A Complete Characterization of the 1-Dimensional Intrinsic Čech Persistence Diagrams for Metric Graphs

Ellen Gasparovic, Maria Gommel, Emilie Purvine, Radmila Sazdanovic, Bei Wang, Yusu Wang, and Lori Ziegelmeier

Abstract Metric graphs are special types of metric spaces used to model and represent simple, ubiquitous, geometric relations in data such as biological networks, social networks, and road networks. We are interested in giving a qualitative description of metric graphs using topological summaries. In particular, we provide a complete characterization of the 1-dimensional intrinsic Čech persistence diagrams for metric graphs using persistent homology. Together with complementary results by Adamaszek et. al, which imply results on intrinsic Čech persistence diagrams in all dimensions for a single cycle, our results constitute important steps toward characterizing intrinsic Čech persistence diagrams for arbitrary metric graphs across all dimensions.

Ellen Gasparovic
Union College, 807 Union Street, Schenectady, NY 12309, e-mail: gasparoe@union.edu

Maria Gommel
University of Iowa, 14 MacLean Hall, Iowa City, IA 52242, e-mail: maria-gommel@uiowa.edu

Emilie Purvine
Pacific Northwest National Laboratory, 1100 Dexter Ave N., Seattle, WA 98109,
e-mail: emilie.purvine@pnnl.gov

Radmila Sazdanovic
NC State University, PO Box 8205, Raleigh NC 27695,
e-mail: rsazdan@ncsu.edu

Bei Wang
University of Utah, 72 South Central Campus Drive, Salt Lake City, UT 84112,
e-mail: beiwang@sci.utah.edu

Yusu Wang
The Ohio State University, 2015 Neil Ave., Columbus, OH 43210, e-mail: yusu@cse.ohio-state.edu

Lori Ziegelmeier
Macalester College, 1600 Grand Avenue, Saint Paul, MN 55104, e-mail: lziegel1@macalester.edu

1 Introduction

Graphs are ubiquitous in data analysis, often used to model social, biological and technological systems. Often, data with a notion of distance can be modeled by a metric graph. A graph is a *metric graph* if each edge is assigned a positive length and if the graph is equipped with a natural metric where the distance between any two points of the graph (not necessarily vertices) is defined to be the minimum path length between them [13]. A metric graph is therefore a special type of metric space that captures simple forms of geometric relations in data that arise in both abstract and practical settings, such as biological networks, social networks and road networks. For example, the movement patterns that GPS systems trace for vehicles can be modeled as a metric graph for location-aware applications. Brain functional networks as metric graphs capture the blood-oxygen-level dependent signal correlations among different areas of the brain [4]. Social networks as metric graphs can encode strengths of influence between social entities (e.g., persons or corporations). Extracting the topological structures of such networks can provide powerful insights for navigating and understanding their underlying data.

Our work aims to describe topological structures of metric graphs by using persistent homology, a fundamental tool in topological data analysis that has been used in many applications to measure topological features of shapes and functions [10]. In this work, we give a qualitative description of information that can be captured from metric graphs using topological, persistence-based summaries. Theorem 1.1, the main theorem in this paper, provides a complete characterization of the persistence diagrams in dimension 1 for metric graphs in a particular intrinsic setting.

Theorem 1.1 *Let G be a metric graph of genus g with shortest cycle basis $\{\gamma_1, \ldots, \gamma_g\}$, and for each $i = 1, \ldots, g$, let $|\gamma_i| = \ell_i$ be the length of the i^{th} cycle, with $\ell_i \leq \ell_j$ for all $i \leq j$. Then the 1-dimensional intrinsic Čech persistence diagram of G, denoted $\mathrm{Dg}_1 IC_G$, consists of the following collection of points on the y-axis:*

$$\mathrm{Dg}_1 IC_G = \left\{ \left(0, \frac{\ell_i}{4}\right) : 1 \leq i \leq g \right\}.$$

Related Work. The work of Adamaszek et. al [3] is most relevant to ours, as it helps to characterize persistence diagrams in all dimensions for a metric graph with a single cycle. In [3], the authors show that the intrinsic Vietoris-Rips or Čech complex of n points in the circle $\mathbb{S}^1$, at any scale r, is homotopy equivalent to either a point, an odd-dimensional sphere, or a wedge sum of spheres of the same even dimension. The results in [3] further imply that the 1-dimensional homology group of a metric graph with a single cycle is either rank 1 (in the case where the associated intrinsic complex is homotopy equivalent to $\mathbb{S}^1$) or rank 0 (in all other cases). One can then show that the 1-dimensional persistence diagram consists of the single point $\left(0, \frac{\ell}{4}\right)$ or $\left(0, \frac{\ell}{6}\right)$ in the case of the Čech or Vietoris-Rips filtration, respectively, where ℓ is the length of the cycle [2].

In this paper, we generalize the above result from [3] from a metric graph with a single cycle to a metric graph containing an arbitrary set of cycles in homological dimension 1. This characterization of persistence diagrams in dimension 1 of an arbitrary metric graph complements the work in [3] and constitutes an important step toward the characterization of the intrinsic Čech persistence diagrams of *arbitrary* metric graphs across *all* dimensions.

In addition to the Čech and Vietoris-Rips complexes, there are a number of other types of complexes or combinatorial structures related to graphs. In [16], the author studies the relationship between properties of a graph G and the homology of an associated neighborhood complex. The paper [18] contains a study of so-called devoid complexes of graphs where simplices correspond to vertex sets whose induced subgraphs do not contain certain forbidden subgraphs. However, the neighborhood and devoid complexes are more related to structural, rather than metric, properties of graphs, so we turn our attention in the remainder of this paper to the more metric-derived Čech complex.

Outline. The outline of the paper is as follows. In Section 2, we recall the necessary background on persistent homology, in particular for the case that the underlying topological space is a metric graph. We prove Theorem 1.1 in Section 3. Finally, we discuss our results and plans for future work in Section 4.

2 Background

2.1 Homology

Homology is an invariant that characterizes properties of a topological space X. In particular, the k-dimensional holes (connected components, loops, trapped volumes, etc.) of a space generate a homology group, $H_k(X)$. The rank of this group is referred to as the *k-th Betti number* β_k and counts the number of k-dimensional holes of X. We provide a brief overview of simplicial homology below. For a comprehensive study, see [12, 15]. For a more categorical viewpoint, see [17], and for a discussion of cubical complexes, see [14].

A *simplicial complex* S is a set consisting of a finite collection of k-simplices where a 0-simplex is a vertex, a 1-simplex is an edge, a 2-simplex is a filled-in triangle, a 3-simplex is a solid tetrahedron, and so on. The k-simplices must satisfy the following: (1) if σ is a simplex in S, then all lower-dimensional subsets of σ, called *subsimplices*, are also in S, and (2) two simplices are either disjoint or intersect in a lower-dimensional simplex.

An algebraic structure of a *vector space* or an *R-module* over some ring R is imposed on the simplicial complex S to uncover the homology of the underlying topological space as follows. The k-simplices form a basis for a vector space, $S^{(k)}$, over some ground field (or ring) $\mathbb{F}$. We call the vector space $S^{(k)}$ the *k-dimensional chain* *group* over simplicial complex S. The finite field $\mathbb{Z}_p$ (where p is a small prime), $\mathbb{Z}$ or $\mathbb{Q}$ are common choices for the ground field or ring. Furthermore, for each pair of consecutive vector spaces there is a linear map, $\delta_k : S^{(k)} \to S^{(k-1)}$, turning the sequence of chain groups into a *chain complex*:

$$\cdots \to S^{(k+1)} \xrightarrow{\delta_{k+1}} S^{(k)} \xrightarrow{\delta_k} S^{(k-1)} \cdots .$$

These maps are known as *boundary operators*, taking each k-simplex to an alternating sum of its $(k-1)$-subsimplices, its boundary. More precisely, if $[v_0, v_1, \ldots, v_k]$ is a k-simplex, the boundary map $\delta_k : S^{(k)} \to S^{(k-1)}$ is defined by

$$\delta_k([v_0, v_1, \ldots, v_k]) = \sum_{i=0}^{k} (-1)^i [v_0, \ldots, \hat{v}_i, \ldots, v_k]$$

where $[v_0, \ldots, \hat{v}_i, \ldots, v_k]$ is the $(k-1)$-simplex obtained from $[v_0, \ldots, v_k]$ by removing vertex v_i.

The simplicial homology, $H_k(S)$, of a simplicial complex S is defined based on two subspaces of the vector space $S^{(k)}$: $Z_k = ker(\delta_k)$ known as *k-cycles*, and $B_k = im(\delta_{k+1}) = \delta_{k+1}(S^{(k+1)})$ known as *k-boundaries*. Since the boundary operator satisfies the property $\delta_k \circ \delta_{k+1} = 0$ for every $0 \leq k \leq \dim(S)$, the set of k-boundaries is contained in the set of k-cycles. Then, $H_k(S) = Z_k / B_k$ consists of the equivalence classes of $k-$cycles that are not $k+1$-boundaries (up to *homotopy*). The elements of $H_k(S)$ are called homology classes and can thus be thought of as equivalence classes represented by cycles enclosing kth order holes that differ by elements of a boundary. The rank of the associated homology group $H_k(S)$ is the number of distinct k dimensional holes, and is referred to as the kth Betti number, denoted β_k.

2.2 *Persistent homology and metric graphs*

In *persistent homology*, rather than studying the topological structure of a single space, X, one considers how the homology changes over an increasing sequence of subspaces. Given a topological space X and a filtration of this space,

$$X_1 \subseteq X_2 \subseteq X_3 \subseteq \ldots \subseteq X_m = X,$$

applying the homology functor gives a sequence of homology groups induced by inclusion of the filtration

$$H_k(X_1) \to H_k(X_2) \to \ldots \to H_k(X_m).$$

A filtration of a topological space may be defined in a number of ways. By considering a continuous function (such as a height function) on the topological space $f : X \to \mathbb{R}$, one may define the *sublevel set filtration*

$$f^{-1}(-\infty,a_0) \to f^{-1}(-\infty,a_1) \to \ldots \to f^{-1}(-\infty,\infty).$$

Another approach is to build a sequence of simplicial complexes on a set of points using, for instance, the *Vietoris-Rips filtration* [11] or the *intrinsic Čech filtration* [6] discussed below. *Persistent homology* [5, 10] then tracks elements of each homology group through the filtration. This information may be displayed in a *persistence diagram* for each homological dimension k. A persistence diagram is a set of points in the plane together with an infinite number of points along the diagonal where each point (x,y) corresponds to a homological element that appears (is 'born') at $H_k(X_x)$ and which no longer remains ('dies') at $H_k(X_y)$. See Figure 1 for an example persistence diagram. Notice that distinct topological features may have the same birth and death coordinates; therefore, a persistence diagram is actually a multiset of points. Since all topological features die after they are born, this is an embedding into the upper half plane above the diagonal $y = x$. Points near the diagonal are often considered noise while those further from the diagonal represent more robust topological features. For a more detailed description of applications of persistent homology to various problems in the experimental sciences, see [5, 8, 11].

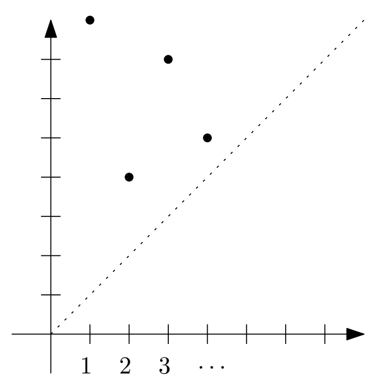

Fig. 1 An example persistence diagram with four points: (1, 8), (2, 4), (3, 7), and (4, 5) corresponding to the birth and death values for distinct topological features.

In this paper, we focus on understanding the topological structure of a *metric graph* in homology dimension $k = 1$. Given a simple graph $G = (V,E)$ we define a metric graph to be a metric space $(|G|,d_G)$ that is homeomorphic to a 1-dimensional finite stratified space consisting of 0-dimensional pieces (i.e. vertices) and 1-dimensional pieces (i.e. edges or loops) glued together, as described in [1, 9]. More formally, any graph G with vertex set V and edge set E, together with a length function, $\mathrm{Len} : E \to \mathbb{R}_{\geq 0}$, on E that assigns lengths to edges in E, gives rise to a metric graph $(|G|,d_G)$ where $|G|$ is a geometric realization of G and d_G is defined in the following manner. Using the notation of [9], let e denote an edge in E with $|e|$ its image in $|G|$, let $e : [0, Len(e)] \to |e|$ be the arclength parametrization, and define $d_G(x,y) = |e^{-1}(y) - e^{-1}(x)|$ for any $x,y \in |e|$. This enables one to define the length of any given path between two points in $|G|$ by first restricting the path to edges in G and then summing the lengths. Then one may define the distance $d_G(u,v)$ between any pair of points $u,v \in |G|$ to be the minimum length of any path in $|G|$ between u and v.

We consider a simplicial complex built on a metric graph as follows. Let (G,d_G) be a metric graph with geometric realization $|G|$. For any point $x \in |G|$, we define the set $B(x,\varepsilon) := \{y \in |G| : d_G(x,y) \leq \varepsilon\}$, and we let $U_\varepsilon := \{B(x,\varepsilon) : x \in |G|\}$ be an open cover. The *nerve* of a family of sets $(Y_i)_{i\in I}$ is the abstract simplicial complex defined on the vertex set I by the rule that a finite set $\sigma \subseteq I$ is in the nerve if and only if $\bigcap_{i\in\sigma} Y_i \neq \emptyset$. We let C_ε denote the nerve of U_ε. The associated *intrinsic Čech filtration* is defined as the set of inclusion maps

$$\{C_\varepsilon \hookrightarrow C_{\varepsilon'}\}_{\forall 0\leq\varepsilon\leq\varepsilon'}.$$

The intrinsic Čech filtration on the metric graph G induces the persistence module

$$\{H_*(C_\varepsilon) \to H_*(C_{\varepsilon'})\}_{\forall 0\leq\varepsilon\leq\varepsilon'}$$

in any dimension, from which we obtain the *intrinsic Čech persistence diagrams*, denoted by $\mathrm{Dg}_* IC_G$. In this paper, we shall only be interested in $\mathrm{Dg}_1 IC_G$.

3 Proof of Main Theorem

In this section, we prove Theorem 1.1 by working with certain mappings on the level of chain groups and their induced mappings on the level of homology groups. These mappings possess important properties from which the statement of the theorem follows.

Proof (Theorem 1.1). Let G be a metric graph of genus g with shortest cycle basis given by $\{\gamma_1,\ldots,\gamma_g\}$, where for each $i = 1,\ldots,g$, $|\gamma_i| = \ell_i$ is the length of the i^{th} cycle, and $\ell_i \leq \ell_j$ for all $i \leq j$. Let C_δ, for a sufficiently small positive value of δ, be the Čech complex which is equivalent to the graph G. For $\varepsilon > \delta$, we consider the chain map $\mu_\varepsilon : C_\delta \to C_\varepsilon$ given by inclusion, and the associated inclusion map $\mu^c_\varepsilon : C^{(1)}_\delta \to C^{(1)}_\varepsilon$ of one dimensional chain groups. The latter induces the map on one dimensional homology $\mu^h_\varepsilon : H^{(1)}_\delta \to H^{(1)}_\varepsilon$, where $H^{(1)}_\delta = H_1(C^{(1)}_\delta)$ and $H^{(1)}_\varepsilon = H_1(C^{(1)}_\varepsilon)$.

First, note that each of the g cycles in G must have been born at δ (≈ 0) since the overlap of δ-balls in U_δ will create a cycle in the associated nerve complex. Furthermore, γ_i will be fully triangulated in C_ε for $\varepsilon = \dfrac{\ell_i}{4}$. This is due to the fact that, for any triple of points $x,y,z \in \gamma_i$, $B\left(x,\dfrac{\ell_i}{4}\right)\bigcap B\left(y,\dfrac{\ell_i}{4}\right)\bigcap B\left(z,\dfrac{\ell_i}{4}\right) \neq \emptyset$. Therefore γ_i must die at $\dfrac{\ell_i}{4}$ or earlier. The rest of the proof consists of showing that:

A) For $i = 1, \dots, g$, the i^{th} cycle does not die before $\varepsilon = \dfrac{\ell_i}{4}$; and
B) No other cycles are created due to interference between cycles.

Notice that A) and B) can be reformulated to the language of bases, where condition A) is equivalent to a linear independence condition, and B) is equivalent to a spanning condition. Therefore, the proof of Theorem 1.1 follows from Proposition 3.1.

Proposition 3.1 *For any $i = 1, \dots, g$, the set*

$$\{[\mu_\varepsilon^c(\gamma_i)], [\mu_\varepsilon^c(\gamma_{i+1})], \dots, [\mu_\varepsilon^c(\gamma_g)]\}$$

is a basis for $H_\varepsilon^{(1)}$ where $\dfrac{\ell_{i-1}}{4} \le \varepsilon < \dfrac{\ell_i}{4}$ and $\ell_0 = 0$.

Proof (Proposition 3.1). We will prove the two conditions A) and B).

For A), we show that $\sum_{j=i}^{g} c_j[\mu_\varepsilon^c(\gamma_j)] = [0]$ implies $c_j = 0$ for all j. Let $\gamma = \sum_{j=i}^{g} c_j \mu_\varepsilon^c(\gamma_j)$ be a cycle representing the trivial class $[0] = [\gamma] \in H_\varepsilon^{(1)}$. Assume, by way of contradiction, that there exists j with $i \le j \le g$ such that $c_j \neq 0$. Since $[\gamma] = [0]$, there exists a 2-dimensional chain $\alpha \in C_\varepsilon$ having γ as its boundary, i.e., $\partial\alpha = \gamma$. Let $\alpha = \sum_{k \in J} \Delta_k$ where, for some index set J, $\{\Delta_k\}_{k \in J}$ is the set of 2-simplices in the triangulation of α, and where for each k, $t_k := \partial\Delta_k \in C_\varepsilon^{(1)}$. Then $\gamma = \partial\alpha = \partial\sum_k \Delta_k = \sum_k \partial\Delta_k = \sum_k t_k$, i.e.

$$\gamma = \sum_{j=i}^{g} c_j \mu_\varepsilon^c(\gamma_j) = \sum_k t_k. \tag{1}$$

We aim to contradict the fact that some $c_j \neq 0$ in the above sum. To this end, we define a map $\rho : C_\varepsilon^{(1)} \to C_\delta^{(1)}$ by specifying its effect on a basis for $C_\varepsilon^{(1)}$, the edges in the Čech complex C_ε, and extending the map linearly to the rest of $C_\varepsilon^{(1)}$. Recall that the existence of the edge $[u,v] \in C_\varepsilon^{(1)}$ implies, by definition, that $B(u,\varepsilon) \cap B(v,\varepsilon) \neq \emptyset$. Thus, we define $\rho([u,v])$ to be a shortest path in $C_\delta^{(1)}$ contained within $B(u,\varepsilon) \cup B(v,\varepsilon)$ which passes through this nontrivial intersection.

Notice that the restriction $\rho|_{C_\delta^{(1)}} : C_\delta^{(1)} \to C_\delta^{(1)}$ is the identity mapping. Clearly $\rho|_{C_\delta^{(1)}}$ is the identity on the basis elements, the edges in the Čech complex $C_\delta^{(1)}$, since there is no shorter path within $C_\delta^{(1)}$ than the edge itself. Then, by linearity, $\rho|_{C_\delta^{(1)}}$ is the identity on all of $C_\delta^{(1)}$. Additionally, $\rho(\mu_\varepsilon^c(\gamma_j)) = \rho(\gamma_j) = \gamma_j$ since $\mu_\varepsilon^c(\gamma_j) = \gamma_j \in C_\varepsilon^{(1)}$. Applying ρ to equation (1) we obtain the following:

$$\rho(\gamma) = \sum_{j=i}^{g} c_j \gamma_j = \sum_k \rho(t_k). \tag{2}$$

Next, we show that for each k, $\rho(t_k)$ is the sum of short cycles. Notice that $t_k = [w_0^k, w_1^k, w_2^k] = \partial \Delta_k$ represents a trivial cycle in $C_\varepsilon^{(1)}$, so there must exist some point $w^k \in \bigcap_{n=0}^{2} B(w_n^k, \varepsilon)$. See Figure 2.

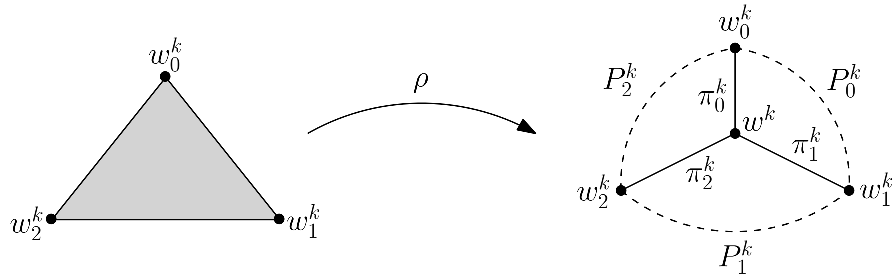

Fig. 2 Action of ρ on the triangle $t_k = [w_0^k, w_1^k, w_2^k]$. Notice the three cycles contained in $\rho(t_k)$.

Consequently, for $n = 0, 1, 2$, there exist the following paths π_n^k and P_n^k in the Cech complex $C_\delta^{(1)}$:

- $\pi_n^k = [w^k, w_n^k]$ which has length less than or equal to ε, and
- $P_n^k = \rho([w_n^k, w_{(n+1 \mod 3)}^k])$, of length at most 2ε, since each P_n^k must be contained within the intersection of two ε-balls.

Therefore, $\rho(t_k) = \sum_{n=0}^{2} \pi_n^k + P_n^k - \pi_{(n+1 \mod 3)}^k$ is the sum of three cycles, each of length at most $\varepsilon + \varepsilon + 2\varepsilon = 4\varepsilon$. Since the length of $\rho(t_k)$ is less than $4\varepsilon < \ell_i$, $\rho(t_k)$ can be expressed in terms of the shortest cycle basis $\{\gamma_j\}_{j=1}^{i-1}$ for $C_\delta^{(1)}$:

$$\rho(\gamma) = \sum_k \rho(t_k) = \sum_k \sum_{j=1}^{i-1} c_j^k \gamma_j = \sum_{j=1}^{i-1} c'_j \gamma_j. \tag{3}$$

$$\implies \sum_{j=i}^{g} c_j \gamma_j \overset{(2)}{=} \sum_{j=1}^{i-1} c'_j \gamma_j \tag{4}$$

$$\implies \sum_{j=1}^{i-1} c_j \gamma_j + \sum_{j=i}^{g} (-c'_j) \gamma_j = 0. \tag{5}$$

As the set $\{\gamma_i\}_{i=1}^{g}$ is a basis for $C_\delta^{(1)}$, the coefficients in the above sums must all be zero, that is $c_j = 0$ for all j, which contradicts our initial assumption. Therefore, the set $\{[\mu_\varepsilon^c(\gamma_i)], [\mu_\varepsilon^c(\gamma_{i+1})], \ldots, [\mu_\varepsilon^c(\gamma_g)]\}$ is linearly independent in $H_\varepsilon^{(1)}$. In particular, γ_i does not become trivial before $\frac{\ell_i}{4}$.

Next, to prove B), we show that the map $\mu_\varepsilon^h : H_\delta^{(1)} \to H_\varepsilon^{(1)}$ is surjective by showing that it has a right inverse up to homotopy. In other words, we will show that for every $[\eta] \in H_\varepsilon^{(1)}$,

$$\mu_\varepsilon^h([\rho(\eta)]) = [(\mu_\varepsilon^c \circ \rho)(\eta)] = [\eta] \in H_\varepsilon^{(1)} \tag{6}$$

where the chain $\eta \in C_\varepsilon^{(1)}$ is a geometric realization of the class $[\eta]$.

Consider a cycle $\eta = \{u_0, u_1, \ldots, u_k, u_0\} \in C_\varepsilon^{(1)}$ representing $[\eta] \in H_\varepsilon^{(1)}$. Let $p_j = \rho([u_j, u_{j+1}]) = \{u_j, v_1^j, \ldots, v_{m_j}^j, u_{j+1}\}$ be a shortest path between u_j and u_{j+1} passing through $B(u_j, \varepsilon) \cap B(u_{j+1}, \varepsilon)$, for $j = 0, 1, \ldots, k$ and $u_{k+1} = u_0$. Then the image $\rho(\eta)$ is just a concatenation of these paths $\rho(\eta) = p_0 + p_1 + \cdots + p_k \in C_\delta^{(1)}$.

To show equation (6) holds, we need to prove that $[\rho(\eta)] = [\eta]$ by showing that p_j is path homotopic to the path determined by $\{u_j, u_{j+1}\}$ for all $j = 0, 1, \ldots, k$ and $u_{k+1} = u_0$.

Since $B(u_j, \varepsilon) \cap B(u_{j+1}, \varepsilon) \neq \emptyset$, there exists some point $v_l^j \in p_j$ such that $v_l^j \in B(u_j, \varepsilon) \cap B(u_{j+1}, \varepsilon)$. Let $p_j^{(1)}$ be the path from u_j to v_l^j and $p_j^{(2)}$ be the path from v_l^j to u_{j+1} (see Figure 3). Because these are shortest paths and v_l^j is in the intersection of $B(u_j, \varepsilon)$ and $B(u_{j+1}, \varepsilon)$, it follows that $p_j^{(1)} \subseteq B(u_j, \varepsilon)$ and $p_j^{(2)} \subseteq B(u_{j+1}, \varepsilon)$. Therefore, each of the following is a 2-dimensional cell: $\Delta(u_j, v_n^j, v_{n+1}^j) \in C_\varepsilon$ for all $0 \le n < l$ and $\Delta(u_{j+1}, v_n^j, v_{n+1}^j) \in C_\varepsilon$ for all $l \le n < m_n$. This shows homotopy equivalence of η and $\rho(\eta)$ and therefore $[\rho(\eta)] = [\eta]$ which establishes equation (6).

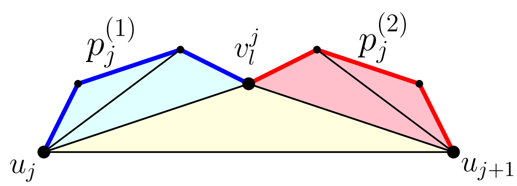

Fig. 3 A part of C_δ used to illustrate that $[\rho(\eta)] = [\eta]$. In particular, each edge $[u_j, u_{j+1}]$ will be mapped by ρ to a chain of edges. $p_j^{(1)}$ is colored in blue, and the path $p_j^{(2)}$ is colored in red. The homotopy is realized in two 2-dimensional cells (represented by the blue/red shading) that exist in C_δ based on the Čech construction.

Notice that $[\rho(\eta)] = \sum_{j=i}^{g} c_j[\gamma_j] \in H_\delta^{(1)}$ since $\frac{\ell_{i-1}}{4} \le \varepsilon < \frac{\ell_i}{4}$. By equation (6) we have

$$[\eta] = \mu_\varepsilon^h([\rho(\eta)]) = \mu_\varepsilon^h(\sum_{j=i}^{g} c_j[\gamma_j])$$
$$= \sum_{j=i}^{g} c_j[\mu_\varepsilon^c(\gamma_j)] \in \mathrm{Span}(\{[\mu_\varepsilon^c(\gamma_j)]\}_{j\geq i}),$$

which completes the proof of the surjectivity of μ_ε^h. This establishes the spanning condition B). In other words, if $[\eta]$ is a homology class in $H_\varepsilon^{(1)}$ then it must be formed only from homology classes $[\mu_\varepsilon^c(\gamma_j)]$ for $j \geq i$, and thus no additional cycles are created.

4 Future Work

The overarching theme of this work is to show how persistence may be used to obtain qualitative-quantitative summaries of metric graphs that reflect the underlying topology of the graphs. We obtained a complete characterization of all possible intrinsic Čech persistence diagrams in homological dimension one for metric graphs. What is currently known regarding the characterization of intrinsic Čech persistence diagrams for metric graphs is summarized in a diagram shown in Figure 4. The horizontal axis represents the homological dimension and the vertical axis represents the genus (number of shortest cycles) of a graph. In this setting, the previous results of Adamaszek, et al. [3] who consider the intrinsic Čech persistence diagrams in all dimensions for a graph that consists of a single cycle, lie on the horizontal line at height one, while the results in this paper constitute the blue vertical line. The rest of the upper-right quadrant is unknown and our hope is to make further progress toward a complete characterization of the intrinsic Čech persistence diagrams associated to arbitrary metric graphs. Moreover, we aim to generalize our results to the Vietoris-Rips complex.

The choice of a particular complex may be inspired by particular graph features that one is interested in. A *graph motif* is usually thought of as a graph on a small number of vertices (in general, any graph pattern can be a motif). Counting the number of small motifs in a graph is equivalent to the subgraph isomorphism problem, which is NP-complete. Since persistence has a polynomial time computational complexity, the question we would like to answer is: can the intrinsic Čech or other related persistence diagrams be used to determine or approximate graph motif counts? Additionally, the local version of this question, the number of graph motifs incident with a particular vertex, may be approached via the local homology at a vertex (homology of the k-neighborhood of a vertex relative to its boundary). As a start, persistence-based characterizations of a class of graph motifs should be obtained. Depending on the type of characterization obtained, we would be interested in determining to what extent our persistence-based summaries could be useful in the classification of the motifs present in a query graph.

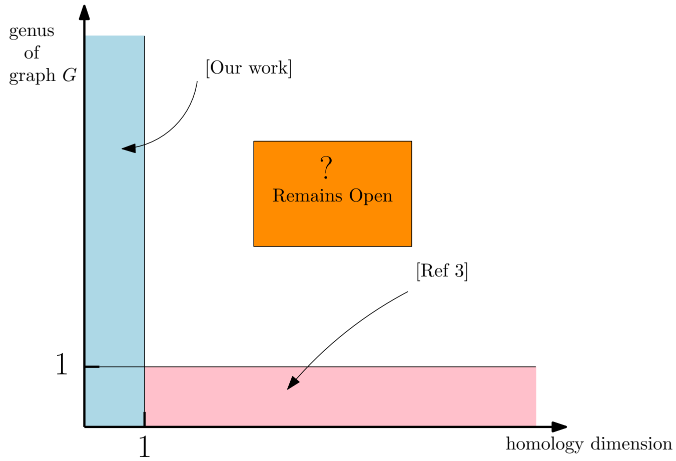

Fig. 4 Figure summarizing the results from this paper and from [3].

Ultimately, the complete or partial characterization of the topological information about a graph that is captured by persistent homology associated to various chain complex constructions is closely related to comparing their discriminative powers. In particular, we are interested in comparing the Čech and persistence distortion distance summaries.

The intrinsic Čech filtration and associated persistence diagrams allow one to define the *intrinsic Čech distance*, d_{IC}, between two metric graphs (G_1, d_{G_1}) and (G_2, d_{G_2}). This distance, introduced in [6], is defined as follows:

$$d_{IC}(G_1, G_2) := d_B(Dg_1 IC_{G_1}, Dg_1 IC_{G_2}),$$

where d_B is the bottleneck distance between the two intrinsic Čech persistence diagrams in dimension 1.

The *persistence distortion distance*, d_{PD}, that was first introduced in [9], is more closely related to the metric properties of a graph. Given a base point $s \in |G|$, define $f_s : |G| \to \mathbb{R}$ to be the geodesic distance to the base point s, i.e, $f_s(x) = d_G(s,x)$ for all $x \in |G|$. Then $Dg_1 f_s$ is the 1st-extended persistence diagram [7] associated to the sublevel set filtration induced by f_s. One may do this for any given base point in the metric graph, yielding a set of persistence diagrams for each graph. Let

$$\begin{aligned} \phi : |G| &\to SpDg \\ s &\mapsto Dgf_s \end{aligned}$$

where $SpDg$ denotes the space of all persistence diagrams. Then $\phi(|G|) \subset SpDg$ is called the *persistence distortion* of G. The *persistence distortion distance* between two metric graphs is defined to be the Hausdorff distance between their persistence distortion sets:

$$d_{PD}(G_1, G_2) := d_H(\phi(|G_1|), \phi(|G_2|)).$$

A natural question to ask is whether or not d_{PD} is more discriminative than d_{IC}, i.e., whether or not there exists a constant $c > 0$ such that

$$d_{IC}(G_1, G_2) \leq c \cdot d_{PD}(G_1, G_2).$$

We are currently working on extending preliminary results that establish the inequality for certain classes of metric graphs to arbitrary input graphs.

Acknowledgements

We are grateful for the Women in Computational Topology (WinCompTop) workshop for initiating our research collaboration. In particular, participant travel support was made possible through an NSF grant (NSF-DMS-1619908), and some additional travel support and social outings throughout the workshop were made possible through a gift from Microsoft Research. The Institute for Mathematics and its Applications generously offered in-kind the use of their facilities, their experienced staff to facilitate conference logistics, and refreshments. We appreciate their continued support of the applied algebraic topology community with regard to the WinCompTop Workshop, the special thematic program *Scientific and Engineering Applications of ALgebraic Topology* held during the 2013-2014 academic year, and the Applied Algebraic Topology Network WebEx talks. Finally, we would like to thank the AWM ADVANCE grant for travel support for organizers and speakers to attend the WinCompTop special session at the AWM Research Symposium in April 2017 .

EP was partially supported by the Asymmetric Resilient Cybersecurity Initiative at Pacific Northwest National Laboratory, part of the Laboratory Directed Research and Development Program at PNNL, a multi-program national laboratory operated by Battelle for the U.S. Department of Energy. During the completion of this project, RS was partially supported by the Simons Collaboration Grant, BW was partially supported by NSF-IIS-1513616, and YW was partially supported by NSF-CCF-1526513.

References

1. Mridul Aanjaneya, Frédéric Chazal, Daniel Chen, Marc Glisse, Leonidas Guibas, and Dmitriy Morozov. Metric graph reconstruction from noisy data. *International Journal of Computational Geometry & Applications*, 22(04):305–325, 2012.
2. Michal Adamaszek and Henry Adams. The Vietoris-Rips complexes of a circle. arXiv 1503.03669, 2015.
3. Michał Adamaszek, Henry Adams, Florian Frick, Chris Peterson, and Corrine Previte-Johnson. Nerve complexes of circular arcs. *Discrete & Computational Geometry*, 56(2):251–

273, 2016.
4. Bharat Biswal, F. Zerrin Yetkin, Victor M. Haughton, and James S. Hyde. Functional connectivity in the motor cortex of resting human brain using echo-planar MRI. *Magnetic Resonance in Medicine*, 34:537–541, 1995.
5. Gunnar Carlsson. Topology and data. *Bulletin of the American Mathematical Society*, 46(2):255–308, 2009.
6. Frédéric Chazal, Vin de Silva, and Steve Oudot. Persistence stability for geometric complexes. *Geometriae Dedicata*, 173:193–214, 2014.
7. David Cohen-Steiner, Herbert Edelsbrunner, and John Harer. Extending persistence using poincaré and lefschetz duality. *Foundations of Computational Mathematics*, 9(1):79–103, 2009.
8. Vin de Silva and Robert Ghrist. Coverage in sensor networks via persistent homology. *Algebraic & Geometric Topology*, 7:339–358, 2007.
9. Tamal K. Dey, Dayu Shi, and Yusu Wang. Comparing graphs via persistence distortion. In Lars Arge and János Pach, editors, *Proceedings 31st International Symposium on Computational Geometry*, volume 34 of *Leibniz International Proceedings in Informatics (LIPIcs)*, pages 491–506, Dagstuhl, Germany, 2015. Schloss Dagstuhl–Leibniz-Zentrum fuer Informatik.
10. Herbert Edelsbrunner and John Harer. Persistent homology - a survey. *Contemporary Mathematics*, 453:257–282, 2008.
11. Robert Ghrist. Barcodes: The persistent topology of data. *Bulletin of the American Mathematical Society*, 45:61–75, 2008.
12. Allen Hatcher. *Algebraic Topology*. Cambridge University Press, 2002.
13. Peter Kuchment. Quantum graphs I. Some basic structures. *Waves in Random Media*, 14(1):S107–S128, 2004.
14. W.S. Massey. *A Basic Course in Algebraic Topology*. Graduate Texts in Mathematics. Springer New York, 1991.
15. J.R. Munkres. *Elements of Algebraic Topology*. Advanced book classics. Perseus Books, 1984.
16. Corrine Previte. *The $\mathscr{D}$-Neighborhood Complex of a Graph*. PhD thesis, Colorado State University, 2014.
17. J. Rotman. *An Introduction to Homological Algebra*. Universitext. Springer New York, 2009.
18. Demet Taylan. Matching trees for simplicial complexes and homotopy type of devoid complexes of graphs. *Order*, 33:459–476, 2016.