcodes that were were previously studied in Figs. 3 and 4.

Observe that the homological patterns of convection cycles significantly change with α. For example, when α is sufficiently small, the homological 1-cycle $\{A, B, C, D\}$ vanishes—it is "washed out" by the introduction of teleportation. In other word, α is a *homology regularizer*. This is further illustrated in Fig. 7(A), where we plot the birth and death times of homological 1-cycles versus α. For comparison, we also plot the PageRanks $\pi_i(\alpha)$ in Fig. 7(B). The vertical line highlights that one of the 1-cycles vanishes when α decreases (approximately) below $\alpha^* = 0.54$

In Appendix A, we present additional experiments that explore convection cycles arising under PageRank with $\alpha = 0.8$. We show that homological 1-cycles arising for EVC filtrations with decreasing filtration parameter ϵ reveal patterns for large-

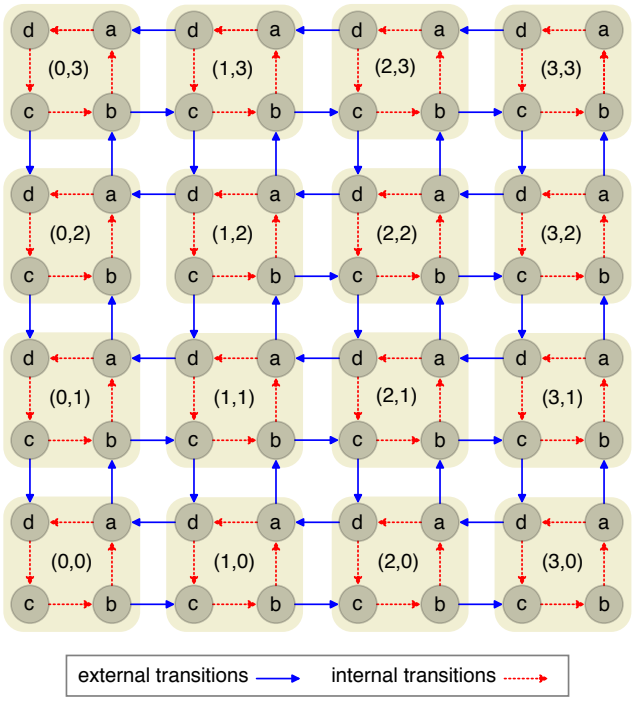

FIG. 8. **MC model for configuration dynamics of a monomer.** The system is contains two monomers of sizes s_1 and s_2, respectively, giving the external state (s_1, s_2). Moreover, there are four internal states: a, b, c, and d. Transitions involving changes to external and internal states occur at rates γ_{ex} and γ_{in}, respectively.

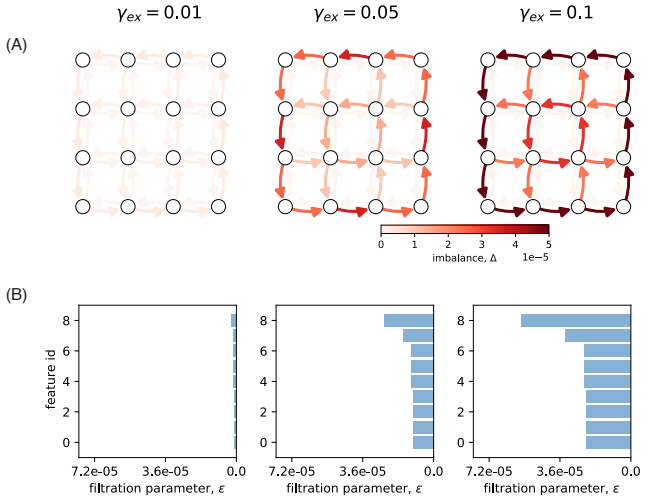

FIG. 9. **Persistent homology chiral edge flow.** (A) Flow imbalances Δ_{ij} between external states for the bi-monomer shown in Fig. 8 with $\gamma_{in} = 0.01$ and different γ_{ex}. In the limit $\gamma_{ex} \gg \gamma_{in}$ [22], there is an emergence of a chiral edge flow, i.e., a convection cycle around the lattice's outer boundary. (B) The corresponding persistence barcodes capture the emergence of this prominent convection cycle and other convection cycles within the lattice.

flow convection cycles. In contrast, when EVC filtrations are constructed with increasing ϵ, we find that the resulting homological 1-cycles relate to small-flow convection cycles, and in particular, those involving low probability teleportation transitions.

B. Persistent homology of chiral edge flows

Our second application investigates homological patterns of convection cycles that arise for an MC that models the stochastic configuration dynamics of two monomers. We adopt the same notation as in [22], which motivated our experiment. The monomer configuration (i.e., "external state") is given by the number of monomers of each type, (s_1, s_2), whereas the "internal state" is one of four possibilities: a, b, c, or d. Transitions that involve a change of internal state occur at rate γ_{in}, whereas transitions between involving external states (i.e., the addition or removal of a monomer) occur at rate γ_{ex}. The resulting MC can be visualized as a 2-dimensional lattice, which we visualize in Fig. 8.

In Fig. 9(A), we visualize flow imbalances for transitions between the external states. We fix $\gamma_{in} = 0.01$ and consider several γ_{ex}. Observe that as γ_{ex} increases, a large counterclockwise convection cycle emerges on the boundary (i.e., "edge") of the lattice. This type of convection cycle is called a *chiral edge flow*, and such convection cycles have important implications for the quantum Hall effect, biological rhythms, and the dynamics of monomers [22]. In Fig. 9(B), we visualize persistence barcodes for EVCs filtrations constructed using the method that we described in Sec. III A. The chiral edge cycle corresponds to the 1-cycle with having the largest lifespan, and its homology becomes more persistent (i.e., prominent) in the limit $\gamma_{ex} \gg \gamma_{in}$.

V. DISCUSSION

In this paper, we examined the patterns of convection cycles that arise under irreversible Markov chains (MCs) from the perspective of persistent homology. Our approach required formalizing a type of filtration (EVC filtration) for scalar functions that are defined on the edges of a graph, and we studied convection cycles by choosing the filtration function to be an MC's flow imbalances in the stationary state. Because Markov chains are crucial to so many diverse applications, we expect our methods to be broadly applicable across the sciences and engineering. Herein, we highlighted two such applications: the PageRank algorithm for centrality analysis and chiral edge flows that arise for the configuration dynamics of monomers. Our experiments revealed how system properties can act as homology regularizers of convection cycles, and we introduced homological bifurcation diagrams to summarize these changes. This approach automates the detection, summary, and examination of convection cycles over networks, places it on stronger mathematical and computational foundations, and paves the way for further investigation into convective flows on networks.

Additionally, our work highlights the need for new persis-

tent homology methods to study convection cycles as well as other functions and signals defined on directed graphs. In Sec. III B, we discuss the relation between convection cycles and homological 1-cycles, and we showed that these are two closely related, but notably different, notions of cycles. Sometimes there is a one-to-one correspondence between these cycles, and sometimes the relation is more complicated, due in part to the fact that a given homological k-cycle can be equivalently represented by possibly more than one homological generator. Such generators may or may not correspond to convection cycles. Moreover, convection cycles can also correspond to the boundaries of 2-simplices, and as such, they will not be identified via the traditional tools of persistent homology. Developing persistent homology techniques that cater to convection cycles, and which specifically account for edge directions, remains an important open challenge for the applied mathematics and physics communities.

Our work opens up several other new lines of research that are also worth noting. First, convection cycles were recently found to be an emergent property of multiplex Markov chains [30] in which a set of (intralayer) Markov chains are coupled together by another set of (interlayer) Markov chains. It would be interesting to employ persistent homology to gain a deeper understanding of this phenomenon. Second, chiral edge flows are known to be important to other applications including the quantum Hall effect and biological rhythms [22], and future work could utilize our methods to investigate these exciting applications. Notably, our methods can reveal convection cycles that exists in addition to a chiral edge flow, which may lead to new insights for these applications and other applications (e.g., reinforcement learning) that rely on irreversible Markov chains.

See [27] for a codebase that reproduces our results and can be used to study the persistent homology for convection cycles arising for other applications.

Appendix A: Convection cycles are better revealed by filtrations that decrease the filtration parameter ϵ versus increase ϵ

In Sec. II B, we defined EVC filtrations in which one decreases a filtration parameter ϵ, retaining edges for which $f(i,j) > \epsilon$. Our numerical experiments that study convection cycles using persistent homology use this approach and let the filtration be given by the flow imbalances $f(i,j) = |\Delta_{ij}|$. By decreasing ϵ, the cycles that are first revealed correspond to large-flow convection cycles, which we consider to be the ones that are more significant. One could also construct EVC filtrations by increasing ϵ and retaining edges for which $f(i,j) < \epsilon$. Here, we show that this latter filtration reveals 1-cycles that relate to small-flow convection cycles, which we consider to be insignificant.

In Fig. 10, we study EVC filtrations applied to flow imbalances arising under the PageRank algorithm with $\alpha = 0.8$ for the same MC that we investigated in Sec. IV A. In Fig. 10(A) and Fig. 10(B), we illustrate EVC filtrations with decreasing and increasing ϵ, respectively. Observe in Fig. 10(A) that the 1-cycles revealed by decreasing ϵ correspond to large-flow convection cycles. In contract, observe in Fig. 10(B) that the 1-cycles revealed by increasing ϵ are small-flow cycles that relate to low-probability transitions that occur due to teleportation.

ACKNOWLEDGMENTS

MQL and DT were supported in part by the National Science Foundation (DMS-2052720 and EDT-1551069) and the Simons Foundation (grant #578333).

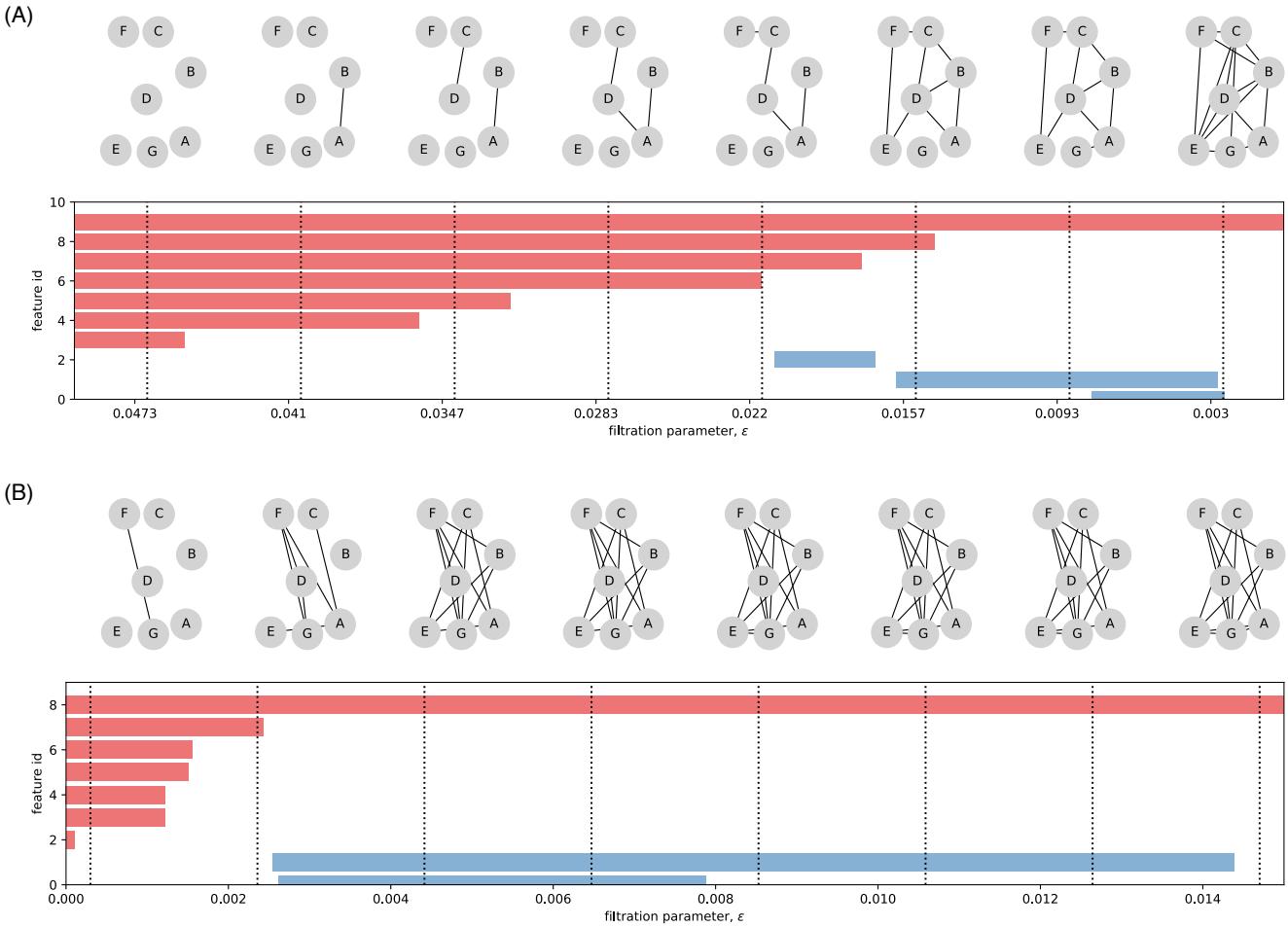

FIG. 10. **Comparing EVC filtrations with decreasing and increasing filtration parameter ϵ.** Extending our study in Sec. IV A that uses persistent homology to study convection cycles arising for a MC under the PageRank algorithm with $\alpha = 0.8$, we now study homological 1-cycles obtained via two different EVC filrations. (A) Similar to our results in Fig. 6, we construct EVC filtrations by including edges for which $|\Delta_{ij}| > \epsilon$ while decreasing ϵ. Observe that the 1-cycles reveal large-flow convection cycles that are associated with large values of $|\Delta_{ij}|$. (B) For comparison, we construct EVC filtrations by including weighted edges $|\Delta_{ij}| < \epsilon$ while increasing ϵ. Observe that these 1-cycles now correspond to small-flow convection cycles that are associated with small values of $|\Delta_{ij}|$. They primarily describe low-probability transitions that occur due to teleportation. In this work we focus on EVC filtrations with decreasing ϵ, since we consider high-flow convection cycles to be the ones that are most important.

[1] H. Edelsbrunner and J. Harer, *Computational topology: an introduction* (American Mathematical Soc., 2010).

[2] N. Otter, M. A. Porter, U. Tillmann, P. Grindrod, and H. A. Harrington, EPJ Data Science **6**, 17 (2017).

[3] F. A. Khasawneh and E. Munch, Mechanical Systems and Signal Processing **70**, 527 (2016).

[4] J. A. Perea and J. Harer, Foundations of Computational Mathematics **15**, 799 (2015).

[5] Y.-J. Xin and Y.-H. Zhou, in *Proceedings of ICSIPNN'94. International Conference on Speech, Image Processing and Neural Networks* (IEEE, 1994) pp. 764–767.

[6] F. Motta, C. Tralie, R. Bedini, F. Bini, G. Bini, H. Eramian, M. Gameiro, S. Haase, H. Haddox, J. Harer, *et al.*, in *2019 18th IEEE International Conference On Machine Learning And Applications (ICMLA)* (IEEE, 2019) pp. 1107–1114.

[7] M. Gabella, arXiv preprint arXiv:1902.08160 (2019).

[8] S. Liu, D. Wang, D. Maljovec, R. Anirudh, J. J. Thiagarajan, S. A. Jacobs, B. C. Van Essen, D. Hysom, J.-S. Yeom, J. Gaffney, *et al.*, IEEE transactions on visualization and computer graphics **26**, 291 (2019).

[9] D. Taylor, F. Klimm, H. A. Harrington, M. Kramár, K. Mischaikow, M. A. Porter, and P. J. Mucha, Nature communications **6**, 7723 (2015).

[10] G. Petri, P. Expert, F. Turkheimer, R. Carhart-Harris, D. Nutt, P. J. Hellyer, and F. Vaccarino, Journal of The Royal Society Interface **11**, 20140873 (2014).

[11] C. Giusti, E. Pastalkova, C. Curto, and V. Itskov, Proceedings of the National Academy of Sciences **112**, 13455 (2015).

[12] B. U. Kilic and D. Taylor, arXiv preprint arXiv:2201.02071 (2022).

[13] L. Kondic, A. Goullet, C. O'Hern, M. Kramar, K. Mischaikow, and R. Behringer, EPL (Europhysics Letters) **97**, 54001 (2012).

[14] M. Kramar, A. Goullet, L. Kondic, and K. Mischaikow, Physical Review E **87**, 042207 (2013).

[15] J. Liang, H. Edelsbrunner, P. Fu, P. V. Sudhakar, and S. Subramaniam, Proteins: Structure, Function, and Bioinformatics **33**, 1 (1998).

[16] T. Ichinomiya, I. Obayashi, and Y. Hiraoka, Biophysical Journal **118**, 2926 (2020).

[17] D. G. Kendall, The Annals of Mathematical Statistics , 338 (1953).

[18] J. F. Kingman, Journal of Applied Probability , 1 (1969).

[19] W. R. Gilks, S. Richardson, and D. J. Spiegelhalter, Markov chain Monte Carlo in practice , 1 (1995).

[20] L. Tierney, the Annals of Statistics , 1701 (1994).

[21] R. Parr and S. Russell, Advances in neural information processing systems , 1043 (1998).

[22] E. Tang, J. Agudo-Canalejo, and R. Golestanian, Physical Review X **11**, 031015 (2021).

[23] T. Kaczynski, K. Mischaikow, and M. Mrozek, *Computational homology*, Vol. 157 (Springer Science & Business Media, 2006).

[24] The GUDHI Project, *GUDHI User and Reference Manual*, 3.4.1 ed. (GUDHI Editorial Board, 2021).

[25] L. Page, S. Brin, R. Motwani, and T. Winograd, *The PageRank citation ranking: Bringing order to the web.*, Tech. Rep. (Stanford InfoLab, 1999).

[26] The notion of "dimension" is more interpretable for the case of a non-abstract simplicial complex, for which the vertices correspond to locations in a Euclidean metric space. In that case, every k-simplex is defined as the k-dimensional surface that is contained by its faces, which themselves are $(k-1)$ dimensional surfaces. For example, a 2-simplex is a triangle defined as the interior of 3 line segments (which are the 3 1-dimensional cofaces of the 2-simplex).

[27] M. Q. Le, "Codebase for persistent homology of convection cycles in network flows `https://github.com/minhquan89/Persistent-Homology-of-Convection-Cycles`,".

[28] L. Lovász *et al.*, Combinatorics, Paul erdos is eighty **2**, 1 (1993).

[29] R. B. Bapat, Linear Algebra and its Applications **275**, 3 (1998).

[30] D. Taylor, Physical Review Research **2**, 033164 (2020).

[31] A. N. Langville and C. D. Meyer, SIAM journal on matrix analysis and applications **27**, 968 (2006).

[32] D. F. Gleich, siam REVIEW **57**, 321 (2015).