

Spatial applications of topological data analysis: Cities, snowflakes, random structures, and spiders spinning under the influence

Michelle Feng¹* and Mason A. Porter²*Department of Mathematics, University of California, Los Angeles, California 90095, USA*

(Received 8 January 2020; accepted 29 June 2020; published 16 September 2020)

Spatial networks are ubiquitous in social, geographical, physical, and biological applications. To understand the large-scale structure of networks, it is important to develop methods that allow one to directly probe the effects of space on structure and dynamics. Historically, algebraic topology has provided one framework for rigorously and quantitatively describing the global structure of a space, and recent advances in topological data analysis have given scholars a new lens for analyzing network data. In this paper, we study a variety of spatial networks—including both synthetic and natural ones—using topological methods that we developed recently for analyzing spatial systems. We demonstrate that our methods are able to capture meaningful quantities, with specifics that depend on context, in spatial networks and thereby provide useful insights into the structure of those networks. We illustrate these ideas with examples of synthetic networks and dynamics on them, street networks in cities, snowflakes, and webs that were spun by spiders under the influence of various psychotropic substances.

DOI: [10.1103/PhysRevResearch.2.033426](https://doi.org/10.1103/PhysRevResearch.2.033426)

I. INTRODUCTION

Many complex systems have a natural embedding in a low-dimensional space or are otherwise influenced by space, and it is often insightful to study such spatial complex systems using the formalism of networks [1,2]. In a spatial network, the location of nodes and edges in space can heavily inform both the structure of the network and the behavior of dynamical processes on it. Indeed, obtaining a meaningful understanding of power grids [3–5], granular systems [6], rabbit warrens [7], and many other systems is impossible without considering the physical relationships between nodes in a network. For example, when studying traffic patterns on a transportation network, it is important to include information both about the physical distances between points and about the locations and directions of paths between heavily trafficked areas [8].

There are a variety of perspectives for studying spatial networks [1,9]. Many important ideas in such studies hail from quantitative geography [10,11]. In the 1970s, geographers were already studying the role of space in the formation of networks and in the activities of individuals and goods in geographical networks. As data have become richer and more readily available, it has become possible to use increasingly intricate computational methods in the study of spatial networks, and a variety of complex-systems approaches have contributed greatly to the literature on spatial networks [1].

Researchers have also proposed various random models for spatial networks, and studying them yields baseline examples to compare to empirical networks [12–15]. There have also been investigations of the effects of certain spatial network properties on the behaviors of several well-known dynamical processes, including the Ising model [16], coupled oscillators [17], and random walks [18].

Although there is much existing work on the properties of spatial networks (e.g., degree distributions, shortest paths, and so on), there are relatively few network tools that leverage “global” structure in the traditional topological sense of the word. Current tools for studying global network structure tend to rely on aggregating local information in some way to paint a global picture of a network. By contrast, methods for understanding the global structure of a topological space rely intrinsically on information about the entire space. To illustrate the difference, consider a sphere. If we sample a neighborhood of any point on a sphere, we obtain a surface with the same properties as a plane. If we consider a collection of a sphere’s neighborhoods (which each resemble a plane) and stitch them together, we are able to obtain a lot of information about the sphere, but we are unable to describe the void in the center of the sphere. (For example, a stereographic projection of a sphere covers the sphere’s entire surface, but it fails to capture the void.) To fully understand the structure of a sphere, we must consider the entire sphere at once. Over the last few decades, algebraic topology has been very useful for characterizing the global structure of mathematical spaces [19,20] through its use of mathematical tools that consider spaces as global objects. By reframing spatial networks using the language of topological spaces, we can leverage existing topological tools to better understand their structures. For a case study with voting data, see our recent paper [21].

Homology groups, which were defined originally in algebraic topology and have been applied insightfully to a broad range of mathematical topics, provide one way to distinguish

*Present address: Department of Computing + Mathematical Sciences, California Institute of Technology, Pasadena, California 91125, USA.

Published by the American Physical Society under the terms of the Creative Commons Attribution 4.0 International license. Further distribution of this work must maintain attribution to the author(s) and the published article’s title, journal citation, and DOI.

between mathematical spaces based on their numbers and types of “holes” [19]. Moreover, the extension of homology to so-called “persistent homology” (PH) allows one to quantify holes in data in a meaningful way and has made it possible to apply homological ideas to a wide variety of empirical data sets [22,23]. PH is helpful for characterizing the “shape” of data, and the myriad applications of it include studies of protein structure [24–27], DNA structure [28], neuronal morphologies [29], computer vision [30], diurnal cycles in hurricanes [31], chaotic dynamics in differential equations [32], spatial percolation problems [33], and many others. Additionally, combining machine-learning approaches with PH has also been very useful for several classification problems [34–37].

Because it is so natural to apply PH to the study of the shape of data, many successful applications of it have been to spatial networks. One particular area of interest has been the study of granular materials, because PH is able to effectively capture geometric properties of granular substances [6,38,39]. In addition to analyzing geometric information, PH methods are also able to describe multiscale spatial relationships. Many biological applications to proteins and DNA rely on the ability of PH to illuminate features at multiple scales, as multiscale structures and compositions of these molecules are extremely important to their function. PH has also been applied to larger-scale biological systems, including leaf-venation patterns [40], aggregation models [41], human migration [42], networks of blood vessels [43], and the effects of psychoactive substances on brain activity [44]. The recent review article [45] includes an extensive discussion of applications of PH to networks.

One confounding factor in the use of PH to study spatial networks is that although PH is able to capture information across scales, traditional distance-based PH constructions can have difficulty with applications in which differences in scale may not be meaningful. For example, in most applications to human geographical data, differences in population densities between urban and rural areas can dominate analyses that employ traditional PH constructions, and they thereby miss signals that do not rely on such variation in density. In a recent paper [21], we examined the shape of voting patterns in the state of California and observed that traditional methods for computing PH are more likely to capture disparities in population density than to detect the presence of interesting voting patterns. To address this issue, we developed two PH constructions—one based on network adjacency and one based on the physical geometry of a map—that were successful at capturing these voting patterns. For a recent analysis of the difficulty of interpreting signal and noise in PH results, see Ref. [46]. For approaches other than PH for analyzing maps while accounting for density variation, see Refs. [47,48].

In the present paper, we apply our PH constructions from [21] to a variety of spatial complex systems to demonstrate the usefulness of these constructions in many domains. We show that our methods are well-suited to capturing interesting structural properties of spatial networks and can thereby yield fascinating insights into such networks, especially with respect to their global structure. Our examples include several synthetic graph models and dynamics on them, city street

networks (which we compare both within a city and across different cities), snowflakes, and webs that were spun by spiders under the influence of various psychotropic substances.

Our paper proceeds as follows. In Sec. II, we give technical background on PH and on our particular constructions. In Sec. III, we discuss our results from computing the PH of (1) Watts threshold model (WTM) dynamics on several well-known examples of synthetic networks and (2) networks that we construct from a variety of empirical data sets from diverse applications. We conclude in Sec. IV. A public repository of the code that we use for our computations is available at Ref. [49].

II. METHODS

A. Computing persistent homology

We now give a brief introduction to PH and tools for computing it. See Refs. [22,50,51] for more details. We begin by defining k -simplices and simplicial complexes. A k -simplex is a k -dimensional polytope that is a convex hull of $k + 1$ nodes. A face of a k -simplex is any subset (of dimension smaller than k) of the k -simplex that is itself a simplex. A simplicial complex K is a set of simplices that satisfies the following requirements: (1) if $\sigma \in K$ is a k -simplex, then every face of σ is in K and (2) if σ and τ are simplices in K , then $\sigma \cap \tau$ is a face of both σ and τ .

Given a data set X , we construct a sequence $X_1 \subseteq X_2 \subseteq \dots \subseteq X_l$ of simplicial complexes of some fixed maximum dimension. We call the sequence $\{X_i\}$ a “filtered simplicial complex” (or simply a “filtration”), and we call each X_i a “subcomplex” of the filtered simplicial complex. We equip each relation $X_i \subseteq X_{i+1}$ with an inclusion map. The filtered simplicial complex, along with its inclusion maps and the chain and boundary maps of each subcomplex, constitutes a “persistence complex.” The inclusion maps $X_i \hookrightarrow X_j$ induce maps $f_{i,j}: H_m(X_i) \rightarrow H_m(X_j)$ between homology groups. The map $f_{i,j}$ allows us to track an element of $H_m(X_i)$ (the m th homology group of the subcomplex X_i) to an element of $H_m(X_j)$. The m th homologies of the persistence complex are given by the pair

$$\left(\{H_m(X_i)\}_{1 \leq i \leq l}, \{f_{i,j}\}_{1 \leq i \leq j \leq l} \right), \quad (1)$$

and we call them the “ m th persistent homology” of X . We refer to the collection of all m th persistent homologies as the “persistent homology” (PH) of X .

Consider a generator $x \in H_m(X_i)$ for some m and i . If x is not in the image of $f_{i-1,i}$, we say that x is “born” at time i . Correspondingly, if $x \in H_m(X_i)$ and $f_{i,i+1}(x) = 0 \in H_m(X_{i+1})$, we say that x “dies” at time $i + 1$. If for every $j \leq l$, we have that $f_{i,j}(x) \neq 0$, then we say that x never dies, and we assign a death time of ∞ to the element (i.e., generator) x . For each element x of the PH of X , there is a birth time b_x and a death time d_x , and the collection of intervals $\{[b_x, d_x]\}$ is the “barcode” of X . Generators with longer associated half-open intervals $[b_x, d_x)$ are more persistent. It is traditional to construe more-persistent intervals as indicators of a signal and less-persistent intervals as indicators of noise, although recent work (see, e.g., [21,52]) indicates that it is not always possible to interpret persistence in this way.

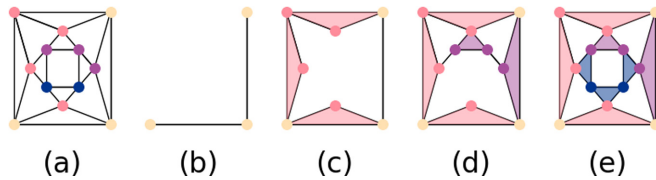


FIG. 1. We illustrate an adjacency construction of persistent homology (PH) on (a) a planar graph, whose nodes we color according to a function value from yellow to dark blue. At each filtration step [see panels (b)–(e)], we add all nodes with a given range of function values. We also add any edges between these new nodes, as well as any edges between these new nodes and existing nodes, and we fill in any triangles that form. Only cycles of length three form triangles, so the graph in panel (a) yields five infinite-length features in H_1 [as one can see from the five holes that remain in panel (e)].

The collection of features in each $H_m(X_i)$ describes the topological properties of the filtration $\{X_i\}$. Intuitively, each feature in H_m corresponds to some m -dimensional void. In H_0 , features are connected components; in H_1 , features are loops. By considering the PHs of $\{X_i\}$, we can examine how the connectedness of $\{X_i\}$ changes for each step (i.e., each X_i) of the filtration in each dimension. For example, a short-persistence feature in H_0 is a connected component that appears and combines quickly with another component. PH records all features and their persistences, allowing us to take a global view of topological changes in each filtration step of $\{X_i\}$.

In the present paper, we use the software package GUDHI [53,54] to compute PH of the filtered simplicial complex $\{X_i\}$. We construct $\{X_i\}$ from X using two different constructions, which we developed recently in a paper on voting data [21].

B. Adjacency construction of PH

We now describe a way to construct a filtered simplicial complex based on network adjacencies. Consider a network in the form of a graph (V, E) , with numerical data $f(v)$ associated with each node v . For a given filtration step X_i , let the 0-simplices of X_i be given by $v \in V$ such that $f(v) \leq \epsilon$ for some value ϵ . For any edge $(u, v) \in E$, if $u \in X_i$ and $v \in X_i$, we add (u, v) to X_i . Finally, to X_i , we add all triangles (u, v, w) such that (u, v) , (v, w) , and (u, w) are in X_i . We repeat this process for X_{i+1} , but now we use a larger value of ϵ . By construction, each $X_i \subseteq X_{i+1}$, and we have a valid filtered simplicial complex. See Fig. 1 for an illustration of such a filtered simplicial complex.

This adjacency construction tracks topological changes in a network as it grows. The homology group $H_m(X_0)$ characterizes the topology of the first filtration step of a filtered simplicial complex. As one adds more nodes, edges, and faces to the simplicial complex, its topology changes and is recorded in $H_m(X_i)$. By choosing f carefully, we can control which subset of a network exists in the first filtration step, and we can also control how the network expands. For example, in Ref. [21], by attaching voting data to a network of precincts of a county, we used this adjacency construction to examine how the topology of the county changes as one expands the range of voting preferences of the precincts that one includes.

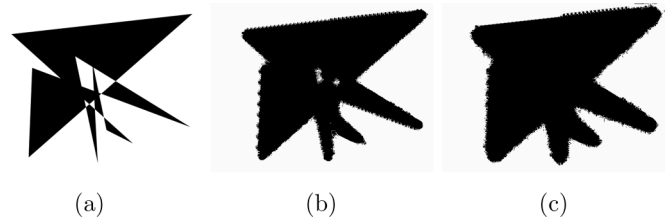


FIG. 2. Illustration of level-set dynamics. Starting from (a) an initial black-and-white image, we apply level-set evolution (2) for several steps to obtain the image in (b) and then the one in (c). In these images, the white space in the center of the image shrinks until eventually it is completely covered by the expanding black surface.

In some of our applications in the present paper, we use an alternative adjacency construction in which we associate data $g(u, v)$ to each edge (u, v) , instead of to the nodes. This construction differs from the one above only in that we define the function $\tilde{f}(v) = \min_{\{u:(u,v) \in E\}} g(u, v)$. We then proceed with the above adjacency construction, but we substitute $\tilde{f}$ for f . We recently introduced our main adjacency construction in Ref. [21], and we introduce this adaptation of it to edge-based data in the present work.

C. Level-set construction of PH

The other PH construction that we use (again see Ref. [21] for details) involves describing data as a manifold, rather than as a graph. Let M denote a two-dimensional (2D) manifold, such as data in an image format. We consider the boundary Γ of M and construct a sequence

$$M_0 \subseteq M_1 \subseteq \dots \subseteq M_n$$

of manifolds (where M_0 is an approximation of M). At each time t , we evolve the boundary Γ_t of M_t outward according to the level-set equation. (See Ref. [55] for a thorough exposition of the level-set equation and level-set dynamics.) Specifically, for a manifold M that is embedded in $\mathbb{R}^2$, we define a function $\phi(\vec{x}, t): \mathbb{R}^2 \times \mathbb{R} \rightarrow \mathbb{R}$, where $\phi(\vec{x}, t)$ is the signed distance function from $\vec{x}$ to Γ_t at time t . We propagate Γ_t outward at velocity v using the level-set equation

$$\frac{\partial \phi}{\partial t} = v|\nabla \phi| \quad (2)$$

until all homological features have died. Because this evolution gives a signed distance function at each time t , we take M_t to be the set of points $\vec{x}$ such that $\phi(\vec{x}, t) > 0$. (This corresponds to points that are surrounded by the boundary Γ_t .) We show an example of this evolution in Fig. 2. Throughout the present paper, we use $v = 1$. Different values of v cause the level set to evolve faster (if $v > 1$) or slower (if $v < 1$), resulting in a different number of time steps (and hence a different number of filtration steps) in our evolution. However, we obtain the same homological features, although with different birth times and death times. If v is sufficiently large, it is possible for all of the features to have the same birth and death time, such that no features occur after the first filtration step. When $v = 0$, there is no evolution.

By imposing $\{M_i\}$ over a triangular grid of points, as described in Ref. [21], we obtain a corresponding simplicial

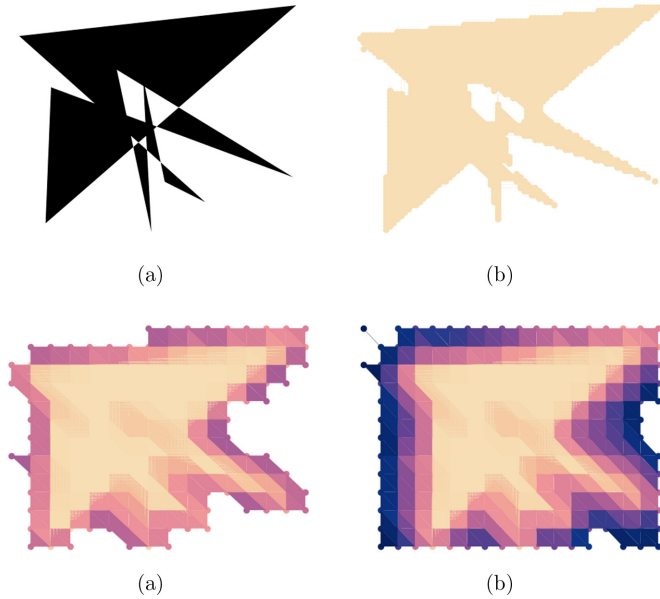


FIG. 3. Illustration of a level-set adjacency construction of PH. In (a), we show a synthetic image that we use as an initial manifold for level-set evolution. In (b)–(d), we show various filtration steps of the filtered simplicial complex that we generate by performing a level-set evolution on the image in panel (a). Panel (b) shows the simplicial complex that we obtain by overlaying the image in panel (a) on a triangular grid. In panels (c) and (d), we add new nodes, edges, and triangles to the image as it evolves outward. Darker colors indicate simplices that enter the filtration at a later time.

complex X_i for each M_i . In Fig. 3, we illustrate this simplicial complex. We construct this level-set complex using a polygon whose points we choose uniformly at random from $[0, 1] \times [0, 1]$ as an initial synthetic image. Because the level-set equation (2) evolves outward, we automatically satisfy that condition that $X_i \subseteq X_{i+1}$, so $\{X_i\}$ is a filtered simplicial complex. Our implementation of the level-set method works with any black-and-white image (or any image that one can describe as a piecewise-constant function $h : \mathbb{R}^2 \rightarrow \{0, 1\}$). We expect our level-set approach to capture information about H_0 and H_1 for any such image. The level-set approach also captures geometric information, which can be useful for some applications; however, this may make it difficult to capture information about holes that are visually irregular. Throughout the present paper, we compare images that have roughly the same resolutions, where we take the image resolution from raw image data. Because image size should primarily affect the computation time of our level-set approach—but not the order in which features appear and disappear as an image evolves—we expect that it is possible to adapt our level-set construction for comparing images of different resolutions. Possible approaches for such an adaptation include normalizing image sizes or adjusting the resolution of the triangular grid that one uses for each image.

III. APPLICATIONS

We now discuss applications of PH to both synthetic networks and empirical spatial networks from a diverse variety of applications.

A. Synthetic networks

In this subsection, we discuss applications of our adjacency PH construction to a dynamical process on synthetic networks in which space plays an important role. For each network (V, E) , we run the WTM [56] on it. Given a graph, we select a fraction $\rho_0 = 0.05$ of its nodes uniformly at random to be “infected” at time 0. At each time (which is discrete), we compute the fraction of each node’s neighbors that are infected. We then synchronously update the states of the nodes [57]. If the fraction of a node’s neighbors that are infected meets or exceeds a threshold (in our case, the threshold is $\varpi = 0.18$ for all nodes), the node becomes infected (if it is not already infected). Once a node is infected, it stays infected forever. We take this implementation of the WTM to be the generator of a function $f : V \rightarrow \mathbb{N}$ [58], where $f(v)$ is the time at which node v becomes infected. We say that infected nodes are in the set I , and we use the term “infection network” for the associated induced subgraph. If v never becomes infected, we set $f(V) = \max_{v \in I} f(v) + 1$, so that we eventually add all nodes to a filtered simplicial complex. The resulting filtered simplicial complex consists of the subgraphs that are generated by I at each time. See Refs. [59–61] for studies of the WTM on spatial networks.

We use the parameter values $\rho_0 = 0.05$ and $\varpi = 0.18$ throughout this section. We expect changes in ρ and ϖ to affect the birth times and death times of features in a filtered simplicial complex. Using a different value of ρ_0 entails considering a different fraction of initially infected nodes. A larger value of ρ_0 yields a larger simplicial complex at the first filtration step, and smaller value of ρ_0 yields a smaller simplicial complex. Using a larger value of ϖ results in fewer nodes becoming infected at each time, and it thus takes more filtration steps before the simplicial complex stops growing. Because our underlying graph is the same for any choice of values of ρ and ϖ , we do not expect changes in the homology of the last filtration step, unless ρ or ϖ are sufficiently small such that some nodes in a graph never become infected. However, one can certainly obtain a different PH for different values of ρ or ϖ , as nodes and edges can join the filtered simplicial complex at different times and (more importantly) in different orders.

We examine topological changes in the infection networks of three different types of synthetic networks (see Fig. 4). We first examine random geometric graphics (RGGs) [62]. For each instance of an RGG, we pick 100 nodes uniformly at random from the unit square. If the Euclidean distance between two nodes is less than or equal to 0.1, we add an edge between them [see Fig. 4(a)]. Our second type of synthetic network is a square lattice with 100 nodes. We arrange the 100 nodes in a 10×10 grid on the unit square, and we then connect the nodes along the grid lines [see Fig. 4(b)]. Our third type of synthetic network is a Watts–Strogatz (WS) small-world network [63,64]. We begin with a ring of 100 nodes, and we then connect each node to its $k = 2$ nearest neighbors on each side. We then rewire each edge uniformly at random with a probability of $p = 0.1$ using the implementation of the WS model in NETWORKX [65]. In this version of the WS model, one removes each rewired edge (u, v) before replacing it with a new edge (u, w) . We show an example of a WS graph in Fig. 4(c).

Sample random graphs with nodes colored by infection time

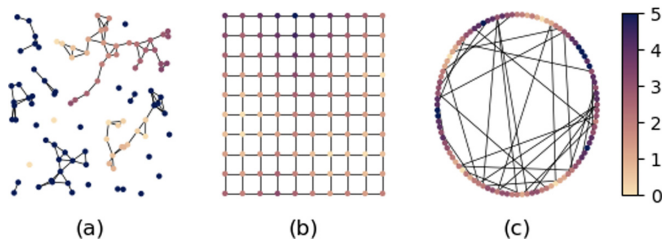


FIG. 4. An instance of each of our synthetic networks with Watts threshold model (WTM) dynamics on it. The corresponding persistence diagrams (PDs) are in Figs. 5, 6, and 7. We color the nodes based on the time that they become infected. The three types of synthetic networks are (a) a random geometric graph, (b) a square lattice network, and (c) a Watts–Strogatz small-world network.

For each type of synthetic network, we consider 100 instances, which we generate using NETWORKX. For the RGG and WS networks, each instance is a different graph; the square lattice network is deterministic. For all three types of networks, each instance has a different initial set of infected nodes. We show visualizations of each of these types of networks (with WTM dynamics on it) in Fig. 4.

Our adjacency construction for the WTM on a network begins by selecting the initially infected nodes and the edges between them to create an infection network. As an infection spreads, we add more nodes and edges to the infection network until eventually we have added all nodes and edges to it. (In the last filtration step, we add any remaining uninfected nodes and their associated edges.)

We show the results of our PH computations using persistence diagrams (PDs). By examining the PHs of the RGGs (see Fig. 5), we see for our parameter values that an infection network tends to have several connected components, resulting in a large number of features in H_0 . However, because of the spreading behavior of the WTM, new nodes can become infected only via their infected neighbors. Features in H_0 record connected components of a graph, so newly infected nodes join existing connected components. Therefore, such features can only be born at time 0 or in the last step, which is when we add all remaining uninfected nodes to our filtered simplicial complex. By contrast, features in H_1 are relatively rare, as most cycles that occur in an RGG are filled because of the uniform probability distribution of the node locations.

For a square lattice network (see Fig. 6 for a PD of an instance of the WTM on such a network), we first note that there is only a single infinite-length feature in H_0 , as the final infection network necessarily consists of a single connected component. Consequently, H_0 consists of a set of features that are born at time 0 and eventually merge (and therefore die), resulting in a single infinite-length feature. Additionally, there are a constant number (81, to be precise) of features in H_1 , because when we construct a simplicial complex, every grid cell of the lattice is a feature in H_1 at the last filtration step. (For this network and other planar networks, we use the word “cell” to refer to the spaces that are surrounded by edges.) However, these features can be born at a variety of times,

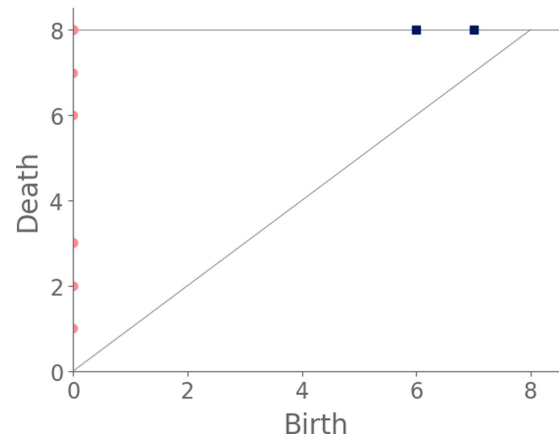


FIG. 5. The PD of an instance of the WTM on an RGG. We plot each feature as a point on the PD, for which the horizontal coordinate represents the birth time and the vertical coordinate represents the death time. We plot features with infinite persistence (i.e., features that do not die within the range of filtration parameters that we use for our PH computation) on a horizontal line at the top of the PD. We plot features in H_0 (which indicates the connected components) as pink circles, and we plot features in H_1 (which indicates the one-dimensional holes) as dark-blue squares.

as the filtration does not include every lattice cell until every node of the graph has entered the filtration.

In Fig. 7, we see that a WS small-world network also eventually has an infection network that consists of a single connected component. However, the WS networks consistently have more features in H_1 than the RGG networks, because the former’s (nongeometric) shortcut edges usually result in splitting an existing cycle (and hence a feature in H_1) into two cycles.

We summarize our observations about the various synthetic networks in Table I, in which we give the means and standard deviations of the number of features during the temporal evolution of the WTM on each type of synthetic graph. We count the features that appear at any time (i.e., we count the total number of features that are born) during the WTM dynamics.

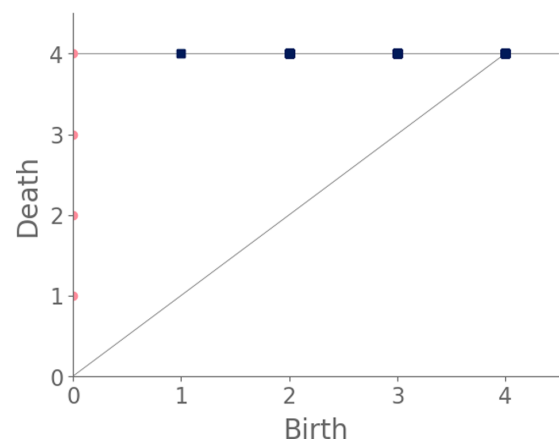


FIG. 6. The PD of an instance of the WTM on a square lattice network.

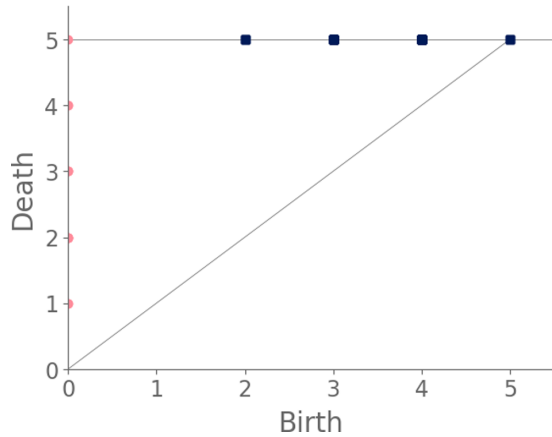


FIG. 7. The PD of an instance of the WTM on a WS network.

B. Street networks in cities

The field of urban analytics has grown rapidly in the last several years [1,11,66]. Increasingly powerful computational tools have allowed researchers to characterize cities in terms of their street networks [67], and a variety of approaches from network analysis have been applied to the study of city street networks [8,68–73]. In the present subsection, we use city street networks as base manifolds in our level-set construction, and we thereby characterize cities based on their PHs. We use these PHs to compare morphologies both within a single city and across a variety of cities. We obtain each of our city street networks with the software package OSMNX [74] using latitude–longitude coordinates and a $1 \text{ km} \times 1 \text{ km}$ square that is centered at specified coordinates. In each example, we indicate how we choose these coordinates.

The first filtration step of a filtered simplicial complex that results from our PH construction consists only of the streets in a network. As we increase the filtration time, we slowly add city blocks to the complex, and the topology changes as those blocks are filled in. More regular city blocks are more likely to be filled in without creating any new homological features, and larger blocks take longer to be filled in. Our construction is thereby able to capture information about the size and regularity of city blocks. The existence of dead ends tends to lead to the “pinching” of blocks into multiple homological features—as dead ends expand, they lengthen and eventually meet with nearby streets, cutting through blocks in the process—so our approach also yields information about dead ends.

TABLE I. Means and standard deviations of the numbers of features in H_0 and H_1 during the temporal evolution of the WTM across all instantiations of each type of synthetic graph that we consider. We conduct 100 simulations for each network model, and we count the features that appear at any time during the WTM dynamics.

	Mean (H_0)	STD (H_0)	Mean (H_1)	STD (H_1)
RGG	23.16	3.1897	1.2	1
Square lattice	4.56	0.5886	81	0
WS	8.29	2.0214	26.95	5.2314

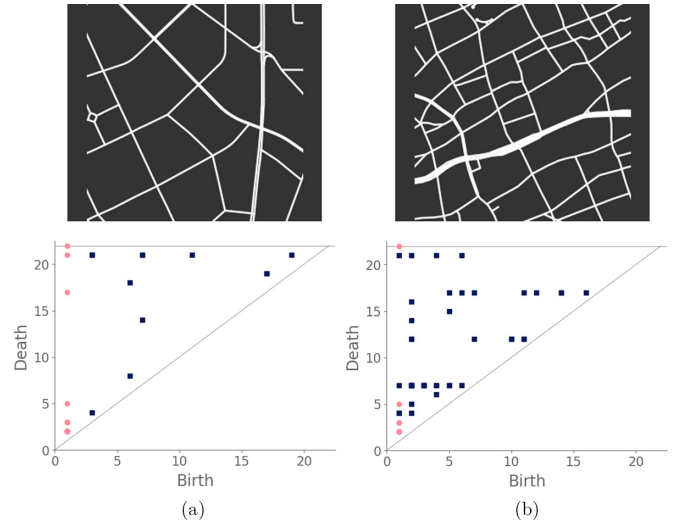


FIG. 8. Sampled street networks from (a) Pudong New Area and (b) the Zhabei district in Shanghai. (We generated both maps using OSMNX [74].)

1. Comparing different regions of the same city

We sample 169 points from the city of Shanghai using a SHAPEFILE of Shanghai’s administrative-district boundaries that we downloaded from ARCGIS [75]. From the SHAPEFILE, we obtain a bounding box for each point. We sample uniformly within this bounding box, discarding points that do not lie within the polygonal district geometry that is defined in the SHAPEFILE. We stop sampling when we reach the desired number of points. In total, we sample 10 points from each administrative district, and we also include nine historical landmarks with coordinates from Google Maps [76]. In Fig. 8, we show maps and their associated PDs for two examples.

After computing PH (which we summarize as a PD) for each map, we compute the bottleneck distance between each pair of maps. The bottleneck distance is a distance function that is defined on the space of PDs. It gives the shortest distance d for which there exists a perfect matching between the points of two PDs (along with all diagonal points), such that any pair of matched points are at most a distance d from each other, where we use the supremum norm in $\mathbb{R}^2$ to compute the distance between points. Once we have pairwise bottleneck distances between PDs, we perform average-linkage hierarchical clustering into three clusters. (We chose to have three clusters after looking at the dendrogram that resulted from hierarchical clustering.) We can replace the bottleneck distance with a different distance function (such as a Wasserstein distance [77]) on PDs or cluster our PDs using a different clustering algorithm. We do not discuss the impact that such choices may have on our results, although we note in passing that we performed k -medoids clustering [78] for our case study of Shanghai and obtained qualitatively similar results.

In Fig. 9, we show the sampled points (which we color according to their cluster assignment). We observe that the three clusters consist largely of historical areas (“City center”), concession-era areas (“Transition areas”), and modern areas (“New construction”). In Fig. 10, we show administrative

Shanghai: Sampled points by cluster assignment

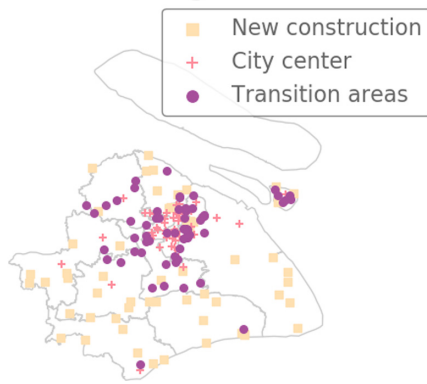


FIG. 9. Sampled points in Shanghai. We color these points according to their cluster assignment from average-linking hierarchical clustering of areas of Shanghai into three clusters.

districts along with the years that they were developed. We partition them based on the percentage of the sample points that are in each cluster.

2. Comparing street networks from different cities

We continue our analysis of cities by characterizing and comparing the structures of street networks of 306 cities across the globe. We downloaded latitude and longitude coordinates from SimpleMaps [79] and selected all cities with a population of at least 1.5 million people. Given these latitude and longitude coordinates, we use OSMNX [74] to obtain street networks. We then compute PH for each city and cluster their PDs using average-linkage hierarchical clustering with

three clusters. We sometimes refer to a city in a given cluster as a city of a certain “type.” Our results depend on the specific latitude and longitude coordinates in our downloaded data set. Accordingly, our results are influenced by the particular location of a city’s coordinates, which are the standard ones in SimpleMaps.

In the following paragraphs, we describe our clusters of cities. We define “blocks” to be the cells of a planar street network. Although our level-set construction for computing PH is not designed explicitly to characterize blocks, we take advantage of the fact that our level-set construction takes the set of streets as its initial manifold. As the streets expand outward according to the level-set equation (2), they fill in the blocks. Larger blocks take longer to fill in, and blocks fill in more evenly when they are closer to circular in shape. Roughly, we characterize block sizes based on the death times of features in H_1 : “small” sizes correspond to early death times (specifically, less than 10), “medium” sizes correspond to death times between 10 and 15, and “large” sizes correspond to late death times (specifically, more than 15). We also designate blocks as “regular” (when they are close to a regular convex polygon) or “irregular” (for blocks that do not resemble a rectangle or some other regular convex polygon). If a block is very irregular, then as its streets expand (in the level-set evolution), it is possible that narrow parts of the block will shrink and close off, such that the block separates into smaller blocks. We refer to this phenomenon as “pinching.” Our three main clusters are dominated by (1) gridlike cities, (2) cities with gridlike patches that are interspersed with larger, nongridlike blocks, and (3) cities that have a large number of nongridlike structures (specifically, dead ends or large holes) that interrupt other structures. We use the term “interrupted grid” to describe cities that either (1) are mostly gridlike, but with some patches that are not gridlike, or (2) consist of patches of disparate grids that are stitched together (with other features between them).

Our first major cluster has 99 cities and is dominated by cities with small, gridlike blocks. All regions of the world have some cities of this type, but North America has the largest percentage (relative to all of the cities that we sample from that continent) of these gridlike cities and Europe has the smallest percentage of them. The block sizes in this cluster tend to be small or of medium size, resulting in filtrations whose maximum filtration value tends to be small in comparison to cities in the other two clusters. In the PDs, we also observe that the distributions of the death times of the features in H_1 tend to be close to uniform and tend to occur over a small range of filtration steps. Such distributions occur because these gridlike cities tend to have even distributions of block sizes, even though they include some areas with slightly smaller and/or slightly larger grid sizes. They do not have large blocks, so they do not have features in H_1 with late death times. In Fig. 11, we show an example of a city in cluster one.

Our second major cluster has cities with patches of grids that are interspersed with structures that are not gridlike. This cluster, with 149 cities, is the largest of our three clusters. The PDs in this cluster tend to have larger maximum death times than the PDs of the cities in our first cluster. In the PDs, gridlike blocks yield collections of features in H_1 with early death times; and the larger, nongridlike structures yield

Administrative districts of Shanghai partitioned by the percentage of sampled points in each cluster

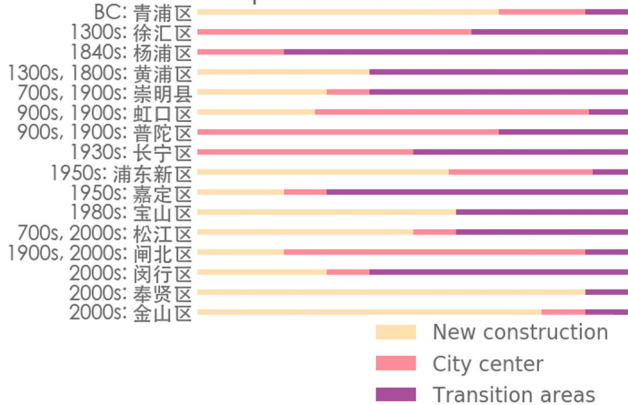


FIG. 10. Partitioning of administrative districts in Shanghai into our three clusters. (We order the districts roughly by their year of development.) Most of the older districts have a larger percentage of points that are assigned to the “City center” cluster, whereas the points in the “Transition areas” cluster tend to occur in districts that included development in the 19th and early 20th centuries. The “New construction” cluster is the most common assignment for administrative districts from the 1950s or later.

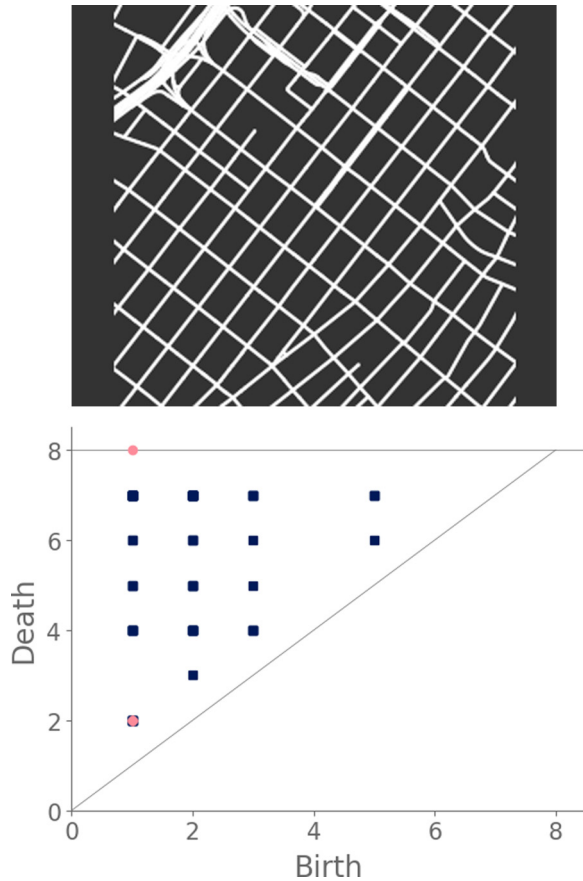


FIG. 11. Cities in our first major cluster have gridlike street layouts. One example of a city in this cluster is Los Angeles. We show its street network in the top row and its associated PD in the bottom row.

features in H_1 with late death times. The nongridlike areas in these cities tend to have fairly regular shapes, resulting in a relatively small number of features in H_1 with late birth times. Such late-birth-time features usually correspond to the pinching of blocks, which can occur either via dead ends or via shape irregularities. By examining the dendrogram from our hierarchical clustering, we see that we can separate the second cluster into two subclusters. We show examples of cities in the two subclusters in Fig. 12. The first of these subclusters consists mostly of cities that have large patches of gridlike structure, with a small number of large blocks that interrupt the grids. The PDs of the cities in this subcluster tend to have a large number of features in H_1 with early death times, and they tend to have only a small number of isolated features in H_1 with late death times. The second subcluster of our second major cluster consists mostly of cities with small patches of grids that are mixed with large irregular blocks. The PDs of the cities in this subcluster tend to have a larger number of features in H_1 with late death times than is the case for the cities in the other subcluster of cluster two.

Our third major cluster, with 58 cities, consists of cities with a large number of nongridlike structures. In particular, many of these cities include a large number of dead ends, rectangular blocks that are not arranged in a grid, or both. We observe streets that do not continue through particular blocks

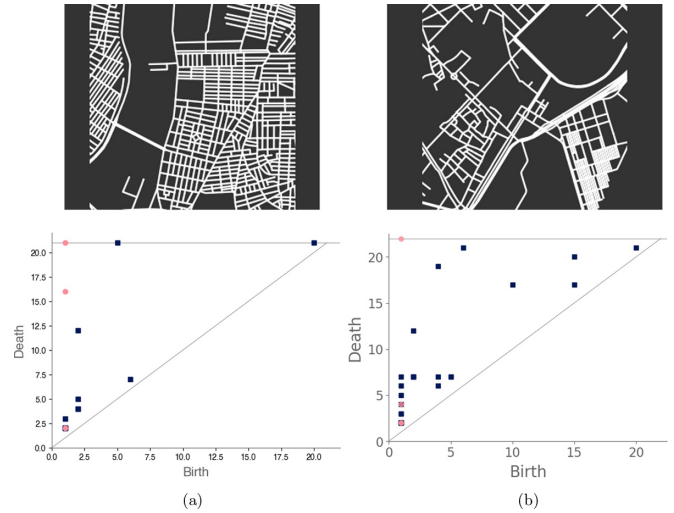


FIG. 12. Cities in our second major cluster have patches of gridlike structure that are mixed with large blocks. Examples of cities in this cluster are (a) Aleppo and (b) Barcelona. We show their street networks in the top row and their associated PDs in the bottom row. Aleppo illustrates the idea of having holes in a large grid and is an example of a city in the first subcluster of cluster two. Barcelona, which is in the second subcluster, is an example of a city with small patches of gridlike structure.

(e.g., there is a street, it is obstructed, but then it continues after the obstruction), which leads to a mixture of block sizes even in areas of a city that tend to have regular blocks. We refer to these situations as “obstructions.” In comparison to the other two clusters, the PDs of the cities in this cluster have a larger number of features in H_1 with medium death times (specifically, in the range 10–15), and many of these features are close to the diagonal. This is common when large blocks are pinched into several regions, as the smaller regions are born at the pinching time, rather than near the beginning of the filtration. Therefore, they do not survive long enough to have a late death time. By examining the dendrogram from our hierarchical clustering, we see two clear subclusters. However, one of these subclusters consists of only two cities (Beirut and Nanyang). The PDs of both of these cities are dominated by two features in H_1 with late death times, and they also have notable features in H_1 with medium death times. In Fig. 13, we show examples of cities in cluster three.

We color our cities according to their major cluster and show them on a world map in Fig. 14. In Fig. 15, we show the breakdown of cities from each continent into the various clusters. We calculate PH for only four major cities in Oceania, so we cannot draw strong conclusions from the cluster assignments of those cities. Among the other regions, we observe that North America has the largest proportion of cities with gridlike street layouts and the smallest proportion of cities with nongridlike layouts. By contrast, Europe has the smallest proportion of cities with gridlike street layouts. This is consistent with the common perception that North American cities are much more gridlike than European cities. In all regions, we also observe that a large fraction of the cities are interrupted grids. Additionally, we observe that South

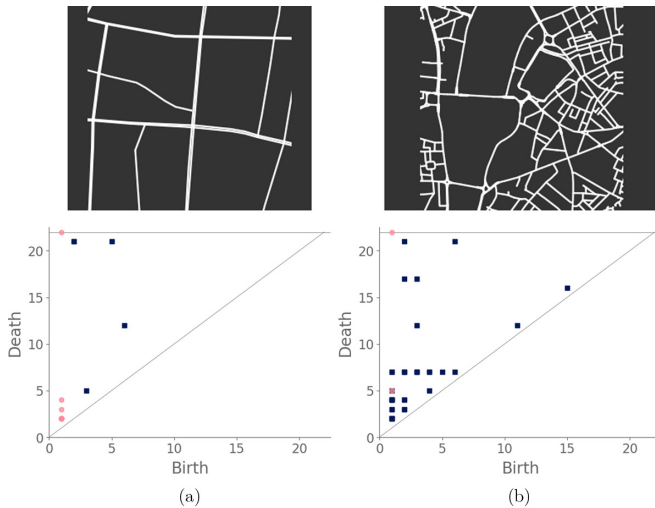


FIG. 13. Cities in our third major cluster include (a) Nanyang and (b) London. We show their street networks in the top row and their associated PDs in the bottom row. Cities in our third major cluster have features like dead ends, irregular blocks, and obstructions. This leads to a large range of block sizes and hence to features in H_1 that have medium death times. (Such features are rare in the other two major clusters.) In this figure, we see that Nanyang has several streets with obstructions and that London has dead ends and a broad distribution of block sizes.

America, Africa, and Asia have similar distributions of city types.

Interestingly, from the map in Fig. 14, South America, Asia, and Africa appear to have areas that are dominated by particular major clusters. We observe nongridlike cities in the northern part of South America, whereas we see gridlike cities

along its east coast. In Africa, most of the nongridlike cities occur along the western coastline. In Asia, most of the gridlike and nongridlike cities occur in East Asia, whereas Southeast Asia is dominated by interrupted grids. Across the map, there appears to be a potential equatorial band of nongridlike cities. We do not have an explanation for these patterns, but they are fascinating and seem worthy of future research efforts.

3. Comparison of our classification of cities to that of Louf and Barthelemy [81]

We compare our results to the city classification of Louf and Barthelemy [81], who associated each city with a conditional probability distribution that captures the areas and shapes of its blocks. We choose their classification as a point of comparison because they studied a wide range of cities and (like us) codified cities from a block-based perspective. They used the word “fingerprint” as a monicker for their block-based representation of cities. In our method, we codify cities according to their PHs, which we generate using the level-set construction of Sec. II C. Both the approach of Ref. [81] and our approach capture information that is based on city blocks, although our PH representation differs substantially from the fingerprints of Ref. [81].

Louf and Barthelemy clustered cities into four groups, whereas we have chosen to cluster our cities into three groups. In Ref. [81], European and North American cities largely inhabit the same cluster (group three in Ref. [81]), but they appear in distinct subclusters, demonstrating that there is a substantive difference between cities in the two regions. Our method finds that North America has the largest proportion of cities with gridlike streets among all of the regions and that Europe has the smallest proportion of such cities.

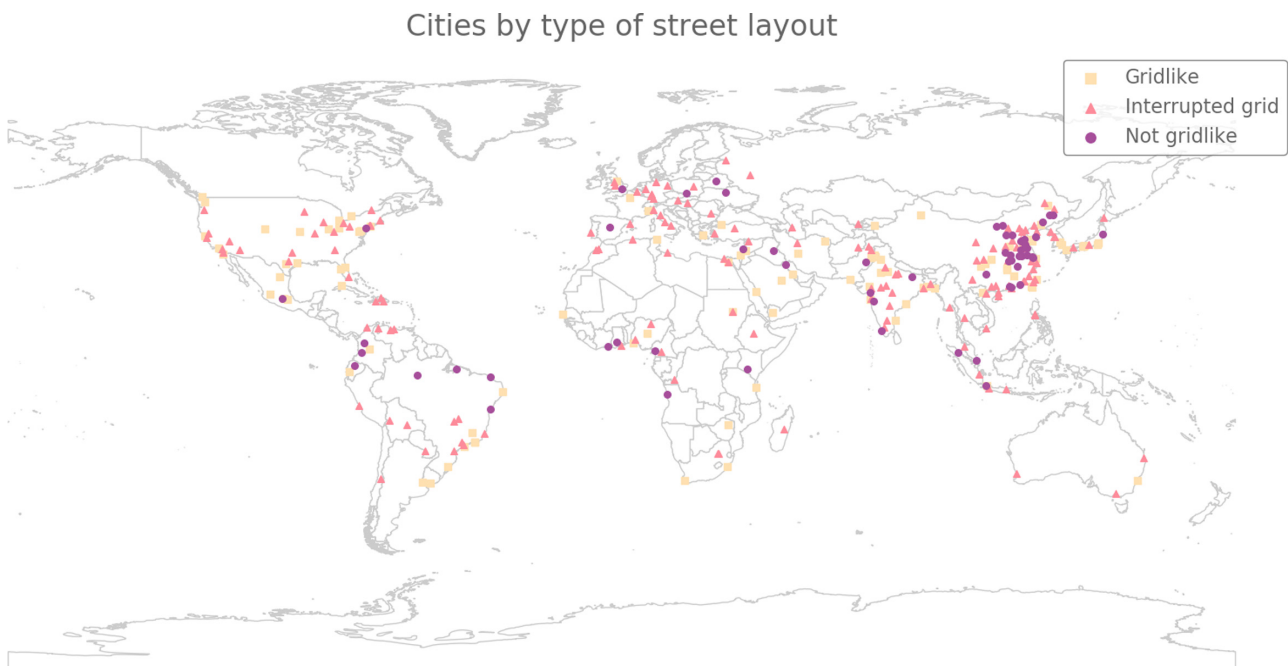


FIG. 14. Cities colored by their cluster assignments from average-linkage hierarchical clustering of cities into three clusters. (The SHAPEFILE of the world map is from Ref. [80].)

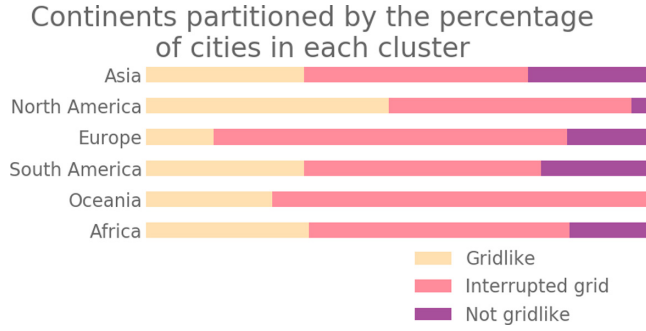


FIG. 15. Continents partitioned based on the percentage of cities in our three major clusters.

In contrast to the above situation, Africa, Asia, and South America have a fairly balanced composition of city types, with a potential equatorial band of nongridlike cities. Louf and Barthelemy observed several clusters (groups one, two, and four in Ref. [81]) that occur predominantly in Africa, Asia and Oceania (which they combined into one entity), and South America. Notably, none of our clusters are as dominant as group three (which was described as having heterogeneous block sizes and shapes) of Ref. [81], although we do observe that our cluster of cities with interrupted grids (such cities are characterized in part by their heterogeneous block sizes) is also our largest cluster.

Now that we have compared our results to those of Ref. [81], we briefly compare and contrast the types of information that the two methods capture. Recall that our level-set construction for PH generates filtered simplicial complexes that first consist of streets and then expand outward to absorb the blocks between them. The PH of such a filtered simplicial complex thereby gives a low-dimensional representation of the original image of a city street network. Because irregularly shaped blocks are absorbed into the surrounding streets at a different rate than regular blocks, we capture information about the regularity of each block. Louf and Barthelemy's method also uses information about the regularity of block shape. See Eq. (3.2) in Ref. [81] for a precise mathematical statement of how they measured the regularity of blocks. It is related to a subset of so-called "compactness measures" [82] (which are used in the study of gerrymandering [83,84]) that compare the area of a shape to the area of the circle in which the shape is inscribed.

Because the original image of a city street network includes information about the spatial relationships between blocks, the PH that results from our approach also encodes some of this information. By contrast, Louf and Barthelemy's fingerprints do not encode information about the spatial relationships of blocks to each other. Additionally, our method captures information from dead ends, which Louf and Barthelemy discarded.

Overall, although both our approach and that of Ref. [81] use a block-based representation to characterize cities, there are subtle differences in the way that the two approaches encode block information. Nevertheless, the commonality of a block-based perspective results in some similarities. For example, the clusters that result from the two approaches seem to be based heavily on block size and regularity. However, our

Hierarchical clustering of snowflake images

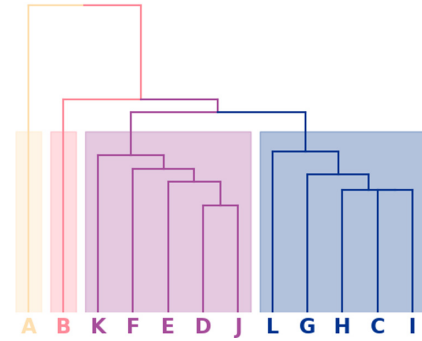


FIG. 16. Dendrogram from clustering the snowflakes in Fig. 20.

approach appears to prioritize spatial relationships between different clusters of blocks (specifically, whether blocks are arranged in a grid); such information is not captured in the approach of Ref. [81]. Consequently, the two approaches capture different city morphologies, and we expect them to be useful as complementary techniques for studying structures in spatial complex systems.

C. Snowflakes

As a second application that uses empirical data, we consider snowflake crystals [85]. We start with 12 different images (from [85]) of snowflakes with different crystalline structures. (See Fig. 20 in the appendix.) Using the GNU Image Manipulation Program [86], we threshold these grayscale images (using a thresholding setting of 205) to create black-and-white images. From the black-and-white images, we compute level-set complexes and PHs, and we then perform average-linkage hierarchical clustering on the PDs to produce the dendrogram in Fig. 16.

The images of snowflakes consist of edges (the black lines in our images) and cells (the white spaces that are bounded by the edges). The edges of a snowflake intersect at nodes to form a network. We refer to the outermost extremes that extend from the center of a snowflake as its "points." The 12 snowflakes have fairly regular crystalline structures, so our computation of PH predominantly records information about the distribution of cell sizes in a snowflake. The inherent hexagonal nature of snowflakes and the regularity of their crystalline structures largely overwhelms our ability to use PH to glean information about their spatial relationships and irregularities. Constructing a simplicial complex that is better suited to capturing information about images with a large number of regular structures may yield better results.

Examining the clusters (see Fig. 16) reveals that snowflake A and snowflake B each reside in their own cluster and that the remaining snowflakes split into two clusters. Snowflake A's PD [see Fig. 17(a)] is dominated by a feature in H_1 with an early birth and a late death. (See the blue square in the top-left corner of the diagram.) This arises from the large feature that is formed by the bold ring near the center of the snowflake. None of the other snowflakes have a bold central ring. More generally, we observe few features in the PD of snowflake

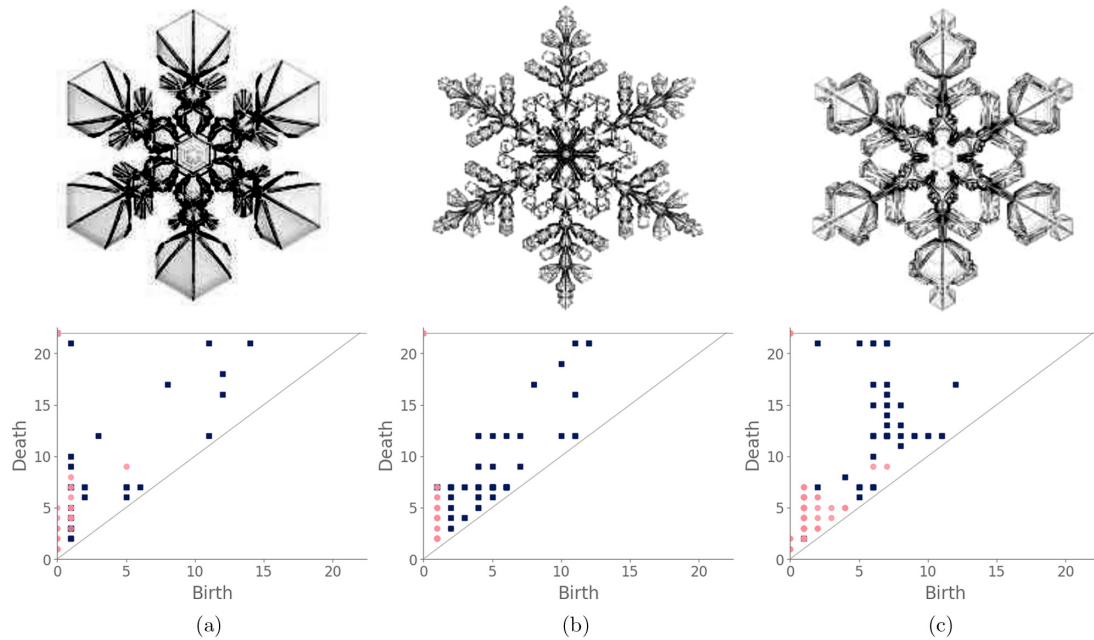


FIG. 17. Snowflakes can have a variety of crystalline structures, as we illustrate with (a) snowflake A, (b) snowflake B, and (c) snowflake D. We show the snowflake structures in the top row and their associated PDs in the bottom row. We show the structures of our full set of snowflakes in Fig. 20. (The images in the top row are from Ref. [85].)

A. By contrast, snowflake B's features are largely close to the diagonal [see Fig. 17(b)] because the initial manifold of the snowflake does not have large holes. Notably, we do not observe any points in the top-left region of its PD. The cell sizes in snowflake B are smaller than those in most of the other snowflakes, and even its central ring structure includes a large number of small holes. The remaining snowflakes either have more large holes than snowflake B, or they do not have a bold central ring like the one in snowflake A. We also note that the

PD of snowflake B is much closer than that of snowflake A to those of the other snowflakes.

D. Spiderwebs

Our final application is to the topology of spiderwebs. In 1948, Peter Witt began research on the effects of drugs on spiders to test whether garden spiders would shift their web-building hours if drugs were administered to them. Witt found that drugs affect the sizes and shapes of the webs that are produced by spiders [87]. He also found that higher doses of most drugs (e.g., $100\text{ }\mu\text{g}$ per spider, as opposed to $10\text{ }\mu\text{g}$ per spider) tend to lead to larger changes in the shapes of webs, including yielding more irregular webs. Witt eventually published more than 100 papers and several books on spiderwebs and the behavior of spiders. For more information on his experiments with psychotropic substances and spiders, see his 1971 review article [88]. In a 1995 technical briefing [89], NASA (which was inspired by Witt's research) proposed that spiders that are exposed to more toxic substances produce webs that are more deformed (in comparison to a web that is spun by a drug-free spider) than spiders that are exposed to less toxic substances. Additionally, using techniques from statistical crystallography, they concluded that spiders fail to complete more sides of their webs when they are under the influence of more toxic substances.

In our case study of PH in spiderwebs, we use five images from the NASA technical briefing [89] and two images from Witt [88] of webs that were spun by spiders under the influence of a variety of psychotropic substances, threshold these grayscale images to turn them into black-and-white images (using a thresholding setting of 205 in the GNU Image Manipulation Program), apply our level-set construction to compute PH, and perform average-linkage hierarchical clustering to

Hierarchical clustering of spiderweb images

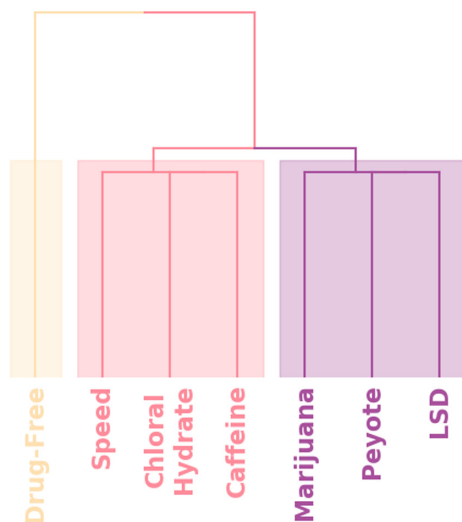


FIG. 18. Classification of webs that were spun by spiders under the influence of various psychotropic substances.

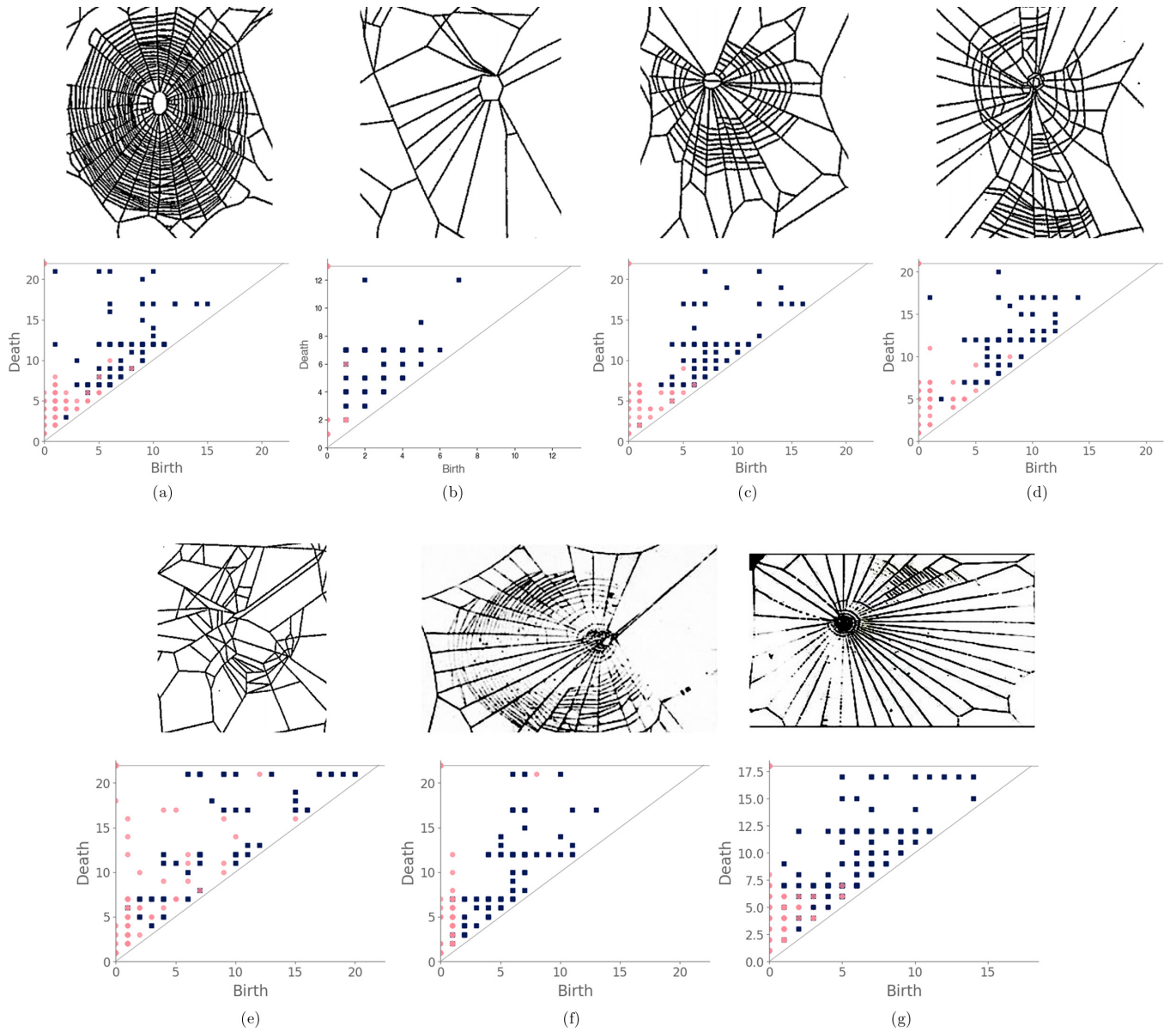


FIG. 19. Webs that were spun by a drug-free spider and spiders that were under the influence of various psychotropic substances, with the associated PD displayed beneath each web. We compare the webs of (a) a drug-free spider with webs that were spun by spiders under the influence of (b) chloral hydrate, (c) marijuana, (d) speed, (e) caffeine, (f) peyote, and (g) LSD. [The images for panels (a)–(e) are from Ref. [89], and the images for panels (f) and (g) are from Ref. [88].]

yield the dendrogram in Fig. 18. We show the images of the spiderwebs and their associated PDs in Fig. 19.

Our classification places the drug-free spider into its own cluster. The spiderweb of the drug-free spider is characterized by a clear central hole, threads that radiate outward at approximately even intervals, and completed rings of threads that surround the center. We place the webs that were spun by spiders under the influence of marijuana, peyote, and LSD into the same cluster. In these webs, there is a clearly identifiable center, and most of the radial threads are evenly spaced, straight, and radiate outward directly from the center. However, for the webs in this cluster, rings of threads are either difficult to see or are incomplete. The final cluster consists of webs that were spun by spiders under the influence

of chloral hydrate, caffeine, and speed. In the caffeinated spider's web, one cannot even clearly identify a center [90]. One can locate a center in the webs of the spiders that were under the influence of speed or chloral hydrate (a sedative that is used in sleeping pills), but many of the radial threads do not join the center and some of the radial threads are not straight. Almost no complete rings of thread are visible in any of the three webs in this cluster.

IV. CONCLUSIONS

It is important to exploit spatial information in the study of spatial complex systems. By using methods for computing persistent homology that take spatial information into

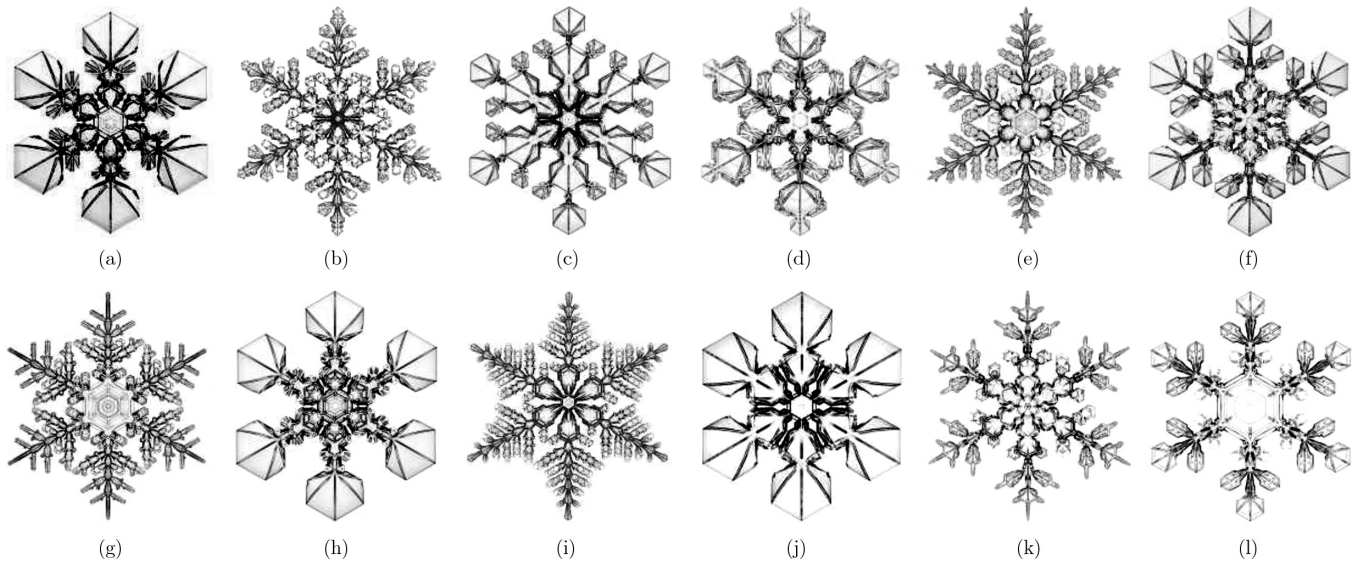


FIG. 20. The full set of 12 snowflake images that we examined in Sec. III C. We label these snowflakes using the panel labels from this figure. We show snowflake A in panel (a), snowflake B in panel (b), and so on. (These images are from Ref. [85].)

account, we presented several applications of topological data analysis to spatial networks. We showed that topological methods are capable of characterizing network structures and detecting structural differences in images of various spatial systems. We also demonstrated, using both synthetic examples and networks from empirical data, that such methods are able to provide insights into large-scale network structures that complement those from traditional techniques of network analysis. As an extended case study, we examined the morphology of street networks in cities, and we used spatial topological data analysis to compare and contrast (1) different regions of the same city and (2) different cities. We hope that our examples help illustrate some ways in which topological methods, especially ones that directly incorporate spatial information in their formulation, can be useful for the analysis of spatial complex systems.

ACKNOWLEDGMENTS

We thank Marc Barthelemy, Heather Zinn Brooks, Hanbaek Lyu, Elizabeth Munch, Stan Osher, Nina Otter, Giovanni Petri, Bernadette Stolz, Katina Vradelis, and an anonymous referee for helpful comments. We are particularly grateful to Joshua Gensler for his many helpful comments on both our paper and our code. We also acknowledge support from the National Science Foundation (Grant No. 1922952) through the Algorithms for Threat Detection (ATD) program.

APPENDIX: ADDITIONAL SNOWFLAKE IMAGES

In Fig. 20, we show the images of all 12 snowflakes that we examined.

- [1] M. Barthelemy, *Morphogenesis of Spatial Networks* (Springer International Publishing, Cham, Switzerland, 2018).
- [2] M. E. J. Newman, *Networks*, 2nd ed. (Oxford University Press, Oxford, UK, 2018).
- [3] R. V. Solé, M. Rosas-Casals, B. Corominas-Murtra, and S. Valverde, *Phys. Rev. E* **77**, 026102 (2008).
- [4] H. Kim, D. Olave-Rojas, E. Álvarez-Miranda, and S.-W. Son, *Sci. Data* **5**, 180209 (2018).
- [5] R. Albert, I. Albert, and G. L. Nakarado, *Phys. Rev. E* **69**, 025103(R) (2004).
- [6] L. Papadopoulos, M. A. Porter, K. E. Daniels, and D. S. Bassett, *J. Complex Networks* **6**, 485 (2018).
- [7] S. H. Lee, M. Cucuringu, and M. A. Porter, *Phys. Rev. E* **89**, 032810 (2014).
- [8] M. Batty, *The New Science of Cities* (MIT Press, Cambridge, MA, USA, 2013).
- [9] M. Barthelemy, *Phys. Rep.* **499**, 1 (2011).
- [10] P. Haggett and R. J. Chorley, *Network Analysis in Geography* (Edward Arnold Publishers, London, UK, 1969).
- [11] D. Pumain, editor, *Theories and Models of Urbanization: Geography, Economics and Computing Sciences* (Springer International Publishing, Cham, Switzerland, 2020).
- [12] A. Liu and M. A. Porter, *Phys. Rev. E* **101**, 062305 (2020).
- [13] S. Nauer, L. Böttcher, and M. A. Porter, *J. Complex Networks* (2019), doi: 10.1093/comnet/cnz037.
- [14] M. Sarzynska, E. A. Leicht, G. Chowell, and M. A. Porter, *J. Complex Networks* **4**, 363 (2016).
- [15] P. Expert, T. S. Evans, V. D. Blondel, and R. Lambiotte, *Proc. Natl. Acad. Sci. U. S. A.* **108**, 7663 (2011).
- [16] D. V. Boulatov and V. A. Kazakov, *Phys. Lett. B* **186**, 379 (1987).
- [17] A. Arenas, A. Díaz-Guilera, J. Kurths, Y. Moreno, and C. Zhou, *Phys. Rep.* **469**, 93 (2008).

- [18] F. Ying, A. O. G. Wallis, M. Beguerisse-Díaz, M. A. Porter, and S. D. Howison, *Phys. Rev. E* **100**, 062304 (2019).
- [19] A. Hatcher, *Algebraic Topology* (Cambridge University Press, Cambridge, UK, 2002).
- [20] H. Edelsbrunner and J. Harer, *Computational Topology: An Introduction* (American Mathematical Society, Providence, RI, USA, 2010).
- [21] M. Feng and M. A. Porter, [arXiv:1902.05911](https://arxiv.org/abs/1902.05911) [SIAM Review (to be published)].
- [22] N. Otter, M. A. Porter, U. Tillmann, P. Grindrod, and H. A. Harrington, *Eur. Phys. J. Data Sci.* **6**, 17 (2017).
- [23] T. Kaczynski, K. M. Mischaikow, and M. Mrozek, *Computational Homology* (Springer-Verlag, Heidelberg, Germany, 2004).
- [24] M. Gameiro, Y. Hiraoka, S. Izumi, M. Kramar, K. Mischaikow, and V. Nanda, *Jpn. J. Ind. Appl. Math.* **32**, 1 (2015).
- [25] K. Xia and G.-W. Wei, *Intl. J. Numer. Methods Biomed. Eng.* **30**, 814 (2014).
- [26] V. Kovacev-Nikolic, P. Bubenik, D. Nikolic, and G. Heo, *Stat. Appl. Genet. Mol. Biol.* **15**, 19 (2016).
- [27] W. Zhou and H. Yan, *Briefings Bioinf.* **15**, 54 (2014).
- [28] K. Emmett, B. Schweinhart, and R. Rabadan, in *Proceedings of the 9th EAI International Conference on Bio-inspired Information and Communications Technologies (Formerly BIONET-ICS)*, BICT '15 ICST (Institute for Computer Sciences, Social-Informatics and Telecommunications Engineering, Brussels, Belgium, 2016), pp. 177–180.
- [29] L. Kanari, P. Dlotko, M. Scolamiero, R. Levi, J. Shillcock, K. Hess, and H. Markram, *Neuroinformatics* **16**, 3 (2018).
- [30] G. Carlsson, T. Ishkhanov, V. de Silva, and A. Zomorodian, *Intl. J. Comput. Vision* **76**, 1 (2008).
- [31] S. Tymochko, E. Munch, J. Dunion, K. Corbosiero, and R. Torn, *Pattern Recognition Letters* **133**, 137 (2020).
- [32] G. Yalnız and N. B. Budanur, *Chaos* **30**, 033109 (2020).
- [33] L. Speidel, H. A. Harrington, S. J. Chapman, and M. A. Porter, *Phys. Rev. E* **98**, 012318 (2018).
- [34] H. Adams, T. Emerson, M. Kirby, R. Neville, C. Peterson, P. Shipman, S. Chepushtanova, E. Hanson, F. Motta, and L. Ziegelmeier, *J. Mach. Learn. Res.* **18**(8), 1 (2017).
- [35] F. A. Khasawneh, E. Munch, and J. A. Perea, *IFAC-PapersOnLine* **51**, 195 (2018), part of the special issue of the 14th IFAC Workshop on Time Delay Systems TDS 2018 (Budapest, Hungary, 28–30 June 2018).
- [36] M. C. Yesilli, S. Tymochko, F. A. Khasawneh, and E. Munch, in *18th IEEE International Conference on Machine Learning And Applications (ICMLA)*, Boca Raton, FL, USA (2019), pp. 1211–1218, doi: [10.1109/ICMLA.2019.00200](https://doi.org/10.1109/ICMLA.2019.00200).
- [37] C. Cai and Y. Wang, [arXiv:2001.06058](https://arxiv.org/abs/2001.06058) (2020).
- [38] M. Kramár, A. Goulet, L. Kondic, and K. Mischaikow, *Phys. Rev. E* **87**, 042207 (2013).
- [39] M. Buchet, Y. Hiraoka, and I. Obayashi, in *Nanoinformatics*, edited by I. Tanaka (Springer-Verlag, Heidelberg, Germany, 2018), pp. 75–95.
- [40] H. Ronellenfitsch and E. Katifori, *Phys. Rev. Lett.* **117**, 138301 (2016).
- [41] C. M. Topaz, L. Ziegelmeier, and T. Halverson, *PLoS ONE* **10**, e0126383 (2015).
- [42] P. S. P. Ignacio and I. K. Darcy, *Eur. Phys. J. Data Sci.* **8**, 1 (2019).
- [43] H. M. Byrne, H. A. Harrington, R. Muschel, G. Reinert, B. J. Stolz, and U. Tillmann, *Math. Today* **55**, 206 (2019).
- [44] G. Petri, P. Expert, F. Turkheimer, R. Carhart-Harris, D. Nutt, P. J. Hellyer, and F. Vaccarino, *J. R. Soc. Interface* **11**, 20140873 (2014).
- [45] F. Battiston, G. Cencetti, I. Iacopini, V. Latora, M. Lucas, A. Patania, J.-G. Young, and G. Petri, *Phys. Rep.* (2020), doi: [10.1016/j.physrep.2020.05.004](https://doi.org/10.1016/j.physrep.2020.05.004).
- [46] P. Bubenik, M. Hull, D. Patel, and B. Whittle, *Inverse Probl.* **36**, 025008 (2020).
- [47] W. R. Tobler, *Geograph. Rev.* **53**, 59 (1963).
- [48] M. T. Gastner and M. E. J. Newman, *Proc. Natl. Acad. Sci. USA* **101**, 7499 (2004).
- [49] M. Feng, Spatial TDA, available at <https://bitbucket.org/mhfeng/spatialtda/src/master/>.
- [50] A. Zomorodian and G. Carlsson, *Discrete Comput. Geometry* **33**, 249 (2005).
- [51] R. Ghrist, *Bull. Am. Math. Soc.* **45**, 61 (2008).
- [52] B. J. Stolz, H. A. Harrington, and M. A. Porter, *Chaos* **27**, 047410 (2017).
- [53] The GUDHI Project, *Gudhi User and Reference Manual* (GUDHI Editorial Board, 2015), Version 3.0.0, available at <http://gudhi.gforge.inria.fr/doc/latest/>.
- [54] C. Maria, in *GUDHI User and Reference Manual* (GUDHI Editorial Board, 2015), Version 3.0.0, available at http://gudhi.gforge.inria.fr/doc/latest/group__simplex__tree.html.
- [55] S. Osher and R. Fedkiw, *Level Set Methods and Dynamic Implicit Surfaces*, Applied Mathematical Sciences, Vol. 153 (Springer-Verlag, Heidelberg, Germany, 2003).
- [56] D. J. Watts, *Proc. Natl. Acad. Sci. U. S. A.* **99**, 5766 (2002).
- [57] M. A. Porter and J. P. Gleeson, *Dynamical Systems on Networks: A Tutorial*, Frontiers in Applied Dynamical Systems: Reviews and Tutorials, Vol. 4 (Springer International Publishing, Cham, Switzerland, 2016).
- [58] We use the convention that $\mathbb{N}$ includes 0.
- [59] D. Taylor, F. Klimm, H. A. Harrington, M. Kramár, K. Mischaikow, M. A. Porter, and P. J. Mucha, *Nat. Commun.* **6**, 7723 (2015).
- [60] B. I. Mahler, U. Tillmann, and M. A. Porter, [arXiv:1812.09806](https://arxiv.org/abs/1812.09806) [SIAM Journal on Applied Mathematics (to be published)].
- [61] F. M. Ying, Dynamical processes on random geometric graphs (2013), available at <https://www.math.ucla.edu/~mason/research/fabian-report-092913.pdf>.
- [62] M. Penrose, *Random Geometric Graphs* (Oxford University Press, Oxford, UK, 2003).
- [63] D. J. Watts and S. H. Strogatz, *Nature (London)* **393**, 440 (1998).
- [64] M. A. Porter, *Scholarpedia* **7**, 1739 (2012).
- [65] A. A. Hagberg, D. A. Schult, and P. J. Swart, in *Proceedings of the 7th Python in Science Conference*, edited by G. Varoquaux, T. Vaught, and J. Millman (Pasadena, CA, 2008), pp. 11–15, available at http://conference.scipy.org/proceedings/scipy2008/SciPy2008_proceedings.pdf.
- [66] M. Barthelemy, *Nat. Rev. Phys.* **1**, 406 (2019).
- [67] G. Boeing, *Intl. J. Inf. Manage.* (2019), doi: [10.1016/j.ijinfomgt.2019.09.009](https://doi.org/10.1016/j.ijinfomgt.2019.09.009).
- [68] G. Boeing, *Environ. Plann. B* **47**, 590 (2020).
- [69] M. Barthelemy, *Environ. Plann. B* **44**, 256 (2017).

- [70] A. Cardillo, S. Scellato, V. Latora, and S. Porta, *Phys. Rev. E* **73**, 066107 (2006).
- [71] M. Ahmed, B. T. Fasy, and C. Wenk, in *Proceedings of the 22nd ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems*, SIGSPATIAL '14 (Association for Computing Machinery, New York, NY, USA, 2014), pp. 43–52.
- [72] Y. Wu, G. Shindnes, V. Karve, D. Yager, D. B. Work, A. Chakraborty, and R. B. Sowers, in *IEEE 20th International Conference on Intelligent Transportation Systems (ITSC)*, Yokohama, Japan, 2017 (Institute of Electrical and Electronics Engineers, 2017), pp. 1–6, doi: [10.1109/ITSC.2017.8317777](https://doi.org/10.1109/ITSC.2017.8317777).
- [73] J. Thompson, M. Stevenson, J. S. Wijnands, K. A. Nice, G. D. P. A. Aschwanden, J. Silver, M. Nieuwenhuijsen, P. Rayner, R. Schofield, R. Hariharan, and C. N. Morrison, *The Lancet: Planetary Health* **4**, E32 (2020).
- [74] G. Boeing, *Comput Environ. Urban Syst.* **65**, 126 (2017).
- [75] E. Song, Administrative district boundaries of city of Shanghai, People's Republic of China, 2017, ARCGIS (2017), available at <https://www.arcgis.com/home/item.html?id=105f92bd1fe54d428bea35eade65691b>.
- [76] Google, Google Maps search for Shanghai, available at <https://www.google.com/maps/place/Shanghai,+China/data=!4m2!3m1!1s0x35b27040b1f53c33:0x295129423c364a1?sa=X&ved=2ahUKEwjSmom9nevmaAhXNuZ4KHangDhIQ8gEwK3oECBkQBA>.
- [77] M. Kerber, D. Morozov, and A. Nigmatov, *J. Exp. Algorithm.* **22**, 1.4 (2017).
- [78] H.-S. Park and C.-H. Jun, *Expert Syst. Appl.* **36**, 3336 (2009).
- [79] SimpleMaps, World cities database (2019), available at <https://simplemaps.com/data/world-cities>.
- [80] M. Belgiu, UIA_Latitude/Longitude Graticules and World Countries Boundaries, ARCGIS (2015), available at <https://www.arcgis.com/home/item.html?id=a21fdb46d23e4ef896f31475217cbb08>.
- [81] R. Louf and M. Barthelemy, *J. R. Soc. Interface* **11**, 20140924 (2014).
- [82] R. Gillman, *Math Horizons* **10**, 10 (2002).
- [83] R. Barnes and J. Solomon, *arXiv:1803.02857* (2018).
- [84] M. Duchin and B. E. Tenner, *arXiv:1808.05860* (2018).
- [85] K. G. Libbrecht, *arXiv:1910.06389* (2019).
- [86] The GIMP Development Team, GNU Image Manipulation Program (GIMP), available at <https://www.gimp.org>.
- [87] Interestingly, whiskey itself produces webs [91].
- [88] P. N. Witt, *Behav. Sci.* **16**, 98 (1971).
- [89] D. A. Noever, R. J. Cronise, and R. A. Relwani, *NASA Tech Briefs* **19**, 82 (1995).
- [90] The web that was produced by the caffeinated spider is always fun to point out when giving presentations.
- [91] S. J. Williams, M. J. Brown, and A. D. Carrithers, *Phys. Rev. Fluids* **4**, 100511 (2019).

Erratum: Spatial applications of topological data analysis: Cities, snowflakes, random structures, and spiders spinning under the influence [Phys. Rev. Research 2, 033426 (2020)]

Michelle Feng  and Mason A. Porter 

(Received 23 April 2021; published 18 May 2021)

DOI: [10.1103/PhysRevResearch.3.029003](https://doi.org/10.1103/PhysRevResearch.3.029003)

Synthetic networks. In Sec. III A of our paper, the image in Fig. 5 is not the one that we intended. We show the correct image in Fig. 1 of this erratum. The image in Fig. 5 in the paper is an example that uses the same dynamical process and family of networks, but it is for an instantiation of the process with $\rho = 0.1$. The correct version of Fig. 5 is from a simulation with $\rho = 0.05$ and corresponds to the example that we showed in Fig. 4(a).

Additionally, in Fig. 4, the specific numerical values that are next to the color bar apply only to panel (c). The qualitative meaning of the colors applies to all panels.

Finally, we clarify the following sentence about Watts–Strogatz (WS) networks: “In Fig. 7, we see that a WS small-world network also eventually has an infection network that consists of a single connected component.” This sentence describes a true feature of our simulation, but the words “we see” imply that one can necessarily see this in the plot without ambiguity, and that is not the case.

Clarification of our procedure for clustering persistence diagrams (PDs). In our hierarchical clustering of PDs, we calculate pairwise distances between PDs of H_1 features only. This is the default choice in the software package GUDHI.

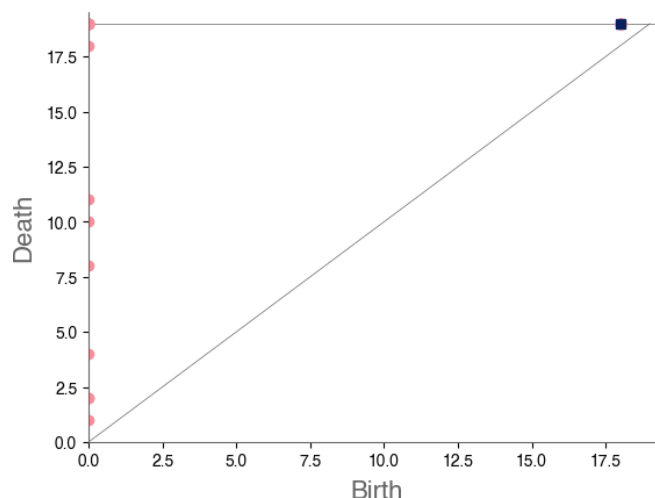


FIG. 1. The persistence diagram (PD) of an instance of the Watts threshold model on a random geometric graph. This figure corrects Fig. 5 of our paper.