# Persistent Path Homology of Directed Networks*

Samir Chowdhury† Facundo Mémoli‡

**Abstract**

While standard persistent homology has been successful in extracting information from metric datasets, its applicability to more general data, e.g. directed networks, is hindered by its natural insensitivity to asymmetry. We extend a construction of homology of digraphs due to Grigoryan, Lin, Muranov and Yau to the persistent framework. The result, which we call persistent path homology or PPH, encodes a rich level of detail about the asymmetric structure of the input directed network. For example, we prove that PPH identifies a class of directed cyclic networks as directed analogues of the circle. In general, PPH produces signatures that differ from natural extensions of Rips or Čech persistence to the directed setting, but we prove that PPH agrees with Čech persistence on symmetric spaces. Additionally, we prove that PPH agrees with Čech persistence on directed networks satisfying a local condition that we call square-freeness. We prove stability of PPH by utilizing a separate theory of homotopy of digraphs that is compatible with path homology. Finally, we study computational aspects of PPH, and derive an algorithm showing that over field coefficients, computing PPH requires the same worst case running time as standard persistent homology.

## 1 Introduction

In recent years, the advent of sophisticated data mining tools has led to rapid growth of network datasets in the sciences. The recently completed Human Connectome Project (2010-2015, http://www.humanconnectome.org/), aimed at mapping the network structure of the human brain, is one example of a large-scale network data acquisition project. The availability of such network data coincides with a time of steady growth of the mathematical theory of *persistent homology*, which aims to study the "shape" of data and thus appears to be a good candidate for analysing network structure. This connection is being developed rapidly [33, 34, 31, 32, 24, 23, 18, 29, 17], but this exploration is far from complete.

There are two problems that arise when studying *directed networks*—complete graphs with *asymmetric*, real-valued weights—via persistent homology: (1) conventional persistence methods take only metric space (i.e. symmetric) or point cloud (i.e. Euclidean) data as input, and (2) these methods typically factor the data through a filtration of simplicial complexes, which are themselves undirected objects. Thus the challenge is to develop persistent homology methods that accept asymmetric data as input, and associate homological signatures without forcing symmetry on the data at any point in the pipeline.

This problem has received some attention in recent literature. Some notable approaches involve computing the homology and Euler characteristic of *directed clique complexes*—see [18, 29] for theoretical and algorithmic details. A persistent homology framework for directed clique complexes was introduced in [35], albeit without an implementation. Yet another approach, using asymmetry-sensitive simplicial complexes called Dowker complexes, has appeared with experimental details in [17].

**Contributions.** In this paper, we address the challenge presented above by constructing the *persistent path homology (PPH)* method for assigning asymmetry-sensitive persistent homology signatures to network data. The key property of PPH is that it factors the input data through a filtration of *directed graphs*, which maintains the asymmetry in the input data. By characterizing the 1-dimensional PPH of a family of directed *cycle networks*, we provide evidence that PPH can appropriately detect directionality information in data.

The main theoretical foundations from which PPH is derived can be traced back to work of Barcelo et al., who developed a notion of *homotopy* for (undirected) graphs [4, 3]. This work has recently been extended by Grigor'yan et al. [26] to a notion of homotopy for directed graphs. Moreover, they proved that this notion of homotopy is consistent with a homology theory on digraphs called *path homology* that they had developed earlier [25]. It is this notion of path homology that we extend to obtain PPH.

*Supported by NSF grants IIS-1422400 and CCF-1526513. Facundo Mémoli was supported by the Mathematical Biosciences Institute at The Ohio State University.

†Department of Mathematics, The Ohio State University. chowdhury.57@osu.edu

‡Departments of Mathematics and Computer Science and Engineering, The Ohio State University. memoli@math.osu.edu

Two main challenges that we faced when attempting to establish the viability of PPH as a data analysis tool were in proving that: (1) PPH is stable to perturbations in the input data, and (2) PPH can be implemented via an algorithm with reasonable complexity. In the persistence literature, standard methods for proving stability invoke results that hold for simplicial complexes. These results are not available in the PPH setting; however, by invoking results regarding homotopy of digraphs obtained in [26], we were still able to prove stability of PPH. The implementation of PPH presented an additional challenge because the chain complex of vector spaces at the core of the persistence machinery does not come equipped with a natural choice of basis. This is in contrast to the setting of simplicial complexes, where the list of all simplices forms a natural basis for the associated chain complex. We resolve this challenge by proving that the required basis can be obtained by the same process of left-to-right Gaussian elimination on the columns of boundary matrices as that used in the general persistent homology algorithm. This observation shows that by using the general persistent homology algorithm, we can compute PPH without any additional overhead for finding the basis.

**Organization of the paper** §3 contains the necessary background on persistent homology, networks, and the network distance $d_{\mathcal{N}}$. §2 contains a description of path homology. In §4 we combine ingredients from the preceding sections to define PPH and prove its stability. In §5 we describe a procedure for implementing PPH in practice.

**Notation** We denote the nonempty elements of the power set of a set $X$ by $\mathrm{Pow}(X)$, and use the convention that the empty set is excluded from $\mathrm{Pow}(X)$. We write $\mathbb{Z}_+, \mathbb{R}_+$ to denote the nonnegative integers and reals, respectively. We will write $\overline{\mathbb{R}}$ to denote the extended real numbers $[-\infty, \infty]$. We fix a field $\mathbb{K}$ and use it throughout the paper. The identity map on a set $X$ is denoted $\mathrm{id}_X$. Given vector spaces $V, V'$, we write $V \cong V'$ to denote isomorphism of vector spaces. Given a finite set $S$, we write $\mathbb{K}[S]$ to denote the free vector space over $\mathbb{K}$ generated by the elements of $S$. When we have a sequence of maps $(f_i)_{i\in I}$ indexed by a set $I$, we will often refer to them collectively as $f_\bullet$, without specifying an index. Given sets $A, B$, a map $f : A \to B$, and subsets $S_A \subseteq A, S_B \subseteq B$, we will write $f(S_A) \subseteq S_B$ to mean that $f(s) \in S_B$ for each $s \in S_A$.

In this paper, a *digraph* is a pair $G = (X, E)$, where $X$ is a finite set (the *vertices*) and $E$ is a subset of $X \times X$ (the *edges*). We always consider digraphs without self-loops. We also make the following remark on notation: given $x, x' \in X$ for a digraph $G = (X, E)$, we write $x \overset{\to}{=} x'$ to mean either $x = x'$, or $(x, x') \in E$.

## 2 Path homology of digraphs

Homology is the formal algebraic construction at the center of our work. For our purposes, we define homology in the setting of general vector spaces, and refer the reader to [30, §1.13] for additional details. Fix a field $\mathbb{K}$. A *chain complex* is defined to be a sequence of vector spaces $(C_k)_{k\in\mathbb{Z}}$ over $\mathbb{K}$ and *boundary maps* $(\partial_k : C_k \to C_{k-1})_{k\in\mathbb{Z}}$ satisfying the condition $\partial_{k-1} \circ \partial_k = 0$ for each $k \in \mathbb{Z}$. We often denote a chain complex as $\mathcal{C} = (C_k, \partial_k)_{k\in\mathbb{Z}}$. Given a chain complex $\mathcal{C}$ and any $k \in \mathbb{Z}_+$, one defines the following subspaces:

$$\begin{aligned} Z_k(\mathcal{C}) &:= \ker(\partial_k) = \{c \in C_k : \partial_k(c) = 0\}, \text{ the } k\text{-cycles},\\ B_k(\mathcal{C}) &:= \mathrm{im}(\partial_{k+1}) = \{c \in C_k : c = \partial_{k+1}(b) \text{ for some } \\ &\qquad b \in C_{k+1}\}, \text{the } k\text{-boundaries}. \end{aligned}$$

The quotient vector space $H_k(\mathcal{C}) := Z_k(\mathcal{C})/B_k(\mathcal{C})$ is called the *$k$-th homology vector space of the chain complex $\mathcal{C}$*. The dimension of $H_k(\mathcal{C})$ is called the *$k$-th Betti number* of $\mathcal{C}$, denoted $\beta_k(\mathcal{C})$.

Given two chain complexes $\mathcal{C} = (C_k, \partial_k)_{k\in\mathbb{Z}}$ and $\mathcal{C}' = (C'_k, \partial'_k)_{k\in\mathbb{Z}}$, a *chain map* $\varphi : \mathcal{C} \to \mathcal{C}'$ is a family of morphisms $(\varphi_k : C_k \to C'_k)_{k\in\mathbb{Z}_+}$ such that $\partial'_k \circ \varphi_k = \varphi_{k-1} \circ \partial_k$ for each $k \in \mathbb{Z}_+$. Such a chain map induces a family of linear maps $(\varphi_\#)_k : H_k(\mathcal{C}) \to H_k(\mathcal{C}')$ for each $k \in \mathbb{Z}_+$ [30].

In what follows, we summarize and condense some concepts that appeared in [25], and attempt to preserve the original notation wherever possible.

**2.1 Elementary paths on a set** Given a finite set $X$ and any integer $p \in Z_+$, an *elementary $p$-path over $X$* is a sequence $[x_0, \ldots, x_p]$ of $p+1$ elements of $X$. For each $p \in \mathbb{Z}_+$, the free vector space consisting of all formal linear combinations of elementary $p$-paths over $X$ with coefficients in $\mathbb{K}$ is denoted $\Lambda_p = \Lambda_p(X) = \Lambda_p(X, \mathbb{K})$. One also defines $\Lambda_{-1} := \mathbb{K}$ and $\Lambda_{-2} := \{0\}$. Next, for any $p \in \mathbb{Z}_+$, one defines a linear map $\partial^{\mathrm{nr}}_p : \Lambda_p \to \Lambda_{p-1}$ to be the linearization of the following map on the generators of $\Lambda_p$:

$$\partial^{\mathrm{nr}}_p([x_0, \ldots, x_p]) := \sum_{i=0}^{p} (-1)^i [x_0, \ldots, \widehat{x_i}, \ldots, x_p],$$

for each elementary $p$-path $[x_0, \ldots, x_p] \in \Lambda_p$.

Here $\widehat{x_i}$ denotes omission of $x_i$ from the sequence. The maps $\partial^{\mathrm{nr}}_\bullet$ are referred to as the *non-regular boundary maps*. For $p = -1$, one defines $\partial^{\mathrm{nr}}_{-1} : \Lambda_{-1} \to \Lambda_{-2}$ to be the zero map. Then $\partial^{\mathrm{nr}}_{p+1} \circ \partial^{\mathrm{nr}}_p = 0$ for any integer $p \geq -1$ [27, Lemma 2.2]. It follows that $(\Lambda_p, \partial^{\mathrm{nr}}_p)_{p\in\mathbb{Z}_+}$ is a chain complex.

For notational convenience, we will often drop the square brackets and commas and write paths of the form $[a, b, c]$ as $abc$. We use this convention in the next example.

*Example.* [Paths on a double edge] We will soon explain the interaction between paths on a set and the edges on a digraph. First consider a digraph on a vertex set $Y = \{a, b\}$ as in Figure 1. Notice that there is a legitimate "path" on this digraph of the form $aba$, obtained by following the directions of the edges. But notice that applying $\partial_2^{\mathrm{nr}}$ to the 2-path $aba$ yields $\partial_2^{\mathrm{nr}}(aba) = ba - aa + ab$, and $aa$ is not a valid path on this particular digraph (self-loops are disallowed). To handle situations like this, one needs to consider *regular paths*, which are explained in the next section.

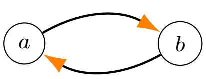



Figure 1: A two-node digraph on the vertex set $Y = \{a, b\}$.

### 2.2 Regular paths on a set

For each $p \in \mathbb{Z}_+$, an elementary $p$-path $[x_0, \ldots, x_p]$ is called *regular* if $x_i \neq x_{i+1}$ for each $0 \leq i \leq p-1$, and *irregular* otherwise. Then for each $p \in \mathbb{Z}_+$, one defines:

$$\mathcal{R}_p = \mathcal{R}_p(X, \mathbb{K}) := \mathbb{K}\big[\{[x_0, \ldots, x_p] : [x_0, \ldots, x_p] \text{ is regular}\}\big]$$
$$\mathcal{I}_p = \mathcal{I}_p(X, \mathbb{K}) := \mathbb{K}\big[\{[x_0, \ldots, x_p] : [x_0, \ldots, x_p] \text{ is irregular}\}\big].$$

One can further verify that $\partial_p^{\mathrm{nr}}(\mathcal{I}_p) \subseteq \mathcal{I}_{p-1}$ [27, Lemma 2.6], and so $\partial_p^{\mathrm{nr}}$ is well-defined on $\Lambda_p/\mathcal{I}_p$. Since $\mathcal{R}_p \cong \Lambda_p/\mathcal{I}_p$ via a natural linear isomorphism, one can define $\partial_p : \mathcal{R}_p \to \mathcal{R}_{p-1}$ as the pullback of $\partial_p^{\mathrm{nr}}$ via this isomorphism [27, Definition 2.7]. Then $\partial_p$ is referred to as the *regular boundary map* in dimension $p$, where $p \in \mathbb{Z}_+$. Now we obtain a new chain complex $(\mathcal{R}_p, \partial_p)_{p \in \mathbb{Z}_+}$.

*Example.* [Regular paths on a double edge] Consider again the digraph in Figure 1. Applying the regular boundary map to the 2-path $aba$ yields $\partial_2(aba) = ba + ab$. This example illustrates the following general principle: *Irregular paths arising from an application of $\partial_\bullet$ are treated as zeros.*

### 2.3 Allowed paths on digraphs

We now expand on the notion of paths on a set to discuss paths on a digraph. We follow the intuition developed in Examples 2.1 and 2.2.

Let $G = (X, E)$ be a digraph. For each $p \in \mathbb{Z}_+$, one defines an elementary $p$-path $[x_0, \ldots, x_p]$ on $X$ to be *allowed* if $(x_i, x_{i+1}) \in E$ for each $0 \leq i \leq p-1$. For each $p \in \mathbb{Z}_+$, the free vector space on the collection of allowed $p$-paths on $(X, E)$ is denoted $\mathcal{A}_p = \mathcal{A}_p(G) = \mathcal{A}_p(X, E, \mathbb{K})$, and is called the *space of allowed $p$-paths*. One further defines $\mathcal{A}_{-1} := \mathbb{K}$ and $\mathcal{A}_{-2} := \{0\}$.

### 2.4 $\partial$-invariant paths and path homology

The allowed paths do not form a chain complex, because the image of an allowed path under $\partial$ need not be allowed. This is rectified as follows. Given a digraph $G = (X, E)$ and any $p \in \mathbb{Z}_+$, the *space of $\partial$-invariant $p$-paths on $G$* is defined to be the following subspace of $\mathcal{A}_p(G)$:

$$\Omega_p = \Omega_p(G) = \Omega_p(X, E, \mathbb{K}) := \{c \in \mathcal{A}_p : \partial_p(c) \in \mathcal{A}_{p-1}\}.$$

One further defines $\Omega_{-1} := \mathcal{A}_{-1} \cong \mathbb{K}$ and $\Omega_{-2} := \mathcal{A}_{-2} = \{0\}$. Now it follows by the definitions that $\mathrm{im}(\partial_p(\Omega_p)) \subseteq \Omega_{p-1}$ for any integer $p \geq -1$. Thus we have a chain complex:

$$\ldots \xrightarrow{\partial_3} \Omega_2 \xrightarrow{\partial_2} \Omega_1 \xrightarrow{\partial_1} \Omega_0 \xrightarrow{\partial_0} \mathbb{K} \xrightarrow{\partial_{-1}} 0$$

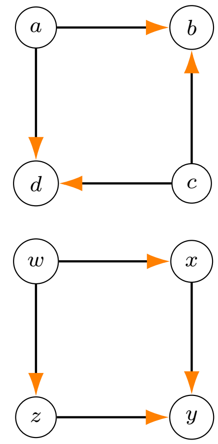



Figure 2: Two types of square digraphs.

For each $p \in \mathbb{Z}_+$, the *$p$-dimensional path homology groups of $\mathfrak{G} = (X, E)$* are defined as:

$$H_p^{\Xi}(\mathfrak{G}) = H_p^{\Xi}(X, E, \mathbb{K}) := \ker(\partial_p)/\operatorname{im}(\partial_{p+1}).$$

*Example.* [Paths on squares] We illustrate the construction of $\Omega_\bullet$ for the digraphs in Figure 2.

For $0 \leq p \leq 2$, we have the following vector spaces of $\partial$-invariant paths:

$$\begin{aligned}
\Omega_0(G_M) &= \mathbb{K}[\{a, b, c, d\}] \\
\Omega_1(G_M) &= \mathbb{K}[\{ab, cb, cd, ad\}] \\
\Omega_2(G_M) &= \{0\} \\
\Omega_0(G_N) &= \mathbb{K}[\{w, x, y, z\}] \\
\Omega_1(G_N) &= \mathbb{K}[\{wx, xy, zy, wz\}] \\
\Omega_2(G_N) &= \mathbb{K}[\{wxy - wzy\}]
\end{aligned}$$

The crux of the $\Omega_\bullet$ construction lies in understanding $\Omega_2(G_N)$. Note that even though $\partial_2^{G_N}(wxy), \partial_2^{G_N}(wzy) \notin \Omega_2(G_N)$ (because $wy \notin \mathcal{A}_1(G_N)$), we still have:

$$\partial_2^{G_N}(wxy - wzy) = xy - wy + wx - zy + wy - wz \in \mathcal{A}_1(G_N).$$

Elementary calculations show $\dim(H_1^{\Xi}(G_M)) = 1$, and $\dim(H_1^{\Xi}(G_N)) = 0$. Thus path homology can successfully distinguish between these two squares.

To compare this with a simplicial approach, consider the directed clique complex homology studied in

[18, 29, 35]. Given a digraph $G = (X, E)$, the directed clique complex is defined to be the *ordered* simplicial complex [30, p. 76] given by writing:

$$\mathfrak{F}_G := X \cup \{(x_0, \ldots, x_p) : (x_i, x_j) \in E \text{ for all } 0 \le i < j \le p\}.$$

Here we use parentheses to denote ordered simplices. For the squares in Figure 2, we have:

$$\mathfrak{F}_{G_M} = \{a, b, c, d, ab, cb, cd, ad\} \text{ and}$$
$$\mathfrak{F}_{G_N} = \{w, x, y, z, wx, xy, wz, zy\},$$

and so their simplicial homologies are equal.

Remark 2.1. **(The challenge of finding a natural basis for $\Omega_\bullet$)** *The digraph $G_N$ in Example 2.4 is a minimal example showing that it is nontrivial to compute bases for the vector spaces $\Omega_\bullet$. Specifically, while it is trivial to read off bases for the allowed paths $\mathcal{A}_\bullet$ from a digraph, one needs to consider* linear combinations *of allowed paths in a systematic manner to obtain bases for the $\partial$-invariant paths.*

*Contrast this with the setting of simplicial homology: here the simplices themselves form bases for the associated chain complex, so there is no need for an extra preprocessing step. Thus when using PPH for asymmetric data, it is important to consider the trade-off between greater sensitivity to asymmetry and increased computational cost.*

*We derive a procedure for systematically computing bases for $\Omega_\bullet$ in §5.*

**2.5 Homotopy of digraphs** The constructions of path homology are accompanied by a theory of homotopy developed in [26]. An illustrated example is provided in Figure 3.

Let $G_X = (X, E_X), G_Y = (Y, E_Y)$ be two digraphs. The *product digraph* $G_X \times G_Y = (X \times Y, E_{X\times Y})$ is defined as follows:

$$X \times Y := \{(x, y) : x \in X, y \in Y\}, \text{ and}$$
$$E_{X\times Y} := \{((x, y), (x', y')) \in (X \times Y)^2 : x = x' \text{ and } (y, y') \in E_Y, \text{ or } y = y' \text{ and } (x, x') \in E_X\}.$$

Next, the *line digraphs* $I^+$ and $I^-$ are defined to be the two-point digraphs with vertices $\{0, 1\}$ and edges $(0, 1)$ and $(1, 0)$, respectively. Two digraph maps $f, g : G_X \to G_Y$ are *one-step homotopic* if there exists a digraph map $F : G_X \times I \to G_Y$, where $I \in \{I^+, I^-\}$, such that:

$$F|_{G_X\times\{0\}} = f \text{ and } F|_{G_X\times\{1\}} = g.$$

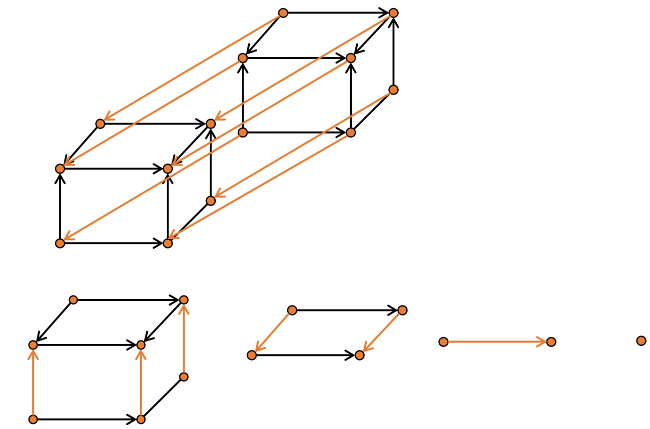


Figure 3: Directed $d$-cubes that are all homotopy equivalent.

This condition is equivalent to requiring:

$$f(x) \overrightarrow{=} g(x) \text{ for all } x \in X, \text{ or } g(x) \overrightarrow{=} f(x) \text{ for all } x \in X.$$

Moreover, $f$ and $g$ are *homotopic*, denoted $f \simeq g$, if there is a finite sequence of digraph maps $f_0 = f, f_1, \ldots, f_n = g : G_X \to G_Y$ such that $f_i, f_{i+1}$ are one-step homotopic for each $0 \le i \le n-1$. The digraphs $G_X$ and $G_Y$ are *homotopy equivalent* if there exist digraph maps $f : G_X \to G_Y$ and $g : G_Y \to G_X$ such that $g \circ f \simeq \mathrm{id}_{G_X}$ and $f \circ g \simeq \mathrm{id}_{G_Y}$.

An example of digraph homotopy equivalence is illustrated in Figure 3. Informally, the homotopy equivalence is given by "crushing" the orange arrows according to the directions they mark. This operation crushes the 4-tesseract to the 3-cube, to the 2-square, to the line, and finally to the point.

The concept of homotopy yields the following theorem on path homology groups:

Theorem 2.1. (Theorem 3.3, [26]) *Let $G, G'$ be two digraphs.*

1. *Let $f, g : G \to G'$ be two homotopic digraph maps. Then these maps induce identical maps on homology vector spaces. More precisely, the following maps are identical for each $p \in \mathbb{Z}_+$:*

$$(f_\#)_p : H_p(G) \to H_p(G')$$
$$(g_\#)_p : H_p(G) \to H_p(G').$$

2. *If $G$ and $G'$ are homotopy equivalent, then $H_p(G) \cong H_p(G')$ for each $p \in \mathbb{Z}_+$.*

## 3 Background on Persistent Homology and Networks

A *persistent vector space* is defined to be a family of vector spaces and linear maps $\{U^\delta \xrightarrow{\mu_{\delta,\delta'}} U^{\delta'}\}_{\delta \le \delta' \in \mathbb{R}}$ such that: (1) $\mu_{\delta,\delta}$ is the identity for each $\delta \in \mathbb{R}$, and (2) $\mu_{\delta,\delta''} = \mu_{\delta',\delta''} \circ \mu_{\delta,\delta'}$ for each $\delta \le \delta' \le \delta'' \in \mathbb{R}$. Persistent *homology* refers to the special case when we have a family of homology vector spaces and induced linear maps arising from chain complexes and chain maps.

By the classification results in [11, §5.2], it is possible to associate a full invariant called a *persistence barcode* to each persistent vector space. This barcode is a multiset of *persistence intervals*, and is represented as a set of lines over a single axis. The barcode of a persistent vector space $\mathcal{V}$ is denoted $\mathbf{Pers}(\mathcal{V})$. An equivalent representation is the *persistence diagram*, which is as a multiset of points lying on or above the diagonal in $\overline{\mathbb{R}}^2$, counted with multiplicity. More specifically,

$$\mathrm{Dgm}(\mathcal{V}) := \left[(\delta_i, \delta_{j+1}) \in \overline{\mathbb{R}}^2 : [\delta_i, \delta_{j+1}) \in \mathbf{Pers}(\mathcal{V})\right],$$

where the multiplicity of $(\delta_i, \delta_{j+1}) \in \overline{\mathbb{R}}^2$ is given by the multiplicity of $[\delta_i, \delta_{j+1}) \in \mathbf{Pers}(\mathcal{V})$.

Persistence diagrams can be compared using the *bottleneck distance*, which we denote by $d_{\mathrm{B}}$. Details about this distance, as well as the other material related to persistent homology, can be found in [13]. Numerous other formulations of the material presented above can be found in [22, 36, 8, 20, 19, 6, 21].

Remark 3.1. *From the definition of bottleneck distance, it follows that points in a persistence diagram* $\mathrm{Dgm}(\mathcal{V})$ *that belong to the diagonal do not contribute to the bottleneck distance between* $\mathrm{Dgm}(\mathcal{V})$ *and another diagram* $\mathrm{Dgm}(\mathcal{U})$*. Thus whenever we describe a persistence diagram as being* trivial, *we mean that either it is empty, or it does not have any off-diagonal points.*

**3.1 Interleavings** Let $\mathcal{U}, \mathcal{V}$ be two persistent vector spaces with linear maps $s_{\delta,\delta'} : U^\delta \to U^{\delta+\varepsilon}$ and $t_{\delta,\delta'} : V^\delta \to V^{\delta+\varepsilon}$, respectively, for each $\delta \le \delta' \in \mathbb{R}$. Given $\varepsilon \ge 0$, $\mathcal{U}$ and $\mathcal{V}$ are said to be *$\varepsilon$-interleaved* [12, 6] if there exist two families of linear maps $\{\varphi_\delta : V^\delta \to U^{\delta+\varepsilon}\}_{\delta\in\mathbb{R}}$ and $\{\psi_\delta : U^\delta \to V^{\delta+\varepsilon}\}_{\delta\in\mathbb{R}}$ such that: (1) $\varphi_{\delta'} \circ s_{\delta,\delta'} = t_{\delta+\varepsilon,\delta'+\varepsilon} \circ \varphi_\delta$, (2) $\psi_{\delta'} \circ t_{\delta,\delta'} = s_{\delta+\varepsilon,\delta'+\varepsilon} \circ \psi_\delta$, (3) $s_{\delta,\delta+2\varepsilon} = \psi_{\delta+\varepsilon} \circ \varphi_\delta$, and (4) $t_{\delta,\delta+2\varepsilon} = \varphi_{\delta+\varepsilon} \circ \psi_\delta$ for each $\delta \le \delta' \in \mathbb{R}$. The Algebraic Stability Theorem of [12] guarantees that if $\mathcal{U}$ and $\mathcal{V}$ are $\varepsilon$-interleaved, then $d_{\mathrm{B}}(\mathrm{Dgm}(\mathcal{U}), \mathrm{Dgm}(\mathcal{V})) \le \varepsilon$. Details on these results are provided in Appendix B.

**3.2 Networks** We follow the framework of [9, 10]. A *network* is a finite set $X$ together with a *weight function* $A_X : X \times X \to \mathbb{R}$. This can be interpreted as a complete graph with asymmetric, real-valued weights, or alternatively, as a generalization of a finite metric space. Note that $A_X$ is not required to satisfy the triangle inequality or any symmetry condition. The collection of all such networks is denoted $\mathcal{N}$.

Given two networks $(X, A_X), (Y, A_Y) \in \mathcal{N}$ and $R \subseteq X \times Y$ any nonempty relation, the *distortion* of $R$ is defined as:

$$\mathrm{dis}(R) := \max_{(x,y),(x',y') \in R} |A_X(x,x') - A_Y(y,y')|.$$

A *correspondence between $X$ and $Y$* is a relation $R$ between $X$ and $Y$ such that $\pi_X(R) = X$ and $\pi_Y(R) = Y$, where $\pi_X : X \times Y \to X$ and $\pi_Y : X \times Y \to Y$ denote the natural projections. The collection of all correspondences between $X$ and $Y$ will be denoted $\mathscr{R}(X,Y)$.

Following prior work in [10], the *network distance* $d_{\mathcal{N}} : \mathcal{N} \times \mathcal{N} \to \mathbb{R}_+$ is then defined as:

$$d_{\mathcal{N}}(X,Y) := \tfrac{1}{2} \min_{R \in \mathscr{R}(X,Y)} \mathrm{dis}(R).$$

It can be verified that $d_{\mathcal{N}}$ as defined above is a pseudometric, and that the networks at 0-distance can be completely characterized [15, 16]. Next we wish to state a reformulation of $d_{\mathcal{N}}$ that will aid our proofs. First we define the distortion of a map between two networks. Given any $(X, A_X), (Y, A_Y) \in \mathcal{N}$ and a map $\varphi : (X, A_X) \to (Y, A_Y)$, the *distortion* of $\varphi$ is defined as:

$$\mathrm{dis}(\varphi) := \max_{x,x' \in X} |A_X(x,x') - A_Y(\varphi(x), \varphi(x'))|.$$

Given maps $\varphi : (X, \omega_X) \to (Y, \omega_Y)$ and $\psi : (Y, \omega_Y) \to (X, \omega_X)$, we define two *co-distortion* terms:

$$C_{X,Y}(\varphi,\psi) := \max_{(x,y) \in X \times Y} |\omega_X(x, \psi(y)) - \omega_Y(\varphi(x), y)|,$$
$$C_{Y,X}(\psi,\varphi) := \max_{(y,x) \in Y \times X} |\omega_Y(y, \varphi(x)) - \omega_X(\psi(y), x)|.$$

Proposition 3.1. ([17, Proposition 4]) *Let* $(X, A_X), (Y, A_Y) \in \mathcal{N}$*. Then,*

$$d_{\mathcal{N}}(X,Y) = \tfrac{1}{2} \min \big\{ \max\big( \mathrm{dis}(\varphi), \mathrm{dis}(\psi), C_{X,Y}(\varphi,\psi), C_{Y,X}(\psi,\varphi)\big) : \varphi : X \to Y, \psi : Y \to X \text{ any maps} \big\}.$$

Remark 3.2. *Proposition 3.1 is analogous to a result of Kalton and Ostrovskii [28, Theorem 2.1] where—instead of $d_{\mathcal{N}}$—one has the Gromov-Hausdorff distance between metric spaces. We remark that when* *restricted to the special case of networks that are also metric spaces, the network distance $d_{\mathcal{N}}$ agrees with the Gromov-Hausdorff distance. Details on the Gromov-Hausdorff distance can be found in [7].*

*An important remark is that in the Kalton-Ostrovskii formulation, there is only one co-distortion term. When Proposition 3.1 is applied to metric spaces, the two co-distortion terms become equal by symmetry, and thus the Kalton-Ostrovskii formulation is recovered. But* a priori, *the lack of symmetry in the network setting requires us to consider both terms.*

**3.3 Rips and Dowker complexes** A standard tool in persistent homology is the *Vietoris-Rips complex*. For any network $(X, A_X)$ and any $\delta \in \mathbb{R}$, this is defined as $\mathfrak{R}^\delta_X := \{\sigma \subseteq X : \max_{x,x' \in X} A_X(x, x') \leq \delta\}$. In [17], it was shown that Rips complexes are insensitive to asymmetry in data. The *Dowker sink/source complexes* introduced in [17] were, however, shown to be sensitive to asymmetry in data. The sink complex is defined as $\mathfrak{D}^{\text{si}}_{\delta,X} := \{\sigma \in X : A_X(x, p) \leq \delta$ for all $x \in \sigma$, for some $p \in X\}$, and the source complex is defined analogously, by swapping the positions of $x$ and $p$. It was shown in [17] that both types of complexes produce the same persistence diagram. The $k$-dimensional Dowker persistence diagram of $(X, A_X)$ is then denoted $\mathrm{Dgm}^{\mathfrak{D}}_k(X)$. In the setting of metric spaces, the Dowker complex coincides with the Čech complex. In such cases, we refer to Dowker/Čech persistence interchangeably.

## 4 The Persistent Path Homology of a Network

Let $\mathcal{X} = (X, A_X) \in \mathcal{N}$. For any $\delta \in \mathbb{R}$, the digraph $\mathfrak{G}^\delta_{\mathcal{X}} = (X, E^\delta_X)$ is defined as follows:

$$E^\delta_X := \{(x, x') \in X \times X : x \neq x', A_X(x, x') \leq \delta\}.$$

Note that for any $\delta' \geq \delta \in \mathbb{R}$, we have a natural inclusion map $\mathfrak{G}^\delta_{\mathcal{X}} \hookrightarrow \mathfrak{G}^{\delta'}_{\mathcal{X}}$. Thus we may associate to $\mathcal{X}$ the *digraph filtration* $\{\mathfrak{G}^\delta_{\mathcal{X}} \hookrightarrow \mathfrak{G}^{\delta'}_{\mathcal{X}}\}_{\delta \leq \delta' \in \mathbb{R}}$.

The functoriality of the path homology construction (Appendix A, Proposition A.2) enables us to obtain a persistent vector space from a digraph filtration. Thus we make the following definition:

Definition 4.1. *Let $\mathfrak{G} = \{\mathfrak{G}^\delta \hookrightarrow \mathfrak{G}^{\delta'}\}_{\delta \leq \delta' \in \mathbb{R}}$ be a digraph filtration. Then for each $p \in \mathbb{Z}_+$, we define the* $p$-dimensional persistent path homology of $\mathfrak{G}$ *to be the following persistent vector space:*

$$\mathcal{H}^\Xi_p(\mathfrak{G}) := \{H^\Xi_p(\mathfrak{G}^\delta) \xrightarrow{(\iota_{\delta,\delta'})_\#} H^\Xi_p(\mathfrak{G}^{\delta'})\}_{\delta \leq \delta' \in \mathbb{R}}.$$

*The diagram associated to $\mathcal{H}^\Xi_p(\mathfrak{G})$ is denoted* $\mathrm{Dgm}^\Xi_p(\mathfrak{G})$. *In particular, given $(X, A_X) \in \mathcal{N}$ and its digraph filtration, we write $\mathrm{Dgm}^\Xi_p(X)$ to denote its path persistence diagram in dimension $p$.*

The first main theorem of this section, which shows that the persistent path homology construction is stable to perturbations of input data, and hence amenable to data analysis, is below:

Theorem 4.1. (Stability) *Let $\mathcal{X} = (X, A_X), \mathcal{Y} = (Y, A_Y) \in \mathcal{N}$. Let $p \in \mathbb{Z}_+$. Then,*

$$d_{\mathrm{B}}(\mathrm{Dgm}_p(\mathcal{X}), \mathrm{Dgm}_p(\mathcal{Y})) \leq 2 d_{\mathcal{N}}(\mathcal{X}, \mathcal{Y}).$$

The proof uses results on the *interleaving distance*, for which we provide details in Appendix B.

*Proof.* [Proof of Theorem 4.1] Let $\eta = 2d_{\mathcal{N}}(\mathcal{X}, \mathcal{Y})$. By virtue of Proposition 3.1, we obtain maps $\varphi : X \to Y$ and $\psi : Y \to X$ such that $\mathrm{dis}(\varphi) \leq \eta, \mathrm{dis}(\psi) \leq \eta$, $C_{X,Y}(\varphi, \psi) \leq \eta$, and $C_{Y,X}(\psi, \varphi) \leq \eta$.

*Claim.* For each $\delta \in \mathbb{R}$, the map $\varphi$ induces a digraph map $\varphi_\delta : \mathfrak{G}^\delta_{\mathcal{X}} \to \mathfrak{G}^{\delta+\eta}_{\mathcal{Y}}$ given by $x \mapsto \varphi(x)$, and the map $\psi$ induces a digraph map $\psi_\delta : \mathfrak{G}^\delta_{\mathcal{Y}} \to \mathfrak{G}^{\delta+\eta}_{\mathcal{X}}$ given by $y \mapsto \psi(y)$.

To see the claim, let $\delta \in \mathbb{R}$, and let $(x, x') \in E^\delta_X$. Then $A_X(x, x') \leq \delta$. Because $\mathrm{dis}(\varphi) \leq \eta$, we have $A_Y(\varphi(x), \varphi(x')) \leq \delta + \eta$. Thus $(\varphi(x), \varphi(x')) \in E^{\delta+\eta}_Y$, and so $\varphi_\delta$ is a digraph map. Similarly, $\psi_\delta$ is a digraph map. Since $\delta \in \mathbb{R}$ was arbitrary, the claim now follows.

*Claim.* Let $\delta \leq \delta' \in \mathbb{R}$, and let $s_{\delta,\delta'}, t_{\delta+\eta,\delta'+\eta}$ denote the digraph inclusion maps $\mathfrak{G}^\delta_{\mathcal{X}} \hookrightarrow \mathfrak{G}^{\delta'}_{\mathcal{X}}$ and $\mathfrak{G}^{\delta+\eta}_{\mathcal{Y}} \hookrightarrow \mathfrak{G}^{\delta'+\eta}_{\mathcal{Y}}$, respectively. Then $\varphi_{\delta'} \circ s_{\delta,\delta'}$ and $t_{\delta+\eta,\delta'+\eta} \circ \varphi_\delta$ are one-step homotopic.

To see this claim, let $x \in X$. We wish to show $\varphi_{\delta'}(s_{\delta,\delta'}(x)) \overrightarrow{=} t_{\delta+\eta,\delta'+\eta}(\varphi_\delta(x))$. But notice that:

$$\varphi_{\delta'}(s_{\delta,\delta'}(x)) = \varphi_{\delta'}(x) = \varphi(x),$$

where the second equality is by definition of $\varphi_{\delta'}$ and the first equality occurs because $s_{\delta,\delta'}$ is the inclusion map. Similarly, $t_{\delta+\eta,\delta'+\eta}(\varphi_\delta(x)) = t_{\delta+\eta,\delta'+\eta}(\varphi(x)) = \varphi(x)$. Thus we obtain $\varphi_{\delta'}(s_{\delta,\delta'}(x)) \overrightarrow{=} t_{\delta+\eta,\delta'+\eta}(\varphi_\delta(x))$. Since $x$ was arbitrary, it follows that $\varphi_{\delta'} \circ s_{\delta,\delta'}$ and $t_{\delta+\eta,\delta'+\eta} \circ \varphi_\delta$ are one-step homotopic.

*Claim.* Let $\delta \in \mathbb{R}$, and let $s_{\delta,\delta+2\eta}$ denote the digraph inclusion map $\mathfrak{G}^\delta_{\mathcal{X}} \hookrightarrow \mathfrak{G}^{\delta+2\eta}_{\mathcal{X}}$. Then $s_{\delta,\delta+2\eta}$ and $\psi_{\delta+\eta} \circ \varphi_\delta$ are one-step homotopic.

To see this claim, recall that $C_{X,Y}(\varphi, \psi) \leq \eta$, which means that for any $x \in X, y \in Y$, we have:

$$|A_X(x, \psi(y)) - A_Y(\varphi(x), y)| \leq \eta.$$

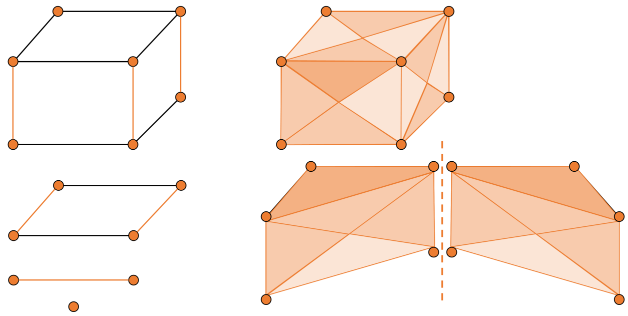


Figure 4: **Left:** $\mathfrak{G}^{\delta}_{\square_3}$ is (digraph) homotopy equivalent to a point at $\delta = 1$, as can be seen by collapsing points along the orange lines. **Right:** $\mathfrak{D}^{\mathrm{si}}_{\delta,\square_3}$ becomes contractible at $\delta = \sqrt{2}$, but has nontrivial homology in dimension 2 that persists across the interval $[1, \sqrt{2})$.

Let $x \in X$, and let $y = \varphi(x)$. Notice that $s_{\delta,\delta+2\eta}(x) = x$ and $\psi_{\delta+\eta}(\varphi_\delta(x)) = \psi(\varphi(x))$. Also note:

$$A_X(x, \psi(\varphi(x))) \leq \eta + A_Y(\varphi(x), \varphi(x)) \leq \delta + 2\eta.$$

Thus $s_{\delta,\delta+2\eta}(x) \overset{\rightarrow}{=} \psi_{\delta+\eta}(\varphi_\delta(x))$, and this holds for any $x \in X$. The claim follows.

By combining the preceding claims and Theorem 2.1, we obtain the following, for each $p \in \mathbb{Z}_+$:

$$((s_{\delta,\delta+2\eta})_\#)_p = ((\psi_{\delta+\eta} \circ \varphi_\delta)_\#)_p,$$
$$((\varphi_{\delta'} \circ s_{\delta,\delta'})_\#)_p = ((t_{\delta+\eta,\delta'+\eta} \circ \varphi_\delta)_\#)_p.$$

By invoking functoriality of path homology (Proposition A.2), we obtain:

$$((s_{\delta,\delta+2\eta})_\#)_p = ((\psi_{\delta+\eta})_\#)_p \circ ((\varphi_\delta)_\#)_p,$$
$$((\varphi_{\delta'})_\#)_p \circ (s_{\delta,\delta'})_\#)_p = ((t_{\delta+\eta,\delta'+\eta})_\#)_p \circ ((\varphi_\delta)_\#)_p.$$

By using similar arguments, we can also obtain, for each $p \in \mathbb{Z}_+$,

$$((t_{\delta,\delta+2\eta})_\#)_p = ((\varphi_{\delta+\eta})_\#)_p \circ ((\psi_\delta)_\#)_p,$$
$$((\psi_{\delta'})_\#)_p \circ (t_{\delta,\delta'})_\#)_p = ((s_{\delta+\eta,\delta'+\eta})_\#)_p \circ ((\psi_\delta)_\#)_p.$$

Thus $\mathcal{H}^{\Xi}_p(\mathcal{X})$ and $\mathcal{H}^{\Xi}_p(\mathcal{Y})$ are $\eta$-interleaved, for each $p \in \mathbb{Z}_+$. The result now follows by an application of the Algebraic Stability Theorem (see §3.1, also Theorem B.1).

Remark 4.1. *The preceding stability result has analogous counterparts in the setting of Rips and Dowker persistence for asymmetric networks. The key difference in the proof technique is that in the Rips/Dowker settings one can use classical results about contiguous simplicial maps, whereas in this setting, we are required to use results on the homotopy of digraphs that were recently developed in [26].*

Having defined PPH, we now answer some fundamental questions related to its characterization. We show that PPH agrees with Čech/Dowker persistence on metric spaces in dimension 1, but not necessarily in higher dimensions. We also show that in the asymmetric case, PPH and Dowker agree in dimension 1 if a certain local condition is satisfied.

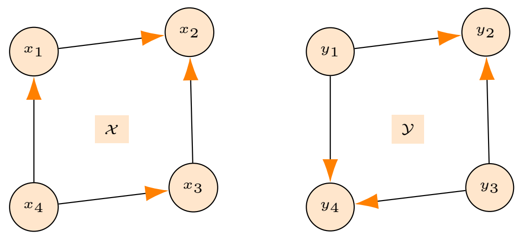
*Example.* [PPH vs Dowker for metric $n$-cubes] In the setting of metric spaces, PPH is generally different from Dowker persistence in dimensions $\geq 2$. To see this, consider $\mathbb{R}^n$ equipped with the Euclidean distance for $n \geq 3$. Define $\square_n := \{(i_1, i_2, \ldots, i_n) : i_j \in \{0, 1\}\ \forall\, 1 \leq j \leq n\}$. Then $\mathfrak{G}^{\delta}_{\square_n}$ has no edges for $\delta < 1$, and for $\delta = 1$, it has precisely an edge between any two points of $\square_n$ that differ on a single coordinate. But at $\delta = 1$, $\mathfrak{G}^{\delta}_{\square_n}$ is homotopy equivalent to $\mathfrak{G}^{\delta}_{\square_{n-1}}$: the homotopy equivalence is given by collapsing points that differ exactly on the $n$th coordinate (see Figure 4). Proceeding recursively, we see that $\mathfrak{G}^{\delta}_{\square_{n-1}}$ is contractible at $\delta = 1$. However, $\mathfrak{D}^{\mathrm{si}}(\square_n)$ is not contractible at $\delta = 1$. Moreover, an explicit verification for the $n = 3$ case shows that $\mathrm{Dgm}^{\mathfrak{D}}_2(\square_3)$ consists of the point $(1, \sqrt{2})$ with multiplicity 7. Thus $\mathrm{Dgm}^{\mathfrak{D}}_2(\square_3) \neq \mathrm{Dgm}^{\Xi}_2(\square_3)$.

Theorem 4.2. *Let $\mathcal{X} = (X, A_X) \in \mathcal{N}$ be a symmetric network, and fix $\mathbb{K} = \mathbb{Z}/p\mathbb{Z}$ for some prime $p$. Then $\mathrm{Dgm}^{\Xi}_1(\mathcal{X}) = \mathrm{Dgm}^{\mathfrak{D}}_1(\mathcal{X})$.*

The preceding result shows that on metric spaces, PPH agrees with Dowker persistence in dimension 1. The converse implication is not true: in §4.1, we provide a family of highly asymmetric networks for which PPH agrees with Dowker persistence in dimension 1. On the other hand, the examples in Figure 5 show that equality in dimension 1 does not necessarily hold for asymmetric networks. Moreover, it turns out that the four-point configurations illustrated in Figure 5 can be used to give another partial characterization of the networks for which PPH and Dowker persistence do agree in dimension 1. We present this statement next.

Definition 4.2. **Squares, triangles, and double edges.** *Let $\mathfrak{G}$ be a finite digraph. Then we define the following local configurations of edges between distinct nodes $a, b, c, d$:*

- *A* double edge *is a pair of edges $(a, b), (b, a)$.*
- *A* triangle *is a set of edges $(a, b), (b, c), (a, c)$.*
- *A* short square *is a set of edges $(a, b)$, $(a, d)$, $(c, b)$, $(c, d)$ such that neither of $(a, c)$, $(c, a)$, $(b, d)$, $(d, b)$ is an edge.*




| | $x_1$ | $x_2$ | $x_3$ | $x_4$ |
|---|---|---|---|---|
| $x_1$ | 0 | 1 | 2 | 2 |
| $x_2$ | 2 | 0 | 2 | 2 |
| $x_3$ | 2 | 1 | 0 | 2 |
| $x_4$ | 1 | 2 | 1 | 0 |

| | $y_1$ | $y_2$ | $y_3$ | $y_4$ |
|---|---|---|---|---|
| $y_1$ | 0 | 1 | 2 | 1 |
| $y_2$ | 2 | 0 | 2 | 2 |
| $y_3$ | 2 | 1 | 0 | 1 |
| $y_4$ | 2 | 2 | 2 | 0 |

Figure 5: Working over $\mathbb{Z}/2\mathbb{Z}$ coefficients, we find that $\mathrm{Dgm}_1^{\Xi}(\mathcal{X})$ and $\mathrm{Dgm}_1^{\mathfrak{D}}(\mathcal{Y})$ are trivial, whereas $\mathrm{Dgm}_1^{\mathfrak{D}}(\mathcal{X}) = \mathrm{Dgm}_1^{\Xi}(\mathcal{Y}) = \{(1,2)\} = \{(1,2)\}$.

- *A* long square *is a set of edges* $(a,b)$, $(b,c)$, $(a,d)$, $(d,c)$ *such that* $(a,c)$ *is not an edge.*

*All these are illustrated in Figure 6. Finally, we define a network* $(X, A_X)$ *to be* square-free *if* $\mathfrak{G}_X^{\delta}$ *does not contain a four-point subset whose induced subgraph is a short or long square, for any* $\delta \in \mathbb{R}$. *An important observation is that to be a square, the subgraph induced by a four-point subset cannot just include one of the configurations pictured above; it must exclude diagonal edges as well.*

Theorem 4.3. *Let* $\mathcal{X} = (X, A_X) \in \mathcal{N}$ *be a square-free network, and fix* $\mathbb{K} = \mathbb{Z}/p\mathbb{Z}$ *for some prime* $p$. *Then* $\mathrm{Dgm}_1^{\Xi}(\mathcal{X}) = \mathrm{Dgm}_1^{\mathfrak{D}}(\mathcal{X})$.

Remark 4.2. *The proofs of Theorems 4.2 and 4.3 both require an argument where simplices are paired up—this requires us to use* $\mathbb{Z}/p\mathbb{Z}$ *coefficients in both theorem statements.*

**4.1 An application: Characterizing the diagrams of cycle networks** For each $n \in \mathbb{N}$, consider the weighted, directed graph $(X_n, E_n, W_{E_n})$ with vertex set $X_n := \{x_1, x_2, \ldots, x_n\}$, edge set $E_n := \{(x_1,x_2),(x_2,x_3),\ldots,(x_{n-1},x_n),(x_n,x_1)\}$, and edge weights $W_{E_n} : E_n \to \mathbb{R}$ given by writing $W_{E_n}(e) = 1$ for each $e \in E_n$. Next let $A_{G_n} : X_n \times X_n \to \mathbb{R}$ denote the shortest path distance induced on $X_n \times X_n$ by $W_{E_n}$. Define the *cycle network on* $n$ *nodes* to be $G_n := (X_n, A_{G_n})$. A cycle network on 6 nodes is illustrated in Figure 7, along with its weight matrix. If

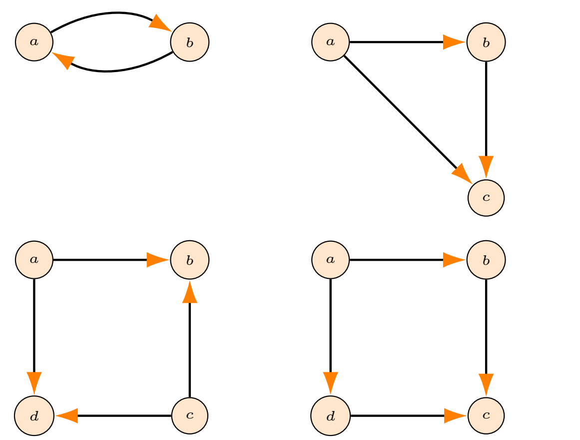
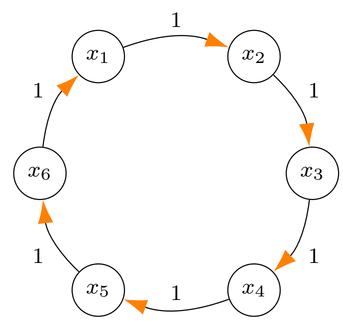



Figure 6: Squares, triangles, and double edges




| | $x_1$ | $x_2$ | $x_3$ | $x_4$ | $x_5$ | $x_6$ |
|---|---|---|---|---|---|---|
| $x_1$ | 0 | 1 | 2 | 3 | 4 | 5 |
| $x_2$ | 5 | 0 | 1 | 2 | 3 | 4 |
| $x_3$ | 4 | 5 | 0 | 1 | 2 | 3 |
| $x_4$ | 3 | 4 | 5 | 0 | 1 | 2 |
| $x_5$ | 2 | 3 | 4 | 5 | 0 | 1 |
| $x_6$ | 1 | 2 | 3 | 4 | 5 | 0 |

Figure 7: A cycle network on 6 nodes, along with its weight matrix. Note that the weights are highly asymmetric.

$x_1, x_2, \ldots, x_k \in X_n$ appear in $G_n$ in this clockwise order, we will write $x_1 \preceq x_2 \preceq \ldots \preceq x_k$.

Notice that cycle networks are square-free. If $a \prec b \prec c \prec d$ are four nodes on a cycle network, then for any $\delta \in \mathbb{R}$ such that we have an edge $a \to d$, we automatically have an edge $a \to c$. Thus the subgraph induced by $\{a,b,c,d\}$ cannot be either a long or a short square.

Cycle networks constitute an interesting family of examples with surprising connections to existing literature [2, 1]. In particular, their Dowker persistence diagrams can be fully characterized by results in [2], [1], and [17]. More specifically, given any $n \geq 3$, we know that $\mathrm{Dgm}_1^{\mathfrak{D}}(G_n)$ consists of the point $(1, \lceil n/2 \rceil)$ with multiplicity 1. In this sense, a cycle network is a directed analogue of the circle.

A natural test to see if PPH detects cyclic behavior in an expected way is to see if it can be characterized for cycle networks. This is the content of the following theorem.

THEOREM 4.4. *Let $G_n$ be a cycle network for some integer $n \geq 3$. Fix a field $\mathbb{K} = \mathbb{Z}/p\mathbb{Z}$ for some prime $p$. Then* $\mathrm{Dgm}_1^{\Xi}(G_n) = \{(1, \lceil n/2 \rceil)\}$.

## 5 Algorithmic Details

The origin of a general persistent homology algorithm for simplicial complexes can be traced back to [22] for $\mathbb{Z}/2\mathbb{Z}$ coefficients, and to [36] for arbitrary field coefficients. Here it was observed that the persistence algorithm has the same running time as Gaussian elimination over fields, i.e. $O(m^3)$ in the worst case, where $m$ is the number of simplices.

The PPH setting is more complicated, due to two reasons: (1) because of directionality, the number of $p$-paths on a vertex set is much larger than the number of $p$-simplices, for any $p \in \mathbb{N}$, and (2) one must first obtain bases for the $\partial$-invariant $p$-paths $\{\Omega_p : p \geq 2\}$. The first item is unavoidable, and even desirable—we capture the asymmetry in the data, thus retaining more information. For the second item, note that $\Omega_0$ and $\Omega_1$ are just the allowed 0 and 1-paths, so their bases can just be read off from the network weight function. After obtaining compatible bases for the filtered chain complex $\{\Omega_\bullet^i \to \Omega_\bullet^{i+1}\}_{i \in \mathbb{N}}$, however, one can use the general persistent homology algorithm [22, 36, 14]. By *compatible* bases, we mean a set of bases $\{B_p^i \subseteq \Omega_p^i : 0 \leq p \leq D+1,\ i \in \mathbb{N}\}$ such that $B_p^i \subseteq B_p^{i+1}$ for each $i$, and relative to which the transformation matrices $M_p$ of $\partial_p$ are known. Here $D$ is the dimension up to which we compute persistence.

We now present a procedure for obtaining compatible bases for the $\partial$-invariant paths. Fix a network $(X, A_X)$. We write $\mathcal{R}_p$ to denote $\mathcal{R}_p(X, \mathbb{K})$, for each $p \in \mathbb{Z}_+$. Given a digraph filtration on $X$, we obtain a filtered vector space $\{\mathcal{A}_\bullet^i \to \mathcal{A}_\bullet^{i+1}\}_{i=1}^N$ and a filtered chain complex $\{\Omega_\bullet^i \to \Omega_\bullet^{i+1}\}_{i=1}^N$ for some $N \in \mathbb{N}$. For any $p$-path $v$, define its *allow time* as $\mathfrak{at}(v) := \min\{k \geq 0 : v \in \mathcal{A}_p^k\}$. Similarly define its *entry time* as $\mathfrak{et}(v) := \min\{k \geq 0 : v \in \Omega_p^k\}$. The allow time and entry time coincide when $p = 0, 1$, but are not necessarily equal in general. In Figure 5, for example, we have $\mathfrak{at}(x_4x_1x_2) = 1 < 2 = \mathfrak{et}(x_4x_1x_2)$.

Now fix $p \geq 2$, and consider the map $\partial_p : \mathcal{R}_p \to \mathcal{R}_{p-1}$. Let $M_p$ denote the matrix representation of $\partial_p$, relative to an arbitrary choice of bases $B_p$ and $B_{p-1}$ for $\mathcal{R}_p$ and $\mathcal{R}_{p-1}$. For convenience, we write the bases as $B_p = \{v_i^p : 1 \leq i \leq \dim(\mathcal{R}_p)\}$ and $B_{p-1} = \{v_i^{p-1} : 1 \leq i \leq \dim(\mathcal{R}_{p-1})\}$, respectively. Each basis element has an allow time that can be computed efficiently, and the allow times belong the set $\{1, 2, \ldots, N\}$. By performing row and column swaps as needed, we can arrange $M_p$ so that the basis vectors for the domain are in increasing allow time, and the basis vectors for the codomain are in decreasing allow time. This is illustrated in Figure 8.

A special feature of $M_p$ is that it is stratified into horizontal strips given by the allow times of the codomain basis vectors. For each $1 \leq i \leq N$, we define the *height range* $i$ as:

$$hr(i) := \{1 \leq j \leq \dim(\mathcal{R}_{p-1}) : \mathfrak{at}(v_j^{p-1}) = i\}.$$

In words, $hr(i)$ lists the codomain basis vectors that have allow time $i$. Next we transform $M_p$ into a *column echelon form* $M_{p,G}$, using left-to-right Gaussian elimination. In this form, all nonzero columns are to the left of any zero column, and the leading coefficient (the topmost nonzero element) of any column is strictly above the leading coefficient of the column on its right. The leading coefficients are usually called *pivots*. An illustration of $M_{p,G}$ is provided in Figure 8. To obtain this column echelon form, the following elementary column operations are used:

1. swap columns $i$ and $j$,
2. replace column $j$ by $(\mathrm{col}\ j - k(\mathrm{col}\ i))$, where $k \in \mathbb{K}$.

The basis for the domain undergoes corresponding changes, i.e. we replace $v_j^p$ by $(v_j^p - kv_i^p)$ as necessary. We write the new basis $B_{p,G}$ for $\mathcal{R}_p$ as $\{\widehat{v}_i^p : 1 \leq i \leq \dim(\mathcal{R}_p)\}$. Moreover, we can write this basis as a union $B_{p,G} = \cup_{i=1}^N B_{p,G}^i$, where each $B_{p,G}^i := \{\widehat{v}_k^p : 1 \leq k \leq \dim(\mathcal{R}_p),\ \mathfrak{et}(\widehat{v}_k^p) \leq i\}$. This follows easily from the column echelon form: for each basis vector $v$ of the domain, the corresponding column vector is $\partial_p(v)$, and $\mathfrak{at}(\partial_p(v))$ can be read directly from the height of the column. Specifically, if the row index of the topmost nonzero entry of $\partial_p(v)$ belongs to $hr(i)$, then $\mathfrak{at}(\partial_p(v)) = i$, and if $\partial_p(v) = 0$, then $\mathfrak{at}(\partial_p(v)) = 0$. Then we have $\mathfrak{et}(v) = \max(\mathfrak{at}(v), \mathfrak{at}(\partial_p(v)))$.

REMARK 5.1. *In the Gaussian elimination step above, we only eliminate entries by adding paths that have already been allowed in the filtration. This means that for any operation of the form $v_j^p \leftarrow v_j^p - kv_i^p$, we must have $\mathfrak{at}(v_i^p) \leq \mathfrak{at}(v_j^p)$. Thus $\mathfrak{at}(v_j^p - kv_i^p) = \mathfrak{at}(v_j^p)$. It follows that the allow times of the domain basis vectors do not change as we pass from $M_p$ to $M_{p,G}$, i.e. $M_p$ and $M_{p,G}$ have the same number of domain basis vectors corresponding to any particular allow time.*

Now we repeat the same procedure for $\partial_{p+1} : \mathcal{R}_{p+1} \to \mathcal{R}_p$, taking care to use the basis $B_{p,G}$ for $\mathcal{R}_p$. Because we never perform any row operations on $M_{p+1}$, the computations for $M_{p+1}$ do not affect $M_{p,G}$. We claim that for each $1 \leq i \leq N$ and each $p \geq 0$, $B_{p,G}^i$ is a basis for $\Omega_p^i$. The correctness of the procedure amounts to proving this claim. Assuming the claim for now, we obtain compatible bases for the chain complex $\{\Omega_\bullet^i \to \Omega_\bullet^{i+1}\}_{i=1}^N$. Applying the general persistence algorithm with respect to the bases we just found now yields the PPH diagram.

**Correctness** Note that *all* paths become allowed eventually, so $\dim(\Omega_p^N) = \dim(\mathcal{R}_p)$. We claim that $B_{p,G}^i$ is a basis for $\Omega_p^i$, for each $1 \le i \le N$. To see this, fix $1 \le i \le N$ and let $v \in B_{p,G}^i$. By the definition of $B_{p,G}^i$, $\mathfrak{et}(v) \le i$, so $v \in \Omega_p^i$. Each $B_{p,G}^i$ was obtained by performing linear operations on the basis $B_p$ of $\mathcal{R}_p$, so it is a linearly independent collection of vectors in $\Omega_p^i$. Towards a contradiction, suppose $B_{p,G}^i$ does not span $\Omega_p^i$. Let $\tilde{u} \in \Omega_p^i$ be linearly independent from $B_{p,G}^i$, and let $\tilde{v} \in B_{p,G} \setminus B_{p,G}^i$ be linearly dependent on $\tilde{u}$ (such a $\tilde{v}$ exists because $B_{p,G}$ is a basis for $\mathcal{R}_p$).

Consider the basis $B_p^{\tilde{u}}$ obtained from $B_{p,G}$ after replacing $\tilde{v}$ with $\tilde{u}$. Let $M_p^{\tilde{u}}$ denote the corresponding matrix, with the columns arranged in the following order from left to right: the first $|B_{p,G}^i|$ columns agree with those of $M_{p,G}$, the next column is $\partial_p(\tilde{u})$, and the remaining columns appear in the same order that they appear in $M_{p,G}$. Notice that $M_{p,G}$ differs from $M_p^{\tilde{u}}$ by a change of (domain) basis, i.e. a sequence of elementary column operations. Next perform another round of left-to-right Gaussian elimination to arrive at a column echelon form $M_p^u$, where $u$ is the domain basis vector obtained from $\tilde{u}$ after performing all the column operations. Let $B_p^u$ denote the corresponding domain basis. It is a standard theorem in linear algebra that the reduced column echelon form of a matrix is unique. Since $M_{p,G}$ and $M_p^u$ were obtained from $M_p$ via column operations, they both have the same unique reduced column echelon form, and it follows that they have the same pivot positions.

Now we arrive at the contradiction. Since $\tilde{v} \notin B_{p,G}^i$, we must have either $\mathfrak{at}(\tilde{v}) > i$, or $\mathfrak{at}(\partial_p(\tilde{v})) > i$. Suppose first that $\mathfrak{at}(\tilde{v}) > i$. Since $\tilde{u} \in \Omega_p^i$, we must have $\mathfrak{et}(\tilde{u}) \le i$, and so $\mathfrak{at}(\tilde{u}) \le i$. By the way in which we sorted $M_p^{\tilde{u}}$, we know that $u$ is obtained by adding terms from $B_{p,G}^i$ to $\tilde{u}$. Each term in $B_{p,G}^i$ has allow time $\le i$, so $\mathfrak{at}(u) \le i$ by Remark 5.1. But then $B_p^u$ has one more basis vector with allow time $\le i$ than $B_p$, i.e. one fewer basis vector with allow time $> i$. This is a contradiction, because taking linear combinations of linearly independent vectors to arrive at $B_p^u$ can only increase the allow time. Next suppose that $\mathfrak{at}(\partial_p(\tilde{v})) > i$. Then, because $M_{p,G}$ is already reduced, the column of $\tilde{v}$ has a pivot at a height that does *not* belong to $hr(i)$. Now consider $\partial_p(u)$. Suppose first that $\partial_p(u) = 0$. Then the column of $u$ clearly does not have a pivot, and it does not affect the pivots of the columns to its right in $M_p^u$. Thus $M_p^u$ has one fewer pivot than $M_{p,G}$, which is a contradiction because both matrices have the same reduced column echelon form and hence the same pivot positions. Finally, suppose $\partial_p(u) \ne 0$. Since $u$ is obtained from $\tilde{u}$ by reduction, we also have $\mathfrak{at}(\partial_p(u)) \le \mathfrak{at}(\partial_p(\tilde{u})) \le i$. Thus $M_p^u$ has one more pivot at height range $i$ than $M_{p,G}$, which is again a contradiction. Thus $B_{p,G}^i$ spans $\Omega_p^i$. Since $1 \le i \le N$ was arbitrary, the result follows.

**Data structure** Our work shows that left-to-right column reduction is sufficient to obtain compatible bases for the filtered chain complex $\{\Omega_\bullet^i \to \Omega_\bullet^{i+1}\}_{i=1}^N$. As shown in [36], this is precisely the operation needed in computing persistence intervals, so we can compute PPH with little more work. It is known that there are simple ways to optimize the left-to-right persistence computation [14, 5], but in this paper we follow the classical treatment. Following [22, 36], our data structure is a linear array $T$ labeled by the elementary regular $p$-paths, $0 \le p \le D+1$, where $D$ is the dimension up to which homology is computed. For completeness, in Appendix D we show how to modify the algorithms in [36] to obtain PPH.

**Analysis** The running time for this procedure is the same as that of Gaussian elimination over fields, i.e. it is $O(m^3)$, where $m$ is the number of $D$-paths (if we compute persistence up to dimension $D-1$). This number is large: the number of regular $D$-paths over $n$ points is $n(n-1)^D$. Computing persistence also requires $O(m^3)$ running time. Thus, to compute PPH in dimension $D-1$ for a network on $n$ nodes, the worst case running time is $O(n^{3+3D})$.

Compare this with the problem of producing simplicial complexes from networks, and then computing simplicial persistent homology. For a network on $n$ nodes, assume that the simplicial filtration is such that every $D$-simplex on $n$ points eventually enters the filtration (see [17] for such filtrations). The number of $D$-simplices over $n$ points is $\binom{n}{D+1}$, which is of the same order as $n^{D+1}$. Thus computing simplicial persistent homology in dimension $D-1$ via such a filtration (using the general algorithm of [36]) still has complexity $O(n^{3+3D})$.

## 6 Discussion

We presented persistent path homology (PPH) as a novel tool for performing topological data analysis on directed networks. We proved its stability by appealing to a homotopy theory for digraphs. We proved some fundamental characterization results, i.e. that PPH agrees with Čech/Dowker persistence in metric spaces (more generally in symmetric/undirected networks) in dimen-

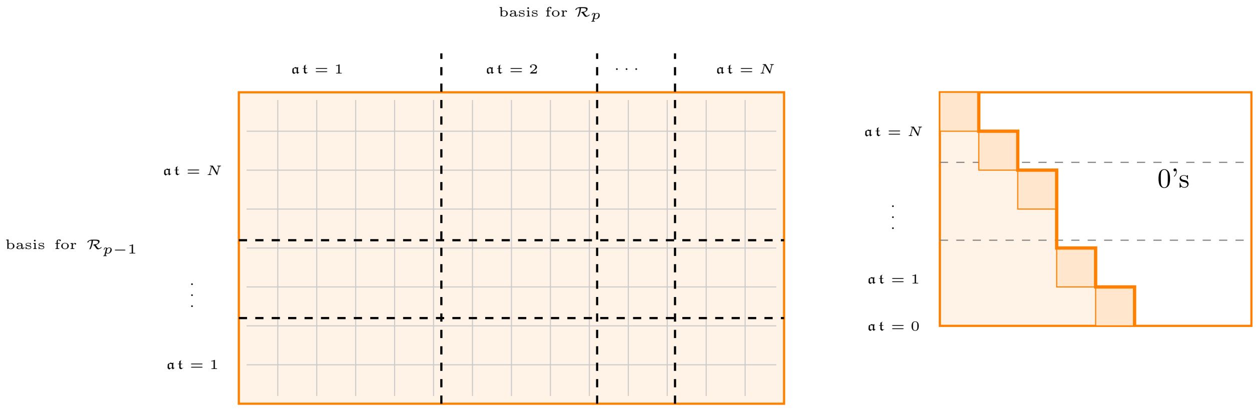



Figure 8: **Left:** The rows and columns of $M_p$ are initially arranged so that the domain and codomain vectors are in increasing and decreasing allow time, respectively. If there are no domain (codomain) vectors having a particular allow time, then the corresponding vertical (horizontal) strip is omitted. **Right:** After converting to column echelon form, the domain vectors of $M_{p,G}$ need not be in the original ordering. But the codomain vectors are still arranged in decreasing allow time.

sion 1 (but not necessarily in higher dimensions), and that PPH recognizes the periodic structure of a cycle network as a directed analogue of a circle, as it should. For this last result, we developed a separate characterization result showing that even on asymmetric/directed networks, PPH agrees with Dowker persistence if the network is square-free.

From the computational standpoint, we proved that the problem of finding a natural basis when computing PPH is automatically solved inside the general persistent homology algorithm, and thus does not cost any additional overhead. Future work includes optimizing the computation of PPH, perhaps in the same way that tools from matroid theory or discrete Morse theory can be employed for efficient computation of simplicial persistent homology.

## References

[1] Michal Adamaszek and Henry Adams. The Vietoris-Rips complexes of a circle. *arXiv preprint arXiv:1503.03669*, 2015.

[2] Michal Adamaszek, Henry Adams, Florian Frick, Chris Peterson, and Corrine Previte-Johnson. Nerve complexes of circular arcs. *Discrete & Computational Geometry*, 56(2):251–273, 2016.

[3] Eric Babson, Hélene Barcelo, Mark Longueville, and Reinhard Laubenbacher. Homotopy theory of graphs. *Journal of Algebraic Combinatorics*, 24(1):31–44, 2006.

[4] Hélene Barcelo, Xenia Kramer, Reinhard Laubenbacher, and Christopher Weaver. Foundations of a connectivity theory for simplicial complexes. *Advances in Applied Mathematics*, 26(2):97–128, 2001.

[5] Ulrich Bauer, Michael Kerber, and Jan Reininghaus. Clear and compress: Computing persistent homology in chunks. In *Topological Methods in Data Analysis and Visualization III*, pages 103–117. Springer, 2014.

[6] Ulrich Bauer and Michael Lesnick. Induced matchings of barcodes and the algebraic stability of persistence. In *Proceedings of the thirtieth annual symposium on Computational geometry*, page 355. ACM, 2014.

[7] Dmitri Burago, Yuri Burago, and Sergei Ivanov. *A Course in Metric Geometry*, volume 33 of *AMS Graduate Studies in Math.* American Mathematical Society, 2001.

[8] Gunnar Carlsson and Vin De Silva. Zigzag persistence. *Foundations of computational mathematics*, 10(4):367–405, 2010.

[9] Gunnar Carlsson, Facundo Mémoli, Alejandro Ribeiro, and Santiago Segarra. Axiomatic construction of hierarchical clustering in asymmetric networks. In *Acoustics, Speech and Signal Processing (ICASSP), 2013 IEEE International Conference on*, pages 5219–5223. IEEE, 2013.

[10] Gunnar Carlsson, Facundo Mémoli, Alejandro Ribeiro, and Santiago Segarra. Hierarchical quasi-clustering methods for asymmetric networks. In *Proceedings of the 31st International Conference on Machine Learning (ICML-14)*, pages 352–360, 2014.

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

[12] Frédéric Chazal, David Cohen-Steiner, Marc Glisse, Leonidas J Guibas, and Steve Y Oudot. Proximity of persistence modules and their diagrams. In *Proceedings of the twenty-fifth annual symposium on Computational geometry*, pages 237–246. ACM, 2009.

[13] Frédéric Chazal, Vin De Silva, Marc Glisse, and Steve

Oudot. The structure and stability of persistence modules. 2012.

[14] Chao Chen and Michael Kerber. Persistent homology computation with a twist. In *Proceedings 27th European Workshop on Computational Geometry*, volume 11, 2011.

[15] Samir Chowdhury and Facundo Mémoli. Metric structures on networks and applications. In *2015 53rd Annual Allerton Conference on Communication, Control, and Computing (Allerton)*, pages 1470–1472, Sept 2015.

[16] Samir Chowdhury and Facundo Mémoli. Distances between directed networks and applications. In *2016 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 6420–6424. IEEE, 2016.

[17] Samir Chowdhury and Facundo Mémoli. Persistent homology of asymmetric networks: An approach based on Dowker filtrations. *arXiv preprint arXiv:1608.05432*, 2016.

[18] Pawel Dlotko, Kathryn Hess, Ran Levi, Max Nolte, Michael Reimann, Martina Scolamiero, Katharine Turner, Eilif Muller, and Henry Markram. Topological analysis of the connectome of digital reconstructions of neural microcircuits. *arXiv preprint arXiv:1601.01580*, 2016.

[19] Herbert Edelsbrunner. Persistent homology: theory and practice. 2014.

[20] Herbert Edelsbrunner and John Harer. *Computational topology: an introduction*. American Mathematical Soc., 2010.

[21] Herbert Edelsbrunner, Grzegorz Jablonski, and Marian Mrozek. The persistent homology of a self-map. *Foundations of Computational Mathematics*, 15(5):1213–1244, 2015.

[22] Herbert Edelsbrunner, David Letscher, and Afra Zomorodian. Topological persistence and simplification. *Discrete and Computational Geometry*, 28(4):511–533, 2002.

[23] Chad Giusti, Robert Ghrist, and Danielle S Bassett. Two's company, three (or more) is a simplex: Algebraic-topological tools for understanding higher-order structure in neural data. *Journal of Computational Neuroscience*, 41(1):doi–10, 2016.

[24] Chad Giusti, Eva Pastalkova, Carina Curto, and Vladimir Itskov. Clique topology reveals intrinsic geometric structure in neural correlations. *Proceedings of the National Academy of Sciences*, 112(44):13455–13460, 2015.

[25] Alexander Grigor'yan, Yong Lin, Yuri Muranov, and Shing-Tung Yau. Homologies of path complexes and digraphs. *arXiv preprint arXiv:1207.2834*, 2012.

[26] Alexander Grigoryan, Yong Lin, Yuri Muranov, and Shing-Tung Yau. Homotopy theory for digraphs. *Pure and Applied Mathematics Quarterly*, 10(4), 2014.

[27] Alexander Grigoryan, Yuri Muranov, and Shing-Tung Yau. Homologies of digraphs and the Künneth formula. 2015.

[28] Nigel J Kalton and Mikhail I Ostrovskii. Distances between Banach spaces. *arXiv preprint math/9709211*, 1997.

[29] Paolo Masulli and Alessandro EP Villa. The topology of the directed clique complex as a network invariant. *SpringerPlus*, 5(1):1–12, 2016.

[30] James R Munkres. *Elements of algebraic topology*, volume 7. Addison-Wesley Reading, 1984.

[31] Giovanni Petri, Paul Expert, Federico Turkheimer, Robin Carhart-Harris, David Nutt, Peter Hellyer, and Francesco Vaccarino. Homological scaffolds of brain functional networks. *Journal of The Royal Society Interface*, 11(101):20140873, 2014.

[32] Giovanni Petri, Martina Scolamiero, Irene Donato, and Francesco Vaccarino. Topological strata of weighted complex networks. *PloS one*, 8(6):e66506, 2013.

[33] Ann Sizemore, Chad Giusti, and Danielle Bassett. Classification of weighted networks through mesoscale homological features. *arXiv preprint arXiv:1512.06457*, 2015.

[34] Ann Sizemore, Chad Giusti, Richard F Betzel, and Danielle S Bassett. Closures and cavities in the human connectome. *arXiv preprint arXiv:1608.03520*, 2016.

[35] Katharine Turner. Generalizations of the Rips filtration for quasi-metric spaces with persistent homology stability results. *arXiv preprint arXiv:1608.00365*, 2016.

[36] Afra Zomorodian and Gunnar Carlsson. Computing persistent homology. *Discrete & Computational Geometry*, 33(2):249–274, 2005.

## A Digraph maps and functoriality

A *digraph map* between two digraphs $G_X = (X, E_X)$ and $G_Y = (Y, E_Y)$ is a map $f : X \to Y$ such that for any edge $(x, x') \in E_X$, we have $f(x) \overset{\rightarrow}{=} f(x')$. Recall that this notation means:

$$\text{either } f(x) = f(x'), \text{ or } (f(x), f(x')) \in E_Y.$$

To extend path homology constructions to a persistent framework, we need to verify the *functoriality* of path homology. As a first step, one must understand how digraph maps transform into maps between vector spaces. Some of the material below can be found in [26]; we contribute a statement and verification of the functoriality of path homology (Proposition A.2) that is central to the PPH framework (Definition 4.1).

Let $X, Y$ be two sets, and let $f : X \to Y$ be a set map. For each dimension $p \in \mathbb{Z}_+$, one defines a map $(f_*)_p : \Lambda_p(X) \to \Lambda_p(Y)$ to be the linearization of the following map on generators: for any generator $[x_0, \ldots, x_p] \in \Lambda_p(X)$,

$$(f_*)_p([x_0, \ldots, x_p]) := [f(x_0), f(x_1), \ldots, f(x_p)].$$

Note also that for any $p \in Z_+$ and any generator

$[x_0, \ldots, x_p] \in \Lambda_p(X)$, we have:

$$\begin{aligned}
&\big((f_*)_{p-1} \circ \partial_p^{\mathrm{nr}}\big)([x_0, \ldots, x_p]) \\
&= \sum_{i=0}^{p} (-1)^i (f_*)_{p-1}\big([x_0, \ldots, \widehat{x_i}, \ldots, x_p]\big) \\
&= \sum_{i=0}^{p} (-1)^i [f(x_0), \ldots, \widehat{f(x_i)}, \ldots, f(x_p)] \\
&= \big(\partial_p^{\mathrm{nr}} \circ (f_*)_p\big)([x_0, \ldots, x_p]).
\end{aligned}$$

It follows that $f_* := ((f_*)_p)_{p\in\mathbb{Z}_+}$ is a chain map from $(\Lambda_p(X), \partial_p^{\mathrm{nr}})_{p\in\mathbb{Z}_+}$ to $(\Lambda_p(Y), \partial_p^{\mathrm{nr}})_{p\in\mathbb{Z}_+}$.

Let $p \in \mathbb{Z}_+$. Note that $(f_*)_p(\mathcal{I}_p(X)) \subseteq \mathcal{I}_p(Y)$, so $(f_*)_p$ descends to a map on quotients

$$(\widetilde{f_*})_p : \Lambda_p(X)/\mathcal{I}_p(X) \to \Lambda_p(Y)/\mathcal{I}_p(Y)$$

which is well-defined. For convenience, we will abuse notation to denote the map on quotients by $(f_*)_p$ as well. Thus we obtain an induced map $(f_*)_p : \mathcal{R}_p(X) \to \mathcal{R}_p(Y)$. Since $p \in \mathbb{Z}_+$ was arbitrary, we get that $f_*$ is a chain map from $(\mathcal{R}_p(X), \partial_p)_{p\in\mathbb{Z}_+}$ to $(\mathcal{R}_p(Y), \partial_p)_{p\in\mathbb{Z}_+}$. The operation of this chain map is as follows: for each $p \in Z_+$ and any generator $[x_0, \ldots, x_p] \in \mathcal{R}_p(X)$,

$$(f_*)_p([x_0, \ldots, x_p]) = [f(x_0), f(x_1), \ldots, f(x_p)]$$

if $f(x_0), f(x_1), \ldots, f(x_p)$ are all distinct, and is 0 otherwise. We refer to $f_*$ as *the chain map induced by the set map* $f : X \to Y$.

Now given two digraphs $G_X = (X, E_X)$, $G_Y = (Y, E_Y)$ and a digraph map $f : G_X \to G_Y$, one may use the underlying set map $f : X \to Y$ to induce a chain map $f_* : \mathcal{R}_\bullet(X) \to \mathcal{R}_\bullet(Y)$. As one could hope, the restriction of the chain map $f_*$ to the chain complex of $\partial$-invariant paths on $G_X$ maps into the chain complex of $\partial$-invariant paths on $G_Y$, and moreover, is a chain map. We state this result as a proposition below, and provide a reference for the proof.

Proposition A.1. (Theorem 2.10, [26]) *Let $G_X = (X, E_X), G_Y = (Y, E_Y)$ be two digraphs, and let $f : G_X \to G_Y$ be a digraph map. Let $f_* : \mathcal{R}_\bullet(X) \to \mathcal{R}_\bullet(Y)$ denote the chain map induced by the underlying set map $f : X \to Y$. Let $(\Omega_p(G_X), \partial_p^{G_X})_{p\in\mathbb{Z}_+}$, $(\Omega_p(G_Y), \partial_p^{G_Y})_{p\in\mathbb{Z}_+}$ denote the chain complexes of the $\partial$-invariant paths associated to each of these digraphs. Then $(f_*)_p(\Omega_p(G_X)) \subseteq \Omega_p(G_Y)$ for each $p \in \mathbb{Z}_+$, and the restriction of $f_*$ to $\Omega_\bullet(G_X)$ is a chain map.*

Henceforth, given two digraphs $G, G'$ and a digraph map $f : G \to G'$, we refer to the chain map $f_*$ given by Proposition A.1 as the *chain map induced by the digraph map* $f$. Because $f_*$ is a chain map, we then obtain an induced linear map $(f_\#)_p : H_p(G) \to H_p(G')$ for each $p \in \mathbb{Z}_+$.

The preceding concepts are necessary for developing the theory of path homology. We use this set up to state and prove the following result, which is used in defining PPH (Definition 4.1) and also for proving stability (Theorem 4.1).

Proposition A.2. **Functoriality of path homology.** *Let $G, G', G''$ be three digraphs.*

1. *Let $\mathrm{id}_G : G \to G$ be the identity digraph map. Then $(\mathrm{id}_{G\#})_p : H_p(G) \to H_p(G)$ is the identity linear map for each $p \in \mathbb{Z}_+$.*
2. *Let $f : G \to G', g : G' \to G''$ be digraph maps. Then $((g \circ f)_\#)_p = (g_\#)_p \circ (f_\#)_p$ for any $p \in \mathbb{Z}_+$.*

*Proof.* Let $p \in \mathbb{Z}_+$. In each case, it suffices to verify the operations on generators of $\Omega_p(G)$. Let $[x_0, \ldots, x_p] \in \Omega_p(G)$. We will write $\mathrm{id}_{G*}$ to denote the chain map induced by the digraph map $\mathrm{id}_G$. First note that

$$\begin{aligned}
(\mathrm{id}_{G*})_p([x_0, \ldots, x_p]) &= [\mathrm{id}_G(x_0), \ldots, \mathrm{id}_G(x_p)] \\
&= [x_0, \ldots, x_p].
\end{aligned}$$

It follows that $(\mathrm{id}_{G*})_p$ is the identity linear map on $\Omega_p(G)$, and thus $(\mathrm{id}_{G\#})_p$ is the identity linear map on $H_p(G)$. For the second claim, suppose first that $g(f(x_0)), \ldots, g(f(x_p))$ are all distinct. This implies that $f(x_0), \ldots, f(x_p)$ are also all distinct, and we observe:

$$\begin{aligned}
((g \circ f)_*)_p([x_0, \ldots, x_p]) &= [g(f(x_0)), \ldots, g(f(x_p))] \\
&= (g_*)_p([f(x_0), \ldots, f(x_p)]) \\
&= (g_*)_p\big((f_*)_p([x_0, \ldots, x_p])\big).
\end{aligned}$$

Next suppose that for some $0 \le i \ne j \le p$, we have $g(f(x_i)) = g(f(x_j))$. Then we obtain:

$$((g \circ f)_*)_p([x_0, \ldots, x_p]) = 0 = (g_*)_p\big((f_*)_p([x_0, \ldots, x_p])\big).$$

It follows that $((g \circ f)_*)_p = (g_*)_p \circ (f_*)_p$. The statement of the proposition now follows.

## B Interleaving distance and stability of persistent path homology.

Given $\varepsilon \ge 0$, two $\mathbb{R}$-indexed persistent vector spaces $\mathcal{V} = \{V^\delta \xrightarrow{\nu_{\delta,\delta'}} V^{\delta'}\}_{\delta\le\delta'}$ and $\mathcal{U} = \{U^\delta \xrightarrow{\mu_{\delta,\delta'}} U^{\delta'}\}_{\delta\le\delta'}$ are said to be *$\varepsilon$-interleaved* [12, 6] if there exist two families of linear maps

$$\{\varphi_\delta : V^\delta \to U^{\delta+\varepsilon}\}_{\delta\in\mathbb{R}},$$
$$\{\psi_\delta : U^\delta \to V^{\delta+\varepsilon}\}_{\delta\in\mathbb{R}}$$

such that the following equalities hold for all $\delta' \geq \delta \in \mathbb{R}$:

$$\begin{aligned}
\varphi_{\delta'} \circ \nu_{\delta,\delta'} &= \mu_{\delta+\varepsilon,\delta'+\varepsilon} \circ \varphi_\delta \\
\nu_{\delta+\varepsilon,\delta'+\varepsilon} \circ \psi_\delta &= \psi_{\delta'} \circ \mu_{\delta,\delta'} \\
\nu_{\delta,\delta+2\varepsilon} &= \psi_{\delta+\varepsilon} \circ \varphi_\delta \\
\mu_{\delta,\delta+2\varepsilon} &= \varphi_{\delta+\varepsilon} \circ \psi_\delta .
\end{aligned}$$

The purpose of introducing $\varepsilon$-interleavings is to define a pseudometric on the collection of persistent vector spaces. The *interleaving distance* between two $\mathbb{R}$-indexed persistent vector spaces $\mathcal{V}, \mathcal{U}$ is given by:

$$d_{\mathrm{I}}(\mathcal{U}, \mathcal{V}) := \inf \{\varepsilon \geq 0 : \mathcal{U} \text{ and } \mathcal{V} \text{ are } \varepsilon\text{-interleaved}\} .$$

One can verify that this definition induces a pseudometric on the collection of persistent vector spaces [12, 6]. The interleaving distance can then be related to the bottleneck distance as follows:

THEOREM B.1. (ALGEBRAIC STABILITY, [12]) *Let $\mathcal{U}, \mathcal{V}$ be two $\mathbb{R}$-indexed persistent vector spaces. Then,*

$$d_{\mathrm{B}}(\mathrm{Dgm}(\mathcal{U}), \mathrm{Dgm}(\mathcal{V})) \leq d_{\mathrm{I}}(\mathcal{U}, \mathcal{V}).$$

A special case of the Algebraic Stability Theorem is the Persistence Equivalence Theorem [20]. This particular version follows from the *isometry theorem* [6], and we refer the reader to [13, Chapter 5] for an expanded presentation of this material.

THEOREM B.2. (PERSISTENCE EQUIVALENCE) *Consider two persistent vector spaces $\mathcal{U} = \{U^\delta \xrightarrow{\mu_{\delta,\delta'}} U^{\delta'}\}_{\delta \leq \delta' \in \mathbb{R}}$ and $\mathcal{V} = \{V^\delta \xrightarrow{\nu_{\delta,\delta'}} V^{\delta'}\}_{\delta \leq \delta' \in \mathbb{R}}$ with connecting maps $f_\delta : U^\delta \to V^{\delta'}$.*

$$\begin{tikzcd}
\cdots \arrow[r] & U^\delta \arrow[r] \arrow[d, "f_\delta"] & U^{\delta'} \arrow[r] \arrow[d, "f_{\delta'}"] & U^{\delta''} \arrow[r] \arrow[d, "f_{\delta''}"] & \cdots \\
\cdots \arrow[r] & V^\delta \arrow[r] & V^{\delta'} \arrow[r] & V^{\delta''} \arrow[r] & \cdots
\end{tikzcd}$$

*If the $f_\delta$ are all isomorphisms and each square in the diagram above commutes, then:*

$$\mathrm{Dgm}(\mathcal{U}) = \mathrm{Dgm}(\mathcal{V}).$$

## C PPH and Dowker persistence

DEFINITION C.1. **Type I and II Dowker simplices** *Let $(X, A_X) \in \mathcal{N}$, fix $\delta \in \mathbb{R}$, and let $\sigma$ be a simplex in $\mathfrak{D}^{\mathrm{si}}_{\delta,X}$. Then we define $\sigma$ to be a* Type I *simplex if some $x \in \sigma$ is a $\delta$-sink for $\sigma$. Otherwise, $\sigma$ is a* Type II *simplex. Notice that if $\sigma$ is a Type II simplex, then there exists $x \notin \sigma$ such that $x$ is a $\delta$-sink for $\sigma$.*

*We define analogous notions at the chain complex level: a chain $\sigma \in C_\bullet(\mathfrak{D}^{\mathrm{si}}_{\delta,X})$ is of Type I if each element in its expression corresponds to a Type I simplex. Otherwise, $\sigma$ is of Type II.*

LEMMA C.1. (PROPOSITION 2.9, [26]) *Let $\mathfrak{G}$ be a finite digraph. Then any $v \in \Omega_2(\mathfrak{G})$ is a linear combination of the following three types of $\partial$-invariant 2-paths:*

1. *$aba$ with edges $(a, b), (b, a)$ (a double edge),*
2. *$abc$ with edges $(a, b), (b, c), (a, c)$ (a triangle), and*
3. *$abc - adc$ with edges $(a, b), (b, c), (a, d), (d, c)$, where $a \neq c$ and $(a, c)$ is not an edge (a long square).*

LEMMA C.2. (PARITY LEMMA) *Fix a simplicial complex $K$ and a field $\mathbb{Z}/p\mathbb{Z}$ for some prime $p$. Let $w := \sum_{i \in I} b_i \tau_i$ be a 2-chain in $C_2(K)$ where $I$ is a finite index set, each $b_i \in \mathbb{Z}/p\mathbb{Z}$, and each $\tau_i$ is a 2-simplex in $K$. Let $\sigma$ be a 1-simplex contained in some $\tau_i$ such that $\sigma$ does not appear in $\partial^\Delta_2(w)$. Define $J_\sigma := \{j \in I : \sigma \text{ a face of } \tau_j\}$. Then there exists $n(\sigma) \in \mathbb{N}$ such that:*

$$w = \sum_{i \in I \setminus J_\sigma} b_i \tau_i + \sum_{j=1}^{n(\sigma)} (\tau_j^+ + \tau_j^-),$$

*where $\partial^\Delta_2(\tau_j^+ + \tau_j^-)$ is independent of $\sigma$ for each $1 \leq j \leq n(\sigma)$.*

*Proof.* [Proof of Lemma C.2] Since we are working over $\mathbb{Z}/p\mathbb{Z}$, we adopt the convention that $b_i \in \{0, 1, \dots, p-1\}$ for each $i \in I$. Then for each $j \in J_\sigma$, we know that $\partial^\Delta_2(\tau_j)$ contributes either $+\sigma$ or $-\sigma$ with multiplicity $b_j$. Write $w = \sum_{i \in I \setminus J_\sigma} b_i \tau_i + \sum_{j \in J_\sigma} b_j \tau_j$.

Since $\sigma$ is not a summand of $\partial^\Delta_2(w)$, it follows that $\sum_{j \in J_\sigma} b_j = 0$. Define:

$$\begin{aligned}
J_\sigma^+ &:= \{j \in J_\sigma : \tau_j \text{ contributes } +\sigma\} , \\
J_\sigma^- &:= \{j \in J_\sigma : \tau_j \text{ contributes } -\sigma\} .
\end{aligned}$$

Then $w = \sum_{i \in I \setminus J_\sigma} b_i \tau_i + \sum_{j \in J_\sigma^+} b_j \tau_j + \sum_{j \in J_\sigma^-} b_j \tau_j$.

Also define $k := |J_\sigma^+|$, and enumerate $J_\sigma^+$ as $\{j_1, \dots, j_k\}$. Write $n^+(\sigma) := \sum_{m=1}^{k} b_{j_k}$, where the sum is taken over $\mathbb{Z}$ (not $\mathbb{Z}/p\mathbb{Z}$). Next define a finite sequence $(\tau_1^+, \dots, \tau_{n^+(\sigma)}^+)$ as follows:

$$\begin{aligned}
\tau_i^+ &:= \tau_{j_1} \text{ for } i \in \{1, \dots, b_{j_1}\} , \\
\tau_i^+ &:= \tau_{j_2} \text{ for } i \in \{b_{j_1} + 1, \dots, b_{j_1} + b_{j_2}\} , \dots , \\
\tau_i^+ &:= \tau_{j_k} \text{ for } i \in \left\{ \sum_{m=1}^{k-1} b_{j_m} + 1, \dots, \sum_{m=1}^{k} b_{j_m} \right\} .
\end{aligned}$$

Here the indexing element $i$ is of course taken over $\mathbb{Z}$ and not $\mathbb{Z}/p\mathbb{Z}$. Similarly we define a sequence $(\tau_1^-, \ldots, \tau_{n^-(\sigma)}^-)$. Then $w = \sum_{i\in I\setminus J_\sigma} b_i\tau_i + \sum_{m=1}^{n^+(\sigma)} \tau_m^+ + \sum_{m=1}^{n^-(\sigma)} \tau_m^-$.

The expression for $\partial_2^\Delta(w)$ contains $+\sigma$ with multiplicity $n^+(\sigma)$ and $-\sigma$ with multiplicity $n^-(\sigma)$, such that the total multiplicity is 0, i.e. is a multiple of $p$. Thus we have $n^+(\sigma) - n^-(\sigma) \in p\mathbb{Z}$. There are two cases: either $n^+(\sigma) \geq n^-(\sigma)$ or $n^+(\sigma) \leq n^-(\sigma)$. Both cases are similar, so we consider the first. Let $q$ be a nonnegative integer such that $n^+(\sigma) = n^-(\sigma) + pq$. We pad the $\tau^-$ sequence by defining $\tau_i^- := \tau_{n^-(\sigma)}^-$ for $i \in \{n^-(\sigma)+1, \ldots, n^-(\sigma)+pq\}$. Then we have:

$$\begin{aligned} w &= \sum_{i\in I\setminus J_\sigma} b_i\tau_i + \sum_{m=1}^{n^+(\sigma)} \tau_m^+ + \sum_{m=1}^{n^-(\sigma)} \tau_m^- \\ &= \sum_{i\in I\setminus J_\sigma} b_i\tau_i + \sum_{m=1}^{n^+(\sigma)} \tau_m^+ + \sum_{m=1}^{n^-(\sigma)} \tau_m^- + \sum_{m=n^-(\sigma)+1}^{n^-(\sigma)+pq} \tau_m^- \\ &= \sum_{i\in I\setminus J_\sigma} b_i\tau_i + \sum_{m=1}^{n^+(\sigma)} \tau_m^+ + \sum_{m=1}^{n^+(\sigma)} \tau_m^-. \end{aligned}$$

THEOREM C.1. *Let $\mathcal{X} = (X, A_X) \in \mathcal{N}$ be a square-free network, and fix $\mathbb{K} = \mathbb{Z}/p\mathbb{Z}$ for some prime $p$. Then* $\mathrm{Dgm}_1^\Xi(\mathcal{X}) = \mathrm{Dgm}_1^{\mathfrak{D}}(\mathcal{X})$.

*Proof.* [Proof of Theorem 4.3] Let $\delta \in \mathbb{R}$. First we wish to find an isomorphism $\varphi_\delta : H_1^\Xi(\mathfrak{G}_X^\delta) \to H_1^\Delta(\mathfrak{D}_{\delta,X}^{\mathrm{si}})$. We begin with the basis $B$ for $\Omega_1(\mathfrak{G}_X^\delta)$. We claim that $B$ is just the collection of allowed 1-paths in $\mathfrak{G}_X^\delta$. To see this, let $ab$ be an allowed 1-path. Then $\partial_1(ab) = b - a$, which is allowed because the vertices $a$ and $b$ are automatically allowed. Thus $ab \in \Omega_1(\mathfrak{G}_X^\delta)$, and so $B$ generates $\Omega_1(\mathfrak{G}_X^\delta)$.

Whenever $ab$ is an allowed 1-path, we have a directed edge $(a, b)$ in $\mathfrak{G}_X^\delta$, and so $A_X(a, b) \leq \delta$ by the definition of $\mathfrak{G}_X^\delta$. Thus the simplex $[a, b]$ belongs to $\mathfrak{D}_{\delta,X}^{\mathrm{si}}$, with $b$ as a $\delta$-sink. Hence $[a, b]$ is a 1-chain in $C_1(\mathfrak{D}_{\delta,X}^{\mathrm{si}})$. Define a map $\widetilde{\varphi}_\delta : \Omega_1(\mathfrak{G}_X^\delta) \to C_1(\mathfrak{D}_{\delta,X}^{\mathrm{si}})$ by setting $\widetilde{\varphi}_\delta(ab) = [a, b]$ and extending linearly. The image of $\widetilde{\varphi}_\delta$ restricted to $B$ is linearly independent because any linear dependence relation would contradict the independence of $B$. Furthermore, $\widetilde{\varphi}_\delta$ induces a map $\widetilde{\varphi}'_\delta : \ker(\partial_1^\Xi) \to \ker(\partial_1^\Delta)$. We need to check that this descends to a map $\varphi_d : \ker(\partial_1^\Xi)/\operatorname{im}(\partial_2^\Xi) \to \ker(\partial_1^\Delta)/\operatorname{im}(\partial_2^\Delta)$ on quotients. To see this, we need to verify that $\widetilde{\varphi}'_\delta(\operatorname{im}(\partial_2^\Xi)) \subseteq \operatorname{im}(\partial_2^\Delta)$.

By Lemma C.1, we have a complete characterization of $\Omega_2(\mathfrak{G}_X^\delta)$. Thus we know that any element of $\operatorname{im}(\partial_2^\Xi)$ is of the form $ba + ab$, $bc - ac + ab$, or $bc + ab - dc - ad$. In the first case, we have $\widetilde{\varphi}'_\delta(ba + ab) = [b, a] + [a, b] = [b, a] - [b, a] = 0 \in \operatorname{im}(\partial_2^\Delta)$. The next case corresponds to the situation where we have $abc \in \Omega_2(\mathfrak{G}_X^\delta)$ with edges $(a, b), (b, c), (a, c)$ in $\mathfrak{G}_X^\delta$. In this case, $[a, b, c]$ is a 2-simplex in $\mathfrak{D}_{\delta,X}^{\mathrm{si}}$, with $c$ as a $\delta$-sink. Thus $[b, c] - [a, c] + [a, b] = \widetilde{\varphi}'_\delta(bc - ac + ab)$ belongs to $\operatorname{im}(\partial_2^\Delta)$.

The final case cannot occur because $\mathfrak{G}_X^\delta$ is square-free. It follows that $\widetilde{\varphi}'_\delta(\operatorname{im}(\partial_2^\Xi)) \subseteq \operatorname{im}(\partial_2^\Delta)$, and so we obtain a well-defined map $\varphi_\delta : H_1^\Xi(\mathfrak{G}_X^\delta) \to H_1^\Delta(\mathfrak{D}_{\delta,X}^{\mathrm{si}})$.

Next we check that $\varphi_\delta$ is injective. Let $v = \sum_{i=0}^k a_i\sigma_i \in \ker(\varphi_\delta)$, where the $a_i$ terms belong to the field $\mathbb{K}$ and each $\sigma_i$ is a 1-path in $\mathfrak{G}_X^\delta$. Then $\varphi_\delta(v) = \varphi_\delta(\sum_{i=0}^k a_i\sigma_i) = \partial_2^\Delta(\sum_{j=0}^m b_j\tau_j)$, where the $b_j$ terms belong to $\mathbb{K}$ and each $\tau_j$ is a 2-simplex in $\mathfrak{D}_{\delta,X}^{\mathrm{si}}$.

*Claim.* $w := \sum_{j=0}^m b_j\tau_j$ is homologous to a 2-cycle $\sum_{k=0}^n b'_k\tau'_k$ in $C_2(\mathfrak{D}_{\delta,X}^{\mathrm{si}})$, where each $\tau'_k$ is of the form $[a, b, c]$ and $abc$ is a triangle in $\mathfrak{G}_X^\delta$.

Suppose the claim is true. Then we immediately see that $v \in \operatorname{im}(\partial_2^\Xi)$. Thus $\ker(\varphi_\delta) = \operatorname{im}(\partial_2^\Xi)$, and hence $\ker(\varphi_\delta)$ is trivial in $H_1^\Xi(\mathfrak{G}_X^\delta)$. This shows that $\varphi_\delta$ is injective.

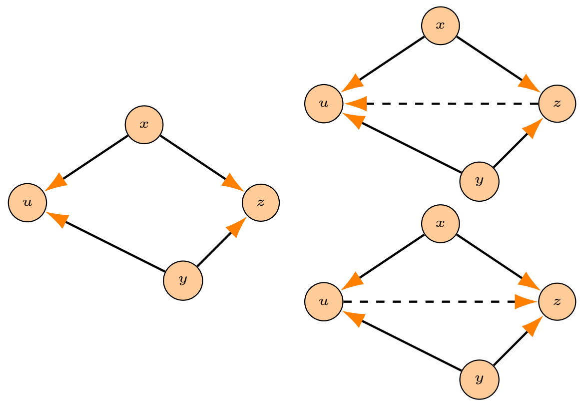
Let us now prove the claim. Suppose $\tau_j$ is a Type II simplex, for some $0 \leq j \leq m$. Write $\tau_j = [u, x, y]$. Then there exists $z \in X$ such that $z$ is a $\delta$-sink for $\tau_j$. But then $[u, x, y, z] \in \mathfrak{D}_{\delta,X}^{\mathrm{si}}$, and $\partial_3^\Delta([u, x, y, z]) = [x, y, z] - [u, y, z] + [u, x, z] - [u, x, y]$. Since $\partial_2^\Delta \circ \partial_3^\Delta = 0$, it follows that $[u, x, y]$ is homologous to $[x, y, z] - [u, y, z] + [u, x, z]$, each of which is a Type I simplex. Using this argument, we first replace all Type II simplices in $w$ by Type I simplices.

Next let $\tau$ be a Type I simplex in the rewritten expression for $w$. By taking a permutation and appending a $(-1)$ coefficient if needed, we can write $\tau = [x, y, z]$, where $z$ is the $\delta$-sink for $\tau$. Thus $(x, z), (y, z)$ are edges in $\mathfrak{G}_X^\delta$. If $(x, y)$ or $(y, x)$ is also an edge, then $xyz$ is a triangle, and we are done. Suppose that neither is an edge, i.e. neither of $xy, yx$ is in $\Omega_1(\mathfrak{G}_X^\delta)$. Then, since $xy$ is not a summand of $v$, we know that $[x, y]$ is not a summand of $\varphi_\delta(v)$. Thus we are in the setting of Lemma C.2, because $\partial_2^\Delta(w) = \varphi_\delta(v)$. Define $J := \{0 \leq j \leq m : [x, y] \text{ a face of } \tau_j\}$. By applying Lemma C.2, we can rewrite $w$:

$$w = \sum_{i\notin J, i=0}^{m} b_i\tau_i + \sum_{j=1}^{n([x,y])} (\tau_j^+ + \tau_j^-),$$

where all the summands of $w$ containing $[x, y]$ as a face are paired in the latter term. Each $\tau^+ + \tau^-$ summand has the following form: $[x, y]$ is a face of both $\tau^+$ and $\tau^-$, and both $\tau^+$ and $\tau^-$ are Type I simplices. Fix

$1 \leq j \leq n([x,y])$. Then for some $z, u \in X$, $\tau_j^+ = [x,y,z]$ and $\tau_j^- = [x,u,y]$ have the following arrangement:




Since $(X, A_X)$ is square-free, we must have at least one of the edges $(z,u)$ or $(u,z)$ in $\mathfrak{G}_X^\delta$. Suppose $(z,u)$ is an edge. Because we have

$$\partial_3^\Delta([x,y,z,u]) = [y,z,u] - [x,z,u] + [x,y,u] - [x,y,z],$$

it follows that $[x,y,z] - [x,y,u] = [x,y,z] + [x,u,y] = \tau_j^+ + \tau_j^-$ is homologous to $[y,z,u] - [x,z,u]$, where $yzu$ and $xzu$ are both triangles in $\mathfrak{G}_X^\delta$.

For the other case, suppose $(u,z)$ is an edge. Because we have $\partial_3^\Delta([x,y,u,z]) = [y,u,z] - [x,u,z] + [x,y,z] - [x,y,u]$, we again know that $\tau_j^+ + \tau_j^-$ is homologous to $[x,u,z] - [y,u,z]$, where $xuz$ and $yuz$ are both triangles in $\mathfrak{G}_X^\delta$.

We can repeat this argument to replace all summands of $w$ containing $[x,y]$ as a face. Since $\tau = [x,y,z]$ was arbitrary, this proves the claim.

It remains to verify that $\varphi_\delta$ is surjective. Let $v = \sum_{i=0}^m a_i \tau_i$ be a 1-cycle in $C_1(\mathfrak{D}_{\delta,X}^{\mathrm{si}})$. First we wish to show that $v$ is homologous to a 1-cycle $v' = \sum_{i=0}^n b_i \tau_i'$ of Type I. Let $\tau_i$ be a Type II simplex in the expression for $v$, for some $0 \leq i \leq m$. Write $\tau_i = [x,y]$, and let $z$ be a $\delta$-sink for $\tau_i$. Then $[x,y,z]$ is a simplex in $\mathfrak{D}_{\delta,X}^{\mathrm{si}}$, and $\partial_2^\Delta([x,y,z]) = [y,z] - [x,z] + [x,y]$. Thus $[x,y]$ is homologous to $[x,z] - [y,z]$, each of which is a Type I simplex. This argument shows that $v$ is homologous to a 1-cycle $v'$ of Type I.

Next let $\tau'$ be a 1-simplex in the expression for $v'$. Write $\tau' = [x,y]$. If $x$ is the $\delta$-sink for $\tau'$, then we replace the $\tau' = [x,y]$ in the expression of $v'$ with $-[y,x]$. This does not change $v$, since we have $\tau' = [x,y] = -[y,x]$ in $C_1(\mathfrak{D}_{\delta,X}^{\mathrm{si}})$. After repeating this procedure for each element of $v'$, we obtain a rewritten expression for $v'$ in terms of elements $[x,y]$ where $y$ is the $\delta$-sink for $[x,y]$. Let $v' = \sum_{i=0}^n b_i'[x_i,y_i]$ denote this new expression.

Finally, observe that for each $[x_i,y_i]$ in the rewritten expression for $v'$, we also have $(x_i,y_i)$ as an edge in $\mathfrak{G}_X^\delta$. Thus $\sum_{i=0}^n b_i' x_i y_i$ is a 1-cycle in $H_1^\Xi(\mathfrak{G}_X^\delta)$ that is mapped to $v'$ by $\varphi_\delta$. It follows that $\varphi_\delta$ is surjective, and hence is an isomorphism.

To complete the proof, let $\delta \leq \delta' \in \mathbb{R}$. Consider the inclusion maps $\iota_{\mathfrak{G}} : \mathfrak{G}_X^\delta \hookrightarrow \mathfrak{G}_X^{\delta'}$ and $\iota_{\mathfrak{D}} : \mathfrak{D}_{\delta,X}^{\mathrm{si}} \hookrightarrow \mathfrak{D}_{\delta',X}^{\mathrm{si}}$, and let $(\iota_{\mathfrak{G}})_\#$, $(\iota_{\mathfrak{D}})_\#$ denote the induced maps at the respective homology levels. Let $v = \sum_{i=0}^n a_i x_i y_i$ be a 1-cycle in $H_1^\Xi(\mathfrak{G}_X^\delta)$. Then we have:

$$\begin{aligned}
&(\varphi_{\delta'} \circ (\iota_{\mathfrak{G}})_\#)\left(\sum_{i=0}^n a_i x_i y_i\right) \\
&= \varphi_{\delta'}\left(\sum_{i=0}^n a_i x_i y_i\right) \\
&= \sum_{i=0}^n a_i [x_i, y_i] \\
&= (\iota_{\mathfrak{D}})_\#\left(\sum_{i=0}^n a_i [x_i, y_i]\right) \\
&= ((\iota_{\mathfrak{G}})_\# \circ \varphi_\delta)\left(\sum_{i=0}^n a_i [x_i, y_i]\right).
\end{aligned}$$

Thus the necessary commutativity relation holds, and the theorem follows by the Persistence Equivalence Theorem.

Theorem C.2. *Let $G_n$ be a cycle network for some integer $n \geq 3$. Fix a field $\mathbb{K} = \mathbb{Z}/p\mathbb{Z}$ for some prime $p$. Then* $\mathrm{Dgm}_1^\Xi(G_n) = \{(1, \lceil n/2 \rceil)\}$.

*Proof.* [Proof of Theorem 4.4] From [17], we know that $\mathrm{Dgm}_1^{\mathfrak{D}}(G_n) = \{(1, \lceil n/2 \rceil)\}$. Thus by Theorem 4.3, it suffices to show that $G_n$ is square-free. Suppose $n \geq 4$, and let $a, b, c, d$ be four nodes that appear in $G_n$ in clockwise order. First let $\delta \in \mathbb{R}$ be such that $(a,b), (b,c), (a,d), (d,c)$ are edges in $\mathfrak{G}_{G_n}^\delta$. Then $\omega_{G_n}(d,c) \leq \delta$, and because of the clockwise orientation $d \preceq a \preceq c$, we automatically $\omega_{G_n}(a,c) \leq \delta$. Hence $(a,c)$ is an edge in $\mathfrak{G}_{G_n}^\delta$, and so the subgraph induced by $a,b,c,d$ is not a long square.

Next suppose $\delta \in \mathbb{R}$ is such that $(a,b), (c,b), (a,d), (c,d)$ are edges in $\mathfrak{G}_{G_n}^\delta$. Since $\omega_{G_n}(c,b) \leq \delta$ and $c \preceq a \preceq b$ in $G_n$, we have $\omega_{G_n}(c,a) \leq \delta$. Hence $(c,a)$ is an edge in $\mathfrak{G}_{G_n}^\delta$, and so the subgraph induced by $a,b,c,d$ is not a short square.

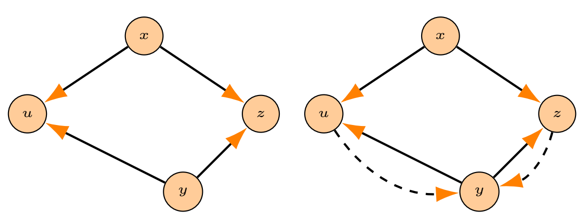
Theorem C.3. *Let $\mathcal{X} = (X, A_X) \in \mathcal{N}$ be a symmetric network, and fix $\mathbb{K} = \mathbb{Z}/p\mathbb{Z}$ for some prime $p$. Then* $\mathrm{Dgm}_1^\Xi(\mathcal{X}) = \mathrm{Dgm}_1^{\mathfrak{D}}(\mathcal{X})$.

*Proof.* [Proof of Theorem 4.2] The proof is similar to that of Theorem 4.3; instead of repeating all details, we will show how the argument changes when the square-free assumption is replaced by the symmetry assumption. Let $\delta \in \mathbb{R}$, and consider the map $\widetilde{\varphi}'_\delta : \ker(\partial_1^{\Xi}) \to \ker(\partial_1^{\Delta})$ defined as in Theorem 4.3. As before, we need to check that this descends to a map $\varphi_d : \ker(\partial_1^{\Xi})/\operatorname{im}(\partial_2^{\Xi}) \to \ker(\partial_1^{\Delta})/\operatorname{im}(\partial_2^{\Delta})$ on quotients. For this we need to verify that $\widetilde{\varphi}'_\delta(\operatorname{im}(\partial_2^{\Xi})) \subseteq \operatorname{im}(\partial_2^{\Delta})$.

By Lemma C.1, we know that any element of $\operatorname{im}(\partial_2^{\Xi})$ is of the form $ba+ab$, $bc-ac+ab$, or $bc+ab-dc-ad$. For the first two cases, we can repeat the argument used in Theorem 4.3. The final case corresponds to the situation where we have a long square in $\mathfrak{G}_X^\delta$ consisting of edges $(a,b),(b,c),(a,d)$, and $(d,c)$. This gives the 2-chain $abc-adc$. Now by the symmetry condition, we also have edges $(c,d)$ and $(c,b)$. Thus $[a,b,c]$ is a 2-simplex in $\mathfrak{D}^{\text{si}}_{\delta,X}$, with $b$ as a $\delta$-sink, and $[a,d,c]$ is a 2-simplex with $d$ as a $\delta$-sink. Hence $[a,b,c]-[a,d,c]$ is a 2-chain in $C_2(\mathfrak{D}^{\text{si}}_{\delta,X})$. Thus $\widetilde{\varphi}'_\delta(bc+ab-dc-ad) = [b,c]+[a,b]-[d,c]-[a,d]$ belongs to $\operatorname{im}(\partial_2^{\Delta})$. Thus we obtain a well-defined map $\varphi_\delta : H_1^{\Xi}(\mathfrak{G}_X^\delta) \to H_1^{\Delta}(\mathfrak{D}^{\text{si}}_{\delta,X})$.

Next we need to check that $\varphi_\delta$ is injective. As in Theorem 4.3, let $v \in \ker(\varphi_\delta)$. Then $\varphi_\delta(v) = \varphi_\delta(\sum_{i=0}^k a_i\sigma_i) = \partial_2^{\Delta}(\sum_{j=0}^m b_j\tau_j)$, where the $a_i, b_j$ terms belong to the field $\mathbb{K}$, each $\sigma_i$ is a 1-path in $\mathfrak{G}_X^\delta$, and each $\tau_j$ is a 2-simplex in $\mathfrak{D}^{\text{si}}_{\delta,X}$. We proceed by proving an analogue of the claim in Theorem 4.3 in the symmetric setting. Write $w := \sum_{j=0}^m b_j\tau_j$. We need to show that $w$ is homologous to a 2-cycle $\sum_{k=0}^n b'_k\tau'_k$ in $C_2(\mathfrak{D}^{\text{si}}_{\delta,X})$, where each $\tau'_k$ is of the form $[a,b,c]$ and $abc$ is either a triangle or part of a square in $\mathfrak{G}_X^\delta$.

As in the proof of Theorem 4.3, we first replace all Type II simplices in $w$ by Type I simplices. Next let $\tau = [x,y,z]$ be a Type I simplex in $w$, and suppose $z$ is the $\delta$-sink for $\tau$, but neither of $(x,y),(y,x)$ is an edge. As in the proof of Theorem 4.3, we apply Lemma C.2 to separate the summands of $w$ containing $[x,y]$ as a face into pairs of the form $(\tau^+ + \tau^-)$. Writing $\tau^+ = [x,z,y]$ and $\tau^- = [x,y,u]$, we obtain the following arrangement:




By the symmetry assumption, $(z,y)$ and $(u,y)$ are also edges in $\mathfrak{G}_X^\delta$, and so $xuy$, $xzy$ are both allowed 2-paths. Since $\tau^- = [x,y,u] = -[x,u,y]$, we can replace $\tau^+ + \tau^-$ by $[x,z,y]-[x,u,y]$, where $xzy - xuy$ is a square in $\mathfrak{G}_X^\delta$. Proceeding in this way, we replace each summand of $w$ containing $[x,y]$ as a face. We repeat this argument for each choice of $\tau = [x,y,z]$ in the expression for $w$.

Finally, we obtain an expression of $w$ such that there exists $v' \in \Omega_2(\mathfrak{G}_X^\delta)$ satisfying $\varphi_\delta(v') = w$. Then we have $\partial_2^{\Xi}(v') = v$, and so $v = 0$ in $H_1^{\Xi}(\mathfrak{G}_X^\delta)$. Thus $\varphi_\delta$ is injective.

We omit the remainder of the argument, because it is a repeat of the corresponding part of the proof of Theorem 4.3. In summary, it turns out that $\varphi_\delta$ is surjective, hence an isomorphism, and furthermore that it commutes with the linear maps induced by the canonical inclusions. This concludes the proof.

## D The modified persistence algorithm

We use the notation introduced in §5. By our observations in §5, computing bases for the filtered chain complex $\{\Omega^i_\bullet \to \Omega^{i+1}_\bullet\}_{i=1}^N$ can be done simultaneously while performing the column reduction operations needed for persistence, and this does not cause any additional overhead. For notational convenience, we use a collection $T_0,\ldots,T_{D+1}$ of linear arrays, where each $T_p$ contains a slot for each elementary regular $p$-path. Specifically, for each $v_j^p \in B_p$ (the chosen basis for $\mathcal{R}_p$), $T_p$ contains a slot labeled $(v_j^p, \mathfrak{et}(v_j^p), \mathfrak{at}(v_j^p))$ which can store a linked list of $(p-1)$-paths and an integer corresponding to an entry time. We sort each $T_p$ according to increasing allow time and relabel $B_p$ if needed so that $v_j^p$ is the label of $T_p[j]$. Thus it makes sense to talk about the *index* of $v_j^p$ in $T_p$: we define $\operatorname{index}(v_j^p) = j$, and $T_p[\operatorname{index}(v_j^p)]$ is labeled by $(v_j^p, \mathfrak{et}(v_j^p), \mathfrak{at}(v_j^p))$. Note that if $v, v' \in B_p$ are such that $\operatorname{index}(v) \le \operatorname{index}(v')$, then $\mathfrak{at}(v) \le \mathfrak{at}(v')$.

Below we present a modified version of the algorithm in [36] that computes PPH. We make one last remark, based on an observation in [36]: because of the relation $\partial_p \circ \partial_{p+1}$, a pivot column of the reduced boundary matrix $M_{p,G}$ corresponds to a zero row in $M_{p+1,G}$. Thus whenever we compute $\partial_p(v)$ in our algorithm, we can immediately remove the summands that correspond to pivot terms in $M_{p-1,G}$. This is done in Algorithm 2.

**Algorithm 1** Computing persistent path homology

**procedure** ComputePPH($\mathcal{X}, D+1$) ▷ Compute PPH of network $\mathcal{X}$ up to dimension $D$
**for** $p = 0, \ldots, D$ **do**
**Pers**$_p = \varnothing$; ▷ Store intervals here
**for** $j = 1, \ldots, \dim(\mathcal{R}_{p+1})$ **do**
$[u, i, et] =$ BasisChange($v_j^{p+1}, p+1$);
**if** $u = 0$ **then** Mark $T_{p+1}[j]$;
**else**
$T_p[i] \leftarrow (u, et)$;
Add $(\mathfrak{et}(v_i^p), et)$ to **Pers**$_p$;
**for** $j = 1, \ldots, \dim(\mathcal{R}_p)$ **do**
**if** $T_p[j]$ marked and empty **then**
Add $(\mathfrak{et}(v_j^p), \infty)$ to **Pers**$_p$;
**return** **Pers**$_0, \ldots,$ **Pers**$_D$;

**Algorithm 2** Left-to-right column reduction

**procedure** BasisChange($v, \dim$) ▷ Find pivot or zero columns
$p \leftarrow \dim$; $u = \partial_p(v)$; Remove unmarked (pivot) terms from $u$;
**while** $u \neq 0$ **do**
$\sigma \leftarrow \operatorname{argmax}\{\operatorname{index}(\tau) : \tau$ is a summand of $u\}$;
$i \leftarrow \operatorname{index}(\sigma)$;
$et \leftarrow \max(\mathfrak{at}(v), \mathfrak{at}(\sigma))$;
**if** $T_{p-1}[i]$ is unoccupied **then** break;
Let $a, b$ be coefficients of $\sigma$ in $T_{p-1}[i]$ and $u$, respectively;
$u \leftarrow u - (a/b)T_{p-1}[i]$; ▷ Column reduction step
**return** $u, i, et$;