Foundations of Computational Mathematics (2025) 25:1711–1776
https://doi.org/10.1007/s10208-024-09679-2

FOUNDATIONS OF COMPUTATIONAL MATHEMATICS
The Journal of the Society for the Foundations of Computational Mathematics

Grounded Persistent Path Homology: A Stable, Topological Descriptor for Weighted Digraphs

Thomas Chaplin[1] · **Heather A. Harrington**[1,2,3,4] · **Ulrike Tillmann**[1,5]

Received: 8 November 2022 / Revised: 25 July 2024 / Accepted: 31 July 2024 /
Published online: 23 August 2024

Abstract

Weighted digraphs are used to model a variety of natural systems and can exhibit interesting structure across a range of scales. In order to understand and compare these systems, we require stable, interpretable, multiscale descriptors. To this end, we propose grounded persistent path homology (GrPPH)—a new, functorial, topological descriptor that describes the structure of an edge-weighted digraph via a persistence barcode. We show there is a choice of circuit basis for the graph which yields geometrically interpretable representatives for the features in the barcode. Moreover, we show the barcode is stable, in bottleneck distance, to both numerical and structural perturbations.

Keywords Weighted directed graphs · Topological data analysis · Persistent homology · Path homology

Mathematics Subject Classification 55N31 (Primary) · 05C20 (Secondary) · 05C22 (Secondary)

Communicated by Herbert Edelsbrunner.

✉ Thomas Chaplin
thomas.chaplin@maths.ox.ac.uk

1 Mathematical Institute, University of Oxford, Oxford, UK

2 Max Planck Institute of Molecular Cell Biology and Genetics, Dresden, Germany

3 Centre for Systems Biology Dresden, Dresden, Germany

4 Faculty of Mathematics, Technische Universität Dresden, Dresden, Germany

5 Isaac Newton Institute, University of Cambridge, Cambridge, UK

1 Introduction

Directed graphs with positive edge-weights arise both as natural objects of mathematical study and as useful models of real-world systems (e.g. [3, 4, 30, 36]). A common task is to distinguish between weighted digraphs. Frequently, this is achieved by defining an invariant, i.e. a map $\mathcal{I} : \mathbf{WDgr} \to X$ from weighted digraphs into some set X, together with a metric d on X. The metric allows us to quantitatively measure to what extent a pair of weighted digraphs differ. When $\mathcal{I}$ is well understood we may be able to explain *why* the weighted digraphs differ. In order to determine desirable characteristics of such an $\mathcal{I}$, consider the examples shown in Fig. 1, in which edge weights correspond to length as drawn.

Consider each G_i from the perspective of a particle flowing through the digraph, such that the particle may only traverse an edge in the direction specified and the time it takes corresponds to the weight. To the particle, loops (or circuits) in the graph are significant features. However, loops can vary greatly based on the orientation and weight of constituent edges.

Despite sharing the same underlying undirected graph, G_1 and G_2 support very different flows since G_1 has a single source and a single sink whereas G_2 has 4 sources and 2 sinks. To reflect this, $d(\mathcal{I}(G_1), \mathcal{I}(G_2))$ should be large. In contrast, G_3 has a different undirected graph but can be obtained from G_1 by simply subdividing each edge. In applications, this may arise from a finer resolution image of the same system. A suitable invariant should be relatively stable to such subdivisions, ideally converging to a limiting value upon iterated subdivision. Finally, G_4 has a higher circuit rank but the new loops are on a small scale, whilst the large scale organisation is mostly similar to G_1. Therefore, $d(\mathcal{I}(G_1), \mathcal{I}(G_4))$ should be small and the difference between $\mathcal{I}(G_1)$ and $\mathcal{I}(G_4)$ should reflect this multiscale comparison.

For successful application, any invariant should be stable to a reasonable noise model. A typical requirement is that $\mathcal{I}$ is continuous (or better yet Lipschitz), with respect to a choice of metric on **WDgr**. Designing metrics for graphs is an active area of research but a common choice is the graph edit distance [19]. For this metric, costs are assigned to operations such as deleting an edge or modifying a weight, then the distance between two graphs is the minimal cumulative cost of modifying one into the other. Since assigning costs to graph operations is somewhat arbitrary, it is reasonable instead to require a bound on $d(\mathcal{I}(G), \mathcal{I}(\mathbb{O}G))$, over a range of graph operations, $\mathbb{O} : \mathbf{WDgr} \to \mathbf{WDgr}$.

Finally, in many applications (particularly in biology), it is important that any invariant $\mathcal{I}$ is interpretable. That is, one must be able to explain *why* the invariant

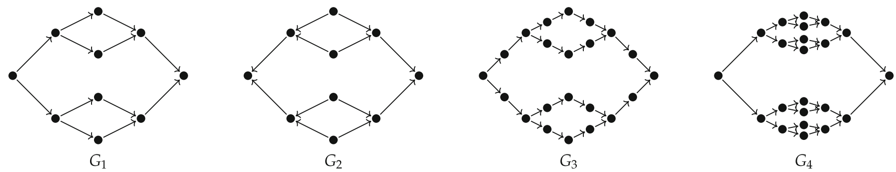

Fig. 1 Four example weighted digraphs; weights correspond to length as drawn

has the value it does. Typically this is achieved through the identification of key contributing subgraphs.

In summary, we seek an invariant for weighted digraphs which

(a) distinguishes graphs with different flow profiles due to directionality;
(b) can detect and describe features (i.e. loops) across a range of scales;
(c) is stable to reasonable perturbations and converges under iterated subdivision; and
(d) is interpretable, e.g. through the identification of important subgraphs.

Pursuant to these goals, we employ two tools from topological data analysis (TDA) – path homology and persistent homology. A number of homology theories for digraphs have been developed (see summary in Sect. 2 of [7]). Path homology is one such theory [24], which is sensitive to directionality and has useful functorial properties [13, 23]. Persistent homology is a tool for developing stable descriptors [1] that extract relevant information in multiscale scenarios. As such, persistent homology has seen successful applications to fields including neuroscience [6, 21, 22, 37], vasculature [32, 35] and financial networks [26], to name but a few [20]. A theory of persistent path homology (PPH) was proposed by Chowdhury and Mémoli [13] and is a stable descriptor for directed networks. In a search to develop an interpretable invariant for weighted digraphs, which respects the inert topology of the underlying digraph, we are lead to an alteration of PPH which we prove meets goals (a)-(d).

1.1 Contributions and Outline

In Sect. 2 we give an overview of path homology and persistent homology, and set up the categorical framework for the rest of the paper. In particular, we define a category of weighted digraphs **ContWDgr** in which morphisms are digraph maps of the underlying digraphs as well as contractions of the natural, shortest-path quasimetric.

In Sect. 3.1 we review a standard pipeline for extracting a topological invariant of a weighted digraph, via PPH. Evaluating this invariant against our stated goals, motivates an alteration of this pipeline, which we call the 'grounded pipeline'. We define this new pipeline and describe categories upon which the resulting invariant is functorial in Sect. 3.2.

Arising from the grounded pipeline, our main contribution is Definition/ Theorem 3.16, wherein we define grounded persistent path homology (GrPPH). This new invariant is a functor

$$ {}^{g}\mathcal{H}_1 : \mathbf{ContWDgr} \to \mathbf{PersVec} \tag{1.1}$$

where **PersVec** is the category of persistent vector spaces. Couching this definition in category theory yields a strong framework for comparing weighted digraphs through the invariant. Indeed, we use this functoriality later in the paper to aid the proof of decomposition and stability results.

Section 4 is devoted to developing an interpretation of GrPPH. The early subsections are dedicated to understanding the features detected by ${}^{g}\mathcal{H}_1(G)$; the following theorem summarises our findings.

Theorem 1.1 *Given a weighted digraph $G \in \mathbf{WDgr}$, denote the underlying undirected graph by $\mathcal{U}(G)$.*

(a) All features in ${}^{g}\mathcal{H}_1(G)$ are born at $t = 0$;
(b) ${}^{g}\mathcal{H}_1(G)$ at time $t = 0$ is the cycle space of $\mathcal{U}(G)$; and moreover
(c) there is a choice of circuits in $\mathcal{U}(G)$ whose homology classes generate ${}^{g}\mathcal{H}_1(G)$.

These results demonstrate how GRPPH is sensitive to circuits in the digraph at all scales, meeting goal (b), and can be interpreted through a persistence basis of such circuits, meeting goal (d). We also discuss how to use ${}^{g}\mathcal{H}_1(G)$ to assign a 'scale' to any circuit in $\mathcal{U}(G)$ in Sect. 4.2. In Sect. 4.4, we prove that if G can be decomposed into smaller parts, then GRPPH also decomposes.

Theorem 1.2 *Given a weighted digraph $G \in \mathbf{WDgr}$, if G decomposes as a wedge decomposition $G = G_1 \vee_{\hat{v}} G_2$ or a disjoint union $G = G_1 \sqcup G_2$ then*

$$ {}^{g}\mathcal{H}_1(G) \cong {}^{g}\mathcal{H}_1(G_1) \oplus {}^{g}\mathcal{H}_1(G_2). \tag{1.2}$$

In Sect. 5 we investigate the stability of ${}^{g}\mathcal{H}_1$. We employ path homotopy theory to prove the main stability theorem (Theorem 5.8), which provides a strategy for obtaining bounds on the bottleneck distance of the barcode, upon perturbing the input weighted digraph. Applying this result, we find local stability to weight perturbation (Theorem 5.11), edge subdivision (Theorem 5.16) and certain classes of edge collapses (Theorem 5.28) and edge deletions (Theorem 5.38). In particular, edge subdivision stability automatically implies that our invariant converges under iterated subdivision (Corollary 5.22). In contrast, the descriptor is unstable to generic edge collapses (Theorem 5.35) and edge deletions (Theorem 5.42), which we demonstrate through a number of counter-examples. We argue that these stability properties suffice to meet goal (c) and indeed stability to larger classes of edge collapses and edge deletion would be undesirable. For a summary of all stability results obtained, please consult Table 1.

In order to build intuition for what is measured by GRPPH, we compute a number of illustrative examples in Sect. 6. In particular, in Sect. 6.1, we consider a simple, cycle graph and determine the limiting value of GRPPH under iterated edge subdivision. In Sect. 6.2, we compute the invariant for a number of small square digraphs with varying edge orientations, illustrating sensitivity to directionality, as required by goal (a). Finally, in Sect. 6.5, we describe properties of ${}^{g}\mathcal{H}_1(G)$ when G is a complete digraph, with weights forming a finite quasimetric space, and compute the descriptor for a point sample of the unit circle.

1.2 Computations

An algorithm for computing PPH in arbitrary degrees was proposed by Chowdhury and Mémoli [13]; a more efficient algorithm for computing PPH in degree 1 was later proposed by Dey, Li, and Wang [17]. We modify the latter algorithm to compute GRPPH; an implementation, capable of computing representatives, is available as a Python package [8].

2 Background

2.1 Basic Notation and Category Theory

We introduce some basic language for graphs and categories, and then recall the definitions of path homology and persistent homology, the two theories we combine later in a new way to define our invariant.

Notation 2.1 For $d \in \mathbb{N}$, the standard d-simplex is

$$\Delta^d := \left\{ (x_0, \dots x_d) \in \mathbb{R}^{d+1} \;\middle|\; \sum x_i = 1, \text{ and } x_i \geq 0\ \forall i \right\}. \tag{2.1}$$

Notation 2.2 Given a category $\mathcal{C}$, we denote the collection of objects $\mathrm{Obj}(\mathcal{C})$ and the collection of morphisms $\mathrm{Mor}(\mathcal{C})$. For two objects $X, Y \in \mathrm{Obj}(\mathcal{C})$, we denote the set of morphisms $X \to Y$ by $\mathrm{Mor}_{\mathcal{C}}(X, Y)$. Where it is clear from context whether α is an object or a morphism, we simply write $\alpha \in \mathcal{C}$.

Notation 2.3 Fix any three categories $\mathcal{C}$, $\mathcal{D}$ and $\mathcal{D}'$.

(a) We denote the category of functors $\mathcal{C} \to \mathcal{D}$, where morphisms are natural transformations, by $[\mathcal{C}, \mathcal{D}]$.
(b) For a morphism $f \in \mathrm{Mor}_{[\mathcal{C}, \mathcal{D}]}(M, N)$, we denote the components of the natural transformation by $f_x : M(x) \to N(x)$ for each $x \in \mathcal{C}$.
(c) Given a functor $\mu : \mathcal{D} \to \mathcal{D}'$, there is a functor $[\mathcal{C}, \mu] : [\mathcal{C}, \mathcal{D}] \to [\mathcal{C}, \mathcal{D}']$. Given $F \in \mathrm{Obj}([\mathcal{C}, \mathcal{D}])$, we map $[\mathcal{C}, \mu](F) := \mu \circ F$. Given a natural transformation $\nu : F \Rightarrow F'$ between functors $F, F' \in [C, D]$, we map $[\mathcal{C}, \mu](\nu) := \mu \circ \nu$ which one can confirm is a natural transformation $\mu \circ F \Rightarrow \mu \circ F'$.

Notation 2.4 (a) We let $\mathbf{R}$ denote the poset of $\mathbb{R}$ equipped with the $\leq$ relation, viewed as a category.
(b) We let **Vec** denote the category of $\mathbb{R}$-vector spaces and **vec** denote the full subcategory of finite-dimensional $\mathbb{R}$-vector spaces.
(c) We let **Ch** denote the category of chain complexes over $\mathbb{R}$.
(d) Given a category $\mathcal{C}$, a filtration is a functor $F : \mathbf{R} \to \mathcal{C}$ where the morphisms $F(s \leq t)$ are inclusions (assuming such a notion is defined in the category $\mathcal{C}$).

Definition 2.5 Given a chain complex $C_\bullet \in \mathbf{Ch}$, we denote the k^{th} chain group by C_k and the boundary map by $\partial_k : C_k \to C_{k-1}$. For each $k \in \mathbb{N}$, the k^{th} homology group, $H_k(C)$ is the quotient

$$H_k(C) := \frac{\ker \partial_k}{\mathrm{im}\, \partial_{k+1}}. \tag{2.2}$$

We can view H_k as a functor $\mathbf{Ch} \to \mathbf{Vec}$.

2.2 Directed Graphs

Definition 2.6 (a) A (simple) digraph is a tuple $G = (V, E)$ where V, the set of vertices, is a finite set and E, the set of edges, is a subset of $V \times V \setminus \Delta_V$ where

$$\Delta_V := \{(v, v) \in V \times V \mid v \in V\}. \tag{2.3}$$

(b) A directed acyclic graph (DAG) is a simple digraph $G = (V, E)$ such that there is a partial order $<$ on the nodes such that $(i, j) \in E \implies i < j$.
(c) An oriented graph is a simple digraph $G = (V, E)$ with no double edges, i.e. $(i, j) \in E \implies (j, i) \notin E$.
(d) A weighted (digraph/DAG/oriented graph) is a triple $G = (V, E, w)$ such that (V, E) is a (simple digraph/DAG/oriented graph) and $w : E \to \mathbb{R}_{>0}$ is a positively-valued function on the edges.
(e) An undirected graph is a tuple $G = (V, E)$ where E is a multiset of 2-element subsets of V.
(f) Given a digraph $G = (V, E)$, the underlying undirected graph is $\mathcal{U}(G) := (V, \mathcal{U}(E))$ where

$$(i, j) \in E, (j, i) \notin E \implies \{i, j\} \in \mathcal{U}(E) \text{ with multiplicity 1}, \tag{2.4}$$
$$(i, j), (j, i) \in E \implies \{i, j\} \in \mathcal{U}(E) \text{ with multiplicity 2}. \tag{2.5}$$

(g) The weakly connected components of G are the connected components of $\mathcal{U}(G)$, as a partition of $V(G)$. If a and b belong to the same weakly connected component, we say they are weakly connected, else we say they are weakly disconnected.
(h) Given two digraphs $G_1 = (V_1, E_1)$, $G_2 = (V_2, E_2)$ with $V_1, V_2 \subseteq V$, their union is $G_1 \cup G_2 := (V_1 \cup V_2, E_1 \cup E_2)$. If V_1, V_2 are disjoint, then we denote this $G_1 \sqcup G_2$.

Notation 2.7 Fix a weighted digraph $G = (V, E, w)$.

(a) We denote $V(G) := V$, $E(G) := E$, $w(G) := w$.
(b) For an edge $e = (i, j) \in E$, we write $\mathrm{st}(e) := i$ and $\mathrm{fn}(e) := j$, to denote the 'start' and 'finish' of e. Furthermore, we say that i and j are incident to e.
(c) For an edge $e = (i, j) \in E$, we write $w(i, j) := w(e)$.
(d) We write $i \to j$ to mean there is an edge $(i, j) \in E$.

Definition 2.8 Fix a weighted digraph $G = (V, E, w)$.

(a) Given $V' \subseteq V$, the induced subgraph on V' is (V', E', w'), where $E' = E \cap (V' \times V')$ and w' is w restricted to E'. Note that the induced subgraph on V is all of G.
(b) Given $E' \subseteq E$, the induced subgraph on E' is (V', E', w'), where V' is the set of all vertices incident to some edges in E' and w' is w restricted to E'. Note that the induced subgraph on E may be a proper subgraph of G.

(c) For $v \in V$,

$$\mathcal{N}_{in}(v; G) := \{a \in V \mid a \to v\}, \tag{2.6}$$

$$\mathcal{N}_{out}(v; G) := \{b \in V \mid v \to b\}, \tag{2.7}$$

$$\mathcal{N}(v; G) := \mathcal{N}_{in}(v) \cup \mathcal{N}_{out}(v) \tag{2.8}$$

and the vertex-neighbourhood graph, $\mathcal{NG}(v; G)$, is the induced subgraph on $\mathcal{N}(v)$. Where G is clear from context, we omit it from notation.

(d) For $F \subseteq E$,

$$\mathcal{N}(F; G) := \{\tau \in E \mid \exists\, e \in F \text{ such that } \tau, e \text{ are incident to a common vertex}\} \tag{2.9}$$

and if $F = \{e\}$ is a single edge, we use the notation $\mathcal{N}(e; G)$. The edge-neighbourhood graph, $\mathcal{NG}(F; G)$, is the induced subgraph on $\mathcal{N}(F; G)$. Where G is clear from context, we omit it from notation.

(e) Given a vertex $v \in V$, the closed star of v is the induced subgraph on those edges that are incident to v, which we denote $\mathrm{Star}(v; G)$. Where G is clear from context, we omit it from notation. Note that $\mathrm{Star}(v) \subseteq \mathcal{NG}(v)$.

(f) For two vertices $a, b \in V$, a (directed) trail from a to b is an alternating sequence of vertices, $v_i \in V$, and forward edges, $e_i \in E$,

$$p = (v_0 = a, e_1, v_1, e_2, \ldots, e_k, v_k = b) \tag{2.10}$$

such that $e_i = (v_{i-1}, v_i)$. We write that p is a trail $a \rightsquigarrow b$. In a simple digraph, the e_i uniquely determine the v_i (and vice versa) so we occasionally omit one from the notation.

(g) A trail with no self-intersections is a path, i.e. $v_i = v_j \implies i = j$.

(h) An undirected circuit is an alternating sequence of vertices, $v_i \in V$, and edges, $e_i \in E$,

$$p = (v_0 = a, e_1, v_1, e_2, \ldots, e_k, v_k = a) \tag{2.11}$$

such that $\{\mathrm{st}(e_i), \mathrm{fn}(e_i)\} = \{v_{i-1}, v_i\}$ and $v_0 = v_k$. Unlike trails, we also require $k > 1$ to ensure the circuit contains at least one edge.

(i) Given an undirected circuit p, as above, if $v_0, \ldots, v_{k-1}$ are all distinct then we say p is simple.

Remark 2.9 We allow trails and paths to contains zero edges, i.e. $p = (a)$ is a trail $a \rightsquigarrow a$ but not an undirected circuit.

Notation 2.10 Fix a weighted digraph $G = (V, E, w)$.

(a) Given a trail p, the length of p is defined as

$$\mathrm{len}(p) := \sum_{i=1}^{k} w(e_i). \tag{2.12}$$

(b) We denote the set of all paths $i \rightsquigarrow j$ by $\mathcal{P}(i, j)$.

Definition 2.11 For a weighted digraph G, the shortest-path quasimetric $d : V(G) \times V(G) \to \mathbb{R} \sqcup \{\infty\}$ is defined by

$$d(i, j) := \begin{cases} \min_{p \in \mathcal{P}(i,j)} \mathrm{len}(p) & \text{if } \mathcal{P}(i, j) \neq \emptyset, \\ \infty & \text{otherwise.} \end{cases} \tag{2.13}$$

Note that for any vertex $v \in V(G)$, $d(v, v) = 0$.

There is a notion of morphisms between (weighted) digraphs.

Definition 2.12 (a) Given two simple digraphs G, H, a digraph map (or simply map), $f : G \to H$, is a map on vertices, $f : V(G) \to V(H)$, such that

$$i \to j \implies f(i) \to f(j) \text{ or } f(i) = f(j). \tag{2.14}$$

Given a vertex map $f : V(G) \to V(H)$ satisfying condition (2.14), we say f induces a digraph map $G \to H$.

(b) A digraph map $f : G \to H$ is called an inclusion if $V(G) \subseteq V(H)$ and f is induced by the inclusion vertex map.

(c) Given two weighted digraphs G, H, a digraph map $f : G \to H$ is called a contraction if for all nodes $i, j \in V(G)$, we have

$$d_H(f(i), f(j)) \le d_G(i, j) \tag{2.15}$$

where d_G and d_H are the shortest-path quasimetrics on G and H respectively.

Notation 2.13 (a) For $e = (i, j) \in E(G)$, we denote $f(e) := (f(i), f(j))$. Note that $f(e) \in E(H) \sqcup \Delta_{V(H)}$.

(b) Given a path $p = (v_0, e_1, \dots, e_k, v_k)$ and a digraph map $f : G \to H$, the image of p is the path, $f(p)$, obtained from

$$(f(v_0), f(e_1), f(v_1), \dots, f(e_k), f(v_k)) \tag{2.16}$$

by removing $f(e_i)$ and $f(v_i)$ from the sequence if $f(e_i)$ is a self-loop, i.e. $f(v_{i-1}) = f(v_i)$.

Remark 2.14 Suppose $f : G \to H$ is a digraph map such that $w(H)(f(e)) \le w(G)(e)$ for every edge $e \in E(G)$ with $f(e) \notin \Delta_{V(H)}$. Then for any path $p : i \rightsquigarrow j$ in G, $f(p)$ is a path $f(i) \rightsquigarrow f(j)$ in H and $\mathrm{len}(f(p)) \le \mathrm{len}(p)$. Therefore, f is a contraction. Whilst this is a sufficient condition to form a contraction, it is not a necessary condition; the shortest path joining $f(i) \rightsquigarrow f(j)$ in H need not be the image of the shortest path joining $i \rightsquigarrow j$.

Given these ways of mapping between (weighted) digraphs, a number of categories naturally arise.

Definition 2.15 (a) We denote the category of simple digraphs, directed acyclic graphs and oriented graphs, where the morphisms are all digraph maps, by **Dgr**, **Dag** and **Dor** respectively.

(b) We use the prefix **W** to denote the corresponding categories of *weighted* digraphs where a morphism is any digraph map of the underlying, unweighted digraphs. For example, **WDgr** is a category of weighted simple digraphs.

(c) For a category of weighted digraphs, we use the prefix **Cont** to denote the subcategory, containing all objects, with the additional restriction that morphisms must be contractions.

(d) For a category of weighted or unweighted digraphs, we use the prefix **Incl** to denote the wide subcategory, containing all objects, with the additional restriction that morphisms must be inclusions.

Finally, in order to align our terminology with that of [13] we make the following definition.

Definition 2.16 Given a finite set V, a directed network is a non-negative function $A : V \times V \to [0, \infty)$ such that $A(v_1, v_2) = 0 \iff v_1 = v_2$.

2.3 Path Homology

Path homology is a homology theory for directed graph, which was first introduced by Grigor'yan et al. [24]. Subsequent papers prove Künneth theorems for Cartesian products and joins [25], and invariance under an appropriate notation of digraph homotopy [23]. A directed network gives rise to a natural filtration of digraphs which leads to a stable theory of persistent path homology [13]. (Persistent) path homology has also been extended to vertex-weighted digraphs [28]. Path homology can be defined for an arbitrary path complex; here we present the definition for a digraph.

Fix a ring R and a simple directed graph $G = (V, E)$.

Definition 2.17 The following definitions classify sequences of vertices in V:

(a) An elementary p-path is any sequence $v_0 \dots v_p$ of $(p + 1)$ vertices, $v_i \in V$.

(b) An elementary p-path, $v_0 \dots v_p$, is regular if $v_i \neq v_{i+1}$ for every i. Otherwise, we say it is non-regular.

(c) An elementary p-path, $v_0 \dots v_p$, is allowed if $(v_i, v_{i+1}) \in E$ for every i.

Definition 2.18 We freely generate R-modules from these sequences of vertices, for each $p \geq 0$.

$$\Lambda_p := \Lambda_p(G; R) := R\Big\langle \big\{ v_0 \dots v_p \text{ elementary } p\text{-path on } V \big\} \Big\rangle \tag{2.17}$$

$$\mathcal{R}_p := \mathcal{R}_p(G; R) := R\Big\langle \big\{ v_0 \dots v_p \text{ regular } p\text{-path on } V \big\} \Big\rangle \tag{2.18}$$

$$\mathcal{I}_p := \mathcal{I}_p(G; R) := R\Big\langle \big\{ v_0 \dots v_p \text{ non-regular } p\text{-path on } V \big\} \Big\rangle \tag{2.19}$$

$$\mathcal{A}_p := \mathcal{A}_p(G; R) := R\Big\langle \big\{ v_0 \dots v_p \text{ allowed } p\text{-path in } G \big\} \Big\rangle \tag{2.20}$$

For $p = -1$, we let $\Lambda_{-1} := \mathcal{R}_{-1} := \mathcal{A}_{-1} := R$.

Definition 2.19 Given $p \geq 0$, the non-regular boundary map $\partial_p^{\mathrm{nr}} : \Lambda_p \to \Lambda_{p-1}$ is given on the standard basis by

$$\partial_p^{\mathrm{nr}}(v_0 \dots v_p) := \sum_{i=0}^{p} (-1)^i v_0 \dots \hat{v_i} \dots v_p \tag{2.21}$$

where $v_0 \dots \hat{v_i} \dots v_p$ is the $(p-1)$-path obtained by removing v_i from $v_0 \dots v_p$.

Definition 2.20 Since $\Lambda_p = \mathcal{R}_p \oplus \mathcal{I}_p$, let $\pi : \Lambda_p \to \mathcal{R}_p$ denote the projection onto $\mathcal{R}_p$ along $\mathcal{I}_p$ and let $\iota : \mathcal{R}_p \to \Lambda_p$ denote the natural inclusion. The regular boundary map $\partial_p : \mathcal{R}_p \to \mathcal{R}_{p-1}$ is given by

$$\partial_p := \pi \circ \partial_p^{\mathrm{nr}} \circ \iota. \tag{2.22}$$

To justify the name, a standard check confirms that indeed $\partial_{p-1} \circ \partial_p = 0$ [24, Lemma 2.9]. However, the $\{\mathcal{R}_p\}$ are invariant to the edge set and although $\mathcal{A}_p$ is a subspace of each $\mathcal{R}_p$, the boundary map ∂_p does not pass down to a boundary operator. Hence, we must define the following sub-modules.

Definition 2.21 The space of ∂-invariant p-paths is

$$\Omega_p := \Omega_p(G; R) := \left\{ v \in \mathcal{A}_p \;\middle|\; \partial_p v \in \mathcal{A}_{p-1} \right\}. \tag{2.23}$$

Elements of this space are called ∂-invariant p-paths.

Note that ∂_p restricts to a homomorphism $\Omega_p \to \Omega_{p-1}$ and hence forms a chain complex.

Definition 2.22 The regular path (chain) complex is

$$\dots \xrightarrow{\partial_3} \Omega_2 \xrightarrow{\partial_2} \Omega_1 \xrightarrow{\partial_1} \Omega_0 \xrightarrow{\partial_0} R \xrightarrow{\partial_{-1}} 0 \tag{2.24}$$

The homology of the regular path complex is the regular path homology of G, the k^{th} homology group is

$$H_k := H_k(G; R) := \frac{\ker \partial_k}{\operatorname{im} \partial_{k+1}}. \tag{2.25}$$

The k^{th} Betti number is $\beta_k := \operatorname{rank} H_k$.

Remark 2.23 Note that the chain groups of this complex are Ω_p not $\mathcal{R}_p$, we use the adjective regular because Ω_p is constructed via the regular boundary map.

Remark 2.24 Note that when $R = \mathbb{Z}$ or $\mathbb{R}$, $\operatorname{rank}(\ker \partial_1)$ coincides with the circuit rank of $\mathcal{U}(G)$ — those unfamiliar with this notion can take this as the definition.

Definition 2.25 Given a digraph map $f : G \to H$, the induced map $f_\# : \mathcal{R}_p(G) \to \mathcal{R}_p(H)$ is given on the standard basis by

$$f_\#(v_0 \dots v_p) := \begin{cases} f(v_0)\dots f(v_p) & \text{if } f(v_0)\dots f(v_p) \text{ is regular} \\ 0 & \text{otherwise} \end{cases} \tag{2.26}$$

Notation 2.26 Given a digraph map $f : G \to H$, we denote the induced map on homology by $f_* := H_k(f_\#) : H_k(G) \to H_k(H)$ for each $k \in \mathbb{N}$.

Lemma 2.27 ([23, Theorem 2.10], [13, Proposition A.2]) *The induced maps restrict to maps $f_\# : \Omega_p(G) \to \Omega_p(H)$ which commute with ∂_p and hence form chain maps between the regular path complexes. Moreover these chain maps are functorial, i.e.* $(f \circ g)_\# = f_\# \circ g_\#$ *and* $(\mathrm{id})_\# = \mathrm{id}$. *Hence Ω is a functor* $\mathbf{Dgr} \to \mathbf{Ch}$.

We will primarily be interested in the 1^{st} homology group H_1. Hence the following characterisation of the low-dimensional chain groups will be of use.

Proposition 2.28 ([24, § 3.3]) *For any simple digraph $G = (V, E)$, $\Omega_0(G)$ is isomorphic to the R-module freely generated by the vertices and $\Omega_1(G)$ is isomorphic to the R-module freely generated by the edges, i.e.*

$$\Omega_0(G) \cong R\langle V\rangle \quad \textit{and} \quad \Omega_1(G) \cong R\langle E\rangle\,. \tag{2.27}$$

Notation 2.29 Note that an edge $e = (a, b) \in E$ gives rise to an allowed 1-path $ab \in \mathcal{A}_1(G)$. Moreover, $ab \in \Omega_1(G)$. For ease of notation, given an edge $e \in E$, we will also use $e = ab \in \Omega_1(G)$ to refer to the generator in $\Omega_1(G)$.

Since Ω_1 is generated by edges in G, any trail p has a representative.

Notation 2.30 Given a directed trail $p = (e_1, \dots, e_k)$ in a digraph G, the representative of p is

$$\mathfrak{R}(p) := \sum_{i=1}^{k} e_i \in \Omega_1. \tag{2.28}$$

Likewise, there is a representative for any undirected circuit.

Notation 2.31 Given an undirected circuit $p = (v_0 = a, e_1, v_1, \dots, e_k, v_k = a)$ in a digraph G, the representative of p is

$$\mathfrak{R}(p) := \sum_{i=1}^{k} \alpha_i e_i \in \Omega_1. \tag{2.29}$$

where $\alpha_i = 1$ if $e_i = (v_{i-1}, v_i)$, else $\alpha_i = -1$.

Remark 2.32 The representative of a circuit p, does not depend on the starting point, but if p' traverses the circuit in the opposite direction then $\mathfrak{R}(p') = -\mathfrak{R}(p)$.

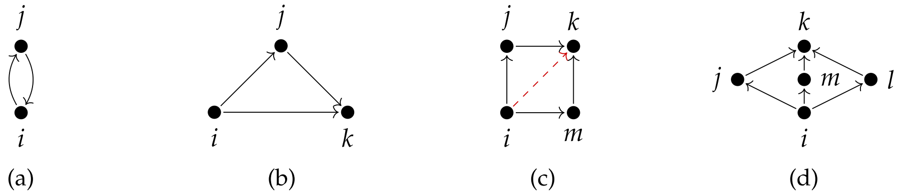

Fig. 2 The three types of generators for $\Omega_2(G;\mathbb{Z})$ and $\Omega_2(G;\mathbb{R})$: **a** A double edge. **b** A directed triangle. **c** A long square (the dashed red edge must not be present). Finally **d** shows a linear dependency between long squares (Color figure online)

Proposition 2.33 ([23, Proposition 2.9], [17, Theorem 3]) *Let G be a finite, simple digraph and $R = \mathbb{R}$ or $\mathbb{Z}$. Any $\omega \in \Omega_2(G; R)$ can be written as a linear combination of ∂-invariant 2-paths of the following three types (represented in Fig. 2):*

(a) (iji) where $i \to j \to i$ (double edge);
(b) (ijk) where $i \to j \to k$, $i \to k$ and $i \neq k$ (directed triangle); and
(c) $(ijk - imk)$ where $i \to j \to k$, $i \to m \to k$, $i \not\to k$ and $i \neq k$ (long square).

First note that all of the elements identified in Proposition 2.33 are elements of $\Omega_2(G; R)$ and hence they form a generating set. However, the generators corresponding to long squares are not necessarily linearly independent. For example, in Fig. 2d, we see

$$(ijk - ilk) = (ijk - imk) + (imk - ilk). \tag{2.30}$$

Removing some long squares to account for these linear relations, we can obtain a basis of $\Omega_2(G;\mathbb{R})$.

2.4 Path Homotopy

Path homotopy, also introduced by Grigor'yan et al. [23], is a homotopy theory under which path homology is invariant. We employ this theory in order to build interleavings between the grounded persistent path homology of two related weighted digraphs. Below, we present the basic definitions and results of path homotopy theory, extending it to a relative version in which we fix a subset of the vertices throughout the homotopy.

Definition 2.34 (a) A line digraph of length n is a digraph on $0, 1, \dots, n$ with exactly one of $(i, i+1)$, $(i+1, i)$ for each i and no other edges.
(b) Denote the two length 1 line digraph by I_+ for $0 \longrightarrow 1$ and I_- for $0 \longleftarrow 1$.
(c) Given two digraphs G, H, their box product is the digraph $G \boxdot H$ where

$$V(G \boxdot H) := V(G) \times V(H),$$
$$E(G \boxdot H) := \left\{((x, y), (x', y')) \;\middle|\; (x \to x' \text{ and } y = y') \text{ or } (x = x' \text{ and } y \to y')\right\}.$$

(d) Given two digraph maps $f, g : G \to H$, a homotopy between f and g is a digraph map $F : G \boxdot I \to H$, where I is a line digraph of length n

$$F\Big|_{G\times\{0\}} = f \quad \text{and} \quad F\Big|_{G\times\{n\}} = g. \tag{2.31}$$

(e) If I is length 1 in the above, we say F is a one-step homotopy. Otherwise, we say F is a multi-step homotopy.
(f) Suppose F is a one-step homotopy. If, furthermore, $I = I_+$ we say F is a from f to g else $I = I_-$ and we say F is from g to f.
(g) If there is a homotopy in either direction, we say f and g are homotopic and write $f \simeq g$.
(h) Given $A \subseteq V(G)$, we say F is a homotopy relative A if for each $a \in A$, the partially applied function $F(a, \bullet) : V(I) \to V(H)$ is constant.

Given an arbitrary homotopy $F : G \boxdot I \to H$ between f and g where I is of length n, let $f_i : G \to H$ denote the restriction of F to $G \times \{i\}$ and let I_i denote the induced subgraph of I on $\{i-1, i\}$ (a single edge). By restricting F to each $G \boxdot I_i$, we obtain n *one-step* homotopies $F_i : G \boxdot I_i \to H$ for $i = 1, \dots, n$. Each homotopy F_i is between f_{i-1} and f_i, with the direction dependent on the direction of the edge I_i. Moreover, if F is a relative $A \subseteq V(G)$ then so too is each F_i.

Grigor'yan et al. [23] showed that path homology is invariant with respect to path homotopy. In [23, Theorem 3.3] the chain homotopy which achieves this invariance is only made explicit in the one-step homotopy case. In order to describe the behaviour of relative path homotopies, we make the construction explicit for multi-step homotopies.

Theorem 2.35 *If $f \simeq g$ then there is an induced chain homotopy, i.e a sequence of morphisms $L_p : \Omega_p(G) \to \Omega_{p+1}(H)$, such that for all $c \in \Omega_p(G)$,*

$$g_\#(c) - f_\#(c) = \partial L_p(c) + L_{p-1}\partial(c). \tag{2.32}$$

Proof Suppose F is a one-step homotopy from f to g, then in the proof of [23, Theorem 3.3] it was shown that there exists a chain homotopy L satisfying Eq. (2.32). Suppose instead that F is an arbitrary homotopy between f and g. Then, as described above, there is a sequence of digraph maps

$$f = f_0 =, f_1, \dots, f_n = g \tag{2.33}$$

and a sequence of *one-step* homotopies $F_1, \dots F_n$ such that either F_i is from f_{i-1} to f_i or is from f_i to f_{i-1}. In the first case, we define $\sigma_i{:=}1$ and in the latter we define $\sigma_i{:=}-1$. Each of the one-step homotopies induces a chain homotopy which we denote L^i and the signs ensure that, for each i,

$$(f_i)_\#(c) - (f_{i-1})_\#(c) = \sigma_i \cdot [\partial L^i_p(c) + L^i_{p-1}\partial(c)]. \tag{2.34}$$

Finally, we linearly combine these homotopies in each degree $L{:=}\sum_{i=1}^n \sigma_i L^i$ to obtain the required chain homotopy. This L satisfies Eq. (2.32) thanks to a telescoping sum. □

Corollary 2.36 ([23, Theorem 3.3]) *If $f, g : G \to H$ are homotopic digraph maps then they induce identical maps on homology $f_*, g_* : H_k(G) \to H_k(H)$ for every $k \in \mathbb{N}$.*

If $f, g : G \to H$ are path homotopic relative $A \subseteq V(G)$ then any $c \in \Omega(G)$ supported entirely on vertices of A has the same image under $f_\#$ and $g_\#$. As one might hope, these ∂-invariant paths are in the kernel of the induced chain homotopy. This behaviour will be crucial in the proof of the main stability theorem (Theorem 5.8).

Lemma 2.37 *Suppose $F : G \boxdot I \to H$ is a path homotopy relative $A \subseteq V(G)$, let G_A denote the induced subgraph of G on A and let L denote the induced chain homotopy from Theorem 2.35. Then $L_p(\Omega_p(G_A)) = 0$.*

Proof Due to the linear construction above, it suffices to show this result in the case where F is a one-step homotopy. We also assume $I = I_+$; the $I = I_-$ case admits a similar proof.

To ease notation, for each vertex $v \in V(G)$, let v denote $(v, 0) \in V(G \boxdot I)$ and let v' denote $(v, 1) \in V(G \boxdot I)$. Next, we can write any $c \in \Omega_p(G_A)$ in terms of the standard basis of $\mathcal{A}_p(G_A)$

$$c = \sum c^{v_0 \dots v_p} v_0 \dots v_p \tag{2.35}$$

where the sum is over all p-paths, $v_0 \dots v_p$, with $v_i \in A$ for all i. Following the construction in [23, Theorem 3.3], one can show

$$L_p(c) = \sum c^{v_0 \dots v_p} \sum_k (-1)^k \pi[F(v_0) \dots F(v_k) F(v_k') \dots F(v_p')]. \tag{2.36}$$

In each summand, we note $F(v_k) = F(v_k')$ since F is relative A. Therefore each of the paths $F(v_0) \dots F(v_k) F(v_k') \dots F(v_p')$ is an irregular path which is sent to 0 under π. □

2.5 Persistent Homology

Topological data analysis (TDA) is a field of applied mathematics which employs the powerful, discriminative tools of algebraic topology to study complex datasets. The cornerstone of the field is persistent homology (PH) which yields a stable, discrete, topological invariant, called a barcode (see [5, 9, 33] for an overview). The barcode summarises topological features in the data (e.g. connected components and loops) and measures the range of scales across which they persist.

Definition 2.38 (a) A persistent chain complex is a functor $\mathbf{R} \to \mathbf{Ch}$.

(b) A persistent vector space is a functor $\mathbf{R} \to \mathbf{Vec}$. We denote the category of such functors $\mathbf{PersVec}$:=$[\mathbf{R}, \mathbf{Vec}]$.

(c) A persistent vector space M is pointwise finite-dimensional (p.f.d) if $M(t)$ is finite dimensional for all $t \in \mathbb{R}$, that is $M \in [\mathbf{R}, \mathbf{vec}]$. We denote the category of p.f.d persistent vector spaces $\mathbf{Persvec}$:=$[\mathbf{R}, \mathbf{vec}]$.

(d) A persistent vector space $M : \mathbf{R} \to \mathbf{Vec}$ is tame [31] if

  (i) $M(t)$ is finite dimensional for all $t \in \mathbb{R}$, and

(ii) there are finitely many $t \in \mathbb{R}$ such that there is no $\epsilon > 0$ such that $M(t-\epsilon \leq t+\epsilon)$ is an isomorphism.

(e) For an interval $I \subseteq R$, we define the corresponding interval, $P(I) \in \mathbf{PersVec}$, in which the vector spaces are

$$P(I)(t){:=}\begin{cases}\mathbb{R} & \text{if } t \in I,\\ 0 & \text{otherwise,}\end{cases} \tag{2.37}$$

and $P(I)(s \leq t)$ is the identity if $s, t \in I$ and the trivial map otherwise.

(f) Given $M, N \in \mathbf{PersVec}$, their direct sum $M \oplus N \in \mathbf{PersVec}$ is given pointwise by

$$(M \oplus N)(t){:=}M(t) \oplus N(t), \tag{2.38}$$

$$(M \oplus N)(s \leq t){:=}M(s \leq t) \oplus N(s \leq t). \tag{2.39}$$

(g) A morphism of persistent vector spaces is a morphism in the category $[\mathbf{R}, \mathbf{Vec}]$. That is, for $M, N \in [\mathbf{R}, \mathbf{Vec}]$ a morphism $\phi : M \to N$ is a family of linear maps $\{\phi_t : M(t) \to N(t)\}$ such that

$$N(s \leq t) \circ \phi_s = \phi_t \circ M(s \leq t) \tag{2.40}$$

whenever $s \leq t$. We say ϕ is an isomorphism if each ϕ_t is an isomorphism of vector spaces. If an isomorphism $M \to N$ exists, we write $M \cong N$.

A (p.f.d) persistent vector space can be decomposed as a direct sum of interval modules, the *indecomposable* persistent vector spaces. Moreover, this decomposition is unique, discrete and finite in all practical applications.

Theorem 2.39 (Structure Theorem for p.f.d persistent vector spaces, [16, Theorem 1.1], [10, Theorem 2.8]) *Given $M \in \mathbf{Persvec}$, there is a multiset $\mathcal{B}M$ of intervals of $\mathbb{R}$ such that*

$$M \cong \bigoplus_{I \in \mathcal{B}M} P(I) \tag{2.41}$$

and any such decomposition is unique, up to reordering. We call $\mathcal{B}M$ the barcode of M.

Definition 2.40 (a) A multiset of intervals of $\mathbb{R}$ is called a barcode.

(b) We call an interval I in a barcode a feature. If I starts at a and ends at b, we say the feature is born at time a and dies at time b.

(c) Given a barcode $\mathcal{B}$, the diagram of $\mathcal{B}$ is the multiset of endpoints

$$\mathrm{Dgm}(B){:=}\{(a_k, b_k) \mid I_k \in \mathcal{B} \text{ has endpoints } a_k \leq b_k\}. \tag{2.42}$$

(d) Given $M \in \mathbf{Persvec}$, the persistence diagram of M is $\mathrm{Dgm}(M){:=}\mathrm{Dgm}(\mathcal{B}M)$.

The barcode can be used as a summary of the persistent vector space. When arising as the homology of a filtration of topological spaces, this summary captures how topological features are born and killed throughout the filtration. In order to use this summary for further statistics, it is desirable that this summary is *stable* to noise and perturbations in the input data. To quantify this stability, we require metrics on persistent vector spaces and the resulting barcodes.

Definition 2.41 ([18]) Given two multisets $D_1, D_2 \subseteq \mathbb{R}^2$, the bottleneck distance is

$$d_B(D_1, D_2) := \inf_{\gamma} \sup_{x \in D_1 \cup \Delta_{\mathbb{R}}} \|x - \gamma(x)\|_\infty \tag{2.43}$$

where γ is over all multi-bijections $D_1 \cup \Delta_{\mathbb{R}} \to D_2 \cup \Delta_{\mathbb{R}}$ and $\Delta_{\mathbb{R}} = \{(x, x) \mid x \in \mathbb{R}\}$ is the diagonal with multiplicity 1. Given barcodes $\mathcal{B}_1, \mathcal{B}_2$, we define the bottleneck distance between them to be the bottleneck distance between their diagrams

$$d_B(\mathcal{B}_1, \mathcal{B}_2) := d_B(\mathrm{Dgm}(\mathcal{B}_1), \mathrm{Dgm}(\mathcal{B}_2)). \tag{2.44}$$

Definition 2.42 ([15]) Given $p \geq 1$ and two multisets $D_1, D_2 \subseteq \mathbb{R}^2$, the p-Wasserstein distance is

$$d_{W_p}(D_1, D_2) := \inf_{\gamma} \left(\sum_{x \in D_1 \cup \Delta_{\mathbb{R}}} \|x - \gamma(x)\|_\infty^p \right)^{1/p} \tag{2.45}$$

where γ is over all multi-bijections $D_1 \cup \Delta_{\mathbb{R}} \to D_2 \cup \Delta_{\mathbb{R}}$ and $\Delta_{\mathbb{R}} = \{(x, x) \mid x \in \mathbb{R}\}$ is the diagonal with multiplicity 1. The p-Wasserstein between two barcodes $\mathcal{B}_1, \mathcal{B}_2$, is

$$d_{W_p}(\mathcal{B}_1, \mathcal{B}_2) := d_{W_p}(\mathrm{Dgm}(\mathcal{B}_1), \mathrm{Dgm}(\mathcal{B}_2)). \tag{2.46}$$

Definition 2.43 ([1]) Given a category $\mathcal{C}$, fix $M, N \in [\mathbf{R}, \mathcal{C}]$ and $\delta \geq 0$.

(a) The δ-shift of M is $M[\delta] \in [\mathbf{R}, \mathcal{C}]$ where

$$M[\delta](t) := M(t + \delta) \quad \text{and} \quad M[\delta](s \leq t) := M(s + \delta \leq t + \delta). \tag{2.47}$$

(b) Given a morphism $f \in \mathrm{Mor}_{[\mathbf{R}, \mathcal{C}]}(M, N)$ the δ-shift of f is $f[\delta] : M[\delta] \to N[\delta]$ in which $f_t := f_{t+\delta}$. When clear from context we often denote $f[\delta] = f$.
(c) The δ-transition morphism is a morphism $\mathcal{T}(M, \delta) : M \to M[\delta]$ which at $t \geq 0$ is given by $M(t \leq t + \delta)$.
(d) A δ-interleaving is a pair of morphisms $\phi : M \to N[\delta]$ and $\psi : N \to M[\delta]$ such that

$$\psi[\delta] \circ \phi = \mathcal{T}(M, 2\delta) \quad \text{and} \quad \phi[\delta] \circ \psi = \mathcal{T}(N, 2\delta). \tag{2.48}$$

(e) The interleaving distance of M and N is

$$d_I(M, N) := \inf \{\delta \geq 0 \mid \exists\ \delta\text{-interleaving}\}. \tag{2.49}$$

Remark 2.44 Recall that, given a δ-interleaving ϕ and ψ, in order to constitute morphisms $M \to N[\delta]$ and $N \to M[\delta]$, they must satisfy relations

$$\begin{aligned}\phi_t \circ M(s \leq t) = N(s+\delta \leq t+\delta) \circ \psi_s \quad \text{and} \quad \psi_t \circ N(s \leq t) \\ = M(s+\delta \leq t+\delta) \circ \phi_s\end{aligned} \tag{2.50}$$

for each $s \leq t$.

Now that we have metrics on persistent vector spaces and their barcode summaries, we can state the isometry theorem. This guarantees that the barcode is a stable summary of the input persistent vector space.

Theorem 2.45 (Isometry Theorem [1, Theorem 3.5]) *Given p.f.d persistent vector spaces $M, N \in$ **Persvec**,*

$$d_B(\mathcal{B}M, \mathcal{B}N) = d_I(M, N). \tag{2.51}$$

Finally, when a persistent vector space is tame, there are finitely many critical values $t_0 < \cdots < t_k$ such that if $t_{i-1} < s \leq t < t_i$ then $M(s \leq t)$ is an isomorphism [10]. Hence, all information of the persistent vector space is contained within the maps $M(t_{i-1} \leq t_i)$ for $i = 1, \ldots, k$. In particular, any interval in the barcode must have its endpoints at one of the critical values (or $\pm\infty$). In these scenarios, it suffices to consider M as a functor $\mathbf{[k]_\leq} \to \mathbf{Vec}$, where $\mathbf{[k]_\leq}$ is the sub-poset of $\mathbf{R}$ consisting of the integers $0, \ldots, k$ [31].

3 Motivation and Definition of GrPPH

Firstly, in Sect. 3.1, we describe a standard pipeline for extracting a topological summary from a weighted digraph, and illustrate a number of issues that naturally arise. Motivated by this in Sect. 3.2, we alter the standard pipeline in order to define a 'grounded pipeline'. We prove that this new pipeline is functorial in an appropriate sense, which we will later exploit for stability results. The pipeline is parameterised by two choices; in Sect. 3.3 we fix these choices in order to define our proposed descriptor.

3.1 Standard Pipeline

A typical TDA pipeline for weighted digraphs consists of three ingredients:

1. a map $F : \mathrm{Obj}(\mathbf{WDgr}) \to \mathrm{Obj}([\mathbf{R}, \mathbf{Dgr}])$, which assigns a filtration of digraphs to every weighted digraph;
2. a chain complex functor $C : \mathbf{Dgr} \to \mathbf{Ch}$ which maps each digraph to a chain complex and induces chain map for every digraph map; and finally

3. a choice of homology functor $H_k : \mathbf{Ch} \to \mathbf{Vec}$ in some degree k.

These components can then be combined into the following pipeline.

$$\mathcal{H}_k : \mathbf{WDgr} \xrightarrow{F} [\mathbf{R},\ \mathbf{Dgr}] \xrightarrow{[\mathbf{R},\, C]} [\mathbf{R},\ \mathbf{Ch}] \xrightarrow{[\mathbf{R},\, H_k]} [\mathbf{R},\ \mathbf{Vec}]$$

Examples of this pipeline include persistent homology of the directed flag complex [29] and persistent path homology [13]. However, it is possible to use these pipelines with filtrations of digraphs that do not arise from an initial weighted digraph.

We obtain a map $\mathcal{H}_k : \mathrm{Obj}(\mathbf{WDgr}) \to \mathrm{Obj}([\mathbf{R},\ \mathbf{Vec}])$ given by $\mathcal{H}_k := [\mathbf{R},\ H_k] \circ [\mathbf{R},\ C] \circ F$. Since F is not a priori functorial, neither is $\mathcal{H}_k$. Under mild assumptions on F and C, it is possible to define a subcategory of $\mathbf{WDgr}$ which makes this pipeline functorial.

Notation 3.1 Given $F : \mathrm{Obj}(\mathbf{WDgr}) \to \mathrm{Obj}([\mathbf{R},\ \mathbf{Dgr}])$, $G \in \mathbf{WDgr}$ and $s \le t$ we write

(a) $F^t G := F(G)(t)$, the image of t under the functor $F(G)$; and
(b) $\iota(s, t) := F(G)(s \le t)$, the image of $s \le t$ under the functor $F(G)$.

Definition 3.2 (a) A filtration map is any map $F : \mathrm{Obj}(\mathbf{WDgr}) \to \mathrm{Obj}([\mathbf{R},\ \mathbf{InclDgr}])$ such that $V(F^t G) \subseteq V(G)$ for all $t \in \mathbb{R}$. In particular, $\iota(s, t)$ must always be an inclusion.

(b) Given a filtration map F, a morphism of weighted digraphs $f \in \mathrm{Mor}_{\mathbf{WDgr}}(G, H)$ is called F-compatible if for every $t \in \mathbb{R}$ the underlying vertex map $f : V(G) \to V(H)$ restricts to a vertex map $V(F^t G) \to V(F^t H)$ which in turn yields a digraph map $F^t G \to F^t H$.

(c) Given a filtration map F, the F-compatible category of weighted digraphs, $\mathbf{WDgr}_F$, is the subcategory of $\mathbf{WDgr}$ such that $\mathrm{Obj}(\mathbf{WDgr}_F) = \mathrm{Obj}(\mathbf{WDgr})$ and

$$\mathrm{Mor}(\mathbf{WDgr}_F) = \{f \in \mathrm{Mor}(\mathbf{WDgr}) \mid f \text{ is } F\text{-compatible}\}. \tag{3.1}$$

Lemma 3.3 *Any filtration map $F : \mathrm{Obj}(\mathbf{WDgr}) \to \mathrm{Obj}([\mathbf{R},\ \mathbf{InclDgr}])$ induces a functor $F : \mathbf{WDgr}_F \to [\mathbf{R},\ \mathbf{Dgr}]$, which we call a filtration functor.*

Proof Given $f \in \mathrm{Mor}_{\mathbf{WDgr}_F}(G, H)$, since f is F-compatible, the underlying vertex map induces digraph maps $f : F^t G \to F^t H$ for every $t \in \mathbb{R}$. Given $s \le t$, both $F(G)(s \le t)$ and $F(H)(s \le t)$ are digraph maps induced by the inclusion vertex map. Hence the following square of morphisms in $\mathbf{Dgr}$ commutes.

$$\begin{tikzcd} F^s G \arrow[r, "f"] \arrow[d, "F(G)(s\le t)"'] & F^s H \arrow[d, "F(H)(s\le t)"] \\ F^t G \arrow[r, "f"'] & F^t H \end{tikzcd}$$

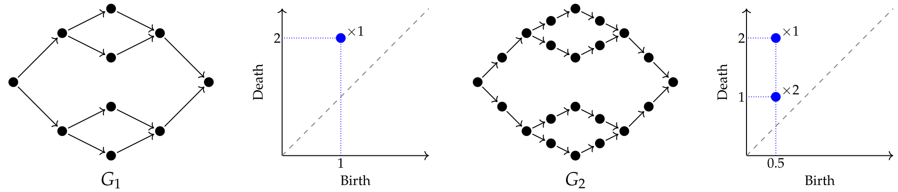

Fig. 3 Persistent path homology of the shortest-path filtration of flow through a bifurcation network, before and after edge subdivision. In G_1 all edges have unit weight, in G_2 all edges have weight 0.5

Hence f induces a natural transformation between $F(G)$ and $F(H)$. Moreover, since each $f : F^t G \to F^t H$ is fully determined by the underlying vertex map $V(G) \to V(H)$, this construction is certainly functorial. □

Remark 3.4 Since any filtration map induces a filtration functor, it suffices to define a filtration functor only as a map on objects.

When F is a filtration map, it induces a functor $\mathbf{WDgr}_F \to [\mathbf{R}, \mathbf{Dgr}]$ and hence $\mathcal{H}_k$ is a functor $\mathbf{WDgr}_F \to [\mathbf{R}, \mathbf{Vec}]$, as desired. We will now consider an illustrative example of this pipeline. Assuming the weight of an edge corresponds to a distance between its endpoints (e.g. the time it takes for a particle to flow down the edge), a natural choice of filtration functor is the following.

Definition 3.5 The shortest-path filtration is a map $F_d : \mathbf{WDgr} \to [\mathbf{R}, \mathbf{Dgr}]$. For $G = (V, E, w) \in \mathbf{WDgr}$ and $t \in \mathbb{R}$, we define

$$F_d(G)(t) := G^t := (V, E^t) \quad \text{where} \quad E^t := \{(i, j) \in (V \times V) \setminus \Delta_V \mid d(i, j) \leq t\} \tag{3.2}$$

and d is the shortest-path quasimetric on G. For $s \leq t$, the digraph map $G^s \to G^t$ is induced by the identity vertex map id_V.

Example 3.6 Choosing $F = F_d$ as above and C to be the regular path complex, we obtain a functor $\mathcal{H}_k : \mathbf{WDgr}_F \to [\mathbf{R}, \mathbf{Vec}]$. The shortest-path quasimetric of a weighted digraph is a directed network and this pipeline measures the persistent path homology of that network. This pipeline was first considered in [13] for cycle networks, alongside a stability analysis of persistent path homology for arbitrary directed networks.

In Fig. 3 we apply this pipeline (with $k = 1$) to a small bifurcating network G_1, a toy model for vasculature networks, in which each edge is given unit weight. The resulting barcode has a single feature with lifetime $[1, 2)$. The second network, G_2, is obtained by subdividing each edge in G_1, giving all edges weight 0.5. The resulting barcode is $\{[0.5, 1), [0.5, 1), [0.5, 2)\}$ which has three features.

This example highlights three key issues with this pipeline:

(a) the number of features in the barcode changes upon subdivision;
(b) loops bounded by triangles or long squares are ‘killed’ as soon as they are born; and

(c) the birth-time of each feature is an artefact of the 'resolution' of the weighted digraph.

A subtler issue arises when we attempt to interpret the diagram. A feature born at time t is supported on edges of the digraph G^t, which may not be edges in the original weighted digraph G. This makes interpretation of features more challenging. Note, this issues does not arise for either G_i shown in Fig. 3 since, in each case, there is some T such that $G_i^T = G_i$ and moreover $E(G_i^t) = \emptyset$ for all $t < T$.

3.2 Grounded Pipeline

We now describe an alteration to the standard pipeline which alleviates these issues by including the underlying digraph G in degree 1 for all $t \in \mathbb{R}$. The main distinction is that we do not factor through a filtration functor F. Instead, we use F and $C \in \mathbf{[Dgr,\ Ch]}$ to construct a new functor ${}^{g}C_F : \mathbf{WDgr}_F \to \mathbf{[R,\ Ch]}$.

$$
\begin{tikzcd}
\mathbf{WDgr} \arrow[r, "F"] \arrow[rr, dashed, bend left, "{}^{g}C_F"] & \mathbf{[R,\ Dgr]} \arrow[r, "{[\mathbf{R}, C]}"] & \mathbf{[R,\ Ch]} \arrow[r, "{[\mathbf{R}, H_1]}"] & \mathbf{[R,\ Vec]}
\end{tikzcd}
$$

In order to define the map on objects, we only need a weaker condition on C.

Definition 3.7 Given a filtration functor $F : \mathbf{WDgr}_F \to \mathbf{[R,\ Dgr]}$, a functor $C : \mathbf{InclDgr} \to \mathbf{Ch}$, $G \in \mathbf{WDgr}$ and $t \in \mathbb{R}$, the chain complex ${}^{g}C_\bullet(G, t; F)$ is the top row of the following diagram.

$$
\begin{tikzcd}
\cdots C_3(F^tG) \arrow[r, "\partial_3"] & C_2(F^tG) \arrow[rr, "\iota_\# \circ \partial_2"] \arrow[dr, dotted, "\partial_2"'] & & C_1(G \cup F^tG) \arrow[rr, "\partial_1"] & & C_0(G \cup F^tG) \cdots \\
& & C_1(F^tG) \arrow[ur, dotted, hook, "\iota_\#"] \arrow[rr, dotted, "\partial_1"'] & & C_0(F^tG) \arrow[ur, dotted, hook, "\iota_\#"'] &
\end{tikzcd}
$$

In the above, $\iota : F^tG \hookrightarrow G \cup F^tG$ is the inclusion digraph map, induced by the inclusion vertex map. Then $\iota_\# = C(\iota)$ is the image of this map under the functor C. The boundary maps ∂_k are derived either from the chain complex $C_\bullet(F^tG)$ or $C_\bullet(G \cup F^tG)$.

We denote the chain groups as ${}^{g}C_k(G, t; F)$ and the boundary maps as ${}^{g}\partial_k^t$. When F and t are clear from context, we omit them from notation

We use the prescript ${}^{g}\Box$ to denote that this chain complex is *grounded*; as we will show in Lemma 4.2, after appropriate choices of F and C, all degree 1 homology classes have representatives in the underlying digraph.

Lemma 3.8 *For each $t \in \mathbb{R}$ and $G \in \mathbf{WDgr}$, $({}^{g}C_\bullet(G, t; F), {}^{g}\partial_\bullet^t)$ is a chain complex.*

Proof Since $\iota_\#$ is a chain map $C(F^tG) \to C(G \cup F^tG)$ we have

$$\partial_1 \circ (\iota_\# \circ \partial_2) = \iota_\# \circ \partial_1 \circ \partial_2 = 0 \tag{3.3}$$

and hence $C_\bullet(G, t; F)$ defines a chain complex. $\square$

Lemma 3.9 *Fix a filtration functor* $F : \mathbf{WDgr}_F \to [\mathbf{R}, \mathbf{Dgr}]$ *and a functor* $C : \mathbf{InclDgr} \to \mathbf{Ch}$. *Given* $G, H \in \mathbf{WDgr}$ *and a vertex map* $f : V(G) \to V(H)$, *suppose that* f *induces digraph maps* $f : F^sG \to F^tH$ *and* $f : G \cup F^sG \to H \cup F^tH$ *for some* $s \le t$. *Then there is a chain map*

$$J(f, s \le t) : {}^gC_\bullet(G, s; F) \to {}^gC_\bullet(G, t; F). \tag{3.4}$$

Moreover, given $g : V(K) \to V(H)$ *which satisfies the necessary conditions to induce* $J(g, r \le s)$,

$$J(f \circ g, r \le t) = J(f, s \le t) \circ J(g, r \le s). \tag{3.5}$$

Finally, for any $t \in \mathbb{R}$, $J(\mathrm{id}_{V(G)}, t \le t)$ *is the identity chain map.*

Proof To distinguish the two digraph maps induced by f, we denote them

$$f : F^sG \to F^tH \quad \text{and} \quad f' : G \cup F^sG \to H \cup F^tH. \tag{3.6}$$

First note that the two digraph maps induced by f commute with the relevant inclusions so that the following square commutes.

$$\begin{tikzcd}
F^sG \arrow[r, hook, "\iota"] \arrow[d, hook, "f"'] & G \cup F^sG \arrow[d, hook, "f'"] \\
F^tH \arrow[r, hook, "\iota"'] & H \cup F^tH
\end{tikzcd}$$

Applying the functor C to the digraph maps f and f' we obtain two chain maps $f_\# : C(F^sG) \to C(F^tH)$ and $f'_\# : C(G \cup F^sG) \to C(H \cup F^tH)$. These chain maps can be combined as in the following diagram.

$$\begin{tikzcd}
\cdots C_3(F^sG) \arrow[r, "\partial_3"] \arrow[ddd, hook, "f_\#"] \arrow[dddr, phantom, "\boxed{A}"] & C_2(F^sG) \arrow[rr, "\iota_\# \circ \partial_2"] \arrow[dr, dotted, "\partial_2"'] \arrow[ddd, hook, "f_\#"] \arrow[ddr, phantom, "\boxed{B}"] & & C_1(G \cup F^sG) \arrow[r, "\partial_1"] \arrow[ddd, hook, "f'_\#"] \arrow[dddr, phantom, "\boxed{D}"] & C_0(G \cup F^sG) \cdots \arrow[ddd, hook, "f'_\#"] \\
& & C_1(F^sG) \arrow[ur, dotted, hook, "\iota_\#"'] \arrow[d, hook, "f_\#"] \arrow[dr, phantom, "\boxed{C}"] & & \\
& & C_1(F^tH) \arrow[dr, dotted, hook, "\iota_\#"] & & \\
\cdots C_3(F^tH) \arrow[r, "\partial_3"'] & C_2(F^tH) \arrow[rr, "\iota_\# \circ \partial_2"'] \arrow[ur, dotted, "\partial_2"] & & C_1(H \cup F^tH) \arrow[r, "\partial_1"'] & C_0(H \cup F^tH) \cdots
\end{tikzcd}$$

Squares $\boxed{A}$ and $\boxed{B}$ and $\boxed{D}$ commute because all vertical maps are components of the same chain map. Square $\boxed{C}$ commutes because it is the image of the initial square of commuting digraph maps under the functor C, restricted to degree 1. Therefore, the vertical solid arrows constitute a chain map between the required chain complexes, which we denote $J(f, s \le t)$. The remaining functorial relations follow automatically because in each degree $J(f, s \le t)$ coincides with either $f_\#$ or $f'_\#$, which each satisfy the relations. $\square$

Lemma 3.10 *Given a filtration functor $F : \mathbf{WDgr}_F \to [\mathbf{R}, \mathbf{Dgr}]$ and a functor $C : \mathbf{InclDgr} \to \mathbf{Ch}$, the chain complex ${}^gC_\bullet(G, t; F)$ is functorial in t.*

Proof For each $s \leq t$ we require chain maps ${}^g\iota(s,t)_\# : {}^gC_\bullet(G, s) \to {}^gC_\bullet(G, t)$ which satisfy the usual functorial axioms in t. First note that that the identity map $\mathrm{id} : V(G) \to V(G)$ induces a digraph map $\mathrm{id} : G \to G$. Moreover, for each $s \leq t$ the identity induces digraph maps $\iota(s,t) : F^sG \to F^tG$ since F is a filtration functor. Hence, the identity also induces digraph maps $G \cup F^sG \to G \cup F^tG$. Applying Lemma 3.9, we obtain a chain map $J(\mathrm{id}, s \leq t) : {}^gC_\bullet(G, s) \to {}^gC_\bullet(G, t)$. Now, given $s \leq t \leq r$, Lemma 3.9 implies that

$$J(\mathrm{id}, s \leq r) = J(\mathrm{id}, t \leq r) \circ J(\mathrm{id}, s \leq t). \tag{3.7}$$

and moreover $J(\mathrm{id}, t \leq t)$ is the identity chain map on ${}^gC(G, t)$. Therefore, we can take ${}^g\iota(s,t)_\# := J(\mathrm{id}, s \leq t)$. □

Definition 3.11 Given a filtration functor $F : \mathbf{WDgr}_F \to [\mathbf{R}, \mathbf{Dgr}]$ and a functor $C : \mathbf{InclDgr} \to \mathbf{Ch}$, the map ${}^gC_F : \mathrm{Obj}(\mathbf{WDgr}) \to \mathrm{Obj}([\mathbf{R}, \mathbf{Ch}])$ is given on objects by

$${}^gC_F(G)(t) := ({}^gC_\bullet(G, t; F), {}^g\partial^t_\bullet) \tag{3.8}$$

and the morphism ${}^gC_F(G)(s \leq t)$ is as constructed in the proof of Lemma 3.10.

Notation 3.12 (a) We denote the induced map on chain complexes by ${}^g\iota(s,t)_\# := {}^gC_F(G)(s \leq t)$.
(b) In each homology degree k, we denote the induced map on homology by ${}^g\iota(s,t)_* := [\mathbf{R}, H_k]({}^g\iota(s,t)_\#)$.

Thanks to Lemma 3.10, ${}^gC_F$ gives us a map $\mathrm{Obj}(\mathbf{WDgr}) \to \mathrm{Obj}([\mathbf{R}, \mathbf{Ch}])$, which we can compose with homology to obtain a persistent vector space. Under additional functorial assumptions on C, when we restrict to the appropriate category ${}^gC_F$ becomes a functor.

Theorem 3.13 *Given a filtration functor $F : \mathbf{WDgr}_F \to [\mathbf{R}, \mathbf{Dgr}]$ and a functor $C : \mathbf{Dgr} \to \mathbf{Ch}$, ${}^gC_F$ is a functor ${}^gC_F : \mathbf{WDgr}_F \to [\mathbf{R}, \mathbf{Ch}]$.*

Proof Given $f \in \mathrm{Mor}_{\mathbf{WDgr}_F}(G, H)$ and $t \in \mathbb{R}$, we need a chain map ${}^gf_\# : {}^gC_\bullet(G, t) \to {}^gC_\bullet(H, t)$. Moreover, f must satisfy the usual functorial axioms, and given $s \leq t$, commute with the chain maps ${}^g\iota(s,t)_\#$.

Note that f is given by a vertex map $f : V(G) \to V(H)$ which induces a digraph map $G \to H$. Moreover, since f is F-compatible, it induces digraph maps $F^tG \to F^tH$. Therefore f must also induce digraph maps $G \cup F^tG \to H \cup F^tH$. Hence, using Lemma 3.9, we can take ${}^gf_\# := J(f, t \leq t)$. The functorial axioms follow immediately from the functorial relations shown in Lemma 3.9. Moreover, since ${}^g\iota(s,t)_\#$ is also constructed via Lemma 3.9, Eq. (3.5) implies

$${}^gf_\# \circ {}^g\iota(s,t)_\# = J(f \circ \mathrm{id}_{V(G)}, s \leq t) = J(\mathrm{id}_{V(H)} \circ f, s \leq t) = {}^g\iota(s,t)_\# \circ {}^gf_\# \tag{3.9}$$

and hence ${}^g f_\#$ constitutes a morphism ${}^g C_F(G) \to {}^g C_F(H)$. □

Notation 3.14 Given a morphism $f \in \mathrm{Mor}(\mathbf{WDgr}_F)$, we denote the induced map on chain complexes, constructed above, by ${}^g f_\# := {}^g C_F(f)$ and the induced map on homology in degree k by ${}^g f_* := [\mathbf{R}, H_k]({}^g f_\#)$.

3.3 Definition of GrPPH

In order to investigate properties and stability of the grounded pipeline, we make choice for both F and C. As we have already discussed, since we interpret edge-weights as a measure of distance, a natural choice for F is the shortest-path filtration. Choices for C include the regular path complex, non-regular path complex and the directed flag complex. However, the latter two constructions are *not* functors $\mathbf{Dgr} \to \mathbf{Ch}$. **Henceforth, for the rest of the paper, we fix F to be the shortest-path filtration and C to be the regular path complex**,

$$F = F_d \quad \text{and} \quad C = \Omega. \tag{3.10}$$

Since F is fixed, we will largely remove it from notation. We also use C instead of Ω.

Lemma 3.15 *The F_d-compatible category of weighted digraphs is the contraction category of weighted digraphs (see Definition 2.15),* $\mathbf{WDgr}_{F_d} = \mathbf{ContWDgr}$.

Proof First note $f \in \mathrm{Mor}(\mathbf{WDgr})$ is precisely a vertex map $f : V(G) \to V(H)$ which induces a digraph map $G \to H$. A vertex map $f : V(G) \to V(H)$ induces a digraph map $G^t \to H^t$ for every $t \in \mathbb{R}$ if and only if

$$d(f(i), f(j)) \leq d(i, j) \tag{3.11}$$

for every $i, j \in V(G)$. Hence, $f \in \mathrm{Mor}(\mathbf{WDgr})$ is F_d-compatible if and only if it is a contraction map. So a morphism $f \in \mathrm{Mor}(\mathbf{WDgr}_{F_d})$ is precisely a contraction digraph map $G \to H$. □

Now that we understand the category $\mathbf{WDgr}_{F_d}$, applying Theorem 3.13 yields a functor ${}^g C : \mathbf{ContWDgr} \to [\mathbf{R}, \mathbf{Ch}]$. Taking the first homology yields a persistent vector space in a functorial way; this functor is our proposed invariant for weighted digraphs.

Definition/Theorem 3.16 *Grounded persistent path homology* (GrPPH) *is the functor*

$${}^g\mathcal{H}_1 := [\mathbf{R}, H_1] \circ {}^g C : \mathbf{ContWDgr} \to \mathbf{PersVec} \tag{3.12}$$

from the contraction category of weighted digraphs (see Definition 2.15) to the category of persistent vector spaces (see Definition 2.38).

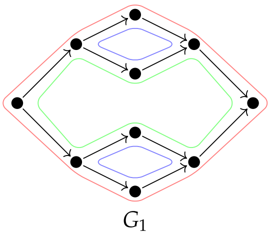

Fig. 4 A bifurcation network, before and after subdivision; all edges in G_1 have unit weight, all edges in G_2 have weight 0.5 (as in Fig. 3). Highlighted in red, green and blue are circuits whose representatives generate ${}^g\mathcal{H}_1(G_i)$ (Color figure online)

Notation 3.17 Given $G \in \mathbf{WDgr}$, in degree k at filtration step t, we denote

$$\begin{aligned}
&\text{the space of grounded } k\text{-cycles by} && {}^gZ(G,t) := \ker {}^g\partial_k^t; && (3.13)\\
&\text{the space of grounded } k\text{-boundaries by} && {}^gB(G,t) := \operatorname{im} {}^g\partial_{k+1}^t && (3.14)\\
&\text{the (degree } k\text{) grounded homology by} && {}^gH_k(G,t) := \frac{{}^gZ_k(G,t)}{{}^gB_k(G,t)}. && (3.15)
\end{aligned}$$

Remark 3.18 Note that for any $t \in \mathbb{R}$,

$$\begin{aligned}
k > 1 &\implies {}^gH_k(G,t) \cong H_k(G^t),\\
k < 1 &\implies {}^gH_k(G,t) \cong H_k(G \cup G^t).
\end{aligned}$$

Therefore, the only new homology occurs in degree $k = 1$, since it compares 1-cycles in $G \cup F^tG$ with 1-boundaries from G^t. This justifies our focus on degree 1 homology in Definition/Theorem 3.16.

Example 3.19 In Fig. 4, we consider again the bifurcating network example of Fig. 3, in which all edges have weight 1. We see the barcodes of the grounded persistent homology are both

$$\mathcal{B}^g\mathcal{H}_1(G_1) = \mathcal{B}^g\mathcal{H}_1(G_2) = \{[0,1),[0,1),[0,2)\}\,. \tag{3.16}$$

In G_1 the two $[0, 1)$ features correspond to the smaller 4-node circuits in the centre of the network, coloured in blue. These circuits birth homological cycles in $G \cup G^t$ at $t = 0$, which are then killed by long squares when the edges appear in G^t at $t = 1$.

The $[0, 2)$ feature corresponds to the large inner circuit (coloured in green). Again this circuit births a homological cycle at $t = 0$ which then becomes null-homologous at $t = 2$ when shortcut edges give rise to a new long square.

The outer red cycle is a linear combination of the inner green and blue cycles, hence it does not give rise to a fourth feature in the barcode. Moreover, at $t = 1$ the red and green cycles becomes homologous.

In G_2 the features correspond to the same circuits (once subdivided).

4 Interpretation of GrPPH

4.1 Decreasing Betti Curves

In the first example we considered (Example 3.19), we saw that all features were born at $t = 0$. Indeed, this is always the case and ${}^{g}H_1(G, 0)$ is in fact the cycle space of $\mathcal{U}(G)$ (with $\mathbb{R}$-valued coefficients).

Lemma 4.1 *Given a digraph $G = (V, E, w)$, two distinct nodes $a, b \in V$ and a trail $p : a \rightsquigarrow b$, then for all $t \geq \mathrm{len}(p)$*

$$\mathfrak{R}(p) = \sum_{\tau \in E(p)} \tau = ab \pmod{{}^{g}B_1(G, t)}. \tag{4.1}$$

Proof Fix arbitrary $t \geq \mathrm{len}(p)$ and denote the vertices of the path as $a = v_0, \dots, v_m = b$. Whenever $i < j$ we can truncate p to obtain a path $v_i \rightsquigarrow v_j$ of length at most t and so $(v_i, v_j) \in E(G^t)$. Hence, whenever $i < j < k$, there is a directed triangle $v_i v_j v_k \in C_2(G^t)$ and hence $v_i v_k = v_i v_j + v_j v_k \pmod{{}^{g}B_1(G, t)}$. Therefore, inductively we can write

$$\begin{aligned} ab = v_0 v_m &= v_0 v_1 + v_1 v_m = v_0 v_1 + v_1 v_2 + v_2 v_m \\ &= \cdots = \sum_{i=1}^{m} v_{i-1} v_i \pmod{{}^{g}B_1(G, t)} \end{aligned} \tag{4.2}$$

as required. □

Proposition 4.2 *Fix a weighted digraph (G, w) and $t \geq 0$. For any cycle $v \in {}^{g}Z_1(G, t)$, there is an initial cycle $v' \in {}^{g}Z_1(G, 0)$, supported on the edges of G, such that v is homologous to ${}^{g}\iota(0, t)_{\#} v'$.*

Proof Given any edge $\tau = (a, b) \in E(G^t)$, there is a path $p : a \rightsquigarrow b$ in G of length at most t. Denoting the edges of p by $(\tau_1, \dots, \tau_m)$, Lemma 4.1 tells us, $\tau = \sum_{i=1}^{m} \tau_i \pmod{{}^{g}B_1(G, t)}$. Now, since $\tau_i \in E(G)$, we see ${}^{g}\iota(0, t)_{\#} \tau_i = \tau_i$ for each edge in p. □

Corollary 4.3 *Given $G \in$ **WDgr**,*

(a) any interval in $\mathcal{B}{}^{g}\mathcal{H}_1(G)$ has birth time 0;
(b) ${}^{g}H_1(G, 0)$ has a basis of simple undirected circuits; and
(c) $\#\,\mathrm{Dgm}({}^{g}\mathcal{H}_1(G))$ coincides with the circuit rank of the underlying undirected graph.

Proof The first point follows immediately from Proposition 4.2. To see the final two points, consider the chain complex, ${}^{g}C_\bullet(G, 0)$ at the start of the filtration. Since G^0 has no edges, the chain complex is simply

$$\cdots 0 \longrightarrow 0 \longrightarrow C_1(G) \longrightarrow C_0(G) \cdots$$

Since G is an orientation of $\mathcal{U}(G)$, the first homology of this chain complex is precisely the real cycle space of $\mathcal{U}(G)$. Fix an arbitrary spanning forest T of $\mathcal{U}(G)$, and label the remaining edges $e_1, \dots, e_k$. Note k is the circuit rank of $\mathcal{U}(G)$. Let p_i denote the simple undirected circuit in G which traverses e_i and then returns to $\mathrm{st}(e_i)$ through the unique path in T. Then $\{\mathfrak{R}(p_1), \dots, \mathfrak{R}(p_k)\}$ is a basis for ${}^{g}H_1(G, 0)$. □

4.2 Circuit Lifetimes

While all features in the barcode (and hence all cycles) are born at time $t = 0$, their death times generally differ. We can assign a death-time to any cycle $v \in {}^{g}Z_1(G, 0)$, as the first time v becomes null-homologous.

Definition 4.4 Given $v \in {}^{g}Z_1(G, 0)$, the death-time of v is

$$\mathcal{D}(v) := \inf \left\{ t \geq 0 \;\middle|\; {}^{g}\iota(0, t)_*[v] = 0 \right\} \tag{4.3}$$

where we let $\mathcal{D}(v) := \infty$ if there is no such t. The lifetime of v is the interval $\mathcal{L}(v) := [0, \mathcal{D}(v))$.

Remark 4.5 Let p be an undirected circuit in G and p' the same circuit, traversed in the opposite direction so that $\mathfrak{R}(p') = -\mathfrak{R}(p)$. Since ${}^{g}\iota(0, t)_*$ is linear, $\mathcal{D}(\mathfrak{R}(p)) = \mathcal{D}(\mathfrak{R}(p'))$. Also, since $\mathfrak{R}(p)$ does not depend on the starting vertex of p, neither does $\mathcal{D}(\mathfrak{R}(p))$.

This pipeline gives us a method for associating a lifetime to an undirected circuit in G which is ‘geometric’ in the sense that it does not depend on the starting vertex or direction. The length of this lifetime gives us a ‘size’ to the circuit from the perspective of the filtration. Since we use the shortest-path filtration, we interpret this size as the time it takes for the flow to ‘fill in’ the circuit.

Lemma 4.6 *Given $G = (V, E, w) \in$ **WDgr** and two directed paths $p_1, p_2 : a \rightsquigarrow b$ between distinct vertices $a, b \in V$, let p_c denote the undirected circuit which traverses p_1 forwards and then p_2 in reverse. For $i = 1, 2$, define*

$$h_i := \min \left\{ t \geq 0 \mid \exists v_i \in V \text{ along } p_i \text{ such that } d(a, v_i) \leq t \text{ and } d(v_i, b) \leq t \right\}. \tag{4.4}$$

Then $\mathcal{D}(p_c) \leq \max(h_1, h_2)$.

Proof Denote $T := \max(h_1, h_2)$. First assume that there are at least 2 edges in each p_i. Then, by the definition of h_i, there exists $v_i \in V \setminus \{a, b\}$ along each p_i such that $d(a, v_i) \leq T$ and $d(v_i, b) \leq T$. Hence there is a long square $av_1b - av_2b \in {}^{g}C_2(G, T)$. By Lemma 4.1,

$$0 = {}^{g}\partial_2(av_1b - av_2b) = \mathfrak{R}(p_c) \pmod{{}^{g}B_1(G, T)}. \tag{4.5}$$

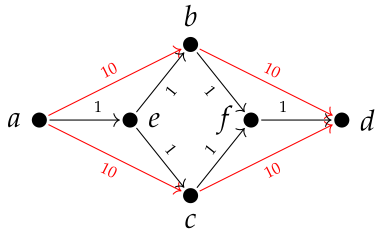

Fig. 5 A simple example of a weighted digraph for which the bound of Lemma 4.6 fails to be sharp

Finally, if p_1 contains one edge then $h_1 = \mathrm{len}(p_1)$ so $d(a, b) \leq h_1 \leq T$. The definition of h_2 ensures there is $v_2 \in V$ along p_2 such that $d(a, v_2), d(v_2, b) \leq T$. Hence there is a directed triangle $av_2b \in {}^g C_2(G, T)$. By Lemma 4.1,

$$0 = {}^g\partial_2(av_2b) = \mathfrak{R}(p_c) \pmod{{}^g B_1(G, T)} \tag{4.6}$$

which concludes the proof. □

Example 4.7 Note that the bound of Lemma 4.6 is by no means sharp. Consider for example Fig. 5. Let p_1 be the outer red path (a, b, d), p_2 the lower red path (a, c, d) and p_c the undirected circuit which traverse p_1 forward then p_2 in reverse. Then $h_1 = h_2 = 10$ but $\mathcal{D}(\mathfrak{R}(p_c)) = 2$.

To see this is the correct death time, first note that ${}^g H_1(G, t)$ can only change at integer values. At $t = 1$, the only edges present in G^t are the black ones drawn in in Fig. 5. Hence, $C_2(G, 1)$ is generated by the long square $ebf - ecf$, whose boundary is not $\mathfrak{R}(p_c)$.

However, at $t = 2$ the edges (a, b), (b, d), (a, c) and (c, d) also appear in G^t. These generate additional long squares and directed triangles. In particular $abd - acd \in C_2(G, 2)$ and

$${}^g\partial_2(abd - acd) = \mathfrak{R}(p_c). \tag{4.7}$$

4.3 Representatives

After computing persistent homology, it is common to compute homological cycles which represent the intervals in the barcode. In general, these "representatives" are not unique and may be quite complicated. In practice, one can often compute representatives with integer (and even unit) coefficients [27] and these representatives are frequently used for interpreting features (e.g. [2, 34]).

Definition 4.8 A persistence basis for ${}^g\mathcal{H}_1(G)$ is a choice of initial cycles $B = \{b_i \in {}^g Z_1(G, 0)\}$ such that for each $t \geq 0$, the set $\{{}^g\iota(0, t)_*[b_i]\} \setminus \{0\}$ yields a basis for ${}^g H_1(G, t)$. We call elements of a persistence basis representatives.

Lemma 4.9 *Given any $G \in$ **WDgr**, a persistence basis for ${}^g\mathcal{H}_1(G)$ always exists.*

Proof This follows from the structure theorem (Theorem 2.39) and Corollary 4.3. □

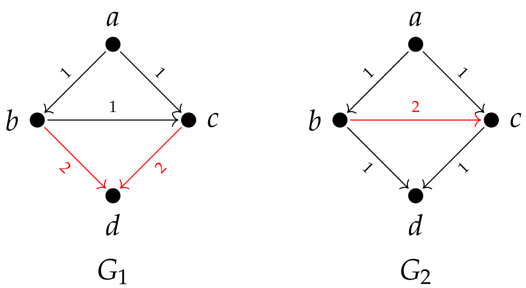

Fig. 6 Two weighted digraphs with the same underlying digraph but different persistence bases, illustrating that not every circuit basis of ${}^{g}H_1(G, 0)$ yields a persistence basis for ${}^{g}\mathcal{H}_1(G)$

Obtaining a persistence basis $B \subseteq {}^{g}Z_1(G, 0)$ is desirable because the constituent cycles *represent* the features of the barcode, in the following sense. If the barcode is $\mathcal{B}^{g}\mathcal{H}_1(G) = \{\!\{I_1, \ldots, I_m\}\!\}$ then there is an ordering on the cycles $B = \{b_1, \ldots, b_m\}$ such that $\mathcal{L}(b_i) = I_i$ and

$$\bigoplus_{i=1}^{m} P(I_i) \cong {}^{g}\mathcal{H}_1(G). \tag{4.8}$$

Moreover, the isomorphism $\phi : \oplus_{i=1}^{m} P(I_i) \to {}^{g}\mathcal{H}_1(G)$ is given by mapping

$$1 \in I_i(t) \mapsto {}^{g}\iota(0, t)_*[b_i] \text{ whenever } t \in I_i. \tag{4.9}$$

This mapping gives an isomorphism because $\{{}^{g}\iota(0, t)_*[b_i]\} \setminus \{0\}$ is always a basis for ${}^{g}H_1(G, t)$. In this sense, the representatives in B generate ${}^{g}\mathcal{H}_1(G)$.

Representatives live in ${}^{g}Z_1(G, 0)$ so they are just $\mathbb{R}$-linear combinations of edges in G. However, a priori, the coefficients of these linear combinations may be arbitrarily complicated. The goal of this section is to show that grounded persistent homology always admits a *geometrically interpretable* persistence basis, in the following sense.

Theorem 4.10 *Given $G \in$ **WDgr** with circuit rank m there exist undirected circuits $p_1, \ldots, p_m$ in G, such that $\{\mathfrak{R}(p_1), \ldots, \mathfrak{R}(p_m)\}$ is a persistence basis for ${}^{g}\mathcal{H}_1(G)$.*

Example 4.11 First, we note that it does not suffice to chose *any* basis of undirected circuits for ${}^{g}Z_1(G, 0)$. For example, consider the two weighted digraphs pictured in Fig. 6. In both digraphs, ignoring choice of direction, there are three undirected simple circuits, whose representatives we denote

$$\gamma_1^i := ab + bc - ac, \tag{4.10}$$
$$\gamma_2^i := bc + cd - bd, \tag{4.11}$$
$$\gamma_3^i := ab + bd - cd - ac. \tag{4.12}$$

where $\gamma_j^i \in {}^{g}Z_1(G_i, 0)$. Note $\gamma_3^i = \gamma_1^i - \gamma_2^i$. The GRPPH of these two weighted digraphs is

$$\mathcal{B}^{g}\mathcal{H}_1(G_1) = \{\!\{[0, 2), [0, 1)\}\!\} \quad \text{and} \quad \mathcal{B}^{g}\mathcal{H}_1(G_2) = \{\!\{[0, 2), [0, 1)\}\!\}. \tag{4.13}$$

A persistence basis for ${}^g\mathcal{H}_1(G_1)$ is $\{\gamma_1^1, \gamma_2^1\}$ with $\mathcal{L}(\gamma_1^1) = [0, 1)$ and $\mathcal{L}(\gamma_2^1) = [0, 2)$. Note that $\{\gamma_2^1, \gamma_3^1\}$ is *not* a persistence basis for ${}^g\mathcal{H}_1(G_1)$ because $\mathcal{L}(\gamma_3^1) = [0, 2)$.

However $\mathcal{L}(\gamma_1^2) = \mathcal{L}(\gamma_2^2) = [0, 2)$, hence $\{\gamma_1^2, \gamma_2^2\}$ does *not* form a persistence basis for ${}^g\mathcal{H}_1(G_2)$. At $t = 2$, we see ${}^g\iota(0, 2)_*\gamma_3^2 = 0$ and hence ${}^g\iota(0, 2)_*\gamma_1^2 = {}^g\iota(0, 2)_*\gamma_2^2$. Instead, a persistence basis for ${}^g\mathcal{H}_1(G_2)$ is $\{\gamma_2^2, \gamma_3^2\}$.

This illustrates that an arbitrary choice of undirected circuit basis for ${}^g H_1(G, 0)$ may not yield a persistence basis of ${}^g\mathcal{H}_1(G)$. Moreover, a correct choice of basis does not depend only on $\mathcal{U}(G)$; we must incorporate information about how cycles in ${}^g Z_1(G, 0)$ die, in order to choose a persistence basis.

To begin tackling Theorem 4.10, since G is finite, we note there are finitely many critical values $t_1 = 0, \ldots, t_m$ where the chain complex ${}^g C_\bullet(G, t)$ changes. Therefore, it suffices to study the following finite persistent chain complex instead.

$$\begin{array}{ccccccc}
\cdots C_2(G^{t_1}) & \longrightarrow & C_1(G \cup G^{t_1}) & \longrightarrow & C_0(G \cup G^{t_1}) \cdots & & {}^g H_1(G, t_1) \\
\downarrow & & \downarrow & & \downarrow & & \downarrow\!\!\downarrow {}^g\iota(t_1, t_2)_* \\
\cdots C_2(G^{t_2}) & \longrightarrow & C_1(G \cup G^{t_2}) & \longrightarrow & C_0(G \cup G^{t_2}) \cdots & & {}^g H_1(G, t_2) \\
\downarrow & & \downarrow & & \downarrow & & \downarrow\!\!\downarrow {}^g\iota(t_2, t_3)_* \\
\vdots & & \vdots & & \vdots & & \vdots \\
\downarrow & & \downarrow & & \downarrow & & \downarrow\!\!\downarrow {}^g\iota(t_{m-1}, t_m)_* \\
\cdots C_2(G^{t_m}) & \longrightarrow & C_1(G \cup G^{t_m}) & \longrightarrow & C_0(G \cup G^{t_m}) \cdots & & {}^g H_1(G, t_m)
\end{array}$$

To the right of the chain complex we show the induced maps on homology ${}^g\iota(t_{i-1}, t_i)_*$. By Lemma 4.2, these maps on homology are always surjective. Our strategy is to find undirected circuit bases for the kernel of each of these maps; Lemma 4.12 achieves this and is the key result. We then collect these elements into a basis for $\ker {}^g\iota(0, t_m)_*$. Together, these elements form representatives for the homology classes with finite lifetime. To obtain representatives for the infinite feature, we extend this to a basis for all of ${}^g H_1(G, 0)$ and show that we obtain a persistence basis.

Lemma 4.12 *For each $i = 2, \ldots, m$, there is a basis $\{b_1, \ldots, b_{k_i}\}$ of $\ker {}^g\iota(t_{i-1}, t_i)_*$ such that $b_j = {}^g\iota(0, t_{i-1})_*[\mathfrak{R}(p_{i,j})]$ for some undirected circuit $p_{i,j}$ in G.*

Proof For notational convenience, we define $r{:=}t_{i-1}$ and $s{:=}t_i$. When the filtration increases from $t = r$ to $t = s$, some number of edges are added to G^t which yield new generators for both $C_1(G \cup G^t)$ and $C_2(G^t)$. Our approach is to decompose ${}^g\iota(r, s)_\#$ into a sequence of chain maps. In the first, all the new generators of $C_1(G \cup C^s)$ are added, along with sufficient new generators in $C_2(G^s)$ to make the new edges homologous to a sum of edges already present in $C_1(G \cup G^r)$. Therefore, on homology this first map is an isomorphism. We then add the remaining generators of $C_2(G^s)$

one at a time in order to find a basis for $\ker {}^g\iota(r,s)_*$. Since we are only interested in degree 1 homology, it suffices to restrict our attention to degrees 0, 1 and 2. □ □

Denote the set of new edges $E_{new} := E(G \cup G^s) \setminus E(G \cup G^r)$. Given an edge $e = (a,b) \in E_{new}$, there is some directed path $p : a \rightsquigarrow b$ in G, of length at most s. Moreover this path must have at least one vertex distinct from the endpoints of e, otherwise $e \in G$. Choose arbitrary such $v_e \in V(p)$. Then (a, v_e, b) is a directed triangle in G^s so $w_e := a v_e b$ is a new generator of $C_2(G^s)$. Repeating this for all new edges we obtain a set of generators

$$U_{new} := \{w_e \mid e \in E_{new}\} \tag{4.14}$$

which were not present in $C_2(G^r)$ and are linearly independent. Define $W_0 := \langle U_{new} \rangle$ and let Q_0 denote the degree 1 homology of the chain complex

$$C_2(G^r) \oplus W_0 \xrightarrow{{}^g\partial_2} C_1(G \cup G^s) \xrightarrow{{}^g\partial_1} C_0(G \cup G^s)$$

which is a subcomplex of ${}^gC_\bullet(G,s)$.

Claim 4.13 *The inclusion chain map*

$$\begin{array}{ccccccc} C_2(G^r) & \longrightarrow & C_1(G \cup G^r) & \longrightarrow & C_0(G \cup G^r) & & {}^gH_1(G,r) \\ \downarrow q_2^0 & & \downarrow q_1^0 & & \downarrow q_0^0 & & q^0 \downarrow \cong \\ C_2(G^r) \oplus W_0 & \longrightarrow & C_1(G \cup G^s) & \longrightarrow & C_0(G \cup G^s) & & Q_0 \end{array}$$

induces an isomorphism on degree 1 *homology,* $q^0 : {}^gH_1(G,r) \to Q_0$.

Proof of Claim. We define a chain map $c_\#$ in the opposite direction and a homotopy $P : C_1(G \cup G^s) \to C_2(G^r) \oplus W_0$ such that $c_1 \circ q_1^0 = \mathrm{id}$ while $q_1^0 \circ c_1 - \mathrm{id} = {}^g\partial_2 \circ P$. Hence, on degree 1 homology, $c_\#$ induces an inverse to q^0. The chain map in degrees $k \neq 1$ is given by

$$c_k(v) := \begin{cases} v & \text{if } v \in {}^gC_k(G,r), \\ 0 & \text{otherwise,} \end{cases} \tag{4.15}$$

and c_1 is defined on the basis of $C_1(G \cup G^s)$ by

$$c_1(ab) := \begin{cases} a v_{(a,b)} + v_{(a,b)} b & \text{if } (a,b) \in E_{new}, \\ ab & \text{otherwise.} \end{cases} \tag{4.16}$$

The homotopy is given on the basis of $C_1(G \cup G^s)$ by

$$P(ab) := \begin{cases} w_{(a,b)} & \text{if } (a,b) \in E_{new}, \\ 0 & \text{otherwise.} \end{cases} \tag{4.17}$$

where $w_{(a,b)} \in U_{new}$. A standard check of the two cases verifies that $c_\#$ is a chain map and the relations $c_1 \circ q_1^0 = \mathrm{id}$ and $q_1^0 \circ c_1 - \mathrm{id} = {}^g\partial_2 \circ P$ hold.

Intuitively, $c_\#$ collapses a new edge $e = (a, b) \in E_{new}$ onto the sum of edges $av_e + v_e b$ in $G \cup G^r$. The homotopy P shows the two elements are homologous thanks to the presence of the directed triangle $w_e \in U_{new}$. □

By Proposition 2.33, there exists $u_1, \dots, u_n \in C_2(G^s)$ such that

$$C_2(G^s) = C_2(G^r) \oplus W_0 \oplus \langle u_1, \dots, u_n \rangle \tag{4.18}$$

where each u_i is amongst the generators identified in Proposition 2.33. Define $W_i := W_0 \oplus \langle u_1, \dots, u_i \rangle$. This gives a sequence of inclusion chain maps

$$\begin{array}{ccccccc}
C_2(G^r) \oplus W_{i-1} & \longrightarrow & C_1(G \cup G^s) & \longrightarrow & C_0(G \cup G^s) & & Q_{i-1} \\
\big\downarrow q_2^i & & \big\downarrow q_1^i & & \big\downarrow q_0^i & & \big\downarrow q^i \\
C_2(G^r) \oplus W_i & \longrightarrow & C_1(G \cup G^s) & \longrightarrow & C_0(G \cup G^s) & & Q_i
\end{array}$$

where each rows is a subcomplex of ${}^gC_\bullet(G, s)$. We denote the degree 1 homology groups by Q_i, with $Q_n := {}^gH_1(G, S)$, and the induced homology maps by $q^i : Q_{i-1} \to Q_i$. Together these chain maps decompose ${}^g\iota(r, s)_\#$ and hence the q^i decompose ${}^g\iota(r, s)_*$, as show in the following diagram.

$$\begin{tikzcd}
C_2(G^r) \arrow[r] \arrow[d] & C_1(G \cup G^r) \arrow[r] \arrow[d] & C_0(G \cup G^r) \arrow[d] & & {}^gH_1(G, r) \arrow[dl, "\cong"', "q^0"] \arrow[ddddd, two heads, "{}^g\iota(r,s)_*"] \\
C_2(G^r) \oplus W_0 \arrow[r] \arrow[d] & C_1(G \cup G^s) \arrow[r] \arrow[d] & C_0(G \cup G^s) \arrow[d] & Q_0 \arrow[d, two heads, "q^1"'] \arrow[ddddr, dashed, two heads, "q^n \circ \cdots \circ q^1"] & \\
C_2(G^r) \oplus W_1 \arrow[r] \arrow[d] & C_1(G \cup G^s) \arrow[r] \arrow[d] & C_0(G \cup G^s) \arrow[d] & Q_1 \arrow[d, two heads, "q^2"'] & \\
\vdots \arrow[d] & \vdots \arrow[d] & \vdots \arrow[d] & \vdots \arrow[d, two heads, "q^{n-1}"'] & \\
C_2(G^n) \oplus W_{n-1} \arrow[r] \arrow[d] & C_1(G \cup G^s) \arrow[r] \arrow[d] & C_0(G \cup G^s) \arrow[d] & Q_{n-1} \arrow[dr, two heads, "q^n"'] & \\
C_2(G^s) \arrow[r] & C_1(G \cup G^s) \arrow[r] & C_0(G \cup G^s) & & {}^gH_1(G, s)
\end{tikzcd}$$

Note that $\dim Q_i$ drops by at most 1 at each step, since $\dim W_{i+1} = \dim W_i + 1$. Let J denote the subset of indices where the dimension drops, i.e.

$$J := \{ i \in \mathbb{N} \mid \dim Q_i = \dim Q_{i-1} - 1 \}. \tag{4.19}$$

Then for each $i \in J$, the map on homology q^i has nullity 1 and a basis for $\ker q^i$ is $\{[b_i]\}$ where $b_i := {}^g\partial_2 u_i \in C_1(G \cup G^s)$.

Note that $[b_i] \in Q_0$ and $(q^{i-1} \circ \cdots \circ q^1)[b_i] = [b_i]$ in Q_{i-1}. Since each $[b_i]$ for $i \in J$ dies in a different Q_i, they must be linearly independent in Q_0. Moreover, the

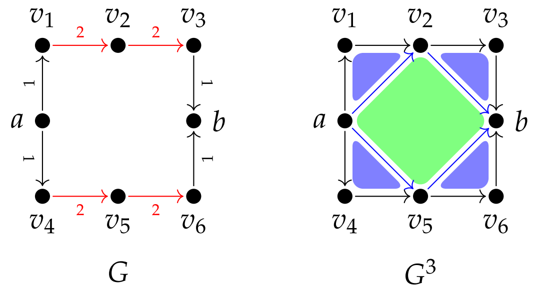

Fig. 7 An example weighted digraph with a single feature which dies at $t = 3$. To the right we show G^3, colouring the new edges in G^3 in blue and highlighting the new generators in $C_2(G^3)$ via blue and green polygons (Color figure online)

nullity of $q^n \circ \cdots \circ q^1$ is $\#J$ so $\{[b_i] \mid i \in J\}$ gives a basis for $\ker(q^n \circ \cdots \circ q^1)$. Hence $\left\{(q^0)^{-1}[b_i] \,\middle|\, i \in J\right\}$ forms a basis for $\ker {}^g\iota(r,s)_*$. It now remains to prove that each $(q^0)^{-1}[b_i]$ has a undirected circuit representative.

Claim 4.14 *For each $i \in J$, there exists an undirected circuit p_i in G such that ${}^g\iota(0,r)_*[\mathfrak{R}(p_i)] = (q^0)^{-1}[b_i]$.*

Proof of Claim. First we recall that $b_i = {}^g\partial_2 u_i$ where u_i is either a double edge, a directed triangle or a long square in G^s. Some of the boundary edges may be edges in G^r but at least one boundary edge is new in G^s. By the previous claim, a representative for $(q^0)^{-1}[b_i]$ is $c_1(b_i)$.

We can write $b_i = \mathfrak{R}(p)$ where p is the undirected circuit which traces the outline of the generator u_i. Now c_1 maps edges in G^r to themselves and edges not in G^r to a sum of two edges in G^r with the same boundary. Therefore, no matter which type of generator u_i is, we can write $c_1(b_i) = \mathfrak{R}(\widetilde{p})$ for some undirected circuit $\widetilde{p}$ through G^r. Using Lemma 4.1, for each edge $\tau \in E(\widetilde{p})$ in this circuit there is a directed path T_τ through G of length at most r such that $\tau = \mathfrak{R}(T_\tau) \pmod{{}^g B_1(G,r)}$. Concatenating the T_τ we obtain an undirected circuit p_i through G such that $\mathfrak{R}(p) = \mathfrak{R}(p_i) \pmod{{}^g B_1(G,r)}$. □

This claim concludes the proof. □

Remark 4.15 Note that the undirected circuits $p_{i,j}$ may not be simple. We conjecture that it should be possible to choose every $p_{i,j}$ to be simple but do not, as yet, have a proof.

Example 4.16 To illustrate how the generators are added in the proof of Lemma 4.12, consider Fig. 7. First note that there is a single feature with representative

$$\gamma := (av_1 + v_1v_2 + v_2v_3 + v_3b) - (av_4 + v_4v_5 + v_5v_6 + v_6b) \tag{4.20}$$

which dies at $t = 3$. Further note that new edges appear at integer values in the shortest-path filtration and $\mathcal{B}^g\mathcal{H}_1(G) = \{[0,3)\}$.

The new edges which appear at $t = 3$ are highlighted in blue. The new generators of $C_2(G^3)$ are the four blue directed triangles and the central green long square. The blue directed triangles form the elements of U_{new} and the central long square is the sole remaining generator u_1. So a basis for $\ker q^1$ is $\{[{}^g\partial_2 u_1]\}$.

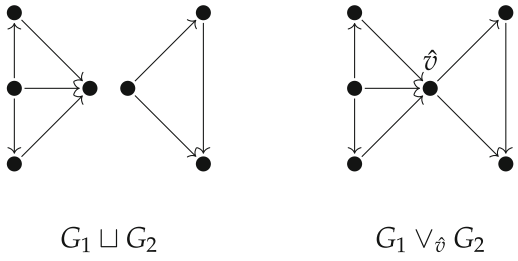

Fig. 8 Illustrations of the disjoint union and wedge decomposition considered in Sect. 4.4

To find a basis for $\ker {}^g\iota(2,3)_*$ we must compute $c_1({}^g\partial_2 u_1)$. Firstly ${}^g\partial_2 u_1 = av_2 + v_2 b - v_5 b - av_5$. Then the chain map c_1 maps each of the blue edges to a sum of edges in G^2. In fact $c_1({}^g\partial_2 u_1) = \gamma$ which is the representative of the sole simple undirected circuit in G.

The following Lemma follows by a standard linear algebra argument, since each ${}^g\iota(t_{i-1}, t_i)_*$ is surjective.

Lemma 4.17 *Given $B_i \subseteq {}^g H_1(G, 0)$ such that ${}^g\iota(0, t_i)_*(B_i)$ is a basis for* $\ker {}^g\iota(t_i, t_{i+1})_*$*, the union $\cup_{i=1}^{m-1} B_i$ is a basis for* $\ker {}^g\iota(0, t_m)_*$.

Certainly a basis of undirected circuits for ${}^g H_1(G, 0)$ exists (by Corollary 4.3). Therefore, we can always extend linearly independent undirected circuits to a basis of such circuits for ${}^g H_1(G, 0)$.

Lemma 4.18 *Given undirected circuits $p_1, \ldots, p_k$ such that $[\Re(p_1)], \ldots, [\Re(p_k)]$ are linearly independent in ${}^g H_1(G, 0)$, there exists undirected circuits $p_{k+1}, \ldots, p_N$ such that $\{[\Re(p_i)]\}_{i=1}^N$ is a basis for ${}^g H_1(G, 0)$.*

We now have all the ingredients we need to prove the main theorem.

Proof of Theorem 4.10 Using Lemmas 4.12 and 4.17, we obtain undirected circuits $p_1, \ldots, p_l$ such that $\{\Re(p_1), \ldots, \Re(p_l)\}$ forms a basis for $\ker {}^g\iota(0, t_m)_*$. Using Lemma 4.18, we extend this to a basis $\{\Re(p_1), \ldots, \Re(p_l), \Re(p_{l+1}), \ldots, \Re(p_N)\}$ for ${}^g H_1(G, 0)$. This is a persistence basis for ${}^g\mathcal{H}_1(G)$. □

Remark 4.19 While we are guaranteed a basis of undirected circuits, this choice of basis is by no means unique. As a simple example, consider again Example 4.11 and Fig. 6. Two possible persistence bases for ${}^g\mathcal{H}_1(G_2)$ are $\left\{\gamma_1^2, \gamma_3^2\right\}$ and $\left\{\gamma_2^2, \gamma_3^2\right\}$. The non-uniqueness of the basis arises in the proof of Lemma 4.12. Namely, there is a choice of v_e and w_e for each $e \in E_{new}$, and a choice of order on the remaining u_i.

4.4 Decomposition

In order to more easily compute GRPPH, it is desirable to understand how decompositions of the input weighted digraphs give rise to decompositions of the descriptor. The simplest such decomposition is a disjoint union; as one might expect, the descriptor decomposes as a direct sum (Fig. 8).

Theorem 4.20 *Suppose $G \in$ **WDgr** decomposes as a disjoint union, $G = G_1 \sqcup G_2$, then*

$$ {}^g\mathcal{H}_1(G) \cong {}^g\mathcal{H}_1(G_1) \oplus {}^g\mathcal{H}_1(G_2). \tag{4.21}$$

Proof Note that for each $t \geq 0$,

$$ G^t = G_1^t \sqcup G_2^t \quad \text{and} \quad G \cup G^t = (G_1 \cup G_1^t) \sqcup (G_2 \cup G_2^t). \tag{4.22}$$

For each degree $k \geq 0$, if $H = H_1 \sqcup H_2$ then $C_k(H) = C_k(H_1) \oplus C_k(H_2)$. Therefore, for each $k \geq 0$, ${}^gC_k(G, t)$ splits as direct sum ${}^gC_k(G_1, t) \oplus {}^gC_k(G_2, t)$. The boundary operator respects this split, mapping ${}^gC_k(G_i, t) \to {}^gC_{k-1}(G_i, t)$ and the maps ${}^g\iota(s, t)_\#$ also respect this split, mapping ${}^gC(G_i, s) \to {}^gC(G_i, t)$. Taking homology in degree 1 maintains this direct sum decomposition. □

Definition 4.21 (a) Given a weighted digraph $G = (V, E, w)$, a wedge vertex is a vertex $\hat{v} \in V$ such that there is a decomposition

$$ V = V_1 \cup V_2 \tag{4.23}$$

with $V_1 \cap V_2 = \{\hat{v}\}$ such that $E \subseteq (V_1 \times V_1) \cup (V_2 \times V_2)$.

(b) Given a wedge vertex, $\hat{v}$ as above the corresponding wedge decomposition of G is the pair (G_1, G_2) where G_1 and G_2 are the induced subgraphs on V_1 and V_2 respectively. We write $G = G_1 \vee_{\hat{v}} G_2$.

(c) Given a wedge decomposition as above, a pair of vertices $a, b \in V$ are called separated if they do not lie in a common V_i.

Remark 4.22 Given a wedge decomposition $G = G_1 \vee_{\hat{v}} G_2$ note that $G = G_1 \cup G_2$.

In the case of a wedge decomposition $G = G_1 \vee_{\hat{v}} G_2$, since each simple circuit is contained either entirely in G_1 or entirely in G_2, one expects that GRPPH also decomposes. The proof is more complicated because, in general, $G^t \neq G_1^t \vee_{\hat{v}} G_2^t$, since there may be paths between separated vertices, through $\hat{v}$. However, using a chain homotopy, we can show that these edges do not affect the homology.

Theorem 4.23 *For a weighted digraph $G = (V, E, w)$ and a wedge decomposition $G = G_1 \vee_{\hat{v}} G_2$,*

$$ {}^g\mathcal{H}_1(G) \cong {}^g\mathcal{H}_1(G_1) \oplus {}^g\mathcal{H}_1(G_2). \tag{4.24}$$

Proof There are natural inclusion digraph maps $j_i : G_i \to G$ which are also contractions. Less obviously, there are contraction digraph maps $f_i : G \to G_i$, where

$$ f_i(v) := \begin{cases} v & \text{if } v \in V_i, \\ \hat{v} & \text{otherwise.} \end{cases} \tag{4.25}$$

Since these are all morphisms in **ContWDgr**, we obtain induced morphisms ${}^g j_{i\#}$ and ${}^g f_{i\#}$. We combine these morphisms to get two morphisms as follows

$$J : {}^gC(G_1) \oplus {}^gC(G_2) \to {}^gC(G), \quad J(\gamma_1, \gamma_2) := {}^g j_{1\#}\gamma_1 + {}^g j_{2\#}\gamma_2; \tag{4.26}$$

$$F : {}^gC(G) \to {}^gC(G_1) \oplus {}^gC(G_2), \quad F(\gamma) := ({}^g f_{1\#}\gamma, {}^g f_{2\#}\gamma). \tag{4.27}$$

Composing with homology in degree 1, denote $J_* := [\mathbf{R}, H_1] \circ J$ and $F_* := [\mathbf{R}, H_1] \circ F$. In the rest of the proof, we show that J_* and F_* are mutually inverse.

Claim 4.24 *In degree* 1, $F \circ J = \mathrm{id}$ *is the identity map on* ${}^gC(G_1) \oplus {}^gC(G_2)$.

Proof of Claim. First note that, $f_i \circ j_i$ is the identity digraph map $\mathrm{id}_i : G_i \to G_i$. However, $f_{3-i} \circ j_i$ is the constant digraph map $c_i : G_i \to G_{3-i}$ which maps all of G_i to the vertex $\hat{v}$. Hence, in matrix form, we can write $F \circ J$ as

$$F \circ J = \begin{pmatrix} {}^g\mathrm{id}_{1\#} & {}^g c_{2\#} \\ {}^g c_{1\#} & {}^g\mathrm{id}_{2\#} \end{pmatrix}. \tag{4.28}$$

Since the constant maps c_i send all vertices to a single vertex, ${}^g c_{i\#}$ maps every edge to 0. Hence, in degree 1, ${}^g c_{i\#}$ is the zero map. Whereas, in degree 1, ${}^g\mathrm{id}_{i\#}$ is the identity map on ${}^gC_1(G_i, t)$ at each t. Therefore, on degree 1, $F \circ J$ is the identity map on ${}^gC(G_1) \oplus {}^gC(G_2)$. □

Composing with homology in degree 1, we see $F_* \circ J_* = \mathrm{id}$ is the identity map on ${}^g\mathcal{H}_1(G_1) \oplus {}^g\mathcal{H}_1(G_2)$.

Claim 4.25 *In degree* 1, $J_* \circ F_* = \mathrm{id}$ *is the identity map on* ${}^g\mathcal{H}_1(G)$.

Proof of Claim. First, we compute $J \circ F$ in degree 1. Recall that ${}^gC_1(G, t)$ is freely generated by the edges in $G \cup G^t$. Given an edge $e = (a, b) \in E(G \cup G^t)$, if the vertices a, b lie in a common V_i then $(J \circ F)(e) = e$. However, if a, b are separated then $(J \circ F)(e) = a\hat{v} + \hat{v}b$. So we see, at the level of chains, $J \circ F$ does not compose to the identity.

If $e = (a, b) \in E(G \cup G^t)$ but the endpoints are separated then we must have $e \in E(G^t)$. Hence, there is a path $p : a \rightsquigarrow b$ in G of length at most t. Moreover, this path must traverse the vertex $\hat{v}$. Hence, p decomposes into two paths $a \rightsquigarrow \hat{v}$ and $\hat{v} \rightsquigarrow b$, each of length at most t. Therefore, the directed triangle $a\hat{v}b$ is present in G^t and is a generator of ${}^gC_2(G, t)$. Note that the boundary of $a\hat{v}b$ is

$$\partial_2(a\hat{v}b) = a\hat{v} + \hat{v}b - (ab) = (J \circ F)(e) - e. \tag{4.29}$$

This discussion show that we can define a map $P : {}^gC_1(G, t) \to {}^gC_2(G, t)$ by

$$P(ab) := \begin{cases} a\hat{v}b & \text{if } a, b \text{ are separated,} \\ 0 & \text{otherwise.} \end{cases} \tag{4.30}$$

Then, $J \circ F - \mathrm{id} = \partial_2 P$ as maps on ${}^gC_1(G)$. Composing with homology, we see $J_* \circ F_* = \mathrm{id}$ is the identity on ${}^g\mathcal{H}_1(G)$. □

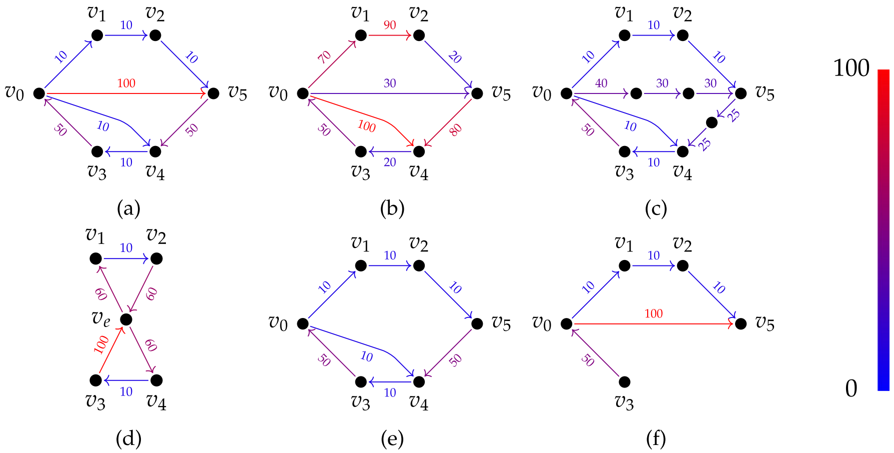

Fig. 9 Illustration of a number of operations for altering a weighted digraph. For formal definitions, see the corresponding subsections of Sect. 5. **a** The initial weighted digraph G. **b** Weight perturbation $\mathbb{O}^{p}_{w'}G$. **c** Edge subdivision $\mathbb{O}^{s}_{S}G$ where the subdivision is $S : \{(v_0, v_5), (v_5, v_4)\} \to \Delta^2$ where $S((v_0, v_5)) = (4/10, 3/10, 3/10)$ and $S((v_5, v_4)) = (1/2, 1/2, 0)$. **d** Edge collapse $\mathbb{O}^{c}_{e}G$ where $e = (v_0, v_5)$. **e** Edge deletion $\mathbb{O}^{d}_{e}G$ where $e = (v_0, v_5)$. **f** Vertex deletion $\mathbb{O}^{d}_{v}G$ where $v = v_4$

Since J_* and F_* are mutually inverse, they induce isomorphisms of persistent vector spaces. □

5 Stability Analysis of GrPPH

It is important that GrPPH is stable with respect to a reasonable noise model. Typically this is shown by proving that ${}^{g}\mathcal{H}_1$ is Lipschitz with respect to reasonable metrics on **WDgr** and the bottleneck distance on **PersVec**. A common choice of metric on graphs is the graph edit distance. However, assigning costs to operations such as edge deletion or edge subdivision is somewhat arbitrary.

Therefore, in this section, we consider operations $\mathbb{O}^{T}_{\theta} : \mathrm{Obj}(\mathbf{WDgr}) \to \mathrm{Obj}(\mathbf{WDgr})$ for editing weighted digraphs, where T is the type of operation θ is the parameter of the operation. For each type T, we derive bounds of the form

$$d_B(\mathcal{B}{}^{g}\mathcal{H}_1(G), \mathcal{B}{}^{g}\mathcal{H}_1(\mathbb{O}^{T}_{\theta}G)) \leq f(G, \theta) \tag{5.1}$$

and then say GrPPH is stable with respect to operations of type T.

Often the operations only alter the graph at a subset of vertices or edges. We say that GrPPH is locally stable to the operation if we obtain a bound as in (5.1) and f depends only on the neighbourhood graph around the altered vertices/edges and θ. If we can show that no such local f exists (for general G and θ) then we say GrPPH is locally unstable. If, in general, f depends on all of G then we say GrPPH is non-locally stable to the operation. Occasionally, some operations do not change the descriptor and we can find an isomorphism ${}^{g}\mathcal{H}_1(G) \cong {}^{g}\mathcal{H}_1(\mathbb{O}^{T}_{\theta}G)$.

Table 1 Stability and instability theorems for ${}^{g}\mathcal{H}_1$, under various digraph operations. ♦ Denotes a theorem which only applies to a subset of such operations

Operation	Locally stable	Non-locally stable	Locally unstable	Isomorphism
Weight perturbation	Theorem 5.11			
Edge subdivision	Theorem 5.16			
Edge collapse	Theorem 5.28♦		Theorem 5.35	
Edge deletion	Corollary 5.41♦	Theorem 5.38	Theorem 5.42	Theorem 5.45♦
Vertex deletion		Theorem 5.50	Corollary 5.49	Corollary 5.48♦

Figure 9 illustrates all of the operations we consider, the precise definitions of which are provided in the relevant subsection. Table 1 summaries our findings.

5.1 Main Stability Theorem

In order to prove stability, we will have to build interleaving chain maps. We construct these via maps of the underlying vertex sets.

Definition 5.1 For $\delta \geq 0$, a δ-shifting vertex map, between two weighted digraphs G and H, is a vertex map $f : V(G) \to V(H)$ such that f induces digraph maps $G^t \to H^{t+\delta}$ and $G \cup G^t \to H \cup H^{t+\delta}$ for all $t \geq 0$.

Remark 5.2 By Lemma 3.15, a 0-shifting vertex map is precisely a contraction digraph map.

Lemma 5.3 *Any δ-shifting vertex map $f : V(G) \to V(H)$ induces a morphism*

$$\mathcal{S}(f, \delta)_\# : {}^{g}C(G) \to {}^{g}C(H)[\delta]. \tag{5.2}$$

Given another ϵ-shifting vertex map $g : V(H) \to V(K)$,

$$\mathcal{S}(g \circ f, \epsilon + \delta)_\# = \mathcal{S}(g, \epsilon)_\# \circ \mathcal{S}(f, \delta)_\#. \tag{5.3}$$

Moreover, if f is 0-shifting then $\mathcal{S}(f, 0)_\# = {}^{g}f_\#$.

Proof Since f is δ-shifting, it induces the necessary digraph maps so that Lemma 3.9 yields a chain map $J(f, t \leq t+\delta) : {}^{g}C_\bullet(G, t) \to {}^{g}C_\bullet(H, t+\delta)$ for any $t \geq 0$. Given $s \leq t$, the chain map ${}^{g}\iota(s, t)_\#$ is induced, through Lemma 3.9, by the identity vertex map. Hence, by Eq. (3.5), we see

$$\begin{aligned} J(f, t \leq t+\delta) \circ {}^{g}\iota(s, t)_\# &= J(f, s \leq t+\delta) \\ &= {}^{g}\iota(s+\delta, t+\delta)_\# \circ J(f, s \leq s+\delta). \end{aligned} \tag{5.4}$$

Therefore we can combine the chain maps $J(f, t \leq t+\delta)$ into the required morphism $\mathcal{S}(f, \delta)_\# : {}^{g}C(G) \to {}^{g}C(H)[\delta]$.

Equation (5.3) follows immediately from Eq. (3.5). Finally, when $\delta = 0$, we note that in fact $f \in \mathbf{WDgr}_{F_d}$ and the construction of $\mathcal{S}(f, 0)_\#$ is identical to the construction of ${}^g f_\#$, as done in Theorem 3.13. □

Recall that $\mathcal{T}({}^g C(G), \epsilon)$ at each $t \geq 0$ is the chain map ${}^g\iota(t, t+\epsilon)_\# : {}^g C(G, t) \to {}^g C(G, t+\epsilon)$. Shifting this by δ, $\mathcal{T}({}^g C(G), \epsilon)[\delta]$ is given at $t \geq 0$ by the chain map ${}^g\iota(t+\delta, t+\delta+\epsilon)_\#$.

Lemma 5.4 *A δ-shifting vertex map is a δ'-shifting vertex map for any $\delta' \geq \delta$ and*

$$\mathcal{S}(f, \delta')_\# = \mathcal{T}({}^g C(H), \delta' - \delta)[\delta] \circ \mathcal{S}(f, \delta)_\#. \tag{5.5}$$

Proof At each $t \geq 0$, each of the morphisms in Eq. (5.5) are induced by Lemma 3.9. Namely, $\mathcal{S}(f, \delta')_\#$ is given by $J(f, t \leq t+\delta')$, $\mathcal{T}({}^g C(H), \delta' - \delta)[\delta]$ is given by $J(\mathrm{id}_{V(H)}, t+\delta \leq t+\delta')$ and $\mathcal{S}(f, \delta)_\#$ is given by $J(f, t \leq t+\delta)$. Since $\mathrm{id}_{V(H)} \circ f = f$ as vertex maps and $t \leq t+\delta \leq t+\delta'$, Eq. (5.5) follows immediately from Eq. (3.5). □

Definition 5.5 Fix weighted digraphs G, H and two vertex maps $f : V(G) \to V(H)$ and $g : V(H) \to V(G)$.

(a) Denote

$$V_{fix}(g, f) := \{v \in V(G) \mid g(f(v)) = v\}, \quad V_{diff}(g, f) := V(G) \setminus V_{fix}(g, f), \tag{5.6}$$

$$E_{fix}(g, f) := \{v \in E(G) \mid g(f(e)) = e\}, \quad E_{diff}(g, f) := E(G) \setminus E_{fix}(g, f). \tag{5.7}$$

Also, denote the following unweighted digraph $G_{diff}(g, f) := (V(G), E_{diff}(g, f))$.

(b) If $\mathrm{id} : G_{diff}(g, f) \to G^{2\delta}$ and $g \circ f : G_{diff}(g, f) \to G^{2\delta}$ are both digraph maps and furthermore are path homotopic relative $V_{fix}(g, f)$ then we say the ordered pair (g, f) has grounded codistortion $\leq \delta$.

(c) If f and g are both δ-shifting vertex maps and the pairs (g, f) and (f, g) both have grounded codistortion $\leq \delta$ then we say they form a δ-grounded interleaving.

Remark 5.6 (a) Fix $\delta' \geq \delta$. Since $t \mapsto G^t$ is an increasing filtration, if a pair of vertex maps (g, f) has grounded codistortion $\leq \delta$, then they certainly have grounded codistortion $\leq \delta'$.

(b) If $g \circ f = \mathrm{id}$ as a vertex map then (g, f) has grounded codistortion ≤ 0.

(c) A 0-grounded interleaving is precisely a mutually inverse pair of isomorphisms of the underling digraphs which also induce isometries of the shortest-path quasimetrics.

Remark 5.7 For the interested reader, this definition is strongly inspired by the Kalton-Ostrovskii characterisation of network distance, which first appeared for directed networks in [12]. In particular, this characterisation was subsequently used to prove the stability of persistent path homology in [13].

Theorem 5.8 (Main stability theorem) *Given two weighted digraphs G, H, if there is a δ-grounded interleaving between them then $d_B(\mathcal{B}^g\mathcal{H}_1(G), \mathcal{B}^g\mathcal{H}_1(H)) \leq \delta$.*

Proof Denote the δ-grounded interleaving by $f : V(G) \to V(H)$ and $g : V(H) \to V(G)$. Since the vertex maps are δ-shifting they induce chain maps $\mathcal{S}(f, \delta)_\#$ and $\mathcal{S}(g, \delta)_\#$. In the following we will show they induce a δ-interleaving between ${}^g\mathcal{H}_1(G)$ and ${}^g\mathcal{H}_1(H)$. The result then follows by the isometry theorem.

Since (g, f) has grounded codistortion $\leq \delta$ there is a homotopy $F : G_{diff}(g, f) \to G^{2\delta}$ between id and $g \circ f$, relative $V_{fix}(g, f)$. By Theorem 2.35, this induces a chain homotopy $L : C_\bullet(G_{diff}(g, f)) \to C_{\bullet+1}(G^{2\delta})$ between $\mathrm{id}_\#$ and $(g \circ f)_\#$. Observe that

$$C_1(G) = R\langle E_{diff}(g, f)\rangle \oplus R\langle E_{fix}(g, f)\rangle, \tag{5.8}$$

$$C_1(G_{diff}(g, f)) = R\langle E_{diff}(g, f)\rangle, \tag{5.9}$$

$$C_0(G_{diff}(g, f)) = C_0(G) = R\langle V_{diff}(g, f)\rangle \oplus R\langle V_{fix}(g, f)\rangle. \tag{5.10}$$

Hence, from the chain homotopy, we obtain the following diagram where the top row is ${}^gC_\bullet(G, 0)$ and the bottom row is ${}^gC_\bullet(G, 2\delta)$. To ease notation, we drop the ring R and we use blue to denote the domain of the L_i maps.

$$\begin{tikzcd}
0 \arrow[rr] & & \langle E_{diff}(g,f)\rangle \oplus \langle E_{fix}(g,f)\rangle \arrow[rr, "\partial"] \arrow[dd, shift right, "id_\#"'] \arrow[dd, shift left, "(g\circ f)_\#"] \arrow[ddll, bend right, "L_1"'] & & \langle V_{diff}(g,f)\rangle \oplus \langle V_{fix}(g,f)\rangle \arrow[dll, bend left, "L_0"] \\
& C_1(G^{2\delta}) \arrow[dr, dashed, "\iota_\#"] & & & \\
C_2(G^{2\delta}) \arrow[ur, dashed, "\partial"] \arrow[rr, "\iota_\#\circ\partial"'] & & C_1(G \cup G^{2\delta}) \arrow[rr, "\partial"'] & & C_0(G \cup G^{2\delta})
\end{tikzcd}$$

Now take a degree-1 cycle in the top row, i.e. some chain $c \in C_1(G)$ such that $\partial c = 0$. Thanks to Eq. (5.8), we can decompose $c = c_1 + c_2$ where $c_1 \in R\langle E_{diff}(g, f)\rangle$ and $c_2 \in R\langle E_{fix}(g, f)\rangle$.

Since c_1 is supported only on edges of $G_{diff}(g, f)$, the chain homotopy equation for L yields

$$(g \circ f)_\#(c_1) - \mathrm{id}_\#(c_1) = \partial L_1(c_1) + L_0\partial(c_1). \tag{5.11}$$

Since c_2 is supported only on edges in $E_{fix}(g, f)$, one can easily verify that $\mathrm{id}_\#(c_2) = (g \circ f)_\#(c_2)$. Moreover ∂c_2 is supported only on vertices in $V_{fix}(g, f)$ and the homotopy is relative $V_{fix}(g, f)$ so Lemma 2.37 ensures that $L_0(\partial c_2) = 0$. Piecing these together we obtain

$$\begin{aligned}
(g \circ f)_\#(c) - \mathrm{id}_\#(c) &= (g \circ f)_\#(c_1) - \mathrm{id}_\#(c_1) \\
&= \partial L_1(c_1) + L_0\partial(c_1) \\
&= \partial L_1(c_1) + L_0\partial(c_1 + c_2) \\
&= \partial L_1(c_1) + L_0(0) \\
&= \partial L_1(c_1).
\end{aligned}$$

Note that $(g \circ f)_\#$ is the degree-1 component of the chain map $\mathcal{S}(g, \delta)_\# \circ \mathcal{S}(f, \delta)_*$, at t=0. Similarly $\mathrm{id}_\#$ is the degree-1 component of the chain map ${}^g\iota(0, 2\delta)_\#$. Therefore on the level of homology, $\mathcal{S}(g, \delta)_* \circ \mathcal{S}(f, \delta)_* = {}^g\iota(0, 2\delta)_*$. Moreover, Proposition 4.2 implies that ${}^g\iota(0, t)_*$ is a surjection for all $t \geq 0$ from which it follows that $\mathcal{S}(g, \delta)_* \circ \mathcal{S}(f, \delta)_* = {}^g\iota(t, t + 2\delta)_*$ for all $t \geq 0$. The same argument holds with the opposite composition and hence we obtain a δ-interleaving as required. □

Remark 5.9 Given δ-shifting vertex maps f and g, we obtain chain maps $\mathcal{S}(f, \delta)_\#$ and $\mathcal{S}(g, \delta)_\#$. Rather than showing the pairs have grounded codistortion $\leq \delta$, it is possible to show they form an interleaving on homology by instead considering their action on a basis of undirected circuits for ${}^g Z_1(G, 0)$. One must show that for any undirected circuit c, there is some $b \in C_2(G^{2\delta})$ such that

$$(\mathcal{S}(g, \delta)_\# \circ \mathcal{S}(f, \delta)_\#)(\mathfrak{R}(c)) - {}^g\iota(0, 2\delta)_\#(\mathfrak{R}(c)) = \partial b. \tag{5.12}$$

This approach was taken in a previous version of this manuscript, available on the arXiv. Note that if (g, f) has grounded codistortion $\leq \delta$ then one can take $b = L_1(\mathfrak{R}(c))$, where L_1 is the degree-1 component of the induced chain homotopy between $g \circ f$ and id.

5.2 Weight Perturbation

The classical stability theorem of persistent homology (first shown in [14]) is that for two continuous tame function $f, g : X \to \mathbb{R}$ of a triangulable topological space X, denoting the persistence barcode of their sub-level set filtration by $\mathcal{B}_f$ and $\mathcal{B}_g$ respectively,

$$d_B(\mathcal{B}_f, \mathcal{B}_g) \leq \|f - g\|_\infty. \tag{5.13}$$

In our setting, the closet analogy to changing the function is changing the weighting, as well as the corresponding effect that has on the shortest-path quasimetric. We find that GrPPH is stable to perturbations of the edge weights. Moreover, the stability is local since it depends only of the weights of the perturbed edges.

Definition 5.10 Given a weighted digraph $G = (V, E, w)$ and a new weight function $w' : E(G) \to \mathbb{R}_{>0}$, we define $\mathbb{O}^p_{w'} G := (V, E, w')$.

Theorem 5.11 *Given a weighted digraph $G = (V, E, w) \in$ **WDgr** and a new weighting function $w' : E(G) \to \mathbb{R}_{>0}$, let d and d' denote the shortest-path quasimetric on G and $\mathbb{O}^p_{w'} G$ respectively. Then*

$$d_B\big(\mathcal{B}^g\mathcal{H}_1(G), \mathcal{B}^g\mathcal{H}_1(\mathbb{O}^p_{w'} G)\big) \leq \max_{i,j \in V} \left| d(i, j) - d'(i, j) \right| \leq \sum_{e \in E} \left| w(e) - w'(e) \right|. \tag{5.14}$$

Proof For brevity, denote $G' := \mathbb{O}^p_{w'} G$ and $\delta := \max_{i,j \in V} \left| d(i,j) - d'(i,j) \right|$. First note that for any (i,j) and any path $p \in \mathcal{P}(i \to j)$

$$\left| \sum_{e \in p} w(e) - \sum_{e \in p} w'(e) \right| \leq \sum_{e \in p} \left| w(e) - w'(e) \right| \leq \sum_{e \in E} \left| w(e) - w'(e) \right| =: W_1. \tag{5.15}$$

So the cost of p differs by at most W_1. Minimising over $\mathcal{P}(i \to j)$, we see $\left| d(i,j) - d'(i,j) \right| \leq W_1$.

Since $V(G) = V(G')$, there are identity vertex maps $i_1 : V(G) \to V(G')$ and $i_2 : V(G') \to V(G)$. Now i_1 defines a digraph map $G \to G'$ since $G = G'$ as digraphs. Moreover, given $(i,j) \in E(G^t)$, then $d(i,j) \leq t$ so $d'(i,j) \leq t + \delta$ and hence $(i,j) \in E((G')^{t+\delta})$. This shows i_1 defines a digraph map $G^t \to (G')^{t+\delta}$ for all $t \geq 0$. Therefore i_1 (and likewise i_2) is a δ-shifting vertex map. Moreover, since composing i_1 and i_2 in either ordered yields the identity vertex map, these morphisms certainly constitute a δ-grounded interleaving. The first inequality then follows by the main stability theorem (Theorem 5.8). □

Remark 5.12 Continuing the analogy to the classical stability theorem, note that the sharper bound obtained by Theorem 5.11 is $\|d - d'\|_\infty$ while the weaker bound is $\|w - w'\|_1$.

5.3 Edge Subdivision

Weighted digraphs arising in applications are subject not only to numerical noise (i.e. weight perturbation) but also *structural noise*. For the remainder of this section, we investigate the effects of various structural perturbations.

First, we consider edge subdivision, in which one or more parent edge is split into multiple child edges with the weight distributed amongst them. Since we are interpreting edge weights as corresponding to a length, it is natural to require that the sum of the weights of the child edges equals the weight of the parent edge. In order to formalise how the weight of an edge is subdivided amongst its children, we use maps into the standard d-simplex, where d is the number of child edges.

Definition 5.13 Given a weighted digraph $G = (V, E, w)$, a subdivision S of G is a choice of edges $F \subseteq E$, along with a map $S : F \to \sqcup_{d \in \mathbb{N}} \operatorname{int}(\Delta^d)$ from edges in F to the formal disjoint union of the interiors of each standard d-simplices.

Intuitively, a subdivision gives us a recipe for subdividing the edges of F where $S(e)_i$ describes the fraction of $w(e)$ which the i^{th} child edge of e should receive..

Notation 5.14 Given a subdivision $S : F \to \sqcup_{d \in \mathbb{N}} \operatorname{int}(\Delta^d)$,

(a) Let $d(e)$ denote the simplex dimension such that $S(e) \in \operatorname{int}(\Delta^{d(e)})$. Note $S(e)$ is a $(d(e)+1)$-tuple whose components we denote $S(e) = (S(e)_0, \dots S(e)_{d(e)})$.
(b) Let $CS(e)$ denote the $(d(e)+1)$-tuple of cumulative sums, i.e. $CS(e)_i := \sum_{j=0}^{i} S(e)_j$.

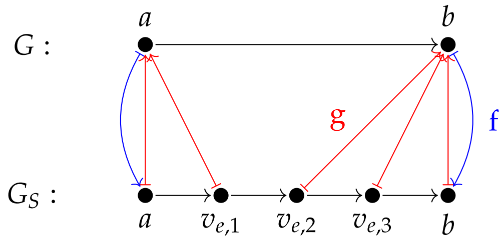

Fig. 10 Visualising the vertex maps f and g under the subdivision $S(e) = (1/4, 1/4, 1/4, 1/4)$ where $e = (a, b)$

Definition 5.15 Given a subdivision $S : F \to \sqcup_{d\in\mathbb{N}} \operatorname{int}(\Delta^d)$, define

$$V_S := V_{old} \sqcup V_{new} := V \sqcup \bigsqcup_{e\in F} \{v_{e,1}, \ldots, v_{e,d(e)}\}$$

$$E_S := E_{old} \sqcup E_{new} := (E \setminus F) \sqcup \bigsqcup_{e\in F} \{\tau_{e,0}, \ldots, \tau_{e,d(e)}\}$$

$$w_S(\tau) := \begin{cases} w(\tau) & \text{if } \tau \in E_{old} \\ S(e)_i \cdot w(e) & \text{if } \tau = \tau_{e,i} \end{cases}$$

where $\tau_{e,i} = (v_{e,i}, v_{e,i+1})$ and we denote $v_{e,0} := \mathrm{st}(e)$ and $v_{e,d(e)+1} := \mathrm{fn}(e)$. We then define $\mathbb{O}^s_S G := (V_S, E_S, w_S)$.

We show that the descriptor is stable to arbitrary subdivisions of arbitrary subsets of edges. Moreover, this stability is local since the bound depends only on the weight of subdivided edges.

Theorem 5.16 *Given a weighted digraph $G = (V, E, w) \in$ **WDgr** and any subdivision $S : F \to \sqcup_{d\in\mathbb{N}} \operatorname{int}(\Delta^d)$,*

$$d_B\left(\mathcal{B}^g\mathcal{H}_1(G), \mathcal{B}^g\mathcal{H}_1(\mathbb{O}^s_S G)\right) \le \max_{e\in F} w(e). \tag{5.16}$$

Proof To begin, denote $\delta := \max_{e\in F} w(G)(e)$. We will construct a δ-grounded interleaving and then employ the main stability theorem (Theorem 5.8). First, we setup some notation. Denote $G_S := \mathbb{O}^s_S G = (V_{old} \sqcup V_{new}, E_{old} \sqcup E_{new}, w_S)$. We let d and d_S denote the shortest-path quasimetric on G and G_S respectively. Define the following vertex maps

$$f : V(G) \to V(G_S) \quad v \mapsto v; \tag{5.17}$$

$$g : V(G_S) \to V(G) \quad v \mapsto \begin{cases} v & \text{if } v \in V_{old}, \\ \mathrm{st}(e) & \text{if } v_{e,i} \in V_{new} \text{ and } CS(e)_i < 1/2, \\ \mathrm{fn}(e) & \text{if } v_{e,i} \in V_{new} \text{ and } CS(e)_i \ge 1/2, \end{cases} \tag{5.18}$$

which are visualized in Fig. 10.

Claim 5.17 *For vertices $i, j \in V_{old}$, there is a path $i \rightsquigarrow j$ in G of length t if and only if there is one in G_S.*

Proof of Claim. This is clear to see, since the weight of an edge is shared amongst its child edges in the subdivision. □

Claim 5.18 *For any $t \geq 0$, f defines a digraph map $G^t \to G_S^{t+\delta}$ and $G \cup G^t \to G_S \cup G_S^{t+\delta}$.*

Proof of Claim. Since f is just the inclusion vertex map, Claim 5.17 shows that f defines a digraph map $G^t \to G_S^t$ and so certainly $G^t \to G_S^{t+\delta}$. For the second map, pick an edge $e \in E$ and note $f(e) = e$. If $e \notin F$ then it is undivided and $e \in E(G_S)$. Otherwise $e \in F$ and $e \notin E(G_S)$, however we note $d_S(\mathrm{st}(e), \mathrm{fn}(e)) \leq w(e) \leq \delta$. Therefore, for any $t \geq 0$, $e \in G_S^{t+\delta}$ and hence f defines a digraph map $G \cup G^t \to G_S \cup G_S^{t+\delta}$. □

Claim 5.19 *For any $t \geq 0$, g defines a digraph map $G_S^t \to G^{t+\delta}$ and $G_S \cup G_S^t \to G \cup G^{t+\delta}$.*

Proof of Claim. Given an edge $\tau = (a, b) \in E(G_S^t)$ there is a path $p : a \rightsquigarrow b$ in G_S of length at most t. We may assume that $g(a) \neq g(b)$, else there is nothing to check for this edge. If $a = v_{e,i}$ is a new vertex from subdividing an edge $e \in F$ then $g(a)$ is either $\mathrm{st}(e)$ or $\mathrm{fn}(e)$. Either by adding or removing relevant child edges of e to/from the start of p, we obtain a new path $g(a) \rightsquigarrow b$ in G_S. By construction, this will add at most $w(e)/2 \leq \delta/2$ to the length of p. Likewise we can alter the end of p to obtain a path $g(a) \rightsquigarrow g(b)$ in G_S of length at most $t + \delta$. By Claim 5.17, we see $g(\tau) \in E(G^{t+\delta})$.

Finally, given an edge $\tau \in G_S$ there are two cases. If $\tau \in E_{old}$ then the edge is preserved under g. Else $\tau = \tau_{e,i} \in E_{new}$ in which case either τ is collapsed to one of the endpoints of e, or it is mapped to e. Hence g is digraph map $G_S \to G$ and the final requirement follows. □

These claims show that f and g are δ-shifting vertex maps. Note that, as vertex maps, $g \circ f = \mathrm{id}_{V(G)}$. Therefore (g, f) certainly has grounded codistortion $\leq \delta$. Composing vertex maps in the opposite order, we do not obtain the identity. Note that $E_{diff}(f, g) = E_{new}$ and $V_{diff}(f, g) = V_{new}$.

For each divided edge $e = (a, b) \in F$, we construct a homotopy $F_e : \mathrm{Ch}(e) \boxdot I \to {G_S}^{2\delta}$ where $\mathrm{Ch}(e)$ is the induced subgraph of G_S on the child edges of e, namely $\{\tau_{e,0}, \tau_{e,1}, \ldots, \tau_{e,d(e)}\}$. It is a two-step homotopy; in the first step all $v_{e,i}$ with $CS(e)_i < 1/2$ are mapped to a and the second step the remaining $v_{e,i}$ are mapped to b. The homotopy for each edge is shown schematically in Fig. 11.

This is indeed a digraph map $F_e : \mathrm{Ch}(e) \boxdot I \to G_S^{2\delta}$ because the original edge e had weight $\leq \delta$ and all edges point along the path $(a, v_{e,1}, \ldots, b)$. Note that the top of the diagram is the map $f \circ g$ and the bottom is the map id, when restricted to $\mathbf{Ch}(e)$. Moreover, for each edge $e \in F$, F_e fixes the vertices $\mathrm{st}(e), \mathrm{fn}(e)$. Also note that $G_{diff}(f, g) = \cup_{e \in F} \mathrm{Ch}(e)$. Hence the F_e can be combined into a homotopy $F : G_{diff}(f, g) \boxdot I \to G_S^{2\delta}$ by the formula

$$F(v, j) := \begin{cases} F_e(v, j) & \text{if } v = v_{e,i} \in V_{new}, \\ v & \text{otherwise.} \end{cases} \quad Hub \tag{5.19}$$

$$
\begin{tikzcd}
 & 2 \bullet & & a \arrow[r] & a \arrow[r] & b \arrow[r] & b \arrow[r] & b & \\
I & 1 \bullet \arrow[u] \arrow[d] & & a \arrow[r] \arrow[u] \arrow[d] & a \arrow[r] \arrow[u] \arrow[d] & v_{e,2} \arrow[r] \arrow[u] \arrow[d] & v_{e,3} \arrow[r] \arrow[u] \arrow[d] & b \arrow[u] \arrow[d] & \mathrm{Ch}(e) \boxdot I \\
 & 0 \bullet & & a \arrow[r] & v_{e,1} \arrow[r] & v_{e,2} \arrow[r] & v_{e,3} \arrow[r] & b &
\end{tikzcd}
$$

$$
\begin{tikzcd}
\underset{a}{\bullet} \arrow[r] & \underset{v_{e,1}}{\bullet} \arrow[r] & \underset{v_{e,2}}{\bullet} \arrow[r] & \underset{v_{e,3}}{\bullet} \arrow[r] & \underset{b}{\bullet} & \mathrm{Ch}(e)
\end{tikzcd}
$$

Fig. 11 Schematic for the homotopy between $f \circ g$ (on the top row) and id (on the bottom row). The label on each node in $\mathrm{Ch}(e) \boxdot I$ is its image under the homotopy F_e

This is a homotopy between $f \circ g$ and id on all of $G_{diff}(f, g)$ and is relative $V_{old} = V_{fix}(f, g)$. Hence we see that (f, g) has grounded codistortion $\leq \delta$, concluding the proof. □

Remark 5.20 Since subdividing an edge does not effect circuit rank of $\mathcal{U}(G)$, the number of features does not change upon subdivision (by Corollary 4.3).

Definition 5.21 Fix a weighted digraph $G = (V, E, w) \in \mathbf{WDgr}$.

(a) The medial subdivision, $S_{med}(G) : E(G) \to \Delta^2$, is given by $S(e) = (1/2, 1/2)$ for every $e \in E(G)$.
(b) The n^{th} iterated medial subdivision of G, $\mathrm{IMS}_n(G)$, is defined iteratively as follows. Firstly, $\mathrm{IMS}_0(G) := G$ then for each n, we define $\mathrm{IMS}_n(G) := \mathbb{O}^s_S \mathrm{IMS}_{n-1}(G)$ where $S = S_{med}(\mathrm{IMS}_{n-1}(G))$.

Corollary 5.22 *Given a weighted digraph* $G \in \mathbf{WDgr}$, *the sequence of barcodes* $(\mathcal{B}^g \mathcal{H}_1(\mathrm{IMS}_n(G)))_{n \in \mathbb{N}}$ *converges under the bottleneck distance.*

Proof We first note that

$$\max_{e \in E(\mathrm{IMS}_n(G))} w(e) = \frac{1}{2^n} \max_{e \in E(G)} w(e). \tag{5.20}$$

Hence, Theorem 5.16 implies that the sequence of barcodes is Cauchy. The space of persistence diagrams with the bottleneck distance is complete [11] and hence the sequence of barcodes converges. □

For an example, we refer the reader forward to Proposition 6.1. While we have bottleneck stability, we do *not* have p-Wasserstein stability for any $p \in [1, \infty)$.

Proposition 5.23 *Fix* $1 \leq p < \infty$. *There exists no function* $f : \mathbf{WDgr} \to \mathbb{R}$ *such that for any weighted digraph* $G = (V, E, w) \in \mathbf{WDgr}$ *and any subdivision* $S : F \to \mathrm{int}(\Delta^d)$ *we have (Fig. 12).*

$$d_{W_p}\big(\mathcal{B}^g \mathcal{H}_1(\mathbb{O}^s_S G), \mathcal{B}^g \mathcal{H}_1(G)\big) \leq f\big(\mathcal{NG}(F; G)\big). \tag{5.21}$$

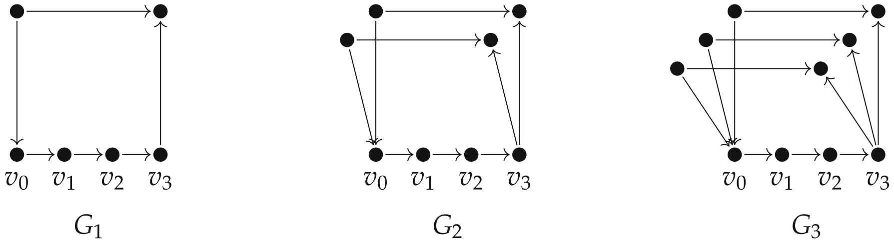

Fig. 12 A sequence of weighted digraphs G_n, in which all edges have unit weight and $\#\mathcal{B}^g\mathcal{H}_1(G_n) = n$

Proof Suppose such f exists and consider the following sequence of digraphs in which each edge has unit weight.

Intuitively, G_n is constructed by gluing n disjoint copies of G_1 along the path (v_0, v_1, v_2, v_3). Note that each copy of G_1 introduces a feature which dies at $t = 3$ so

$$\mathcal{B}^g\mathcal{H}_1(G_n) = \{[0, 3) \text{ with multiplicity } n\}. \tag{5.22}$$

Upon subdividing the edge $e = (v_1, v_2)$ via $S(e) = (1/2, 1/2)$, each feature changes to $[0, 2.5)$. Hence

$$d_{W_p}\big(\mathcal{B}^g\mathcal{H}_1(\mathbb{O}^s_S G_n), \mathcal{B}^g\mathcal{H}_1(G_n)\big) = \left(n \cdot \frac{1}{2^p}\right)^{1/p} = \frac{1}{2}n^{1/p} \tag{5.23}$$

which eventually exceeds the constant $f(\mathcal{NG}(e; G))$. □

5.4 Edge Collapse

Another potential structural perturbation is that of edge collapses, in which the two end points of an edge are identified and the edge deleted. In applications, this may happen particularly to low-weight edges, which cannot be discerned by the imaging method and hence collapsed to a vertex instead. Since we interpret edge-weights as corresponding to distance, we add half the weight of the collapsed edge to each of its neighbours so that the length of paths through the collapsed edge are not changed.

Definition 5.24 Given a weighted digraph $G = (V, E, w)$ we say a subset of edges $F \subseteq E$ is collapsible if the edge sets of the neighbourhoods $\mathcal{NG}(e; G)$ for each $e \in F$ are pairwise disjoint.

Definition 5.25 Given a weighted digraph $G = (V, E, w)$ and $F \subseteq E$ collapsible we define the edge collapse $\mathbb{O}^c_F G := (V_F, E_F, w_F)$ as follows.

(a) We define an equivalence relation $\sim$ on V such that $i \sim j \iff i = j$ or $(i, j) \in F$ or $(j, i) \in F$. The vertex set is the set of equivalence classes of this relation $V_F := {}^{V}\!/\!_{\sim}$.
(b) Given two distinct vertices $I, J \in V_F$, we include an edge $(I, J) \in E_F$ if and only if there is some $i \in I$ and $j \in J$ such that there is an edge $i \to j$ in G.

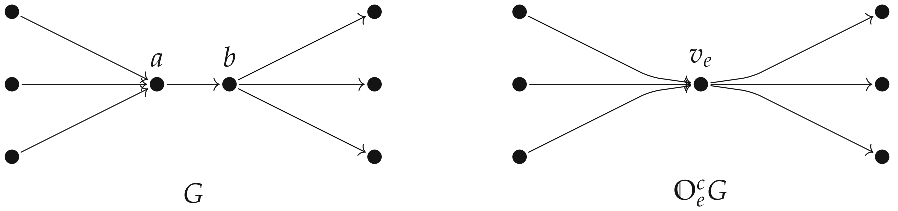

Fig. 13 Schematic of $\mathcal{NG}(e) \subseteq G$ and $\mathcal{NG}(v_e) \subseteq \mathbb{O}_e^c G$, under the assumptions of Theorem 5.28

(c) Finally, given an edge $(I, J) \in E_F$, we define the weight by

$$w_F((I, J)) := \min_{i \in I, j \in J \text{ s.t. } i \to j} \left(w(i, j) + \sum_{e \in F} \frac{w(e)}{2} \cdot \mathbb{1}_{(i,j) \in \mathcal{N}(e)} \right) \tag{5.24}$$

where $\mathbb{1}_{\tau \in \mathcal{N}(e)} = 1 \iff \tau \in \mathcal{N}(e)$, else $\mathbb{1}_{\tau \in \mathcal{N}(e)} = 0$.

Remark 5.26 The minimum is required in Eq. (5.24) because there may be an edge $i \to a$ and an edge $i \to b$ with different weights.

Notation 5.27 Thanks to the collapsible condition on F, all equivalence classes in V_F are singletons except for those of the form $\{a, b\}$ for $e = (a, b) \in F$. Therefore, we denote any singleton class $\{i\}$ by its sole representative i and the class $\{a, b\}$ by the symbol v_e.

Arbitrary edge collapses can drastically change the connectivity of the digraph and in turn alter the shortest-path quasimetric. However, given some control on the local neighbourhood of collapsed edges, it is possible to bound these effects and hence get local stability.

Theorem 5.28 *Given a weighted digraph, $G = (V, E, w) \in$ **WDgr** and $F \subseteq E$ collapsible, suppose that for each $e \in F$, $\mathcal{N}_{out}(\mathrm{st}(e)) = \{\mathrm{fn}(e)\}$ and $\mathcal{N}_{in}(\mathrm{fn}(e)) = \{\mathrm{st}(e)\}$ (as in Fig. 13). Then, for each $e \in F$, define*

$$\delta_e := w(e) + \min\left(\max_{v \in \mathcal{N}_{in}(\mathrm{st}(e))} w(v, \mathrm{st}(e)), \max_{v \in \mathcal{N}_{out}(\mathrm{fn}(e))} w(\mathrm{fn}(e), v) \right) \tag{5.25}$$

then $d_B(\mathcal{B}^g\mathcal{H}_1(G), \mathcal{B}^g\mathcal{H}_1(\mathbb{O}_F^c G)) \leq \max_{e \in F} \delta_e$.

Proof Denote $G_F := \mathbb{O}_F^c G = (V_F, E_F, w_F)$ and $\delta := \max_{e \in F} \delta_e$. Note that the condition on each $e \in F$ ensures given any edge $(I, J) \in E_F$, there is exactly one $i \in I$ and $j \in J$ such that $i \to j$.

We define two vertex maps. Firstly $f : V \to V_F$ is given by $a, b \mapsto v_e$ for each $e = (a, b) \in F$ and $v \mapsto v$ otherwise. Secondly, we define $g : V_F \to V$ as follows. For most elements of V_e we choose the only representative of the equivalence class $v \mapsto v$. The only classes containing more than one element are of the form $v_e = \{\mathrm{st}(e), \mathrm{fn}(e)\}$ for some $e \in F$. For this class, we choose $g(v_e) = \mathrm{st}(e)$ if

$$\max_{v \in \mathcal{N}_{in}(\mathrm{st}(e))} w(v, \mathrm{st}(e)) \geq \max_{v \in \mathcal{N}_{out}(\mathrm{fn}(e))} w(\mathrm{fn}(e), v), \tag{5.26}$$

else we choose $g(v_e) = \mathrm{fn}(e)$. We show that f and g define a δ-grounded interleaving.

Claim 5.29 *f defines a digraph map $G \to G_e$.*

Proof of Claim. All edges of G are mapped to edges of G_e, with the exception of $e = (a, b)$. The two endpoints of e are mapped to the same point, v_e. Therefore, f define a digraph map as required. □

Given a path $p : i \rightsquigarrow j$ in G, $f(p)$ is a path $f(i) \rightsquigarrow f(j)$ in G_e. Suppose the edges of p are all contained in $\mathcal{NG}(e; G)$ for some $e = (a, b) \in F$. If $i \in \mathcal{N}_{in}(a)$ and $j \in \mathcal{N}_{out}(b)$ then length of $f(p)$ equals that of p, thanks to the weight distribution. In any other case $f(p)$ can increase in length by at most $w(e)/2$.

Claim 5.30 *f defines a digraph map $G^t \to G_F^{t+\delta}$ for all $t \geq 0$.*

Proof of Claim. Given a path $p : i \rightsquigarrow j$ in G, $f(p)$ is a path $f(i) \rightsquigarrow f(j)$ in G_F. Suppose the edges of p are all contained in $\mathcal{NG}(e; G)$ for some $e = (a, b) \in F$. If $i \in \mathcal{N}_{in}(a)$ and $j \in \mathcal{N}_{out}(b)$ then length of $f(p)$ equals that of p, thanks to the weight distribution. Otherwise one of i or j must be a or b and in these cases $f(p)$ can increase in length by at most $w(e)/2 \leq \delta_e/2 \leq \delta/2$.

If $(i, j) \in E(G^t)$, then there is a path p joining $i \rightsquigarrow j$ of length at most t in G. We can decompose p into a sequence of maximal sub-paths $p_1, \ldots, p_m$ such that each p_k is either entirely contained in $\mathcal{NG}(e; G)$ for some $e \in F$ or contains no edge in any such neighbourhood. Sub-paths not contained in any $\mathcal{NG}(e; G)$ for $e \in F$ have the same length after mapping through f. Sub-paths p_i contained in some $\mathcal{NG}(e; G)$ for $e \in F$ where $1 < i < m$ must fully traverse from some in-neighbour of $\mathrm{st}(e)$ to an out-neighbour of $\mathrm{fn}(e)$ and thus $f(p_i)$ has the same length as p_i. The remaining p_1, p_m can increase in length by at most $\delta/2$ each and hence $f(p)$ is a path $f(i) \rightsquigarrow f(j)$ of length at most $t + \delta$. Therefore $(i, j) \in E(G_F^{t+\delta})$. □

Claim 5.31 *g defines a digraph map $G_F^t \to G^{t+\delta}$.*

Proof of Claim. Suppose $(i, j) \in E(G_F^t)$, then there is a path $p : i \rightsquigarrow j$ in G_F of length at most t. We construct a path p' in G as follows: replace any singleton class $i \in V_e$ with its single representative $i \in V$ and replace any other class v_e for $e = (a, b) \in F$ with the vertices a, b. We claim that p' is a path of length at most $t + \delta$ that contains a sub-path $g(i) \rightsquigarrow g(j)$.

As in the previous claim, we can decompose p into a sequence of maximal sub-paths $p_1, \ldots, p_m$ such that each p_k is either entirely contained in $\mathcal{NG}(v_e; G_F)$ for some $e \in F$ or contains no edge in any such neighbourhood. We let p_i' denote the corresponding sub-paths of p'.

If p_i is not contained in any $\mathcal{NG}(v_e; G_F)$ for $e \in F$ then p_i' has the same length as p_i. If p_i is contained in some $\mathcal{NG}(v_e; G_F)$ for $e \in F$ where $1 < i < m$, then p_i

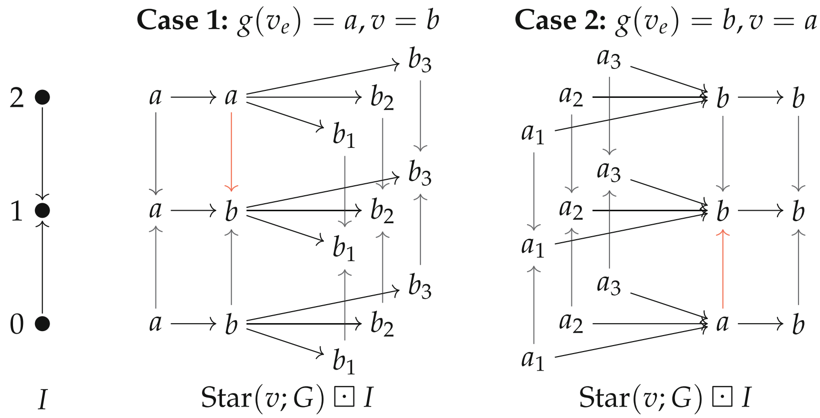

Fig. 14 Schematic for the homotopy between $g \circ f$ (on the top row) and id (on the bottom row). The label on each node of $\mathrm{Star}(v; G) \boxdot I$ is its image under the homotopy F_v. There are two cases depending on whether $g(v_e) = a$ or $g(v_e) = b$. The red edges denote the only time the image of a vertex differs between the layers (Color figure online)

must fully traverse from some in-neighbour of v_e to an out-neighbour of v_e. Then p'_i must also traverse from the same in-neighbour of $\mathrm{st}(v_e)$ to the same out-neighbour of $\mathrm{fn}(v_e)$. Since the weight of e is distributed evenly to the neighbouring edges, p'_i has the same length as p_i. The remaining sub-paths p_1, p_m can increase in length by at most $\delta/2$ each and hence p' is a path of length at most $t + \delta$.

Finally, note that if p traverses v_e for some $e \in F$ then both vertices $\mathrm{st}(e), \mathrm{fn}(e)$ appear in the path p'. Hence, no matter which vertex of e is chosen for $g(v_e)$, the path p' will traverse it. Therefore, p' contains a sub-path $g(i) \rightsquigarrow g(j)$. □

Claim 5.32 *g defines a digraph map $G_F \cup G_F^t \to G \cup G^{t+\delta}$.*

Proof of Claim. Thanks to the previous claim, we only need to check edges $(i, j) \in E(G_F)$. Any edge $(i, j) \in E(G_F)$ which is not incident to some v_e is mapped by g to itself $(i, j) \in E(G)$. For the remaining cases, thanks to the collapsibility condition on F, we can assume that exactly one of i, j is a vertex of the form $v_e \in V_F$ for some $e \in F$. Further assume that $g(v_e) = \mathrm{st}(e)$; the case where $g(v_e) = \mathrm{fn}(e)$ admits a similar proof.

If $j = v_e$ then $i \in V_e$ is a singleton and $i \in \mathcal{N}_{in}(\mathrm{st}(e))$ and hence $(g(i), g(v_e)) = (i, \mathrm{st}(e)) \in E(G)$. Else, if $i = v_e$ then j is a singleton and $j \in \mathcal{N}_{out}(\mathrm{fn}(e))$. Note that the path $(\mathrm{st}(e), \mathrm{fn}(e), j)$ in G is of length $w(e) + w(b, j) \leq \delta_e \leq \delta$. Therefore $(g(v_e), g(j)) = (\mathrm{st}(e), j) \in E(G^{t+\delta})$ for all $t \geq 0$. □

As vertex maps we note that $f \circ g = \mathrm{id}_{V_e}$ and hence it only remains to show that (g, f) has grounded codistortion $\leq \delta$. First, we note that the vertices $v \in V_{diff}(g, f)$ are precisely the endpoints of each $e \in F$ such that $g(v_e) \neq v$. Hence, $G_{diff}(g, f)$ is just the union of the closed stars of such v and moreover the edge sets of these neighbourhoods are disjoint thanks to the collapsibility condition. For a fixed line digraph I and each $v \in V_{diff}(g, f)$ we will define a path homotopy $F_v : \mathrm{Star}(v; G) \boxdot$

$I \to G^{2\delta}$ that is relative all the neighbours of v. We can then combine these homotopies according to the formula

$$F(v, j) := \begin{cases} F_v(v, j) & \text{if } v \in V_{diff}(g, f), \\ v & \text{otherwise} \end{cases} \tag{5.27}$$

to obtain the desired homotopy $F : G_{diff}(g, f) \to G^{2\delta}$.

The homotopies F_v and line digraph I are shown schematically in Fig. 14 for a vertex v which is the endpoint of some collapsed edge $e = (a, b) \in F$. All vertices are fixed through the homotopy except for v, whose image varies with $j \in I$ as shown in the schematic. There are two cases, depending on whether $g(v_e) = a$ or $g(v_e) = b$. Focusing on the first case, the vertex $v = b$ has a single in-neighbour a and a finite number of out-neighbours $b_1, b_2, \dots, b_k$. The fact that $g(v_e) = a$ implies that

$$w(e) + \max_{v \in \mathcal{N}_{out}(b)} w(b, v) = \delta_e \le \delta \tag{5.28}$$

and hence (a, b, b_i) is always a path of length at most δ. Therefore, all of the edges (a, b), (a, b_i) and (b, b_i) are contained within $G^{2\delta}$. A similar proof works in the case $g(v_e) = b$. □

Remark 5.33 Suppose G is a DAG and contains an edge $e = (a, b)$ that satisfies the condition of Theorem 5.28. Then one can number the vertices of G as $v_1, \dots, v_n$ such that $v_i \to v_j \implies i < j$ and moreover a and b are adjacent. However, note that this is *not* a sufficient condition for local stability to edge collapse. For an example, we refer the reader to Fig. 21 and the subsequent discussion in Sect. 6.2. In the digraph G_2, note that a, d, c, b is a valid ordering of the vertices but ${}^g\mathcal{H}_1(G_2) = \{[0, \infty)\}$ whilst ${}^g\mathcal{H}_1(\mathbb{O}^c_{(d,c)} G_2) = \{[0, 1.5)\}$.

Remark 5.34 Collapses of the sort described in Theorem 5.28 remove exactly $\#F$ vertices and $\#F$ edges, and do not change the number of weakly connected components. Therefore, the circuit rank of $\mathcal{U}(G)$ does not change and hence, by Corollary 4.3 we have $\#\mathcal{B}^g\mathcal{H}_1(G) = \#\mathcal{B}^g\mathcal{H}_1(\mathbb{O}^c_e G)$.

Given a collapse of the sort required by Theorem 5.28 and two vertices v, w that are not adjacent to a collapsed edge, there is a path $v \rightsquigarrow w$ in G if and only if there is such a path in $\mathbb{O}^c_F G$. In general this is not the case (see Fig. 15); indeed this leads to local instabilities.

Theorem 5.35 *There exists no function $f : \mathbf{WDgr} \to \mathbb{R}$ such that for any weighted digraph $G = (V, E, w) \in \mathbf{WDgr}$ and any edge $e \in E$ therein we have*

$$d_B\big(\mathcal{B}^g\mathcal{H}_1(G), \mathcal{B}^g\mathcal{H}_1(\mathbb{O}^c_e G)\big) \le f\big(\mathcal{NG}(e; G)\big). \tag{5.29}$$

Proof Suppose such f exists then consider the weighted digraphs illustrated in Fig. 15. First, denote the edges $e_1 := (v_0, v_5)$ and $e_2 := (v_2, v_3)$ in all three graphs (wherein those edges exist). Note that G' is independent of the weight W; hence we can safely define

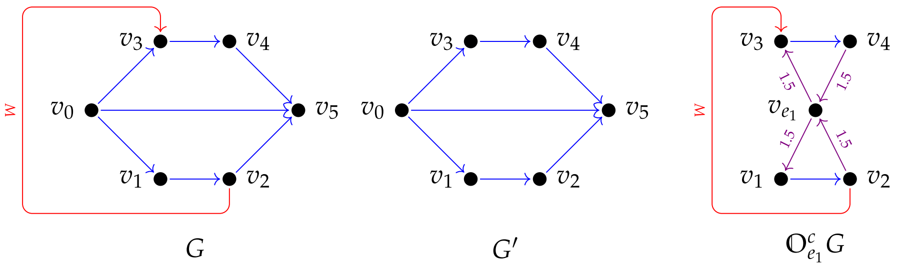

Fig. 15 An example weighted digraph which illustrates that ${}^g\mathcal{H}_1$ is locally unstable to arbitrary edge collapses. Unlabelled edges have weight 1. Note that there is no path $v_4 \rightsquigarrow v_1$ in G but there is such a path in $\mathbb{O}^c_{e_1}G$

$W{:=}2f(\mathcal{NG}(e_1; G')) + 4$ to be the weight of e_2 in G. Observe that $\mathcal{NG}(e_1; G) = \mathcal{NG}(e_1; G')$ and hence

$$W = 2f(\mathcal{NG}(e_1; G')) + 4 = 2f(\mathcal{NG}(e_1; G)) + 4. \tag{5.30}$$

Initially, $\mathcal{B}{}^g\mathcal{H}_1(G)$ has 3 features, which die at 2, 2 and W. After collapsing e_1, $\mathcal{B}{}^g\mathcal{H}_1(\mathbb{O}^c_{e_1}G)$ has 3 features, which die at 2.5, 2.5 and 3. The longer feature, supported on the red edge (v_2, v_3), has a reduced death-time in $\mathbb{O}^c_{e_1}G$ because there is a shortcut (v_2, v_{e_1}, v_3), of length 3. Any bijection between these features (and the diagonals) must have bottleneck cost at least $\min(W/2, W-3) > f(\mathcal{NG}(e_1; G))$. □

While this seems like a serious problem for our descriptor, note that the collapse in Fig. 15 makes significant changes to the topology of the underlying digraph. Originally, G was a DAG with source v_0 and sink v_5; the edge collapse identified these two nodes and introduced directed cycles. Moreover, in G the only path $v_2 \rightsquigarrow v_3$ was via the costly red edge but in $\mathbb{O}^d_e G$ there is a shortcut via v_e. Therefore, since the profile of paths has changed drastically, it is arguably desirable that our descriptor changes too.

5.5 Edge Deletion

5.5.1 General Case

Another important class of structural perturbations is edge deletion. Intuitively, as with edge collapse, some edge deletions can have drastic impact on the descriptor whereas some deletions are minor events.

Definition 5.36 Given a weighted digraph $G = (V, E, w)$ and a subset of edges $F \subseteq E$, we define $\mathbb{O}^d_F G{:=}(V, E\backslash F, w_F)$ where w_F is obtained by restricting w to $E\backslash F$. If $F = \{e\}$ is a single edge, we denote this $\mathbb{O}^d_e G$.

Remark 5.37 We do not distribute the weight of e to neighbouring edges. In applications, we anticipate these errors may appear due to, for example, imaging errors, in which case the neighbouring edges would not change weight.

We find that our descriptor is stable to deletions but the bound depends on the minimum length of a possible diversion. In general, this diversion cost may be infinite.

Theorem 5.38 *Given a weighted digraph $G = (V, E, w) \in$ **WDgr** and a subset of edges $F \subseteq E$, let d and d_F denote the shortest-path quasimetric for G and $\mathbb{O}^d_F G$ respectively. Then*

$$d_B\big(\mathcal{B}^g\mathcal{H}_1(G), \mathcal{B}^g\mathcal{H}_1(\mathbb{O}^d_F G)\big) \le \max\left(\max_{i,j\in V} |d_F(i,j) - d(i,j)|, \max_{(i,j)\in F} d_F(i,j)\right). \tag{5.31}$$

Proof Denote $G_F := \mathbb{O}^d_F G = (V_e, E \setminus F, w_F)$,

$$\delta_1 := \max_{i,j\in V} |d_F(i,j) - d(i,j)|, \quad \delta_2 := \max_{(i,j)\in F} d_F(i,j) \tag{5.32}$$

and $\delta := \max(\delta_1, \delta_2)$. We claim that id_V constitutes a δ-shifting vertex map $G \to G_F$ and $G_F \to G$. Then, since the vertex maps are both the identity, they constitute a δ-grounded interleaving and thus we obtain the result via the main stability theorem.

The shortest-path distance between any two nodes in G increases by at most δ_1 upon deleting the edges of F. Hence, since $\delta \ge \delta_1$, id_V defines a digraph map $G^t \to G_F^{t+\delta}$ and $G_F^t \to G^{t+\delta}$ for all $t \ge 0$. Then $E \setminus F \subseteq E$, so id_V certainly defines a digraph map $G_F \to G$ and thus $G_F \cup G_F^t \to G \cup G^{t+\delta}$ for all $t \ge 0$.

In the other direction, id_V does *not* define a digraph map $G \to G_F$. However, given any deleted edge $e \in F$, $e \in E(G_F^t)$ for all $t \ge \delta_2$. Hence id_V *does* define a digraph map $G \cup G^t \to G_F \cup G_F^{t+\delta}$ for all $t \ge 0$. □

Remark 5.39 The above bound is infinite if and only if there is some deleted edge $e = (i, j) \in F$ such that the only path $i \rightsquigarrow j$ in G is through e.

Corollary 5.40 *Given a weighted digraph $G = (V, E, w) \in$ **WDgr** and an edge $e = (a, b) \in E$ let d and d_e denote the shortest-path quasimetric for G and $\mathbb{O}^d_e G$ respectively. Then*

$$d_B\big(\mathcal{B}^g\mathcal{H}_1(G), \mathcal{B}^g\mathcal{H}_1(\mathbb{O}^d_e G)\big) \le d_e(a, b). \tag{5.33}$$

Proof Denote $G_e := \mathbb{O}^d_e G$ and $\delta := d_e(a, b)$. We aim to show that

$$\max_{i,j\in V} |d_e(i,j) - d(i,j)| \le |d_e(a,b) - d(a,b)| \le d_e(a,b) = \delta. \tag{5.34}$$

The result then follows by Theorem 5.38. To see this inequality, note that for arbitrary $i, j \in V$ we have $d_e(i, j) \ge d(i, j)$ since any path $i \rightsquigarrow j$ in G_e is also a path $i \rightsquigarrow j$ in G of the same length. Moreover, there is path $p_e : a \rightsquigarrow b$ in G_e of length at most δ. Then, given a path $p : i \rightsquigarrow j$ in G of length t, the path contains e at most once. We can replace e with p_e to obtain a new path $i \rightsquigarrow j$ in G_e of length at most $t + (\delta - w(e)) \le t + \delta$. Therefore, $d_e(i, j) \le d(i, j) + \delta$. □

In general, Theorem 5.38 is a non-local bound, but if an edge has an alternative route in its local neighbourhood then the bound becomes local.

Corollary 5.41 *Given a weighted digraph $G = (V, E, w)$ and an edge $e = (a, b) \in E$ such that there exists a vertex $v \in V$ such that $a \to v \to b$ then*

$$d_B\big(\mathcal{B}^g\mathcal{H}_1(G), \mathcal{B}^g\mathcal{H}_1(\mathbb{O}_e^d G)\big) \le w(a, v) + w(v, b). \tag{5.35}$$

In general, the shortest-path distance between the endpoints of an edge, upon its deletion, can depend on *all* remaining edges in the graph. Therefore, the bound of Theorem 5.38 is non-local and indeed no generic, local stability theorem is possible.

Theorem 5.42 *There exists no function $f : \mathbf{WDgr} \to \mathbb{R}$ such that for any digraph $G = (V, E, w) \in \mathbf{WDgr}$ and any edge $e \in E$ therein we have*

$$d_B\big(\mathcal{B}^g\mathcal{H}_1(G), \mathcal{B}^g\mathcal{H}_1(\mathbb{O}_e^d G)\big) \le f\big(\mathcal{NG}(e; G)\big). \tag{5.36}$$

Proof Suppose such f exists then consider the following weighted digraph G, where $e{:=}(v_1, v_3)$ and $W{:=}2f(\mathcal{NG}(e; G)) + 2$ (Fig. 16).
Note, G has a single feature which dies at time W, whereas $\mathbb{O}_e^d G$ has no features. Therefore, the bottleneck distance is $W/2 > f(\mathcal{NG}(e; G))$. □

5.5.2 Separating Edges

Since all features are born at $t = 0$ and ${}^gZ_1(G, 0)$ has a basis of simple undirected circuits, one might expect that edges never involved in such circuits can be safely deleted without changing the descriptor. Indeed, this is the case and is a direct consequence of the wedge decomposition theorem.

Definition 5.43 In a weighted digraph $G = (V, E, w)$, an edge $e = (a, b) \in E$ is called a separating edge if a and b are not weakly connected in $\mathbb{O}_e^d G$.

Remark 5.44 An edge is separating if and only if there are no simple undirected circuits containing it.

Corollary 5.45 *Given a weighted digraph $G = (V, E, w)$ and a separating edge $e = (a, b) \in E$,*

$${}^g\mathcal{H}_1(G) \cong {}^g\mathcal{H}_1(\mathbb{O}_e^d G). \tag{5.37}$$

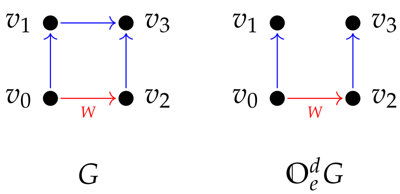

Fig. 16 An example weighted digraph which illustrates that ${}^g\mathcal{H}_1$ is locally unstable to arbitrary edge deletions. Unlabelled edges have weight 1

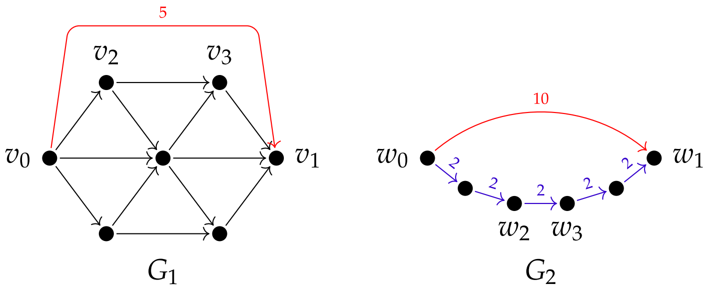

Fig. 17 Some example weighted digraphs, used to interpret the consequences of the stability theorems obtained in Sect. 5.5. Unlabelled edges have weight 1

Proof First note that a and b are both wedge vertices. Let V_1 denote the vertices in the weak connected component of a in $\mathbb{O}^d_e G$. Define $V_2 := \{a, b\}$. Finally, define $V_3 := (V(G) \setminus V_1)$. Let G_1, G_2 and G_3 denote the induced subgraph of G on V_1, V_2 and V_3 respectively.

Then a wedge decomposition of G is $G = (G_1 \vee_a G_2) \vee_b G_3$ and a disjoint union decomposition of $\mathbb{O}^d_e G$ is $\mathbb{O}^d_e G = G_1 \sqcup G_3$. Note that G_2 is just a single edge connecting two vertices so ${}^g\mathcal{H}_1(G_2)$ is the trivial persistent vector space. Using Theorems 4.20 and 4.23, we see

$$\begin{aligned} {}^g\mathcal{H}_1(G) &\cong {}^g\mathcal{H}_1(G_1) \oplus {}^g\mathcal{H}_1(G_2) \oplus {}^g\mathcal{H}_1(G_3) \cong {}^g\mathcal{H}_1(G_1) \oplus {}^g\mathcal{H}_1(G_3) \\ &\cong {}^g\mathcal{H}_1(\mathbb{O}^d_e G) \end{aligned} \tag{5.38}$$

as required. □

5.5.3 Interpretation

Theorem 5.38 tells us that we are stable to deleting edges which have fast diversions. That is, if there is a path $p : i \rightsquigarrow j$, not involving the edge $e := (i, j)$, of length δ, then removing e changes the barcode by at most δ in bottleneck distance. Note, this bound is independent of the weight of the deleted edge $w(e)$.

To illustrate this point, consider G_1 in Fig. 17. Removing (v_2, v_3) incurs a bottleneck cost of at most 2, since there is a diversion of length 2. Likewise, despite being a highly-weighted edge, we can also remove (v_0, v_1) for a bottleneck cost of at most 2.

On the other hand, consider now G_2 in Fig. 17. The barcode contains a single feature, $\mathcal{B}{}^g\mathcal{H}_1(G_2) = \{[0, 10)\}$. The edge (w_0, w_1) has a high weight and the only diversion is via the black edges, of length 10. Deleting the edge (w_0, w_1) removes the single feature, which can be matched with $[5, 5)$ on the diagonal and thus incurs a bottleneck cost of 5. Moreover, deleting one of the smaller edges (for example (w_2, w_3)) also incurs a bottleneck cost of 5 since it deletes the same feature.

5.6 Vertex Deletion

Definition 5.46 Given a weighted digraph $G = (V, E, w)$ and a vertex $v \in V$, we define $\mathbb{O}^d_v G := (V \setminus \{v\}, E_v, w_v)$ where $E_v := E \cap (V\setminus\{v\}) \times (V\setminus\{v\})$ and w_v is obtained by restricting w to E_v. Given a subset of vertices $W \subseteq V$, we define $\mathbb{O}^d_W G$ iteratively by choosing an arbitrary $w \in W$ then setting $\mathbb{O}^d_W G := \mathbb{O}^d_{W\setminus\{w\}} \mathbb{O}^d_w G$.

Remark 5.47 As with edge deletion, we do not distribute the weight of deleted edges because we anticipate these errors occurring due to imaging, in which case the neighbouring edges would not change weight.

Since a single vertex graph has trivial GRPPH the disjoint union decomposition theorem (Theorem 4.20) allows us to delete isolated vertices.

Corollary 5.48 *Given a weighted digraph $G = (V, E, w)$ and an isolated vertex $v_i \in V$ (i.e. $\mathcal{N}(v) = \emptyset$), then*

$$\mathcal{B}^g\mathcal{H}_1(G) \cong \mathcal{B}^g\mathcal{H}_1(\mathbb{O}^d_{v_i} G). \tag{5.39}$$

However, in general, deleting a vertex from a digraph can drastically change its topology. This follows immediately from Theorems 5.16 and 5.42 since a local vertex deletion stability theorem would imply a local edge deletion stability theorem.

Corollary 5.49 *There exists no function $f : \mathbf{WDgr} \to \mathbb{R}$ such that for any digraph $G = (V, E, w)$ and any vertex $v \in V$ therein we have*

$$d_B\big(\mathcal{B}^g\mathcal{H}_1(G), \mathcal{B}^g\mathcal{H}_1(\mathbb{O}^d_v G)\big) \leq f\big(\mathcal{NG}(v; G)\big). \tag{5.40}$$

As with edge deletion, it is possible to bound the effect of a deleting a vertex v_0, but the bound depends on the choice of an alternative vertex v_1 and the length of a number of diversions. In general, this is a non-local bound.

Theorem 5.50 *Given a weighted digraph $G = (V, E, w)$ and a vertex $v_0 \in V$, let d denote the shortest-path quasimetric. Suppose that for some $\delta > 0$ there is some vertex $v_1 \in V \setminus \{v_0\}$ such that, for all $a \in \mathcal{N}_{in}(v_0)$ and $b \in \mathcal{N}_{out}(v_0)$*

(a) $d(a, v_0) \leq 2\delta$ and $d(v_0, b) \leq 2\delta$;
(b) there are paths $p_a : a \rightsquigarrow v_1$ and $q_b : v_1 \rightsquigarrow b$ each of length at most δ and not traversing v_0;
(c) there is a path $p_{a,b} : a \rightsquigarrow b$ of length at most $w(a, v_0) + w(v_0, b) + \delta$ and not traversing v_0;
(d) there is a path q connecting $v_0 \rightsquigarrow v_1$ or $v_1 \rightsquigarrow v_0$ of length at most 2δ.

Then $d_B(\mathcal{B}^g\mathcal{H}_1(G), \mathcal{B}^g\mathcal{H}_1(\mathbb{O}^d_{v_0} G)) \leq \delta$.

Proof Note there is an inclusion map $g : V(\mathbb{O}^d_{v_0} G) \to V(G)$ and define a map in the opposite direction $f : V(G) \to V(\mathbb{O}^d_{v_0} G)$ by

$$f(v) := \begin{cases} v_1 & \text{if } v = v_0, \\ v & \text{otherwise.} \end{cases} \tag{5.41}$$

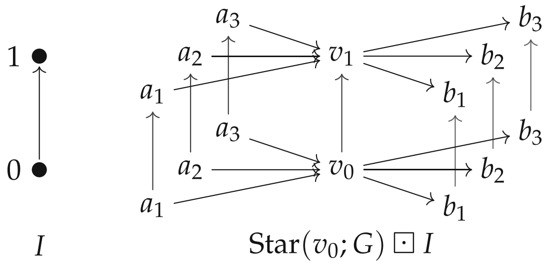

Fig. 18 Schematic for the homotopy between $g \circ f$ (on the top row) and id (on the bottom row). The label on each node in $\mathrm{Star}(v_0; G) \boxdot I$ is its image under the homotopy F

Denote the altered digraph by $G_d := \mathbb{O}^d_{v_0} G$. We claim that, under the conditions of the theorem, f and g constitute a δ-grounded interleaving.

First note that g is a digraph map $G_d \to G$ because G_d is a subgraph. Moreover, g induces a contraction on the shortest-path quasimetric and hence g is certainly a δ-shifting vertex map.

Claim 5.51 *f induces digraph maps $G^t \to G_d^{t+\delta}$.*

Proof of Claim. Suppose $(i, j) \in E(G^t)$; hence there is a path $p : i \rightsquigarrow j$ of length at most t in G. We require a path $f(i) \rightsquigarrow f(j)$ of length at most $t + \delta$ in G_d.

Assume that $i, j \neq v_0$ so that $f(i) = i$ and $f(j) = j$. If p does not traverse v_0 then p is a path in G_d and we are done. Else p contains a sequence of vertices (a, v_0, b) for some $a \in \mathcal{N}_{in}(v_0)$ and $b \in \mathcal{N}_{out}(v_0)$. Thanks to condition (c) of the theorem, we can replace this sequence with the path $p_{a,b}$ to obtain a new path $i \rightsquigarrow j$ in G_d, of length at most $t + \delta$.

Next, suppose $j = v_0$ but $i \neq v_0$. Then p must finish with the sequence (a, v_0) for some $a \in \mathcal{N}_{in}(v_0)$. We can replace this sequence with p_a to obtain a new path $i \rightsquigarrow v_1 = f(v_0)$ in G_d of length at most δ. A similar proof works for the final case $i = v_0$, $j \neq v_0$. □

Claim 5.52 *f induces digraph maps $G \cup G^t \to G_d \cup G_d^{t+\delta}$.*

Proof of Claim. Thanks to the previous claim we only need to consider edges in $E(G)$. Given an edge $(i, j) \in E(G)$ suppose that $i, j \neq v_0$. Then $f(i) = i$ and $f(j) = j$ and (i, j) is still an edge in G_d. Suppose $(a, v_0) \in E(G)$, then $f(a) = a$ and $f(v_0) = v_1$; the existence of the path p_a shows that $(a, v_1) \in E(G_d^{t+\delta})$ for all $t \geq 0$. Likewise, if $(v_0, b) \in E(G)$ then the existence of q_b shows that $(v_1, b) \in E(G_d^{t+\delta})$ for all $t \geq 0$. □

Now note that $f \circ g = \mathrm{id}$ is the identity on $V(G_d)$ and hence (f, g) has grounded codistortion $\leq \delta$.

Claim 5.53 *(g, f) has grounded codistortion $\leq \delta$.*

Proof of Claim. First note that $G_{diff}(g, f) = \mathrm{Star}(v_0; G)$ so we require a homotopy $F : \mathrm{Star}(v_0) \boxdot I \to G^{2\delta}$ for some line digraph I. Note further that $g \circ f$ fixes all vertices except v_0 which gets mapped to v_1. Assume that q is a path $v_0 \rightsquigarrow v_1$ of length at most 2δ, then set $I = I_+$ and define the homotopy via

$$F(v, j) := \begin{cases} v_1 & \text{if } (v, j) = (v_0, 1), \\ v & \text{otherwise}, \end{cases} \tag{5.42}$$

which is shown in Fig. 18. Each of the edges in the lower layer belong to $G^{2\delta}$ thanks to condition (a) of the theorem. Each of the edges in the upper layer belong to $G^{2\delta}$ thanks to condition (b) of the theorem. The only non-trivial vertical edge is $v_0 \to v_1$ which exists in $G^{2\delta}$ thanks to our assumption on q. Moreover, F is relative all vertices except v_0, as required. In the case where q is a path $v_1 \rightsquigarrow v_0$ we follow the same construction and proof, except we take $I = I_-$. □

The bound on bottleneck distance now follows from the main stability theorem. □

This theorem can be extended to the case of deleting multiple vertices, $W \subseteq V$. In order for the proof above to easily extend, each of the neighbourhoods must be sufficiently isolated, i.e. the subgraphs $\mathrm{Star}(w; G)$ for $w \in W$ should be pairwise edge-disjoint. This allows us to join the homotopies around each $w \in W$ together, as in the proof of Theorem 5.16.

Next, one must choose a replacement vertex for each of the deleted vertices, i.e. a map $f : W \to V \setminus W$. The resultant bound on the bottleneck distance will then depend on a number of factors, such as:

1. the extent to which f changes the quasi-metric, i.e. the minimum δ such that f induces a digraph map $G^t \to (\mathbb{O}^d_W G)^{t+\delta}$ for all $t \geq 0$;
2. the maximum distance in $\mathbb{O}^d_W G$ amongst $d(a, f(v))$ and $d(f(v), b)$ over $v \in W$, $a \in \mathcal{N}_{in}(v; G)$, $b \in \mathcal{N}_{out}(v; G)$;
3. the maximum distance in G amongst $d(a, v)$ and $d(v, b)$ over $v \in W$, $a \in \mathcal{N}_{in}(v; G)$, $b \in \mathcal{N}_{out}(v; G)$;
4. the maximum distance between v and $f(v)$ (in either direction) in $\mathbb{O}^d_W G$, over $v \in W$.

Once these factors are controlled, by the existence of sufficiently short paths in G and $\mathbb{O}^d_W G$, then a similar proof yields a bound on the bottleneck distance. However, one may have to replace F with a two-step homotopy to allow for the possibility that there are only short paths $w_1 \rightsquigarrow f(w_1)$ and $f(w_2) \rightsquigarrow w_2$ for some $w_1, w_2 \in W$.

5.7 Combining Operations

Given $G_1, G_2, G_3 \in \mathbf{WDgr}$, suppose that the previous stability theorems give

$$d_b(\mathcal{B}^g\mathcal{H}_1(G_1), \mathcal{B}^g\mathcal{H}_1(G_2)) \leq \delta_1 \quad \text{and} \quad d_b(\mathcal{B}^g\mathcal{H}_1(G_2), \mathcal{B}^g\mathcal{H}_1(G_3)) \leq \delta_2. \tag{5.43}$$

One immediately obtains the bound $d_B(\mathcal{B}^g\mathcal{H}_1(G_1), \mathcal{B}^g\mathcal{H}_1(G_3)) \leq \delta_1 + \delta_2$. A natural question arises as to when this bound can be improved upon. Suppose that $\{f, g\}$ is a δ_1-grounded interleaving and $\{p, q\}$ is a δ_2-grounded interleaving arranged as follows:

$$V(G_1) \underset{g}{\overset{f}{\rightleftarrows}} V(G_2) \underset{q}{\overset{p}{\rightleftarrows}} V(G_3).$$

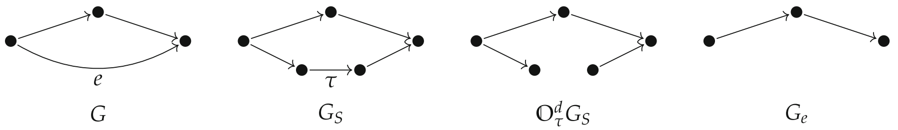

Fig. 19 Schematic of the four weighted digraphs in Theorem 5.54, where e is the curved bottom edge and τ is as labelled

It is easy to see that both $p \circ f$ and $g \circ q$ form $(\delta_1 + \delta_2)$-shifting vertex maps. However, it is not clear, a priori, that either of the ordered pairs have grounded codistortion $\leq (\delta_1 + \delta_2)$. Nevertheless, depending on the combination of operations, it is often possible to show that $\{p \circ f, g \circ q\}$ is a δ-grounded interleaving for some $\delta \leq \delta_1 + \delta_2$.

As an illustrative example, given $G \in \mathbf{WDgr}$, suppose an edge $e \in E(G)$ is subdivided according to some subdivision S, to obtain $G_S := \mathbb{O}^s_S G$. Then suppose one of the child edges $\tau := \tau_{e,i}$ is deleted (as shown in Fig. 19), to obtain $\mathbb{O}^d_\tau G_S$. In the final digraph, $\mathbb{O}^d_\tau G_S$, there is no alternative path between the endpoints of τ so Theorem 5.38 would yield an infinite bound

$$d_B\big(\mathcal{B}^g\mathcal{H}_1(G_S), \mathcal{B}^g\mathcal{H}_1(\mathbb{O}^d_\tau G_S)\big) < \infty. \tag{5.44}$$

However, the remaining child edges and vertices from the subdivision can be further deleted from $\mathbb{O}^d_\tau G_S$ to obtain $G_e := \mathbb{O}^d_e G$. Then, Corollaries 5.45 and 5.48 imply that ${}^g\mathcal{H}_1(\mathbb{O}^d_\tau G_S) \cong {}^g\mathcal{H}_1(G_e)$. Combining Theorems 5.16 and 5.38, we can bound

$$\begin{aligned} d_B\big(\mathcal{B}^g\mathcal{H}_1(G_S), \mathcal{B}^g\mathcal{H}_1(\mathbb{O}^d_\tau G_S)\big) &= d_B\big(\mathcal{B}^g\mathcal{H}_1(G_S), \mathcal{B}^g\mathcal{H}_1(G_e)\big) &(5.45)\\ &\leq d_B\big(\mathcal{B}^g\mathcal{H}_1(G_S), \mathcal{B}^g\mathcal{H}_1(G)\big) \\ &\quad + d_B\big(\mathcal{B}^g\mathcal{H}_1(G), \mathcal{B}^g\mathcal{H}_1(G_e)\big) &(5.46)\\ &\leq w(e) + d_e(\mathrm{st}(e), \mathrm{fn}(e)) &(5.47)\end{aligned}$$

where d_e is the shortest-path quasimetric in G_e.

Recall that the grounded interleaving used to prove edge deletion stability (Theorem 5.38) was a pair of identity vertex maps. Therefore, in the following, we show that the grounded interleaving used to prove subdivision stability directly yields a sharper bound for this combination of operations.

Theorem 5.54 *Given a weighted digraph $G = (V, E, w) \in$ **WDgr**, an edge $e = (a, b) \in E$ and a subdivision $S : \{e\} \to \mathrm{int}(\Delta^d)$, denote $G_S := \mathbb{O}^s_S G$ and $G_e := \mathbb{O}^d_e G$. Let d_e denote the shortest-path quasimetric in G_e. Choose any of the child edges $\tau := \tau_{e,i} \in E(G_S)$, then.*

$$d_B\big(\mathcal{B}^g\mathcal{H}_1(G_S), \mathcal{B}^g\mathcal{H}_1(\mathbb{O}^d_\tau G_S)\big) \leq \max\left(d_e(a, b), \frac{w(e)}{2}\right). \tag{5.48}$$

Proof As discussed above, it remains to prove

$$d_B\big(\mathcal{B}^g\mathcal{H}_1(G_S), \mathcal{B}^g\mathcal{H}_1(G_e)\big) \leq \max\left(d_e(a,b), \frac{w(e)}{2}\right) =: \delta. \tag{5.49}$$

Define the following vertex maps

$$f : V(G_e) \to V(G_S) \quad v \mapsto v; \tag{5.50}$$

$$g : V(G_S) \to V(G_e) \quad v \mapsto \begin{cases} v & \text{if } v \in V_{old}, \\ \mathrm{st}(e) & \text{if } v_{e,i} \in V_{new} \text{ and } CS(e)_i < 1/2, \\ \mathrm{fn}(e) & \text{if } v_{e,i} \in V_{new} \text{ and } CS(e)_i \geq 1/2. \end{cases} \tag{5.51}$$

We note that f induces a contraction digraph map $G_e \to G_S$ and is hence a δ-shifting vertex map.

Claim 5.55 *g induces a digraph map $G_S^t \to G_e^{t+\delta}$, for every $t \geq 0$.*

Proof of Claim. Given $(i, j) \in E(G_S^t)$, there is a path $p : i \rightsquigarrow j$ in G_S of length at most t. Since $d_e(a, b) \leq \delta$ there is a path $p_e : a \rightsquigarrow b$ in G_e of length at most δ. We construct a new trail p' in G_e as follows.

If the entire sequence of child edges $(\tau_{e,1}, \ldots, \tau_{e,d(e)})$ appears in p then we replace that sequence with p_e. If $i \in V_{new}$ and $g(i) = a$ then we replace the initial sequence of child edges with p_e. If $i \in V_{new}$ and $g(i) = b$ then we simply remove the initial sequence of child edges. Likewise, if $j \in V_{new}$ and $g(j) = b$ then we replace the final sequence of child edges with p_e. If $j \in V_{new}$ and $g(j) = a$ then we simply remove the final sequence of child edges. This yields a trail $p' : g(i) \rightsquigarrow g(j)$. Since p cannot repeat edges, this construction inserts p_e at most once and hence the length of p' is at most $t + \delta$. Therefore $(i, j) \in E(G_e^{t+\delta})$. □

Claim 5.56 *g induces a digraph map $G_S \cup G_S^t \to G_e \cup G_e^{t+\delta}$, for every $t \geq 0$.*

Proof of Claim. It remains to check the image of edge $e \in E(G_S)$. Any un-subdivided edge $e \in E_{old}$ is preserved under g. Given an edge $\tau_{e,i} = (x, y) \in E_{new}$ then $\tau = (v_{e,i-1}, v_{e,i})$ for some i and there are three cases

$$(g(v_{e,i-1}), g(v_{e,i})) = (\mathrm{st}(e), \mathrm{st}(e)) \text{ or } (\mathrm{st}(e), \mathrm{fn}(e)) \text{ or } (\mathrm{fn}(e), \mathrm{fn}(e)). \tag{5.52}$$

Hence either $g(x) = g(y)$ or $(g(x), g(y)) = e$. The edge e does not appear in G_e but it does appear in $G_e^{t+\delta}$ for all $t \geq 0$. Therefore g defines a digraph map as required. □

First observe $g \circ f = \mathrm{id}_{V(G)}$ and hence (g, f) has grounded codistortion ≤ 0. The homotopy used to show (f, g) has grounded codistortion $\leq \delta$ is identical to the corresponding homotopy in the proof of Theorem 5.16. However, note that we require $2\delta \geq w(e)$ so that all edges in the image of the homotopy are edges of $G_S^{2\delta}$. □

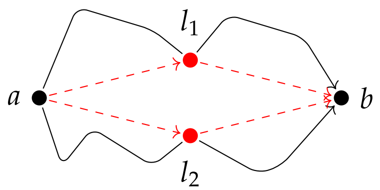

Fig. 20 Illustration of the weighted digraph G, considered in Proposition 6.1. The top path p_1 has length l_1 and the bottom path p_2 has length l_2. The limiting death time of the sole feature corresponds to the limiting value of the earliest time that a long square appears of the form $av_1b - av_2b$ where v_1 is along p_1 and v_2 is along p_2 (as drawn in red) dashed lines) (Color figure online)

6 Examples

6.1 Iterated Medial Subdivision

In order to develop intuition for how the descriptor behaves under iterative subdivision, we explicitly derive the limiting diagram for a DAG with exactly one loop (shown in Fig. 20). Certainly, the diagram contains exactly one feature which is born at $t = 0$. Intuitively, the death time corresponds to the earliest time that a long square can appear between the source and sink nodes, filling in the central hole.

Proposition 6.1 *Suppose $G \in$ **WDag** is the union two directed paths p_1, p_2 from a source to a sink, with lengths $l_1 \geq l_2$ respectively. Recall the definition of iterated medial subdivision (Definition 5.21). Then*

$$\lim_{n\to\infty} \mathcal{B}^g\mathcal{H}_1(\mathrm{IMS}_n(G)) = \left\{\!\!\left\{ \left[0, \frac{1}{2}l_1\right) \right\}\!\!\right\} \tag{6.1}$$

Proof For brevity we denote $G_n := \mathrm{IMS}_n(G)$. By Corollary 4.3, the barcode $\mathcal{B}^g\mathcal{H}_1(G_n)$ has exactly one feature. Let $p_i^{(n)}$ denote the path $a \rightsquigarrow b$ in G_n arising from subdividing the edges of p_i. For each n, let $c^{(n)}$ denote the simple undirected circuit in G_n which follows $p_1^{(n)}$ and then $p_2^{(n)}$ in reverse. Clearly $\{\mathfrak{R}(c^{(n)})\}$ is a persistence basis for ${}^g\mathcal{H}_1(G_n)$. Therefore, it suffices to show $\mathcal{D}(\mathfrak{R}(c^{(n)})) \to \frac{1}{2}l_1$ as $n \to \infty$. Fix some natural n.

Using Lemma 4.6 we see $\mathcal{D}(\mathfrak{R}(c^{(n)})) \leq \max(h_1^{(n)}, h_2^{(n)})$ where

$$h_i^{(n)} := \min\left\{ t \geq 0 \;\middle|\; \exists v_i \in V \text{ along } p_i^{(n)} \text{ such that } d(a, v_i) \leq t \text{ and } d(v_i, b) \leq t \right\}. \tag{6.2}$$

Note that $h_1^{(n)} \to \frac{1}{2}l_1$ and $h_2^{(n)} \to \frac{1}{2}l_2$ and hence $\max(h_1^{(n)}, h_2^{(n)}) \to \frac{1}{2}l_1$ as $n \to \infty$.

Next, we wish to show $\mathcal{D}(\mathfrak{R}(c^{(n)})) \geq \frac{1}{2}l_1$. Choose arbitrary $t_2 < \frac{1}{2}l_1$, then it suffices to show that $\dim {}^gH_1(G_n, t_2) \neq 0$. In order to do so, we claim the inclusion chain map

$$\begin{array}{ccccc}
0 & \longrightarrow & C_1(G_n) & \xrightarrow{{}^g\partial_1} & C_0(G_n) \\
\downarrow{\scriptstyle j_2} & & \downarrow{\scriptstyle j_1} & & \downarrow{\scriptstyle j_0} \\
C_2(G_n^{t_2}) & \xrightarrow{{}^g\partial_2} & C_1(G_n \cup G_n^{t_2}) & \xrightarrow{{}^g\partial_1} & C_0(G_n \cup G_n^{t_2})
\end{array}$$

induces an isomorphism on homology in degree 1. It then follows that $\dim {}^g H_1(G_n, t_2) = 1$ because the first homology of top row is the real cycle space of G_n.

We define a chain map q in the opposite direction to j. In degree 2, q_2 is the zero map and in degree 0, q_0 is the identity map. Finally in degree 1, given $(i, j) \in C_1(G_n \cup G_n^{t_2})$, if $(i, j) = (a, b)$ then let $p_{i,j} := p_2$, otherwise let $p_{i,j}$ denote the unique path $i \rightsquigarrow j$ in G_n Then q_1 is given by $q_1(ij) := \mathfrak{R}(p_{i,j})$. This is a chain map because there is no 2-path avb where v is somewhere along $p_1^{(n)}$.

It is certainly the case that, at the level of chain maps, $q_1 j_1 = \mathrm{id}$. Choose arbitrary $(i, j) \in C_1(G_n \cup G_n^{t_2})$ and note that $j_1 q_1(ij) - (ij) = \mathfrak{R}(p_{i,j}) - ij$. By Lemma 4.1, there is some $u_{i,j} \in C_2(G_n^{t_2})$ such that ${}^g\partial_2 u_{i,j} = \mathfrak{R}(p_{i,j}) - ij$. Define $P : C_1(G_n \cup G_n^{t_2}) \to C_2(G_n^{t_2})$ by $ij \mapsto u_{i,j}$. Then, by construction, we see $j_1 q_1 - \mathrm{id} = {}^g\partial_2 P$. Hence, j and q are mutually inverse on homology in degree 1.

To conclude, we have shown for each n,

$$\frac{1}{2} l_1 \leq \mathcal{D}(\mathfrak{R}(c^{(n)})) \leq \max(h_1^{(n)}, h_2^{(n)}). \tag{6.3}$$

Taking the limit $n \to \infty$ finishes the proof. □

Note that description of Proposition 6.1 is not unique to this descriptor, indeed the same result holds for the standard pipeline. As discussed in Sect. 3.1, for the standard pipeline, as the weighted digraph is subdivided, the birth times of all features tend to 0. When the digraph is sufficiently subdivided, all edges enter the filtration very early on and the effect of adding the edges from G at $t = 0$ has negligible effect. Hence, in the subdivision limit, the diagrams obtained from the two pipelines coincide.

Theorem 6.2 *Given $G \in \mathbf{WDgr}$, let $\mathcal{H}_1$ denote the 'standard pipeline' with $C = \Omega$ and $F = F_d$, as used in Example 3.6. Then*

$$\lim_{n\to\infty} \mathcal{B}^g\mathcal{H}_1(\mathrm{IMS}_n(G)) = \lim_{n\to\infty} \mathcal{B}\mathcal{H}_1(\mathrm{IMS}_n(G)). \tag{6.4}$$

Proof For brevity, we denote $G_n := \mathrm{IMS}_n(G)$. Fix $\epsilon > 0$ and choose N sufficiently large that for any $n \geq N$ we have $w(e) < \epsilon$ for all $e \in E(G_n)$. For any $t \geq 0$, define

$$\begin{aligned}
i_1 &: C_1(G_n^t) \to C_1(G_n \cup G_n^{t+\epsilon}) \\
i_2 &: C_2(G_n^t) \to C_2(G_n^{t+\epsilon})
\end{aligned}$$

where each i_k is taken from the chain map induced by the relevant inclusion of digraphs. It can be easily checked that $i_1 \partial_2 = {}^g\partial_2 i_2$ and hence i_1 induces a map on homology

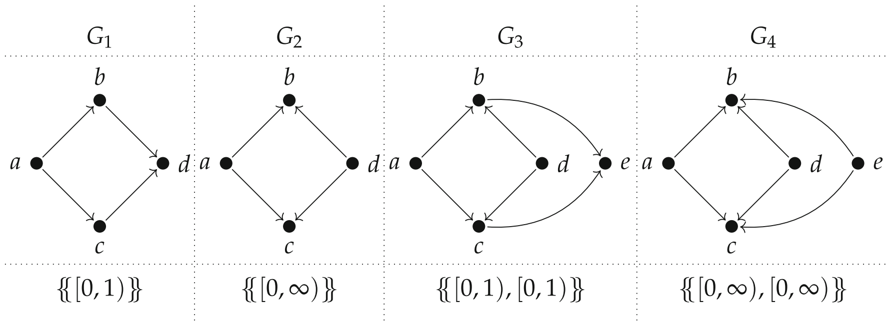

Fig. 21 Interpreting ${}^g\mathcal{H}_1(G)$ via differences between directed paths in square motifs. Top row: name of weighted digraph; middle row: diagram where all weights are 1; bottom row: barcode of GrPPH

$i_* : H_1(G_n^t) \to {}^g H_1(G_n, t+\epsilon)$. Similarly, for any $t \geq 0$ define

$$\begin{aligned} j_1 &: C_1(G_n \cup G_n^t) \to C_1(G_n^{t+\epsilon}) \\ j_2 &: C_2(G_n^t) \to C_2(G_n^{t+\epsilon}) \end{aligned}$$

where each j_k is likewise taken from the chain map induced by the relevant inclusion of digraphs. Note, in particular, given an edge $e \in E(G_n)$, we know $d(\mathrm{st}(e), \mathrm{fn}(e)) < \epsilon$ and hence $e \in E(G_n^{t+\epsilon})$. Again $j_1 {}^g\partial_2 = \partial_2 j_2$ and hence j_1 induces a map on homology $j_* : {}^g H_1(G,t) \to H_1(G^{t+\epsilon})$.

Clearly $i_* \circ j_* = {}^g\iota(t, t+2\epsilon)_*$ and $j_* \circ i_* = \iota(t, t+2\epsilon)_*$. Therefore, by the algebraic stability theorem, we see

$$d_B\left(\mathcal{B}^g\mathcal{H}_1(G_n), \mathcal{B}\mathcal{H}_1(G_n)\right) \leq \epsilon \tag{6.5}$$

for all $n \geq N$. □

6.2 Square Motifs

Example 6.3 Further to the interpretation developed in Proposition 6.1, consider the four weighted digraphs in Fig. 21. All edges are given unit weight and the barcodes are indicated under each digraph. Homology representatives for each of the features are given by

$$\begin{aligned} &ab + bd - cd - ac; \\ &ab - db + dc - ac; \\ &ab + be - ce - ac \quad , \quad db + be - ce - dc; \\ &ab - eb + ec - ac \quad , \quad db - eb + ec - dc. \end{aligned}$$

In G_1, note that the flow starting at a recombines at d after flowing for $t = 2$ seconds. In contrast, the flow in G_2 splits from the sources and then never recombines. This is reflected in the lifetime of the feature changing from [0, 1) to [0, ∞).

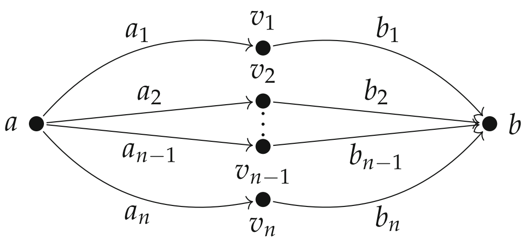

Fig. 22 A weighted digraph with many paths (of 2 edges each) from source to sink for which we can compute $\mathcal{B}^g\mathcal{H}_1(G)$

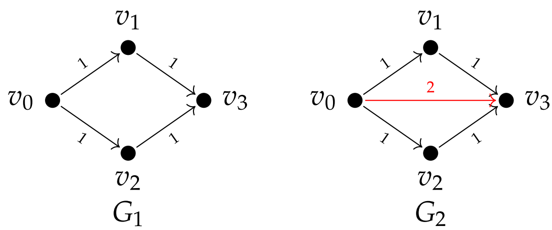

Fig. 23 Two weighted digraphs with identical shortest-path quasimetric network but differing GRPPH which can be explained by the difference in circuit rank

If we add additional edges to G_2 to recombine the flow at a new vertex (as in G_3), we add an additional feature but all features now have finite lifetime. Finally, reversing these additional edges (as in G_4) prevents the flow from recombining again and the features return to lifetime $[0, \infty)$.

This further emphasises the interpretation that features arise when flow is split between two paths and the lifetime of the feature is related to the time it takes for the flow to recombine.

6.3 Multiple Paths

Example 6.4 Consider Fig. 22, in which there is a single source and a single sink but multiple paths between. Define $\alpha_i := \max(a_i, b_i)$ and assume that $\alpha_1 \leq \alpha_2 \leq \dots \alpha_{n-1} \leq \alpha_n$. Then, the barcode is

$$\mathcal{B}^g\mathcal{H}_1(G) = \{[0, \alpha_2), [0, \alpha_3), \dots, [0, \alpha_{n-1}), [0, \alpha_n)\}\,. \tag{6.6}$$

A persistence basis for ${}^g\mathcal{H}_1(G)$ is $\{c_2, \dots, c_n\}$ where $c_i := av_1 + v_1b - v_ib - av_i$ and $\mathcal{D}(c_i) = \alpha_i$.

6.4 Digraphs with Identical Quasimetrics

Example 6.5 Finally, consider the two weighted digraphs illustrated in Fig. 23. Since they both have the same shortest-path quasimetric, they yield the same barcode under the standard pipeline. More formally, $F_d(G_1) = F_d(G_2)$ and hence $\mathcal{H}_1(G_1) = \mathcal{H}_1(G_2)$. Moreover, $\mathcal{B}\mathcal{H}_1(G_1)$ is empty because the circuit (v_0, v_1, v_3, v_2) is filled-in with a long square as soon as it appears in the filtration. In contrast,

$$\mathcal{B}^g\mathcal{H}_1(G_1) = \{[0, 1)\} \quad \text{and} \quad \mathcal{B}^g\mathcal{H}_1(G_2) = \{[0, 1), [0, 2)\}\,. \tag{6.7}$$

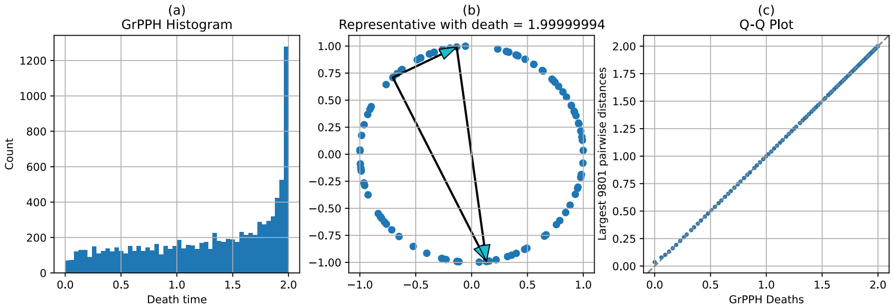

Fig. 24 GrPPH of 100 points sampled on the unit circle. (a) A histogram of the death times of each of the features in ${}^{g}\mathcal{H}_1$. (b) Quantile-quantile plot of of GrPPH deaths against the distribution of the 9801 largest pairwise distances. (c) The point sample along with a directed triangle which is a representative of the largest feature, with death time almost 2

A persistence basis for ${}^{g}\mathcal{H}_1(G_1)$ is $\{v_0v_1 + v_1v_3 - (v_0v_2 + v_2v_3)\}$ while a persistence basis for ${}^{g}\mathcal{H}_1(G_2)$ is $\{v_0v_1 + v_1v_3 - (v_0v_2 + v_2v_3) \quad , \quad v_0v_1 + v_1v_3 - v_0v_3.\}$ Note that at $t = 1$ the two triangular cycles becomes homologous in ${}^{g}\mathcal{H}_1(G_2)$ but are still non-trivial, until they die at $t = 2$.

6.5 Finite (quasi)metric Space

Given a finite quasimetric space $d : X \times X \to [0, \infty)$, one can construct a complete digraph, weighted by d as $MG(X) := (X, X \times X \backslash \Delta_X, w)$ where

$$w((v_i, v_j)) := d(v_i, v_j) \quad \text{for all } (v_i, v_j) \in X \times X \setminus \Delta_X. \tag{6.8}$$

Naturally one can ask what $\mathcal{B}{}^{g}\mathcal{H}_1(MG(X))$ measures, and how it compares to persistent path homology.

Firstly, since the input digraph has $n(n-1)$ edges, n nodes and 1 weakly connected component, the resulting barcode contains

$$\#\text{edges} - \#\text{vertices} + \#\text{weak components} = n(n-1) - n + 1 = (n-1)^2 \tag{6.9}$$

features. Also given any three distinct vertices $v_0, v_1, v_2 \in X$, all of the edges (v_0, v_1), (v_1, v_2) and (v_0, v_2) are present in G and thus form a circuit, c, with death time

$$\mathcal{D}(\mathfrak{R}(c)) = \max(d(v_0, v_1), d(v_1, v_2), d(v_0, v_2)). \tag{6.10}$$

Indeed, any cycle c' supported on an edge (v_i, v_j) must have death time $\mathcal{D}(\mathfrak{R}(c')) \geq d(v_i, v_j)$.

Defining, the diameter of X via $\operatorname{diam}(X) := \max_{v_i, v_j \in X} d(v_i, v_j)$, we note that $MG(X)^t$ is the complete graph for $t \geq \operatorname{diam}(X)$. Since any complete graph has trivial path homology [23, Example 3.11], all features must have death time at most

$\mathrm{diam}(X)$. Therefore, the largest feature in $\mathcal{B}^g\mathcal{H}_1(MG(X))$ must have death time equal to $\mathrm{diam}(X)$.

In Fig. 24, we have sampled 100 points on the unit circle $X \subseteq S^1$ and computed $\mathcal{B}^g\mathcal{H}_1(MG(X))$, using the ambient Euclidean metric. We see a large quantity of features with death time of approximately 2, the diameter of the circle. A representative of the largest feature is shown in the second panel, which is indeed a directed triangle supported on an edge which spans almost the full diameter of the circle.

The distribution of death times is well-aligned with the distribution of the top $(100-1)^2 = 9801$ pairwise distances, as illustrated by the quantile-quantile plot in Fig. 24. The main deviation between the two distribution occurs at the small scales whilst the larger quantiles are almost identical. This illustrates that GrPPH has similar descriptive power to the collection of pairwise distances. That is, not much information is lost, but consequently the output is not very interpretable and provides little additional insight.

In contrast, computing standard PPH, the degree 1-barcode contains a single feature with lifetime approximately $[0.291, 1.418)$. The birth time corresponds to the length of the largest gap in the point sample (at the top of the circle). Meanwhile, the death time is approximately $\sqrt{2}$, the side length of a square inscribing the circle.

Grounded persistence was designed as a descriptor of sparse weighted digraphs where the existence of an edge is an important signal from the input data and the circuits are of interest. We enrich the circuits already present in G with an intrinsic, directed notion of scale, using the weighting and an appropriate filtration. In the finite (quasi)metric space scenario, all possible edges are present in $MG(d)$ and thus the circuit representatives are not as interpretable as in the sparse case. As such, our recommendation would be to use non-grounded persistent path homology [13] (viewing the data as a directed network) or traditional, symmetric TDA methods.

Acknowledgements The first author would like to thank H. Byrne, A. Goriely, A. Ó hEachteirn and T. Thompson for valuable discussions which motivated and aided the early stages of this work. The authors are thankful for the detailed and helpful comments of the reviewers of this manuscript. In particular, the definition of a δ-grounded interleaving in Sect. 5.1 was greatly motivated by a suggestion to adapt the Kalton-Ostrovskii characterisation of network distance. HAH gratefully acknowledges funding from a Royal Society University Research Fellowship. The authors are members of the Centre for Topological Data Analysis, which is funded by the EPSRC grant 'New Approaches to Data Science: Application Driven Topological Data Analysis' EP/R018472/1. For the purpose of Open Access, the authors have applied a CC BY public copyright licence to any Author Accepted Manuscript (AAM) version arising from this submission.

References

1. U. Bauer and M. Lesnick, Induced matchings and the algebraic stability of persistence barcodes, Journal of Computational Geometry **6**(2) (2015), 162–191. https://doi.org/10.20382/jocg.v6i2a9
2. K. Benjamin, L. Mukta, G. Moryoussef, C. Uren, H.A. Harrington, U. Tillmann, A. Barbensi, Homology of homologous knotted proteins, Journal of the Royal Society Interface **20**(201) (2023), 20220727. https://doi.org/10.1098/rsif.2022.0727
3. A. Bittner, B.T. Fasy, M. Grudzien, S. Ghosh Hajra, J. Huang, K. Pelatt, C. Thatcher, A. Tumurbaatar and C. Wenk, Comparing directed and weighted road maps, in Research in Computational Topology (E.W. Chambers, B.T. Fasy, and L. Ziegelmeier, eds.), Springer, Cham, 2018, pp. 57–70. https://doi.org/10.1007/978-3-319-89593-2_4
4. E.L. Brown, T.L. Lefebvre, P.W. Sweeney, B.J. Stolz, J. Gröhl, L. Hacker, Z. Huang, D.L. Couturier, H.A. Harrington, H.M. Byrne and S.E. Bohndiek, Quantification of vascular networks in photoacoustic mesoscopy, Photoacoustics **26** (2022), 100357. https://doi.org/10.1016/j.pacs.2022.100357
5. P. Bubenik and J.A. Scott, Categorification of persistent homology, Discrete & Computational Geometry **51**(3) (2014), 600–627. https://doi.org/10.1007/s00454-014-9573-x
6. L. Caputi, A. Pidnebesna and J. Hlinka, Promises and pitfalls of topological data analysis for brain connectivity analysis, NeuroImage **238** (2021), 118245. https://doi.org/10.1016/j.neuroimage.2021.118245
7. L. Caputi and H. Riihimäki, Hochschild homology, and a persistent approach via connectivity digraphs, Journal of Applied and Computational Topology (2023). https://doi.org/10.1007/s41468-023-00118-9
8. T. Chaplin, GrPPHATI. https://github.com/tomchaplin/grpphati
9. F. Chazal and B. Michel, An introduction to topological data analysis: Fundamental and practical aspects for data scientists, Frontiers in Artificial Intelligence **4** (2021). https://doi.org/10.3389/frai.2021.667963
10. F. Chazal, V. de Silva, M. Glisse and S. Oudot, The Structure and Stability of Persistence Modules, Springer International Publishing, Cham, 2016. https://doi.org/10.1007/978-3-319-42545-0_2
11. M. Che, F. Galaz-García, L. Guijarro, I. Membrillo Solis and M. Valiunas, Basic metric geometry of the bottleneck distance, Proceedings of the American Mathematical Society **152** (2024). https://doi.org/10.1090/proc/16776
12. S. Chowdhury and F. Mémoli, A functorial dowker theorem and persistent homology of asymmetric networks, Journal of Applied and Computational Topology **2**(1) (2018), 115–175. https://doi.org/10.1007/s41468-018-0020-6
13. S. Chowdhury and F. Mémoli, Persistent path homology of directed networks, in Proceedings of the 2018 Annual ACM-SIAM Symposium on Discrete Algorithms (SODA) (A. Czumaj, eds.), Society for Industrial and Applied Mathematics, Philadelphia, PA, 2018, pp. 1152–1169. https://doi.org/10.1137/1.9781611975031.75
14. D. Cohen-Steiner, H. Edelsbrunner and J. Harer, Stability of persistence diagrams, in SCG ’05: Proceedings of the twenty-first annual symposium on Computational geometry (J. Mitchell, and G.Rote, eds.), Association for Computing Machinery, New York, NY, USA, 2005, pp. 263–271. https://doi.org/10.1145/1064092.1064133
15. D. Cohen-Steiner, H. Edelsbrunner, J. Harer and Y. Mileyko, Lipschitz functions have Lp-stable persistence, Foundations of Computational Mathematics **10**(2) (2010), 127–139. https://doi.org/10.1007/s10208-010-9060-6
16. W. Crawley-Boevey, Decomposition of pointwise finite-dimensional persistence modules, Journal of Algebra and Its Applications **14**(05) (2015), 1550066. https://doi.org/10.1142/S0219498815500668
17. T.K. Dey, T. Li and Y. Wang, An Efficient Algorithm for 1-Dimensional (Persistent) Path Homology, in 36th International Symposium on Computational Geometry (SoCG 2020) (S. Cabello, and D.Z. Chen, eds.), Dagstuhl–Leibniz-Zentrum für Informatik, Dagstuhl, Germany, 2020, pp. 36:1-36:15. https://doi.org/10.4230/LIPIcs.SoCG.2020.36
18. H. Edelsbrunner and J. Harer, Persistent homology-a survey, in Surveys on Discrete and Computational Geometry: Twenty Years Later (J.E. Goodman, J. Pach, and R. Pollack, eds.), American Mathematical Society, Providence, RI, 2008, pp. 257–282. https://doi.org/10.1090/conm/453
19. X. Gao, B. Xiao, D. Tao and X. Li, A survey of graph edit distance, Pattern Analysis and Applications **13**(1) (2010), 113–129. https://doi.org/10.1007/s10044-008-0141-y
20. B. Giunti, TDA-Applications. https://www.zotero.org/groups/2425412/tda-applications.

21. C. Goodbrake, D. Beers, T.B. Thompson, H.A. Harrington and A. Goriely, Brain chains as topological signatures for alzheimer's disease, Journal of Applied and Computational Topology (2024). https://doi.org/10.1007/s41468-024-00168-7
22. D. Govc, R. Levi and J.P. Smith, Complexes of tournaments, directionality filtrations and persistent homology, Journal of Applied and Computational Topology **5**(2) (2021), 313–337. https://doi.org/10.1007/s41468-021-00068-0
23. A. Grigor'yan, Y. Lin, Y. Muranov and S.T. Yau, Homotopy theory for digraphs, Pure and Applied Mathematics Quarterly **10**(4) (2014), 619–674. https://doi.org/10.4310/PAMQ.2014.v10.n4.a2
24. A. Grigor'yan, Y. Lin, Y. Muranov and S.T. Yau, Homologies of path complexes and digraphs (2012). arXiv:1207.2834
25. A.A. Grigor'yan, Y. Lin, Y.V. Muranov and S.T. Yau, Path complexes and their homologies, Journal of Mathematical Sciences **248**(5) (2020), 564–599. https://doi.org/10.1007/s10958-020-04897-9
26. M.S. Ismail, M.S.M. Noorani, M. Ismail, F.A. Razak and M.A. Alias, Early warning signals of financial crises using persistent homology, Physica A: Statistical Mechanics and its Applications **586** (2022), 126459. https://doi.org/10.1016/j.physa.2021.126459
27. L. Li, C. Thompson, G. Henselman-Petrusek, C. Giusti and L. Ziegelmeier, Minimal cycle representatives in persistent homology using linear programming: an empirical study with user's guide, Frontiers in artificial intelligence **4** (2021), 681117. https://doi.org/10.3389/frai.2021.681117
28. Y. Lin, S. Ren, C. Wang and J. Wu, Weighted path homology of weighted digraphs and persistence (2019). arXiv:1910.09891
29. D. Lütgehetmann, D. Govc, J.P. Smith and R. Levi, Computing persistent homology of directed flag complexes, Algorithms **13**(1) (2020). https://doi.org/10.3390/a13010019
30. J.D. Medaglia, Graph theoretic analysis of resting state functional mr imaging, Neuroimaging clinics of North America **27** (2017), 593–607. https://doi.org/10.1016/j.nic.2017.06.008
31. V. Nanda, Computational algebraic topology — lecture notes (2021). https://people.maths.ox.ac.uk/nanda/cat/TDANotes.pdf
32. J.T. Nardini, B.J. Stolz, K.B. Flores, H.A. Harrington and H.M. Byrne, Topological data analysis distinguishes parameter regimes in the anderson-chaplain model of angiogenesis, PLOS Computational Biology **17**(6) (2021), 1–29. https://doi.org/10.1371/journal.pcbi.1009094
33. N. Otter, M.A. Porter, U. Tillmann, P. Grindrod and H.A. Harrington, A roadmap for the computation of persistent homology, EPJ data science **6** (2017), 1–38. https://doi.org/10.1140/epjds/s13688-017-0109-5
34. A.E. Sizemore, C. Giusti, A. Kahn, J.M. Vettel, R.F. Betzel and D.S. Bassett, Cliques and cavities in the human connectome, Journal of Computational Neuroscience **44**(1) (2018), 115–145. https://doi.org/10.1007/s10827-017-0672-6
35. B.J. Stolz, J. Kaeppler, B. Markelc, F. Braun, F. Lipsmeier, R.J. Muschel, H.M. Byrne, H.A. Harrington, Multiscale topology characterizes dynamic tumor vascular networks, Science Advances **8**(23) (2022), eabm2456. https://doi.org/10.1126/sciadv.abm2456
36. P.W. Sweeney, A. d'Esposito, S. Walker-Samuel and R.J. Shipley, Modelling the transport of fluid through heterogeneous, whole tumours in silico, PLOS Computational Biology **15**(6) (2019), e1006751. https://doi.org/10.1371/journal.pcbi.1006751
37. J. Xing, J. Jia, X. Wu and L. Kuang, A spatiotemporal brain network analysis of alzheimer's disease based on persistent homology, Frontiers in aging neuroscience **14** (2022), 788571. https://doi.org/10.3389/fnagi.2022.788571