Topological estimation of the latent geometry of a complex network

Bukyoung Jhun
1) CCSS, CTP and Department of Physics and Astronomy, Seoul National University, Seoul 08826, Korea
2) Department of Physics, The University of Texas at Austin, Austin, TX 78712, USA

(Dated: September 29, 2021)

Many real-world and synthetic networks are embedded in their latent spaces. If two nodes are embedded closely in the latent space, there is a disproportionately high probability that they are connected by a link. The latent geometry of a complex network is one of the central topics of research in network science with an expansive range of practical applications. Although extensive studies have been devoted to the embedding of networks in latent hyperbolic space, little research has been conducted to develop a method to estimate general unknown latent geometry of complex networks. Here, we develop methods to estimate unknown latent geometry of a given network by removing links with high loads according to certain criteria. The proposed framework can estimate the homology of the latent geometry and provide a simplified map of the latent geometry of the network.

I. INTRODUCTION

Topological data analysis (TDA) has attracted extensive attention in recent years. In the TDA of point cloud data, it is assumed that there is a latent geometry in the configuration space where the point cloud lies, and the point cloud is stochastically generated along the latent geometry. The objective is to estimate the latent geometry and its properties. Many networks have latent geometry, and such networks are called noisy geometric networks. If two nodes are embedded closely in the latent geometry, they have high probability of being connected by a link. For instance, in the Watts-Strogatz small-world networks, each node is embedded at a point on a ring, and if two nodes are close to each other on the ring, there is a disproportionately high probability that they are connected by a link. Another example is the popularity-similarity optimization model, of which the latent geometry is hyperbolic space. The latent geometry of complex networks has various applications, such as efficient routing, link prediction, and brain mapping. It has been reported that certain epidemic processes spread along the latent geometry. Studies on the topological analysis of networks have focused on small-scale loops and voids, which usually appear in a number proportional to the size of the system. Extensive studies have been devoted to embedding a given network in a latent hyperbolic space, but little research has been conducted to develop a method that estimates an unknown latent geometry.

In the TDA of point cloud given with a metric function, one constructs a simplicial complex by connecting close points with certain criteria. Various methods of algebraic topology are then be implemented in the constructed simplicial complex. Persistent homology, which is one of the most commonly used tools in TDA, encodes multiscale topological features of a point dataset. For a given filtration, which is usually the length-scales of interest, a sequence of simplicial complexes is constructed, and the parameters that $d$-dimensional holes are constructed and destructed are encoded in the barcode of the persistent homology diagram. Zero-dimensional homology measures the number of connected components, one-dimensional homology measures the number of loops, two-dimensional homology measures the number of voids. Higher dimensional homology counts higher dimensional holes, but they are not commonly used in applications. Persistent homology has been used as a topological tool to reveal how the cognitive circuits are programmed, analyze the relation between the topology and function of proteins, identify brain disease, analyze the effect of psychedelics on the functional pattern of brain activity, classify nanoporous materials, and analyze structures of social collaborations. Another powerful method in TDA is the mapper. The method summarizes the given high-dimensional point cloud into a mapper graph. Data points are assigned to nodes in the resulting mapper graph, which represents the simplified latent geometry of the data. The mapper has been used in a wide range of disciplines to discover a new subgroup of breast cancer, identify Attention-Deficit/Hyperactivity Disorder, or neurologial damage, reveal dynamical organization of the brain, distinguish resilience to infection in human population, prediction of manufacturing productivity, and mine data in social media. Mapper with a single real-valued filter estimates the Reeb graph, and a higher-dimensional structure is estimated when a higher-dimensional filter is implemented.

One can implement these methods of TDA, persistent homology and mapper, on networks. The distance between two nodes on a network can be used as the metric function for these methods; however, when persistent homology is directly applied, it fails to estimate the homology of the latent geometry. For instance, presence of only 0.1% long-range links in a Watts-Strogatz small-world network ($P = 0.001$) is enough to completely erase the homology signature in the persistent homology [Fig. 1(a, b)]. Also, the mapper can be implemented using the distance from a randomly selected node as the filter. However, this fails to reveal the latent geometry of the network, which is a ring, as shown in Fig. 1(c). If there are no long-range links in the network, both methods can successfully estimate the latent geometry of the network and its properties.

Therefore, if we identify and remove the long-range links...
in a noisy geometric network, we can successfully reveal the latent geometry of the network. One property that can be used to distinguish the long-range links from the short-range links is the load\textsuperscript{14}. Long-range links in a noisy geometric network have significantly higher loads than short-range links. Therefore, we can recover the latent geometry by removing links with high loads. We found that when an appropriate percentage of the links with the highest loads are removed, topological analysis can successfully estimate the latent geometry of the given network. Subnetworks consisting of high-load links have been extensively studied\textsuperscript{15,20,27}, but subnetworks consisting of links with low loads have attracted little to no attention.

Here, we propose a set of systematic methods that can estimate the arbitrary latent geometry of a given network: the modified persistent homology diagram and the map of the latent geometry. In the modified persistent homology diagram, we focus on one-dimensional homology, but this method can be implemented to study higher-dimensional homology of the latent geometry. In the modified persistent homology diagram, we can recover the latent geometry by removing links with high loads. We found that when an appropriate percentage of the links with the highest loads are removed, topological analysis can successfully estimate the latent geometry of the given network. Subnetworks consisting of high-load links have been extensively studied\textsuperscript{15,20,27}, but subnetworks consisting of links with low loads have attracted little to no attention.

One property that can be used to distinguish the long-range links from the short-range links is the load\textsuperscript{14}. Long-range links in a noisy geometric network have significantly higher loads than short-range links. Therefore, we can recover the latent geometry by removing links with high loads. We found that when an appropriate percentage of the links with the highest loads are removed, topological analysis can successfully estimate the latent geometry of the given network. Subnetworks consisting of high-load links have been extensively studied\textsuperscript{15,20,27}, but subnetworks consisting of links with low loads have attracted little to no attention.

Here, we propose a set of systematic methods that can estimate the arbitrary latent geometry of a given network: the modified persistent homology diagram and the map of the latent geometry. In the modified persistent homology diagram, we focus on one-dimensional homology, but this method can be implemented to study higher-dimensional homology of the latent geometry. In the modified persistent homology diagram, we can recover the latent geometry by removing links with high loads. We found that when an appropriate percentage of the links with the highest loads are removed, topological analysis can successfully estimate the latent geometry of the given network. Subnetworks consisting of high-load links have been extensively studied\textsuperscript{15,20,27}, but subnetworks consisting of links with low loads have attracted little to no attention.

II. NOISY GEOMETRIC NETWORKS

Many network models have latent geometries. A network with a latent geometry is called a noisy geometric network\textsuperscript{12}. In a noisy geometric network, each node is assigned to a position in a metric space, which is called the latent geometry of the network. If two nodes are close in the latent geometry, they have a high probability of being connected by a link. Short-range links in the latent geometry are called geometric edges, and long-range links are called non-geometric edges\textsuperscript{12}. Examples of noisy geometric network models include the Watts-Strogatz small-world network\textsuperscript{26}, and its variants\textsuperscript{21,28}, and some scale-free network models\textsuperscript{19,37}.

Noisy geometric networks have been the subject of extensive studies in the network science community. Latent geometry has been used to efficiently navigate in networks\textsuperscript{4,21}, predict missing links\textsuperscript{10,27}, and map brains\textsuperscript{5}. Certain epidemic phenomena have been reported to spread along the latent geometry\textsuperscript{21}. Embedding a given network to a hyperbolic geometric space is an extensively studied subject\textsuperscript{12,23,28} with a wide variety of applications. Other topological studies of networks have focused on the number of small-scale homology, which is usually proportional to system size. However, estimation
of the arbitrary latent geometry of a complex network has attracted little attention. In this study, we used three noisy geometric network models with various latent geometries along with real-world networks to test whether our methods estimate the correct properties of a given latent geometry.

The Watts-Strogatz small-world network was proposed as a model that manifests a high clustering coefficient and short average path lengths and has served as a classic network model in the research community along with its variants. The original Watts-Strogatz model had a ring as its latent geometry. We generated networks with three types of latent geometries: ring, double-ring, and triple-line [Fig. 2a–c]. The latent geometry serves as a metric space: the distance between two nodes is measured along the structure. The connection probability of two nodes in a Watts-Strogatz network is solely determined by whether the distance between the two nodes is smaller or larger than $K/2$. If the parameters of the network are $N, K,$ and $P,$ two points that have a distance of less than $K/2$ between them are directly connected by a link with probability of $1 - P + KP^2/(N - 1) \approx 1 - P$. If the distance is larger than $K/2$, the probability is $KP/(N - 1)$, which approaches zero as $N$ approaches infinity while $K$ and $P$ are fixed. A small fraction $P$ of long-range links gives the small-world property to the network.

Short-range links are highly regular in the Watts-Strogatz model. To show that our methods do not rely on this regularity, we constructed a model with stochastically generated short-range links are generated stochastically: the exponentially connected network. An exponentially connected network is constructed on a latent geometry, and each node is assigned to a point in the structure. The probability that two nodes $i$ and $j$ are connected by a link is given by an exponential function of the distance between them:

$$P_{ij} = e^{-d_{ij}/\xi},$$

where $d_{ij}$ is the distance between the two nodes in the latent geometry, and $\xi$ is the correlation length that characterizes the network. An exponentially connected network with a one-dimensional latent geometry has a mean degree $\langle k \rangle = 2 \sum_{i=1}^{\infty} e^{-i/\xi} = 2/(e^{1/\alpha} - 1)$. We then construct a multiplex network with one layer consisting of a exponentially connected network with correlation length a $\xi$ and another consisting of an Erdős–Rényi network with mean degree $\langle k_{LR} \rangle$. The links in the exponentially connected network serve as short-range links and the links in the Erdős–Rényi network serve as long-range links. The fraction of long-range links in the multiplex network is $\langle k_{LR} \rangle \langle (k_{LR}) + (k_{SR}) \rangle = \langle k_{LR} \rangle / (2/(e^{1/\alpha} - 1) + \langle k_{LR} \rangle)$.

Long-range links and short-range links are generated by different processes in the two aforementioned noisy geometric network models. In the power-law connected network, long-range links and short-range links are created by the same process. Similar to the exponentially connected network, each node of the network is assigned a position on the latent geometry. Two points $i$ and $j$ in a power-law connected network are connected with probability

$$P_{ij} = \begin{cases} \alpha d_{ij}^{-\sigma} & \text{if } d_{ij} \geq \alpha^{1/\sigma} \\ 1 & \text{if } d_{ij} < \alpha^{1/\sigma} \end{cases},$$

where $\alpha$ is used.
where $d_{ij}$ is the distance between nodes $i$ and $j$ in the latent geometry, and $\sigma$ is the power that controls the decay of probability over distance. The mean degree of a power-law connected network with a one-dimensional latent geometry is

$$\langle k \rangle = 2\alpha \left( \xi(\sigma) - \sum_{n=1}^{\lfloor \alpha/|R| \rfloor + 1} n^{-\sigma} \right) + 2\alpha^{1/\sigma} + 2.$$

III. MODIFIED PERSISTENT HOMOLOGY DIAGRAM

The methods of TDA investigate the latent shape of data using the distance between each pair of points. One then constructs a simplicial complex by connecting set of points that are close to each other with simplices using certain criteria. A simplicial complex is a class of hypergraph that satisfies a certain property: if a hyperedge $E$ is in a simplicial complex $H$, any hyperedge $E' \subseteq E$ is also in $H$. Once a simplicial complex is constructed, we can implement the methods of algebraic topology to calculate the properties of the simplicial complex. We then use the obtained results to estimate the properties of the latent geometry of the given point dataset. The most common simplicial complex used in TDA is the Vietoris-Rips complex (often called Rips complex). For a radius $t$, we include one-dimensional simplices (link) $S_1 = \{i_0, i_1\}$ in the Rips complex if and only if the distance between the nodes $i_0$ and $i_1$ is no greater than $t$: $d(i_0, i_1) \leq t$. A higher $d$-dimensional simplex $S_d = \{i_0, i_1, \cdots, i_d\}$ is included in the Rips complex if and only if the distance between all pairs in the simplex is less than $t$: $\forall x, y \in S_d$, $d(x, y) \leq t$. The Rips complex generated by this rule satisfies the properties of the simplicial complex: if a simplex $S_d$ is in a simplicial complex $S$, any simplex $S' \subseteq S_d$ is also in $S$.

Persistent homology, one of the most commonly used tools in TDA, encodes multiscale topological features of data. For a standard $\epsilon$-ball persistent homology, one constructs a sequence of Rips complex for a range of radius $t$. As the radius $t$ increases, loops, voids, and higher-dimensional holes appear and disappear in the corresponding Rips complex. We track the radius that holes appear and disappear, and this provides a barcode structure in the persistent homology diagram. Zero-dimensional homology measures the number of connected components, one-dimensional homology measures the number of loops, two-dimensional homology measures the number of voids in the simplicial complex. Higher-dimensional homology measures the number of higher dimensional holes in the simplicial complex, but it is not commonly used in applications. If $n$ number of bars persist in the persistent $d$-dimensional homology diagram for a wide range of $t$, it is assumed that there are $n$ $d$-dimensional holes in the latent geometry of the data.

One only needs the distances between the pairs of points in the data to implement persistent homology. Therefore, we can apply the method to a complex network using the shortest-distance between the nodes in the network as the metric function. However, when applied to a Watts-Strogatz small-world network with a ring latent geometry, it fails to estimate the latent geometry of the network [Fig. III.a, b]. Presence of only 0.1% long-range links ($P = 0.001$) is enough to completely erase the topological signature of the latent geometry. The result counts the number of connected components, loops, and higher-dimensional holes in the network instead of estimating the latent geometry. Such loops and voids are formed stochastically, and their number is often proportional to the number of nodes in the network. In a network, the birth radius of zero- and one-dimensional homology is always one. For one-dimensional homology, it suffices to only keep track of the homology at each radius, instead of keeping track of all the birth-death radii. Direct implementation of persistent homology on a network can reveal the number and lengths of loops in the network. While the direct implementation of persistent homology can be useful to study many novel properties of a network, it cannot reveal the properties of the latent geometry.

If there are no long-range links in the network, persistent homology can successfully estimate the homology of the latent geometry. Therefore, if we identify and remove the long-range links in a network, then we can implement persistent homology to estimate the number of loops and voids in the latent geometry. One property that distinguishes long-range links from short-range links is the load, as long-range links have significantly higher loads than short-range links. We distill the network by removing a certain fraction $\delta$ of the links with the highest loads. We vary the distillation rate $\delta$ along with the radius $t$ for a range of values. Here, we focus on the one-dimensional homology and study the number of loops in the latent geometry, but the method can be generalized to study higher-dimensional homology. If there are $n$ loops in the latent geometry, for a significant range of values of distillation rate $\delta$, the one-dimensional homology is $n$ for a significant range of radius $t$. Therefore, the one-dimensional homology should persistently appear as $n$ in a wide region on the $t$-$\delta$ plane (Fig. III). This allows us to estimate the homology of the latent geometry of the network.

IV. MAP OF THE LATENT GEOMETRY

Another common topological signature studied in TDA is the mapper. The purpose of the mapper is to simplify the latent geometry of a given point cloud as a network. Unlike the persistent homology, this requires the user to make a few choices of arbitrary parameters. The first thing that must be chosen is the filter function. The filter function is used to partition data points into groups. A one-dimensional filter function $f$ assigns a real number to each data point $x \in X$: $f : X \rightarrow \mathbb{R}$. A one-dimensional filter is often used to reveal a one-dimensional filamentary structure, but the target space for the filter function can be chosen as a higher-dimensional space: $f : X \rightarrow \mathbb{R}^d$. This choice reveals a higher-dimensional latent geometry. We then divide the range of the function $f : X \rightarrow \mathbb{R}^d$ into a set of smaller overlapping intervals $I_j$’s. In this step, we must choose two parameters: the length of the smaller interval and the percentage overlap between the successive intervals. Then, for each interval $I_j \in S$ and $X_j = \{x | f(x) \in I_j\}$, we implement a clustering algorithm to find clusters $X_{jk}$ in each $X_j$. In this step, we must choose a
Figure 3. Modified persistent homology diagrams for Watts-S impression small-world network with (a) a ring latent geometry, (b) a double-ring latent geometry, (c) a triple-line latent geometry; (d) multiplex network of exponentially connected network and Erdős–Rényi network with a double-ring latent geometry, and (e) power-law connected network with a double-ring latent geometry. The diagram in (a) shows a persistent region of homology $H_1 = 1$ and correctly estimates the homology of a ring. The diagram in (b, d, e) shows a persistent region of homology $H_1 = 2$, and the diagram in (c) shows no persistent region of non-zero homology $H_1$. The number of nodes is 1200 and the mean degree is 12 in each network. The fraction of the long-range links $P = 0.1$ in small-world networks and the multiplex network. The power of connectivity $\sigma = 2$ in the power-law connected network.

Figure 4. The map of the latent geometry of Watts-Strogatz small-world network with (a) a ring latent geometry, (b) a double-ring latent geometry, (c) a triple-line latent geometry; (d) multiplex network of exponentially connected network and Erdős–Rényi network with a double-ring latent geometry, and (e) power-law connected network with a double-ring latent geometry. The size of a vertex in the map is proportional to the number of network nodes in the cluster, and the color encodes value of the filter function, with red indicating high and blue indicates low value. The number of nodes is 1200 and the mean degree is 12 in each network. The fraction of the long-range links is 0.1 in the small-world networks and the multiplex network. The power of connectivity $\sigma = 2$ in the power-law connected network. The map correctly represents the latent geometry in each case. (f) Scatter plot of the distances between pairs of nodes in latent geometry and the distilled network. The Pearson correlation coefficient is 0.9995, hence the distilled network successfully recovers the latent geometry. The Pearson correlation coefficient of distances on distilled networks $r(\delta_1, \delta_2)$ of (g) Erdős–Rényi network and (h) Watts-Strogatz small-world network used in (a). A clear persistent interval appears in the small-world network but not in the Erdős–Rényi network. (i) The persistence measured by $\chi_{0.025}$. A clear peak at $\delta = 0.31$ appears.

parameter for the clustering algorithm of choice (for instance, the $k$ value in the $k$-nearest neighbor clustering or the number of clusters in agglomerative clustering). The mapper does not impose any conditions on the clustering algorithm. We treat each cluster $X_{jk}$ as a vertex in the mapper graph. If there exists an overlapping point in two clusters (with distinct $j$’s), we connect the two clusters in the mapper graph.

One can implement the mapper to a geometric network, using the network distance as the metric function. Distance from a selected node is defined for each node that is in the same connected component as the selected node and can therefore be used as the filter function. At each layer of the filter function $X_j$, we define each connected component in the subnetwork consisting of $X_j$ as $X_{jk}$. We applied this algorithm to a Watts-Strogatz small-world network with $N = 1200$, $P = 0.001$, and $K = 12$ (Fig. 2c). The mapper does not show a graph representing the latent geometry of the network, which is a ring. Long-range links constituting 0.1% of the total links
completely erase the topological signature of the latent geometry. Neither can network coarse-graining methods show the latent structure in the presence of long-range links.

In the same spirit as the modified persistent homology diagram, we can remove a fraction of the links in the network with high loads and then apply the mapper algorithm or coarse-graining methods. We distill the given network by removing a fraction $\delta$ of the links with the highest loads. For an appropriate distillation rate $\delta$, which can be determined using a method that will be elaborated later, the mapper returns a graph that successfully represents the latent geometry of the given networks (Fig. 4). We use the shortest-distance between nodes in the distilled network as the metric function. For each connected component in the distilled network, we select a node and use the distances from other nodes to the selected node as the filter function within the connected component. We select one of the two nodes that are farthest apart in the connected component of the distilled network. We then partition the interval of the filter function into $I_j \in S$ and partition $X$ into $X_j = \{x \mid f(x) \in I_j\}$. The length of the interval is the length-scale of the resulting map of the latent geometry (note that the filter function is a distance in this method). In each subnetwork consists of $X_j$, we define each connected component $X_{jk}$ in the subnetwork as a node in the resulting map of the latent geometry. We connect two nodes in the map if the intersection between the connected components $X_{jk}$ and $X_{j'k'}$ that each corresponds to the nodes is non-empty.

We choose the optimal distillation rate that persistently returns a consistent network structure for a wide-ranging distillation rate $\delta$. We use the distances between pairs of nodes to quantify the similarity between two network structures. Suppose that $G_1$ and $G_2$ are distilled networks of $G$ with distillation rates $\delta_1$ and $\delta_2 > \delta_1$. The two networks have exactly the same set of nodes, and if a link $\{i, j\}$ is in $G_2$, it is also in $G_1$. The distance between two nodes $i$ and $j$ in $G_1$, $d_1(i, j)$ is no greater than the distance in $G_2$, $d_2(i, j)$. We quantify the similarity between the two networks $G_1$ and $G_2$ using the Pearson correlation coefficient of distances that are finite in $G_2$ (two nodes are in the same connected component in $G_2$ and thus in $G_1$) and larger than one in $G_2$: $r(\delta_1, \delta_2)$. We exclude the pairs that are directly linked in $G_2$ because it is guaranteed that they are linked in $G_1$ and thus do not provide any meaningful information regarding the structure. We illustrated $r(\delta_1, \delta_2)$ for a Watts-Strogatz small-world network with a ring latent geometry (Fig. 4h). The plot shows a persistent interval of distillation rate $\delta$, and in this interval, the map yields a correct representation of the latent structure. We selected the parameter that yields the highest consistency, quantified by

$$\chi_{0.025} = \min\{a| r(\delta, \delta + a) > 0.975, r(\delta - a, \delta) > 0.975\} \quad (3)$$

This method could potentially be used for parameter selection in the TDA of point cloud, such as k-nearest neighbor graph construction.

V. TOPOLOGICAL ANALYSIS OF SYNTHETIC AND EMPIRICAL NETWORKS

To test our methods, we applied them to various types of synthetic noisy geometric networks with various latent geometries. We constructed Watts-Strogatz small-world network with latent geometry of a ring, double-ring, and triple-line (Fig. 2). We also constructed a multiplex network of exponentially connected network and Erdős–Rényi network and a power-law connected network with a double-ring latent geometry. The modified persistent homology diagrams illustrated in Fig. 3 clearly show one-dimensional homology of zero for the triple-line network, one for the ring network, and two for the double-ring networks. The modified persistent homology for complex networks successfully estimates the homology of the latent geometries of synthetic noisy geometric networks.

We also mapped the latent geometries of the same set of networks. The results are illustrated in Fig. 4 (a–e). As illustrated in Fig. 4f), the Pearson correlation coefficient between the distance in the latent geometry $d(x, y)$ and the distance in the distilled network $d_\delta(x, y)$ is 0.99955 at the optimal distillation rate. The correlation of distances between pairs of nodes in two distilled networks is illustrated in Fig. 4g), which clearly shows a persistent interval of the distillation rate $\delta$, of which the distilled network yields a consistent network structure. The consistency defined in Eq. 3 shows a clear peak at $\delta = 0.31$, and we choose this value as the optimal distillation rate. In contrast, no persistent interval appears in the $r(\delta_1, \delta_2)$ diagram for the Erdős–Rényi network, which has no latent geometry. The same process has been used to find the optimal distillation rates for other networks.

To further test our methods, we applied them to a paradigmatic network dataset: the coauthorship structure of network science researchers. In this network, there are 1589 scientists from variety of fields who published a paper on the topic of networks, 379 of which fall into the largest connected component. If there exists a paper (by May 2006) between two researchers, two are connected in the network. The network is undirected and unweighted. Figure 5(a) illustrates the modified persistent homology diagram of the network. There is no persistent region of non-zero homology in this diagram, and we conclude that there is no loop in the latent geometry of the network. The Pearson correlation coefficient of the distances in the distilled networks $r(\delta_1, \delta_2)$ is illustrated in Fig. 5b). A clear persistent interval appears around $\delta = 0.1$ appears in the diagram. This suggests that the latent geometry of the network is not completely random. The map of the latent geometry is illustrated for the optimal distillation rate $\delta = 0.1$ in Fig. 5d). We showed the five largest connected components in the map and displayed the researcher with the highest degree within the cluster as the representative.

VI. CONCLUSIONS

Many synthetic and empirical networks have latent geometry, and if two nodes are close in the latent geometry, they
are connected by a link with high probability. Extensive studies have focused on the embedding of complex networks in latent hyperbolic geometry\textsuperscript{1,4,22,28}. In such studies, it is assumed that the latent geometry of the given network is a hyperbolic space. This assumption is not true for many synthetic network models, and possibly for many real-world networks. However, little research has been devoted to the development of a method to estimate arbitrary unknown latent geometry\textsuperscript{42}.

Here, we developed methods that successfully estimates the latent geometry of complex networks: the modified persistent homology diagram and the map of the latent geometry. By exploiting the fact that long-range links have disproportionately higher loads than short-range links, we can remove links with high loads to obtain a distilled network. We use two-dimensional filtration for the persistent homology. A simplicial complex is determined by the distillation rate $\delta$ and the radius $t$. We estimate the homology of the latent geometry by the persistent region in the $\delta$-$t$ plane that yields a constant homology in the modified persistent homology diagram. This method is completely non-parametric. We also implement the mapper to the distilled network to obtain the map of the latent geometry. One must choose only two parameters to map the latent geometry of the network: the length-scale of interest and the overlapping percentage of the intervals. The number of parameters is smaller compared to the standard mapper algorithm of point cloud data, for which the user must choose the length of the interval, the overlapping percentage, the clustering algorithm and its parameters, and the filter function. We reduced the number of arbitrary parameters by using a method to choose the optimal distillation rate: we choose the distillation rate that yields the most consistent network structure for a wide interval of the parameter. We used the Pearson correlation coefficient of the distances between pairs of nodes to quantify the consistency of the network structures.

Mapping networks to hyperbolic spaces has a wide range of applications. Such applications include efficient navigation in a complex network such as the internet\textsuperscript{4}, missing link prediction\textsuperscript{44} and brain mappings\textsuperscript{5}. If the latent geometry of a network is not a hyperbolic space, hyperbolic mapping can lead to a faulty conclusion. In such cases, the latent geom-

Figure 5. (a) The modified persistent homology diagram of the coauthorship structure of network science researchers. No persistent region appears in the diagram, and this indicates that there is no loop in the latent geometry. (b) The Pearson correlation coefficient of distances on distilled networks $r(\delta_1, \delta_2)$. A clear persistent interval appears, and this indicates that the latent geometry of the network is not completely random. (c) The persistent measured by $\chi_{0.025}$. A clear peak appears at $\delta = 0.1$. (d) The map of the latent geometry of the coauthorship in network science. The size of a vertex in the map is proportional to the number of network nodes in the cluster, and the color encodes value of the filter function, with red indicates high and blue indicates low value. Five largest connected components in the map are shown. Name of the researcher who has the highest degree within the cluster is shown. A researcher can simultaneously be in two adjacent clusters, therefore, a single researcher can appear multiple times in the map.
etry of the network must be revealed first, and then navigation, link prediction, and mapping should be performed. Our methods can estimate arbitrary unknown latent geometry and make contributions similar to the hyperbolic mapping in such networks.

VII. ACKNOWLEDGMENTS

This research was supported by the NRF, Grant No. NRF-2014R1A3A2069005.

REFERENCES

1. G. Alanis-Lobato, P. Mier, and M. A. Andrade-Navarro. Efficient embedding of complex networks to hyperbolic space via their laplacian. Sci Rep, 6:30108, 2016.

2. Khaled Almgren, Minkyu Kim, and Jeongkyu Lee. Mining social media data using topological data analysis. In 2017 IEEE International Conference on Information Reuse and Integration (IRI), pages 144–153. IEEE, 2017.

3. O. Bobrowski and P. Skraba. Homological percolation and the euler characteristic. Phys Rev E, 101(3-1):032304, 2020.

4. M. Boguna, F. Papadopoulos, and D. Krzakouk. Sustaining the internet with hyperbolic mapping. Nat Commun, 1:62, 2010.

5. Alberto Cacciola, Alessandro Muscoloni, Vaibhav Narula, Alessandro Calamuneri, Salvatore Nigro, Emeran A Mayer, Jennifer S Labus, Giuseppe Anastasi, Aldo Quattrone, and Angelo Quarratone. Coalescent embedding in the hyperbolic space unsupervisedly discloses the hidden geometry of the brain. arXiv preprint arXiv:1705.04192, 2017.

6. Gunnar Carlsson. Topology and data. Bulletin of the American Mathematical Society, 46(2):255–308, 2009.

7. Gunnar Carlsson, Afra Zomorodian, Anne Collins, and Leonidas J. Guibas. Persistence barcodes for shapes. International Journal of Shape Modeling, 11(02):149–187, 2005.

8. C. J. Carstens and K. J. Horadam. Persistent homology of collaboration networks. Mathematical Problems in Engineering, 2013:1–7, 2013.

9. R. Chaudhuri, B. Gercek, B. Pandey, A. Peyrache, and I. Fiete. The intrinsic attractor manifold and population dynamics of a canonical cognitive circuit across waking and sleep. Nat Neurosci, 22(9):1512–1520, 2019.

10. A. Clauset, C. Moore, and M. E. Newman. Hierarchical structure and the prediction of missing links in networks. Nature, 453(7191):98–101, 2008.

11. Thomas MJ Fruchterman and Edward M Reingold. Graph drawing by force-directed placement. Software: Practice and experience, 21(11):1129–1164, 1991.

12. Tianchong Gao and Feng Li. De-anonymization of dynamic online social networks via persistent structures. In ICC 2019-2019 IEEE International Conference on Communications (ICC), pages 1–6. IEEE, 2019.

13. D. Gfeller and P. De Los Rios. Spectral coarse graining of complex networks. Phys Rev Lett, 99(3):038701, 2007.

14. K-I Goh, Byungnam Kang, and Doochul Kim. Universal behavior of load distribution in scale-free networks. Physical review letters, 87(27):278701, 2001.

15. K. I. Goh, G. Salvi, B. Kahng, and D. Kim. Skeleton and fractal scaling in complex networks. Phys Rev Lett, 96(1):018701, 2006.

16. Wei Guo and Ashis G Banerjee. Toward automated prediction of manufacturing productivity based on feature selection using topological data analysis. In 2016 IEEE international symposium on assembly and manufacturing (ISAM), pages 31–36. IEEE, 2016.

17. Danijela Horak, Shibodan Muleti, and Milan Rajković. Persistent homology of complex networks. Journal of Statistical Mechanics: Theory and Experiment, 2009(03), 2009.

18. Buyk Jung Hun. Effective epidemic containment strategy in hypergraphs. Phys Rev Res, 3(3):033282, sep 2021.

19. B. J. Kim. Geographical coarse graining of complex networks. Phys Rev Lett, 93(16):168701, 2004.
examining spreading processes on networks. Nat Commun, 6:7723, 2015.

41Brenda Y Torres, Jose Henrique M Oliveira, Ann Thomas Tate, Poonam Rath, Katherine Cumnock, and David S Schneider. Tracking resilience to infections by mapping disease space. PLoS biology, 14(4):e1002436, 2016.

42Zuxi Wang, Yao Wu, Qingguang Li, Fengdong Jin, and Wei Xiong. Link prediction based on hyperbolic mapping with community structure for complex networks. Physica A: Statistical Mechanics and its Applications, 450:609–623, 2016.

43Larry Wasserman. Topological data analysis. Annual Review of Statistics and Its Application, 5(1):501–532, 2018.

44D. J. Watts and S. H. Strogatz. Collective dynamics of 'small-world' networks. Nature, 393(6684):440–2, 1998.

45Z. Wu, L. A. Braunstein, S. Havlin, and H. E. Stanley. Transport in weighted networks: partition into superhighways and roads. Phys Rev Lett, 96(14):148702, 2006.

46K. Xia and G. W. Wei. Persistent homology analysis of protein structure, flexibility, and folding. Int J Numer Method Biomed Eng, 30(8):814–44, 2014.

47Afra Zomorodian and Gunnar Carlsson. Computing persistent homology. Discrete & Computational Geometry, 33(2):249–274, 2004.