PERSISTENT HOMOLOGY WITH K-NEAREST-NEIGHBOR FILTRATIONS REVEALS TOPOLOGICAL CONVERGENCE OF PAGERANK

MINH QUANG LE[✉*1] AND DANE TAYLOR[✉2]

[1]Faculty of Information Technology,
Ho Chi Minh City Open University, Viet Nam

[2]School of Computing; and the Department of Mathematics and Statistics,
University of Wyoming, USA

(Communicated by Vasileios Maroulas)

ABSTRACT. Graph-based representations of point-cloud data are widely used in data science and machine learning, including ϵ-graphs that contain edges between pairs of data points that are nearer than ϵ and kNN-graphs that connect each point to its k nearest neighbors. Recently, topological data analysis has emerged as a family of mathematical and computational techniques to investigate topological features of data using simplicial complexes. These are a higher-order generalization of graphs and many techniques such as Vietoris-Rips (VR) filtrations are also parameterized by a distance ϵ. Here, we develop kNN complexes as a generalization of kNN graphs, leading to kNN-based persistent homology techniques for which we develop stability and convergence results. We apply this technique to characterize the convergence properties PageRank, highlighting how the perspective of discrete topology complements traditional geometrical-based analyses of convergence. Specifically, we show that convergence of relative positions (i.e., ranks) is captured by kNN persistent homology, whereas persistent homology with VR filtrations coincides with vector-norm convergence. In general, kNN-based persistent homology is expected to be useful to other data-science applications in which the relative positioning of data points is more important than their precise locations.

1. **Introduction.** Topological data analysis (TDA) is a rapidly growing field of applied mathematics in which techniques from computational and applied topology are applied to extract structural information about the "shape" of data. TDA has been applied to numerous contexts ranging from visualization and dimensionality reduction [9, 40] and time series analyses [59] to applications in cosmology [65, 71], physical processes over networks [43, 68, 39, 38], neuroscience [61, 26, 13, 54], and systems biology [64, 44, 35, 37]. One of the main tools is the study of persistent homology, which can effectively reveal multiscale topological properties of data. This approach relies on examining how the homology of a topological space evolves as
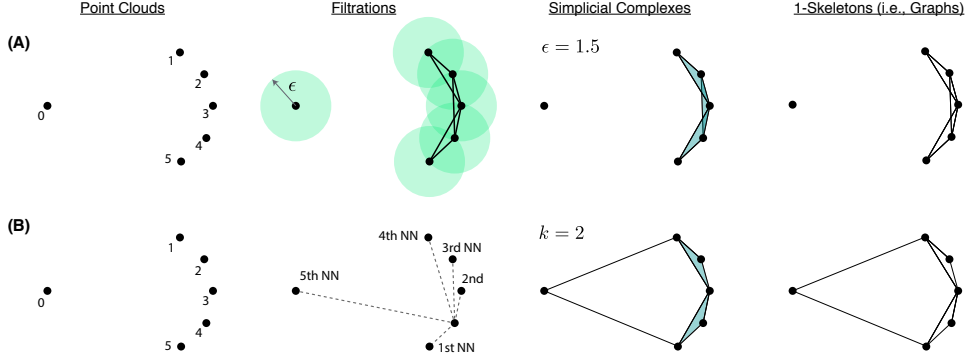
FIGURE 1. Visualizations of simplicial complexes and graphs resulting from (A) a Vietoris-Rips (VR) filtration and (B) our proposed k-nearest-neighbor (kNN) filtration. As shown in the second column, VR filtrations are parameterized by the radius ϵ of ϵ-balls that are centered at the points, whereas kNN filtrations are parameterized by the number k of nearest neighbors. (Dotted lines depict the nearest-neighbor orderings for node $i = 4$.) The third column depicts simplicial complexes that are obtained at some ϵ and k. The fourth columns shows their 1-skeletons, which are graphs in which k-simplices of $k > 1$ are discarded.

one applies a filtration, which is one of the basic notions in topology. There are many types of filtrations [2, 63, 31, 12, 24], however we will focus on the widely used Vietoris-Rips (VR) filtration [19]. In general, different filtrations reveal complementary insights, and it is important to develop additional filtrations that cater to different applications.

In the prototypical setting, one aims to construct and study empirical topological features for a set of data points, or *point cloud*. The approach involves using points to construct a filtered topological space, which is often represented by a simplicial complex in which the data points are 0-simplices, and higher-dimensional k-simplices are constructed via some set of rules. Often, k-simplices are added according to the pairwise distances between 0-simplices. As an example, in Fig. 1(A), we visualize one of the most commonly studied point-cloud filtrations, the Vietoris-Rips (VR) filtration. As shown, VR filtrations are constructed by considering simplicial complexes in which k-simplices are added between 0-simplices that are less than $\epsilon > 0$ distance apart.

In Fig. 1(B), we illustrate a different filtration that we propose and develop in this paper: *kNN filtrations*, whereby k-simplices are created according to k-nearest-neighbor sets. We note that while ϵ and kNN graphs are both very prevalent in the data-science and machine-learning literatures, TDA methods largely focus on ϵ-based simplicial complexes and filtrations, thereby limiting their potential utility to new applications. We develop kNN-based simplicial complexes as a generalization of kNN graphs, allowing us to develop kNN-based TDA methods including kNN
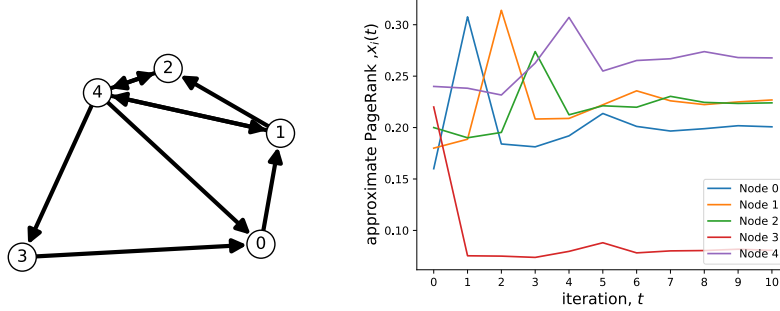
FIGURE 2. (left) An example graph. (right) Convergence of approximate PageRank values $x_i(t) \to \pi_i$ with t iterations.

nodes	$R_i(0)$	$R_i(1)$	$R_i(2)$	$R_i(3)$	$R_i(4)$	$R_i(5)$	$R_i(6)$	$R_i(7)$	$R_i(8)$	$R_i(9)$	$R_i(\infty)$
$i = 0$	5	1	4	4	4	4	4	4	4	4	4
$i = 1$	4	4	1	3	3	2	2	3	3	2	2
$i = 2$	3	3	3	1	2	3	3	2	2	3	3
$i = 3$	2	5	5	5	5	5	5	5	5	5	5
$i = 4$	1	2	2	2	1	1	1	1	1	1	1

TABLE 1. The relative orderings (i.e., node ranks $R_i(t)$) converge at $t \ge 9$. For each t, we indicate the top-ranked node by $R_i(t) = 1$.

persistent homology. We formulate and study persistent homology under kNN filtrations, constructing persistence diagrams and studying their robustness properties. We also define and study several local version of kNN complexes, filtrations, and persistent homology. By constructing filtrations using discrete sets, our approach relies on discrete topology, thereby contrasting approaches that are tied to continuous topological spaces [37]. In addition, we further analyze and explore homological features for converging sequences of point sets, exploring how the convergence of persistent diagrams for kNN filtrations contrasts that for VR filtrations.

We apply kNN persistent homology to study the ranking of nodes in graphs, and in particular, we study the convergence of rankings given by approximate PageRank values that are obtained by the power iteration method. Our work establishes a new connection between TDA and PageRank (see also [43]), and provides a topological approach for determining how many iterations are required for the node rankings to convergence.

We visualize this application and motivation in Fig. 2. In this application, approximate PageRank values $x_i(t) \to \pi_i$ asymptotically converge to their final values and the normed error $||\mathbf{x}(t) - \boldsymbol{\pi}||$ vanishes. However, as highlighted in Table 1, the orderings (i.e., node ranks) according to $x_i(t)$ values converge after only $t = 9$ iterations, even though $||\mathbf{x}(t) - \boldsymbol{\pi}|| > 0$ for all t. We apply kNN-based persistent homology to study of convergence for these relative orderings, which is a property that is crucial to PageRank and which is not revealed through VR filtrations. Although we focus here on PageRank, we expect kNN-based TDA to be widely useful to other applications in which the relative positioning of points—as opposed to the precise locations—is a property of main interest.

This paper is organized as follows. We provide background information in Sec. 2 and our main findings in Sec. 3. In Sec. 4, we apply this approach to develop a discrete topological perspective for the convergence of the PageRank algorithm, examining both empirical and synthetic networks. We provide a discussion in Sec. 5.

2. **Background information.** We first provide background information. In Sec. 2.1, we discuss simplicial complexes that are derived from point clouds. In Sec. 2.2, we define homology and persistent homology for simplicial complexes. In Sec. 2.3, we discuss the notions of stability and convergence for persistence diagrams.

2.1. **Simplicial complexes derived from enumerated point sets.**

2.1.1. *Point clouds.* We will define simplicial complexes in a geometric way by considering a finite set of enumerated points, or "point cloud." That is, consider a finite set $\mathcal{Y} = \{\mathbf{y}^{(i)}\}_{i=1}^N$ of cardinality $N = |\mathcal{Y}|$ that contains p-dimensional points $\mathbf{y}^{(i)} \in \mathbb{R}^p$ and assume distinct points $\mathbf{y}^{(i)} \neq \mathbf{y}^{(j)}$ (i.e., repeated points are disallowed). We enumerate the points $i \in \mathcal{V} = \{1, \ldots, N\}$ (i.e., we assign a surjective labeling $\mathcal{V} \mapsto \mathcal{Y}$) and equip $\mathbb{R}^p$ with the Euclidean metric. Note that we could have equivalently defined the point cloud using an N-tuple with elements from $\mathbb{R}^p$. To facilitate later discussion, it is helpful to further define $\mathcal{A}_p^N = (\mathbb{R}^p)^N$ to be the space of enumerated point sets in $\mathbb{R}^p$ containing N elements so that $\mathcal{Y} \in \mathcal{A}_p^N$.

Definition 2.1 (Euclidean Metric Space). Let $\mathbb{R}^p$ be the set of all ordered p-tuples, or vectors, over the real numbers and $d : \mathbb{R}^p \times \mathbb{R}^p$ be the Euclidean metric

$$d(\mathbf{x}, \mathbf{y}) = ||\mathbf{x} - \mathbf{y}||_2 = \sqrt{\sum_{q=1}^p (x_q - y_q)^2},$$

where $\mathbf{x} = (x_1, .., x_p), \mathbf{y} = (y_1, .., y_p) \in \mathbb{R}^p$. Then the **Euclidean metric space** is given by $(\mathbb{R}^p, d)$.

Definition 2.2 (Euclidean Ball, or ϵ-ball). A p-dimensional Euclidean ball $\mathcal{B}_\epsilon(\mathbf{x})$ centered at $\mathbf{x}$ with radius ϵ is defined by

$$\mathcal{B}_\epsilon(\mathbf{x}) = \{\mathbf{y} \in \mathbb{R}^p : d(\mathbf{x}, \mathbf{y}) \leq \epsilon\}.$$

Definition 2.3 (Pairwise Distance Map). Let $\mathcal{Y} = \{\mathbf{y}^{(i)}\}_{i=1}^N \in \mathcal{A}_p^N$ be an enumerated point set. Then we define the **pairwise distance map** $f : \mathcal{A}_p^N \to \mathbb{R}_+^{N \times N}$ to encode the distances between all pairs of points so that $f_{ij}(\mathcal{Y}) = d(\mathbf{y}^{(i)}, \mathbf{y}^{(j)})$. We similarly define a map $f_i : \mathcal{A}_p^N \to \mathbb{R}_+^N$ for each row i of matrix $[f_{ij}]$ so that $[f_i(\mathcal{Y})]_j = f_{ij}(\mathcal{Y})$.

2.1.2. *Simplicial complexes.* Given an enumerated point set $\mathcal{Y} = \{\mathbf{y}^{(i)}\}$ indexed by $i \in \mathcal{V} = \{1, 2, \ldots, N\}$, we will define simplicial complexes in which the locations and connectivity among k-simplices are derived from $\mathcal{Y}$. That is, we define a vertex $v_i = \{i\}$ at each location $\mathbf{y}^{(i)}$. A k-dimensional simplex $\mathcal{S}^k$, or **k-simplex**, is defined by a subset of $\mathcal{V}$ having cardinality $k + 1$. For example, each 0-simplex is a vertex, each 1-simplex is an edge, 2-simplices are informally "triangles" that must be defined using $k + 1 = 3$ nodes. Because the nodes are spatially embedded in $\mathbb{R}^p$, each k-simplex is a k-dimensional geometrical object. It is then natural to consider their $(k-1)$-dimensional faces, which can be defined as follows. A face of a k-simplex S^k is a subset of S^k with cardinality k, i.e , with one of the elements of S^k omitted. If S_f^{k-1} is a face of simplex S^k, then S^k is called *coface* of S_f^{k-1}. For

example, the 0-simplices at the start and end of a 1-simplex are its faces, and the set of 1-simplices that are adjacent to a 0-simplex are its cofaces.

A simplicial complex X is a collection of k-simplices with the property that if $S^k \in X$, then all the faces of S^k are also members of X. Note that the k-simplices in X can have different dimension k, and we define the dimensionality of X to be the maximum dimension of X's simplices. The notions of face and coface can also be extended to *abstract simplicial complexes* that lack spatial coordinates, and one can also define connections between k-simplices of the same dimension by considering if their faces and cofaces overlap. Note that with a given abstract simplicial complex X, we can construct a **geometric realization** of X. As in [17], this realization is defined as an injective map $f : V(X) \to \mathbb{R}^p$ with $V(X)$ denoted as the union of members in X. Then, f is called a geometric realization of X in $\mathbb{R}^p$ if $f(X)$ is a simplicial complex in $\mathbb{R}^p$. By abuse of notation, now we also consider $f(X)$ as a geometric realization of X and denoted as $|X|$. Two k-simplices S_i^k and S_j^k are said to be lower adjacent if they have a common face, and they are upper adjacent if they are both faces of a common $(k+1)$-simplex. For any $S^k \subset X$ we define its degree, denoted by $deg(S^k)$, to be the number of cofaces of S^k. We use X^k to denote the subset of k-simplices in X with dimension that is exactly k.

Note that a graph, while typically defined via two sets (nodes, edges), can also be interpreted as a 1-dimensional abstract simplicial complex, since it contains k-simplices of dimension $k \leq 1$. Moreover, for any simplicial complex, one can obtain an associated graph that is its 1-skeleton, whereby one discards any k-simplices of dimension $k > 1$. More generally, a κ-skeleton of a simplicial complex X can be obtained by discarding all k-simplices of dimension $k > \kappa$. Thus, a simplicial complex can be understood as a generalization of a graph that allows for higher-order relationships between nodes. To emphasize this connection, we will make no distinction between 1-simplices in a simplicial complex X and the edges of a graph, and we will refer to them interchangeably.

2.1.3. *Two simplicial complexes based on ϵ.* There are many ways in which one can construct a set of simplices involving the vertices $\mathcal{V}$ associated with a given a set of points $\mathcal{Y} = \{\mathbf{y}^{(i)}\}_{i=1}^N \in \mathcal{A}_p^N$. While we focus on enumerated point sets, these constructions can also be applied to non-enumerated point sets. Most approaches stem from considering ϵ-balls centered at the points $\mathcal{Y}$. Here, we present two closely related simplicial complexes that are parameterized by a distance threshold ϵ and are often motivated by the assumption that the point cloud lies on a low-dimensional manifold.

Definition 2.4 (Čech Complex [6])**.** Given a collection of points $\mathcal{Y} = \{\mathbf{y}^{(i)}\}_{i=1}^N \in \mathcal{A}_p^N$, the Čech complex, $\mathcal{C}_\epsilon$, is the abstract simplicial complex whose k-simplices are determined by the $(k+1)$-tuples of points $\{\mathbf{y}^{(i)}\}_{i=1}^{k+1}$ whose closed $(\epsilon/2)$-ball neighborhoods have a point of common intersection.

The *Vietoris-Rips (VR) complex* is closely related and is defined as follows.

Definition 2.5 (Vietoris-Rips Complex [3])**.** Given a collection of points $\mathcal{Y} = \{\mathbf{y}^{(i)}\}_{i=1}^N \in \mathcal{A}_p^N$, the Rips complex, $\mathcal{R}_\epsilon$, is the abstract simplicial complex whose k-simplices correspond to $(k+1)$-tuples of points $C = \{\mathbf{y}^{(i)}\}_{i=1}^{k+1}$, with $\mathbf{y}^{(i)} \in \mathcal{R}_\epsilon, i = 1, .., k+1$, whose pairwise distance satisfy $||\mathbf{y}^{(i)} - \mathbf{y}^{(j)}||_2 \leq \epsilon$ for any $\mathbf{y}^{(i)}, \mathbf{y}^{(j)} \in C$.

Notably, Čech complexes are sometimes preferred over VR complexes because of their relation to a continuous topological space that is associated with unions of ϵ-balls, as established in the following Theorem.

Theorem 2.6 (Čech theorem [37])**.** *The **Čech theorem** (or, equivalently, the "Nerve theorem") states that a Čech complex $\mathcal{C}_\epsilon$ has the homotopy type of the union of closed $(\epsilon/2)$-balls about the point set $\mathcal{Y} = \{\mathbf{y}^{(i)}\}_{i=1}^N$.*

This result is somewhat surprising since a Čech complex $\mathcal{C}_\epsilon$ is an abstract simplicial complex, which is a discrete topological space of potentially high dimension. In contrast, the union of closed $(\epsilon/2)$-balls is a continuous topological space (i.e., subset of $\mathbb{R}^p$), and so it is interesting that these two spaces have the same homotopy type. While an identical relation has not been identified for VR complexes, it has been shown that they are closely related to Čech complexes through the following interleaving result.

Lemma 2.7 (Relation Between Čech and Vietoris-Rips Complexes [37])**.** *For any $\epsilon > 0$, there is a chain of inclusion maps for Čech complexes and VR complexes*

$$\mathcal{R}_\epsilon \subseteq \mathcal{C}_{\epsilon\sqrt{2}} \subseteq \mathcal{R}_{\epsilon\sqrt{2}}.$$

From a practical perspective, applications involving TDA often focus on VR complexes because they are more computationally efficient to compute than Čech complexes. That is, it is easier to compute whether the distance between two points is less than ϵ versus check whether the intersection of ϵ-balls is nonempty. As such, TDA methods based on VR complexes are very popular in applied settings, and we will later conduct numerical experiments that use them as a baseline for comparison. Given the relations between VR and Čech complexes and the unions of ϵ-balls, these all may approximate the structure of a manifold, assuming all data points lie on the manifold. Such an assumption, however, is not appropriate for every data set.

Before continuing, we note that simplicial complexes can be constructed in ways that do not require a set of points. For example, one can construct abstract simplicial complexes based on a given graph.

Definition 2.8 (Clique Complex [73])**.** The **clique complex** $Cl(G)$ of an undirected graph $G = (\mathcal{V}, \mathcal{E})$ is a simplicial complex where $\mathcal{V}$ are nodes of G and each k-clique (i.e. a complete subgraph with k nodes) in G corresponds to a $(k-1)$-simplex in $Cl(G)$. More precisely, it is the simplicial complex

$$Cl(G) = \mathcal{V} \cup \mathcal{E} \cup \left\{ \sigma \Big| \binom{\sigma}{2} \subseteq \mathcal{E} \right\},$$

where $\sigma \subset \mathcal{V}$ denotes a simplex and $\binom{\sigma}{2} \in \mathcal{V} \times \mathcal{V}$ denotes the set of pairs that can be obtained by selecting two 0-simplices from σ (which must be an edge in the graph).

It is worth noting that the *1-skeleton* (i.e., 0-simplices and 1-simplices) of any clique complex recovers its associated graph $G = (\mathcal{V}, \mathcal{E})$. That is, the construction of a graph's clique complex is an invertible transformation.

Similar to the construction in Def. 2.5, we can easy observe the relation between the Rips complex and the corresponding clique complex. The Vietoris-Rips complex $\mathcal{R}_\epsilon$ of a collection of points $\mathcal{Y} = \{\mathbf{y}^{(i)}\}_{i=1}^N \in \mathcal{A}_p^N$ is considered as a clique complex of the graph whose vertex set is $\mathcal{Y}$, and each pair $\mathbf{y}^{(i)}, \mathbf{y}^{(j)}$ of points is connected when $||\mathbf{y}^{(i)} - \mathbf{y}^{(j)}||_2 \leq \epsilon$.

2.2. **Topological data analysis (TDA) and persistent homology.** TDA usually refers to the study of a point cloud by examining its multiscale features from the perspective of persistent homology. That is, after representing a data set by a simplicial complex, one can study its topological structure via its homology. To obtain multiscale insights, it is also useful to study how that homology changes and persists as one varies a parameter, such as a distance threshold ϵ.

2.2.1. *Homology of simplicial complexes.* We formulate homology by considering functions defined over k-simplices within a simplicial complex. A chain complex over a field $\mathbf{F}$ is a tuple $(\mathcal{C}, \partial)$ where $\mathcal{C}$ is a collection $\{\mathcal{C}_k\}_{k \in \mathbf{N}}$ of vector spaces together with a collection ∂ of $\mathbf{F}$-linear maps $\{\partial_k : C_k \longrightarrow C_{k-1}\}_{k \in \mathbf{N}}$ such that $\partial_{k-1} \circ \partial_k = 0$ for all integers k. Here, a vector space $\mathcal{C}_k$ is associated with a set of simplicial k-chains, where a *simplicial k-chain* is defined as a formal linear combination over $\mathbf{F}$ of an abstract simplicial complex K. The maps ∂_k are called boundary maps. The k-cycles of the complex are the elements that are sent to zero by the map ∂_k; the k-boundaries are the elements in the image of ∂_{k+1}. A map of chain complexes $f : (C, \partial) \longrightarrow (C', \partial')$ is a collection $\{f_k : C_k \longrightarrow C_{k'}\}_{k \in \mathbf{N}}$ of $\mathbf{F}$-linear map such that $f_{k-1} \circ \partial_k = \partial'_k \circ f_k$ for all natural numbers k .

The k-cycles form a vector space, and so do the k-boundaries; we denote these vector spaces by Z_k and B_k, respectively. The kth homology of a chain complex (C, ∂) over a field $\mathbf{F}$ is the quotient vector space

$$H_k((C, \partial)) = Z_k / B_k.$$

The number

$$\beta_k(C) \equiv \mathsf{dim}(H_k(C, \partial)) = \mathsf{dim} Z_k - \mathsf{dim} B_k$$

is called the k-th Betti number. For a geometric intuition, the dimension of $H_k(X)$ can be thought of as the number of 'k-dimensional holes' of X:

- The 0th Betti number β_0 is the number of connected components.
- The 1st Betti number β_1 counts the number of loops, or 1-cycles.
- The 2nd Betti number β_2 counts the number of voids, or 2-cycles.

The dimension of a simplicial complex is the maximum over the dimensions of its simplices. If X is a simplicial complex of dimension d then $H_k(X) = 0$ for all $k \geq d$.

Any map of chain complexes $\Xi : (C, \partial) \longrightarrow (C', \partial')$ induces a linear map on homology [55]:

$$H(\Xi) : H_k((C, \partial)) \longrightarrow H_k((C', \partial'))$$

To simplify notation, we later use Ξ in place of $H(\Xi)$.

2.2.2. *Filtrations of simplicial complexes.* To discuss how homology changes and persists, it is necessary to define filtrations. While one can study filtrations for general topological spaces, we will focus here on filtrations of simplicial complexes. Given a simplicial complex X, a *filtration* is a totally ordered set of subcomplexes X_i of X, indexed by the real numbers. Commonly, a filtration is constructed with an ascending filtation parameters [2]; however, they can also be useful to construct them by allow the filtration parameter to be descending [43]. The totally ordering itself is called a *filter*. We call a simplicial complex together with a filtration a *filtered simplicial complex*.

There are different ways to define a filtration [2], and we will focus herein on the popular *Vietoris-Rips filtration*.

Definition 2.9 (Vietoris-Rips Filtration [2])**.** Let $G = (\mathcal{V}, \mathcal{E})$ be an undirected, all-to-all graph in which each node $i \in \mathcal{V}$ has a position $\mathbf{y}^{(i)} \in \mathbb{R}^p$ in a metric space with metric d. Consider a weight function $W : \mathcal{V} \times \mathcal{V} \longrightarrow \mathbb{R}$ defined on the edges $\mathcal{E}$ so that $W(i,j) = d(\mathbf{y}^{(i)}, \mathbf{y}^{(j)})$ encodes the distance between $\mathbf{y}^{(i)}$ and $\mathbf{y}^{(j)}$. Let $Cl(G)$ be the associated clique complex for $G(\mathcal{V}, \mathcal{E})$. For any $\delta \in \mathbb{R}$, the 1-skeleton $G_\delta = (V, E_\delta)$ is defined as the subgraph of G where $\mathcal{E}_\delta \subseteq \mathcal{E}$ includes only the edges such that $W(i,j) \leq \delta$. Then, for any $\delta, \delta' \in \mathbb{R}$, we define the *Vietoris-Rips filtration* by

$$\{Cl(G_\delta) \longrightarrow Cl(G_{\delta'})\}_{0 \leq \delta \leq \delta'}.$$

In this work, we will only consider when the distance function is the Euclidean metric, although one can in principle use other metrics. That said, we note in passing that one can also construct filtrations in a variety of ways including functional metric filtrations [2], node-based clique filtrations [63], k-clique filtrations [63], weighted simplex filtrations [31], node function based filtrations [2], Dowker sink and source filtrations [12], and the intrinsic Čech filtrations [24].

2.2.3. *Persistent homology and persistence diagrams.* Persistent homology involves studying how homology changes (or more precisely, identifies when it does not change) during a filtration. Consider a simplicial complex, for all $i \leq j$ the inclusion maps $X_i \hookrightarrow X_j$ induce $\mathbf{F}$-linear maps $\Xi_{i,j} : H_k(X_i) \longrightarrow H_k(X_j)$ on simplicial homology. For a generator $x \in H_k(X_i)$ with $x \neq 0$, we say that x dies in $H_k(X_j)$ if $j > i$ is the smallest index for which $\Xi_{i,j}(x) = 0$. We similarly say that $x \in H_k(X_i)$ is born in $H_k(X_i)$ if $x \notin \mathbf{im}(\Xi_{t,i})$ for all $t < i$. We can represent the lifetime of x by the half open interval $[i, j)$. If $\Xi_{i,j}(x) \neq 0$ for all $i < j \leq l$, then we say that x lives forever and we represent its lifetime by the interval $[i; \infty)$.

The k-th persistent homology vector spaces of a filtered simplicial complex X are defined as $H_k^{i,j} = \mathbf{im}(\Xi_{i,j})$, and the total kth persistent homology of X is defined as $\oplus_{i=1}^l H_k(X_i)$. By the Correspondence Theorem of Persistent Homology [74], for each k we can assign to the total k-th persistent homology vector space a finite well-defined collection of half open intervals, i.e., its so called *barcode*. An alternative way to represent persistent homology graphically is given by persistence diagrams, in which case each open interval $[i, j)$ is represented by the point (i, j) in $\mathbb{R}^2$. See Fig. 3 for an example persistence diagram, which we will more formally define in the next section.

To proceed, we must introduce some definitions. Let X be a simplicial complex and $g : X \longrightarrow \mathbb{R}$ be a tame function. A homological critical value for a function g is defined as follows:

Definition 2.10 ([14])**.** A homological critical value of g is a real number a for which there exists an integer k such that for all sufficiently small $\epsilon > 0$, the map $H_k(g^{-1}(-\infty, a - \epsilon]) \longrightarrow H_k(g^{-1}(-\infty, a + \epsilon])$ induced by inclusion is not an isomorphism.

Now, we need a special class of functions that is associated with the number of it's homological critical values to discuss the stability theorem in 2.16.

Definition 2.11 (Tame [14])**.** A function $g : X \longrightarrow \mathbb{R}$ is tame if it has a finite number of homological critical values and the homology groups $H_k(g^{-1}(-\infty, a))$ are finite-dimensional for all $k \in \mathbb{Z}$ and $a \in \mathbb{R}$.
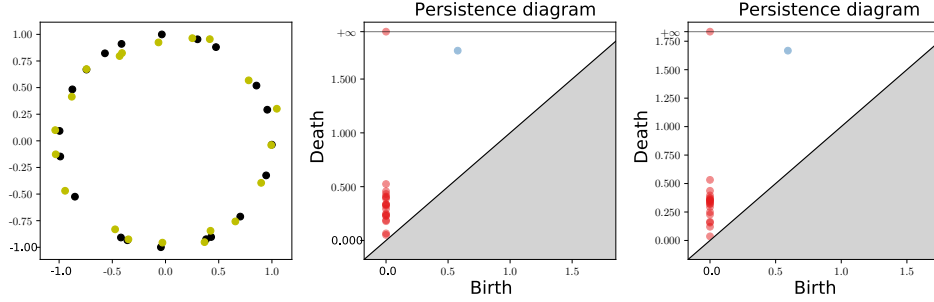
FIGURE 3. Example of stability for persistence diagrams resulting from Vietoris-Rips filtrations. (left) Two point-cloud sets $\mathcal{Y} = \{\mathbf{y}^{(i)}\}_{i=1}^{N}$ and $\mathcal{Z} = \{\mathbf{z}^{(i)}\}_{i=1}^{N}$ are close with respect to the L_∞ norm. The center and right panels depict their associated persistence diagrams $D_\mathcal{Y}$ and $D_\mathcal{Z}$, which are also close with respect to the bottleneck distance.

In particular, Morse functions on compact manifolds are tame, as well as piecewise linear functions on finite simplicial complexes and, more generally, Morse functions on compact Whitney-stratified spaces [28]. Assuming a fixed integer k, we define $G_x = H_k(g^{-1}(-\infty, x])$, and for any $x < y$, we let $g_x^y : G_x \longrightarrow G_y$ be the map induced by inclusion of the sub-level set of x in that of y.

Lemma 2.12 (Critical Value Lemma [14]). *If some closed interval $[x, y]$ contains no homological critical value of g, then g_x^y is an isomorphism for every integer k.*

We now formally define a **persistence diagram**. Using the same notation as above, we write $G_x^y = \mathbf{img}(g_x^y)$ for the image of G_x in G_y. By convention, we set $G_x^y = \{0\}$ whenever x or y is infinite. The group G_x^y is called the persistent homology group.

Let $g : X \longrightarrow \mathbb{R}$ be a tame function and denote $(a_i)_{i=1,\dots,n}$ as its homological critical values, and let $(b_i)_{i=0..n}$ be an interleaved sequence in which $b_{i-1} < a_i < b_i$ for all i. We set $b_{-1} = a_0 = -\infty$ and $b_{n+1} = a_{n+1} = +\infty$. For two integers $0 \leq i < j \leq n + 1$, we define the multiplicity of the pair (a_i, a_j) by:

$$\mu_i^j = \beta_{b_{i-1}}^{b_j} - \beta_{b_i}^{b_j} + \beta_{b_i}^{b_{j-1}} - \beta_{b_{i-1}}^{b_{j-1}}$$

where $\beta_x^y = \mathbf{dim}(G_x^y)$ denotes the persistent Betti numbers for all $-\infty \leq x \leq y \leq +\infty$.

Definition 2.13 (Persistence Diagram [14]). The persistence diagram $D(\mathcal{T}) \subset \overline{\mathbb{R}}^2$ of a filtered topological space $\mathcal{T}$, such as X, is the set of points $[a_i, a_j)$, which are counted with multiplicity μ_i^j for $0 \leq i < j \leq n + 1$, along with all points on the diagonal, which are counted with infinite multiplicity.

The points $[a_i, a_j)$ are often visualized as a scatter plot, but can also be visualized as "barcodes", in which each interval $[a_i, a_j)$ is depicted by a thick horizontal line.

2.3. **Stability and convergence of persistence diagrams.** Persistence diagrams and barcodes are widely used as a concise representation for the multiscale

homological features of a point-cloud data set. As such, it is important to understand their robustness to perturbations, which can represent data error or noise [43]. One can also study for a converging sequence of point clouds whether their associated persistence diagrams also converge. In this section, we present basic results related to stability and convergence.

2.3.1. *Stability.* We first discuss stability in a general setting, which requires several definitions. Given two persistence diagrams, we can quantify their dissimilarity using various metrics. In particular, we will consider the following metric.

Definition 2.14 (Bottleneck Distance [14]). The bottleneck distance between two multisets of points, $\mathcal{Y}$ and $\mathcal{Z}$, is given by

$$d_B(\mathcal{Y}, \mathcal{Z}) = \inf_{\gamma} \sup_{\mathbf{y}} \|\mathbf{y} - \gamma(\mathbf{y})\|_{\infty} \ , \tag{1}$$

where $\mathbf{y} \in \mathcal{Y}$ and γ ranges over all bijections from $\mathcal{Y}$ to $\mathcal{Z}$. Here, we interpret each point with multiplicity k as k individual points, and the bijection is between the resulting sets.

Next, we present triangulable spaces. As in [21], an *underlying space* of a simplicial complex K is defined as $\bigcup K = \bigcup_{\sigma \in K} |\sigma|$, where $|\sigma|$ is the geometric realization of σ. If there exists a simplicial complex K such that $\bigcup K$ is homeomorphic to X, then X is triangulable.

Next, we highlight a filtration that is important for understanding the stability of persistence diagrams.

Definition 2.15 (Height filtration [69]). Let $X \subset \mathbb{R}^p$ be a finite simplicial complex. With any unit vector $v \in \mathbb{S}^{p-1}$, we define a filtration $X(v)_{v \in \mathbb{R}}$, called *height filtration*, that is parameterized by a height $r \in \mathbb{R}$ where:

$$X(v)_{v \in \mathbb{R}} = \{x \in X : x \cdot v \le r\}$$

In Fig. 3, we show that it two point clouds are similar, then their associated persistence diagrams are similar. That is, persistence diagrams are 'stable' with respect to small changes, which can be quantified through the bottleneck distance.

Theorem 2.16 (Stability Theorem [8]). *Let X be a triangulable space with continuous tame functions $g, h : X \longrightarrow \mathbb{R}$, and let D_g and D_h denote their associated persistence diagrams obtained using a height filtration. Then the bottleneck distance between these persistence diagrams satisfies a global uniform bound*

$$d_B(D_g, D_h) \le \|g - h\|_{\infty}.$$

While Thm. 2.16 describes the stability of persistence diagrams with respect to changes to the functions to which filtrations are applied, one can also use it to equivalently prove the stability of persistence diagrams for when the function changes due to perturbations of points. We present the following example.

Corollary 1 (Stability of VR Persistence Diagrams to Point Perturbations). *Let $\mathcal{Y} = \{\mathbf{y}^{(i)}\}_{i \in \mathcal{V}} \in \mathcal{A}_p^N$ with $\mathcal{V} = \{1, ..., N\}$ denote a set of N points, and let $\mathcal{Z} = \{\mathbf{z}^{(i)}\}_{i \in \mathcal{V}} \in \mathcal{A}_p^N$ be a set of perturbed points such that $\|\mathbf{y}^{(i)} - \mathbf{z}^{(i)}\|_2 \le \epsilon$ for all i. Further, let $D_{\mathcal{Y}}$ and $D_{\mathcal{Z}}$ be their associated persistence diagrams for VR filtrations. It then follows that $d_B(D_{\mathcal{Y}}, D_{\mathcal{Z}}) \le 2\epsilon$.*

Proof. Our proof follows from Theorem 2.16. Let $f : \mathcal{A}_p^N \longrightarrow \mathbb{R}_+^{N \times N}$ be the pairwise distance function in which $f_{ij}(\mathcal{Y}) = ||\mathbf{y}^{(i)} - \mathbf{y}^{(j)}||_2$. Consider two standard, complete simplicial complexes K and M embedded in $\mathbb{R}^{N+1}$ that are correspondingly constructed from two vertices sets $\mathcal{Y}$ and $\mathcal{Z}$. Let g and h be two functions that are defined on 1-dimensional simplices within K and M such that $g_{ij} = f_{ij}(\mathcal{Y})$ and $h_{ij} = f_{ij}(\mathcal{Z})$. Here, g_{ij} and h_{ij} are denoted for two entries of two matrices that representing for two functions g and h respectively. Over higher-dimensional simplices, the functions g and h are inherited from the maximum value of the adjacent 1-simplices. It follows that the persistence diagrams $D_\mathcal{Y}$ and $D_\mathcal{Z}$ are equivalent to those resulting from height filtrations for tame functions g and h defined on the clique complex of an all-to-all graph with edge weights $\{g_{ij}\}$ and $\{h_{ij}\}$. We find

$$d_B(D_\mathcal{Y}, D_\mathcal{Z}) \leq ||g - h||_\infty = \max_{i,j \in \mathcal{V}} |f(\mathbf{y}^{(i)}, \mathbf{y}^{(j)}) - f(\mathbf{z}^{(i)}, \mathbf{z}^{(j)})|$$

$$= \max_{i,j \in \mathcal{V}} \left| ||\mathbf{y}^{(i)} - \mathbf{y}^{(j)}||_2 - ||\mathbf{z}^{(i)} - \mathbf{z}^{(j)}||_2 \right|.$$

However, if we define $\mathbf{z}^{(i)} = \mathbf{y}^{(i)} + \mathbf{e}^{(i)}$, then $||\mathbf{e}^{(i)}||_2 \leq \epsilon$ for any $i \in \mathcal{V}$ and

$$||\mathbf{z}^{(i)} - \mathbf{z}^{(j)}||_2 = ||(\mathbf{y}^{(i)} - \mathbf{y}^{(j)}) + (\mathbf{e}^{(i)} - \mathbf{e}^{(i)})||_2$$

$$\leq ||(\mathbf{y}^{(i)} - \mathbf{y}^{(j)})||_2 + ||(\mathbf{e}^{(i)} - \mathbf{e}^{(j)})||_2.$$

It follows that

$$\left| ||\mathbf{y}^{(i)} - \mathbf{y}^{(j)}||_2 - ||\mathbf{z}^{(i)} - \mathbf{z}^{(j)}||_2 \right| \leq ||(\mathbf{e}^{(i)} - \mathbf{e}^{(j)})||_2 \leq 2\epsilon.$$

$\square$

2.3.2. *Convergence.* The stability of persistence diagrams also has important consequences for the convergence of a sequence of persistence diagrams that is associated with convergent sequence of point clouds. Focusing on VR filtrations, consider a sequence of point clouds $\mathcal{Y}(t) = \{\mathbf{y}^{(i)}(t)\}_{i=1}^N$ of fixed size $N = |\mathcal{Y}(t)|$ for each $t \in \mathbb{N}_+$ with $\mathbf{y}^{(i)}(t) \in \mathbb{R}^p$. Assume for each i that the sequence converges $\mathbf{y}^{(i)}(t) \to \mathbf{y}^{(i)}$ such that $\|\mathbf{y}^{(i)}(t) - \mathbf{y}^{(i)}\|_2 \leq 1/t$. Note that such a convergence criterion can be ensured by considering a subsequence in which each subsequent element is chosen so that the bound is true for all $i \in \{1, \ldots, N\}$. Let $D_{\mathcal{Y}(t)}$ and $D_\mathcal{Y}$ be the persistence diagrams for VR filtrations applied to both point clouds. By using Corollary 1, we obtain convergence of the associated persistence diagrams since

$$d_B(D_{\mathcal{Y}(t)}, D_\mathcal{Y}) \leq 2\|\mathbf{y}^{(i)}(t) - \mathbf{y}^{(i)}\|_2 \leq \frac{2}{t},$$

which converges to 0 as $t \to \infty$.

3. **Topological data analysis using k nearest neighbors.** We now present our main findings: an approach for persistent homology that is based on the relative positioning of points according to their k-nearest-neighbor (kNN) sets. Our approach introduces kNN complexes and filtrations using a filtration parameter $k \in \mathbb{N}_+$, thereby contrasting VR and other filtrations that use a distance threshold ϵ as the filtration parameter. We will show that kNN-based persistent homology has certain advantages that can benefit applications for which the relative positioning of data points is important, i.e., as opposed to their precise locations.

This section is organized as follows. In Sec. 3.1, we define kNN orderings of points. In Sec. 3.2, we develop kNN filtrations, use them to study kNN persistent

homology, and compare them to the study of VR filtrations. In Sec. 3.3, we analyze the stability and convergence of persistence diagrams resulting from kNN filtrations.

3.1. **kNN graphs and kNN complexes.** We begin with a definition for kNN orderings. One complication is that the ordering of nearest neighbors is not necessarily symmetric, and so we will also describe additional transformations that we will use to symmetrize orderings prior to constructing kNN graphs, kNN complxes, and kNN filtrations.

3.1.1. *kNN orderings.* We first formalize the concept of k-nearest neighbors.

Definition 3.1 (k-Nearest-Neighbor Orderings)**.** Let $\mathcal{V} = \{1, \ldots, N\}$ enumerate a set of points $\mathcal{Y} = \{\mathbf{y}^{(i)}\}_{i \in \mathcal{V}}$ with $\mathbf{y}^{(i)}(t) \in \mathbb{R}^p$ in a Euclidean metric space and $\{f_{ij} = ||\mathbf{y}^{(i)} - \mathbf{y}^{(j)}||_2\}$ be their pairwise distances resulting from Definition 2.3. For each i, let k_{ij} denote the **nearest-neighbor order** of $\mathbf{y}^{(j)}$ with respect to $\mathbf{y}^{(i)}$ so that $\mathbf{y}^{(j)}$ is the (k_{ij})-th nearest neighbor of $\mathbf{y}^{(i)}$ ($k_{ii} \equiv 0$). The ordering is formally defined for each fixed i by $\{k_{ij}\}_{j=1}^N = \phi(\{f_{ij}\}_{j=1}^N)$, where $\phi : \mathbb{R}^N \longrightarrow \mathbb{N}^N$ is the "argsort function".

In the above, we assume that the orderings are well-defined in that for a given $i \in \mathcal{V}$, the entries in the set $\{f_{ij}\}_{j=1}^N$ are unique. We will refer to such point clouds as being **non-degenerate**. While we don't focus on degenerate point clouds herein, they can in principle occur. If two distances f_{ij} and $f_{ij'}$ are the same, and they both correspond to order k, then we assign an ordering at random so that either $k_{ij} = k$ and $k_{ij'} = k + 1$, or vice versa. If even more distances are the same, then we similarly assign them a relative ordering uniformly at random. In principle, the argsort function ϕ can be defined in a variety ways to handle the ordering of repeated entries.

Given the above definition of kNN orderings, we now introduce an associated map between a set of points and an associated matrix with entries k_{ij}.

Definition 3.2 (kNN Ordering Function)**.** Let $\mathcal{A}_p^N$ be the space of enumerated N-point sets in $\mathbb{R}^p$ and recall the row-defined pairwise distance function $f_i : \mathcal{A}_p^N \longrightarrow \mathbb{R}_+^N$ given in Definition 2.3. Letting ϕ be the argsort function, we define the map **k-nearest-neighbor (kNN) ordering function** $F : \mathcal{A}_p^N \longrightarrow \mathbb{N}_+^{N \times N}$ as the matrix-valued function $F = [F_1, \ldots, F_N]^T$ in which each row F_i is defined by $F_i = \phi \circ f_i : \mathcal{A}_p^N \longrightarrow \mathbb{N}_+^N$.

3.1.2. *kNN Neighborhoods.* We will use kNN orderings of points to define local neighborhoods $\mathcal{N}_{ik}$ for each point $i \in \mathcal{V}$, noting that any nested sequence of neighborhoods defines a local filtration of the nodes $\mathcal{V}$.

Definition 3.3 (kNN neighborhoods)**.** Given a set of kNN orderings $\{k_{ij}\}$ for $i, j \in \mathcal{V}$, we define the **kNN neighborhoods** as the sets $\mathcal{N}_{ik} = \{j | k_{ij} \leq k\}$ for each $i \in \mathcal{V}$ and $k \in \{0\} \cup \mathcal{V} \setminus \{N\}$.

3.1.3. *kNN symmetrization.* Importantly, the kNN orderings and kNN neighborhoods are not symmetric—that is, $j \in \mathcal{N}_{ik}$ does not imply $i \in \mathcal{N}_{jk}$. Because we would like to use the kNN neighborhood sets to construct filtrations of simplicial complexes that contain undirected simplices, we will now describe how to construct symmetric kNN neighborhoods.

Definition 3.4 (Symmetrized kNN Orderings and Neighborhoods)**.** We define three types of **symmetrized kNN neighborhoods**—$\mathcal{N}_{ik}^{min}$, $\mathcal{N}_{ik}^{trans}$ and $\mathcal{N}_{ik}^{max}$—that use, respectively, the following three symmetrizations of kNN orderings $\{k_{ij}\}$:

$$\tilde{k}_{ij} = \min\{k_{ij}, k_{ji}\}$$
$$\overline{k}_{ij} = (k_{ij} + k_{ji})/2$$
$$\hat{k}_{ij} = \max\{k_{ij}, k_{ji}\}. \tag{2}$$

Notably, these kNN neighborhoods satisfy the following nestedness relation.

Lemma 3.5 (Nestedness of Symmetrized kNN Neighborhoods)**.** *Consider fixed k and i. Then the neighborhood sets $\mathcal{N}_{ik}^{min}, \mathcal{N}_{ik}^{trans}, \mathcal{N}_{ik}^{max}$ satisfy the following nestedness relationships:*

$$\mathcal{N}_{ik}^{max} \subseteq \mathcal{N}_{ik}^{trans} \subseteq \mathcal{N}_{ik}^{min}.$$

Proof. For any $k_{ij}, k_{ji} \in \mathbb{R}$, one has $\tilde{k}_{ij} \leq \overline{k}_{ij} \leq \hat{k}_{ij}$ by definition of the min and max functions. If $j \in \mathcal{N}_{ik}^{max}$, then $\hat{k}_{ij} \leq k$. This implies $\tilde{k}_{ij} \leq \overline{k}_{ij} \leq \hat{k}_{ij} \leq k$, and so $j \in \mathcal{N}_{ik}^{min} \subseteq \mathcal{N}_{ik}^{trans} \subseteq \mathcal{N}_{ik}^{max}$. □

3.1.4. *kNN graphs and complexes.* We now define graphs and clique complexes using symmetrized kNN orderings and neighborhoods.

Definition 3.6 (Symmetrized kNN Graph)**.** Given a set of points enumerated by $\mathcal{V}$, let $\mathcal{E}_k = \{\{i, j\} | i \in \mathcal{V}, j \in \mathcal{N}_{ik}\}$ be a set of undirected edges connecting pairs of nearest neighbors, which are symmetrically defined by choosing $\mathcal{N}_{ik} \in \{\mathcal{N}_{ik}^{min}, \mathcal{N}_{ik}^{trans}, \mathcal{N}_{ik}^{max}\}$. We denote the respective graphs $\mathcal{G}_k(\mathcal{V}, \mathcal{E}_k)$, depending on symmetrization method, by $\mathcal{G}_k^{min}$, $\mathcal{G}_k^{trans}$, and $\mathcal{G}_k^{max}$.

Definition 3.7 (Symmetrized kNN Clique Complex)**.** Given a symmetrized kNN graph $\mathcal{G}_k \in \{\mathcal{G}_k^{min}, \mathcal{G}_k^{trans}, \mathcal{G}_k^{max}\}$ with nodes $\mathcal{V}$ and edges $\mathcal{E}_k$, we construct its associated clique complex $X_k = Cl(\mathcal{G}_k)$, and we similarly use a superscript to denote the method of symmetrization, i.e., X_k^{min}, X_k^{trans}, and X_k^{max}.

Corollary 2 (Nestedness of Symmetrized kNN Graphs and kNN Complexes)**.** *Symmetrized kNN graphs and kNN complexes satisfy the following nestedness relations*

$$\mathcal{G}_k^{max} \hookrightarrow \mathcal{G}_k^{trans} \hookrightarrow \mathcal{G}_k^{min}$$
$$X_k^{max} \hookrightarrow X_k^{trans} \hookrightarrow X_k^{min}. \tag{3}$$

Proof. The results follow immediately from Lemma 3.5, which proved the nestedness of kNN neighborhoods. □

3.2. Filtrations and persistent homology using kNN complexes. We now formulate filtrations and persistent homology using symmetrized kNN complexes.

3.2.1. *kNN filtrations.* Varying k gives rise to a sequence of nested sets called a filtration, and we will define several types based on the different symmetrization methods. We will also define local and global versions of filtrations.

Definition 3.8 (Global kNN Filtration)**.** Let $X_k, X_{k'} \in \{X_k^{min}, X_k^{trans}, X_k^{max}\}$ be kNN complexes with the same symmetrization method. Then we define a **kNN-filtered simplicial complex**

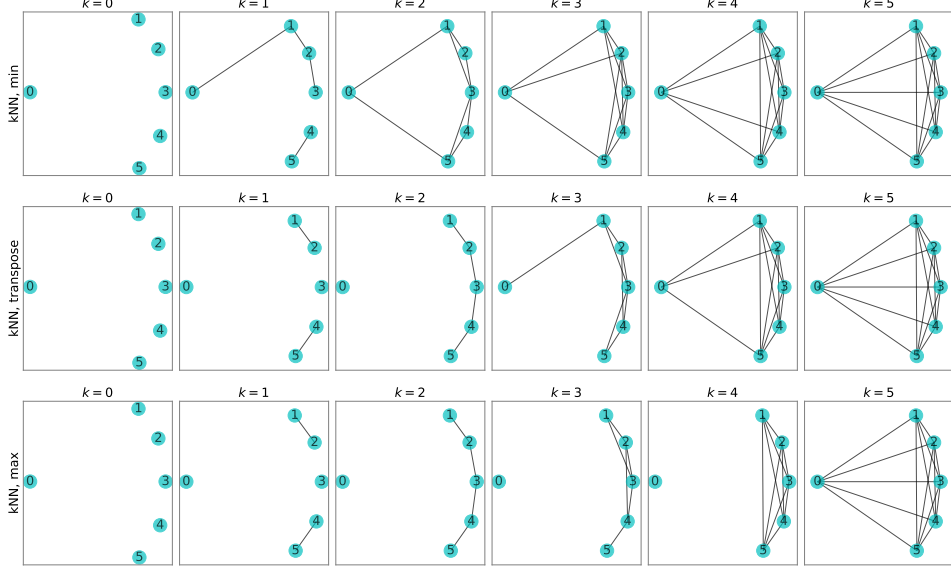$$\{X_k \longrightarrow X_{k'}\}_{0 \leq k \leq k' \leq N-1}.$$

FIGURE 4. Visualization of kNN-filtered simplicial complexes for a point cloud using the three types of symmetrization given by Definition 3.4. Comparing across the columns, observe the nestedness property given by Corollary 2.

In Fig. 4, we illustrate global kNN filtrations with the three different symmetrization methods given by Definition 3.4. For simplicity, we only visualize 0-simplices and 1-simplices. By comparing the kNN complexes across a given column, one can observe the nestedness relations defined by Corollary 2.

3.2.2. *kNN persistent homology.* We formulate persistent homology for kNN filtrations of a simplicial complex anaologous to that defined for a VR filtrations (recall Sec. 2.2), and in Fig. 5, we visualize persistence diagrams for both (left) a kNN,min filtration (kNN,min denotes for using the symmetrization $\tilde{k}_{ij}$ of kNN ordering) and (right) a VR filtration for an example point cloud. The red and blue persistence barcodes indicate 0-dimensional and 1-dimensional cycles, respectively. Observe that the kNN filtration reveals a 1-cycle, whereas the VR filtration does not. Hence, filtrations of simplicial complexes (specifically clique complexes) according to pairwise distances and kNN provide complementary homological information.

In the next sections, we will further compare persistence diagrams resulting from kNN and VR filtrations; however, it's worth highlighting several differences here. First, one benefit of using kNN sets versus a distance threshold ϵ is that filtrations are standardized. That is, the filtration parameter range $k \in [0, N]$ for kNN filtrations is always the same for a set of N points. In contrast, different point clouds have different lengths scales, and so the range of a distance-based filtration parameter is in general not standardized. One could seek to standardize the ranges of VR filtration parameters by standardizing distances; however, there are many ways to normalize a point cloud, such as dividing by the mean distance or by the distances' standard deviation. Given that one could implement a variety of normalization approaches, comparing VR filtrations across different point clouds that have different length
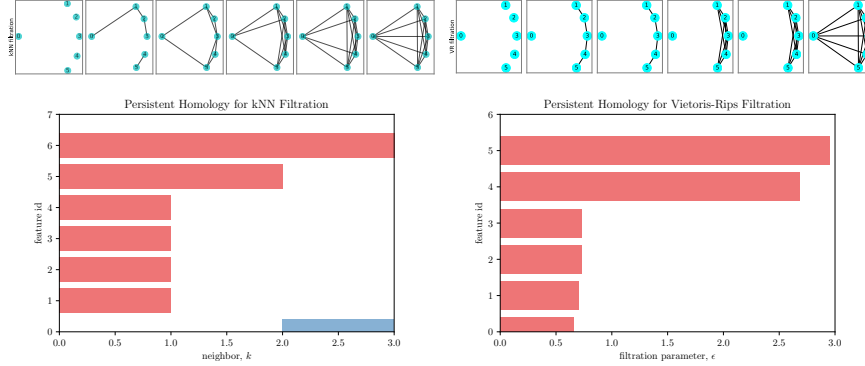
FIGURE 5. Comparison of persistent homology for an example point cloud using two different filtrations: (left) our proposed kNN,min filtration; and (right) a VR filtration. The red and blue persistence barcodes indicate 0-dimensional and 1-dimensional cycles, respectively. Observe that the kNN filtration reveals a 1-cycle that was born at $k = 2$ and die at $k = 3$, whereas the VR filtration does not.

scales or different dimensions is not straightforward. In contrast, kNN filtrations are standardized by definition. The highlighting feature of kNN filtrations is that focusing on the structure within k^{th}-nearest neighborhood of each member in a given point cloud dataset, so homological features will be changed if we add more points into our data. However, these locally associated neighborhoods demonstrated vital roles in many applications, such as in Topological Deep Learning [29], [47], RNA sequence analysis [16], etc. Therefore, the explorations under homological features via this filtration is potentially necessary. In our scenario, the functions of this filtration are carved via considering some "neighborhood-convergence" processes, such as Google's PageRank convergence in 4.2, when only putting our notification in the kth-nearest neighborhood of each point in a row vector.

Second, the filtration parameter $\epsilon \in \mathbb{R}_+$ for VR filtrations can in principle take on any positive number. In contrast, $k \in \{0, \ldots, N-1\}$ can only take on integer values (or in the case of the symmetrization method of $trans$, half integers as well). Potentially, this reduced space of possible filtration parameters could benefit the computationally efficiency of implementing kNN filtrations, although we don't explore that pursuit herein. That said, our experiments generally find kNN filtrations to have higher computational complexity, since they require both computing pairwise distances as well as their orderings.

3.3. **Stability and convergence of kNN homology.** Here, we will study two key properties for the space of persistence diagrams that follow from kNN filtrations: stability and convergence. We begin by by formalizing when point sets have identical or similar kNN orderings.

3.3.1. *kNN equivalance of enumerated point sets.* We begin by defining an equivalence class over the space of enumerated point sets $\mathcal{A}_p^N$.

Definition 3.9 (kNN Equivalence of Point Sets)**.** Let F denote the neighbor-ordering function given by Def. 3.2 and consider two enumerated point sets $\mathcal{Y}, \mathcal{Z} \in$

$\mathcal{A}_p^N$ such that $\mathcal{Y} = \{\mathbf{y}^{(i)}\}_{i \in \mathcal{V}}$ and $\mathcal{Z} = \{\mathbf{z}^{(i)}\}_{i \in \mathcal{V}}$. Point sets $\mathcal{Y}$ and $\mathcal{Z}$ are said to be **kNN equivalent** if $F(\mathcal{Y}) = F(\mathcal{Z})$. Elements $\mathbf{y}^{(i)}$ and $\mathbf{z}^{(i)}$ are said to be **locally kNN equivalent** if $F_i(\mathcal{Y}) = F_i(\mathcal{Z})$. Letting $\sim$ represent kNN equivalence, we further define $[\mathcal{Y}] = \{\mathcal{Y}' | \mathcal{Y}' \sim \mathcal{Y}\}$ as the kNN equivalence class of $\mathcal{Y}$.

Indeed, one can show
- **Reflexivity: $\mathcal{Y} \sim \mathcal{Y}$.**
  It is trivial, since $F(\mathcal{Y}) = F(\mathcal{Y})$.
- **Symmetry: if $\mathcal{Y} \sim \mathcal{Z}$ then $\mathcal{Z} \sim \mathcal{Y}$.**
  It is also trivial, since $F(\mathcal{Y}) = F(\mathcal{Z})$ implies $F(\mathcal{Z}) = F(\mathcal{Y})$.
- **Transitivity: if $\mathcal{Y} \sim \mathcal{Z}$ and $\mathcal{Z} \sim \mathcal{W}$ then $\mathcal{Y} \sim \mathcal{W}$.**
  This follows from the observation that $F(\mathcal{Y}) = F(\mathcal{Z}) = F(\mathcal{W})$.

Notably, kNN equivalences can be generalized in a few ways. First, the equivalence relation could be defined using the symmetrized versions of kNN orderings (i.e., $\tilde{k}$, $\bar{k}$, and $\hat{k}$). We can also define a slight variation in which the preservation of kNN orderings is only required for the nearest neighbor orderings up to a finite size $k \leq K$, which we will refer to as "K-bounded" kNN equivalence. In that case, one has $\mathcal{Y} \sim \mathcal{Z}$ only if $[F(\mathcal{Y})]_{ij} = [F(\mathcal{Z})]_{ij}$ for any $[F(\mathcal{Y})]_{ij}, [F(\mathcal{Z})]_{ij} \leq K$.

We also highlight that in the above definitions, the two enumerated point clouds $\mathcal{Y}$ and $\mathcal{Z}$ are matched so that for each i, the neighbor orderings of $\mathbf{y}^{(i)}$ are identical to those for $\mathbf{z}^{(i)}$, i.e., $F_i(Y) = F_i(Z)$. One can also generalize the equivalence relation to consider all possible permutations so that $\mathcal{Y} \sim \mathcal{Z}$ if $F_i(Y) = F_{j_i}(Z)$ for some permutation $\{j_1, j_2, \ldots, j_N\}$ of indices $\mathcal{V} = \{1, 2, \ldots, N\}$. We refer to such a kNN equivalence as "permutation-invariant," and it could also act on the space of non-enumerated point sets; however, we restrict our attention to enumerated point sets in this paper. Unless stated otherwise, we will assume multiple point sets to be optimally "matched" so that $\sum_i ||F_i(Y) - F_{j_i}(Z)||$ is minimized when $j_i = i$ for each i.

3.3.2. *kNN-preserving transformations.* We observe that if points undergo perturbations that are sufficiently small, then it is possible to move points without changing any of the neighbor orderings. To make this more precise, we define a family of point-cloud transformations that have this property.

Definition 3.10 (Local and Global kNN-Preserving Transformations)**.** Consider an enumerated point set $\mathcal{Y} = \{\mathbf{y}^{(i)}\}_{i=1}^N \in \mathcal{A}_p^N$, and let $F : \mathcal{A}_p^N \longrightarrow \mathbb{N}_+^{N \times N}$ and $F_i : \mathcal{A}_p^N \longrightarrow \mathbb{N}_+^N$ be the full and row-wise kNN ordering functions given by Def. 3.2. We define a function $h : \mathcal{A}_p^N \longrightarrow \mathcal{A}_p^N$ that takes the form

$$h(\mathcal{Y}) = \{h_1(\mathbf{y}^{(1)}), \ldots, h_N(\mathbf{y}^{(N)})\}$$

to be **global kNN preserving** for $\mathcal{Y}$ if $F(h(\mathcal{Y})) = F(\mathcal{Y})$. That is, the action of F on $\mathcal{Y}$ is invariant to the transformation h. Similarly, we define h to be **local kNN preserving** for point $\mathbf{y}^{(i)} \in \mathcal{Y}$ if F_i is invariant to the transformation, i.e., $F_i(h(\mathcal{Y})) = F_i(\mathcal{Y})$.

Note for a given $\mathcal{Y}, \mathcal{Z} \in \mathcal{A}_p^N$ that there exists a global kNN preserving transformation h for $\mathcal{Y}$ such that $\mathcal{Z} = h(\mathcal{Y})$ if and only if $\mathcal{Y}$ and $\mathcal{Z}$ are kNN equivalent, $\mathcal{Y} \sim \mathcal{Z}$. Like kNN equivalence, kNN-preserving transformations can also be generalized, e.g., to preserve kNN equivalence with respect to symmetrization, K-boundedness, or point-label permutations.

As one example, consider a set of three points $x_i \in \{0, 2, 3\} \subset \mathbb{R}$ and the transformation $h(\{0, 2, 3\}) = \{0, 2.1, 3\}$. It is both local and global kNN-preserving, since the perturbation $2 \to 2.1$ is sufficiently small such that none of the neighbor orderings change. In fact, all isometric transformations including (i.e., translation, rotation, reflection and glide reflection) are global kNN-preserving transformations for any $\mathcal{Y} \in \mathcal{A}_p^N$ because they leave the pairwise distances unchanged and thus the kNN orderings are also unchanged.

However, there are many transformations that are not kNN preserving. In Fig. 6, we illustrate an example where a small perturbation to one point can change the neighbor ordering in a discontinuous way. Consider a set of points $\mathcal{V} = \{a, b, c\}$ with locations $\mathcal{Y} = \{x_a, x_b, x_c\}$ where $x_a = -1$, $x_c = 1$ and $x_b = -\epsilon \in \mathbb{R}$ (for some small value $0 < \epsilon \ll 1$). Then the neighbor ordering matrix is given in Table 2. Now consider the transformation $h_1(x_a) = x_a$, $h_2(x_b) = x_b + 2\epsilon$, and $h_3(x_c) = x_c$, then the neighbor ordering matrix changes, as shown in Table 3.

(A)

(B)

FIGURE 6. Example with 3 points $\mathcal{Y} = \{x_a, x_b, x_c\}$ with $x_a = -1$, $x_c = 1$ and either (A) $x_b = -\epsilon$ or (B) $x_b = \epsilon$. Note that the perturbation can be made arbitrarily small for any $\epsilon > 0$, and the nearest-neighbor orderings are different. (See Tables 2 and 3.)

	point a	point b	point c
point a	0	1	2
point b	1	0	2
point c	2	1	0

TABLE 2. kNN orderings before transformation h.

	point a	point b	point c
point a	0	1	2
point b	2	0	1
point c	2	1	0

TABLE 3. kNN orderings after transformation h.

3.3.3. *Stability of kNN persistence diagrams.* As shown in Sec. 2.3, Theorem 2.16 can be used to describe the stability and convergence of persistence diagrams created using VR filtrations. Here, we show that similar results do not directly result for persistent diagrams created using kNN filtrations.

Theorem 3.11 (kNN Persistence Diagrams are not Uniformly Globally Stable to Point Changes)**.** *Consider two enumerated point clouds $\mathcal{Y}, \mathcal{Z} \in \mathcal{A}_p^N$ and the metric $\|\mathcal{Y} - \mathcal{Z}\|_\infty = \max_i \|\mathbf{y}^{(i)} - \mathbf{z}^{(i)}\|$ over the space of enumerated point clouds. Letting F be the kNN ordering function from Def. 3.2, define matrices $G = F(\mathcal{Y})$*

and $H = F(\mathcal{Z})$ with entries G_{ij} and H_{ij}. *Further denote their associated kNN persistence diagrams $D_{\mathcal{Y}}$ and $D_{\mathcal{Z}}$. Then there exists a uniform global bound of the form*

$$d_B(D_{\mathcal{Y}}, D_{\mathcal{Z}}) \leq \|G - H\|_\infty \equiv \max_{ij} |G_{ij} - H_{ij}|;$$

however, there does not exist a bound of the form:

$$d_B(D_{\mathcal{Y}}, D_{\mathcal{Z}}) \leq \|G - H\|_\infty \not\leq L \|\mathcal{Y} - \mathcal{Z}\|_\infty$$

for some constant L.

Proof. The first inequality is satisfied by applying the original statement of the stability theorem: persistence diagrams are uniformly bound by the maximal difference between the functions that are filtered. That is, the difference in persistence diagrams is bound by the maximum difference in neighbor orderings. However, the stability of persistent homology with respect to neighbor-ordering changes does not imply stability with respect to point-location changes. This occurs because kNN orderings are not stable to point perturbations. To disprove the second inequality, we provide a counter example. Recall the sets of $N = 3$ points in Fig. 6 with locations $\mathcal{Y} = \{-1, -\epsilon, 1\}$ and $\mathcal{Z} = h(\mathcal{Y}) = \{-1, \epsilon, 1\}$, with $\epsilon \in \mathbb{R}$. By construction, $\|G - H\|_\infty = 1$ for any small ϵ, but $\|\mathcal{Y} - \mathcal{Z}\|_\infty = 2\epsilon$. Suppose there did exist a uniform bound with Lipschitz constant L, then the points $\mathcal{Y}$ and $\mathcal{Z}$ with any $\epsilon < 1/(2L)$ yields a contradiction since one would must have both $\|G - H\|_\infty = 1$ and $\|G - H\|_\infty < 1$. $\qquad\square$

3.3.4. *Topological convergence of kNN orderings.* We now use the kNN ordering function to define a discrete-topological notion of convergence for a sequence of point sets using kNN persistent homology.

Definition 3.12. Consider a sequence of point sets $\mathcal{Y}(t) = \{\mathbf{y}^{(i)}(t) | i \in \mathcal{V}\}$ with $t \in \mathbb{N}$ and $\mathcal{V} = \{1, ..., N\}$ and $\mathbf{y}^{(i)}(t) \in \mathbb{R}^p$. We say that the sequence $\{\mathcal{Y}(t)\}$ has **global convergence in kNN topology** to the limit $\mathcal{Y}$ iff $\lim_{t\to\infty} F(\mathcal{Y}(t)) = F(\mathcal{Y})$, where F is the neighbor-ordering function. More precisely, for any ϵ, there exists a t^* such that $\max_{ij} |F_{ij}(\mathcal{Y}) - F_{ij}(\mathcal{Y}(t))| < \epsilon$ for all $t > t^*$.

Definition 3.13. Consider a sequence of point sets $\mathcal{Y}(t) = \{\mathbf{y}^{(i)}(t) | i \in \mathcal{V}\}$ with $t \in \mathbb{N}$ and $\mathcal{V} = \{1, ..., N\}$ and $\mathbf{y}^{(i)}(t) \in \mathbb{R}^p$. Let $\mathcal{N}_{ik} \in \{\mathcal{N}_{ik}^{min}, \mathcal{N}_{ik}^{trans}, \mathcal{N}_{ik}^{max}\}$ be symmetrized neighborhoods that are defined in Def. 3.3. We say that the sequence $\{\mathcal{Y}(t)\}$ has **κ-bounded convergence in kNN,min topology** to a limit $\mathcal{Y}$ iff $\lim_{t\to\infty} F_{ij}(\mathcal{Y}(t)) = F_{ij}(\mathcal{Y})$ for all $i \in \mathcal{V}$ and $j \in \mathcal{N}_{ik}^{min}$, where $k \leq \kappa$. Similarly, we have **κ-bounded convergence in kNN,trans topology**, **κ-bounded convergence in kNN,max topology** for $\mathcal{N}_{ik} \in \{\mathcal{N}_{ik}^{trans}, \mathcal{N}_{ik}^{max}\}$ respectively. More precisely, for any ϵ, there exists a t^* such that $\max_{i\in\mathcal{V}, j\in\mathcal{N}_{ik}} |F_{ij}(\mathcal{Y}) - F_{ij}(\mathcal{Y}(t))| < \epsilon$ for all $t > t^*$ and $j \in \mathcal{N}_{ik}$ where $\mathcal{N}_{ik} \in \{\mathcal{N}_{ik}^{min}, \mathcal{N}_{ik}^{trans}, \mathcal{N}_{ik}^{max}\}$.

Note that κ-bounded convergence is inclusive so that convergence for a given κ also implies convergence for any $\kappa' \leq \kappa$. In particular, global convergence in kNN topology implies κ-bounded convergence for any κ.

We now define local variants of kNN topological convergence.

Definition 3.14. Consider a sequence of point sets $\mathcal{Y}(t) = \{\mathbf{y}^{(i)}(t) | i \in \mathcal{V}\}$ with $t \in \mathbb{N}$ and $\mathcal{V} = \{1, ..., N\}$ and $\mathbf{y}^{(i)}(t) \in \mathbb{R}^p$. We say that the sequence $\{\mathcal{Y}(t)\}$ has **$\mathcal{U}$-local convergence in kNN topology** to a limit $\mathcal{Y}$ iff $\lim_{t\to\infty} F_{ij}(\mathcal{Y}(t)) =$

$F_{ij}(\mathcal{Y})$ for all $i, j \in \mathcal{U} \subseteq \mathcal{V}$. More precisely, for any ϵ, there exists a t^* such that $\max_{i,j \in \mathcal{U}} |F_{ij}(\mathcal{Y}) - F_{ij}(\mathcal{Y}(t))| < \epsilon$ for all $t > t^*$.

Definition 3.15. Consider a sequence of point sets $\mathcal{Y}(t) = \{\mathbf{y}^{(i)}(t) | i \in \mathcal{V}\}$ with $t \in \mathbb{N}$ and $\mathcal{V} = \{1, ..., N\}$ and $\mathbf{y}^{(i)}(t) \in \mathbb{R}^p$. We say that the sequence $\{\mathcal{Y}(t)\}$ has κ-**bounded $\mathcal{U}$-local convergence in kNN topology** to a limit $\mathcal{Y}$ iff $\lim_{t \to \infty} F_{ij}(\mathcal{Y}(t)) = F_{ij}(\mathcal{Y})$ for all $i \in \mathcal{U} \subseteq \mathcal{V}$ and $j \in \mathcal{N}_{ik}$, where the symmetrized neighborhoods $\mathcal{N}_k \in \{\mathcal{N}_{ik}^{min}, \mathcal{N}_{ik}^{trans}, \mathcal{N}_{ik}^{max}\}$ are given in Def. 3.3 and $k \leq \kappa$. More precisely, for any ϵ, there exists a t^* such that $\max_{i \in \mathcal{U}, j \in \mathcal{V}} |F_{ij}(\mathcal{Y}) - F_{ij}(\mathcal{Y}(t))| < \epsilon$ for all $t > t^*$.

Note for any point set that convergences in global kNN topology (or k-bounded convergence in kNN topology), that there exists a t^* such that $F_{ij}(\mathcal{Y}) = F_{ij}(\mathcal{Y}(t))$ for all $t > t^*$ and $i, j \in \mathcal{V}$ (or $i \in \mathcal{V}$ and $j \in N_{ik}$). Specifically, we can choose $\epsilon \in (0, 1)$. Since $F_{ij}(\mathcal{Y}) \in \mathbb{N}$ for any point set $\mathcal{Y}$, the condition $|F_{ij}(\mathcal{Y}) - F_{ij}(\mathcal{Y}(t))| < \epsilon$ implies $F_{ij}(\mathcal{Y}) - F_{ij}(\mathcal{Y}(t)) = 0$ for all $t > t^*$. That is, kNN convergence is a discrete property that is exactly obtained for sufficiently large $t > t^*$. This significantly contrasts the notion of convergence in norm, which is often asymptotically approached rather than exactly obtained. The next theorem more precisely establishes a relation between convergence in a normed metric space and convergence in kNN topology.

Theorem 3.16 (Convergence in Norm Implies Convergence in kNN Topology for non-degenerate point clouds). *Let $\mathcal{Y}(t) = \{\mathbf{y}^{(i)}(t) | i \in \mathcal{V}\}$, with $\mathbf{y}^{(i)}(t) \in \mathbb{R}^p$, be an element of a sequence of point clouds in which each point converges in norm to some limit, $\lim_{t \to \infty} ||\mathbf{y}^{(i)}(t) - \mathbf{y}^{(i)}|| = 0$ for each $i \in \mathcal{V}$. We assume the points $\{\mathbf{y}^{(i)}\}$ to be non-degenerate as described following Definition 3.1 (i.e., we assume $||\mathbf{y}^{(i)} - \mathbf{y}^{(j)}||_2 \neq ||\mathbf{y}^{(i)} - \mathbf{y}^{(k)}||_2$ for any $k \neq j$). Define $\mathcal{Y} = \{\mathbf{y}^{(i)} | i \in \mathcal{V}\} \in \mathcal{A}_p^N$ as that limit, and assume uniform convergence $\mathcal{Y}(t) \to \mathcal{Y}$. Then $\mathcal{Y}(t) \to \mathcal{Y}$ in kNN topology, and there exists a t^* such that $F(\mathcal{Y}(t)) = F(\mathcal{Y})$ for any $t > t^*$, where F is the neighbor-ordering given by Def. 3.2.*

Proof. Let $d_{ij}(t) = ||\mathbf{y}^{(i)}(t) - \mathbf{y}^{(j)}(t)||_2$, $d_{ij}(\infty) = ||\mathbf{y}^{(i)} - \mathbf{y}^{(j)}||_2$, and $d_{ik}(\infty) = ||\mathbf{y}^{(i)} - \mathbf{y}^{(k)}||_2$, and define $\delta = \min_{i,j,k \neq j} |d_{ij}(\infty) - d_{ik}(\infty)|$. Therefore, not only is it the case that the points are non-degenerate, $d_{ij}(\infty) \neq d_{ik}(\infty)$, but that the distances to any two neighbors of a given point $\mathbf{y}^{(i)}$ differ by at least δ. Since $\mathbf{y}^{(i)}(t) \to \mathbf{y}^{(i)}(\infty)$ uniformly, then for any ϵ there exists a t^* such that $\max_i ||\mathbf{y}^{(i)}(t) - \mathbf{y}^{(i)}|| < \epsilon$ for all i and $t > t^*$. It then follows that $|d_{ij}(t) - d_{ij}(\infty)| < 2\epsilon$. Further, choose t^* such that the convergence bound holds for $\epsilon = \delta/4$, implying that $|d_{ij}(t) - d_{ij}(\infty)| < \delta/2$ for all i, j and $t > t^*$. That is, the distances to neighbors differ by at most $\delta/2$ from their limiting values, and at the same time, any two neighbor distances differ by at least δ. It then follows that the neighbor orderings and kNN topology are identical for points $\mathcal{Y}(t)$ and $\mathcal{Y}$ [i.e., $F(\mathcal{Y}(t)) = F(\mathcal{Y})$] for any $t > t^*$. In Appendix A, we provide an extended discussion on why the limiting point cloud $\mathcal{Y}$ must be non-degenerate. $\square$

Finally, we prove that the reverse property does not necessarily hold.

Theorem 3.17 (Convergence in kNN Topology Does Not Imply Convergence in Norm). *Let $\mathcal{Y}(t) = \{\mathbf{y}^{(i)}(t) | i \in \mathcal{V}\}$, with $\mathbf{y}^{(i)}(t) \in \mathbb{R}^p$, be an element of a sequence of point sets that converges in kNN topology. Then each $\mathbf{y}^{(i)}(t)$ may or may not converge to some limit $\mathbf{y}^{(i)}$ as $t \to \infty$.*

Proof. The proof to Thm. 3.16 describes the case in which both converge. Here, we complete the proof by giving a counter-example in which a sequence of point

sets converges in kNN topology but does not converge pointwise. Let $\mathcal{V} = \{1, 2, 3\}$, $\mathcal{Y}(t) = \{\mathbf{y}^{(i)}(t) = it | i \in \mathcal{V}, t \in \mathbf{N}\}$, and note that $\mathbf{y}^{(i)}(t) = it$ diverges with t for all $i \in \mathcal{V}$. However,

$$F(\mathcal{Y}(t)) = \begin{pmatrix} 0 & 1 & 2 \\ 1 & 0 & 2 \\ 2 & 1 & 0 \end{pmatrix} \tag{4}$$

for any $t > 0$, implying that $\mathcal{Y}(t)$ converges in kNN topology to an element $\mathcal{Y} \in \mathcal{A}_1^3$— e.g., one can choose $\mathcal{Y} = \{1, 2, 3\}$ so that $F(\mathcal{Y}(t)) = F(\mathcal{Y})$ for any $t > 0$. Thus, convergence in kNN topology has no bearing on pointwise convergence. $\qquad\square$

4. kNN persistent homology reveals discrete topological convergence of PageRank algorithm.
In this section, we study the convergence of an iterative method for approximating Google's PageRank an use kNN persistent homology to develop a perspective from discrete topology—that is, as opposed to the typical geometric perspective of convergence under a vector norm. In Sec. 4.1, we review the PageRank algorithm. In Sec. 4.2, we propose a method of using kNN homology to study PageRank convergence, applying our method to an empirical network encoding social interactions among dolphins. In Sec. 4.3, we extend these experiments to study kNN homological convergence for three families of random network.

4.1. PageRank: A Review.
Google developed *PageRank* to solve the problem of web search and ranking for the World Wide Web [3]. Their aim was to create an importance measure for each webpage distinguish highly recognizable, relevant pages from those that are less known. There are many derivations of PageRank [67, 36], all of which stem from modeling 'websurfing' (i.e., how people navigate the web) as a Markov chain. In this analogy, the fraction of random websurfers at a particular web page is given by the stationary distribution of a Markov chain.

A main challenge for this formulation is that in practice, a network connecting webpages via their hyperlinks does not usually consist of a single connected component. Instead, there are isolated webpages that cannot be navigated to, or navigated from. To address this issue, the Google founders introduced 'teleportation' so that with probability α, websurfers click a hyperlink to move between webpages, and with probability $1 - \alpha$, websurfers randomly jump to webpage i with probability v_i. Parameter α is called as *teleportation parameter* or *damping factor*. In the original formulation, a walker would jump uniformly at random to another webpage so that the transition probability to each webpage is the same: $v_i = v_j = 1/N$, where N is the number of webpages. It is also beneficial to allow the v_i values to be heterogeneous to bias the random websurfing to remain near a particular set of webpages. Such dynamics is called 'personalized' PageRank, and $\mathbf{v}$ is called the *personalization vector*.

Both PageRank and personalized PageRank can be formulated as a discrete-time Markov chain in which the transition matrix is given by the Google matrix:

$$\mathbf{G} = \alpha \mathbf{P} + (1 - \alpha)\mathbf{e}\mathbf{v}^T, \tag{5}$$

where $\mathbf{e}$ is a vector of ones and each entry P_{ij} gives the probability of transitioning from webpage i to webpage j following a hyperlink. The matrices P, $\mathbf{e}\mathbf{v}^T$ and G are all row-stochastic transition matrices. The stationary distribution of the Markov chain with transition matrix $\mathbf{G}$ is called the PageRank vector $\pi \in \mathbb{R}^N$, which is the limit of the iterative equation

$$\mathbf{x}(t + 1)^T = \mathbf{x}(t)^T \mathbf{G}, \tag{6}$$

for which the fixed point is the solution to the eigenvalue problem $\mathbf{G}^T\boldsymbol{\pi} = \boldsymbol{\pi}$.

The PageRank algorithm has contributed toward Google's rise as a leading technology company, but it's worth noting that it has also been applied in a wide variety of domains beyond web search [27]. In any context, the practical usage of PageRank comes with various challenges. For example, what α value should be used? The PageRank vector can be integrated with a machine learning framework for web search [53], used to help fit functions over directed graphs [72], and can be used to predict missing genes and protein functions [32, 52]. In these various cases, α must be separately chosen as appropriate. As in [57], the author suggests choosing $\alpha = 0.15$ for correlated discovery in a multimedia database. Google historically set $\alpha = 0.85$, which often remains as the default choice in the literature.

In our paper, we explore a different challenge that arises when considering how many times to iterate Eq. (6). The iterated values $\mathbf{x}(t) \to \boldsymbol{\pi}$ converge as $t \to \infty$, but a more practical questions involves studying how many iterations are required for the associated ranks to converge. That is, if one ranks webpages from top to bottom based on their π_i values, then one only needs to iterate Eq. (6) until those ranks converge. Understanding the asymptotic convergence rate and asymptotic error is not immediately relevant when considering this practical question, and we propose kNN persistent homology as a mathematically principled technique to study the convergence of rankings (and relative orderings more generally).

Below, we will study the convergence of persistence diagrams associated with converging approximate PageRank values $\mathbf{x}(t) \to \boldsymbol{\pi}$, comparing the results for kNN filtrations to those of VR filtrations. We will show that the convergence of persistence diagrams for VR filtrations closely relates to the convergence in vector norm (i.e., due to the stability theorem). In contrast, we the convergence of persistence diagrams for kNN filtrations more closely resembles the convergence of rank orderings. In other words, convergence of kNN persistent homology can be used to predict how many iterations are required for $\mathbf{x}(t)$ to be sufficiently close to $\boldsymbol{\pi}$ such that the ranking of nodes—i.e., from 1 to N—has converged.

We formally define the converged rank order according to PageRank by

$$R_i(\boldsymbol{\pi}) = [\phi(-\boldsymbol{\pi})]_i,$$

where $\phi : \mathbb{R}^N \to \mathbb{N}_+^N$ is the 'argsort' function. Recall that the function ϕ was previously defined to sort the pairwise distances in ascending order. By multiplying $\boldsymbol{\pi}$ by negative one, we now sort the π_i values in descending order so that $R_i(\boldsymbol{\pi}) = 1$ for the top-ranked node $i = \text{argmax}_j \pi_j$, which is considered to be the most important node in the graph. Similarly, $R_i(\boldsymbol{\pi}) = N$ for the lowest-ranked node. For each time t, we similarly define the approximate rank orderings

$$R_i(\mathbf{x}(t)) = [\phi(-\mathbf{x}(t))]_i.$$

We note that the rank orderings satisfy $R_i(\mathbf{x}(t)), R_i(\boldsymbol{\pi}) \in \{1, \ldots, N\}$. Because $\mathbf{x}(t) \to \boldsymbol{\pi}$, the approximate rank orderings $R_i(\mathbf{x}(t))$ converge to their final rank orderings $R_i(\boldsymbol{\pi})$. Moreover, for each node i, the rank $R_i(\mathbf{x}(t))$ can converge to $R_i(\boldsymbol{\pi})$ at a different time step t. Therefore, we define t_i^* to be the **iteration of rank convergence**, which is the value of t at which $R_i(\mathbf{x}(t)) = R_i(\boldsymbol{\pi})$ for all $t \geq t_i^*$.

4.2. kNN homological convergence of PageRank for a dolphin social network. We first study an empirical network in which undirected, unweighted edges
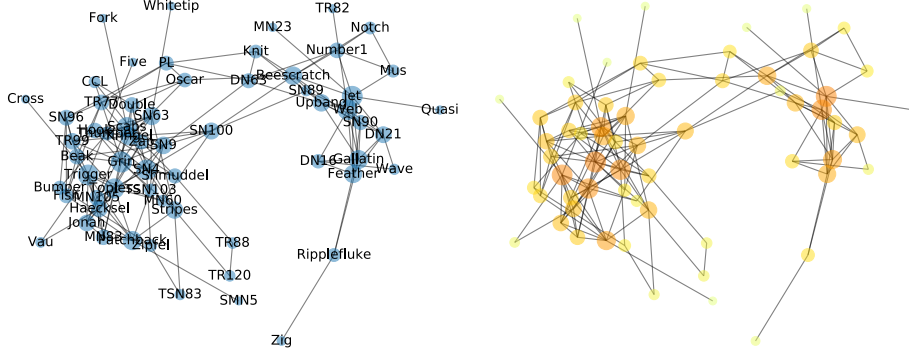
FIGURE 7. (left) A social network of interactions among $N = 62$ dolphins in New Zealand. (right) node colors indicate the nodes' respective PageRank values. In both panels, nodes with larger/smaller size have more/fewer connecting edges.
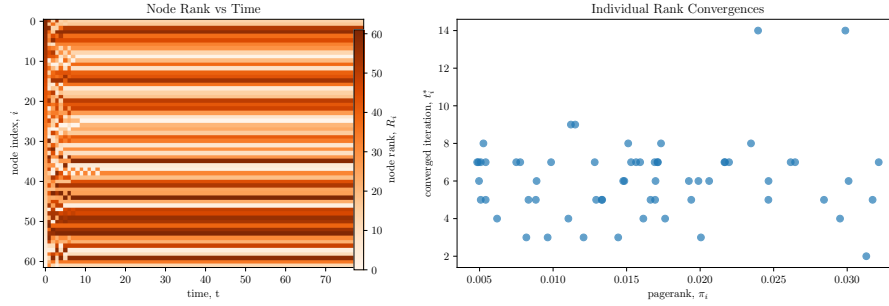

FIGURE 8. (left) Convergence of nodes' rank orderings $R_i(\mathbf{x}(t)) \to R_i(\boldsymbol{\pi})$ versus time step t. (right) Scatter plot comparing t_i^* and π_i across the nodes i.

encode social interactions among $N = 62$ bottlenose dolphins living near Doubtful Sound in New Zealand [48]. We apply the iterative PageRank algorithm with teleportation parameter $\alpha = 0.85$ to the network with an initial condition given by $x_i(0) = i/(\sum_{j=1}^{N} j)$. In the Figure 7, we illustrate the dolphin network and indicate the nodes' converged PageRank values by node color.

4.2.1. *Convergence of ranks for dolphin network.* In Fig. 8, we study the convergence of the rank orderings $R_i(\mathbf{x}(t)) \to R_i(\boldsymbol{\pi})$ for the dolphin network. In Fig. 8(left), observe that the $R_i(\mathbf{x}(t))$ values change up until a time step t_i^*, which is potentially different for each node i. In Fig. 8(right), we depict a scatter plot comparing t_i^* and π_i, noting that we do not see any strong correlation. From a practical perspective, one is often most interested in the top-ranked nodes, and so one is primarily interested in how many iterations are required for the top-ranks to converge. However, there is no guarantee that the top-rank nodes converge before the lower-ranked ones do, or vice versa.
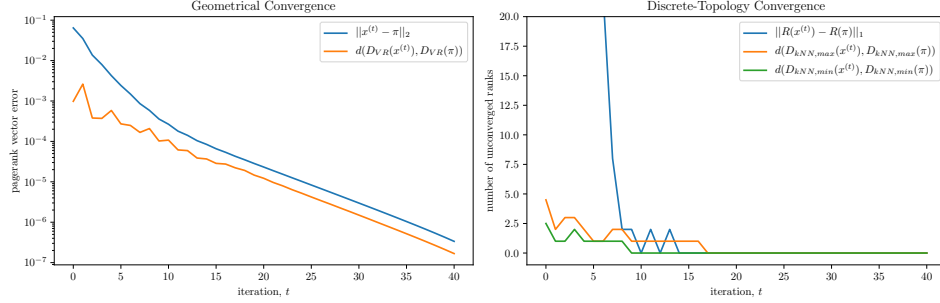
FIGURE 9. Homological convergence of an iterative algorithm for PageRank for the dolphin social network. (left) Convergence of persistent homology for VR filtrations coincides with a geometrical notion of convergence $||\mathbf{x}(t) - \boldsymbol{\pi}|| \to 0$ due to the stability theorem. Both asymptotically approach 0 with exponential decay. (right) In contrast, convergence of persistent homology for kNN filtrations more closely resembles the convergence of the rank ordering, which exactly converges after $t = 14$ time steps in this case. Observe the kNN persistence diagrams for the max and min methods of symmetrization for kNN sets converge at around the same number of iterations.

4.2.2. *Convergence of kNN homology for dolphin network.* We now study the convergence of persistence diagrams for the converging $x_i(t) \to \pi_i$ values, comparing persistence diagrams resulting from kNN filtrations to those of VR filtrations. More precisely, we define $D_{VR}(\mathbf{x})$ and $D_{kNN}(\mathbf{x})$ to be the persistence diagrams according to VR and kNN filtrations, respectively. We will also study kNN filtrations with two types of symmetrization for k-nearest neighbor sets: the minimum and maximum approaches. Then, we let $D_{kNN,max}(\mathbf{x})$ and $D_{kNN,min}(\mathbf{x})$ that are the corresponding persistence diagrams in the symmetrization $\tilde{k}_{ij}, \hat{k}_{ij}$ of kNN ordering. Given the persistence diagrams for $\boldsymbol{\pi}$ and $\mathbf{x}(t)$, we study homological convergence through the bottleneck distance, e.g., $d(D_{VR}(\mathbf{x}(t)), D_{VR}(\boldsymbol{\pi}))$.

In Fig. 9, we illustrate the convergence of persistence diagrams for (left) VR filtrations and (right) kNN filtrations. We compare these two converging topological spaces, respectively, with the normed approximation error, $||\mathbf{x}(t) - \boldsymbol{\pi}||$, and the total difference in rank orderings, $\sum_i |R_i(\boldsymbol{\pi}) - R_i(\mathbf{x}(t))|$. Observe in Fig. 9(left) that VR persistent homology converges similarly to $||\mathbf{x}(t) - \boldsymbol{\pi}||$, whereas kNN persistent homology converges similarly to $\sum_i |R_i(\boldsymbol{\pi}) - R_i(\mathbf{x}(t))|$. That is, kNN topological convergence can be used as a proxy to estimate the number of iterations required for the rank orderings to converge to exactly their final values.

4.2.3. *Impact of α on convergence for kNN homology and PageRank.* Next, we extend the previous experiment for the dolphin network by studying the convergence of the iterative PageRank algorithm and how it changes as one varies the teleportation parameter α. Specifically, in Fig. 10 we plot results that are similar to those shown in Fig. 9 (which used $\alpha = 0.85$), except now different columns depict results for different choices of α. That is, the top row show geometrical convergence
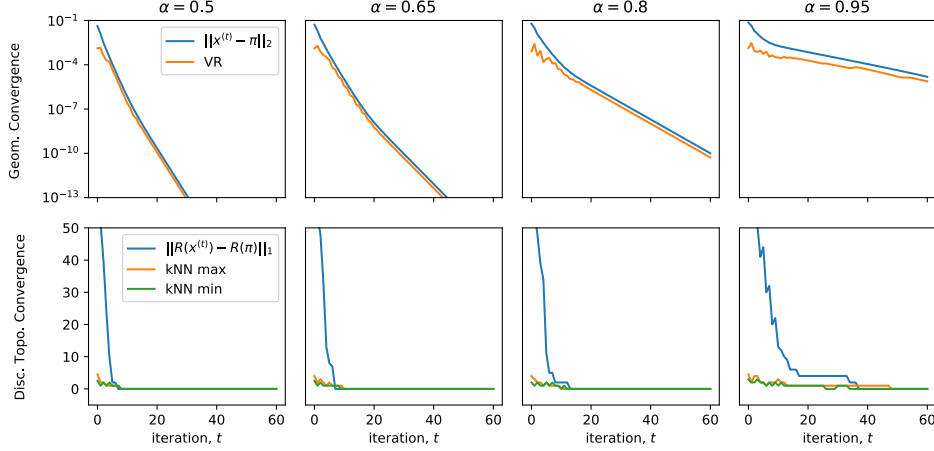
FIGURE 10. Impact of teleportation parameter α on the geometrical and (discrete) topological convergence of PageRank for the dolphin network. The top and bottom rows are similar to the left and right columns of Fig. 9, respectively, except now different columns depict different choices for α.
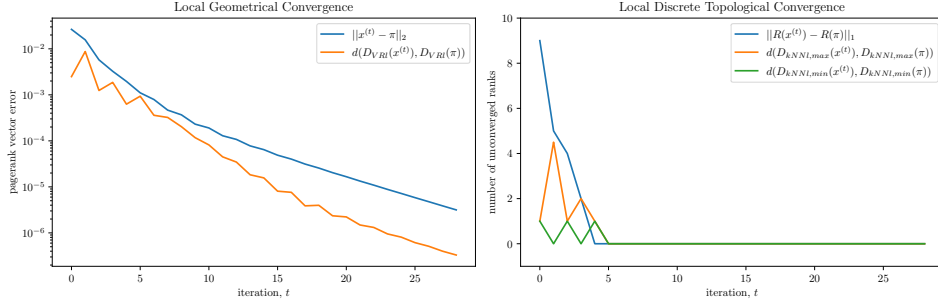

FIGURE 11. $\mathcal{U}$-local topological convergence for the following subset of nodes: $\mathcal{U}_1 = \{$Ripplefluke, Zig, Feather, Gallatin, SN90, DN16, Wave, DN21, Web, Upbang$\}$. Similar to Fig. 9, the left and right panels depict convergence of persistence diagrams for VR and kNN filtrations, respectively.

(i.e., convergence in norm), whereas the bottom row depicts discrete topological convergence using kNN persistent homology.

First, observe in the top row that convergence slows as alpha increases, which is expected since it is known that the iterative method has an asymptotic convergence rate of α^t [56]. More importantly, observe in the bottom row of Fig. 10 that for all values of α the ranks converge at approximately the same iteration that kNN topological convergence occurs. In contrast, as was previously noted, examining VR homology is not immediately informative about the convergence of ranks, since it coincides with geometrical convergence.
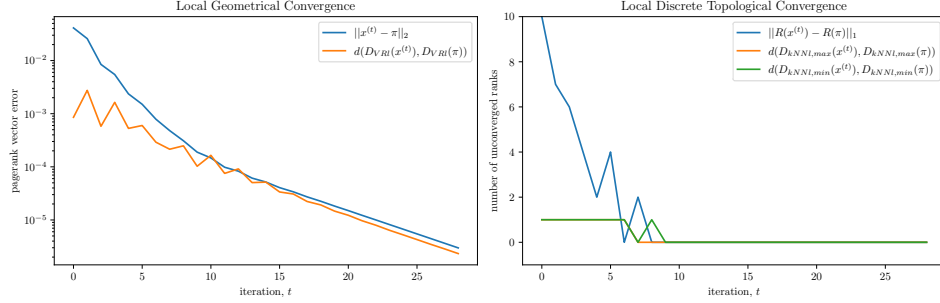
FIGURE 12. Same information as in Fig. 11, except we consider the subset $\mathcal{U}_1$ of nodes that have the largest PageRank values.

4.2.4. *$\mathcal{U}$-local convergence of dolphin social network.* Next, we further study the homological convergence of PageRank for the dolphin network, except we now focus on our notion of $\mathcal{U}$-local topological convergence, which we presented in Sec. 3.3.4. In this context, we focus on convergence for a subset $\mathcal{U} \subset \mathcal{V}$ of the nodes, and we will consider two subsets: a randomly selected set of nodes—$\mathcal{U}_1 = \{$Ripplefluke, Zig, Feather, Gallatin, SN90, DN16, Wave, DN21, Web, Upbang$\}$ and $\mathcal{U}_2$ is the set of 10 nodes that have the top PageRank values. Note that we still compute the iterative approximation to PageRank in the usual way, except that we only consider the values π_i and $x_i(t)$ for which $i \in \mathcal{U}$.

In Fig. 11(left) and (right), we depict convergence of VR and kNN persistence homology, respectively for the subset of nodes $\mathcal{U}_1$. Observe that convergence for VR persistent homology is similar to geometrical (i.e., vector-norm) convergence. In contrast, observe in Fig. 11(right) that the kNN homology and rank orderings converge after approximately $t = 5$ iterations for this subset.

In Fig. 12, we depict the same information, except that we now consider the subset $\mathcal{U}_2$ of nodes with top PageRank values. In this case, the VR convergence more closely aligns with $\|\mathbf{x}(t) - \boldsymbol{\pi}\|$ than when we considered the full set $\mathcal{V}$ of nodes or the subset $\mathcal{U}_1$. Moreover, the kNN homology and the rank orderings converge after approximately $t = 9$ iterations in this case.

4.3. **kNN homological convergence of PageRank for random graphs.** In this section, we further use kNN persistent homology to study the PageRank algorithm's convergence from the perspective of discrete topology. Specifically, we will extend our findings for the empirical dolphin network by considering three families of generative models for random graphs: the Erdős-Rényi model [22]; the Watts-Strogatz model for small-world networks [70]; and the Barabasi-Albert model for scale-free networks [4]. All experiments support our finding that convergence in kNN homology closely predicts the convergence in rank according to PageRank.

4.3.1. *Erdős-Rényi (ER) random graphs.* We first study kNN homological convergence for random graphs generated by the ER $G(N, p)$ model with N nodes and undirected edges that are created with probability p. In Fig. 13 we compare the measure $\|R(\mathbf{x}(t) - R(\pi)\|$ for the difference in ranks to the bottleneck distance $d(D_{kNN}(\mathbf{x}(t)), D_{kNN}(\pi(t)))$ that relates the persistence diagram $D_{kNN}(\mathbf{x}(t))$ according to kNN homology for the approximate PageRank $\mathbf{x}(t)$ to that for its asymptotic limit $D_{kNN}(\pi)$. We consider kNN homology with both the max and min

methods for symmetrizing the nearest neighbors. Different columns and rows in Fig. 13 depict results for different choices of N and p, respectively. Observe in all panels that convergence in rank closely aligns with convergence in kNN homology.
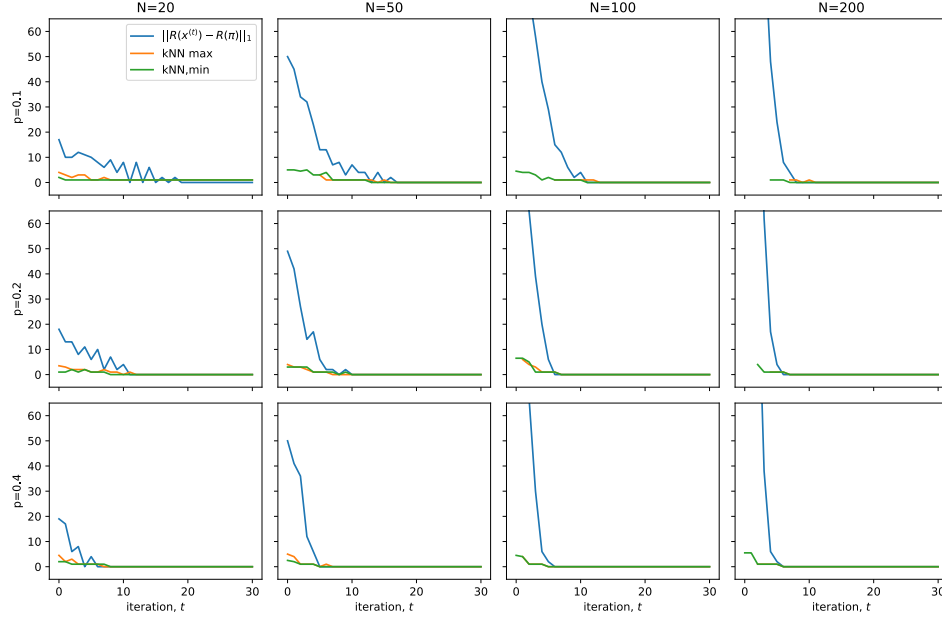

FIGURE 13. kNN homological convergence of PageRank for Erdős-Rényi random graphs [22].

4.3.2. *Small-world random graphs.* Next, we study kNN homological convergence for small-world random graphs generated by the Watts-Strogatz model [70]. Each graph is constructed as follows. First, N nodes are uniformly dispersed along a ring manifold, and each node is connected by k undirected edges to its four nearest-neighbor nodes. (All experiments assume $k = 4$.) Then, each edge is rewired uniformly at random with probability p. In Fig. 14, we compare the convergence of rankings and kNN homology similarly to Fig. 13, and as before, different columns and rows reflect different choices for N and p, noting that p has a different meaning for this model. Again, observe in all panels that the convergence in rank and kNN homology occur at approximately the same iteration t.

4.3.3. *Scale-free random graphs.* Finally, we repeat this experiment for scale-free graphs using the Barabasi-Albert model [4]. In this model, one starts with an initial "core" graph and then nodes are iteratively connected to the core using m undirected edges. This repeats until the graph contains N nodes. In Fig. 15 we compare the convergence of rankings and kNN homology similarly to Figs. 13 and 14, except now different columns and rows reflect different choices for N and m. As before, the convergence in rank and kNN homology occur at approximately the same iteration.
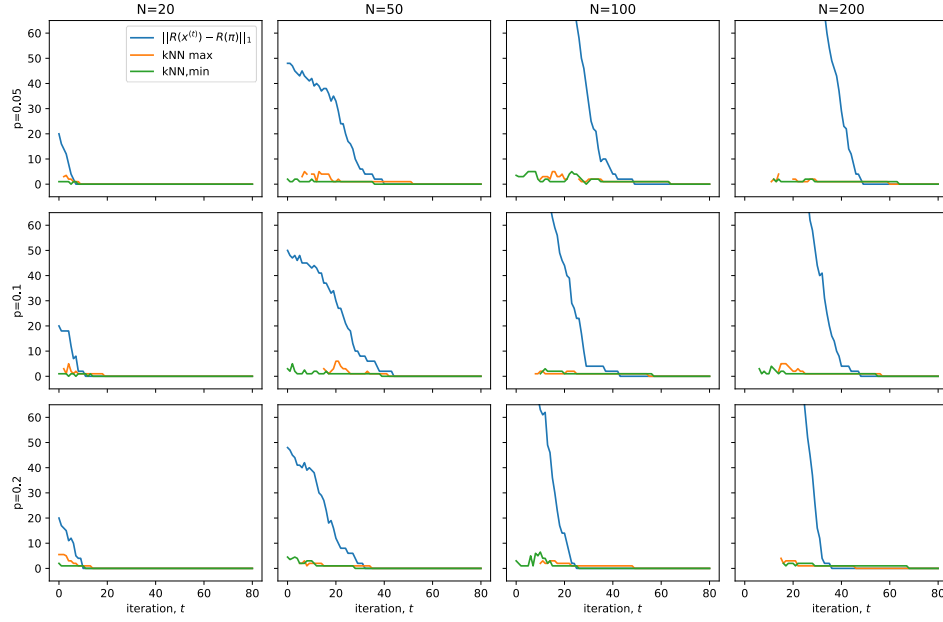
FIGURE 14. kNN homological convergence of PageRank for small-world networks generated using the Watts-Strogatz model [70].
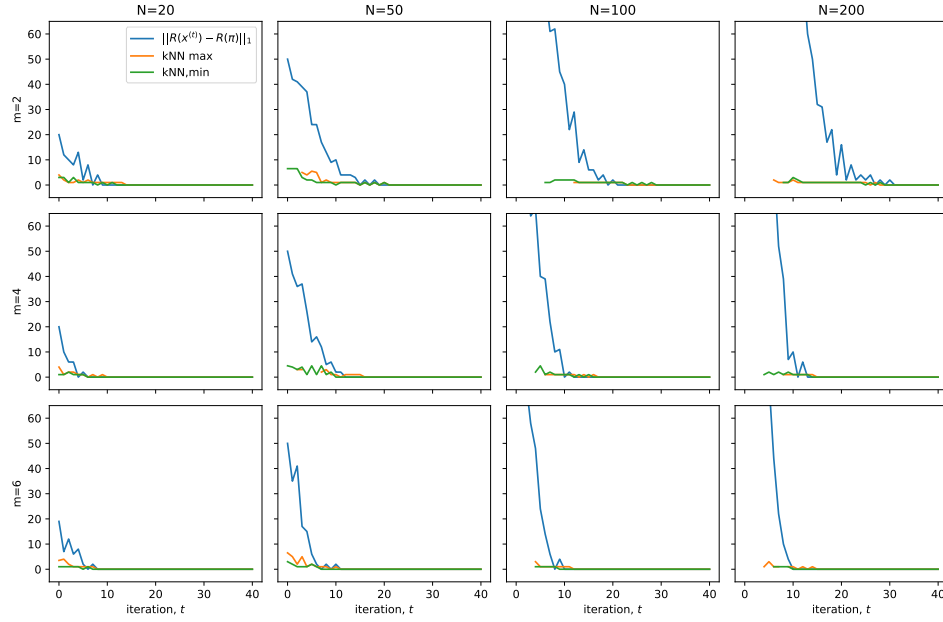

FIGURE 15. kNN homological convergence of PageRank for scale-free networks generated using the Barabasi-Albert model [4].

5. **Discussion.** In this paper, we developed an approach for topological data analysis (TDA) that utilizes k-nearest neighbor sets to define kNN complexes, kNN

filtrations, and kNN persistent homology. Our approach was developed in the spirit of discrete topology and by examining the relative ordering of points, as opposed to the precise distances between points. Although kNN orderings are related to pairwise distances, we provided theory and many experiments highlighting important differences between persistent homology that is based on kNN filtrations versus Vietoris-Rips (VR) filtrations.

To gain theoretical insights into kNN-based TDA, we investigated stability properties for the resulting persistence diagrams in Sec. 3.3.3 and convergence properties in Sec. 3.3.4. While persistence diagrams resulting from kNN filtrations do not satisfy a stability theorem involving a universal bound on perturbed point sets (see Theorem 3.11), we identified and characterized different notions of stability and convergence by identifying equivalence classes as well as bounds on the bottleneck distance between persistence diagrams for kNN homology that are based on the maximum difference for a kNN ordering. Our formulation of convergence also led to several types including global convergence, κ-bounded convergence and $\mathcal{U}$-local convergence. Moreover, we identified some relations among these types as in Theorem 3.16 and Theorem 3.17. Notably, TDA methods have already been insightful for myriad interdisciplinary applications, and we expect kNN-based TDA to similarly find broad applications as this methodology is further developed and applied. For example, our proposed kNN filtrations have already been extended to develop a topological PCA algorithm that was applied to RNA-sequence data [16].

Herein, rather that apply the TDA toolset to data, we proposed to utilize TDA to develop a topological perspective for converging data-science algorithms, thereby complementing the more traditional geometrical perspective (i.e., convergence in norm). Focusing on the PageRank algorithm, we showed that the convergence of persistence diagrams for VR filtrations closely relates to the convergence in vector norm (i.e., due to the stability theorem). In contrast, the convergence of persistence diagrams for kNN filtrations more closely resembles the convergence of rank ordering. In other words, convergence of kNN persistent homology can be used to predict how many iterations are required for $\mathbf{x}(t)$ to be sufficiently close to $\boldsymbol{\pi}$ such that the ranking of nodes—i.e., from 1 to N—has converged. Although we have focused on the PageRank algorithm, iterative algorithms for solving systems (e.g., root finding) are some of the most widely used numerical algorithms. We have shown that the existing TDA approach of VR filtrations coincides with geometrical (i.e., normed) convergence and thereby provides one perspective for the convergence of numerical algorithms. In contrast, kNN filtrations provide complementary insights from the perspective of a discrete topological space that is associated with the relative positioning of points. As such, we expect kNN complexes, filtrations, and persistent homology to have many applications for converging numerical algorithms beyond PageRank.

Appendix A. Extended discussion for Theorem 3.16. The following sequence of point clouds illustrates the necessity of the limiting point cloud being non-degenerate in Theorem 3.16. It is a slight adaptation of an example presented in Fig. 6.

For each $t \in \mathbb{N}$, define $\mathcal{Y}(t) = \{y_1^{(t)}, y_2^{(t)}, y_3^{(t)}\}$ with $y_1^{(t)} = -1$, $y_3^{(t)} = 1$ and

$$y_2^{(t)} := \begin{cases} \frac{1}{t} & \text{if } t \text{ is even} \\ \frac{-1}{t} & \text{if } t \text{ is odd.} \end{cases}$$

Note that each point converges in norm to some limit:

$$y_1^{(t)} \longrightarrow -1 =: y_1 \qquad y_2^{(t)} \longrightarrow 0 =: y_2 \qquad y_3^{(t)} \longrightarrow +1 =: y_3$$

and the limiting point cloud is $\mathcal{Y} := \{y_1, y_2, y_3\}$. However, in this example

$$F(\mathcal{Y}(t)) = \begin{pmatrix} 0 & 1 & 2 \\ 2 & 0 & 1 \\ 2 & 1 & 0 \end{pmatrix}$$

if t is even, but

$$F(\mathcal{Y}(t)) = \begin{pmatrix} 0 & 1 & 2 \\ 1 & 0 & 2 \\ 2 & 1 & 0 \end{pmatrix}$$

if t is odd. Thus the sequence of $F(\mathcal{Y}(t))$ and the associated kNN persistence diagrams do not converge with t. In this example, $|y_1 - y_2| = |y_2 - y_3|$, and so the enumerated point cloud $\mathcal{Y}$ is degenerate. Assuming non-degeneracy for the point cloud ensures the sequence $F(\mathcal{Y}^{(}t))$ to converge with t (and in fact, the limit $F(\mathcal{Y})$ is well defined only under this assumption).

REFERENCES

[1] H. Adams, T. Emerson, M. Kirby, R. Neville, C. Peterson, P. Shipman, S. Chepushtanova, E. Hanson, F. Motta and L. Ziegelmeier, Persistence images: A stable vector representation of persistent homology, *The Journal of Machine Learning Research*, **18** (2017), Paper No. 8, 35 pp.

[2] M. E. Aktas, E. Akbas and A. El Fatmaoui, Persistence homology of networks: Methods and applications, *Applied Network Science*, **4** (2019), Article number: 61.

[3] D. Attali, A. Lieutier and D. Salinas, Vietoris–Rips complexes also provide topologically correct reconstructions of sampled shapes, *Computational Geometry*, **46** (2013), 448-465.

[4] A.-L. Barabási and R. Albert, Emergence of scaling in random networks, *Science*, **286** (1999), 509-512.

[5] S. Brin and L. Page, The anatomy of a large-scale hypertextual web search engine, *Computer Networks and ISDN Systems*, **30** (1998), 107-117.

[6] P. Bubenik, Statistical topological data analysis using persistence landscapes, *The Journal of Machine Learning Research*, **16** (2015), 77-102.

[7] P. Bubenik and J. A. Scott, Categorification of persistent homology, *Discrete & Computational Geometry*, **51** (2014), 600-627.

[8] A. Bukkuri, N. Andor and I. K. Darcy, Applications of topological data analysis in oncology, *Frontiers in Artificial Intelligence*, **4** (2021), 38.

[9] G. Carlsson, Topology and data, *Bulletin of the American Mathematical Society*, **46** (2009), 255-308.

[10] G. Carlsson, T. Ishkhanov, V. De Silva and A. Zomorodian, On the local behavior of spaces of natural images, *International Journal of Computer Vision*, **76** (2008), 1-12.

[11] F. Chazal, V. De Silva, M. Glisse and S. Oudot, The structure and stability of persistence modules, Peprint, 2012, `arXiv:1207.3674`.

[12] S. Chowdhury and F. Mémoli, A functorial Dowker theorem and persistent homology of asymmetric networks, *Journal of Applied and Computational Topology*, **2** (2018), 115-175.

[13] M. K. Chung, P. Bubenik and P. T. Kim, Persistence diagrams of cortical surface data, *International Conference on Information Processing in Medical Imaging*, Springer, **5636** (2009), 386-397.

[14] D. Cohen-Steiner, H. Edelsbrunner and J. Harer, Stability of persistence diagrams, *Discrete & Computational Geometry*, **37** (2007), 103-120.

[15] S. Cottrell, Y. Hozumi and G.-W. Wei, K-nearest-neighbors induced topological PCA for since cell RNA-sequence data analysis, preprint, 2023, `arXiv:2310.14521`.

[16] S. Cottrell, Y. Hozumi and G.-W. Wei, K-nearest-neighbors induced topological PCA for single cell RNA-sequence data analysis, *Computers in biology and medicine*, **175** (2024), 108497.

[17] P. O. de Mendez, Geometric realization of simplicial complexes, *Graph Drawing: 7th International Symposium, GD'99 Štiřín Castle, 1999*, Lecture Notes in Comput. Sci., Springer-Verlag, Berlin, **1731** (1999), 323-332.

[18] V. De Silva and R. Ghrist, Coverage in sensor networks via persistent homology, *Algebraic & Geometric Topology*, **7** (2007), 339-358.

[19] H. Edelsbrunner and J. Harer, Persistent homology-a survey, *Contemporary Mathematics*, **453** (2008), 257-282.

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

[21] H. Edelsbrunner and N. R. Shah, Triangulating topological spaces, *International Journal of Computational Geometry & Applications*, **7** (1997), 365-378.

[22] P. Erdős, A. Rényi, and others, On the evolution of random graphs, *Publ. Math. Inst. Hung. Acad. Sci.*, **5** (1960), 17-60.

[23] J. Gamble and G. Heo, Exploring uses of persistent homology for statistical analysis of landmark-based shape data, *Journal of Multivariate Analysis*, **101** (2010), 2184-2199.

[24] E. Gasparovic, M. Gommel, E. Purvine, R. Sazdanovic, B. Wang, Y. Wang and L. Ziegelmeier, The relationship between the intrinsic Čech and persistence distortion distances for metric graphs, *Journal of Computational Geometry*, **10** (2019), 477-499.

[25] R. Ghrist, Barcodes: The persistent topology of data, *Bulletin of the American Mathematical Society*, **45** (2008), 61-75.

[26] C. Giusti, E. Pastalkova, C. Curto and V. Itskov, Clique topology reveals intrinsic geometric structure in neural correlations, *Proceedings of the National Academy of Sciences*, **112** (2015), 13455-13460.

[27] D. F. Gleich, PageRank beyond the Web, *SIAM Review*, **57** (2015), 321-363.

[28] M. Goresky and R. MacPherson, Stratified morse theory, *Stratified Morse Theory*, Springer, 1988, 3-22.

[29] M. Hajij, G. Zamzmi, T. Papamarkou, N. Miolane, A. Guzmán-Sáenz, K. N. Ramamurthy, T. Birdal, T. K. Dey, S. Mukherjee, S. N. Samaga and others, Topological deep learning: Going beyond graph data, preprint, 2022, `arXiv:2206.00606`.

[30] D. J. Higham, Google PageRank as mean playing time for pinball on the reverse web, *Applied Mathematics Letters*, **18** (2005), 1359-1362.

[31] W. Huang and A. Ribeiro, Persistent homology lower bounds on high-order network distances, *IEEE Transactions on Signal Processing*, **65** (2017), 319-334.

[32] T. Ichinomiya, I. Obayashi and Y. Hiraoka, Protein-Folding Analysis Using Features Obtained by Persistent Homology, *Biophysical Journal*, **118** (2020), 2926-2937.

[33] M. Kahle, Topology of random simplicial complexes: A survey, *Algebraic Topology: Applications and New Directions,* Contemp. Math., American Mathematical Society, Providence, RI, **620** (2014), 201-222.

[34] H. Kannan, E. Saucan, I. Roy and A. Samal, Persistent homology of unweighted complex networks via discrete Morse theory, *Scientific Reports*, **9** (2019), Article number: 1, 18pp.

[35] P. M. Kasson, A. Zomorodian, S. Park, N. Singhal, L. J. Guibas and V. S. Pande, Persistent voids: A new structural metric for membrane fusion, *Bioinformatics*, **23** (2007), 1753-1759.

[36] F. A. Khasawneh and E. Munch, Chatter detection in turning using persistent homology, *Mechanical Systems and Signal Processing*, **70** (2016), 527-541.

[37] J. Kim, J. Shin, F. Chazal, A. Rinaldo and L. Wasserman, Homotopy reconstruction via the cech complex and the vietoris-rips complex, *SoCG 2020 - 36th International Symposium on Computational Geometry*, LIPIcs. Leibniz Int. Proc. Inform., Schloss Dagstuhl. Leibniz-Zentrum für Informatik, Wadern, **164** (2020), Art. No. 54, 19 pp.

[38] L. Kondic, A. Goullet, C. S. O'Hern, M. Kramar, K. Mischaikow and R. P. Behringer, Topology of force networks in compressed granular media, *EPL (Europhysics Letters)*, **97** (2012), 54001.

[39] M. Kramar, A. Goullet, L. Kondic and K. Mischaikow, Persistence of force networks in compressed granular media, *Physical Review E*, **87** (2013), 042207.

[40] G. Kusano, Y. Hiraoka and K. Fukumizu, Persistence weighted Gaussian kernel for topological data analysis, *International Conference on Machine Learning*, PMLR, 2016, 2004-2013.

[41] R. Lambiotte, M. Rosvall and I. Scholtes, From networks to optimal higher-order models of complex systems, *Nature Physics*, **15** (2019), 313-320.

[42] A. N. Langville and C. D. Meyer, Updating Markov chains with an eye on Google's PageRank, *SIAM Journal on Matrix Analysis and Applications*, **27** (2006), 968-987.

[43] M. Q. Le and D. Taylor, Persistent homology of convection cycles in network flows, *Physical Review E*, **105** (2021), Paper No. 044311, 10 pp.

[44] J. Liang, H. Edelsbrunner, P. Fu, P. V. Sudhakar and S. Subramaniam, Analytical shape computation of macromolecules: I. Molecular area and volume through alpha shape, *Proteins: Structure, Function, and Bioinformatics*, **33** (1998), 1-17.

[45] N. Linial and Y. Peled, On the phase transition in random simplicial complexes, *Annals of Mathematics*, **184** (2016), 745-773.

[46] 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 and others, Scalable topological data analysis and visualization for evaluating data-driven models in scientific applications, *IEEE Transactions on Visualization and Computer Graphics*, **26** (2020), 291-300

[47] E. R. Love, B. Filippenko, V. Maroulas and G. Carlsson, Topological deep learning, preprint, 2021, `arXiv:2101.05778`.

[48] D. Lusseau, K. Schneider, O. J. Boisseau, P. Haase, E. Slooten and S. M. Dawson, The bottlenose dolphin community of Doubtful Sound features a large proportion of long-lasting associations, *Behavioral Ecology and Sociobiology*, **54** (2003), 396-405.

[49] M. R. McGuirl, A. Volkening and B. Sandstede, Topological data analysis of zebrafish patterns, *Proceedings of the National Academy of Sciences*, **117** (2020), 5113-5124.

[50] R. Meshulam and N. Wallach, Homological connectivity of random k-dimensional complexes, *Random Structures & Algorithms*, **34** (2009), 408-417.

[51] K. Mischaikow and V. Nanda, Morse theory for filtrations and efficient computation of persistent homology, *Discrete & Computational Geometry*, **50** (2013), 330-353.

[52] J. L. Morrison, R. Breitling, D. J. Higham and D. R. Gilbert, GeneRank: Using search engine technology for the analysis of microarray experiments, *BMC Bioinformatics*, **6** (2005), Article number: 233.

[53] M. A. Najork, H. Zaragoza and M. J. Taylor, HITS on the Web: How does it Compare?, *Proceedings of The 30th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, 2007, 471-478.

[54] J. L. Nielson, S. R. Cooper, J. K. Yue, M. D. Sorani, T. Inoue, E. L. Yuh, P. Mukherjee, T. C. Petrossian, J. Paquette, P. Y. Lum and others, Uncovering precision phenotype-biomarker associations in traumatic brain injury using topological data analysis, *PLOS One*, **12** (2017), e0169490.

[55] N. Otter, M. A. Porter, U. Tillmann, P. Grindrod and H. A. Harrington, A roadmap for the computation of persistent homology, *EPJ Data Science*, **6** (2017), Article number: 17.

[56] L. Page, S. Brin, R. Motwani and T. Winograd, The PageRank citation ranking: Bringing order to the web., Stanford InfoLab, 1999.

[57] J.-Y. Pan, H.-J. Yang, C. Faloutsos and P. Duygulu, Automatic multimedia cross-modal correlation discovery, *Proceedings of The Tenth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 2004, 653-658.

[58] A. D. Parks and D. J. Marchette, Persistent homology in graph power filtrations, *Royal Society Open Science*, **3** (2016), 160228.

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

[60] G. Petri and A. Barrat, Simplicial activity driven model, *Physical Review Letters*, **121** (2018), 228301.

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

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

[63] B. Rieck, U. Fugacci, J. Lukasczyk and H. Leitte, Clique community persistence: A topological visual analysis approach for complex networks, *IEEE Transactions on Visualization and Computer Graphics*, **24** (2017), 822-831.

[64] M. Rucco, L. Falsetti, D. Herman, T. Petrossian, E. Merelli, C. Nitti and A. Salvi, Using topological data analysis for diagnosis pulmonary embolism, *Journal of Theoretical and Applied Computer Science*, **9** (2015), 41-55.

[65] T. Sousbie, The persistent cosmic web and its filamentary structure–I. Theory and implementation, *Monthly Notices of the Royal Astronomical Society*, **414** (2011), 350-383.

[66] B. J. Stolz, H. A. Harrington and M. A. Porter, Persistent homology of time-dependent functional networks constructed from coupled time series, *Chaos: An Interdisciplinary Journal of Nonlinear Science*, **27** (2017), 047410.

[67] E. Tang, J. Agudo-Canalejo and R. Golestanian, Topology protects chiral edge currents in stochastic systems, *Physical Review X*, **11** (2021), 031015.

[68] 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 Communications*, **6** (2015), Article number: 7723.

[69] K. Turner, S. Mukherjee and D. M. Boyer, Persistent homology transform for modeling shapes and surfaces, *Information and Inference: A Journal of the IMA*, **3** (2014), 310-344.

[70] D. J. Watts and S. H. Strogatz, Collective dynamics of 'small-world'networks, *Nature*, **393** (1998), 440-442.

[71] R. van de Weygaert, G. Vegter, H. Edelsbrunner, B. J. T. Jones, P. Pranav, C. Park, and W. A. Hellwing, B. Eldering, N. Kruithof, E. G. P. Bos and others, Alpha, betti and the megaparsec universe: On the topology of the cosmic web, *Transactions On Computational Science XIV*,Springer, **6970** (2011), 60-101.

[72] D. Zhou, J. Huang and B. Schölkopf, Learning from labeled and unlabeled data on a directed graph, *Proceedings of The 22nd International Conference on Machine Learning*, 2005, 1036-1043.

[73] A. Zomorodian, Fast Construction of the Vietoris-Rips complex,aaaaaaaaaaaa *Computers & Graphics*, **34** (2010), 263-271.

[74] A. Zomorodian and G. Carlsson, Computing persistent homology, *Discrete & Computational Geometry*, **33** (2005), 249-274.