Homology Theories for Digraphs: Statistics, Applications and Isomorphisms



Thomas Chaplin
New College
University of Oxford

A thesis submitted for the degree of
Doctor of Philosophy

Michaelmas 2024

Acknowledgements

The work in this thesis would not have been possible without the support of my two academic supervisors, Prof. Heather A. Harrington and Prof. Ulrike Tillmann. Arriving in Oxford with little to no research experience, I was uneasy about what the next years would hold, but their steady guidance helped me to discover my interests, improve my work and left me with a life-long passion for research. Most of all, the talented, passionate and friendly community that they, along with countless others, have fostered in Oxford and beyond has been the most invaluable resource.

Through their foundation of support, my family gave me the freedom and excitement to pursue both mathematics and my doctoral studies. They kept my spirts high through dark times and celebrated in my successes. To my parents Mark and Patricia, to my sister Naomi and brother-in-law Aidan, and to my entire extended family, I am eternally grateful. To Eliza and George, your joy for learning never fails to inspire me.

Finally, I owe the greatest debt of gratitude to my partner, Nikki. There were many moments when, without her unwavering kindness and support, this thesis would not have come to pass. Amongst numerous other lessons, she taught me the value in hard work, self-care and good food.

Abstract

The persistent homology of a point cloud or finite metric space summarises the pertinent topological features in the data, as they appear and disappear when we view the data at a range of scales. Recently, there has been interest in developing persistence pipelines for directional data, such as finite quasimetric spaces or weighted digraphs, in a way that is sensitive to the data's intrinsic asymmetry. A key ingredient in developing such a pipeline is a homology theory for digraphs. In this thesis we study two such theories: directed flag complex homology and path homology. The primary aim of this thesis is to further develop our mathematical and statistical understanding of these two homology theories, so that we may more appropriately apply them to data.

First, we seek to understand the statistical behaviour of these theories in an Erdős-Rényi random digraph model. We describe both low-density and high-density regimes in which one should expect trivial degree-1 path homology with high probability. In contrast, we show there is a medium-density regime in which one should expect non-trivial degree-1 path homology. We repeat this analysis for the directed flag complex and find similar results.

Secondly, we develop a persistence pipeline which is more directly applicable to data which takes the form of a weighted digraph. We alter persistent path homology to ensure that the number of features in the barcode coincides with the circuit rank of the digraph, and that representatives correspond to circuits in the original digraph. Using path homotopy theory, we show that the resulting pipeline is stable to a range of numerical and structural perturbations to the input weighted digraph.

Finally, through comparison with path homotopy theory, we develop a homotopy-like equivalence relation for digraph maps such that equivalent maps induce the same map on directed flag complex homology. This allows us to identify conditions under which two digraphs have isomorphic directed

flag complex homology. Moreover, this equivalence allows us to obtain stability theorems for the persistent directed flag complex homology of a filtration of digraphs.

Contents

1 Introduction **1**

1.1 Structure of this thesis 4

1.1.1 Chapter 2 4

1.1.2 Chapter 3 5

1.1.3 Chapter 4 6

1.2 Other contributions 7

2 First Betti number of the path homology of random directed graphs **9**

2.1 Introduction 9

2.1.1 Summary of results 11

2.1.2 Data Availability 14

2.2 Background 14

2.2.1 Graph theory definitions and assumptions 14

2.2.2 Analytic and algebraic definitions 15

2.2.3 Erdős-Rényi random graphs 16

2.2.4 Symmetrisation 16

2.3 Path homology of directed graphs 19

2.3.1 Definition 19

2.3.2 Proof of Theorem 2.1.4 22

2.3.3 Chain group generators 23

2.4 Asymptotic results for path homology 26

2.4.1 Proof of Theorem 2.1.3(1) 27

2.4.2 Proof of Theorem 2.1.3(2) 29

2.4.3 Proof of Theorem 2.1.3(3) 29

2.4.4 Proof of Theorem 2.1.3(4) 30

2.5 Directed flag complex of random directed graphs 38

2.6 Discussion 41

2.A Explicit probability bounds . . . 42
2.A.1 Positive Betti numbers at low densities . . . 42
2.A.2 Zero Betti numbers . . . 43
2.A.3 Positive Betti numbers at high densities . . . 47
2.B Experimental results . . . 51
2.B.1 Data collection . . . 51
2.B.2 Illustrations . . . 51
2.B.3 Finding boundaries . . . 53
2.B.4 Testing for normal distribution . . . 56

3 Grounded persistent path homology: a stable, topological descriptor for weighted digraphs 58
3.1 Introduction . . . 58
3.1.1 Contributions and outline . . . 60
3.1.2 Computations . . . 62
3.2 Background . . . 62
3.2.1 Basic notation and category theory . . . 62
3.2.2 Directed graphs . . . 63
3.2.3 Path homology . . . 67
3.2.4 Path homotopy . . . 70
3.2.5 Persistent homology . . . 73
3.3 Motivation and definition of GrPPH . . . 76
3.3.1 Standard pipeline . . . 76
3.3.2 Grounded pipeline . . . 79
3.3.3 Definition of GRPPH . . . 82
3.4 Interpretation of GrPPH . . . 84
3.4.1 Decreasing Betti curves . . . 84
3.4.2 Circuit lifetimes . . . 85
3.4.3 Representatives . . . 87
3.4.4 Decomposition . . . 94
3.5 Stability analysis of GrPPH . . . 96
3.5.1 Main stability theorem . . . 98
3.5.2 Weight perturbation . . . 101
3.5.3 Edge subdivision . . . 102
3.5.4 Edge collapse . . . 107

3.5.5 Edge deletion . . . 113
3.5.6 Vertex deletion . . . 116
3.5.7 Combining operations . . . 119
3.6 Examples . . . 122
3.6.1 Iterated medial subdivision . . . 122
3.6.2 Square motifs . . . 124
3.6.3 Multiple paths . . . 125
3.6.4 Digraphs with identical quasimetrics . . . 126
3.6.5 Finite (quasi)metric space . . . 126

4 A notion of homotopy for directed graphs and their flag complexes 129
4.1 Introduction . . . 129
4.1.1 Summary . . . 130
4.1.2 Related work . . . 132
4.1.3 Notation . . . 133
4.2 Combinatorial and algebraic complexes . . . 133
4.2.1 Path complexes . . . 133
4.2.2 Regular path complexes from digraphs . . . 136
4.2.3 Chain complexes from regular path complexes . . . 139
4.3 Systems of one-step homotopies . . . 142
4.4 Homotopies for ordered simplicial complexes . . . 146
4.4.1 Homotopies for path complexes . . . 146
4.4.2 Ordered simplicial complexes as path complexes . . . 149
4.4.3 Ordered simplicial complexes as simplicial sets . . . 153
4.5 Homotopies for digraphs . . . 157
4.5.1 Pull-back along the allowed path functor . . . 158
4.5.2 Pull-back along the directed flag complex functor . . . 160
4.5.3 Comparing the two pull-backs . . . 164
4.6 Stability of persistent directed flag complex homology . . . 170
4.6.1 The interleaving distance . . . 171
4.6.2 Stability theorems . . . 172
4.6.3 Instabilities . . . 177
4.A Mapping cylinders . . . 178
4.B Grounded pipelines . . . 182
4.B.1 Background . . . 182

4.B.2 Stability theorems 183

Bibliography **186**

List of Figures

2.1 10
2.2 24
2.3 25
2.4 31
2.5 32
2.6 46
2.7 47
2.8 50
2.9 52
2.10 52
2.11 54
2.12 54
2.13 56
2.14 57
3.1 58
3.2 70
3.3 78
3.4 83
3.5 86
3.6 88
3.7 92
3.8 94
3.9 97
3.10 104
3.11 105
3.12 106
3.13 108
3.14 110
3.15 112
3.16 115
3.17 116
3.18 118
3.19 121
3.20 122
3.21 125
3.22 125
3.23 126
3.24 127
4.1 138
4.2 140
4.3 147
4.4 160
4.5 162
4.6 165
4.7 166
4.8 167
4.9 174
4.10 177
4.11 178

List of Tables

2.1 41
2.2 51
3.1 98
4.1 185

Chapter 1

Introduction

A fundamental challenge in data analysis is how to combine insights into phenomena that occur at different scales but are nevertheless connected. As an extreme example, the behaviour of a whole organism *is* the collective behaviour of its cells, but integrating new understanding of subcellular mechanics into a model of the entire organism is a formidable task. Topological data analysis (TDA) proposes an approach to this challenge: functorial statistics.

At a high level, any statistic can be viewed as a function

$$S : \boldsymbol{\mathcal{D}} \times \boldsymbol{\mathcal{P}} \to \boldsymbol{\mathcal{C}}, \tag{1.1}$$

taking as input data from some domain $\boldsymbol{\mathcal{D}}$, depending on parameters from some space $\boldsymbol{\mathcal{P}}$, and producing a summary in some space $\boldsymbol{\mathcal{C}}$. If $\boldsymbol{\mathcal{D}} \times \boldsymbol{\mathcal{P}}$ is given the structure of a discrete category then any categorical structure on $\boldsymbol{\mathcal{C}}$ makes S into a functor. However, the utility and interpretability of S increases if we can maintain functoriality whilst adding more morphisms to the categorical structure on $\boldsymbol{\mathcal{D}} \times \boldsymbol{\mathcal{P}}$ and restricting $\boldsymbol{\mathcal{C}}$ to a category of well-understood, informative morphisms. This increased functoriality is helpful because it allows us to characterise and control how the output of S changes in response to a broader class of changes in both the data and the parameters.

The prototypical example from TDA is the offset filtration. This function takes a finite subset $X \subseteq \mathbb{R}^n$ (known as a point cloud) and a radius parameter $r \geq 0$, and constructs the union of balls of radius r around each point in X, i.e.

$$\mathcal{O} : \textbf{PointClouds} \times [0, \infty] \to \textbf{Top}, \quad (X, r) \mapsto \bigcup_{x \in X} B(x, r). \tag{1.2}$$

This construction is functorial in the sense that if $Y \subseteq X$ and $r \leq s$ then $\mathcal{O}(Y, r) \subseteq \mathcal{O}(X, s)$.

Whilst the resulting topological space $\mathcal{O}(X, r)$ is an elegant model for the data X viewed at some particular scale r, it is still too complicated to analyse directly. As such, it is prudent

to further summarise this space by a topological invariant — most typically, homology. Composing with the homology functor H and fixing the point cloud X, we obtain a functor

$$\mathrm{PH}(X) : [0,\infty] \to \mathbf{grVec}, \quad r \mapsto (H \circ \mathcal{O})(X, r), \tag{1.3}$$

from the category on $[0,\infty]$ generated by the $\leq$-relation, to the category of graded vector spaces. This functor is commonly referred to as the persistent homology of X.

At any particular scale r, the invariant $\mathrm{PH}(X)(r)$ is already a useful summary of X. The 0-dimensional component captures the clusters in the data whilst the 1-dimensional component detects looping structures. However, the true power of $\mathrm{PH}(X)$ is that it also captures how these homological features relate to each other as we vary our choice of radius parameter. The linear map $\mathrm{PH}(X)(r \leq s)$ summarises how homology appears and disappears as we change our viewpoint to a larger scale. It is precisely this additional structure that allows us, via the structure theorem [36], to canonically decompose the isomorphism type of $\mathrm{PH}(X)$ into its barcode. The barcode consists of a finite multiset of half-open intervals $\{\!\{ [b_i, d_i) \}\!\}_{i=1}^{m}$ with each interval representing a homological feature in the data that first appears when viewing the data at scale b_i, that can be *consistently* identified at scales $r \in [b_i, d_i)$ via the map $\mathrm{PH}(X)(b_i \leq r)$, and that finally disappears at scale d_i.

The primary advantage of the barcode is that it segments the rich, multi-scale information captured by the functor $\mathrm{PH}(X)$ in a principled way, thus achieving the goal of reconciling homological phenomena that appear across a range of scales. As such, it has seen success across a wide range of application domains (see [40] for a database of examples). Furthermore, the offset filtration also exhibits functoriality in its data category that we have yet to fully exploit. Namely, given any subset of the original point cloud, $Y \subseteq X$, there is an induced map on persistent homology, i.e. a natural transformation $\mathrm{PH}(Y) \Rightarrow \mathrm{PH}(X)$. These maps allow us to interrogate how subsets of the data contribute to the persistent homology of the whole. Moreover, these maps admit summaries that are similar to the barcode [35], and recent innovations in computing these statistics [37, 82] promise to provide fresh insights, particularly into biological data.

However, not all data can be adequately modelled as a finite point cloud in $\mathbb{R}^n$. Consider, instead, the flow of blood through a network of vessels, the synaptic connections between neurons in the brain or the passing of messages through a social network. At the core of each of these systems is directionality. Blood flows through vessels in a particular direction, chemical synapses consist of a presynaptic and a postsynaptic terminal, and authorities in

social networks can be characterised by the asymmetry in their interactions. Accordingly, these systems are better modelled by mathematical objects which are also directional, such as finite *quasimetric* spaces or weighted digraphs. In order to obtain insight into these systems, we require functorial statistics of directional data types that admit interpretable summaries and are sensitive to the intrinsic asymmetry in the data.

If we wish to follow a similar approach, leveraging the structure theorem to obtain a barcode, we require two ingredients:

1. a filtration of the data, i.e. a map $F : \boldsymbol{\mathcal{D}} \times [0, \infty] \to \boldsymbol{\mathcal{C}}$, where $\boldsymbol{\mathcal{D}}$ is the domain of the data and $\boldsymbol{\mathcal{C}}$ is some target category, such that F is functorial in the second parameter;
2. a homology functor $H : \boldsymbol{\mathcal{C}} \to \mathbf{grVec}$ for the target category of the filtration.

There are numerous choices one could make for F, $\boldsymbol{\mathcal{C}}$ and H; the correct choice is highly domain specific. A first approach, most frequently used in practice, is to symmetrise your data (for example into a finite metric space or a weighted, undirected graph) and then apply traditional filtrations and homology theories from TDA. When the asymmetry in the data is not critical to the task at hand, this approach often suffices and indeed emphasises this lack of dependence on directionality (e.g. [97]). A second approach is to construct a filtration of non-directional objects, such as simplicial complexes, so that the filtration is sensitive to the asymmetry in the data (e.g. [79]). With this approach, one can leverage the literature and software for pre-existing homology theories. A third approach, to which this thesis primarily attends, is to construct a filtration of simpler directed objects. This approach is potentially advantageous because it maintains the directionality in the data throughout the pipeline up until homology is computed. Arguably the simplest type of directed object one could use for this purpose is the unweighted, directed graph.

In order to take this approach, we therefore require a homology theory for digraphs, i.e. a functor $H : \mathbf{Dgr} \to \mathbf{grVec}$ from some category of digraphs. Despite being intrinsically 1-dimensional, digraphs can contain complicated substructures; in applications, these substructures may be evidence of higher-order interactions in the original data. As such, it is often desirable that such a theory be capable of producing non-trivial homology in higher dimensions. In this thesis, we focus on two theories that achieve this aim: directed flag complex homology and path homology.

We delay the precise definition of these homology theories to the later chapters. However, it is worth justifying the use of the term 'homology theory' for each of these functors. First, both functors can be written as a composition, $H \circ C$, where C constructs a chain complex from relevant substructures in the digraph and then H is the usual homology functor

for chain complexes. The directed flag complex is in fact an ordered simplicial complex and thus is analogous to the more classical clique complex of an undirected graph. Path homology comes in two variants: regular path homology and non-regular path homology. For each of these variants the chain complex C is constructed directly from paths in the digraph, rather than via a simplicial complex.

In general, the two variants of path homology are not isomorphic. As a minimal example, the 1-dimensional path homology (over $\mathbb{R}$) of a complete digraph on 2 nodes is trivial for the regular variant but is isomorphic to $\mathbb{R}$ for the non-regular variant. The non-regular variant is slightly simpler to define, but the regular variant enjoys much stronger theoretical properties, including invariance with respect to a homotopy theory for digraphs [55], analogues of the Eilenberg–Steenrod axioms [56] and a relationship to magnitude homology [3]. Accordingly, the regular variant is more frequently studied in the literature and is often referred to simply as 'path homology'. Where possible, we attempt to treat both variants in this thesis. However, due to these desirable properties, in Chapter 3 we quickly fall back to regular variant and Chapter 4 exclusively discusses the regular variant.

In this thesis, we seek to develop our understanding of both directed flag complex homology and path homology, with a particular focus on theoretical advancements that might aid in the application of these theories to data.

1.1 Structure of this thesis

Each of the subsequent chapters of this thesis are self-contained, containing the relevant definitions, literature review and notation. Below we provide brief overviews of each chapter.

1.1.1 Chapter 2

For our first study of these two homology theories, we set aside persistence and ask a simpler, statistical question: when is it unusual to observe a directed graph with trivial, or non-trivial, homology? In order to answer this question, one requires a null model of digraphs; as a starting point, we take the Erdős-Rényi model. This random graph model depends on two parameters: $n \in \mathbb{N}$, the number of nodes, and $p \in [0, 1]$, the probability with which each edge appears.

This question has already been well-treated for the clique complex of an undirected graph, wherein it is known that homology is typically non-trivial in precisely one dimension (see [66] for a summary). Returning to directed graphs, we find empirical evidence for three regions of (n, p)-parameter space: a low-density region, a high-density region and a goldilocks zone. Focusing on degree-1 path homology, in the first two regions one typically observes trivial homology, whilst in the final region, one typically observes non-trivial homology. Moreover, the boundaries in parameter space between these regions appear to follow a power law.

To investigate this analytically, we traverse power-law trajectories through (n, p)-parameter space and show that, depending on the power-law exponent, one observes either trivial or non-trivial homology with probability tending to 1. Our analysis leaves open an interval of exponents, which we conjecture to be in the high-density region of trivial homology. We repeat the study for the directed flag complex and obtain similar results. Finally, we use a computer-aided search to obtain sharper probability bounds, which could be used for hypothesis testing.

Authorship. The work in this chapter was fully undertaken by the author of this thesis, under the supervision of his two supervisors, to whom he is most grateful. The content of this chapter was published in a special issue of the Journal of Applied and Computational Topology [24].

1.1.2 Chapter 3

Returning to persistence, we develop a persistent homology pipeline for data taking the form of a weighted digraph. Persistent path homology (PPH) is a pre-existing methodology that provides a stable summary of a finite quasimetric space [32]. The natural way to apply this pipeline to a weighted digraph would be to first compute its shortest-path quasimetric. However, doing so discards the underlying topology and combinatorics of the digraph. For example, it is possible to construct a sequence of weighted digraphs with identical PPH barcodes but strictly increasing circuit rank.

We present a modification of PPH that is targeted towards weighted digraphs and alleviates these issues, called grounded persistent path homology (GrPPH). In particular, we show that the number of features in the barcode of GrPPH is equal to the circuit rank of the underlying digraph. Moreover, we show that any weighted digraph admits a choice of combinatorial circuits that forms both a circuit basis for the underlying graph and a persistence basis for GrPPH. Although this choice is not unique, these circuits can be used

in practice to interpret the features in the barcode. Finally, we consider how GRPPH varies as the input digraph is modified through operations such as weight perturbation, edge subdivision and edge deletion. We show that, under sufficient constraints, the barcode is stable to these modifications, thus making the methodology practical for real-world applications.

Authorship. The work in this chapter was a collaboration between the author of this thesis and his two supervisors, Prof. Heather A. Harrington and Prof. Ulrike Tillmann. The results were obtained by the author and writing was undertaken collaboratively, but the majority of the writing was the author's. Notably, the introduction was partially written by both collaborators. Particular assistance was provided in deciding on the appropriate operations to consider in the stability section, and in deciding on the level of generality in which to present the main pipeline. The content of this chapter was published in Foundations of Computational Mathematics [26].

1.1.3 Chapter 4

In the final chapter, we turn our attention towards the directed flag complex. In proving the stability results in the previous chapter, we lean heavily on the path homotopy theory of digraphs. This theory provides an equivalence relation of digraph maps such that equivalent digraph maps induce identical maps on path homology. Our goal in this chapter is to develop a similar relation such that equivalent digraph maps induce identical maps on directed flag complex homology. Such a relation will also allow us to identify an equivalence relation for digraphs such that equivalent digraphs have isomorphic directed flag complex homology.

We will derive the relation for digraph maps by viewing the directed flag complex as a 'path complex' — the main object of study in path homology. However, one could equally view the directed flag complex as a simplicial set. This alternative perspective comes equipped with a similar equivalence relation, via left homotopies of simplicial maps. We show that these two viewpoints yield identical 'pull-back' equivalence relations for digraph maps.

Authorship. The work in the final chapter of content is also a collaboration between the author and his two supervisors, Prof. Heather A. Harrington and Prof. Ulrike Tillmann. The results were obtained by the author and writing was undertaken collaboratively, but the majority of the writing was the author's. Collaborators provided assistance in writing the introduction, and in the decision to pay particular attention to the system of one-step

homotopies for ordered simplicial complexes. The content of this chapter is currently available as a pre-print [25] and will be submitted in due course.

1.2 Other contributions

For completeness, we note that during the course of his doctoral studies the author of this thesis also contributed to the pre-print [82], concerning the chromatic alpha filtration. The author also developed numerous open-source software packages, of which we list the main contributions below:

The CAT List — `cat-list.github.io`

Developed alongside fellow DPhil candidate Otto Sumray, this is an online, searchable repository of open-source software for TDA. Individual packages are tagged by features and language, and the database has received numerous contributions from the wider TDA community.

phlite — `github.com/tomchaplin/phlite`

This package aims to bring some of the advances of `Ripser` [6] to a wider class of persistent homology pipelines. In particular, `phlite` provides a framework for implementing lazy oracles into sparse matrices D, with indexes and coefficients of arbitrary type. Under sufficient type constraints, the package provides functions for decomposing these matrices in the form $R = DV$, enabling the computation of persistent (co)homology. This approach avoids constructing the entire (co)boundary matrix and thus significantly reduces memory usage. An implementation of GRPPH was developed in this framework.

lophat — `github.com/tomchaplin/lophat`

The primary contribution of this package is an implementation of the parallel, lock-free persistent homology algorithm, presented by Morozov and Nigmetov [78]. Unlike the original, the `lophat` implementation uses epoch-based memory management. Python bindings are provided, with a similar API to the PHAT package [7].

phimaker — `github.com/tomchaplin/phimaker`

This is a package for computing the barcodes of the images, kernels and cokernels for maps of persistent vector spaces. The package implements the algorithm proposed by Cohen-Steiner et al. in [35] and all matrix decompositions are completed via `lophat`.

This package was used in the `chalc` software package for the chromatic alpha complex, developed by fellow DPhil candidate Abhinav Natarajan (see [82]).

gramag — `github.com/tomchaplin/gramag`

This software package computes the magnitude homology of a directed graph (see [59, 3]). Computation is accelerated by a parallelised depth-first search for generators of the magnitude chain groups, and homology calculations are conducted either via `lophat` or `phlite`.

Chapter 2

First Betti number of the path homology of random directed graphs

2.1 Introduction

In applications, networks often arise with asymmetry and directionality. Chemical synapses in the brain have an intrinsic direction (see [86, §5]); gene regulatory networks record the causal effects between genes (e.g. [1]); communications in social networks have a sender and a recipient (e.g. [70]). A common hypothesis is that the structure of a network determines its function [61, 87], at least in part. In order to investigate such a claim, one requires a topological invariant which describes the structure of the network. To obtain such a summary for a digraph, one often symmetrises to obtain an undirected graph, before applying traditional tools from TDA (e.g. [58]). This potentially inhibits the predictive power of the descriptor, since the pipeline becomes blind to the direction of edges. In recent years, particularly in applications related to neuroscience (e.g. [17, 87]), researchers have explored the use of topological methods which are sensitive to the asymmetry of directed graphs.

A much-studied construction, for undirected graphs, is the clique complex (or flag complex) – a simplicial complex in which the k-simplices are the $(k+1)$-cliques in the underlying graph. An obvious extension to the case of directed graphs is the directed flag complex [74]. This is an *ordered* simplicial complex in which the *ordered* k-simplices are the $(k+1)$-directed cliques: $(k+1)$-tuples of distinct vertices $(v_0, \ldots, v_k)$ such that $v_i \to v_j$ whenever $i < j$. An important property of this construction is that is able to distinguish between directed graphs with identical underlying, undirected graphs; it is sensitive to the asymmetry of the digraph.

Path homology (first introduced by Grigor'yan et al. [54]) provides an alternative

construction which, while more computationally expensive, is capable of distinguishing between digraphs which are indistinguishable to the directed flag complex (e.g. Figure 2.1, which was first shown in [32, Figure 2]). Path homology admits two variants, namely regular and non-regular path homology. The regular and non-regular chain complexes, underlying the respective variants, each contain the directed flag complex as a subcomplex. Intuitively, the generators of the k^{th} chain group of the directed flag complex are all the directed paths, of length k, such that all shortcut edges are present in the graph. Whereas, the k^{th} chain group of the non-regular chain complex consists of all *linear combinations* of directed paths, of length k, such that any missing shortcuts of length $(k-1)$ are cancelled out.



(a) (b)

Figure 2.1: Two motifs which are indistinguishable to the directed flag complex but have different path homology.

Other desirable features of regular path homology include good functorial properties in an appropriate digraph category [55, 52] and invariance under an appropriate notion of path homotopy [55, Theorem 3.3]. Furthermore, path homology is a particularly novel method since it operates directly on directed paths within the digraph, rather than first constructing a simplicial complex. Rather than being freely generated by distinguished motifs, the chain groups for path homology are formed as the pre-images of the boundary maps. As such, finding a basis for the chain groups is often non-trivial, which complicates the understanding of how homology arises in a random digraph. Hence, it is desirable to develop an understanding of the statistical behaviour of path homology, both from an applied perspective and from independent interest.

Key questions include (as discussed for the clique complex in [64, 65, 67]): when should one expect homology to be trivial or non-trivial; when homology is non-trivial, what are the expected Betti numbers; and how are the Betti numbers distributed? In order to obtain concise, qualitative answers to these questions, one often makes assumptions about how null model parameters depend on the number of nodes n. Then, one can show that a property P holds with probability tending to 1 as $n \to \infty$. Under these conditions, we say

that the property P holds with high probability. Moreover, for two random variables X, Y, we write $X \sim Y$ with high probability if for any $\epsilon > 0$

$$\mathbb{P}\left[1 - \epsilon \leq \frac{X}{Y} \leq 1 + \epsilon\right] \to 1 \quad \text{as } n \to \infty. \tag{2.1}$$

To date, traditional topological invariants enjoy a greater statistical understanding in the context of basic null models. In particular, Kahle showed the following for the Erdős-Rényi random undirected graph model $G(n, p)$, in which every graph has n nodes and each edge appears independently with probability p.

Theorem 2.1.1 (Kahle [64, 65]). *For an Erdős-Rényi random undirected graph $G \sim G(n, p)$, denote the k^{th} Betti number (over a field of characteristic 0) of its clique complex $\mathrm{X}(G)$ by β_k. Let f_k denote the number of k-cliques then $\mathbb{E}[f_k] = \binom{n}{k+1} p^{\binom{k+1}{2}}$. Assume $p = n^{\alpha}$, then*

1. *if $-1/k < \alpha < -1/(k+1)$ then $\beta_k \sim \mathbb{E}[\beta_k] \sim \mathbb{E}[f_k] \sim f_k$ with high probability;*
2. *if $-1/k < \alpha < -1/(k+1)$ then $\beta_k > 0$ with high probability;*
3. *if $\alpha < -1/k$ then $\beta_k = 0$ with high probability;*
4. *if $\alpha > -1/(k+1)$ then $\beta_k = 0$ with high probability.*

In essence, this characterises the understanding that, in any given degree, random graphs only have non-trivial, clique complex homology in a 'goldilocks' region, wherein graph density is neither too big nor too small. Moreover, the boundaries of this region are dependent on the number of nodes in the graph, scaling as a power law. Our primary contribution is a similar description for two different flavours of path homology, in degree 1.

2.1.1 Summary of results

In order to derive useful probability bounds, it is often necessary to prescribe a null model which is highly symmetric and depends on few parameters. Therefore, throughout this paper we will be focusing on an Erdős-Rényi random directed graph model, in which the number of nodes is fixed (at n) and each possible directed edge appears independently, with some probability p. Note, this model allows for the existence of a reciprocal pair of directed edges.

Although individual results are potentially stronger, the following theorems characterise the theoretical understanding that we will develop. Denote the k^{th} Betti number of the non-regular path homology of a digraph by $\overrightarrow{\beta}_k$. Firstly, $(\overrightarrow{\beta}_0 + 1)$, is the number of weakly connected components. This coincides with the number of connected components of the

flat symmetrisation of the digraph (see Definition 2.2.10). If $G \sim \overrightarrow{G}(n,p)$ is an Erdős-Rényi random directed graph then its symmetrisation is an Erdős-Rényi random undirected graph, $\bar{G} \sim G(n,\bar{p})$, where $\bar{p} = 1-(1-p)^2$. Thus, we use a standard result due to Erdős and Rényi [43, §9] (see [64, Therorem 1.1] for a modern summary) to prove the following.

Theorem 2.1.2. *For an Erdős-Rényi random directed graph $G \sim \overrightarrow{G}(n,p(n))$, let $\overrightarrow{\beta}_0$ denote the 0^{th} Betti number of its non-regular path homology over $\mathbb{Z}$. Assume $1-(1-p(n))^2 = (\log(n)+f(n))/n$, then*

1. *if $\lim_{n\to\infty} f(n) = -\infty$ then $\overrightarrow{\beta}_0 > 0$ with high probability;*
2. *if $\lim_{n\to\infty} f(n) = \infty$ then $\overrightarrow{\beta}_0 = 0$ with high probability.*

The same result holds for regular path homology.

Our primary contribution identifies a similar 'goldilocks' region for the first Betti number of path homology, $\overrightarrow{\beta}_1$.

Theorem 2.1.3. *For an Erdős-Rényi random directed graph $G \sim \overrightarrow{G}(n,p(n))$, let $\overrightarrow{\beta}_1$ denote the 1^{st} Betti number of its non-regular path homology over $\mathbb{Z}$. Let N_1 denote the number of edges, $N_1 = \#E(G)$, then $\mathbb{E}[N_1] = n(n-1)p$. Assume $p(n) = n^\alpha$, then*

1. *if $-1 < \alpha < -2/3$ then $\overrightarrow{\beta}_1 \sim \mathbb{E}[\overrightarrow{\beta}_1] \sim \mathbb{E}[N_1] \sim N_1$ with high probability;*
2. *if $-1 < \alpha < -2/3$ then $\overrightarrow{\beta}_1 > 0$ with high probability;*
3. *if $\alpha < -1$ then $\overrightarrow{\beta}_1 = 0$ with high probability;*
4. *if $\alpha > -1/3$ then $\overrightarrow{\beta}_1 = 0$ with high probability.*

The same result holds for regular path homology.

By way of justifying the assumption $p(n) = n^\alpha$ with $\alpha \leq 0$, in Figure 2.9a we plot $\mathbb{P}[\overrightarrow{\beta}_1(G) = 0]$, for $G \sim \overrightarrow{G}(n,p)$, in colour against $\log(n)$ and $\log(p)$ along the two spatial axes. We observe two transitions between three distinct regions in parameter space. There is an interim region, in which we observe mostly $\overrightarrow{\beta}_1 > 0$; when p becomes too small we suddenly observe mostly $\overrightarrow{\beta}_1 = 0$, and likewise when p becomes too large. On this plot, the boundaries between the three regions appear as straight lines. Hence a reasonable conjecture is that these boundaries follow a power-law relationship $\log(p) = \alpha\log(n) + c$. Therefore, following power-law trajectories through parameter space will allow us to derive either $\mathbb{P}[\overrightarrow{\beta}_1(G) > 0] \to 1$ or $\mathbb{P}[\overrightarrow{\beta}_1(G) = 0] \to 1$. Taking α large (i.e. close to 0) yields a 'high density' regime whilst taking lower α causes p to shrink faster and hence corresponds to a 'low density' regime. We will see that the numbers -1 and $-2/3$ result from comparisons

of the growth rates of chain groups (Proposition 2.4.2). The number $-1/3$ appears due to the combinatorics of certain subgraphs called directed centres, which we will show are abundant in the high density regime and imply vanishing $\overrightarrow{\beta}_1$ (Remark 2.4.19).

Turning our attention to higher degrees, we provide weak guarantees for the asymptotic behaviour of $\overrightarrow{\beta}_k$, for arbitrary $k \geq 1$, at low densities.

Theorem 2.1.4. *For an Erdős-Rényi random directed graph $G \sim \overrightarrow{G}(n, p(n))$, let $\overrightarrow{\beta}_k$ denote the k^{th} Betti number of its non-regular path homology over $\mathbb{Z}$. Assume $p(n) = n^{\alpha}$ with $\alpha < -\frac{N+1}{N}$ for some $N \in \mathbb{N}$. Then, $\overrightarrow{\beta}_k = 0$ with high probability for every $k \geq N$. The same result holds for regular path homology.*

For comparison, in Section 2.5, we apply the techniques used to prove Theorem 2.1.3 in order to obtain analogous results for the directed flag complex.

Theorem 2.1.5. *For an Erdős-Rényi random directed graph $G \sim \overrightarrow{G}(n, p(n))$, let β_k denote the k^{th} Betti number of its directed flag complex homology over $\mathbb{Z}$. Let N_k denote the number of directed $(k+1)$-cliques, $N_k = \operatorname{rank} \overrightarrow{X}_k(G)$, then $\mathbb{E}[N_k] = \binom{n}{k+1}(k+1)!\, p^{\binom{k+1}{2}}$. Assume $p(n) = n^{\alpha}$, then for each $k \geq 0$*

1. *if $-1/k < \alpha < -1/(k+1)$ then $\mathbb{E}[\beta_k] \sim \mathbb{E}[N_k]$;*
2. *if $-1 < \alpha < -1/2$ then $\beta_1 \sim \mathbb{E}[\beta_1] \sim \mathbb{E}[N_1] \sim N_1$ with high probability;*
3. *if $-1 < \alpha < -1/2$ then $\beta_1 > 0$ with high probability;*
4. *if $\alpha < -1$ then $\beta_1 = 0$ with high probability;*
5. *if $\alpha > -1/4$ then $\beta_1 = 0$ with high probability.*

In Section 2.6, we summarise these results and compare path homology and the directed flag complex to more traditional symmetric methods. We provide Table 2.1 in which we record, for each of the homologies under consideration, the α-region in which we know β_1 is either zero or positive, with high probability (assuming $p = n^{\alpha}$).

In Appendix 2.A, with a view towards hypothesis testing, we derive a tighter explicit bound for $\mathbb{P}(\overrightarrow{\beta}_1(G) > 0)$, which becomes useful when p is large. In order to identify a given Betti number as statistically significant, against a Erdős-Rényi null model, one would usually resort to a Monte Carlo permutation test (e.g. [41]). This would require the computation of path homology for a large number of random graphs. For large graphs ($n \geq 100$ nodes), this is often infeasible, due to the computational complexity of path homology. However, if graph density falls into one of the regions identified by the results in Appendix 2.A, one can potentially circumvent this costly computation.

2.1.2 Data Availability

The code and data for the experiments of Appendix 2.B, as well as an implementation of the algorithm described in Lemma 2.A.7, are available on both the OSF repository [22] and GitHub [21]. All code is written in `MATLAB` and data from the experiments is available in the `.mat` format.

2.2 Background

2.2.1 Graph theory definitions and assumptions

For clarity, we present a number of standard definitions, and assumptions that we will use throughout this paper. First, we fix our notation for graphs.

Definition 2.2.1.
1. A (undirected) graph is a pair $G = (V, E)$, where V is an arbitrary set and E is a set of 2-element subsets of V.
2. A directed graph (or digraph) is a pair $G = (V, E)$, where V is an arbitrary set and $E \subseteq V \times V$.
3. A (resp. directed) multigraph is a (resp. directed) graph $G = (V, E)$ in which E is allowed to be a multiset.
4. In all cases, we call $V(G) := V$ the set of nodes or vertices and $E(G) := E$ the set of edges.
5. A digraph $G = (V, E)$ is simple if $E \subseteq (V \times V) \setminus \Delta$, where $\Delta := \{(i, i) \mid i \in V\}$.
6. The density of a simple digraph $G = (V, E)$ is the ratio of edges present, relative to the maximum number of possible edges:

$$\text{density}(G) := \frac{\#E}{\#V(\#V - 1)}. \tag{2.2}$$

Assumption 2.2.2. Throughout this paper, unless stated otherwise, we assume that all digraphs $G = (V, E)$ are simple. This means that they contain no self loops and contain at most one edge between any ordered pair of vertices.

Given a directed graph G, we make the following definitions to refer to subgraphs within G.

Definition 2.2.3. Given a digraph $G = (V, E)$, we make the following definitions.

1. A subgraph is another graph $G' = (V', E')$ such that $V' \subseteq V$ and $E' \subseteq E$; we denote this as $G' \subseteq G$.

2. Given a subgraph $G_1 \subseteq G$ and a subset of edges $E_2 \subseteq E(G)$ we let $G_1 \cup E_2$ denote a new graph with edges
$$E(G_1 \cup E_2) = E(G_1) \cup E_2. \tag{2.3}$$
and node-set $V(G_1 \cup E_2)$, the smallest superset of $V(G_1)$ that contains all endpoints of edges in E_2.
3. A (combinatorial) undirected walk is an alternating sequences of vertices and edges
$$\rho = (v_0, e_1, v_1, e_2, \ldots, v_{n-1}, e_n, v_n) \tag{2.4}$$
such that edges connect adjacent vertices, in either direction. That is, for each i, either $e_i = (v_{i-1}, v_i)$ or $e_i = (v_i, v_{i-1})$.
4. A (combinatorial) directed walk is an undirected walk such that all edges are forward edges, that is $e_i = (v_{i-1}, v_i)$ for every i.
5. A (combinatorial) directed/undirected path is a directed/undirected walk which never repeats vertices or edges, that is $v_i = v_j$ or $e_i = e_j$ implies $i = j$.
6. A (combinatorial) directed/undirected cycle is a directed/undirected walk such that
$$v_i = v_j, i \neq j \iff \{i, j\} = \{0, n\}. \tag{2.5}$$
7. The length of a walk is the number of edges it traverses, e.g. the length of ρ in equation (2.4) is n.
8. A double edge is an unordered pair of vertices $\{i, j\} \subseteq V$ such that both directed edges are in the graph, i.e. $(i, j), (j, i) \in E$.

Notation 2.2.4. 1. For vertices $i, j \in V$, we write $i \to j$ if $(i, j) \in E$.
2. If $E_2 = \{e\}$ is a singleton then we define $G_1 \cup e := G_1 \cup E_2$.

Remark 2.2.5. Assumption 2.2.2 allows for the existence of double edges.

2.2.2 Analytic and algebraic definitions

Next, we provide definitions of 'Landau symbols', which we use describe the asymptotic behaviour of two functions, relative to one another.

Definition 2.2.6. Given two functions $f, g : \mathbb{R} \to \mathbb{R}$ we write

1. $f(x) = o(g(x))$ if $\lim_{x\to\infty} \frac{f(x)}{g(x)} = 0$;
2. $f(x) = \omega(g(x))$ if $\lim_{x\to\infty} \frac{g(x)}{f(x)} = 0$;
3. $f(x) \sim g(x)$ if $\lim_{x\to\infty} \frac{f(x)}{g(x)} = 1$.

Remark 2.2.7. There is an equivalence, $f(x) = \omega(g(x)) \iff g(x) = o(f(x))$.

Finally, we make a formal, algebraic definition, which will be required later in order to define path homology.

Definition 2.2.8. Given a ring R and a set V, we let $R\langle V\rangle$ denote the R-module of formal R-linear combinations of elements of V. That is,

$$R\langle V\rangle := \left\{ \sum_{i=1}^{n} \alpha_i e_{v_i} \;\middle|\; n \in \mathbb{N},\ \alpha_i \in R,\ v_i \in V \right\} \tag{2.6}$$

where $\{e_v \mid v \in V\}$ are formal symbols which form a basis of the free R-module $R\langle V\rangle$.

2.2.3 Erdős-Rényi random graphs

Throughout this paper, we will primarily be investigating random directed graphs under an Erdős-Rényi model.

Definition 2.2.9. 1. The Erdős-Rényi random undirected graph model, $G(n,p)$, is a probability space of undirected graphs. Each graph has exactly n nodes $\{1,\dots,n\}$ and each directed edge is included, independently, with probability p. A given graph G on n nodes with m edges appears with probability

$$p^m(1-p)^{\binom{n}{2}-m}. \tag{2.7}$$

2. The Erdős-Rényi random directed graph model, $\overrightarrow{G}(n,p)$, is a probability space of directed graphs. Each graph has exactly n nodes $\{1,\dots,n\}$ and the each directed edge is included, independently, with probability p. A given digraph G on n nodes with m edges appears with probability

$$p^m(1-p)^{n(n-1)-m}. \tag{2.8}$$

2.2.4 Symmetrisation

Definition 2.2.10. Given a directed graph $G = (V,E)$,

1. the flat symmetrisation is an undirected graph, $\bar{G} := (V,\bar{E})$, where

$$\{i,j\} \in \bar{E} \iff (i,j) \in E \text{ or } (j,i) \in E \text{ or both.} \tag{2.9}$$

2. the weak symmetrisation is an undirected multigraph $\mathring{G} := (V,\mathring{E})$, where $\{i,j\}$ appears in $\mathring{E}$ with multiplicity 2 if both $(i,j) \in E$ and $(j,i) \in E$, or with multiplicity if 1 if only one of these edges is present.

Remark 2.2.11. We can view $\bar{G}$ and $\mathring{G}$ as topological spaces by giving them the natural structure of a simplicial complex and delta complex respectively. Both of these structures have no simplices above dimension 1, so clearly $\beta_k(\bar{G}) = \beta_k(\mathring{G}) = 0$ for all $k > 1$.

Lemma 2.2.12. *Given a random directed graph $G \sim \overrightarrow{G}(n,p)$, the flat symmetrisation is distributed as $\bar{G} \sim G(n,\bar{p})$ where*

$$\bar{p} := 1 - (1-p)^2 = 2p - p^2. \tag{2.10}$$

Proof. A given undirected edge $\{i,j\}$ appears in $\bar{G}$ if and only if at least one of (i,j) or (j,i) is in G. Therefore

$$\mathbb{P}(\{i,j\} \notin \bar{E}) = \mathbb{P}\big((i,j) \notin E \text{ and } (j,i) \notin E\big) = (1-p)^2. \tag{2.11}$$

Hence, the undirected edge appears with probability $1-(1-p)^2$. The existence of each undirected edge depends on the existence of a distinct pair of directed edges. Hence each undirected edge appears independently. □

Remark 2.2.13. Note that $p \to 0 \iff \bar{p} \to 0$ hence $\bar{p} \sim 2p$ and therefore, for any $\alpha \leq 0$, $p = o(n^\alpha) \iff \bar{p} = o(n^\alpha)$.

Definition 2.2.14. Throughout this paper, we define $\bar{p}$ as in (2.10), whenever the underlying p is clear from context.

Definition 2.2.15. Given an undirected graph $G = (V,E)$,

1. a k-clique is a subset of vertices $V' \subseteq V$, such that $\#V' = k$ and for any two, distinct vertices, $i,j \in V'$, the edge between them is present, i.e. $\{i,j\} \in E$;
2. the clique complex, $\mathrm{X}(G)$ is a simplicial complex where the k-simplices are the $(k+1)$-cliques in G.

We now investigate the behaviour of these 'symmetric methods' on random directed graphs. Since the flat symmetrisation of a random digraph $\overrightarrow{G}(n,p)$ is a random graph $G(n,\bar{p})$ and the asymptotics of $\bar{p}$ do not differ greatly from those of p, Theorem 2.1.1 can be restated immediately for $\beta_k(\mathrm{X}(\bar{G}))$ with the only change being that $\mathbb{E}[f_k] = \binom{n}{k+1}\bar{p}^{\binom{k+1}{2}}$.

Alternatively, we can view $\bar{G}$ and $\mathring{G}$ directly as 1-dimensional cell complexes without taking the clique complex. In this setup, we prove that if $p = p(n)$ shrinks too quickly then β_1 will vanish for $\bar{G}$ and $\mathring{G}$, with high probability. This is a special case of the proof given by Kahle [64, Theorem 2.6]. We repeat the proof to illustrate that it can be applied to $\beta_1(\bar{G})$, $\beta_1(\mathring{G})$ and, later on, path homology $\overrightarrow{\beta}_1(G)$.

Proposition 2.2.16. *If $p = p(n) = o(n^{-1})$ then, given a random directed graph $G \sim \overrightarrow{G}(n,p)$, we have*

$$\lim_{n\to\infty} \mathbb{P}(\beta_1(\bar{G}) = 0) = \lim_{n\to\infty} \mathbb{P}(\beta_1(\mathring{G}) = 0) = 1. \tag{2.12}$$

Proof. Note that the existence of an undirected cycle in $\bar{G}$ of length $L \in [3,n]$ is a necessary condition for $\beta_1(\bar{G}) > 0$. For each L, by a union bound (i.e. given events $(A_i)_{i\in I}$, we can bound the probability of the union by $\mathbb{P}[\cup_{i\in I} A_i] \leq \sum_{i\in I} \mathbb{P}[A_i]$), the probability of there being an undirected cycle of length L is at most $\binom{n}{L} L! (\bar{p})^L \leq (n\bar{p})^L$. Hence, the probability that there is an undirected cycle of any length is at most $(n\bar{p})^3/(1-(n\bar{p}))$. The assumption $p = o(n^{-1})$ implies that $\lim_{n\to\infty}(n\bar{p}) = 0$ and hence the bound converges to 0 as $n \to \infty$.

To prove $\overrightarrow{\beta}_1(\mathring{G}) = 0$ with high probability, all that remains is to bound probability of there being an undirected cycle on 2 nodes (i.e. a double edge) in G. The probability that there is some double edge is at most $\binom{n}{2} p^2 \leq n^2 p^2$ which tends to 0 since $p = o(n^{-1})$. □

Finally, we investigate conditions under which we expect $\beta_1(\bar{G}) > 0$ and $\beta_1(\mathring{G}) > 0$ with high probability, and determine the growth rate of $\mathbb{E}[\beta_1]$ in each situation. Standard techniques, as employed for the clique complex in [64], shows the following.

Proposition 2.2.17. *If $p = p(n) = \omega(n^{-1})$ then, given a random directed graph $G \sim \overrightarrow{G}(n,p)$,*

$$\mathbb{E}[\beta_1(\bar{G})] \sim \binom{n}{2}\bar{p} \quad \text{and} \quad \mathbb{E}[\beta_1(\mathring{G})] \sim n(n-1)p. \tag{2.13}$$

Moreover, $\beta_1(\bar{G}) \sim \mathbb{E}[\beta_1(\bar{G})]$ and $\beta_1(\mathring{G}) \sim \mathbb{E}[\beta_1(\mathring{G})]$ with high probability and hence

$$\lim_{n\to\infty} \mathbb{P}(\beta_1(\bar{G}) > 0) = \lim_{n\to\infty} \mathbb{P}(\beta_1(\mathring{G}) > 0) = 1. \tag{2.14}$$

Proof. Denoting the original digraph $G = (V,E)$, we deal with the flat symmetrisation first. For convenience, we define $N_1 := \#\bar{E}$ and $N_0 := \#V$. Note that $\mathbb{E}[N_1] = \binom{n}{2}\bar{p}$ and $\mathbb{E}[N_0] = n$. A standard application of the Euler characteristic shows

$$N_1 - N_0 \leq \beta_1 \leq N_1. \tag{2.15}$$

The assumption $p(n) = \omega(n^{-1})$ yields $\mathbb{E}[N_0] = o(\mathbb{E}[N_1])$ and hence $\mathbb{E}[\beta_1] \sim \mathbb{E}[N_1]$.

Now we show $\beta_1 \sim \mathbb{E}[\beta_1]$ with high probability. Since $\mathbb{E}[\beta_1] \sim \mathbb{E}[N_1] \to \infty$, by an application of Chebyshev's inequality [2, Theorem 4.1.1], it suffices to show that $\mathrm{Var}(\beta_1) = o(\mathbb{E}[\beta_1]^2)$. Using the inequalities (2.15) we can (eventually) bound

$$\frac{\mathrm{Var}(\beta_1)}{\mathbb{E}[\beta_1]^2} = \frac{\mathbb{E}[\beta_1^2] - \mathbb{E}[\beta_1]^2}{\mathbb{E}[\beta_1]^2} \leq \frac{\mathbb{E}\left[N_1^2\right] - \mathbb{E}[N_1 - N_0]^2}{\mathbb{E}[N_1 - N_0]^2}. \tag{2.16}$$

Then since N_1 is a binomial random variable with mean $\mathbb{E}[N_1] \to \infty$ we have $\mathbb{E}\left[N_1^2\right] \sim \mathbb{E}[N_1]^2$. We have already seen that $\mathbb{E}[N_1]^2 \sim \mathbb{E}[N_1 - N_0]^2$ and hence the bound in (2.16) tends to 0.

Now since $\beta_1 \sim \mathbb{E}[\beta_1]$ with high probability and eventually $\mathbb{E}[\beta_1] > 0$, the final conclusion $\beta_1 > 0$ with high probability follows because for any $\epsilon \in (0,1)$ we can eventually bound

$$\mathbb{P}[\beta_1(\bar{G}) > 0] \geq \mathbb{P}[\beta_1(\bar{G}) \geq (1-\epsilon)\mathbb{E}[\beta_1]] \to 1. \tag{2.17}$$

The case for the weak symmetrisation has an identical proof, except that $\mathbb{E}[\#\mathring{E}] = \mathbb{E}[\#E] = n(n-1)p$. □

2.3 Path homology of directed graphs

2.3.1 Definition

Path homology was first introduced by Grigor'yan et al. [54, 52]. The key concept behind path homology is that, in order to capture the asymmetry of a digraph, we should not construct a simplicial complex, but instead a *path complex*. In a simplicial complex, one can remove any vertex from a simplex and obtain a new simplex in the complex. This property may not hold for directed paths in digraphs; if we bypass a vertex in the middle of a path then we may not obtain a new path. However, we can always remove the initial or final vertex of a path and obtain a new path. This is the defining property of a path complex [54, §1].

Path homology can be defined on any path complex but for this paper we focus on the natural path complex associated to a digraph. Throughout this section we fix a ring R and a simple digraph $G = (V, E)$.

Definition 2.3.1. We make the following definitions to classify sequences of vertices in V:

1. Any sequence $v_0 \dots v_p$ of $(p+1)$ vertices $v_i \in V$ is an elementary p-path.
2. An elementary path is regular if no two consecutive vertices are the same, i.e. $v_i \neq v_{i+1}$ for every i. Otherwise, the path is called non-regular or irregular.
3. An elementary path is allowed if subsequent vertices are joined by a directed edge in the graph, i.e. $(v_i, v_{i+1}) \in E$ for every i.

Remark 2.3.2. An allowed path coincides with a combinatorial, directed walk.

Definition 2.3.3. The following R-modules are defined to be freely generated by the generators specified, for $p \geq 0$:

$$\Lambda_p := \Lambda_p(G; R) := R\langle \{v_0 \dots v_p \text{ elementary } p\text{-path on } V\} \rangle \tag{2.18}$$

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

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

For $p = -1$, we let $\Lambda_{-1} := \mathcal{R}_{-1} := \mathcal{A}_{-1} := R$. Given an elementary p-path $v_0 \dots v_p$, the corresponding generator of Λ_p is denoted $e_{v_0 \dots v_p}$. For convenience, given an edge $\tau = (a, b) \in E(G)$ we define $e_\tau := e_{ab}$ as an alias for the basis element of $\mathcal{A}_1$.

We can construct homomorphisms $\Lambda_p \to \Lambda_{p-1}$ for each p.

Definition 2.3.4. Given $p > 0$, we can define the non-regular boundary map $\partial_p : \Lambda_p \to \Lambda_{p-1}$ by setting

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

where $v_0 \dots \hat{v_i} \dots v_p$ denotes the elementary $(p-1)$-path $v_0 \dots v_p$ with the vertex v_i omitted. This defines ∂_p on a basis of Λ_p, from which we extend linearly. In the case $p = 0$, we define $\partial_0 : \Lambda_0 \to R$ by

$$\partial_0 \left(\sum_{v \in V} \alpha_v e_v \right) := \sum_{v \in V} \alpha_v \tag{2.22}$$

which yields an element of R.

Remark 2.3.5.

1. A standard check verifies that $\partial_{p-1} \circ \partial_p = 0$ [54, Lemma 2.4] and hence $\{\Lambda_p, \partial_p\}$ forms a chain complex.
2. Since we assume all digraphs are simple, there are no self-loops. Therefore, any allowed path must be regular and hence

$$\mathcal{A}_p \subseteq \mathcal{R}_p \subseteq \Lambda_p. \tag{2.23}$$

In order to incorporate information about paths in the graph we would like a boundary operator between the $\mathcal{A}_p$. However, the boundary of an allowed path may not itself be allowed, because it involves removing vertices from the middle of paths. To resolve this, we define a R-module, for each $p \geq 0$, called the space of ∂-invariant p-paths

$$\Omega_p := \Omega_p(G; R) := \{v \in \mathcal{A}_p \mid \partial_p v \in \mathcal{A}_{p-1}\} = \mathcal{A}_p \cap \partial_p^{-1}(\mathcal{A}_{p-1}). \tag{2.24}$$

Since $\partial_{p-1} \circ \partial_p = 0$, we see that $\partial_p(\Omega_p) \subseteq \Omega_{p-1}$. Hence, we can make the following construction.

Definition 2.3.6. The non-regular chain complex is

$$\dots \xrightarrow{\partial_3} \Omega_2 \xrightarrow{\partial_2} \Omega_1 \xrightarrow{\partial_1} \Omega_0 \xrightarrow{\partial_0} R \xrightarrow{\partial_{-1}} 0 \tag{2.25}$$

where each ∂_p is the restriction of the non-regular boundary map to Ω_p.

Definition 2.3.7. The homology of the non-regular chain complex (2.25) is the non-regular path homology of G. The p^{th} homology group is denoted

$$H_p(G;R) := \frac{\ker \partial_p}{\operatorname{im} \partial_{p+1}}. \tag{2.26}$$

The rank of the p^{th} homology group is p^{th} Betti number, denoted $\overrightarrow{\beta}_p(G;R)$.

When computing Ω_p, one often encounters paths $v \in \mathcal{A}_p$ with irregular summands in their boundary. For example,

$$\partial_2(e_{iji}) = e_{ji} - e_{ii} + e_{ij}. \tag{2.27}$$

Since irregular summands are never allowed, these must be cancelled to obtain an element of Ω_p. An alternative construction, which is featured more frequently in the literature, alters the boundary operator to remove these irregularities.

There is a projection map $\pi : \Lambda_p \to \mathcal{R}_p$ which sends every irregular path to 0. This allows us to make the following construction:

Definition 2.3.8. For each $p \geq 0$, the regular boundary operator $\partial_p^{\mathcal{R}} : \mathcal{R}_p \to \mathcal{R}_{p-1}$ is defined by

$$\partial_p^{\mathcal{R}} := \pi \circ \partial_p. \tag{2.28}$$

With this new boundary operator we still have the issue that the boundary of an allowed path may not be allowed. Therefore, we again construct an R-module, for each $p \geq 0$, called the space of $\partial^{\mathcal{R}}$-invariants p-paths.

$$\Omega_p^{\mathcal{R}} := \Omega_p^{\mathcal{R}}(G;R) := \left\{ v \in \mathcal{A}_p \;\middle|\; \partial_p^{\mathcal{R}} v \in \mathcal{A}_{p-1} \right\} = \mathcal{A}_p \cap (\partial_p^{\mathcal{R}})^{-1}(\mathcal{A}_{p-1}). \tag{2.29}$$

One can check that, given any irregular path v, either $\partial v = 0$ or ∂v is a sum of irregular paths [54, Lemma 2.9] and hence

$$\partial_{p-1}^{\mathcal{R}} \circ \partial_p^{\mathcal{R}} = \pi \circ \partial_{p-1} \circ \partial_p = \pi \circ 0 = 0. \tag{2.30}$$

Definition 2.3.9. The regular chain complex is

$$\dots \xrightarrow{\partial_3^{\mathcal{R}}} \Omega_2^{\mathcal{R}} \xrightarrow{\partial_2^{\mathcal{R}}} \Omega_1^{\mathcal{R}} \xrightarrow{\partial_1^{\mathcal{R}}} \Omega_0^{\mathcal{R}} \xrightarrow{\partial_0^{\mathcal{R}}} R \xrightarrow{\partial_{-1}^{\mathcal{R}}} 0 \tag{2.31}$$

where each $\partial_p^{\mathcal{R}}$ is the restriction of the non-regular boundary map to $\Omega_p^{\mathcal{R}}$.

Definition 2.3.10. The homology of the regular chain complex is the regular path homology of G and the k^{th} homology group is denoted

$$H_k^{\mathcal{R}}(G;R) := \frac{\ker \partial_k^{\mathcal{R}}}{\operatorname{im} \partial_{k+1}^{\mathcal{R}}}. \tag{2.32}$$

We denote the Betti numbers for these homology groups by $\overrightarrow{\beta}_k^{\mathcal{R}}(G;R)$.

Remark 2.3.11.

1. If R is also a field, then the homology groups H_k and $H_k^{\mathcal{R}}$ are vector spaces and so fully characterised, up to isomorphism, by $\overrightarrow{\beta}_k$ and $\overrightarrow{\beta}_k^{\mathcal{R}}$ respectively.
2. Since we augment the chain complex with R in dimension -1, this is technically a reduced homology, but we omit additional notation for simplicity.
3. As noted in [54, §5.1], given a subgraph $G' \subseteq G$ then, for every $p \geq 0$,

$$\Omega_p(G') \subseteq \Omega_p(G) \quad \text{and} \quad \Omega_p^{\mathcal{R}}(G') \subseteq \Omega_p^{\mathcal{R}}(G). \tag{2.33}$$

Notation 2.3.12. When G is clear from context, we shall omit it from notation. If the coefficient ring R is omitted from notation, assume that $R = \mathbb{Z}$.

Note that the primary difference between the regular and non-regular chain complex is the boundary operator. The difference between the boundary operators ∂_p and $\partial_p^{\mathcal{R}}$ affects the difference between the R-modules Ω_p and $\Omega_p^{\mathcal{R}}$.

2.3.2 Proof of Theorem 2.1.4

As an easy first step, we show that, when graph density is too low, it is very unlikely that there are any long paths within the digraph. Therefore, for large k, $\mathcal{A}_k$ becomes trivial and consequently $\overrightarrow{\beta}_k = 0$.

Proposition 2.3.13. *Given $N \in \mathbb{N}$, if $p = p(n) = o(n^{-(N+1)/N})$ for some $N \in \mathbb{N}$ then, given a random directed graph $G \sim \overrightarrow{G}(n,p)$, for all $k \geq N$ we have*

$$\lim_{n\to\infty} \mathbb{P}(\overrightarrow{\beta}_k(G) = 0) = 1. \tag{2.34}$$

Proof. Note that it suffices to show that $\mathbb{P}(\mathcal{A}_N = \{0\}) \to 1$ as $n \to \infty$ because, if there are no allowed N-paths, then there are certainly no allowed k-paths. If there are no allowed k-paths then $\Omega_k = \{0\}$ and so $\overrightarrow{\beta}_k = 0$.

For $\mathcal{A}_N$ to be non-trivial there must be some combinatorial, directed walk of length N. Equivalently, there must exist a combinatorial, directed cycle or a combinatorial, directed path of length N (or both).

If $p = o(n^{-(N+1)/N})$ then certainly $p = o(n^{-1})$ and hence, following the proof of Proposition 2.2.16, the probability that there is a directed cycle tends to 0 as $n \to \infty$.

A combinatorial, directed path is a sequence of $N+1$ distinct nodes, each joined by an edge in the forward direction. By a union bound, the probability that there exists such a sequence is at most

$$\binom{n}{N+1}(N+1)!\, p^N \leq n^{N+1} p^N \tag{2.35}$$

which, by the assumption on p, tends to 0 as $n \to \infty$. □

Proof of Theorem 2.1.4. By Proposition 2.3.13, it suffices to note that $n^\alpha = o(n^{-(N+1)/N})$ whenever $\alpha < -\frac{N+1}{N}$. □

This theorem is very weak. For example, to obtain $\overrightarrow{\beta}_1 = 0$ with high probability, we require $p = o(n^{-2})$, in which case the expected number of edges in the digraph tends to 0. The weakness of this result stems from its reliance on the chain of inequalities

$$\overrightarrow{\beta}_k \leq \operatorname{rank}(\ker \partial_k) \leq \operatorname{rank} \Omega_k \leq \operatorname{rank} \mathcal{A}_k. \tag{2.36}$$

There is likely a region of graph densities wherein one or more of these inequalities is strict. Hence, in order to obtain stronger results, we require an understanding of Ω_k, at the very least.

2.3.3 Chain group generators

Proposition 2.3.14 ([54, §3.3])**.** *For any simple digraph $G = (V, E)$,*

$$\Omega_0 = \Omega_0^{\mathcal{R}} = R\langle V \rangle = \mathcal{A}_0 \quad \textit{and} \quad \Omega_1 = \Omega_1^{\mathcal{R}} = R\langle E \rangle = \mathcal{A}_1. \tag{2.37}$$

Proof. Certainly $\Omega_0 \subseteq \mathcal{A}_0$ and $\Omega_1 \subseteq \mathcal{A}_1$. Moreover, the boundary of any vertex is just an element of $R = \mathcal{A}_{-1}$ and hence allowed. The boundary of any edge is a sum of vertices and any vertex is an allowed 0-path. Therefore $\mathcal{A}_0 \subseteq \Omega_0$ and $\mathcal{A}_1 \subseteq \Omega_1$. □

We can also see that the non-regular chain complex is a subcomplex of the regular chain complex. This subcomplex relation was first noted by Grigor'yan et al. [54, Proposition 3.16]. Since the chain groups agree in dimension 1, this yields an inequality between the first Betti numbers.

Proposition 2.3.15. *For any simple digraph G, the non-regular chain complex is a subcomplex of the regular chain complex. In particular, for each $p \geq 0$, we have*

$$\Omega_p(G) \subseteq \Omega_p^{\mathcal{R}}(G). \tag{2.38}$$

Proof. Suppose $v \in \Omega_p$, then $\partial_p(v) \in \mathcal{A}_{p-1}$. We have seen that $\mathcal{A}_{p-1} \subseteq \mathcal{R}_{p-1}$. Hence, if we project $\partial_p(v)$ onto $\mathcal{R}_{p-1}$ via π, we do not remove any summands. Therefore

$$\partial_p^{\mathcal{R}}(v) = \pi\left(\partial_p(v)\right) = \partial_p(v) \in \mathcal{A}_{p-1}. \tag{2.39}$$

Certainly $v \in \mathcal{A}_p$ and hence $v \in \Omega_p^{\mathcal{R}}$. Since the two operators, ∂_p and $\partial_p^{\mathcal{R}}$, agree on Ω_p, the non-regular chain complex is a subcomplex of the regular chain complex. □

Corollary 2.3.16. *For any simple digraph G, $\overrightarrow{\beta}_1^{\mathcal{R}}(G) \leq \overrightarrow{\beta}_1(G)$.*

Proof. By Proposition 2.3.14, the two complexes coincide in dimensions 0 and 1 and hence $\operatorname{rank}\ker\partial_1 = \operatorname{rank}\ker\partial_1^{\mathcal{R}}$. By Proposition 2.3.15, $\operatorname{im}\partial_2 \subseteq \operatorname{im}\partial_2^{\mathcal{R}}$ and hence $\operatorname{rank}\operatorname{im}\partial_2 \leq \operatorname{rank}\operatorname{im}\partial_2^{\mathcal{R}}$. Therefore,

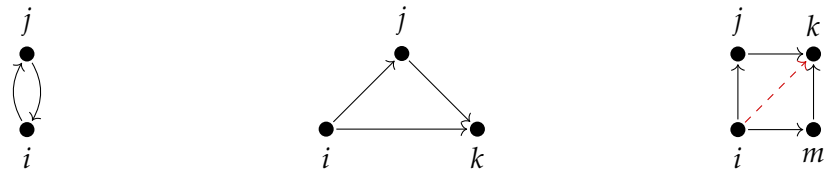
$$\overrightarrow{\beta}_1^{\mathcal{R}}(G) = \operatorname{rank}\ker\partial_1^{\mathcal{R}} - \operatorname{rank}\operatorname{im}\partial_2^{\mathcal{R}} \leq \operatorname{rank}\ker\partial_1 - \operatorname{rank}\operatorname{im}\partial_2 = \overrightarrow{\beta}_1(G). \tag{2.40}$$

□

Note, given a directed edge $\tau = (i, j)$, $\partial_1(e_{ij}) = \partial_1^{\mathcal{R}}(e_{ij}) = e_j - e_i$. From this, it is easy to obtain the characterisation of the lowest Betti number first stated in Section 2.1.

Proof of Theorem 2.1.2. A standard argument shows that $\overrightarrow{\beta}_0 = \overrightarrow{\beta}_0^{\mathcal{R}} = \#C - 1$, where $\#C$ is the number of weakly connected components of the digraph G. Note, $\#C$ coincides with the number of connected components of the symmetrisation $\bar{G}$. The result follows by Lemma 2.2.12 and a standard result due to Erdős and Rényi (see [43, §9] or [64, Therorem 1.1] for a modern summary). □

Unfortunately, higher chain groups do not enjoy such a concise description. However, when working with coefficient over $\mathbb{Z}$, it is possible to write down generators for $\Omega_2^{\mathcal{R}}$, in terms of motifs within the digraph G.



(a) (b) (c)

Figure 2.2: Generators for $\Omega_2^{\mathcal{R}}(G;\mathbb{Z})$ as described in Proposition 2.3.17. (a) A double edge. (b) A directed triangle. (c) A long square. The red, dashed line must not be present for the third motif to constitute a long square.

Proposition 2.3.17 ([55, Proposition 2.9]). *Let G be any finite digraph. Then any $\omega \in \Omega_2^{\mathcal{R}}(G;\mathbb{Z})$ can be represented as a linear combination of 2-paths of the following three types:*

1. *e_{iji} with $i \to j \to i$ (double edges);*
2. *e_{ijk} with $i \to j \to k$ and $i \to k$ (directed triangles);*
3. *$e_{ijk} - e_{imk}$ with $i \to j \to k$, $i \to m \to k$, $i \not\to k$ and $i \neq k$ (long squares).*

Remark 2.3.18. Note that all vertices i, j, k, m in this theorem are distinct, either due to the existence of an edge (e.g. $i \to j$ implies $i \neq j$) or explicit statement (e.g. $i \neq k$).

The following non-regular corollary follows immediately since, by Proposition 2.3.15, $\Omega_2 \subseteq \Omega_2^{\mathcal{R}}$.

Corollary 2.3.19. *Let G be any finite digraph. Then any $\omega \in \Omega_2(G;\mathbb{Z})$ can be represented as a linear combination of 2-paths of the three types enumerated in Proposition 2.3.17.*

Note that each of the generators in Proposition 2.3.17 are elements of $\Omega_2^{\mathcal{R}}$ and hence they form a generating set for $\Omega_2^{\mathcal{R}}(G;\mathbb{Z})$. Note that elements of each type reside in complementary subspaces of $\mathcal{A}_2$ because they are supported on distinct basis elements. That is, we can write

$$\Omega_2^{\mathcal{R}} = D \oplus T \oplus S \tag{2.41}$$

where D is freely generated by all double edges e_{iji} in G, and T is freely generated by all directed triangles e_{ijk} in G. The final component, S, is generated by all long squares $e_{ijk} - e_{imk}$ in G. However, they may not be linearly independent, for example, as seen in Figure 2.3,

$$(e_{ijk} - e_{imk}) + (e_{imk} - e_{ilk}) + (e_{ilk} - e_{ijk}) = 0. \tag{2.42}$$

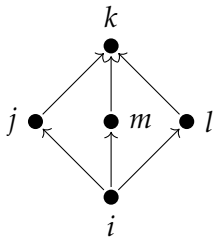


Figure 2.3: Linearly dependent long squares with source i and sink k.

Note that double edges are not ∂-invariant paths, i.e. $e_{iji} \notin \Omega_2$. However there are linear combinations of double edges which do belong to Ω_2. For example, suppose $i \to j \to i$ and $i \to k \to i$, then $e_{iji} - e_{iki} \in \Omega_2$. It is possible to state a non-regular version of

Proposition 2.3.17, in which all generators are elements of Ω_2. This can be achieved by replacing double edge generators with such differences of double edges, which share a common base point. However, we omit this result, as it is not necessary for our main contribution.

Remark 2.3.20.

1. An alternative approach to computing $\operatorname{rank}\Omega_2$ and $\operatorname{rank}\Omega_2^{\mathcal{R}}$ was first seen in [54, Proposition 4.2] and is explored further in Appendix 2.A.2.
2. For the interested reader, more results which characterise relations between the Ω_p are available in [54, 52].
3. We can use Proposition 2.3.17 to obtain an intuition for $H_1^{\mathcal{R}}(G;\mathbb{Z})$. Starting with the cell complex $\mathring{G}$, glue in a 2-cell for each generator identified by Proposition 2.3.17 by identifying its boundary with the corresponding motif in $\mathring{G}$. Then H_1 of this cell complex coincides with $H_1^{\mathcal{R}}(G;\mathbb{Z})$. Unfortunately, since we do not have generators for $\Omega_p^{\mathcal{R}}$ for $p > 2$, developing intuition in higher degrees is much harder.

Example 2.3.21. For further intuition, we reproduce the example given in [54, Proposition 4.7] but compute both regular and non-regular path homology. A cycle graph is a weakly connected digraph, on $n \geq 2$, nodes such that each vertex has degree 2. Fix a cycle graph G. Then, $H_1^{\mathcal{R}}(G;\mathbb{Z}) \cong \mathbb{Z}$ unless G is a double edge, directed triangle or long square, in which case $H_1^{\mathcal{R}}(G;\mathbb{Z}) = 0$. Whereas, $H_1(G;\mathbb{Z}) \cong \mathbb{Z}$ unless G is a directed triangle or long square, in which case $H_1(G;\mathbb{Z}) = 0$. For more examples, please consult [54].

2.4 Asymptotic results for path homology

Intuitively, we expect that the two transitions, identified in Figure 2.9, correspond to two distinct topological phenomena. When density becomes sufficiently large, cycles start to appear in the graph and $\ker\partial_1$ is non-empty for the first time. Then, when density becomes too large, boundaries enter into Ω_2 which begin to cancel out all of the cycles, removing all homology. In the interim period, we expect that the number of cycles and the number of boundaries is approximately balanced. Therefore, in order to understand the lower boundary we should study $\ker\partial_1$ and in order to understand the upper boundary we should study $\operatorname{im}\partial_2$. In order to show that $\overrightarrow{\beta}_1 > 0$ in the 'goldilocks' region we should compare the growth rates of $\operatorname{rank}\ker\partial_1$ and $\operatorname{rank}\operatorname{im}\partial_2$, or some approximation thereof. Moreover we expect reasonable conditions on $p(n)$ to be of the form $p = o(n^\alpha)$ or $p = \omega(n^\alpha)$ for some α, since conditions of this sort constrain $p(n)$ relative to straight lines through Figure 2.9.

2.4.1 Proof of Theorem 2.1.3(1)

In order to characterise the behaviour of $\overrightarrow{\beta}_1$ when it is non-trivial, we will follow the approach of Kahle in [64, §7]. The approach is to use the 'Morse inequalities'. In the context of a digraph G, denote the ranks of the chain groups by $N_k := \operatorname{rank} \Omega_k(G;\mathbb{Z})$ and $N_k^{\mathcal{R}} := \operatorname{rank} \Omega_k^{\mathcal{R}}(G;\mathbb{Z})$. Then we have

$$-N_{k-1} + N_k - N_{k+1} \leq \overrightarrow{\beta}_k \leq N_k \tag{2.43}$$

and a similar set of inequalities between the $N_k^{\mathcal{R}}$ and $\overrightarrow{\beta}_k^{\mathcal{R}}$. It is usually easier to compute the rank of chain groups than the rank homology groups. Hence, we use the limiting behaviour of N_k to investigate the limiting behaviour of $\overrightarrow{\beta}_k$. First we will need estimates for $\mathbb{E}[N_k]$.

Lemma 2.4.1. *For a random directed graph $G \sim \overrightarrow{G}(n,p)$ we have the following expectations*

$$\mathbb{E}[N_0] = \mathbb{E}\left[N_0^{\mathcal{R}}\right] = n \tag{2.44}$$

$$\mathbb{E}[N_1] = \mathbb{E}\left[N_1^{\mathcal{R}}\right] = n(n-1)p \tag{2.45}$$

$$\mathbb{E}[N_2] \leq \mathbb{E}\left[N_2^{\mathcal{R}}\right] \leq n^2p^2 + n^3p^3 + n^4p^4. \tag{2.46}$$

Proof. The first two claims are clear since they count the expected number of nodes and edges in G, respectively. There is no difference between the regular and non-regular chain complex in dimensions 0 and 1.

We use Proposition 2.3.17 to compute bounds for $\mathbb{E}[N_2^{\mathcal{R}}]$ and then the bound on $\mathbb{E}[N_2]$ follows immediately because $\Omega_2 \subseteq \Omega_2^{\mathcal{R}}$ (by Proposition 2.3.15). Since both orientations of a double edge constitute a distinct basis element of $\Omega_2^{\mathcal{R}}$, the expected number of double edges is $n(n-1)p^2$, which is bounded above by n^2p^2. The expected number of directed triangles is $6\binom{n}{3}p^3$, because each subset of 3 vertices can support 6 distinct directed triangles.

Counting linearly independent long squares is more involved. For an upper bound, note that any subset of 4 vertices can support 12 long squares (not double counting for the two orientations since they differ by a factor of ± 1). Each fixed long square appears with probability $p^4(1-p)$. Therefore an upper bound on the number of *linearly independent* long squares is

$$12\binom{n}{4}p^4(1-p) \leq n^4p^4. \tag{2.47}$$

Combining these counts yields the upper bound on $\mathbb{E}[N_2^{\mathcal{R}}]$. □

Proposition 2.4.2. *If $G \sim \overrightarrow{G}(n,p)$ where $p = p(n)$, with $p(n) = \omega(n^{-1})$ and $p(n) = o(n^{-2/3})$, then*

$$\mathbb{E}[-N_0 + N_1 - N_2] \sim \mathbb{E}[N_1] \quad \textit{and hence} \quad \mathbb{E}[\overrightarrow{\beta}_1(G)] \sim \mathbb{E}[N_1]\,. \tag{2.48}$$

Moreover, the same relations hold between the $N_k^{\mathcal{R}}$ and $\overrightarrow{\beta}_k^{\mathcal{R}}$.

Proof. We prove the non-regular case, but the regular case follows from an identical argument. Using our expectations from Lemma 2.4.1, we see $\mathbb{E}[N_0] = o(\mathbb{E}[N_1])$ because

$$\lim_{n\to\infty} \frac{\mathbb{E}[N_0]}{\mathbb{E}[N_1]} = \lim_{n\to\infty} \frac{n}{n(n-1)p} = \lim_{n\to\infty} \frac{1}{np} = 0, \tag{2.49}$$

where the final equality follows from the assumption $p = \omega(n^{-1})$. Next, note that

$$0 \leq \lim_{n\to\infty} \frac{\mathbb{E}[N_2]}{\mathbb{E}[N_1]} \leq \lim_{n\to\infty} \frac{n^2p^2 + n^3p^3 + n^4p^4}{n(n-1)p} = \lim_{n\to\infty} \left(p + np^2 + n^2p^3\right). \tag{2.50}$$

The assumption $p = o(n^{-2/3})$ is equivalent to $n^{2/3}p \to 0$ as $n \to \infty$. This is sufficient to ensure $p \to 0$, $np^2 \to 0$ and $n^2p^3 \to 0$ as $n \to \infty$ and so $\mathbb{E}[N_2] = o(\mathbb{E}[N_1])$. Hence the first relation follows and the latter follows immediately from the Morse inequalities (2.43). □

Remark 2.4.3. If we choose $p = n^\alpha$ to satisfy the hypotheses of Proposition 2.4.2, then we must have $-1 < \alpha < -2/3$ in which case $\mathbb{E}[N_1]$ is of the order $n^2p = n^{\alpha+2}$. Then $\alpha + 2 > 1$ so $\mathbb{E}[\overrightarrow{\beta}_1] \to \infty$ at least linearly as $n \to \infty$.

Proposition 2.4.4. *If $G \sim \overrightarrow{G}(n,p)$ where $p = p(n)$, with $p(n) = \omega(n^{-1})$ and $p(n) = o(n^{-2/3})$, then $\overrightarrow{\beta}_1(G) \sim \mathbb{E}[\overrightarrow{\beta}_1(G)]$ and $\overrightarrow{\beta}_1^{\mathcal{R}}(G) \sim \mathbb{E}[\overrightarrow{\beta}_1^{\mathcal{R}}(G)]$ with high probability.*

Proof. We prove the non-regular case but the regular case follows by an identical argument. As in Proposition 2.2.17, it suffices to show that $\mathrm{Var}(\overrightarrow{\beta}_1) = o(\mathbb{E}[\overrightarrow{\beta}_1]^2)$.

We have seen that $\mathbb{E}[N_1] \sim \mathbb{E}[-N_0 + N_1 - N_2]$ and certainly $\mathbb{E}[N_1] \to \infty$ as $n \to \infty$. Therefore, eventually $\mathbb{E}[-N_0 + N_1 - N_2] \geq 0$ so, by the Morse inequalities, eventually we can bound

$$\mathbb{E}[\overrightarrow{\beta}_1]^2 \geq \mathbb{E}[-N_0 + N_1 - N_2]^2. \tag{2.51}$$

Moreover, we always have $\overrightarrow{\beta}_1^2 \leq N_1^2$ so eventually we can bound

$$\frac{\mathrm{Var}(\overrightarrow{\beta}_1)}{\mathbb{E}[\overrightarrow{\beta}_1]^2} = \frac{\mathbb{E}[\overrightarrow{\beta}_1^2] - \mathbb{E}[\overrightarrow{\beta}_1]^2}{\mathbb{E}[\overrightarrow{\beta}_1]^2} \leq \frac{\mathbb{E}\left[N_1^2\right] - \mathbb{E}[-N_0 + N_1 - N_2]^2}{\mathbb{E}[N_0 + N_1 - N_2]^2}. \tag{2.52}$$

To conclude, it suffices to show $\mathbb{E}\left[N_1^2\right] \sim \mathbb{E}[-N_0 + N_1 - N_2]^2$. By Proposition 2.4.2, we know $\mathbb{E}[-N_0 + N_1 - N_2]^2 \sim \mathbb{E}[N_1]^2$. Then $\mathbb{E}\left[N_1^2\right] \sim \mathbb{E}[N_1]^2$ because N_1 is a binomial random variable with mean $\mathbb{E}[N_1] \to \infty$. □

Proof of Theorem 2.1.3(1). If $p(n) = n^\alpha$ for $-1 < \alpha < -2/3$ then $p = \omega(n^{-1})$ and $p = o(n^{-2/3})$. Moreover, $p = \omega(n^{-1})$ is sufficient to ensure $\mathbb{E}[N_1] \to \infty$ as $n \to \infty$. Since N_1 is a binomial random variable, this implies $N_1 \sim \mathbb{E}[N_1]$ with high probability. Combining Proposition 2.4.2 and Proposition 2.4.4 yields the result. □

2.4.2 Proof of Theorem 2.1.3(2)

Having done the work of showing $\overrightarrow{\beta}_1 \sim N_1$, showing that $\overrightarrow{\beta}_1 > 0$ is now an easy corollary.

Corollary 2.4.5. *If $G \sim \overrightarrow{G}(n,p)$ where $p = p(n)$, with $p(n) = \omega(n^{-1})$ and $p(n) = o(n^{-2/3})$, then*

$$\lim_{n\to\infty} \mathbb{P}(\overrightarrow{\beta}_1(G) > 0) = \lim_{n\to\infty} \mathbb{P}(\overrightarrow{\beta}_1^{\mathcal{R}}(G) > 0) = 1. \tag{2.53}$$

Proof. As in Proposition 2.2.17, this follows immediately because $\overrightarrow{\beta}_1 \sim \mathbb{E}[\overrightarrow{\beta}_1]$ and $\overrightarrow{\beta}_1^{\mathcal{R}} \sim \mathbb{E}[\overrightarrow{\beta}_1^{\mathcal{R}}]$ with high probability and both expectations are eventually positive. □

Proof of Theorem 2.1.3(2). If $p(n) = n^{\alpha}$ for $-1 < \alpha < -2/3$ then $p = \omega(n^{-1})$ and $p = o(n^{-2/3})$. Hence, by Corollary 2.4.5, $\mathbb{P}[\overrightarrow{\beta}_1(G) > 0] \to 1$ and $\mathbb{P}[\overrightarrow{\beta}_1^{\mathcal{R}}(G) > 0] \to 1$ as $n \to \infty$. □

2.4.3 Proof of Theorem 2.1.3(3)

Having understood the behaviour of $\mathbb{E}[\overrightarrow{\beta}_1]$ in the 'goldilocks' region, we turn our attention to the boundaries of this region. As with the symmetric methods, we expect that if p is too small then $\overrightarrow{\beta}_1$ will vanish due to the lack of cycles.

Proposition 2.4.6. *If $p = p(n) = o(n^{-1})$ then, given directed random graphs $G \sim \overrightarrow{G}(n,p)$, we have*

$$\lim_{n\to\infty} \mathbb{P}(\overrightarrow{\beta}_1(G) = 0) = \lim_{n\to\infty} \mathbb{P}(\overrightarrow{\beta}_1^{\mathcal{R}}(G) = 0) = 1. \tag{2.54}$$

Proof. Given a double edge iji, the cycle on its boundary is null-homologous in regular path homology because $\partial_2^{\mathcal{R}}(e_{iji}) = e_{ij} + e_{ji}$. Hence, for the regular case, a necessary condition for $\overrightarrow{\beta}_1^{\mathcal{R}} > 0$ is that there is some undirected cycle, of length at least 3, in the digraph. Whereas, the cycle $e_{ij} + e_{ji}$ may or may not be null-homologous in non-regular path homology. Hence, for the non-regular case, a necessary condition is that there is some undirected cycle, of length at least 2, in the digraph. Therefore, the proof of the regular case is identical to the proof that $\overrightarrow{\beta}_1(\bar{G}) = 0$ with high probability and the proof of the non-regular case is identical to the proof that $\overrightarrow{\beta}_1(\mathring{G}) = 0$ with high probability, as seen in Proposition 2.2.16. □

Proof of Theorem 2.1.3(3). Assume that $p(n) = n^{\alpha}$. If $\alpha < -1$ then $p = o(n^{-1})$ and hence, by Proposition 2.4.6, $\mathbb{P}(\overrightarrow{\beta}_1(G) = 0) \to 1$ and $\mathbb{P}(\overrightarrow{\beta}_1^{\mathcal{R}}(G) = 0) \to 1$ as $n \to \infty$. □

2.4.4 Proof of Theorem 2.1.3(4)

For the previous subsection we chose p small enough to ensure that it is highly likely that $\ker \partial_1$ is empty. We also observe $\overrightarrow{\beta}_1$ vanishing for larger values of p. In these regimes $\ker \partial_1$ is likely non-empty but all cycles are cancelled out by boundaries. Put another way, we wish to show that, when p is large, every cycle $\omega \in \ker \partial_1$ can be shown to satisfy

$$\omega = 0 \qquad (\text{mod } \operatorname{im} \partial_2). \tag{2.55}$$

The strategy is to find conditions under which cycles supported on many vertices can be reduced down to cycles supported on just 3 vertices, and then show that small cycles can be reduced to 0. For this subsection, we will prove that $\mathbb{P}[\overrightarrow{\beta}_1(G) = 0] \to 1$ which then implies, by Corollary 2.3.16, that $\mathbb{P}[\overrightarrow{\beta}_1^{\mathcal{R}}(G) = 0] \to 1$, as $n \to \infty$. First, we need to ensure that we can choose a basis for $\ker \partial_1$ which will be amenable to our reduction strategy.

Definition 2.4.7. Given an element $\omega \in \Omega_1(G)$, we can write ω in terms of the standard basis

$$\omega = \sum_{\tau \in E(G)} \alpha_\tau e_\tau. \tag{2.56}$$

1. We define the support of v to be

$$\operatorname{supp}(\omega) := \{\tau \in E \mid \alpha_\tau \neq 0\}. \tag{2.57}$$

2. We call ω a fundamental cycle if $\omega \in \ker \partial_1$, $\alpha_\tau \in \{\pm 1\}$ for each $\tau \in E$, and $\operatorname{supp}(\omega)$ forms a combinatorial, undirected cycle in G.

Lemma 2.4.8. *Given a simple digraph,* $\ker \partial_1$ *has a basis of fundamental cycles in* G.

Proof. Take an undirected spanning forest T for G, i.e. a subgraph of T in which every two vertices in the same weakly connected component of G can be joined by a unique undirected path through T. One can check that $\partial_1 : \Omega_1(T; R) \to \Omega_0(T; R)$ has trivial kernel, since there are no undirected cycles in T.

Given an edge outside the forest $\tau = (a, b) \in E(G) \setminus E(T)$, there is a unique undirected path ρ through T which joins the endpoints of τ:

$$\rho = (a = v_0, \tau_1, v_1, \ldots, \tau_{k-1}, v_{k-1}, \tau_k, b = v_k). \tag{2.58}$$

for some $v_i \in V(G)$, $\tau_i \in E(G)$. Define

$$\alpha_i := \begin{cases} 1 & \text{if } \tau_i = (v_{i-1}, v_i) \\ -1 & \text{if } \tau_i = (v_i, v_{i-1}) \end{cases} \tag{2.59}$$

and note that

$$\partial_1 \left(e_\tau - \sum_{i=1}^{k} \alpha_i e_{\tau_i} \right) = 0. \tag{2.60}$$

Hence $b_\tau := e_\tau - \sum_{i=1}^{k} \alpha_i e_{\tau_i} \in \ker \partial_1$. Note that b_τ is a fundamental cycle.

The set $B := \{b_\tau \mid \tau \in E(G) \setminus E(T)\}$ is linearly independent because, given $b_\tau \in B$, no other $b_{\tau'} \in B$ involves the basis element e_τ of Ω_1. Note, we can write

$$\Omega_1 = R\langle \{e_\tau \mid \tau \in E(T)\} \rangle \oplus R\langle B \rangle \,. \tag{2.61}$$

Since there are no cycles in the spanning forest T, the kernel of ∂_1 on the first component is trivial. Therefore, $\operatorname{rank} \ker \partial_1 \leq \#B$ and hence B spans $\ker \partial_1$. □

Now we can describe the strategy by which we will systematically reduce long fundamental cycles into smaller ones. We design a combinatorial condition on a directed graph which is more likely to occur at higher densities.

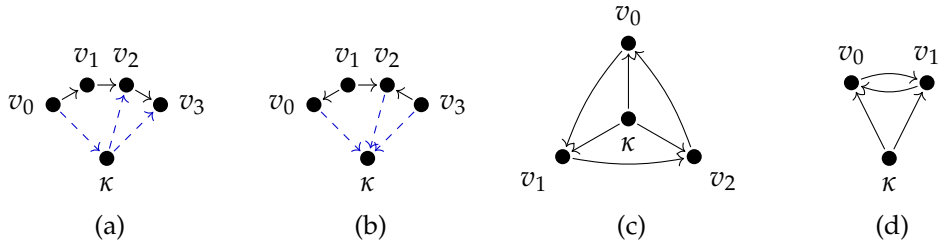


Figure 2.4: (a, b) Directed centres for undirected paths of length 3. The blue, dashed edges constitute $J_{\sigma,\kappa}$. (c) A cycle centre for a 3-cycle. (d) A cycle centre for a 2-cycle.

Definition 2.4.9.

1. An undirected path $\sigma \subseteq G$, on vertices $(v_1, \ldots, v_k)$, is said to be reducible if there is some shortcut edge, $e = (v_i, v_j)$, with $|i - j| > 1$ such that $\overrightarrow{\beta}_1(\sigma \cup e) = 0$. If a path is not reducible then it is called irreducible.
2. Given an undirected path $\sigma \subseteq G$ of length 3, on vertices (v_0, v_1, v_2, v_3), and a vertex $\kappa \in V(G) \setminus V(\sigma)$, define the linking set

$$J_{\sigma,\kappa} := \{e \in E(G) \mid e = (v_i, \kappa) \text{ or } e = (\kappa, v_i) \text{ for some } i\} \,. \tag{2.62}$$

   Such a vertex, κ, is called a directed centre for σ if there is some subset of linking edges $J' \subseteq J_{\sigma,\kappa}$ such that $\overrightarrow{\beta}_1(\sigma \cup J') = 0$ and $\sigma \cup J'$ contains an undirected path, of length 2, on the vertices (v_0, κ, v_3).
3. A cycle centre for a directed cycle of length k, on vertices $(v_0, \ldots, v_{k-1})$, is a vertex $\kappa \in V(G) \setminus \{v_0, \ldots, v_{k-1}\}$ such that $(k, v_i) \in E(G)$ for all $i = 0, \ldots, k-1$ or $(v_i, k) \in E(G)$ for all i.

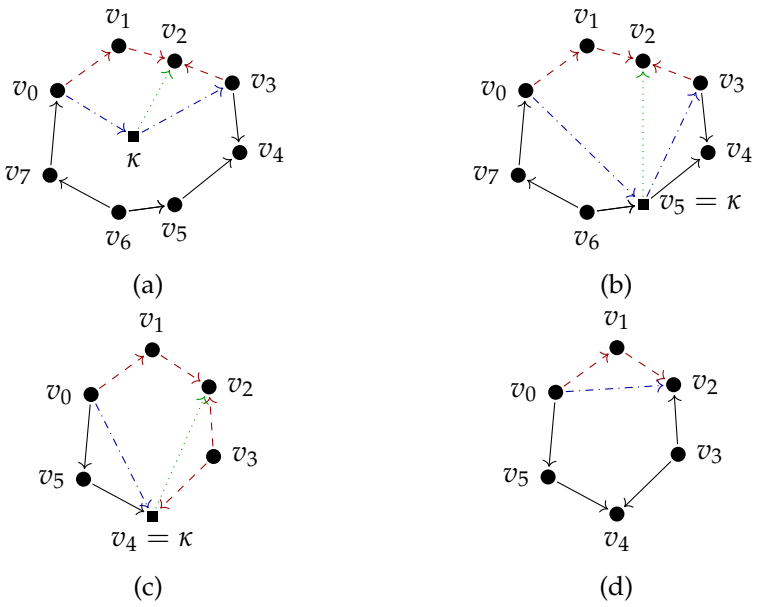


Figure 2.5: Examples of the reductions used in Lemma 2.4.11 which are explained in greater depth in Example 2.4.10. Black, solid edges indicate the initial cycle. Blue, dash-dotted edges are new edges in the reduced cycle. Red, dashed edges are those removed in the reduced cycle. Green, dotted edges must be present in order to do the illustrated reduction. Square nodes symbolise directed centres for the undirected path (v_0, v_1, v_2, v_3).

In the following examples, we demonstrate the utility of directed centres.

Example 2.4.10. Figure 2.5 shows four examples of the reduction strategy described by Lemma 2.4.11. For illustration, we describe these reductions in more detail below.

1. In Figure 2.5a, the initial undirected path of length 3 has a directed centre κ which does not coincide with a vertex in the rest of the cycle. Therefore, we can write

$$\begin{aligned}&[e_{v_0v_1} + e_{v_1v_2} - e_{v_3v_2}] + e_{v_3v_4} - e_{v_5v_4} - e_{v_6v_5} + e_{v_6v_7} + e_{v_7v_0}\\ =&[e_{v_0\kappa} + e_{\kappa v_3}] + e_{v_3v_4} - e_{v_5v_4} - e_{v_6v_5} + e_{v_6v_7} + e_{v_7v_0} \pmod{\operatorname{im}\partial_2}.\end{aligned} \tag{2.63}$$

2. In Figure 2.5b, the path has a directed centre $\kappa = v_5$. Replacing the initial path with the smaller path, via the directed centre, yields a sum of two fundamental cycles:

$$\begin{aligned}&[e_{v_0v_1} + e_{v_1v_2} - e_{v_3v_2}] + e_{v_3v_4} - e_{v_5v_4} - e_{v_6v_5} + e_{v_6v_7} + e_{v_7v_0}\\ =&[e_{v_0\kappa} + e_{\kappa v_3}] + e_{v_3v_4} - e_{v_5v_4} - e_{v_6v_5} + e_{v_6v_7} + e_{v_7v_0} \pmod{\operatorname{im}\partial_2}\\ =&[e_{v_0v_5} - e_{v_6v_5} + e_{v_6v_7} + e_{v_7v_0}] + [e_{v_5v_3} + e_{v_3v_4} - e_{v_5v_4}] \pmod{\operatorname{im}\partial_2}.\end{aligned} \tag{2.64}$$

3. In Figure 2.5c, the path has a directed centre $\kappa = v_4$. Replacing the initial path (v_0, v_1, v_2, v_3) with the smaller path (v_0, κ, v_3) yields a much smaller support since the

edge (v_3, v_4) gets cancelled out:

$$\begin{aligned}
&[e_{v_0v_1} + e_{v_1v_2} - e_{v_3v_2}] + e_{v_3v_4} - e_{v_5v_4} - e_{v_0v_5} \\
=&[e_{v_0\kappa} - e_{v_3\kappa}] + e_{v_3v_4} - e_{v_5v_4} - e_{v_0v_5} \pmod{\operatorname{im}\partial_2} \\
=&e_{v_0v_4} - e_{v_5v_4} - e_{v_0v_5} \pmod{\operatorname{im}\partial_2}.
\end{aligned} \tag{2.65}$$

4. Finally, in Figure 2.5d, the initial path is reducible via the shortcut edge (v_0, v_2) and hence

$$\begin{aligned}
&[e_{v_0v_1} + e_{v_1v_2} - e_{v_3v_2}] + e_{v_3v_4} - e_{v_5v_4} - e_{v_0v_5} \\
=&[e_{v_0v_2} - e_{v_3v_2}] + e_{v_3v_4} - e_{v_5v_4} - e_{v_0v_5} \pmod{\operatorname{im}\partial_2}.
\end{aligned} \tag{2.66}$$

These examples tell the story of each case in the following lemma, in which we confirm that the presence of directed centres allows us to systematically reduce fundamental cycles.

Lemma 2.4.11. *For any simple digraph G, suppose every irreducible, undirected path of length 3 has a directed centre. Given a fundamental cycle $\omega \in \ker\partial_1$ with $\#\operatorname{supp}(\omega) = k \geq 4$, there exists fundamental cycles $\tilde{\omega}_1, \tilde{\omega}_2 \in \ker\partial_1$ such that*

$$\omega = \tilde{\omega}_1 + \tilde{\omega}_2 \pmod{\operatorname{im}\partial_2} \tag{2.67}$$

with $\#\operatorname{supp}(\tilde{\omega}_1) + \#\operatorname{supp}(\tilde{\omega}_2) \leq k - 1$ and, potentially, one or more $\tilde{\omega}_i = 0$.

Proof. Since ω is a fundamental cycle, it is supported on some combinatorial, undirected cycle

$$\rho = (v_0, \tau_1, v_1, \ldots, v_{k-1}, \tau_k, v_k = v_0). \tag{2.68}$$

for some $v_i \in V$ and $\tau_i \in E$ ordered such that

$$\omega = \sum_{i=1}^{k} \alpha_i e_{\tau_i} \tag{2.69}$$

where

$$\alpha_i := \begin{cases} 1 & \text{if } \tau_i = (v_{i-1}, v_i) \\ -1 & \text{if } \tau_i = (v_i, v_{i-1}) \end{cases}. \tag{2.70}$$

Since $k \geq 4$, the vertices $(v_0, \ldots, v_3)$ are distinct and, along with the edges τ_1, τ_2, τ_3, form an undirected path of length 3. Either this is reducible via some shortcut edge $\tau \in E$, or there exists a directed centre $\kappa \in V$. In either case, there is some undirected path, from v_0 to v_3, of length at most 2. This path is represented by some $\eta' \in \Omega_1$ with coefficients in $\{\pm 1\}$, such that $\partial_1\eta' = e_{v_3} - e_{v_1}$ and $\#\operatorname{supp}\eta' \leq 2$.

Since both η' and $\eta := \sum_{i=1}^{3} \alpha_i e_{\tau_i}$ are supported on undirected paths from v_0 to v_3, we have $\partial_1(\eta - \eta') = 0$. Since $\operatorname{supp}(\eta) \subseteq E(G')$ and $\operatorname{supp}(\eta') \subseteq E(G')$ for some subgraph

$G' \subseteq G$ with $\overrightarrow{\beta}_1(G') = 0$ (either due to reducibility or a directed centre), there is some $u \in \Omega_2(G)$ such that $\partial_2 u = \eta - \eta'$. Therefore we can replace the initial undirected path of length 3, in v, with an undirected path of length at most 2, i.e

$$\omega = \eta' + \sum_{i=4}^{k} \alpha_i e_{\tau_i} =: \tilde{\omega} \pmod{\operatorname{im} \partial_2}. \tag{2.71}$$

Certainly $\#\operatorname{supp}(\tilde{\omega}) \leq 2 + (k-3) < 3 + (k-3) = \#\operatorname{supp}(\omega)$ so $\tilde{\omega}$ has a strictly smaller support. It remains to prove that $\tilde{\omega}$ can be decomposed into a sum of at most two fundamental cycles. In the case that the path has a directed centre κ, we split into two further sub-cases.

Case 1.1: If $\kappa \neq v_i$ for any i, $\operatorname{supp}(\omega)$ and $\operatorname{supp}(\eta')$ are disjoint so all coefficients of $\tilde{\omega}$ are still ± 1. Moreover, since κ is distinct from the vertices of ρ, replacing $(v_0, \ldots, v_3)$ with (v_0, κ, v_3) certainly yields an undirected cycle in G.

Case 1.2: If $\kappa = v_i$ for some $i \in \{0, \ldots, k-1\}$ then there are number of possible sub-cases. If $\operatorname{supp}(\omega) \cap \operatorname{supp}(\eta') = \emptyset$ then all coefficients of $\tilde{\omega}$ are still ± 1. However, the replacement procedure has the effect of pinching $\operatorname{supp}(\tilde{\omega})$ into two edge-disjoint, undirected cycles, which share a vertex at κ. Hence, we can easily decompose $\tilde{\omega}$ into a sum of two fundamental cycles $\tilde{\omega}_1$ and $\tilde{\omega}_2$, supported on each of these underlying cycles.

If the intersection is non-empty, then $\operatorname{supp}(\omega) \cap \operatorname{supp}(\eta') \subseteq \{\tau_4, \tau_k\}$, so there are at most two offending edges. Moreover, in order to attain $\partial_1(\eta - \eta') = 0$ these edges must appear with opposite signs in ω and η' respectively. If there are two offending edges then we must have $4 = k - 1$ and the replacements procedure yields $\tilde{\omega} = 0$. If there is only one offending edge then this edge is no longer contained in $\operatorname{supp}(\tilde{\omega})$ and the length of the underlying undirected cycle is further reduced.

Case 2: If the path was reducible, in most cases $\operatorname{supp}(\omega)$ and $\operatorname{supp}(\eta')$ are disjoint and the replacement process simply removes one or two vertices from the undirected cycle. The only remaining case is if $k = 4$ and $\operatorname{supp}(\omega) \cap \operatorname{supp}(\eta') = \{\tau_k\}$, in which case the replacement procedure yields $\tilde{\omega} = 0$. □

Once we have reduced large cycles into smaller ones, we need conditions to ensure that the resulting small cycles are themselves homologous to zero.

Lemma 2.4.12. *Given a fundamental cycle $\omega \in \ker \partial_1$ such that $\operatorname{supp}(\omega)$ is a directed cycle of length k, if $\operatorname{supp}(\omega)$ has a cycle centre $\kappa \in V$ then*

$$\omega = 0 \pmod{\operatorname{im} \partial_2}. \tag{2.72}$$

Proof. For some vertices $v_0, \dots, v_{k-1} \in V$ and edges $\tau_1, \dots, \tau_k \in E$ we can write the underlying cycle as

$$\rho = (v_0, \tau_1, \dots, v_{k-1}, \tau_k, v_0) \tag{2.73}$$

so that

$$\omega = \pm \sum_{i=0}^{k-1} e_{\tau_i}. \tag{2.74}$$

Since κ is a cycle centre, either $\gamma_i := e_{\kappa v_i v_{i+1}} \in \Omega_2$ for every i or $\gamma_i := e_{v_i v_{i+1} \kappa} \in \Omega_2$ for every i (identifying $v_k = v_0$). In either case, by a telescoping sum argument,

$$\partial_1 \left(\sum_{i=0}^{k-1} \gamma_i \right) = \sum_{i=0}^{k-1} e_{\tau_i}. \tag{2.75}$$

After adjusting for a factor of ± 1, this concludes the proof. □

Piecing these lemmas together, gives us a topological condition, which implies $\overrightarrow{\beta}_1(G) = 0$, and which is likely to occur in high density graphs.

Proposition 2.4.13. *For any simple digraph G, if every irreducible, undirected path of length 3 has a directed centre, and every directed cycle of length 2 or 3 has a cycle centre, then $\overrightarrow{\beta}_1(G) = 0$ and $\overrightarrow{\beta}_1^{\mathcal{R}}(G) = 0$.*

Proof. We prove the non-regular case from which the regular case immediately follows by Corollary 2.3.16.

Fix a basis $\{\omega_1, \dots, \omega_k\}$ of fundamental cycles for $\ker \partial_1$, as described by Lemma 2.4.8. Choose an arbitrary $i \in \{1, \dots, k\}$. It suffices to show that $\omega_i = 0 \pmod{\operatorname{im} \partial_2}$. By Lemma 2.4.11, we can reduce each ω_i to a sum of fundamental cycles

$$\omega_i = \sum_{j=1}^{k_i} \tilde{\omega}_{i,j} \pmod{\operatorname{im} \partial_2} \tag{2.76}$$

with $\#\operatorname{supp} \tilde{\omega}_{i,j} \leq 3$ for each j.

If $\operatorname{supp}(\tilde{\omega}_{i,j})$ is a directed triangle then $\tilde{\omega}_{i,j}$ is the boundary of the corresponding basis element in Ω_2; hence $\tilde{\omega}_{i,j} = 0 \pmod{\operatorname{im} \partial_2}$. Otherwise, $\operatorname{supp}(\tilde{\omega}_{i,j})$ must be a directed cycle or length 2 or 3. In either case, the support has a cycle centre and hence, by Lemma 2.4.12, $\tilde{\omega}_{i,j} = 0 \pmod{\operatorname{im} \partial_2}$. Therefore $\omega_i = 0 \pmod{\operatorname{im} \partial_2}$. □

Remark 2.4.14. Every double edge $(i, j), (j, i) \in E$ appears as an allowed 2-path in $\Omega_2^{\mathcal{R}}$ and $\partial_1^{\mathcal{R}}(e_{iji}) = e_{ji} + e_{ij}$. Therefore, the requirement that every cycle of length 2 has a cycle centre is not strictly necessary to ensure $\overrightarrow{\beta}_1^{\mathcal{R}}(G) = 0$.

Definition 2.4.15. For each $n \in \mathbb{N}$, the complete directed graph on n-nodes, K_n, is defined by

$$V(K_n) := \{1, 2, \ldots, n\}, \quad E(K_n) := \{(i, j) \mid i \neq j\}. \tag{2.77}$$

Moreover, we define the following collection of subgraphs contained within K_n:

1. $P_3^n := \{\text{subgraphs } \sigma \subseteq K_n \mid \sigma \text{ is an undirected path of length 3}\}$;
2. For each $k \geq 2$, $C_k^n := \{\text{subgraphs } \sigma \subseteq K_n \mid \sigma \text{ is a directed cycle of length } k\}$.

Given a random graph $G \sim G(n, p)$, we define the following events:

1. for $\sigma \in P_3^n$ or $\sigma \in C_k^n$ for some $k \geq 2$, S_σ is the event that σ is a subgraph of G;
2. for $\sigma \in P_3^n$, I_σ is the event that σ is irreducible in the graph $G \cup \sigma$;
3. for $\sigma \in P_3^n$, $A_{\sigma,\kappa}$ is the event that κ is a directed centre for σ in the graph $G \cup \sigma$;
4. for $\sigma \in C_k^n$ for some $k \geq 2$, $B_{\sigma,\kappa}$ is the event that κ is a cycle centre for σ in the graph $G \cup \sigma$.

Remark 2.4.16. For a fixed $\sigma \in P_3^n$, the events S_σ, I_σ and $A_{\sigma,\kappa}$ for every $\kappa \in V(G) \setminus V(\sigma)$ are mutually independent. For a fixed $\sigma \in C_k^n$ for some $k \geq 2$, the events S_σ and $B_{\sigma,\kappa}$ for every $\kappa \in V(G) \setminus V(\sigma)$ are mutually independent.

Proposition 2.4.17. *If* $G \sim \overrightarrow{G}(n, p)$, *where* $p = p(n) = \omega\left((n/\log(n))^{-1/3}\right)$, *then*

$$\lim_{n\to\infty} \mathbb{P}(\overrightarrow{\beta}_1(G) = 0) = \lim_{n\to\infty} \mathbb{P}(\overrightarrow{\beta}_1^{\mathcal{R}}(G) = 0) = 1. \tag{2.78}$$

Proof. By Proposition 2.4.13, it suffices to show that the probability that there exists an irreducible, undirected path of length 3 without directed centre, or a cycle of length 2 or 3 without directed centre, tends to 0 as $n \to \infty$. The probability that there is an irreducible, undirected path of length 3 without a directed centre is at most

$$\sum_{\sigma \in P_3^n} \left(\mathbb{P}[S_\sigma] \cdot \mathbb{P}[I_\sigma] \cdot \mathbb{P}\left[\bigcap_{\kappa \in V(G)\setminus V(\sigma)} A_{\sigma,\kappa}{}^c \right] \right). \tag{2.79}$$

Because the events $\{A_{\sigma,\kappa} \mid \kappa \in V(G) \setminus V(\sigma)\}$ are independent, $\mathbb{P}[\cap_\kappa A_{\sigma,\kappa}{}^c] = \prod_\kappa (1 - \mathbb{P}[A_{\sigma,\kappa}])$. Note that $\#P_3^n = \binom{n}{4} 4!\, 2^2$. This count arises because an undirected path of length 3 is determined by a choice of 4 nodes, an order on the nodes and a choice of orientation on each edge. However, this counts each path twice: once in each direction. Also, each path arises in G with probability $\mathbb{P}[S_\sigma] = p^3$ and clearly $\mathbb{P}[I_\sigma] \leq 1$.

For each $\sigma \in P_3^n$ and $\kappa \in V(G) \setminus V(\sigma)$, there is at least one choice of 3 directed edges, from κ to the vertices of the path, which forms a directed centre. Namely, label the vertices

of σ by $(v_0, \dots, v_3)$. Then we can always choose an edge between κ and v_0 and another edge between κ and v_2 so that there is a long square on $\{\kappa, v_0, v_1, v_2\}$, as illustrated in Figure 2.4. The third edge can then be chosen to ensure that there is a directed triangle on $\{\kappa, v_2, v_3\}$. If these three edges are present in G, they constitute $J' \subseteq J_{\sigma,\kappa}$ with the properties required to form a directed centre and hence $\mathbb{P}[A_{\sigma,\kappa}] \geq p^3$. Therefore we can bound the probability (2.79) further by

$$\binom{n}{4} 4!\, 2^2 p^3 \left[1 - p^3\right]^{n-4} \leq 4n^4 p^3 \exp\left(-p^3(n-4)\right). \tag{2.80}$$

We wish to show that this bound tends to 0 as $n \to \infty$. Since $p \leq 1$, it suffices to show $\lim_{n\to\infty} n^4 \exp(-p^3 n) = 0$. By Lemma 2.4.18, the condition on p ensures that $\lim_{n\to\infty}(4\log(n) - p^3 n) = -\infty$. By the continuity of the exponential function, $\lim_{n\to\infty} n^4 \exp(-p^3 n) = 0$.

Note $\#C_2^n = \binom{n}{2}$ and $\#C_3^n = 2\binom{n}{3}$. By another union bound, we see that the probability that there is a directed cycle, of length 2, without cycle centre, is at most

$$\sum_{\sigma \in C_2^n} \left(\mathbb{P}[S_\sigma] \cdot \prod_{\kappa \in V(G)\setminus V(\sigma)} \mathbb{P}[B_{\sigma,\kappa}{}^c] \right) = \binom{n}{2} p^2 [1 - p^2]^{2n-4}. \tag{2.81}$$

Similarly, the probability that there is a directed cycle, of length 3, without cycle centre is at most

$$2\binom{n}{3} p^3 [1 - p^3]^{2n-6}. \tag{2.82}$$

Again, by Lemma 2.4.18, the condition on p suffices to ensure that these two bounds also tend to 0 as $n \to \infty$. □

Lemma 2.4.18. *Given $k \in \mathbb{N}_{>0}$ and $A, B > 0$, if $p = \omega\left((n/\log(n))^{-1/k}\right)$ then*

$$\lim_{n\to\infty} (A\log(n) - Bnp^k) = -\infty. \tag{2.83}$$

Proof. The condition on p is equivalent to $\lim_{n\to\infty} \frac{np^k}{\log(n)} = \infty$, which implies $(A\frac{\log(n)}{np^k} - B) \to -B$ and $np^k \to \infty$ as $n \to \infty$. □

Proof of Theorem 2.1.3(4). Assume that $p(n) = n^\alpha$. If $\alpha > -1/3$ then an application of L'Hôpital's rule shows $p = \omega\left((n/\log(n))^{-1/3}\right)$ and hence, by Proposition 2.4.17, $\mathbb{P}(\overrightarrow{\beta}_1(G) = 0) \to 1$ and $\mathbb{P}(\overrightarrow{\beta}_1^{\mathcal{R}}(G) = 0) \to 1$ as $n \to \infty$. □

Remark 2.4.19. Lemma 2.4.18 reveals the origin of the ratio 1/3, which appears in Theorem 2.1.3(4) and Proposition 2.4.17. In particular, it arises as the ratio between the power of n and the power of p inside the exponential of equation (2.80). The power of n is 1 because

there are on the order of n^1 possible directed centres for an undirected path of length 3. The power of p is 3 because we require at least 3 edges from κ to the path, in order for κ to form a directed centre. In Lemma 2.A.7, we will see that this is indeed the minimal number of edges required to form a directed centre.

The bounds used in the proof of Proposition 2.4.17 are by no means the best possible. Indeed, by splitting P_3^n into four isomorphism classes, it is possible to get exact values for $\mathbb{P}[I_\sigma]$ and $\mathbb{P}[A_{\sigma,\kappa}]$. We explore this further in Appendix 2.A.3 in order to obtain tighter bounds, useful for hypothesis testing.

Moreover, the topological condition for $\overrightarrow{\beta}_1(G) = 0$ presented in Proposition 2.4.13 was chosen since it is likely to occur at high densities. However, there may (and indeed probably does) exist weaker topological conditions which imply $\overrightarrow{\beta}_1(G) = 0$ and occur at somewhat lower densities. This could potentially allow for a weaker hypothesis on Proposition 2.4.17. In order to conjecture the weakest possible hypothesis, we conduct a number of experiments in Appendix 2.B.

2.5 Directed flag complex of random directed graphs

For comparative purposes, we now apply the techniques of Section 2.4 to the directed flag complex, which features more readily in the literature.

Definition 2.5.1. [74, Definition 2.2] An ordered simplicial complex on a vertex set V is a collection of ordered subsets of V, which is closed under taking non-empty, ordered subsets (with the induced order). A subset in the collection consisting of $(k+1)$ vertices is called a k-simplex.

Definition 2.5.2. [74, Definition 2.3] Given a directed graph $G = (V, E)$,

1. a directed $(k+1)$-clique is a $(k+1)$-tuple of distinct vertices $(v_0, \dots, v_k)$ such that $(v_i, v_j) \in E$ whenever $i < j$;
2. the directed flag complex, $\overrightarrow{X}(G)$, (often denoted $\mathrm{dFl}(G)$) is an ordered simplicial complex, whose k-simplicies are the directed $(k+1)$-cliques.

Given a ring R, the directed flag chain complex is $\{\overrightarrow{X}_k(G), \partial_k\}_{k\geq -1}$ where, for $k \geq 0$,

$$\overrightarrow{X}_k(G) := \overrightarrow{X}_k(G; R) := R\langle\{(v_0, \dots, v_k) \mid \text{directed } (k+1)\text{-clique in } G\}\rangle\,, \tag{2.84}$$

$$\partial_k(e_{(v_0,\dots,v_k)}) := \sum_{i=0}^{k}(-1)^i e_{(v_0,\dots,\hat{v_i},\dots,v_k)}. \tag{2.85}$$

where $(v_0, \ldots, \hat{v}_i, \ldots, v_k)$ denotes the directed k-clique $(v_0, \ldots, v_k)$ with the vertex v_i removed. This defines ∂_k on a basis of $\overrightarrow{X}_k(G)$, from which we extend linearly. We also define $\overrightarrow{X}_{-1}(G) = R$ and ∂_0 simply sums the coefficients in the standard basis, as in equation (2.22).

The homology of this chain complex is the directed flag complex homology. The Betti numbers are denoted $\beta_k(\overrightarrow{X}(G))$. When the coefficient ring R is omitted from notation, assume $R = \mathbb{Z}$.

Firstly, as with path homology, $\beta_0(\overrightarrow{X}(G))$ captures the weak connectivity of a digraph G and hence Theorem 2.1.2 also holds for the directed flag complex. Next, since we have an explicit list of generators for $\overrightarrow{X}_k(G)$, and they are easy to count, we can calculate the expected rank of the chain groups in every dimension. As before, we denote the ranks of chain groups as $N_k := \operatorname{rank} \overrightarrow{X}_k(G)$ and use these to estimate the Betti numbers.

Lemma 2.5.3. *For an Erdős-Rényi directed random graph $G \sim \overrightarrow{G}(n,p)$, for any $k \geq 0$ we have*

$$\mathbb{E}[N_k] = \binom{n}{k+1}(k+1)!\, p^{\binom{k+1}{2}}. \tag{2.86}$$

Proof. A possible directed clique is uniquely determined by an ordered $(k+1)$-tuple of distinct vertices. Therefore, there are $\binom{n}{k+1}(k+1)!$ possible cliques. For a given clique to be present, one edge must be present in G for every pair of distinct nodes. □

Using the Morse inequalities as before, this allows us to compute the growth rate of the expected Betti numbers, under suitable conditions on $p = p(n)$.

Proposition 2.5.4. *For $k \geq 0$, if $G \sim \overrightarrow{G}(n,p)$ where $p = p(n)$, with $p(n) = \omega(n^{-1/k})$ and $p(n) = o(n^{-1/(k+1)})$, then*

$$\mathbb{E}[-N_{k-1} + N_k - N_{k+1}] \sim \mathbb{E}[N_k] \quad \textit{and hence} \quad \mathbb{E}[\beta_k(\overrightarrow{X}(G))] \sim \mathbb{E}[N_k]\,. \tag{2.87}$$

Proof. It is easy to check that

$$\frac{\mathbb{E}[N_{k-1}]}{\mathbb{E}[N_k]} = \frac{1}{(n-k)}p^{-k} \sim \frac{1}{np^k} \tag{2.88}$$

which tends to 0 thanks to the condition $p(n) = \omega(n^{-1/k})$. It follows that $\mathbb{E}[N_{k+1}]\,/\mathbb{E}[N_k] \sim np^{k+1}$ which tends to 0 thanks to the condition $p(n) = o(n^{-1/(k+1)})$. As in Proposition 2.4.2, the result then follows by the Morse inequalities. □

Proposition 2.5.5. *If $G \sim \overrightarrow{G}(n,p)$ where $p = p(n)$, with $p(n) = \omega(n^{-1})$ and $p(n) = o(n^{-1/2})$, then $\beta_1(\overrightarrow{X}(G)) \sim \mathbb{E}[\beta_1(\overrightarrow{X}(G)]$ with high probability.*

Proof. The proof is identical to Proposition 2.4.4 except we only need $p(n) = o(n^{-1/2})$ to ensure $\mathbb{E}[-N_0 + N_1 - N_2] \sim \mathbb{E}[N_1]$. □

Corollary 2.5.6. *If $G \sim \overrightarrow{G}(n,p)$ where $p = p(n)$, with $p(n) = \omega(n^{-1})$ and $p(n) = o(n^{-1/2})$, then*

$$\lim_{n\to\infty} \mathbb{P}(\beta_1(\overrightarrow{X}(G)) > 0) = 1. \tag{2.89}$$

Proof. The proof is identical to Corollary 2.4.5 except again we only require $p(n) = o(n^{-1/2})$ to invoke Proposition 2.5.5. □

As with path homology, degree 1 homology appears in the directed flag complex with the appearance of undirected cycles in the underlying digraph. Therefore, the same conditions show that $\beta_1(\overrightarrow{X}(G)) = 0$ with high probability, when $p = p(n)$ shrinks too quickly.

Proposition 2.5.7. *If $p = p(n) = o(n^{-1})$ then, given a random directed graph $G \sim \overrightarrow{G}(n,p)$, we have*

$$\lim_{n\to\infty} \mathbb{P}(\beta_1(\overrightarrow{X}(G)) = 0) = 1. \tag{2.90}$$

Proof. The proof is identical to the non-regular case of Proposition 2.4.6. □

The techniques from Section 2.4.4 can be applied, mutatis mutandis, to show $\beta_1(G) = 0$ with high probability, when $p = p(n)$ shrinks too slowly.

Proposition 2.5.8. *If $G \sim \overrightarrow{G}(n,p)$, where $p = p(n) = \omega\left((n/\log(n))^{-1/4}\right)$, then*

$$\lim_{n\to\infty} \mathbb{P}(\beta_1(\overrightarrow{X}(G)) = 0) = 1. \tag{2.91}$$

Proof. This follows from the same argument as Proposition 2.4.17. The only difference is that a directed centre for an undirected path of length 3 requires at least 4 edges, in order to form 3 directed cliques. This results in the ratio $1/4$ instead of $1/3$. □

As in the earlier sections, the results obtain here suffice to prove the summary result, Theorem 2.1.5, presented in the introduction.

2.6 Discussion

We have identified asymptotic conditions on $p = p(n)$ which ensure that a random directed graph $G \sim \overrightarrow{G}(n,p)$ has $\overrightarrow{\beta}_1(G) > 0$ with high probability. Moreover, under these conditions we showed that $\mathbb{E}[\overrightarrow{\beta}_1(G)] \sim n(n-1)p$. Beneath the lower boundary of this positive region, we showed that $\overrightarrow{\beta}_1(G) = 0$ with high probability. Immediately after the upper boundary of the positive $\overrightarrow{\beta}_1$ range, our theory is inconclusive, but experimental results (shown in Appendix 2.B) provide evidence that $\overrightarrow{\beta}_1(G) = 0$ with high probability. Further away from the positive region, e.g. when $p = n^\alpha$ for $\alpha > -1/3$, our theory again guarantees that $\overrightarrow{\beta}_1(G) = 0$ with high probability. For comparison, we applied these techniques to the directed flag complex and found similar results, with minor changes to the gradient of the boundary lines. We summarise these results, along with similar results for 'symmetric methods' in Table 2.1.

Homology	**Expected growth**	**Positive region (α)**	**Zero region (α)**
Symmetric methods			
$\beta_1(\mathring{G})$	$n(n-1)p$	$(-1,0]$	$(-\infty,-1)$
$\beta_1(\bar{G})$	$\binom{n}{2}\bar{p}$	$(-1,0]$	$(-\infty,-1)$
$\beta_1(\mathrm{X}(\bar{G}))$	$\binom{n}{2}\bar{p}$	$(-1,-1/2)$	$(-\infty,-1)\cup(-1/2,0]$
$\beta_k(\mathrm{X}(\bar{G})), k>1$	$\binom{n}{k+1}\bar{p}^{\binom{k+1}{2}}$	$\left(-\frac{1}{k},-\frac{1}{k+1}\right)$	$\left(-\infty,-\frac{1}{k}\right)\cup\left(-\frac{1}{k+1},0\right]$
Directed methods			
$\beta_1(\overrightarrow{X}(G))$	$n(n-1)p$	$(-1,-1/2)$	$(-\infty,-1)\cup(-1/4,0]$
$\overrightarrow{\beta}_1(G)$	$n(n-1)p$	$(-1,-2/3)$	$(-\infty,-1)\cup(-1/3,0]$
$\overrightarrow{\beta}_1^{\mathcal{R}}(G)$	$n(n-1)p$	$(-1,-2/3)$	$(-\infty,-1)\cup(-1/3,0]$
$\overrightarrow{\beta}_k(G), k>1$	**?**	**?**	$\left(-\infty,\frac{k+1}{k}\right)$
$\overrightarrow{\beta}_k^{\mathcal{R}}(G), k>1$	**?**	**?**	$\left(-\infty,\frac{k+1}{k}\right)$

Table 2.1: Given a random directed graphs $G \sim \overrightarrow{G}(n,p)$, assuming $p = n^\alpha$, we record the known regions of α for which various homologies are either positive, or zero, with high probability. Moreover, we describe the growth rate of the expected Betti numbers, in the respective positive regions. That is, for Betti number β, we give a function $f(n)$ such that, when α is in the positive region, $\mathbb{E}[\beta(G)] \sim f(n)$.

These results, along with known results for the clique complex of a random undirected graph, motivate the following research directions:

1. **Tighter upper boundary** – Experimental results, in Appendix 2.B.3, indicate that the condition $\alpha > -1/3$ of Theorem 2.1.3 could be weakened significantly, potentially as

far as $\alpha > -2/3$. Indeed this is the best possible result because $\overrightarrow{\beta}_1 > 0$ with high probability for $-1 < \alpha < -2/3$. We saw how the ratio $1/3$ arose from our proof in Remark 2.4.19, which informs how we might improve this result. Generalising a directed centre to be a set of $k > 1$ nodes, with suitable conditions, might yield better results since there are on the order of n^k sets of k nodes. However, this would complicate the probability bound because two directed centres would no longer be independent. Alternative approaches may include finding a cover of the random graph which can be used to show that the digraph is contractible via path homotopy [55].

2. **Higher degrees** – So far we only have weak guarantees for the behaviour of $\overrightarrow{\beta}_k$ for $k > 1$. One potential avenue for improvement is to find conditions under which $\mathbb{P}[\Omega_k = \{0\}] \to 1$ as $n \to \infty$. In order to get better results for vanishing $\overrightarrow{\beta}_k$ at small p, we require a greater understanding of generators of $\ker \partial_k$. In order to get better results at large p, we need a high-density, topological condition which implies that these generators can be reduced to $0 \pmod{\operatorname{im} \partial_{k+1}}$.
3. **Distributional results** – One direction of research is to show that normalised $\overrightarrow{\beta}_1$ converges to a normal distribution as $n \to \infty$. More evidence for this conjecture is given in Appendix 2.B.4. This could be done, for example, by Stein's method [30], as is done by Kahle and Meckes [67] for $\beta_1(\mathrm{X}(G(n,p)))$.

2.A Explicit probability bounds

The results presented in Section 2.4 provide a broad-stroke, qualitative description of the behaviour of $\overrightarrow{\beta}_1$ on random directed graphs. However, they provide no guarantees for a digraph of a fixed size, such as those arising in applications. For hypothesis testing, it is desirable to have explicit bounds on the $\mathbb{P}[\overrightarrow{\beta}_1(G) > 0]$ and $\mathbb{P}[\overrightarrow{\beta}_1(G) = 0]$ for $G \sim \overrightarrow{G}(n,p)$, given n and p. In this section, we describe such bounds arising from the proofs in Section 2.4 and improve on them where possible.

2.A.1 Positive Betti numbers at low densities

First we refine the bound developed in Proposition 2.4.6, in order to show that it is unlikely to observe $\overrightarrow{\beta}_1(G) > 0$ when graph density is low.

Theorem 2.A.1. *If $G \sim \overrightarrow{G}(n,p)$ then*

$$\mathbb{P}\left[\overrightarrow{\beta}_1(G) > 0\right] \leq \sum_{L=2}^{n} \binom{n}{L} \frac{L!}{2L} (2p)^L. \tag{2.92}$$

The same bound holds for $\mathbb{P}[\beta_1(\overrightarrow{X}(G)) > 0]$. *The same bound holds for* $\mathbb{P}[\overrightarrow{\beta}_1^{\mathcal{R}}(G) > 0]$ *after removing the* $L = 2$ *term.*

Proof. We start with the non-regular and directed flag complex case. We follow the same argument as the proof of Proposition 2.4.6 but make more accurate estimates. A sufficient condition for both $\overrightarrow{\beta}_1(G) = 0$ and $\beta_1(\overrightarrow{X}(G))$ is that there are no undirected cycles of any length $2 \leq L \leq n$ in G. For each L, there are

$$\binom{n}{L} \frac{L!}{2L} 2^L \tag{2.93}$$

possible undirected cycles. This is because an undirected cycle can be determined by a choice of L vertices, an order on those vertices, and a choice of orientation for each edge. However, this over-counts, by a factor of $2L$, since we could traverse the cycle in either direction and start at any vertex. Each cycle of length L appears with probability p^L, so a union bound yields the result.

For regular path homology, the only undirected cycles of length 2 are double edges, which are boundaries in the regular path complex. Therefore, we can remove the $L = 2$ term from the bound. □

The region of parameter space in which this theorem applies is illustrated in Figure 2.6a. For each n, we plot $p_l^t(n)$, the maximum value such that for all $p \leq p_l^t(n)$, Proposition 2.A.1 implies that $\mathbb{P}[\overrightarrow{\beta}_1(G) > 0] \leq 0.05$.

2.A.2 Zero Betti numbers

In order to obtain the best possible estimate, following the second moment method of Corollary 2.4.5, we need an exact value for $\mathbb{E}[\operatorname{rank} \Omega_2]$. We reproduce the approach of [54, Proposition 4.2], making the necessary alterations for the non-regular case. The approach is to determine the number of linearly independent conditions required to describe Ω_2 as a subspace of $\mathcal{A}_2$.

Definition 2.A.2. Given a directed graph $G = (V, E)$,

1. a semi-edge is an ordered pair of distinct vertices $(i, k) \in V^2$, $i \neq k$ such that $i \not\to k$ but there is some other vertex $j \in V$, $j \neq i, k$ such that $i \to j \to k$;
2. a semi-vertex is a vertex $i \in V$ such that there is some other vertex $j \in V$, $j \neq i$ such that $i \to j \to i$.

We denote the set of all semi-edges by $\mathcal{S}_E$ and all semi-vertices by $\mathcal{S}_V$.

Lemma 2.A.3. *If* $G \sim \overrightarrow{G}(n,p)$ *then*

$$\mathbb{E}[\operatorname{rank}\Omega_2(G;\mathbb{Z})] = n(n-1)^2p^2 - n(n-1)(1-p)\left[1-(1-p^2)^{n-2}\right] - n\left[1-(1-p^2)^{n-1}\right] \tag{2.94}$$

$$\mathbb{E}\left[\operatorname{rank}\Omega_2^{\mathcal{R}}(G;\mathbb{Z})\right] = n(n-1)^2p^2 - n(n-1)(1-p)\left[1-(1-p^2)^{n-2}\right] \tag{2.95}$$

Proof. First we deal with the non-regular case. Given $v \in \mathcal{A}_2$, then $v \in \Omega_2$ if and only if $\partial v \in \mathcal{A}_1$. Let A_2 denote the set of all allowed 2-paths in G, then we can write

$$v = \sum_{ijk\in A_2} v^{ijk}e_{ijk} \tag{2.96}$$

so that

$$\partial v = \sum_{ijk\in A_2} v^{ijk}(e_{jk} - e_{ik} + e_{ij}). \tag{2.97}$$

Since ijk is allowed, so too are ij and jk. Therefore,

$$\partial v = -\sum_{ijk\in A_2} v^{ijk}e_{ik} \pmod{\mathcal{A}_1}. \tag{2.98}$$

Now we split off terms corresponding to double edges

$$\partial v = -\sum_{\substack{ijk\in A_2\\ i\neq k}} v^{ijk}e_{ik} - \sum_{iji\in A_2} v^{iji}e_{ii} \pmod{\mathcal{A}_1}. \tag{2.99}$$

Note ii is never an allowed 1-path. However, for $i \neq k$, ik is an allowed 1-path if $i \to k$, so we can remove these summands

$$\partial v = -\sum_{\substack{ijk\in A_2\\ i\neq k, i\not\to k}} v^{ijk}e_{ik} - \sum_{iji\in A_2} v^{iji}e_{ii} \pmod{\mathcal{A}_1}. \tag{2.100}$$

Therefore, $\partial v \in \mathcal{A}_1$ if and only if for each $(i,k) \in V^2$ with $i \neq k$ and $i \not\to k$

$$\sum_{j:\ ijk\in A_2} v^{ijk} = 0 \tag{2.101}$$

and for each $i \in V$

$$\sum_{j:\ iji\in A_2} v^{iji} = 0. \tag{2.102}$$

Some of the indexing sets of these summations may be empty and hence some of these conditions may be trivial. The remaining conditions are linearly independent and hence it remains to count the number of non-trivial equations. Equation (2.101) is non-trivial if and

only if (i,k) is a semi-edge and equation (2.102) is non-trivial if and only if i is a semi-vertex. Therefore,

$$\operatorname{rank}\Omega_2 = \operatorname{rank}\mathcal{A}_2 - \#\mathcal{S}_E - \#\mathcal{S}_V. \tag{2.103}$$

Taking expectations

$$\mathbb{E}[\operatorname{rank}\mathcal{A}_2] = n(n-1)^2p^2, \tag{2.104}$$

$$\mathbb{E}[\#\mathcal{S}_E] = n(n-1)(1-p)\left[1-(1-p^2)^{n-2}\right], \tag{2.105}$$

$$\mathbb{E}[\#\mathcal{S}_V] = n\left[1-(1-p^2)^{n-1}\right] \tag{2.106}$$

which concludes the non-regular case.

For the regular case, note that equation (2.100) becomes

$$\partial^{\mathcal{R}} v = -\sum_{\substack{ijk\in A_2\\ i\neq k, i\not\to k}} v^{ijk}e_{ik} \pmod{\mathcal{A}_1} \tag{2.107}$$

since the e_{ii} terms are removed by the projection. Hence, all the semi-vertex conditions of equation (2.102) are removed. □

Theorem 2.A.4. *If $G \sim \overrightarrow{G}(n,p)$ then*

$$\mathbb{P}\left[\overrightarrow{\beta}_1(G) > 0\right] \geq \frac{\max\left(0, -n + n(n-1)p - \mathbb{E}[N_2]\right)^2}{n(n-1)p(1-p) + n^2(n-1)^2p^2} \tag{2.108}$$

where

$$\mathbb{E}[N_2] = n(n-1)^2p^2 - n(n-1)(1-p)\left[1-(1-p^2)^{n-2}\right] - n\left[1-(1-p^2)^{n-1}\right]. \tag{2.109}$$

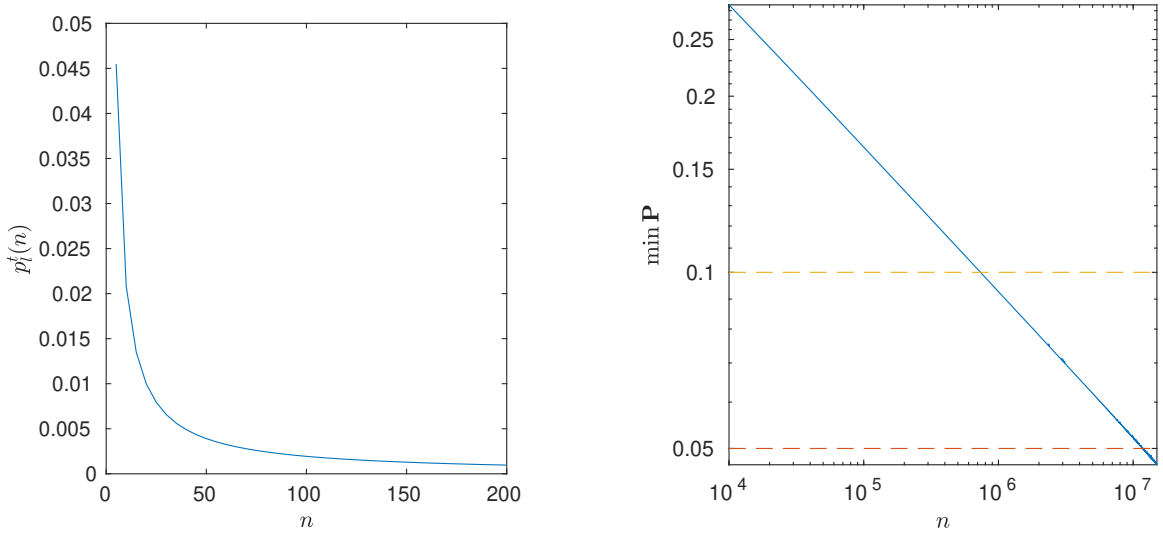
Proof. In theory, we could track back the bound from the theorems invoked by Corollary 2.4.5. Instead, we use a more direct bound. Since $\overrightarrow{\beta}_1(G)$ is a non-negative random variable, an application of the Cauchy-Schwarz inequality to $\mathbb{E}[\overrightarrow{\beta}_1 \mathbb{1}_{\overrightarrow{\beta}_1>0}]$ gives

$$\mathbb{P}[\overrightarrow{\beta}_1(G) > 0] \geq \frac{\mathbb{E}[\overrightarrow{\beta}_1(G)]^2}{\mathbb{E}[\overrightarrow{\beta}_1(G)^2]}. \tag{2.110}$$

Employing the Morse inequalities again, we see

$$\mathbb{P}\left[\overrightarrow{\beta}_1(G) > 0\right] \geq \frac{\max(0, \mathbb{E}[-N_0 + N_1 - N_2])^2}{\mathbb{E}\left[N_1^2\right]} \tag{2.111}$$

where $N_k := \operatorname{rank}\Omega_k(G;\mathbb{Z})$. We obtain the numerator using the expectations for N_0 and N_1 from Lemma 2.4.1 and the expectation of N_2 from Lemma 2.A.3. Then N_1 is a binomial



(a) (b)

Figure 2.6: (a) If (n, p) falls beneath the line $p_l^t(n)$ then Theorem 2.A.1 implies $\mathbb{P}(\overrightarrow{\beta}_1(\overrightarrow{G}(n,p)) > 0) \leq 0.05$. Both axes are linearly scaled. (b) For $n \in [10^4, 1.5 \times 10^7]$, we plot the minimum value of the bound of Theorem 2.A.4 for $p \in [0,1]$. Both axes are logarithmically scaled.

random variable on $n(n-1)$ trials, each with independent probability p, and hence the second moment is

$$\mathbb{E}\left[N_1^2\right] = n(n-1)p(1-p) + n^2(n-1)^2p^2 \tag{2.112}$$

which concludes the proof. □

Remark 2.A.5. Letting N_k denote the rank of the k^{th} chain group in each of the respective chain complexes, it is quick to see that N_1 is the same random variable across all complexes. In order to obtain an analogous theorem to bound $\mathbb{P}[\overrightarrow{\beta}_1^{\mathcal{R}}(G) > 0]$ one need simply replace $\mathbb{E}[N_2]$ with the computation of $\mathbb{E}\left[\operatorname{rank}\Omega_2^{\mathcal{R}}\right]$ from Lemma 2.A.3. To obtain a result for the directed flag complex, one can use the expectations from Lemma 2.5.3.

Unfortunately, this bound is not useful for practical applications. In Figure 2.6b we plot the minimum value of this bound over all $p \in [0,1]$, for a range of n. Note that the bound does not reach a significance level of 0.1, at any p, until approximately $n = 7.4 \times 10^5$ and does not reach a significance level of 0.05 until approximately $n = 1.2 \times 10^7$. At these large graph sizes, computing $\overrightarrow{\beta}_1(G)$ is currently infeasible and hence, without significant advances in computing $\overrightarrow{\beta}_1$, this bound will not serve much practical use.

2.A.3 Positive Betti numbers at high densities

Finally, we tackle bounding $\mathbb{P}[\overrightarrow{\beta}_1(G) > 0]$ when graph density is large. In order to improve upon the naive bound available from Proposition 2.4.17, we provide more avenues for reducing long paths into shorter ones. Specifically, we will obtain better bounds on $\mathbb{P}[I_\sigma]$ and $\mathbb{P}[A_{\sigma,\kappa}{}^c]$. This will not affect the asymptotic behaviour of the bound, but may provide a substantially lower bound, at a fixed n and p.

In order to achieve this, we must partition P_3^n, the set of all possible undirected paths of length 3, on an n-node graph. Every undirected path of length 3 is uniquely determined by a choice of 4 distinct nodes in the graph, an ordering on these nodes (v_0, v_1, v_2, v_3), and a choice of orientation for the edges joining $\{v_0, v_1\}$ and $\{v_2, v_3\}$. The edge joining the middle two nodes is assumed to be 'forward', i.e. via (v_1, v_2), so as to avoid double counting each path. Therefore, we can partition P_3^n, based on the choices of edge orientations, into one of four classes, $P_{3,m}^n$ for $m \in \{0, 1, 2, 3\}$. These classes are visualised in Figure 2.7.

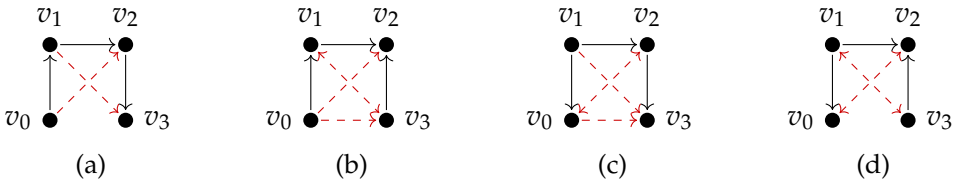


Figure 2.7: The four possible '3-path motifs' which partition P_3^n, shown with black edges. From left to right we see the edge orientations for σ belonging to $P_{3,0}^n$, $P_{3,1}^n$, $P_{3,2}^n$ and $P_{3,3}^n$ respectively. Dashed, red edges indicate edges which must not be present for the path to be considered irreducible.

Lemma 2.A.6. *If $\sigma \in P_{3,m}^n$ then $\mathbb{P}[I_\sigma] = (1-p)^{c_m}$, where $\mathbf{c} = (c_m) = (2, 4, 4, 4)^\mathsf{T}$.*

Proof. The red, dashed edges shown in Figure 2.7 identify 'shortcut' edges; if any one of these edges is present then σ is reducible. Moreover, these are the only shortcut edges, since the addition of any other edge would create a subgraph with $\overrightarrow{\beta}_1 > 0$. Hence, if $\sigma \in P_{3,m}^n$, it is irreducible if and only if none of these c_m edges are present. Since the existence of each edge is independent, the result follows. □

Lemma 2.A.7. *If $\sigma \in P_{3,m}^n$ then*

$$\mathbb{P}[A_{\sigma,\kappa}] = \sum_{l=0}^{8} q_{m,l}\, p^l (1-p)^{8-l}, \tag{2.113}$$

where

$$Q = (q_{m,l}) = \begin{pmatrix} 0 & 0 & 0 & 2 & 11 & 22 & 23 & 8 & 1 \\ 0 & 0 & 0 & 4 & 19 & 33 & 25 & 8 & 1 \\ 0 & 0 & 0 & 4 & 19 & 33 & 25 & 8 & 1 \\ 0 & 0 & 0 & 2 & 16 & 34 & 26 & 8 & 1 \end{pmatrix}. \tag{2.114}$$

Proof. Assume that $\sigma = (v_0, v_1, v_2, v_3) \in P^n_{3,m}$ and $\kappa \in V(G) \setminus V(\sigma)$. Let $q_{m,l}$ denote the number of l-element subsets $J \subseteq L := \{(v, \kappa), (\kappa, v) \mid v \in V(\sigma)\}$ such that κ is a directed centre for σ in the graph $\sigma \cup J$. Then, conditioning on the cardinality of $J_{\sigma,k}$, we can write

$$\mathbb{P}[A_{\sigma,\kappa}] = \sum_{l=0}^{8} \mathbb{P}[A_{\sigma,\kappa} | \, \#J_{\sigma,\kappa} = l] \cdot \mathbb{P}(\#J_{\sigma,\kappa} = l) \tag{2.115}$$

$$= \sum_{l=0}^{8} \frac{q_{m,l}}{\binom{8}{l}} \cdot \binom{8}{l} p^l (1-p)^{8-l}. \tag{2.116}$$

Note that we can write

$$q_{m,l} := \sum_{\substack{J \subseteq L: \\ \#J = l}} \gamma_{m,J}, \tag{2.117}$$

where $\gamma_{m,J}$ is the indicator variable that κ is a directed centre for σ in $\sigma \cup J$. Recalling the definition of directed centres, in order to compute $\gamma_{m,J}$ we must check whether there is any subset of $J' \subseteq J$ such that the graph $\sigma \cup J'$ contains an undirected 2-path from v_0 to v_3 and $\overrightarrow{\beta}_1(\sigma \cup J') = 0$. Hence, if we let $\alpha_{m,J'}$ denote the indicator variable that J' satisfies these conditions, we can decompose $\gamma_{m,J}$ as

$$\gamma_{m,J} = \max_{J' \subseteq J} \alpha_{m,J'}. \tag{2.118}$$

Note that $q_{m,l}$ is independent of the underlying digraph G and the particular choices of σ and κ because, for a fixed m, all $\sigma \cup J$ are isomorphic for $\sigma \in P_{3,m}$ and $\kappa \in V(G) \setminus V(\sigma)$. Therefore, $q_{m,l}$ depends only on m and l, so we can compute an exact value as follows, using equations (2.117) and (2.118):

1. For each $m = 0, 1, 2, 3$, initialise a graph G_m with nodes $V(G_m) = \{v_0, \ldots, v_3, \kappa\}$. The edge set $E(G_m)$ consists solely of an undirected path σ, of length 3, on the vertices $(v_0, \ldots, v_3)$, such that $\sigma \in P_{3,m}$ (the black, solid edges of Figure 2.7).
2. Define the set of all possible linking edges

$$L := \{(v_i, \kappa), (\kappa, v_i) \mid i = 0, 1, 2, 3\}.$$

3. For each subset $J \subseteq L$, construct $G_{m,J} := (V(G_m), E(G_m) \cup J)$.
4. Set $\alpha_{m,J} := 1$ if $G_{m,J}$ contains an undirected path of length 2 from v_0 to v_3 and $\overrightarrow{\beta}_1(G_{m,J}) = 0$. Otherwise, set $\alpha_{m,J} := 0$.
5. For each subset $J \subseteq L$, set $\gamma_{m,J} := 1$ if $\alpha_{m,J'} = 1$ for any $J' \subseteq J$. Otherwise, set $\gamma_{m,J} = 0$.
6. Finally, compute $q_{m,l}$ according to equation (2.117).

This algorithm was implemented as a `MATLAB` script and was subsequently used to compute the matrix Q given in the statement of the theorem. In step 4, Betti numbers are computed via the `pathhomology` package [99], using the `symbolic` option. This uses `MATLAB`'s symbolic computational toolbox in order to avoid any numerical errors. Details on how to access the codebase are available in Section 2.1.2. □

Theorem 2.A.8. *If* $G \sim \overrightarrow{G}(n,p)$ *then*

$$\begin{aligned}\mathbb{P}\left[\overrightarrow{\beta}_1(G) > 0\right] \leq \binom{n}{4} 4! \sum_{m=0}^{3} p^3(1-p)^{c_m}\left(1 - \sum_{l=0}^{8} q_{m,l}\, p^l (1-p)^{8-l}\right)^{n-4} \\ + \binom{n}{2} p^2\left[1-p^2\right]^{2n-4} + 2\binom{n}{3} p^3\left[1-p^3\right]^{2n-6},\end{aligned} \tag{2.119}$$

where $\mathbf{c} = (c_m) = (2,4,4,4)^\intercal$ *and*

$$Q = (q_{m,l}) = \begin{pmatrix} 0 & 0 & 0 & 2 & 11 & 22 & 23 & 8 & 1 \\ 0 & 0 & 0 & 4 & 19 & 33 & 25 & 8 & 1 \\ 0 & 0 & 0 & 4 & 19 & 33 & 25 & 8 & 1 \\ 0 & 0 & 0 & 2 & 16 & 34 & 26 & 8 & 1 \end{pmatrix}. \tag{2.120}$$

Proof. We follow the same approach as Proposition 2.4.17, but with tighter bounds. One can bound the probability that there is an irreducible, undirected 3-path, without a directed centre by

$$\sum_{m=0}^{3} \sum_{\sigma \in P^n_{3,m}} \left(\mathbb{P}[S_\sigma] \cdot \mathbb{P}[I_\sigma] \cdot \prod_{\kappa \in V(G)\setminus V(\sigma)} [1 - \mathbb{P}(A_{\sigma,\kappa})] \right). \tag{2.121}$$

By Lemma 2.A.6 and Lemma 2.A.7 we can bound this further by

$$\binom{n}{4} 4! \sum_{m=0}^{3} p^3(1-p)^{c_m}\left(1 - \sum_{l=0}^{8} q_{m,l}\, p^l (1-p)^{8-l}\right)^{n-4} \tag{2.122}$$

since $\#P^n_{3,m} = \binom{n}{4}4!$ for each m. The same bounds, from Proposition 2.4.17, apply to the probability that there is a directed cycle, of length 2 or 3, without cycle centre. Combining these bounds concludes the proof. □

Remark 2.A.9. By Corollary 2.3.16 the bound identified in Theorem 2.A.8 also holds for $\overrightarrow{\beta}^{\mathcal{R}}_1(G)$. However, as noted in Remark 2.4.14, we can obtain a stronger bound by removing the term

$$\binom{n}{2}\left[1-p^2\right]^{2n-4} \tag{2.123}$$

which corresponds to undirected cycles of length 2.

To demonstrate the utility of this theorem, in Figure 2.8, at each n, we plot the minimum $p_u^t(n)$ such that the bounds from Proposition 2.4.17 (respectively Theorem 2.A.8) imply $\mathbb{P}(\overrightarrow{\beta}_1(\overrightarrow{G}(n,p)) > 0) \leq 0.05$ for all $p \geq p_u^t(n)$. We compute $p_u^t(n)$ via MATLAB's `fzero` root-finding algorithm, with an initial interval of $[0.1, 1]$. For graphs on $n \leq 470$ nodes, the region of p in which Theorem 2.A.8 applies is at least 0.1 larger. With $p_u(t)$ plotted on logarithmic axes, the boundaries both appear to be straight lines with the same gradient. This demonstrates that the bound derived in Theorem 2.A.8 would not allow for weaker asymptotic conditions on $p(n)$ in Proposition 2.4.17.

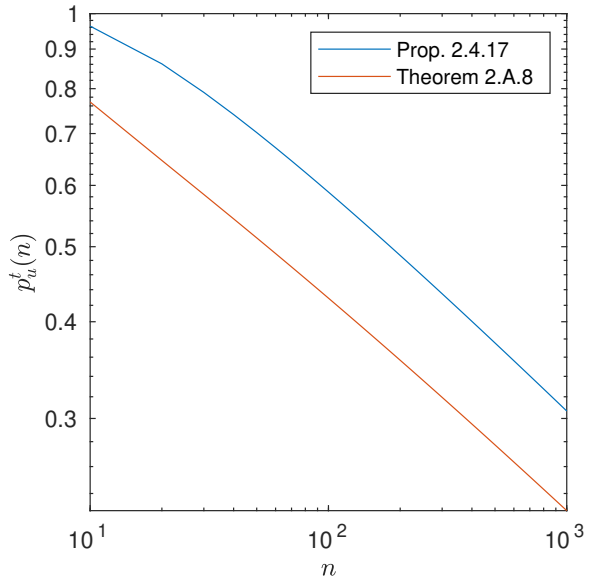


Figure 2.8: Boundaries of the parameter regions in which Proposition 2.4.17 and Theorem 2.A.8 apply. If (n,p) falls above a given line then the corresponding theorem implies $\mathbb{P}(\overrightarrow{\beta}_1(\overrightarrow{G}(n,p)) > 0) \leq 0.05$.

Finally, these same techniques can be applied to the directed flag complex to get a similar explicit bound.

Theorem 2.A.10. *If* $G \sim \overrightarrow{G}(n,p)$ *then*

$$\begin{aligned}\mathbb{P}\left[\beta_1(\overrightarrow{X}(G)) > 0\right] \leq \binom{n}{4} 4! \sum_{m=0}^{3} p^3(1-p)^{c_m} \left(1 - \sum_{l=0}^{8} q_{m,l}\, p^l (1-p)^{8-l}\right)^{n-4} \\ + \binom{n}{2} p^2 \left[1-p^2\right]^{2n-4} + 2\binom{n}{3} p^3 \left[1-p^3\right]^{2n-6},\end{aligned} \tag{2.124}$$

where $\mathbf{c} = (c_m) = (2,3,3,4)^\mathsf{T}$ *and*

$$Q = (q_{m,l}) = \begin{pmatrix} 0 & 0 & 0 & 0 & 5 & 16 & 19 & 8 & 1 \\ 0 & 0 & 0 & 0 & 7 & 20 & 20 & 8 & 1 \\ 0 & 0 & 0 & 0 & 7 & 20 & 20 & 8 & 1 \\ 0 & 0 & 0 & 0 & 8 & 22 & 21 & 8 & 1 \end{pmatrix}. \tag{2.125}$$

Proof. This follows from the same argument as Theorem 2.A.8. The only changes are the counts of possible shortcut edges for each isomorphism class in Lemma 2.A.6, and the output of the algorithm in the proof of Lemma 2.A.7. □

2.B Experimental results

2.B.1 Data collection

In order to further investigate the behaviour of path homology on random directed graphs, we sample empirical distributions of Betti numbers. Table 2.2 records the four experiments that were conducted. In each experiment, a number of random graphs were sampled from $G \sim \overrightarrow{G}(n,p)$, for n evenly spaced in intervals of 5 in the noted n-range, and 50 values of p logarithmically spaced in the noted p-range. Then, we compute the first Betti number of either non-regular path homology or directed flag complex, as noted in the table.

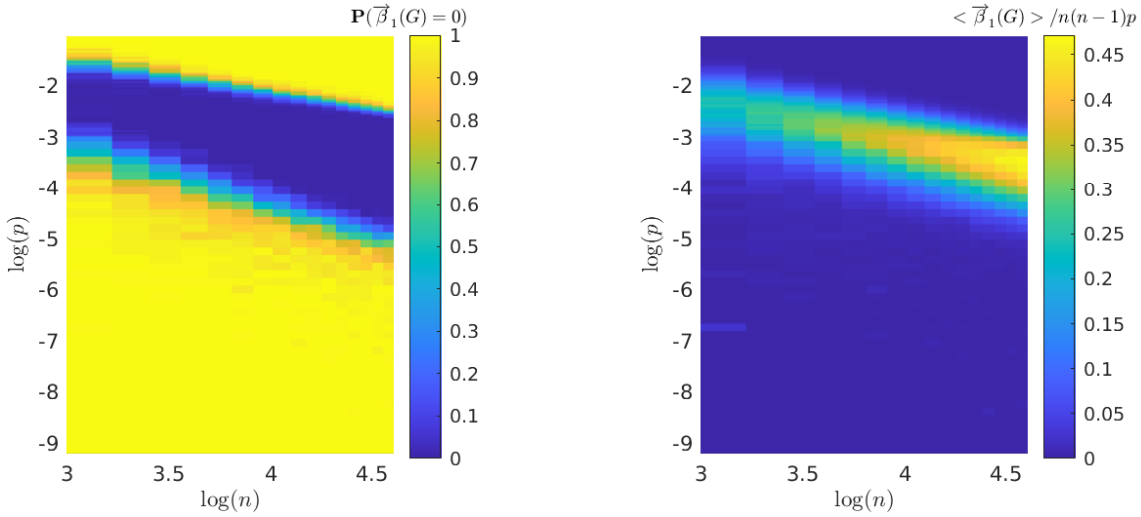
By logarithmically spaced in the range $[a,b]$ we mean that values are chosen evenly spaced between $\log(a)$ and $\log(b)$ and then we apply the exponential function. We discuss the reason for this logarithmic spacing in Appendix 2.B.3. Non-regular path homology is computed with the `pathhomology` package [99]. Unlike in Appendix 2.A, we do not use the `symbolic` option and hence Betti numbers are subject to numerical errors, due to error in rank computations. Directed flag complex homology is computed with the `flagser` package [74], without approximation turned on. Also note that, due to computational restrictions, Experiments 1 and 2 were occasionally stopped and restarted. These experiments were run before `rng` persistence was implemented and hence reproduction attempts may yield slightly different results.

Exp. #	**Homology**	**Samples**	**n-range**	**p-range**
1	$\overrightarrow{\beta}_1(G)$	100	$[20,150]$	$[10^{-4},0.15]$
2	$\overrightarrow{\beta}_1(G)$	100	$[20,100]$	$[0.05,0.35]$
3	$\beta_1(\overrightarrow{X}(G))$	200	$[20,200]$	$[5\times10^{-4},0.3]$
4	$\beta_1(\overrightarrow{X}(G))$	200	$[20,200]$	$[0.1,0.45]$

Table 2.2: Scope of the four data collection experiments.

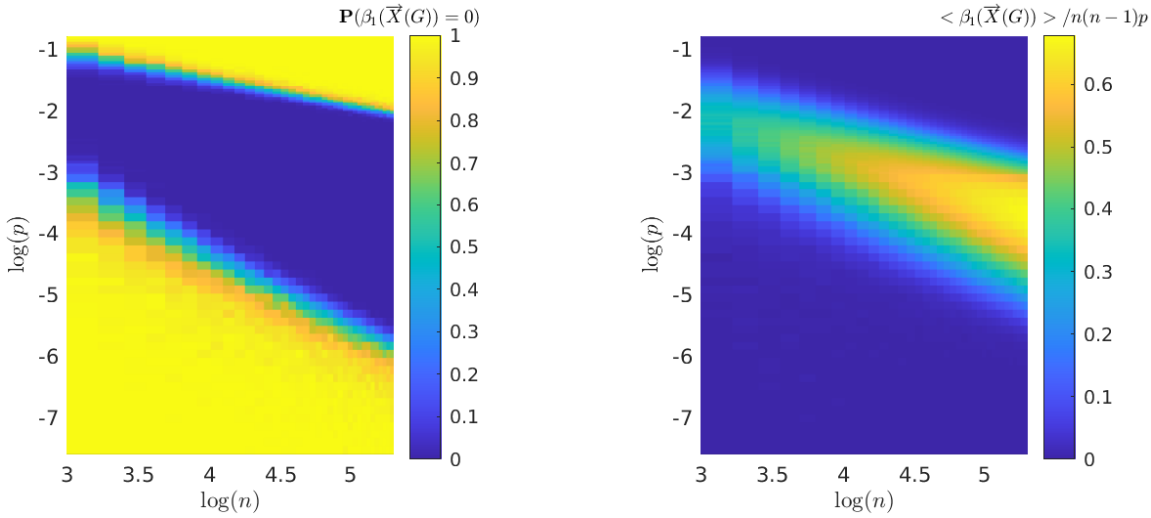
2.B.2 Illustrations

In Figure 2.9, we merge the samples from Experiments 1 and 2 for $n \in [20,100]$. We then use the colour axis to plot statistics for each of these samples, against the logarithm of each



(a) (b)

Figure 2.9: Statistics for the path homology of samples of 100 random directed graphs $G \sim \overrightarrow{G}(n,p)$, sampled at a range of parameter values, plotted against the natural logarithm of the parameters. Our primary contribution is descriptions of the boundaries of the darker, blue region in Figure 2.9a and a limiting value for the lighter, yellow region of Figure 2.9b as $n \to \infty$.



(a) (b)

Figure 2.10: Statistics for the directed flag complex homology of samples of 200 random graphs $G \sim \overrightarrow{G}(n,p)$ sampled at a range of parameter values.

parameter on the two spatial axes. In Figure 2.9a we record the empirical probability that $\overrightarrow{\beta}_1 = 0$ for each of the samples. In Figure 2.9b, we record $\langle \overrightarrow{\beta}_1(G) \rangle / n(n-1)p$, where $\langle \overrightarrow{\beta}_1(G) \rangle$ is the mean $\overrightarrow{\beta}_1$ for each random sample of graphs. In the following informal discussion, we refer to these figures as proxies for the exact probabilities and expectations of the distributions from which we sample.

We observe two distinct transitions between three distinct regions in parameter space. Namely, when graph density is relatively low $\mathbb{P}[\overrightarrow{\beta}_1 = 0]$ is almost 1. Next, there is a 'goldilocks' region in which suddenly $\mathbb{P}[\overrightarrow{\beta}_1 = 0]$ is close to 0 and $\mathbb{E}[\overrightarrow{\beta}_1]$ appears to be growing. Finally, when density is too large, we transition back to a regime in which $\mathbb{P}[\overrightarrow{\beta}_1 = 0]$ is almost 1. Thanks to the logarithmic scaling of the parameter axes, the boundaries between these regions appear to be straight lines.

Theorem 2.1.3 describes the fate of straight line trajectories through this diagram. Theorem 2.1.3(3-4) says that a straight line with gradient $m < -1$ (resp. $m > -1/3$) will eventually cross into and remain in the lower (resp. upper) yellow region of Figure 2.9a, where $\mathbb{P}(\overrightarrow{\beta}_1 = 0)$ is close to 1. Theorem 2.1.3(2) says that a straight line with gradient $-1 < m < -2/3$ will eventually reach the blue region of Figure 2.9a, where $\mathbb{P}(\overrightarrow{\beta}_1 = 0)$ is close to 0. In particular, this implies that the gradient of the lower boundary tends towards -1 and the gradient of the upper boundary is eventually in the region $[-2/3, -1/3]$. Finally, Theorem 2.1.3(1) says that a straight line with gradient $-1 < m < -2/3$ will eventually reach the yellow region of Figure 2.9b and the ratio $\langle \overrightarrow{\beta}_1(G) \rangle / n(n-1)p$ will approach 1.

In Figure 2.10, we merge the samples from Experiments 3 and 4 for $n \in [20, 200]$. This figure provides Theorem 2.1.5 with a similar interpretation, as above, except that the upper boundary is eventually in the region $[-1/2, -1/4]$.

It is worth reiterating that these interpretations and results all hold *in the limit*. That is, Theorem 2.1.3 provides no guarantees for a finite line segment, regardless of its gradients or length. However, we do observe that the boundaries between the three regions converge onto straight lines, of the correct gradient, relatively quickly (i.e. within $n \leq 50$ nodes). We will see empirical evidence for this in the following section.

2.B.3 Finding boundaries

Note, Theorem 2.1.3 says nothing of the region $-2/3 < \alpha < -1/3$. In the following discussion we attempt to determine, empirically, the equations of the boundaries between the positive region and the two zero regions identified in Figure 2.9. This provides evidence

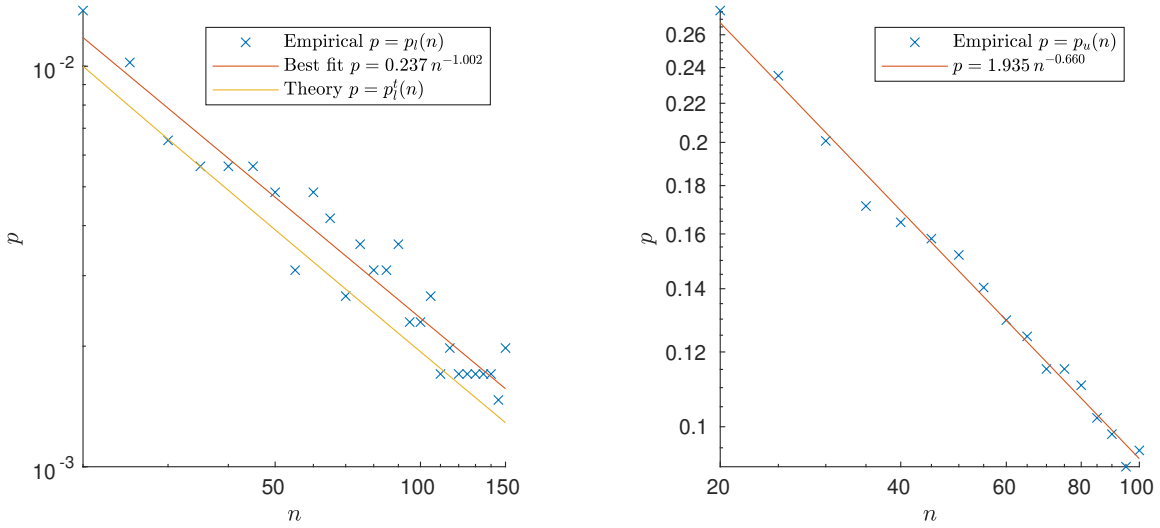


Figure 2.11: Over a range of parameters (n, p) we measure the first Betti number, $\overrightarrow{\beta}_1(G)$, for 100 sampled random graphs $G \sim \overrightarrow{G}(n, p)$. Then, at each n, we determine maximum $p_l(n)$ (and minimum $p_u(n)$) such that for all $p \leq p_l(n)$ (and all $p \geq p_u(n)$) at most 5% of graphs sampled from $\overrightarrow{G}(n, p)$ have $\overrightarrow{\beta}_1(G) > 0$. The figures show log–log plots of these two functions. In both cases, we fit a line of best fit in order to obtain an approximate power-law relationship.

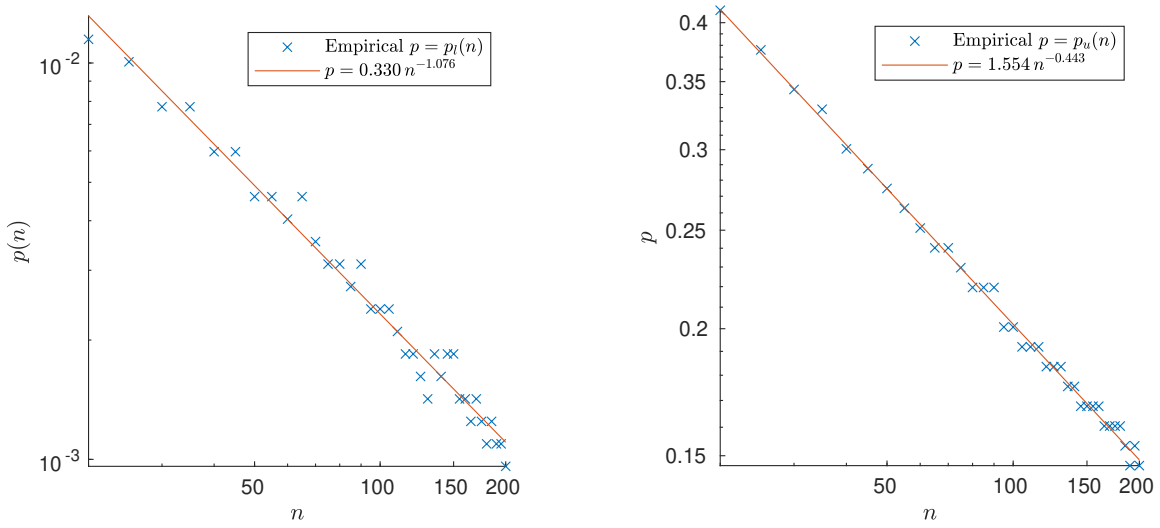


Figure 2.12: As for Figure 2.11 but for 200 samples of $\beta_1(\overrightarrow{X}(G))$.

to support the conjecture that the zero region for path homology can be expanded to $(-\infty, -1) \cup (-2/3, 0]$.

Conjecture 2.B.1. *For an Erdős-Rényi random directed graph $G \sim \overrightarrow{G}(n, p(n))$, let $\overrightarrow{\beta}_1$ denote the 1^{st} Betti number of its non-regular path homology. Assume $p(n) = n^{\alpha}$, if $\alpha > -2/3$ then $\overrightarrow{\beta}_1(G) = 0$ with high probability. The same holds for regular path homology.*

Using the samples from Experiment 1, for each n, we determine the maximum $p_l(n)$ such that for all $p \leq p_l(n)$ we observe $\hat{\mathbb{P}}[\overrightarrow{\beta}_1(G) = 0] \geq 0.95$ for $G \sim \overrightarrow{G}(n, p)$, where $\hat{\mathbb{P}}$ denotes the empirical probability, derived from our sampled distribution. Similarly, using the samples from Experiment 2, we determine the minimum $p_u(n)$ such that for all $p \geq p_u(n)$ we observe $\hat{\mathbb{P}}[\overrightarrow{\beta}_1(G) = 0] \geq 0.95$ for $G \sim \overrightarrow{G}(n, p)$. Since we anticipate a power-law relationship, the logarithmic spacing of p allows us to achieve greater precision as n increases, since the values of $\log(p)$ are evenly spaced. Moreover, we chose the boundaries of the p-region so that precision is greater near the lower boundary in Experiment 1 and near the upper boundary in Experiment 2.

We then compute a least-squares, line of best fit between $\log(p_l(n))$ and $\log(n)$, to obtain a power-law relationship of the form $p_l(n) = An^{\gamma}$. We repeat this for the upper boundary to obtain a similar relationship for $p_u(n)$. The results of this experiment are shown in Figure 2.11.

Figure 2.11a shows that the empirical lower boundary has a similar dependency on n to that predicted by Theorem 2.1.3(2, 3), i.e. $p_l(n) \sim n^{-1}$. Moreover, Figure 2.11a contains a plot of $p_l^t(n)$, as defined in Appendix 2.A.1. For a parameter pair (n, p) falling below this line, Theorem 2.A.1 implies $\mathbb{P}[\overrightarrow{\beta}_1(G) > 0] \leq 0.05$. We observe that this theoretical boundary of significance lies very close to the observed, experimental boundary, indicating that Theorem 2.A.1 is close to the best possible bound.

Conversely, Figure 2.11b predicts an upper boundary of $p_u(n) \sim n^{-0.660}$. This is consistent with Theorem 2.1.34 since $-0.660 < -1/3$, but indicates that there is significant room for improvement. This suggests that the hypothesis of Theorem 2.1.34 can be weakened to $\alpha > -2/3$. However, short of a stronger theoretical result, we require more experiments with graphs on $n > 150$ nodes to confirm this; computational complexity is currently the limiting factor.

In Figure 2.12 we repeat the same analysis with Experiments 3 and 4 respectively, in order to discern the boundaries of the positive region for directed flag complex homology. Again, Figure 2.12a shows that the empirical lower boundary has a similar dependency

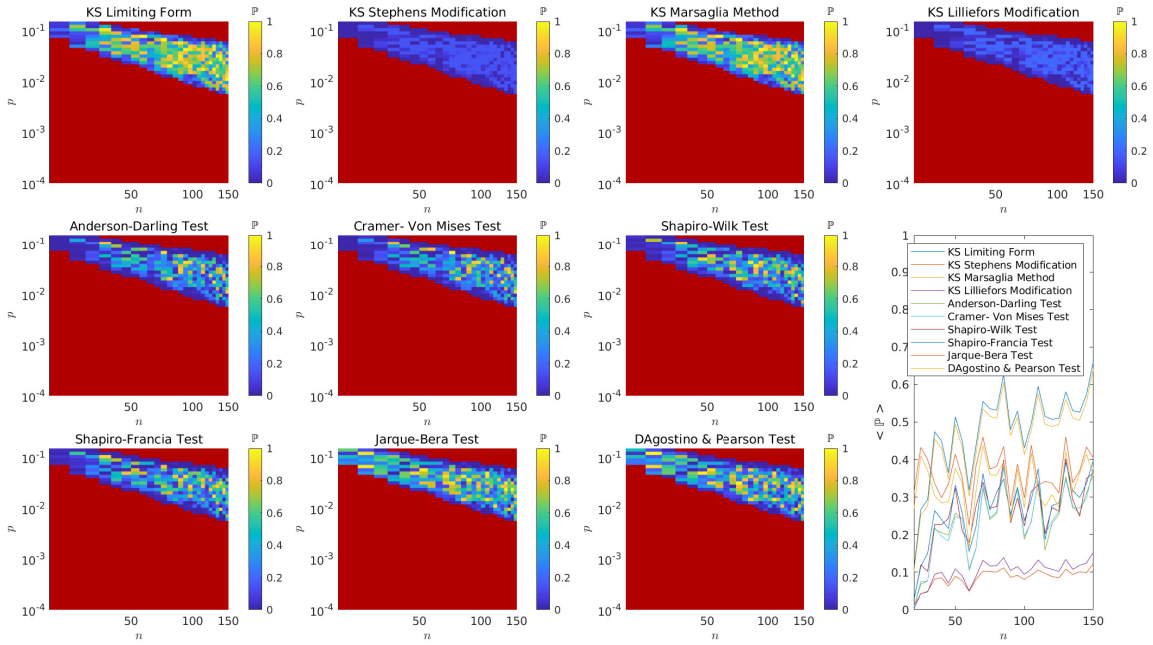


Figure 2.13: Results for 10 normality tests on samples of $\overrightarrow{\beta}_1(G)$ for $G \sim \overrightarrow{G}(n,p)$ at a range of n and p. Red rectangles indicate that at least 5% of samples were zero and hence are excluded from the experiment. Colour indicates the $\mathbb{P}$-value for the hypothesis test in question. Finally, for each test and at each n, we average the $\mathbb{P}$-value of the range of relevant p, which is recorded on the line graph. Note adjacent densities, p, on the horizontal axis are shown with equal width, despite being logarithmically spaced.

on n to that predicted by Theorem 2.1.5(3, 4), i.e. $p_l(n) \sim n^{-1}$. Figure 2.11b shows an upper boundary of $p_u(n) \sim n^{-0.443}$. This is also consistent with Theorem 2.1.55 since $-0.437 < -1/4$. This provides evidence that the zero region for directed flag complex can be expanded as far as $(-\infty, -1) \cup (-1/2, 0]$.

2.B.4 Testing for normal distribution

In analogy to known results for the clique complex [67, Theorem 2.4], one conjecture is that, in the known positive region, the normalised Betti number $\overrightarrow{\beta}_1$ approaches a normal distribution.

Conjecture 2.B.2. *If $G \sim G(n,p)$, where $p = p(n)$ with $p(n) = \omega(n^{-1})$ and $p(n) = o(n^{-2/3})$, then*

$$\frac{\overrightarrow{\beta}_1(G) - \mathbb{E}\left[\overrightarrow{\beta}_1(G)\right]}{\sqrt{\mathrm{Var}\left(\overrightarrow{\beta}_1(G)\right)}} \Longrightarrow \mathcal{N}(0,1) \text{ as } n \to \infty \tag{2.126}$$

where $\mathcal{N}(0,1)$ is the normal distribution with mean 0 and variance 1.

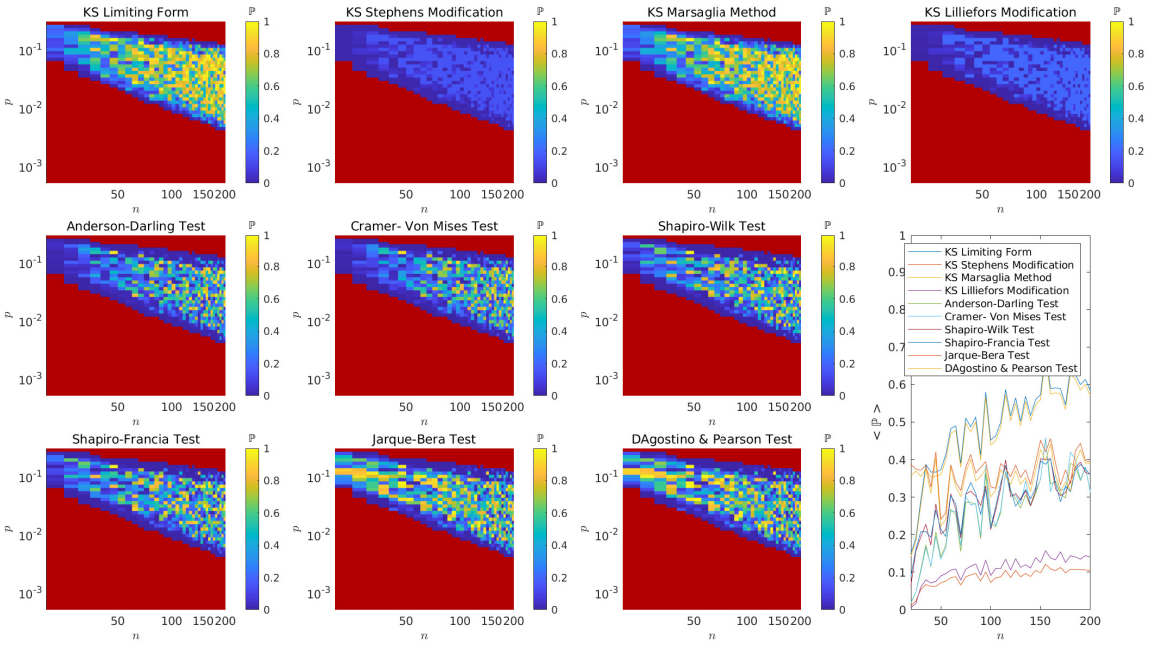


Figure 2.14: As Figure 2.13 but for $\beta_1(\overrightarrow{X}(G))$.

To provide some empirical evidence towards this conjecture, we perform 10 normality tests on the distributions of $\overrightarrow{\beta}_1$, obtained in Experiment 1. We restrict our focus to the samples in which at most 5% of samples were zero, so that we are in a parameter region where we hope our conjecture would apply.

We normalise each of the remaining samples and perform 10 hypothesis tests under the null hypothesis that the samples come from normal distributions. To avoid confusion with the null model parameter, we refer to the significance of these hypothesis tests as $\mathbb{P}$-values. These tests are computed with the `MATLAB` package `normalitytest` [84]. The $\mathbb{P}$-values (and names of the tests) are recorded in Figure 2.13, along with the average $\mathbb{P}$-value against edge-inclusion probability p. In all tests, we see a noisy but consistent trend: there is a decreasing amount of evidence for discarding the null hypothesis as $n \to \infty$.

In Figure 2.14, we repeat this analysis with the distributions of $\beta_1(\overrightarrow{X}(G))$ collected in Experiment 3. Again we observe a similar but stronger trend: there is a decreasing amount of evidence for discarding the null hypothesis as $n \to \infty$.

While no individual test is sufficient to conclude that $\overrightarrow{\beta}_1$ tends towards a normal distribution, the ensemble of tests provide good evidence towards this claim. Larger sample sizes, as well as samples at larger n, are required for more convincing evidence.

Chapter 3

Grounded persistent path homology: a stable, topological descriptor for weighted digraphs

3.1 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. [10, 14, 76, 94]). 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 Figure 3.1, in which edge weights correspond to length as drawn.

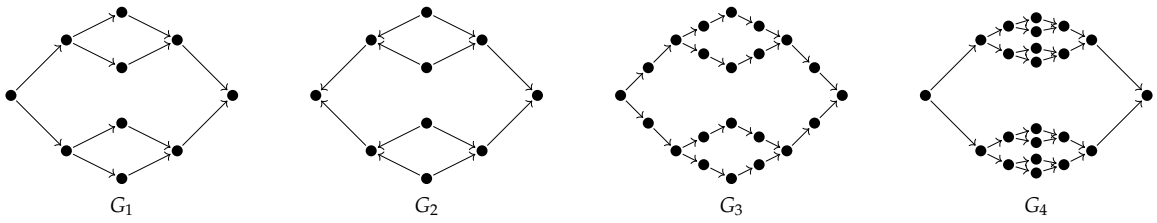


Figure 3.1: Four example weighted digraphs; weights correspond to length as drawn.

The main application motivating this work is the study of blood flow through vascular networks. Flow through such a network can be summarised by a weighted digraph wherein each vessel is represented by an edge, with the direction indicating flow direction and the weight indicating the time it takes for a particle to flow down that vessel. From

the perspective of such a 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. Moreover, consider the sequence of digraphs $(H_n)_{n\in N}$ beginning $(G_1, G_3, \dots)$ where each subsequent digraph is obtained iteratively via the same subdivision procedure, in which the weight of each parent edge is divided equally amongst its children. Ideally $\mathcal{I}(H_n)$ should converge to a limiting value. 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 [46]. 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 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 Section 2 of [18]). Path homology is one such theory [54],

which is sensitive to directionality and has useful functorial properties [32, 55]. Persistent homology is a tool for developing stable descriptors [8] that extract relevant information in multiscale scenarios. As such, persistent homology has seen successful applications to fields including neuroscience [17, 50, 51, 98], vasculature [81, 93] and financial networks [62], to name but a few [47]. A theory of persistent path homology (PPH) was proposed by Chowdhury and Mémoli [32] and is a stable descriptor for directed networks. In a search to develop an interpretable invariant for weighted digraphs, which respects the inherent topology of the underlying graph (viewed as a cell complex), we are lead to an alteration of PPH which we prove meets goals (a)-(d).

3.1.1 Contributions and outline

In Section 3.2 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 Section 3.3.1 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 Section 3.3.2.

Arising from the grounded pipeline, our main contribution is Definition/Theorem 3.3.16, wherein we define grounded persistent path homology (GRPPH). This new invariant is a functor

$$ {}^{g}\mathcal{H}_1 : \mathbf{ContWDgr} \to \mathbf{PersVec} \tag{3.1} $$

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 3.4 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 3.1.1. *Given a weighted digraph $G \in \mathbf{WDgr}$, denote the underlying undirected graph by $\mathcal{U}(G)$, which contains an undirected edge for every directed edge in G, including 2 distinct edges whenever there is a reciprocal pair of edges in 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 $\mathcal{U}(G)$ (i.e. the 1^{st} homology group of $\mathcal{U}(G)$ viewed as a cell complex); 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 Section 3.4.2. In Section 3.4.4, we prove that if G can be decomposed into smaller parts, then GRPPH also decomposes.

Theorem 3.1.2. *Given a weighted digraph $G \in$ **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*

$$ {}^{g}\mathcal{H}_1(G) \cong {}^{g}\mathcal{H}_1(G_1) \oplus {}^{g}\mathcal{H}_1(G_2). \tag{3.2} $$

In Section 3.5 we investigate the stability of ${}^{g}\mathcal{H}_1$. We employ path homotopy theory to prove the main stability theorem (Theorem 3.5.8), 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 3.5.11), edge subdivision (Theorem 3.5.16) and certain classes of edge collapses (Theorem 3.5.25) and edge deletions (Theorem 3.5.31). In particular, edge subdivision stability automatically implies that our invariant converges under iterated subdivision (Corollary 3.5.19). In contrast, the descriptor is unstable to generic edge collapses (Theorem 3.5.28) and edge deletions (Theorem 3.5.35), 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 3.1.

In order to build intuition for what is measured by GRPPH, we compute a number of illustrative examples in Section 3.6. In particular, in Section 3.6.1, we consider a simple, cycle graph and determine the limiting value of GRPPH under iterated edge subdivision. In Section 3.6.2, 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 Section 3.6.5, 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.

3.1.2 Computations

An algorithm for computing PPH in arbitrary degrees was proposed by Chowdhury and Mémoli [32]; a more efficient algorithm for computing PPH in degree 1 was later proposed by Dey, Li, and Wang [38]. We modify the latter algorithm to compute GRPPH; an implementation, capable of computing representatives, is available as a Python package [23].

3.2 Background

3.2.1 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 3.2.1. For $d \in \mathbb{N}$, the standard d-simplex is

$$\Delta^d := \left\{ (x_0, \dots x_d) \in \mathbb{R}^{d+1} \;\middle|\; \sum x_i = 1, \text{ and } x_i \geq 0 \; \forall i \right\}. \tag{3.3}$$

Notation 3.2.2. Given a category $\boldsymbol{\mathcal{C}}$, we denote the collection of objects $\mathrm{Obj}(\boldsymbol{\mathcal{C}})$ and the collection of morphisms $\mathrm{Mor}(\boldsymbol{\mathcal{C}})$. For two objects $X, Y \in \mathrm{Obj}(\boldsymbol{\mathcal{C}})$, we denote the set of morphisms $X \to Y$ by $\mathrm{Mor}_{\boldsymbol{\mathcal{C}}}(X, Y)$. Where it is clear from context whether α is an object or a morphism, we simply write $\alpha \in \boldsymbol{\mathcal{C}}$.

Notation 3.2.3. Fix any three categories $\boldsymbol{\mathcal{C}}$, $\boldsymbol{\mathcal{D}}$ and $\boldsymbol{\mathcal{D}}'$.

(a) We denote the category of functors $\boldsymbol{\mathcal{C}} \to \boldsymbol{\mathcal{D}}$, where morphisms are natural transformations, by $[\boldsymbol{\mathcal{C}}, \boldsymbol{\mathcal{D}}]$.

(b) For a morphism $f \in \mathrm{Mor}_{[\boldsymbol{\mathcal{C}}, \boldsymbol{\mathcal{D}}]}(M, N)$, we denote the components of the natural transformation by $f_x : M(x) \to N(x)$ for each $x \in \boldsymbol{\mathcal{C}}$.

(c) Given a functor $\mu : \boldsymbol{\mathcal{D}} \to \boldsymbol{\mathcal{D}}'$, there is a functor $[\boldsymbol{\mathcal{C}}, \mu] : [\boldsymbol{\mathcal{C}}, \boldsymbol{\mathcal{D}}] \to [\boldsymbol{\mathcal{C}}, \boldsymbol{\mathcal{D}}']$. Given $F \in \mathrm{Obj}([\boldsymbol{\mathcal{C}}, \boldsymbol{\mathcal{D}}])$, we map $[\boldsymbol{\mathcal{C}}, \mu](F) := \mu \circ F$. Given a natural transformation $\nu : F \Rightarrow F'$ between functors $F, F' \in [\boldsymbol{\mathcal{C}}, \boldsymbol{\mathcal{D}}]$, we map $[\boldsymbol{\mathcal{C}}, \mu](\nu) := \mu \circ \nu$ which one can confirm is a natural transformation $\mu \circ F \Rightarrow \mu \circ F'$.

Notation 3.2.4. (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 3.2.5. 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

$$H_k(C) := \frac{\ker \partial_k}{\operatorname{im} \partial_{k+1}}. \tag{3.4}$$

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

3.2.2 Directed graphs

Definition 3.2.6. (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

$$\Delta_V := \{(v, v) \in V \times V \mid v \in V\}. \tag{3.5}$$

(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 $(i, j) \in E \implies i < j$. This is equivalent to G not containing any directed cycles.

(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

$$(i, j) \in E, (j, i) \notin E \implies \{i, j\} \in \mathcal{U}(E) \text{ with multiplicity 1,} \tag{3.6}$$

$$(i, j), (j, i) \in E \implies \{i, j\} \in \mathcal{U}(E) \text{ with multiplicity 2.} \tag{3.7}$$

(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 3.2.7. 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 3.2.8. 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 $v \in V$,

$$\mathcal{N}_{in}(v; G) := \{a \in V \mid a \to v\}, \tag{3.8}$$

$$\mathcal{N}_{out}(v; G) := \{b \in V \mid v \to b\}, \tag{3.9}$$

$$\mathcal{N}(v; G) := \mathcal{N}_{in}(v) \cup \mathcal{N}_{out}(v) \tag{3.10}$$

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 $F \subseteq E$,

$$\mathcal{N}(F; G) := \{\tau \in E \mid \exists\, e \in F \text{ such that } \tau, e \text{ are incident to a common vertex}\} \tag{3.11}$$

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, $e_i \in E$,

$$p = (v_0 = a, e_1, v_1, e_2, \ldots, e_k, v_k = b) \tag{3.12}$$

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, $e_i \in E$,

$$p = (v_0 = a, e_1, v_1, e_2, \ldots, e_k, v_k = a) \tag{3.13}$$

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 3.2.9. 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 3.2.10. Fix a weighted digraph $G = (V, E, w)$.

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

$$\mathrm{len}(p) := \sum_{i=1}^{k} w(e_i). \tag{3.14}$$

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

Definition 3.2.11. For a weighted digraph G, the shortest-path quasimetric $d : V(G) \times V(G) \to \mathbb{R} \sqcup \{\infty\}$ is defined by

$$d(i, j) := \begin{cases} \min_{p \in \mathcal{P}(i,j)} \mathrm{len}(p) & \text{if } \mathcal{P}(i, j) \neq \emptyset, \\ \infty & \text{otherwise.} \end{cases} \tag{3.15}$$

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

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

Definition 3.2.12. (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

$$i \to j \implies f(i) \to f(j) \text{ or } f(i) = f(j). \tag{3.16}$$

Given a vertex map $f : V(G) \to V(H)$ satisfying condition (3.16), 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

$$d_H(f(i), f(j)) \leq d_G(i, j) \tag{3.17}$$

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

Notation 3.2.13. (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 $p = (v_0, e_1, \ldots, e_k, v_k)$ and a digraph map $f : G \to H$, the image of p is the path, $f(p)$, obtained from

$$(f(v_0), f(e_1), f(v_1), \ldots, f(e_k), f(v_k)) \tag{3.18}$$

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 3.2.14. Suppose $f : G \to H$ is a digraph map such that $w(H)(f(e)) \leq w(G)(e)$ 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 $\mathrm{len}(f(p)) \leq \mathrm{len}(p)$. 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 3.2.15. (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 [32] we make the following definition.

Definition 3.2.16. Given a finite set V, a directed network is a non-negative function $A : V \times V \to [0, \infty)$ such that $A(v_1, v_2) = 0 \iff v_1 = v_2$.

Typically, directed networks are a dissimilarity measure, meaning that a high value of $A(x, x')$ represents that x and x' are 'far apart' or 'dissimilar'. A common example is a quasi-metric space, but a directed network need not satisfy the triangle inequality. Note that a directed network could alternatively be defined as a positively-weighted complete digraph.

3.2.3 Path homology

Path homology is a homology theory for directed graphs, which was first introduced by Grigor'yan et al. [54]. Subsequent papers prove Künneth theorems for Cartesian products and joins [52], and invariance under an appropriate notation of digraph homotopy [55]. A directed network gives rise to a natural filtration of digraphs which leads to a stable theory of persistent path homology [32], where the distance between two such networks is defined to be the minimum distortion of a correspondence between their respective domains.. (Persistent) path homology has also been extended to vertex-weighted digraphs [73]. 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 3.2.17. 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 3.2.18. We freely generate R-modules from these sequences of vertices, for each $p \geq 0$.

$$\Lambda_p := \Lambda_p(G;R) := R\langle \{v_0 \dots v_p \text{ elementary } p\text{-path on } V\} \rangle \tag{3.19}$$

$$\mathcal{R}_p := \mathcal{R}_p(G;R) := R\langle \{v_0 \dots v_p \text{ regular } p\text{-path on } V\} \rangle \tag{3.20}$$

$$\mathcal{I}_p := \mathcal{I}_p(G;R) := R\langle \{v_0 \dots v_p \text{ non-regular } p\text{-path on } V\} \rangle \tag{3.21}$$

$$\mathcal{A}_p := \mathcal{A}_p(G;R) := R\langle \{v_0 \dots v_p \text{ allowed } p\text{-path in } G\} \rangle \tag{3.22}$$

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

Definition 3.2.19. Given $p \geq 0$, the non-regular boundary map $\partial_p^{\text{nr}} : \Lambda_p \to \Lambda_{p-1}$ is given on the standard basis by

$$\partial_p^{\text{nr}}(v_0 \dots v_p) := \sum_{i=0}^{p} (-1)^i v_0 \dots \hat{v_i} \dots v_p \tag{3.23}$$

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 3.2.20. 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

$$\partial_p := \pi \circ \partial_p^{\text{nr}} \circ \iota. \tag{3.24}$$

To justify the name, a standard check confirms that indeed $\partial_{p-1} \circ \partial_p = 0$ [54, Lemma 2.9]. However, the $\{\mathcal{R}_p\}$ do not depend on 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 3.2.21. The space of ∂-invariant p-paths is

$$\Omega_p := \Omega_p(G;R) := \{v \in \mathcal{A}_p \mid \partial_p \in \mathcal{A}_{p-1}\}. \tag{3.25}$$

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 3.2.22. The regular path (chain) complex is

$$\dots \xrightarrow{\partial_3} \Omega_2 \xrightarrow{\partial_2} \Omega_1 \xrightarrow{\partial_1} \Omega_0 \xrightarrow{\partial_0} R \xrightarrow{\partial_{-1}} 0 \tag{3.26}$$

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

$$H_k := H_k(G;R) := \frac{\ker \partial_k}{\operatorname{im} \partial_{k+1}}. \tag{3.27}$$

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

Remark 3.2.23. 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 3.2.24. 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 3.2.25. 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

$$f_\#(v_0 \dots v_p) := \begin{cases} f(v_0) \dots f(v_p) & \text{if } f(v_0) \dots f(v_p) \text{ is regular} \\ 0 & \text{otherwise} \end{cases} \tag{3.28}$$

Notation 3.2.26. 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 3.2.27 ([55, Theorem 2.10], [32, 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* **Dgr** $\to$ **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 3.2.28 ([54, §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.*

$$\Omega_0(G) \cong R\langle V \rangle \quad \textit{and} \quad \Omega_1(G) \cong R\langle E \rangle \,. \tag{3.29}$$

Notation 3.2.29. 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 3.2.30. Given a directed trail $p = (e_1, \dots, e_k)$ in a digraph G, the representative of p is

$$\mathfrak{R}(p) := \sum_{i=1}^{k} e_i \in \Omega_1. \tag{3.30}$$

Likewise, there is a representative for any undirected circuit.

Notation 3.2.31. Given an undirected circuit $p = (v_0 = a, e_1, v_1, \ldots, e_k, v_k = a)$ in a digraph G, the representative of p is

$$\mathfrak{R}(p) := \sum_{i=1}^{k} \alpha_i e_i \in \Omega_1 \tag{3.31}$$

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

Remark 3.2.32. The representative of a circuit p, does not depend on the starting point, but if p' traverses the circuit in the opposite direction then $\mathfrak{R}(p') = -\mathfrak{R}(p)$.

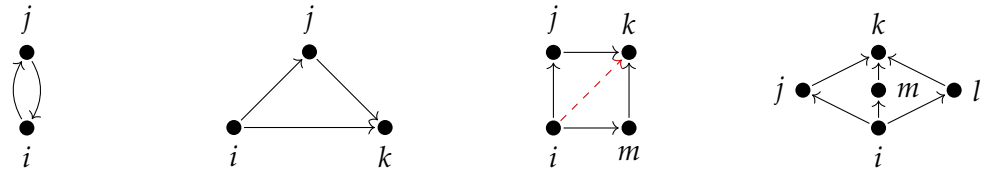


Figure 3.2: The three types of generators for $\Omega_2(G;\mathbb{Z})$ and $\Omega_2(G;\mathbb{R})$: (a) A double edge. (b) A directed triangle. (c) A long square (the dashed red edge must not be present). Finally, (d) shows a linear dependency between long squares.

Proposition 3.2.33 ([55, Proposition 2.9], [38, 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 Figure 3.2):*

(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 3.2.33 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 Figure 3.2d, we see

$$(ijk - ilk) = (ijk - imk) + (imk - ilk). \tag{3.32}$$

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

3.2.4 Path homotopy

Path homotopy, also introduced by Grigor'yan et al. [55], 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 3.2.34. (a) A line digraph of length n is a digraph on $0, 1, \ldots, n$ with exactly one of $(i, i+1)$, $(i+1, i)$ for each i and no other edges.

(b) Denote the two length 1 line digraphs 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

$$V(G \boxdot H) := V(G) \times V(H), \tag{3.33}$$

$$E(G \boxdot H) := \left\{ ((x,y),(x',y')) \;\middle|\; (x \to x' \text{ and } y = y') \text{ or } (x = x' \text{ and } y \to y') \right\}. \tag{3.34}$$

(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

$$F\big|_{G\times\{0\}} = f \quad \text{and} \quad F\big|_{G\times\{n\}} = g. \tag{3.35}$$

(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 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 $i = 1, \ldots, n$. 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 relative $A \subseteq V(G)$ then so too is each F_i.

Grigor'yan et al. [55] showed that path homology is invariant with respect to path homotopy, i.e. path homotopic digraph maps induce identical maps on regular path homology. In [55, 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 3.2.35. *Given a homotopy $F : G \boxdot I \to H$ between two digraph maps $f, g : G \to H$, 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 $c \in \Omega_p(G)$,*

$$g_\#(c) - f_\#(c) = \partial L_p(c) + L_{p-1}\partial(c). \tag{3.36}$$

Proof. Suppose F is a one-step homotopy from f to g, then in the proof of [55, Theorem 3.3] it was shown that there exists a chain homotopy L satisfying equation (3.36). Suppose instead that F is an arbitrary homotopy between f and g. Then, as described above, there is a sequence of digraph maps

$$f = f_0 =, f_1, \ldots, f_n = g \tag{3.37}$$

and a sequence of *one-step* homotopies $F_1, \ldots F_n$ 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,

$$(f_i)_\#(c) - (f_{i-1})_\#(c) = \sigma_i \cdot [\partial L^i_p(c) + L^i_{p-1}\partial(c)]. \tag{3.38}$$

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 equation (3.36) thanks to a telescoping sum. □

Corollary 3.2.36 ([55, 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 3.5.8).

Lemma 3.2.37. *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 3.2.35. 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 $\mathcal{A}_p(G_A)$

$$c = \sum c^{v_0 \ldots v_p} v_0 \ldots v_p \tag{3.39}$$

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 [55, Theorem 3.3], one can show

$$L_p(c) = \sum c^{v_0 \dots v_p} \sum_k (-1)^k \pi[F(v_0) \dots F(v_k) F(v'_k) \dots F(v'_p)]. \tag{3.40}$$

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. □

3.2.5 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 [16, 27, 85] 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 3.2.38. (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 $\mathbf{PersVec} := \mathbf{[R, Vec]}$.

(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 $M \in \mathbf{[R, vec]}$. We denote the category of p.f.d persistent vector spaces $\mathbf{Persvec} := \mathbf{[R, vec]}$.

(d) A persistent vector space $M : \mathbf{R} \to \mathbf{Vec}$ is tame [80, Definition 6.13] if

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

  (ii) there are at most 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

$$P(I)(t) := \begin{cases} \mathbb{R} & \text{if } t \in I, \\ 0 & \text{otherwise,} \end{cases} \tag{3.41}$$

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

$$(M \oplus N)(t) := M(t) \oplus N(t), \tag{3.42}$$

$$(M \oplus N)(s \leq t) := M(s \leq t) \oplus N(s \leq t). \tag{3.43}$$

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

$$N(s \leq t) \circ \phi_s = \phi_t \circ M(s \leq t) \tag{3.44}$$

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 and described by a finite amount of information in all practical applications.

Theorem 3.2.39 (Structure Theorem for p.f.d persistent vector spaces, [36, Theorem 1.1], [28, Theorem 2.8])**.** *Given $M \in$ **Persvec**, there is a multiset $\mathcal{B}M$ of intervals of $\mathbb{R}$ such that*

$$M \cong \bigoplus_{I \in \mathcal{B}M} P(I) \tag{3.45}$$

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

Definition 3.2.40. (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

$$\mathrm{Dgm}(B) := \{\{ (a_k, b_k) \mid I_k \in \mathcal{B} \text{ has endpoints } a_k \leq b_k \}\}. \tag{3.46}$$

(d) Given $M \in$ **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 3.2.41 ([42, §6])**.** Given two multisets $D_1, D_2 \subseteq \mathbb{R}^2$, the bottleneck distance is

$$d_B(D_1, D_2) := \inf_{\gamma} \sup_{x \in D_1 \cup \Delta_{\mathbb{R}}} \|x - \gamma(x)\|_\infty \tag{3.47}$$

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

$$d_B(\mathcal{B}_1, \mathcal{B}_2) := d_B(\mathrm{Dgm}(\mathcal{B}_1), \mathrm{Dgm}(\mathcal{B}_2)). \tag{3.48}$$

Definition 3.2.42 ([34, §3])**.** Given $p \geq 1$ and two multisets $D_1, D_2 \subseteq \mathbb{R}^2$, the p-Wasserstein distance is

$$d_{W_p}(D_1, D_2) := \inf_{\gamma} \left(\sum_{x \in D_1 \cup \Delta_{\mathbb{R}}} \|x - \gamma(x)\|_\infty^p \right)^{1/p} \tag{3.49}$$

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 infinite multiplicity. The p-Wasserstein distance between two barcodes $\mathcal{B}_1$, $\mathcal{B}_2$, is

$$d_{W_p}(\mathcal{B}_1, \mathcal{B}_2) := d_{W_p}(\mathrm{Dgm}(\mathcal{B}_1), \mathrm{Dgm}(\mathcal{B}_2)). \tag{3.50}$$

Definition 3.2.43 ([8, §3.1])**.** 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

$$M[\delta](t) := M(t+\delta) \quad \text{and} \quad M[\delta](s \leq t) := M(s+\delta \leq t+\delta). \tag{3.51}$$

(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

$$\psi[\delta] \circ \phi = \mathcal{T}(M, 2\delta) \quad \text{and} \quad \phi[\delta] \circ \psi = \mathcal{T}(N, 2\delta). \tag{3.52}$$

(e) The interleaving distance of M and N is

$$d_I(M,N) := \inf \{\delta \geq 0 \mid \exists\ \delta\text{-interleaving}\}. \tag{3.53}$$

Remark 3.2.44. 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

$$\phi_t \circ M(s \leq t) = N(s+\delta \leq t+\delta) \circ \psi_s \quad \text{and} \quad \psi_t \circ N(s \leq t) = M(s+\delta \leq t+\delta) \circ \phi_s \tag{3.54}$$

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 3.2.45 (Isometry Theorem [8, Theorem 3.5])**.** *Given p.f.d persistent vector spaces* $M, N \in$ **Persvec**,

$$d_B(\mathcal{B}M, \mathcal{B}N) = d_I(M, N). \tag{3.55}$$

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 [28, §1.1]. 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 $[\boldsymbol{k}]_{\leq} \to$ **Vec**, where $[\boldsymbol{k}]_{\leq}$ is the sub-poset of **R** consisting of the integers $0, \ldots, k$ [80, §6.4].

3.3 Motivation and definition of GrPPH

Firstly, in Section 3.3.1, 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 Section 3.3.2, 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 Section 3.3.3 we fix these choices in order to define our proposed descriptor.

3.3.1 Standard pipeline

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

1. a map $F : \mathrm{Obj}(\mathbf{WDgr}) \to \mathrm{Obj}(\mathbf{[R, Dgr]})$, 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.

$$\mathcal{H}_k : \mathbf{WDgr} \xrightarrow{F} \mathbf{[R, Dgr]} \xrightarrow{[\mathbf{R}, C]} \mathbf{[R, Ch]} \xrightarrow{[\mathbf{R}, H_k]} \mathbf{[R, Vec]} \tag{3.56}$$

Examples of this pipeline include persistent homology of the directed flag complex [74] and persistent path homology [32]. 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 $\mathcal{H}_k : \mathrm{Obj}(\mathbf{WDgr}) \to \mathrm{Obj}(\mathbf{[R, Vec]})$ given by $\mathcal{H}_k := \mathbf{[R}, H_k\mathbf{]} \circ \mathbf{[R}, C\mathbf{]} \circ F$. 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 **WDgr** which makes this pipeline functorial.

Notation 3.3.1. Given $F : \mathrm{Obj}(\mathbf{WDgr}) \to \mathrm{Obj}(\mathbf{[R, Dgr]})$, $G \in \mathbf{WDgr}$ and $s \leq t$ we write

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

Definition 3.3.2. (a) A filtration map is any map $F : \mathrm{Obj}(\mathbf{WDgr}) \to \mathrm{Obj}(\mathbf{[R, InclDgr]})$ 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 **WDgr** such that $\mathrm{Obj}(\mathbf{WDgr}_F) = \mathrm{Obj}(\mathbf{WDgr})$ and

$$\mathrm{Mor}(\mathbf{WDgr}_F) = \{f \in \mathrm{Mor}(\mathbf{WDgr}) \mid f \text{ is } F\text{-compatible}\}. \tag{3.57}$$

Lemma 3.3.3. *Any filtration map $F : \mathrm{Obj}(\mathbf{WDgr}) \to \mathrm{Obj}(\mathbf{[R, InclDgr]})$ induces a functor $F : \mathbf{WDgr}_F \to \mathbf{[R, Dgr]}$, 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 $s \leq t$, both $F(G)(s \leq t)$ and $F(H)(s \leq t)$ are digraph maps induced by the inclusion vertex map. Hence, the following square of morphisms in **Dgr** commutes.

$$\begin{array}{ccc} F^s G & \xrightarrow{f} & F^s H \\ {\scriptstyle F(G)(s\leq t)}\downarrow & & \downarrow{\scriptstyle F(H)(s\leq t)} \\ F^t G & \xrightarrow[f]{} & F^t H \end{array} \tag{3.58}$$

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 3.3.4. 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 $\mathbf{WDgr}_F \to \mathbf{[R, Dgr]}$ and hence $\mathcal{H}_k$ is a functor $\mathbf{WDgr}_F \to \mathbf{[R, Vec]}$, 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 3.3.5. The shortest-path filtration is a map $F_d : \mathbf{WDgr} \to \mathbf{[R, Dgr]}$. For $G = (V, E, w) \in \mathbf{WDgr}$ and $t \in \mathbb{R}$, we define

$$F_d(G)(t) := G^t := (V, E^t) \quad \text{where} \quad E^t := \{(i,j) \in (V \times V) \setminus \Delta_V \mid d(i,j) \leq t\} \tag{3.59}$$

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.

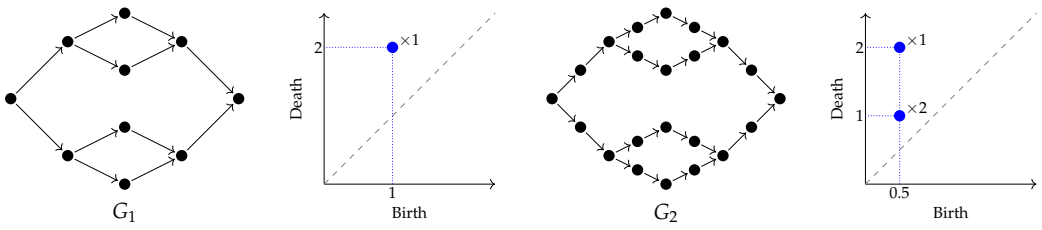


Figure 3.3: Persistent path homology of the shortest-path filtration of flow through a bifurcation network, before and after edge subdivision. In G_1 all edges have unit weight, in G_2 all edges have weight 0.5.

Example 3.3.6. Choosing $F = F_d$ as above and C to be the regular path complex, we obtain a functor $\mathcal{H}_k : \mathbf{WDgr}_F \to \mathbf{[R, Vec]}$. 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 [32] for cycle networks, alongside a stability analysis of persistent path homology for arbitrary directed networks.

In Figure 3.3 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 Figure 3.3 since, in each case, there is some T such that $G_i^T = G_i$ and moreover $E(G_i^t) = \emptyset$ for all $t < T$.

3.3.2 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 $C \in$ **[Dgr, Ch]** to construct a new functor ${}^gC_F : \mathbf{WDgr}_F \to$ **[R, Ch]**.

$$\mathbf{WDgr} \xrightarrow{F} \mathbf{[R,\ Dgr]} \xrightarrow{[\mathbf{R}, C]} \mathbf{[R,\ Ch]} \xrightarrow{[\mathbf{R}, H_1]} \mathbf{[R,\ Vec]} \qquad ({}^gC_F \text{ dashed})$$ (3.60)

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

Definition 3.3.7. Given a filtration functor $F : \mathbf{WDgr}_F \to$ **[R, Dgr]**, a functor $C :$ **InclDgr** $\to$ **Ch**, $G \in$ **WDgr** and $t \in \mathbb{R}$, the chain complex ${}^gC_\bullet(G, t; F)$ is the top row of the following diagram.

$$\cdots C_3(F^tG) \xrightarrow{\partial_3} C_2(F^tG) \xrightarrow{\iota_\# \circ \partial_2} C_1(G \cup F^tG) \xrightarrow{\partial_1} C_0(G \cup F^tG) \cdots$$
$$C_1(F^tG) \xrightarrow{\partial_1} C_0(F^tG) \qquad (\partial_2 : C_2(F^tG) \to C_1(F^tG),\ \iota_\# : C_1(F^tG) \to C_1(G \cup F^tG),\ \iota_\# : C_0(F^tG) \to C_0(G \cup F^tG))$$ (3.61)

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 ${}^gC_k(G, t; F)$ 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 ${}^g\square$ to denote that this chain complex is *grounded*; as we will show in Lemma 3.4.2, after appropriate choices of F and C, all degree 1 homology classes have representatives in the underlying digraph.

Lemma 3.3.8. *For each $t \in \mathbb{R}$ and $G \in \mathbf{WDgr}$, $({}^{g}C_\bullet(G,t;F), {}^{g}\partial^t_\bullet)$ is a chain complex.*

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

$$\partial_1 \circ (\iota_\# \circ \partial_2) = \iota_\# \circ \partial_1 \circ \partial_2 = 0 \tag{3.62}$$

and hence $C_\bullet(G,t;F)$ defines a chain complex. All other adjacent pairs of boundary maps come entirely from $C(F^tG)$ or $C(G \cup F^tG)$ and hence compose to 0. □

Lemma 3.3.9. *Fix a filtration functor $F : \mathbf{WDgr}_F \to$ [**R**, **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*

$$J(f, s \le t) : {}^{g}C_\bullet(G,s;F) \to {}^{g}C_\bullet(G,t;F). \tag{3.63}$$

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

$$J(f \circ g, r \le t) = J(f, s \le t) \circ J(g, r \le s). \tag{3.64}$$

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

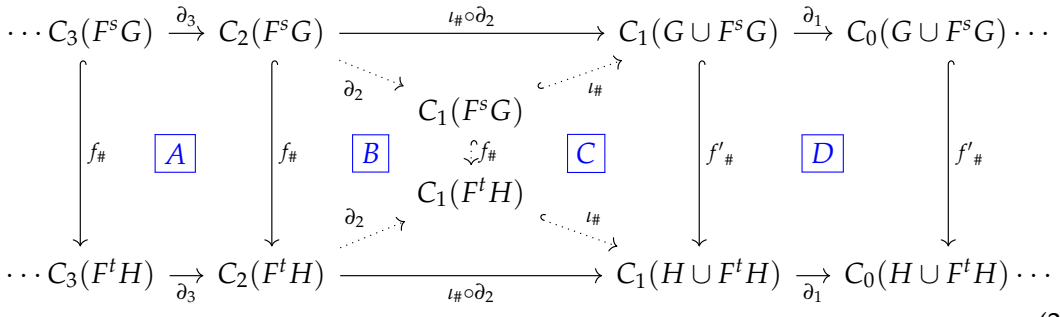
Proof. To distinguish the two digraph maps induced by f, we denote them

$$f : F^sG \to F^tH \quad \text{and} \quad f' : G \cup F^sG \to H \cup F^tH. \tag{3.65}$$

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

$$\begin{array}{ccc} F^sG & \xhookrightarrow{\iota} & G \cup F^sG \\ {\scriptstyle f}\downarrow & & \downarrow{\scriptstyle f'} \\ F^tH & \xhookrightarrow[\iota]{} & H \cup F^tH \end{array} \tag{3.66}$$

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.



(3.67)

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 $J(f, s \leq t)$. The remaining functorial relations follow automatically because in each degree $J(f, s \leq t)$ coincides with either $f_\#$ or $f'_\#$, which each satisfy the relations. □

Lemma 3.3.10. *Given a filtration functor F : $\mathbf{WDgr}_F \to$ [**R**, **Dgr**] and a functor C : **InclDgr** $\to$ **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 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 3.3.9, 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 3.3.9 implies that

$$J(\mathrm{id}, s \leq r) = J(\mathrm{id}, t \leq r) \circ J(\mathrm{id}, s \leq t) \tag{3.68}$$

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 3.3.11. Given a filtration functor F : $\mathbf{WDgr}_F \to$ **[R, Dgr]** and a functor C : **InclDgr** $\to$ **Ch**, we define the functor ${}^gC_F$: **WDgr** $\to$ **[R, Ch]** as follows. The map on objects is given by

$${}^gC_F(G)(t) := ({}^gC_\bullet(G, t; F), {}^g\partial^t_\bullet) \tag{3.69}$$

and the morphism ${}^gC_F(G)(s \leq t)$ is as constructed in the proof of Lemma 3.3.10.

Notation 3.3.12. (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 3.3.10, ${}^gC_F$ 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 3.3.13. *Given a filtration functor F : $\mathbf{WDgr}_F \to$ [**R**, **Dgr**] and a functor C : **Dgr** $\to$ **Ch**, ${}^gC_F$ is a functor ${}^gC_F$: $\mathbf{WDgr}_F \to$ [**R**, **Ch**].*

Proof. Given $f \in \mathrm{Mor}_{\mathbf{WDgr}_F}(G, H)$ and $t \in \mathbb{R}$, we need a chain map ${}^g f_\# : {}^g C_\bullet(G, t) \to {}^g C_\bullet(H, t)$. 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 $F^t G \to F^t H$. Therefore, f must also induce digraph maps $G \cup F^t G \to H \cup F^t H$. Hence, using Lemma 3.3.9, we can take ${}^g f_\# := J(f, t \leq t)$. The functorial axioms follow immediately from the functorial relations shown in Lemma 3.3.9. Moreover, since ${}^g \iota(s, t)_\#$ is also constructed via Lemma 3.3.9, equation (3.64) implies

$$ {}^g f_\# \circ {}^g \iota(s, t)_\# = J(f \circ \mathrm{id}_{V(G)}, s \leq t) = J(\mathrm{id}_{V(H)} \circ f, s \leq t) = {}^g \iota(s, t)_\# \circ {}^g f_\# \tag{3.70} $$

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

Notation 3.3.14. 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 ${}^g f_* := [\mathbf{R},\ H_k]({}^g f_\#)$.

3.3.3 Definition of GRPPH

In order to investigate properties and stability of the grounded pipeline, we make choices 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 **Dgr** $\to$ **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**,

$$ F = F_d \quad \text{and} \quad C = \Omega. \tag{3.71} $$

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

Lemma 3.3.15. *The F_d-compatible category of weighted digraphs is the contraction category of weighted digraphs (see Definition 3.2.15),* $\mathbf{WDgr}_{F_d} = \mathbf{ContWDgr}$.

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

$$ d(f(i), f(j)) \leq d(i, j) \tag{3.72} $$

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 3.3.13 yields a functor ${}^{g}C : \mathbf{ContWDgr} \to \mathbf{[R, Ch]}$. Taking the first homology yields a persistent vector space in a functorial way; this functor is our proposed invariant for weighted digraphs.

Definition/Theorem 3.3.16. Grounded persistent path homology (GRPPH) is the functor

$$ {}^{g}\mathcal{H}_1 := \mathbf{[R}, H_1\mathbf{]} \circ {}^{g}C : \mathbf{ContWDgr} \to \mathbf{PersVec} \tag{3.73} $$

from the contraction category of weighted digraphs (see Definition 3.2.15) to the category of persistent vector spaces (see Definition 3.2.38).

Notation 3.3.17. Given $G \in \mathbf{WDgr}$, in degree k at filtration step t, we denote

$$ \text{the space of grounded } k\text{-cycles by} \qquad {}^{g}Z(G,t) := \ker {}^{g}\partial_k^t; \tag{3.74} $$

$$ \text{the space of grounded } k\text{-boundaries by} \qquad {}^{g}B(G,t) := \mathrm{im}\, {}^{g}\partial_{k+1}^t \tag{3.75} $$

$$ \text{the (degree } k\text{) grounded homology by} \qquad {}^{g}H_k(G,t) := \frac{{}^{g}Z_k(G,t)}{{}^{g}B_k(G,t)}. \tag{3.76} $$

Remark 3.3.18. Note that for any $t \in \mathbb{R}$,

$$ k > 1 \implies {}^{g}H_k(G,t) \cong H_k(G^t), $$
$$ k < 1 \implies {}^{g}H_k(G,t) \cong H_k(G \cup G^t). $$

Therefore, the only new homology occurs in degree $k = 1$, since it compares 1-cycles in $G \cup F^t G$ with 1-boundaries from G^t. This justifies our focus on degree 1 homology in Definition/Theorem 3.3.16.

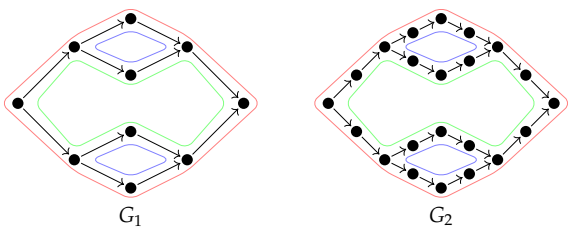


Figure 3.4: A bifurcation network, before and after subdivision; all edges in G_1 have unit weight, all edges in G_2 have weight 0.5 (as in Figure 3.3). Highlighted in red, green and blue are circuits whose representatives generate ${}^{g}\mathcal{H}_1(G_i)$.

Example 3.3.19. In Figure 3.4, we consider again the bifurcating network example of Figure 3.3, in which all edges have weight 1. We see the barcodes of the grounded persistent homology are both

$$\mathcal{B}^g\mathcal{H}_1(G_1) = \mathcal{B}^g\mathcal{H}_1(G_2) = \{\{\, [0,1), [0,1), [0,2) \,\}\}\,. \tag{3.77}$$

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).

3.4 Interpretation of GrPPH

3.4.1 Decreasing Betti curves

In the first example we considered (Example 3.3.19), we saw that all features were born at $t = 0$. Indeed, this is always the case and ${}^gH_1(G, 0)$ is in fact the cycle space of $\mathcal{U}(G)$ (with $\mathbb{R}$-valued coefficients).

Lemma 3.4.1. *Given a digraph $G = (V, E, w)$, two distinct nodes $a, b \in V$ and a trail $p : a \rightsquigarrow b$, then for all $t \geq \mathrm{len}(p)$*

$$\mathfrak{R}(p) = \sum_{\tau \in E(p)} \tau = ab \pmod{{}^gB_1(G,t)}. \tag{3.78}$$

Proof. Fix arbitrary $t \geq \mathrm{len}(p)$ and denote the vertices of the path as $a = v_0, \ldots, v_m = b$. 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 $v_iv_jv_k \in C_2(G^t)$ and hence $v_iv_k = v_iv_j + v_jv_k \pmod{{}^gB_1(G,t)}$. Therefore, inductively we can write

$$ab = v_0v_m = v_0v_1 + v_1v_m = v_0v_1 + v_1v_2 + v_2v_m = \cdots = \sum_{i=1}^{m} v_{i-1}v_i \pmod{{}^gB_1(G,t)} \tag{3.79}$$

as required. □

Proposition 3.4.2. *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 $(\tau_1, \ldots, \tau_m)$, Lemma 3.4.1 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 3.4.3. *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 3.4.2. 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

$$\cdots 0 \longrightarrow 0 \longrightarrow C_1(G) \longrightarrow C_0(G) \cdots \tag{3.80}$$

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 $e_1, \ldots, e_k$. 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 $\{\mathfrak{R}(p_1), \ldots, \mathfrak{R}(p_k)\}$ is a basis for ${}^{g}H_1(G,0)$. □

3.4.2 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 3.4.4. Given $v \in {}^{g}Z_1(G,0)$, the death-time of v is

$$\mathcal{D}(v) := \inf\left\{t \geq 0 \;\middle|\; {}^{g}\iota(0,t)_{*}[v] = 0\right\} \tag{3.81}$$

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 3.4.5. 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 3.4.6. *Given $G = (V, E, w) \in \mathbf{WDgr}$ 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*

$$h_i := \min \{t \geq 0 \mid \exists v_i \in V \text{ along } p_i \text{ such that } d(a, v_i) \leq t \text{ and } d(v_i, b) \leq t\}. \tag{3.82}$$

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 $av_1b - av_2b \in {}^gC_2(G, T)$. By Lemma 3.4.1,

$$0 = {}^g\partial_2(av_1b - av_2b) = \mathfrak{R}(p_c) \pmod{{}^gB_1(G, T)}. \tag{3.83}$$

Finally, if p_1 contains one edge then $h_1 = \text{len}(p_1)$ 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 $av_2b \in {}^gC_2(G, T)$. By Lemma 3.4.1,

$$0 = {}^g\partial_2(av_2b) = \mathfrak{R}(p_c) \pmod{{}^gB_1(G, T)} \tag{3.84}$$

which concludes the proof. □

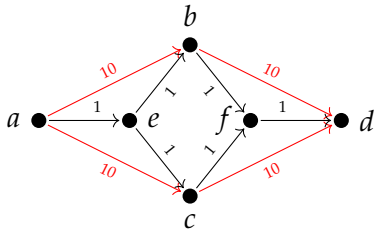


Figure 3.5: A simple example of a weighted digraph for which the bound of Lemma 3.4.6 fails to be sharp.

Example 3.4.7. Note that the bound of Lemma 3.4.6 is by no means sharp. Consider for example Figure 3.5. 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 Figure 3.5. 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

$$ {}^{g}\partial_2(abd - acd) = \mathfrak{R}(p_c). \tag{3.85}$$

3.4.3 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 [71] and these representatives are frequently used for interpreting features (e.g. [9, 92]).

Definition 3.4.8. A persistence basis for ${}^{g}\mathcal{H}_1(G)$ is a choice of initial cycles $\{b_i \in {}^{g}Z_1(G,0)\}$ such that for each $t \geq 0$, the set $\left\{{}^{g}\iota(0,t)_*[b_i]\right\} \setminus \{0\}$ yields a basis for ${}^{g}H_1(G,t)$. We call elements of a persistence basis representatives.

Lemma 3.4.9. *Given any $G \in$ **WDgr**, a persistence basis for ${}^{g}\mathcal{H}_1(G)$ always exists.*

Proof. This follows from the structure theorem (Theorem 3.2.39) and Corollary 3.4.3. □

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

$$ \bigoplus_{i=1}^{m} P(I_i) \cong {}^{g}\mathcal{H}_1(G). \tag{3.86}$$

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

$$ 1 \in I_i(t) \mapsto {}^{g}\iota(0,t)_*[b_i] \text{ whenever } t \in I_i. \tag{3.87}$$

This mapping gives an isomorphism because $\left\{{}^{g}\iota(0,t)_*[b_i]\right\} \setminus \{0\}$ 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 3.4.10. *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 ${}^{g}\mathcal{H}_1(G)$.*

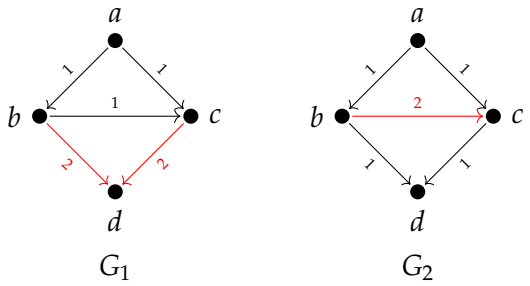


Figure 3.6: Two weighted digraphs with the same underlying digraph but different persistence bases, illustrating that not every circuit basis of ${}^{g}H_1(G,0)$ yields a persistence basis for ${}^{g}\mathcal{H}_1(G)$.

Example 3.4.11. First, we note that it does not suffice to choose *any* basis of undirected circuits for ${}^{g}Z_1(G,0)$. For example, consider the two weighted digraphs pictured in Figure 3.6. In both digraphs, ignoring choice of direction, there are three undirected simple circuits, whose representatives we denote

$$\gamma_1^i := ab + bc - ac, \tag{3.88}$$

$$\gamma_2^i := bc + cd - bd, \tag{3.89}$$

$$\gamma_3^i := ab + bd - cd - ac \tag{3.90}$$

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

$$\mathcal{B}^{g}\mathcal{H}_1(G_1) = \{\{\,[0,2),[0,1)\,\}\} \quad \text{and} \quad \mathcal{B}^{g}\mathcal{H}_1(G_2) = \{\{\,[0,2),[0,1)\,\}\}\,. \tag{3.91}$$

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 3.4.10, 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.

$$\begin{array}{ccccccc}
\cdots C_2(G^{t_1}) & \longrightarrow & C_1(G \cup G^{t_1}) & \longrightarrow & C_0(G \cup G^{t_1}) \cdots & & {}^{g}H_1(G,t_1) \\
\downarrow & & \downarrow & & \downarrow & & \downarrow {\scriptstyle {}^{g}\iota(t_1,t_2)_*} \\
\cdots C_2(G^{t_2}) & \longrightarrow & C_1(G \cup G^{t_2}) & \longrightarrow & C_0(G \cup G^{t_2}) \cdots & & {}^{g}H_1(G,t_2) \\
\downarrow & & \downarrow & & \downarrow & & \downarrow {\scriptstyle {}^{g}\iota(t_2,t_3)_*} \\
\vdots & & \vdots & & \vdots & & \vdots \\
\downarrow & & \downarrow & & \downarrow & & \downarrow {\scriptstyle {}^{g}\iota(t_{m-1},t_m)_*} \\
\cdots C_2(G^{t_m}) & \longrightarrow & C_1(G \cup G^{t_m}) & \longrightarrow & C_0(G \cup G^{t_m}) \cdots & & {}^{g}H_1(G,t_m)
\end{array} \tag{3.92}$$

To the right of the chain complex we show the induced maps on homology ${}^{g}\iota(t_{i-1},t_i)_*$. By Lemma 3.4.2, these maps on homology are always surjective. Our strategy is to find undirected circuit bases for the kernel of each of these maps; Lemma 3.4.12 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 3.4.12. *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 $w_e := av_eb$ is a new generator of $C_2(G^s)$. Repeating this for all new edges we obtain a set of generators

$$U_{new} := \{w_e \mid e \in E_{new}\} \tag{3.93}$$

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

$$C_2(G^r) \oplus W_0 \xrightarrow{{}^g\partial_2} C_1(G \cup G^s) \xrightarrow{{}^g\partial_1} C_0(G \cup G^s) \tag{3.94}$$

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

Claim 3.4.12.1. *The inclusion chain map*

$$\begin{array}{ccccc} C_2(G^r) & \longrightarrow & C_1(G \cup G^r) & \longrightarrow & C_0(G \cup G^r) \\ \downarrow{\scriptstyle q_2^0} & & \downarrow{\scriptstyle q_1^0} & & \downarrow{\scriptstyle q_0^0} \\ C_2(G^r) \oplus W_0 & \longrightarrow & C_1(G \cup G^s) & \longrightarrow & C_0(G \cup G^s) \end{array} \qquad \begin{array}{c} {}^gH_1(G, r) \\ {\scriptstyle q^0}\downarrow{\scriptstyle \cong} \\ Q_0 \end{array} \tag{3.95}$$

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

$$c_k(v) := \begin{cases} v & \text{if } v \in {}^gC_k(G, r), \\ 0 & \text{otherwise,} \end{cases} \tag{3.96}$$

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

$$c_1(ab) := \begin{cases} av_{(a,b)} + v_{(a,b)}b & \text{if } (a, b) \in E_{new}, \\ ab & \text{otherwise.} \end{cases} \tag{3.97}$$

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

$$P(ab) := \begin{cases} w_{(a,b)} & \text{if } (a, b) \in E_{new}, \\ 0 & \text{otherwise.} \end{cases} \tag{3.98}$$

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 $av_e + v_eb$ in $G \cup G^r$. The homotopy P shows the two elements are homologous thanks to the presence of the directed triangle $w_e \in U_{new}$. □ *of Claim.*

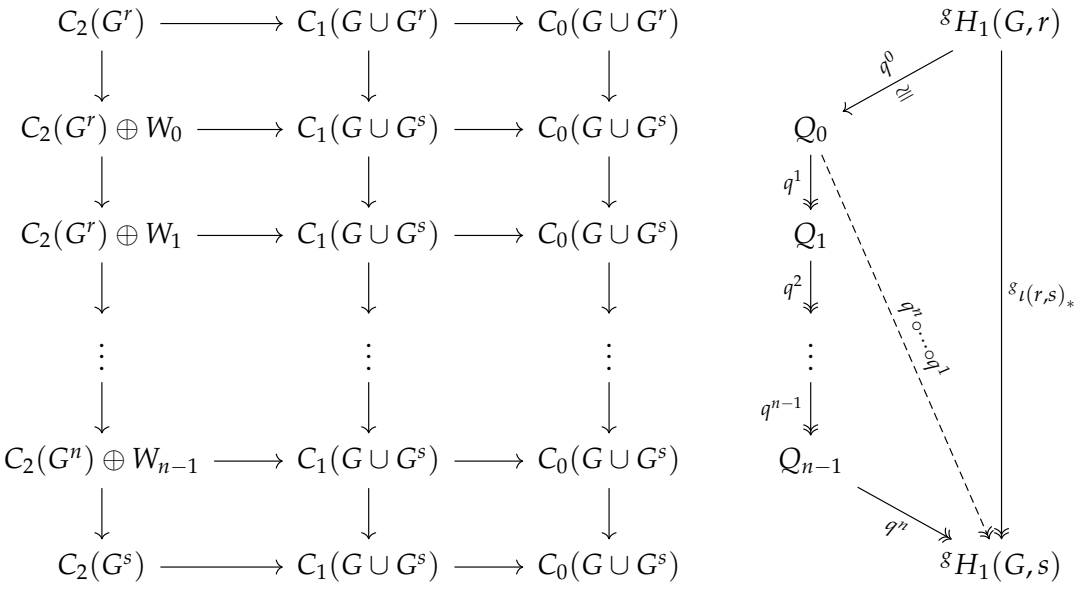
By Proposition 3.2.33, there exists $u_1, \ldots, u_n \in C_2(G^s)$ such that

$$C_2(G^s) = C_2(G^r) \oplus W_0 \oplus \langle u_1, \ldots, u_n \rangle \tag{3.99}$$

where each u_i is amongst the generators identified in Proposition 3.2.33. Define $W_i := W_0 \oplus \langle u_1, \ldots, u_i \rangle$. This gives a sequence of inclusion chain maps

$$\begin{array}{ccccccc} C_2(G^r) \oplus W_{i-1} & \longrightarrow & C_1(G \cup G^s) & \longrightarrow & C_0(G \cup G^s) & \qquad & Q_{i-1} \\ \downarrow{\scriptstyle q_2^i} & & \downarrow{\scriptstyle q_1^i} & & \downarrow{\scriptstyle q_0^i} & & \downarrow{\scriptstyle q^i} \\ C_2(G^r) \oplus W_i & \longrightarrow & C_1(G \cup G^s) & \longrightarrow & C_0(G \cup G^s) & & Q_i \end{array} \tag{3.100}$$

where each row 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.



(3.101)

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.

$$J := \{i \in \mathbb{N} \mid \dim Q_i = \dim Q_{i-1} - 1\}. \tag{3.102}$$

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 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, $\{(q^0)^{-1}[b_i] \mid i \in J\}$ forms a basis for $\ker {}^g\iota(r,s)_*$. It now remains to prove that each $(q^0)^{-1}[b_i]$ has an undirected circuit representative.

Claim 3.4.12.2. *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 3.4.1, 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 $\tau = \mathfrak{R}(T_\tau) \pmod{{}^gB_1(G,r)}$. Concatenating the T_τ we obtain an undirected circuit p_i through G such that $\mathfrak{R}(p) = \mathfrak{R}(p_i) \pmod{{}^gB_1(G,r)}$. □ *of Claim.*

This claim concludes the proof. □

Remark 3.4.13. 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.

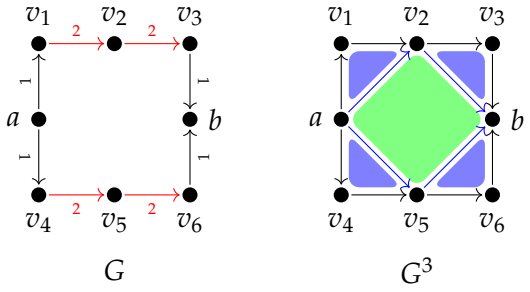


Figure 3.7: An example weighted digraph with a single feature which dies at $t = 3$. To the right we show G^3, colouring the new edges in G^3 in blue and highlighting the new generators in $C_2(G^3)$ via blue and green polygons.

Example 3.4.14. To illustrate how the generators are added in the proof of Lemma 3.4.12, consider Figure 3.7. First note that there is a single feature with representative

$$\gamma := (av_1 + v_1v_2 + v_2v_3 + v_3b) - (av_4 + v_4v_5 + v_5v_6 + v_6b) \tag{3.103}$$

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 $\{[{}^g\partial_2 u_1]\}$.

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 3.4.15. *Given $B_i \subseteq {}^gH_1(G,0)$ 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 $\ker {}^g\iota(0,t_m)_*$.*

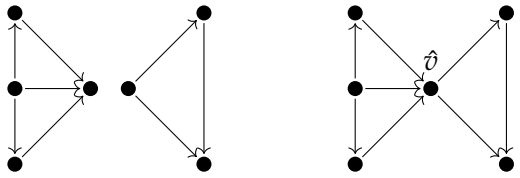
Certainly a basis of undirected circuits for ${}^gH_1(G,0)$ exists (by Corollary 3.4.3). Therefore, we can always extend linearly independent undirected circuits to a basis of such circuits for ${}^gH_1(G,0)$.

Lemma 3.4.16. *Given undirected circuits $p_1, \ldots, p_k$ such that $[\mathfrak{R}(p_1)], \ldots, [\mathfrak{R}(p_k)]$ are linearly independent in ${}^gH_1(G,0)$, there exists undirected circuits $p_{k+1}, \ldots, p_N$ such that $\{[\mathfrak{R}(p_i)]\}_{i=1}^N$ is a basis for ${}^gH_1(G,0)$.*

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

Proof of Theorem 3.4.10. Using Lemma 3.4.12 and Lemma 3.4.15, we obtain undirected circuits $p_1, \ldots, p_l$ such that $\{\mathfrak{R}(p_1), \ldots, \mathfrak{R}(p_l)\}$ forms a basis for $\ker {}^g\iota(0,t_m)_*$. Using Lemma 3.4.16, we extend this to a basis $\{\mathfrak{R}(p_1), \ldots, \mathfrak{R}(p_l), \mathfrak{R}(p_{l+1}), \ldots, \mathfrak{R}(p_N)\}$ for ${}^gH_1(G,0)$. This is a persistence basis for ${}^g\mathcal{H}_1(G)$. □

Remark 3.4.17. 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 3.4.11 and Figure 3.6. Two possible persistence bases for ${}^g\mathcal{H}_1(G_2)$ are $\{\gamma_1^2, \gamma_3^2\}$ and $\{\gamma_2^2, \gamma_3^2\}$. The non-uniqueness of the basis arises in the proof of Lemma 3.4.12. 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.



$G_1 \sqcup G_2$ $\qquad\qquad$ $G_1 \vee_{\hat{v}} G_2$

Figure 3.8: Illustrations of the disjoint union and wedge decomposition considered in Section 3.4.4.

3.4.4 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.

Theorem 3.4.18. *Suppose $G \in$ **WDgr** decomposes as a disjoint union, $G = G_1 \sqcup G_2$, then*

$$ {}^{g}\mathcal{H}_1(G) \cong {}^{g}\mathcal{H}_1(G_1) \oplus {}^{g}\mathcal{H}_1(G_2). \tag{3.104} $$

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

$$ G^t = G_1^t \sqcup G_2^t \quad \text{and} \quad G \cup G^t = (G_1 \cup G_1^t) \sqcup (G_2 \cup G_2^t). \tag{3.105} $$

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 $k \geq 0$, ${}^{g}C_k(G, t)$ splits as direct sum ${}^{g}C_k(G_1, t) \oplus {}^{g}C_k(G_2, t)$. The boundary operator respects this split, mapping ${}^{g}C_k(G_i, t) \to {}^{g}C_{k-1}(G_i, t)$ and the maps ${}^{g}\iota(s, t)_{\#}$ also respect this split, mapping ${}^{g}C(G_i, s) \to {}^{g}C(G_i, t)$. Taking homology in degree 1 maintains this direct sum decomposition. □

Definition 3.4.19. (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

$$ V = V_1 \cup V_2 \tag{3.106} $$

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 3.4.20. 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 3.4.21. *For a weighted digraph $G = (V, E, w)$ and a wedge decomposition $G = G_1 \vee_{\hat{v}} G_2$,*

$$ {}^g\mathcal{H}_1(G) \cong {}^g\mathcal{H}_1(G_1) \oplus {}^g\mathcal{H}_1(G_2). \tag{3.107} $$

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

$$ f_i(v) := \begin{cases} v & \text{if } v \in V_i, \\ \hat{v} & \text{otherwise.} \end{cases} \tag{3.108} $$

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

$$ J : {}^gC(G_1) \oplus {}^gC(G_2) \to {}^gC(G), \quad J(\gamma_1, \gamma_2) := {}^g j_{1\#}\gamma_1 + {}^g j_{2\#}\gamma_2; \tag{3.109} $$

$$ F : {}^gC(G) \to {}^gC(G_1) \oplus {}^gC(G_2), \quad F(\gamma) := ({}^g f_{1\#}\gamma, {}^g f_{2\#}\gamma). \tag{3.110} $$

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

Claim 3.4.21.1. *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

$$ F \circ J = \begin{pmatrix} {}^g\mathrm{id}_{1\#} & {}^g c_{2\#} \\ {}^g c_{1\#} & {}^g\mathrm{id}_{2\#} \end{pmatrix}. \tag{3.111} $$

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 ${}^gC(G_1) \oplus {}^gC(G_2)$. □ *of Claim.*

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 3.4.21.2. *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

$$\partial_2(a\hat{v}b) = a\hat{v} + \hat{v}b - (ab) = (J \circ F)(e) - e. \tag{3.112}$$

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

$$P(ab) := \begin{cases} a\hat{v}b & \text{if } a, b \text{ are separated,} \\ 0 & \text{otherwise.} \end{cases} \tag{3.113}$$

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 ${}^g\mathcal{H}_1(G)$. □ *of Claim.*

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

3.5 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 $\mathbb{O}_\theta^T : \mathrm{Obj}(\mathbf{WDgr}) \to \mathrm{Obj}(\mathbf{WDgr})$ 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

$$d_B(\mathcal{B}^g\mathcal{H}_1(G), \mathcal{B}^g\mathcal{H}_1(\mathbb{O}_\theta^T G)) \leq f(G,\theta) \tag{3.114}$$

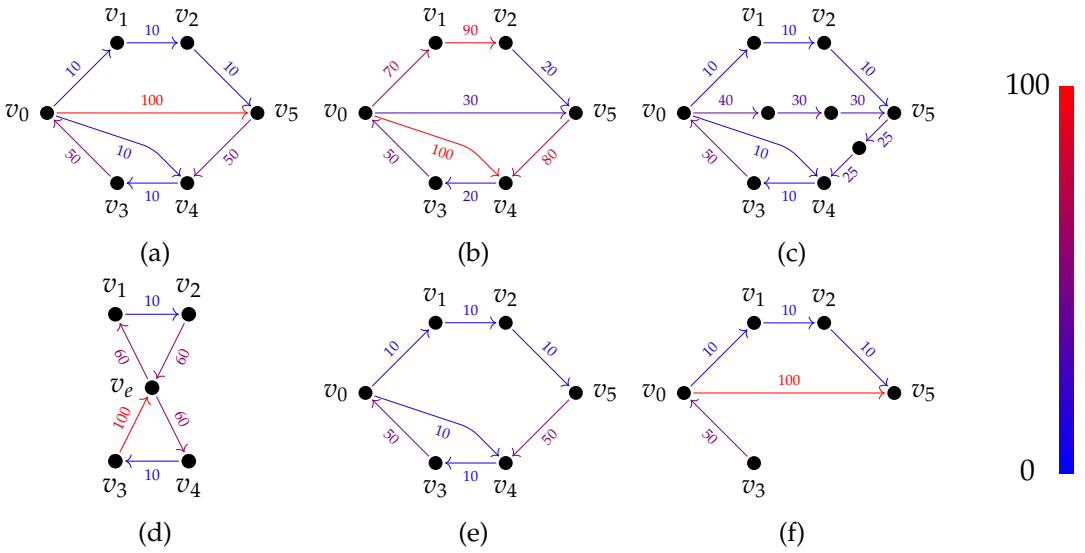


Figure 3.9: Illustration of a number of operations for altering a weighted digraph. For formal definitions, see the corresponding subsections of Section 3.5. **(a)** The initial weighted digraph G. **(b)** Weight perturbation $\mathbb{O}^p_{w'}G$. **(c)** Edge subdivision $\mathbb{O}^s_S G$ where the 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 $v = v_4$.

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 (3.114) 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 ${}^g\mathcal{H}_1(G) \cong {}^g\mathcal{H}_1(\mathbb{O}^T_\theta G)$.

Figure 3.9 illustrates all of the operations we consider, the precise definitions of which are provided in the relevant subsection. Table 3.1 summaries our findings. It is important to emphasise that other operations are possible. In particular, there is often ample choice in how to assign weights to the altered digraph. We choose to focus on the operations presented herein because they are most relevant to the primary application domain of interest: blood flow through vascular networks.

Operation	Locally Stable	Non-locally Stable	Locally Unstable	Isomorphism
Weight perturbation	Thm 3.5.11			
Edge subdivision	Thm 3.5.16			
Edge collapse	Thm 3.5.25$^\blacklozenge$		Thm 3.5.28	
Edge deletion	Cor. 3.5.34$^\blacklozenge$	Thm 3.5.31	Thm 3.5.35	Thm 3.5.38$^\blacklozenge$
Vertex deletion		Thm 3.5.43	Cor. 3.5.42	Cor. 3.5.41$^\blacklozenge$

Table 3.1: Stability and instability theorems for ${}^g\mathcal{H}_1$, under various digraph operations. $\blacklozenge$ Denotes a theorem which only applies to a subset of such operations.

3.5.1 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 3.5.1. 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 3.5.2. By Lemma 3.3.15, a 0-shifting vertex map is precisely a contraction digraph map.

Lemma 3.5.3. *Any δ-shifting vertex map $f : V(G) \to V(H)$ induces a morphism*

$$\mathcal{S}(f,\delta)_\# : {}^gC(G) \to {}^gC(H)[\delta]. \tag{3.115}$$

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

$$\mathcal{S}(g \circ f, \epsilon + \delta)_\# = \mathcal{S}(g,\epsilon)_\# \circ \mathcal{S}(f,\delta)_\#. \tag{3.116}$$

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

Proof. Since f is δ-shifting, it induces the necessary digraph maps so that Lemma 3.3.9 yields a chain map $J(f, t \leq t+\delta) : {}^gC_\bullet(G,t) \to {}^gC_\bullet(H, t+\delta)$ for any $t \geq 0$. Given $s \leq t$, the chain map ${}^g\iota(s,t)_\#$ is induced, through Lemma 3.3.9, by the identity vertex map. Hence, by equation (3.64), we see

$$J(f, t \leq t+\delta) \circ {}^g\iota(s,t)_\# = J(f, s \leq t+\delta) = {}^g\iota(s+\delta, t+\delta)_\# \circ J(f, s \leq s+\delta). \tag{3.117}$$

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

Equation (3.116) follows immediately from equation (3.64). Finally, when $\delta = 0$, we note that in fact $f \in \mathbf{WDgr}_{\Gamma_d}$ and the construction of $\mathcal{S}(f,0)_\#$ is identical to the construction of ${}^g f_\#$, as done in Theorem 3.3.13. □

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

Lemma 3.5.4. *A δ-shifting vertex map is a δ'-shifting vertex map for any $\delta' \geq \delta$ and*

$$\mathcal{S}(f,\delta')_\# = \mathcal{T}({}^gC(H),\delta'-\delta)[\delta] \circ \mathcal{S}(f,\delta)_\#. \tag{3.118}$$

Proof. At each $t \geq 0$, each of the morphisms in equation (3.118) are induced by Lemma 3.3.9. Namely, $\mathcal{S}(f,\delta')_\#$ is given by $J(f, t \leq t+\delta')$, $\mathcal{T}({}^gC(H),\delta'-\delta)[\delta]$ 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 $t \leq t+\delta \leq t+\delta'$, equation (3.118) follows immediately from equation (3.64). □

Definition 3.5.5. Fix weighted digraphs G, H and two vertex maps $f : V(G) \to V(H)$ and $g : V(H) \to V(G)$.

(a) Denote

$$V_{fix}(g,f) := \{v \in V(G) \mid g(f(v)) = v\}, \quad V_{diff}(g,f) := V(G) \setminus V_{fix}(g,f), \tag{3.119}$$
$$E_{fix}(g,f) := \{v \in E(G) \mid g(f(e)) = e\}, \quad E_{diff}(g,f) := E(G) \setminus E_{fix}(g,f). \tag{3.120}$$

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 3.5.6. (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 3.5.7. For the interested reader, this definition is strongly inspired by the Kalton-Ostrovskii characterisation of network distance, which first appeared for directed networks in [31]. In particular, this characterisation was subsequently used to prove the stability of persistent path homology in [32, Theorem 4.1].

Theorem 3.5.8 (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.

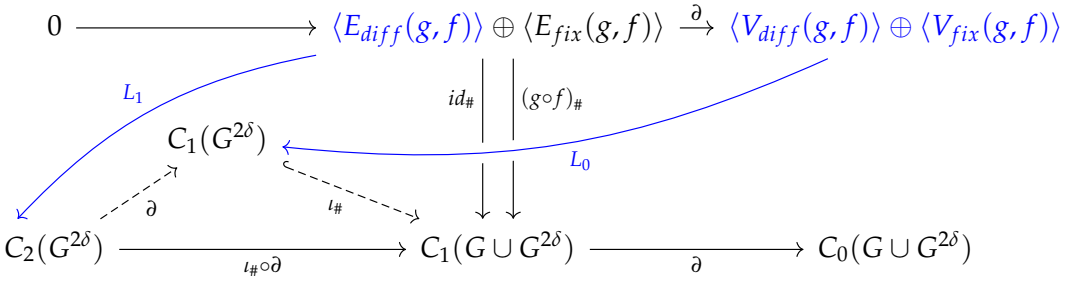
Since (g, f) has grounded codistortion $\leq \delta$ there is a homotopy $F : G_{diff}(g,f) \to G^{2\delta}$ between id and $g \circ f$, relative $V_{fix}(g,f)$. By Theorem 3.2.35, this induces a chain homotopy $L : C_\bullet(G_{diff}(g,f)) \to C_{\bullet+1}(G^{2\delta})$ between $\mathrm{id}_\#$ and $(g\circ f)_\#$. Observe that

$$C_1(G) = R\langle E_{diff}(g,f)\rangle \oplus R\langle E_{fix}(g,f)\rangle, \tag{3.121}$$

$$C_1(G_{diff}(g,f)) = R\langle E_{diff}(g,f)\rangle, \tag{3.122}$$

$$C_0(G_{diff}(g,f)) = C_0(G) = R\langle V_{diff}(g,f)\rangle \oplus R\langle V_{fix}(g,f)\rangle. \tag{3.123}$$

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.



(3.124)

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 equation (3.121), 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

$$(g\circ f)_\#(c_1) - \mathrm{id}_\#(c_1) = \partial L_1(c_1) + L_0\partial(c_1). \tag{3.125}$$

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 3.2.37 ensures that $L_0(\partial c_2) = 0$. Piecing these together we obtain

$$\begin{aligned}(g\circ f)_\#(c) - \mathrm{id}_\#(c) &= (g\circ f)_\#(c_1) - \mathrm{id}_\#(c_1)\\ &= \partial L_1(c_1) + L_0\partial(c_1)\\ &= \partial L_1(c_1) + L_0\partial(c_1+c_2)\\ &= \partial L_1(c_1) + L_0(0)\\ &= \partial L_1(c_1).\end{aligned} \tag{3.126}$$

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 3.4.2 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 3.5.9. 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 ${}^gZ_1(G,0)$. One must show that for any undirected circuit c, there is some $b\in C_2(G^{2\delta})$ such that

$$(\mathcal{S}(g,\delta)_\#\circ\mathcal{S}(f,\delta)_\#)(\mathfrak{R}(c)) - {}^g\iota(0,2\delta)_\#(\mathfrak{R}(c)) = \partial b. \tag{3.127}$$

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.

3.5.2 Weight perturbation

The classical stability theorem of persistent homology (first shown in [33]) is that for two continuous tame functions $f,g: X\to\mathbb{R}$ 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,

$$d_B(\mathcal{B}_f,\mathcal{B}_g)\leq\|f-g\|_\infty. \tag{3.128}$$

In our setting, the closest 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 on the weights of the perturbed edges.

Definition 3.5.10. 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 3.5.11. *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*

$$d_B\big(\mathcal{B}^g\mathcal{H}_1(G), \mathcal{B}^g\mathcal{H}_1(\mathbb{O}^p_{w'} G)\big) \leq \max_{i,j \in V} |d(i,j) - d'(i,j)| \leq \sum_{e \in E} |w(e) - w'(e)|. \tag{3.129}$$

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

$$|\sum_{e \in p} w(e) - \sum_{e \in p} w'(e)| \leq \sum_{e \in p} |w(e) - w'(e)| \leq \sum_{e \in E} |w(e) - w'(e)| =: W_1. \tag{3.130}$$

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

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 3.5.8). □

Remark 3.5.12. Continuing the analogy to the classical stability theorem, note that the sharper bound obtained by Theorem 3.5.11 is $\|d - d'\|_\infty$ while the weaker bound is $\|w - w'\|_1$.

3.5.3 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 3.5.13. 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 3.5.14. 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. $CS(e)_i := \sum_{j=0}^{i} S(e)_j$.

Definition 3.5.15. Given a subdivision $S : F \to \sqcup_{d\in\mathbb{N}} \operatorname{int}(\Delta^d)$, define

$$V_S := V_{old} \sqcup V_{new} := V \sqcup \bigsqcup_{e\in F} \{v_{e,1}, \dots, v_{e,d(e)}\} \tag{3.131}$$

$$E_S := E_{old} \sqcup E_{new} := (E \setminus F) \sqcup \bigsqcup_{e\in F} \{\tau_{e,0}, \dots, \tau_{e,d(e)}\} \tag{3.132}$$

$$w_S(\tau) := \begin{cases} w(\tau) & \text{if } \tau \in E_{old} \\ S(e)_i \cdot w(e) & \text{if } \tau = \tau_{e,i} \end{cases} \tag{3.133}$$

where $\tau_{e,i} = (v_{e,i}, v_{e,i+1})$ and we denote $v_{e,0} := \operatorname{st}(e)$ and $v_{e,d(e)+1} := \operatorname{fn}(e)$. 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 3.5.16. *Given a weighted digraph $G = (V, E, w) \in$ **WDgr** and any subdivision $S : F \to \sqcup_{d\in\mathbb{N}} \operatorname{int}(\Delta^d)$,*

$$d_B\big(\mathcal{B}^g\mathcal{H}_1(G), \mathcal{B}^g\mathcal{H}_1(\mathbb{O}_S^s G)\big) \leq \max_{e\in F} w(e). \tag{3.134}$$

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 3.5.8). First, we set up 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

$$f : V(G) \to V(G_S) \quad v \mapsto v; \tag{3.135}$$

$$g : V(G_S) \to V(G) \quad v \mapsto \begin{cases} v & \text{if } v \in V_{old}, \\ \mathrm{st}(e) & \text{if } v_{e,i} \in V_{new} \text{ and } CS(e)_i < 1/2, \\ \mathrm{fn}(e) & \text{if } v_{e,i} \in V_{new} \text{ and } CS(e)_i \geq 1/2, \end{cases} \tag{3.136}$$

which are visualized in Figure 3.10.

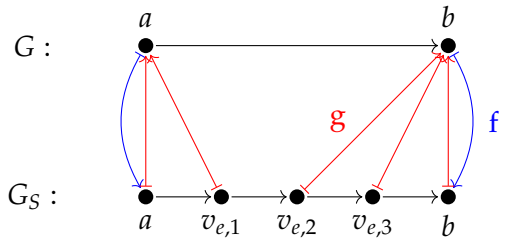


Figure 3.10: Visualising the vertex maps f and g under the subdivision $S(e) = (1/4, 1/4, 1/4, 1/4)$ where $e = (a, b)$.

Claim 3.5.16.1. *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. □ *of Claim.*

Claim 3.5.16.2. *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 3.5.16.1 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 $G \cup G^t \to G_S \cup G_S^{t+\delta}$. □ *of Claim.*

Claim 3.5.16.3. *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 3.5.16.1, 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. $\square$ *of Claim.*

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 Figure 3.11.

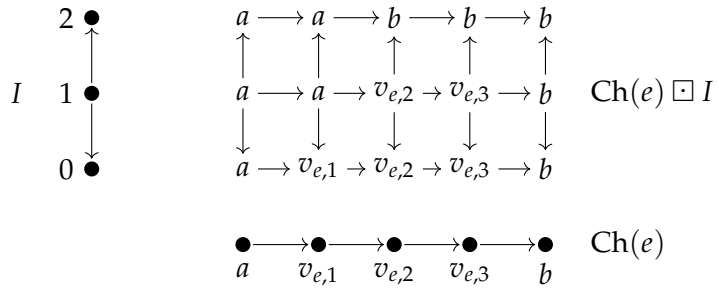


Figure 3.11: Schematic for the homotopy between $f \circ g$ (on the top row) and id (on the bottom row). The label on each node in $\mathrm{Ch}(e) \boxdot I$ is its image under the homotopy F_e.

This is 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

$$F(v, j) := \begin{cases} F_e(v, j) & \text{if } v = v_{e,i} \in V_{new}, \\ v & \text{otherwise.} \end{cases} \tag{3.137}$$

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. $\square$

Remark 3.5.17. Since subdividing an edge does not affect circuit rank of $\mathcal{U}(G)$, the number of features does not change upon subdivision (by Corollary 3.4.3).

Definition 3.5.18. 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 3.5.19. *Given a weighted digraph $G \in \mathbf{WDgr}$, the sequence of barcodes*

$$(\mathcal{B}^g\mathcal{H}_1(\mathrm{IMS}_n(G)))_{n\in\mathbb{N}} \tag{3.138}$$

converges under the bottleneck distance.

Proof. We first note that

$$\max_{e\in E(\mathrm{IMS}_n(G))} w(e) = \frac{1}{2^n} \max_{e\in E(G)} w(e). \tag{3.139}$$

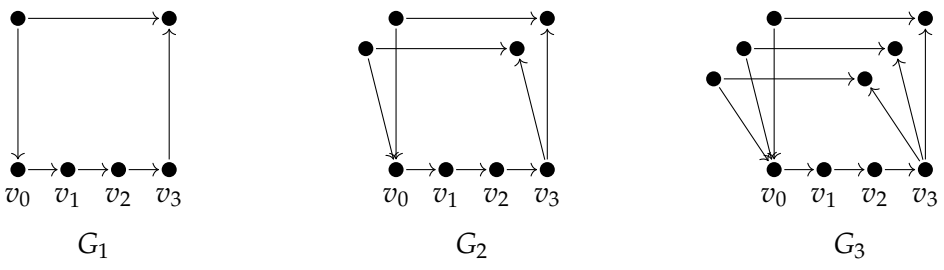
Hence, Theorem 3.5.16 implies that the sequence of barcodes is Cauchy. The space of persistence diagrams with the bottleneck distance is complete [29, Theorem B] and hence the sequence of barcodes converges. □

For an example, we refer the reader forward to Proposition 3.6.1. While we have bottleneck stability, we do *not* have p-Wasserstein stability for any $p \in [1, \infty)$.

Proposition 3.5.20. *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*

$$d_{W_p}\big(\mathcal{B}^g\mathcal{H}_1(\mathbb{O}_S^s G), \mathcal{B}^g\mathcal{H}_1(G)\big) \leq f\big(\mathcal{NG}(F; G)\big). \tag{3.140}$$

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



G_1 G_2 G_3

Figure 3.12: A sequence of weighted digraphs G_n, in which all edges have unit weight and $\#\mathcal{B}^g\mathcal{H}_1(G_n) = n$.

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

$$\mathcal{B}^g\mathcal{H}_1(G_n) = \{\{\, [0,3) \text{ with multiplicity } n \,\}\}\,. \tag{3.141}$$

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

$$d_{W_p}\big(\mathcal{B}^g\mathcal{H}_1(\mathbb{O}_S^s G_n), \mathcal{B}^g\mathcal{H}_1(G_n)\big) = \left(n \cdot \frac{1}{2^p}\right)^{1/p} = \frac{1}{2}n^{1/p} \tag{3.142}$$

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

3.5.4 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 3.5.21. 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 3.5.22. Given a weighted digraph $G = (V, E, w)$ and $F \subseteq E$ collapsible we define the edge collapse $\mathbb{O}_F^c G := (V_F, E_F, w_F)$ 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.

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

$$w_F((I, J)) := \min_{i \in I, j \in J \text{ s.t. } i \to j} \left(w(i, j) + \sum_{e \in F} \frac{w(e)}{2} \cdot \mathbb{1}_{(i,j) \in \mathcal{N}(e)} \right) \tag{3.143}$$

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 3.5.23. The minimum is required in equation (3.143) because, when collapsing an edge $e = (a, b)$, there may be edges $v \to a$ and $v \to b$ with different weights.

Notation 3.5.24. 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 3.5.25. *Given a weighted digraph, $G = (V, E, w) \in \mathbf{WDgr}$ 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 Figure 3.13). Then, for each $e \in F$, define*

$$\delta_e := w(e) + \min\left(\max_{v \in \mathcal{N}_{in}(\mathrm{st}(e))} w(v, \mathrm{st}(e)), \max_{v \in \mathcal{N}_{out}(\mathrm{fn}(e))} w(\mathrm{fn}(e), v)\right) \tag{3.144}$$

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

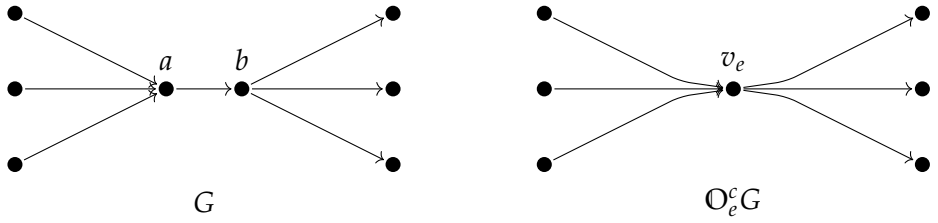


Figure 3.13: Schematic of $\mathcal{NG}(e) \subseteq G$ and $\mathcal{NG}(v_e) \subseteq \mathbb{O}_e^c G$, under the assumptions of Theorem 3.5.25.

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

$$\max_{v \in \mathcal{N}_{in}(\mathrm{st}(e))} w(v, \mathrm{st}(e)) \geq \max_{v \in \mathcal{N}_{out}(\mathrm{fn}(e))} w(\mathrm{fn}(e), v), \tag{3.145}$$

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

Claim 3.5.25.1. *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. $\square$ *of Claim.*

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 3.5.25.2. *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, $(i, j) \in E(G_F^{t+\delta})$. $\square$ *of Claim.*

Claim 3.5.25.3. *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 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 $g(i) \rightsquigarrow g(j)$. □ *of Claim.*

Claim 3.5.25.4. *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 $t \geq 0$. □ *of Claim.*

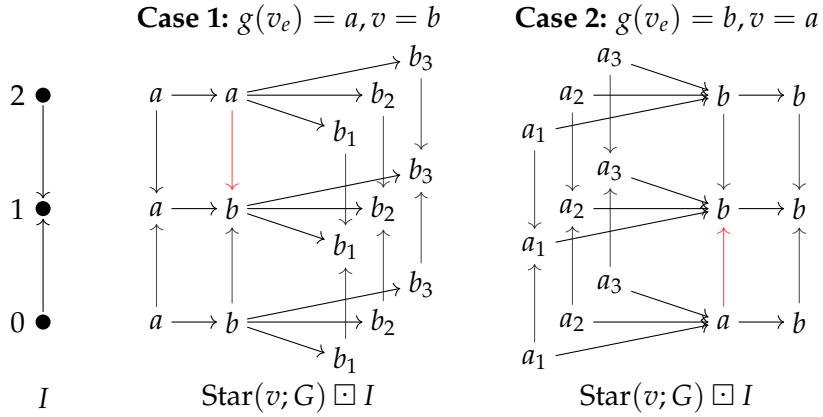


Figure 3.14: Schematic for the homotopy between $g \circ f$ (on the top row) and id (on the bottom row). The label on each node of $\mathrm{Star}(v; G) \boxdot I$ is its image under the homotopy F_v. There are two cases depending on whether $g(v_e) = a$ or $g(v_e) = b$. The red edges denote the only time the image of a vertex differs between the layers.

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

$$F(v, j) := \begin{cases} F_v(v, j) & \text{if } v \in V_{diff}(g, f), \\ v & \text{otherwise} \end{cases} \tag{3.146}$$

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 Figure 3.14 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 $b_1, b_2, \ldots, b_k$. The fact that $g(v_e) = a$ implies that

$$w(e) + \max_{v \in \mathcal{N}_{out}(b)} w(b, v) = \delta_e \leq \delta \tag{3.147}$$

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 3.5.26. Suppose G is a DAG and contains an edge $e = (a, b)$ that satisfies the condition of Theorem 3.5.25. Then one can number the vertices of G as $v_1, \ldots, v_n$ 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 Figure 3.21 and the subsequent discussion in Section 3.6.2. 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 3.5.27. Collapses of the sort described in Theorem 3.5.25 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 3.4.3 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 3.5.25 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 $\mathbb{O}_F^c G$. In general this is not the case (see Figure 3.15); indeed this leads to local instabilities.

Theorem 3.5.28. *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*

$$d_B\big(\mathcal{B}^g\mathcal{H}_1(G), \mathcal{B}^g\mathcal{H}_1(\mathbb{O}_e^c G)\big) \leq f\big(\mathcal{NG}(e; G)\big). \tag{3.148}$$

Proof. Suppose such f exists then consider the weighted digraphs illustrated in Figure 3.15. 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 $W := 2f(\mathcal{NG}(e_1; G')) + 4$ to be the weight of e_2 in G. Observe that $\mathcal{NG}(e_1; G) = \mathcal{NG}(e_1; G')$ and hence

$$W = 2f(\mathcal{NG}(e_1; G')) + 4 = 2f(\mathcal{NG}(e_1; G)) + 4. \tag{3.149}$$

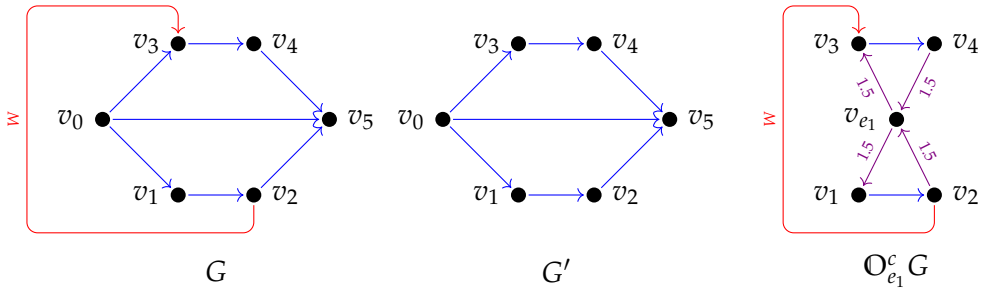


Figure 3.15: An example weighted digraph which illustrates that ${}^g\mathcal{H}_1$ is locally unstable to arbitrary edge collapses. Unlabelled edges have weight 1. Note that there is no path $v_4 \rightsquigarrow v_1$ in G but there is such a path in $\mathbb{O}_{e_1}^c G$.

Initially, $\mathcal{B}^g\mathcal{H}_1(G)$ has 3 features, which die at 2, 2 and W. After collapsing e_1, $\mathcal{B}^g\mathcal{H}_1(\mathbb{O}_{e_1}^c G)$ 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 $\mathbb{O}_{e_1}^c G$ 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 Figure 3.15 makes significant changes to the connectivity of the underlying digraph. That is to say, originally, G was a DAG with source v_0 and sink v_5; the edge collapse identifies these two nodes, introduces directed cycles and makes the digraph strongly connected. Moreover, in G the only path $v_2 \rightsquigarrow v_3$ was via the costly red edge but in $\mathbb{O}_e^d G$ 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.

3.5.5 Edge deletion

3.5.5.1 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 3.5.29. Given a weighted digraph $G = (V, E, w)$ and a subset of edges $F \subseteq E$, we define $\mathbb{O}_F^d G := (V, E \setminus F, w_F)$ where w_F is obtained by restricting w to $E \setminus F$. If $F = \{e\}$ is a single edge, we denote this $\mathbb{O}_e^d G$.

Remark 3.5.30. 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 3.5.31. *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 $\mathbb{O}_F^d G$ respectively. Then*

$$d_B\big(\mathcal{B}^g\mathcal{H}_1(G), \mathcal{B}^g\mathcal{H}_1(\mathbb{O}_F^d G)\big) \leq \max\left(\max_{i,j\in V} |d_F(i,j) - d(i,j)|, \max_{(i,j)\in F} d_F(i,j)\right). \tag{3.150}$$

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

$$\delta_1 := \max_{i,j\in V} |d_F(i,j) - d(i,j)|, \quad \delta_2 := \max_{(i,j)\in F} d_F(i,j) \tag{3.151}$$

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 $\delta \geq \delta_1$, id_V defines a digraph map $G^t \to G_F^{t+\delta}$ and $G_F^t \to G^{t+\delta}$ for all $t \geq 0$. 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 $t \geq 0$.

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 $t \geq \delta_2$. Hence, id_V *does* define a digraph map $G \cup G^t \to G_F \cup G_F^{t+\delta}$ for all $t \geq 0$. □

Remark 3.5.32. 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 3.5.33. *Given a weighted digraph $G = (V, E, w) \in \mathbf{WDgr}$ and an edge $e = (a, b) \in E$ let d and d_e denote the shortest-path quasimetric for G and $\mathbb{O}_e^d G$ respectively. Then*

$$d_B\big(\mathcal{B}^g\mathcal{H}_1(G), \mathcal{B}^g\mathcal{H}_1(\mathbb{O}_e^d G)\big) \leq d_e(a, b). \tag{3.152}$$

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

$$\max_{i,j \in V} |d_e(i, j) - d(i, j)| \leq |d_e(a, b) - d(a, b)| \leq d_e(a, b) = \delta. \tag{3.153}$$

The result then follows by Theorem 3.5.31. To see this inequality, note that for arbitrary $i, j \in V$ we have $d_e(i, j) \geq d(i, j)$ 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 $t + (\delta - w(e)) \leq t + \delta$. Therefore, $d_e(i, j) \leq d(i, j) + \delta$. □

In general, Theorem 3.5.31 is a non-local bound, but if an edge has an alternative route in its local neighbourhood then the bound becomes local.

Corollary 3.5.34. *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*

$$d_B\big(\mathcal{B}^g\mathcal{H}_1(G), \mathcal{B}^g\mathcal{H}_1(\mathbb{O}_e^d G)\big) \leq w(a, v) + w(v, b). \tag{3.154}$$

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 3.5.31 is non-local and indeed no generic, local stability theorem is possible.

Theorem 3.5.35. *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*

$$d_B\big(\mathcal{B}^g\mathcal{H}_1(G), \mathcal{B}^g\mathcal{H}_1(\mathbb{O}_e^d G)\big) \leq f\big(\mathcal{NG}(e; G)\big). \tag{3.155}$$

Proof. Suppose such f exists then consider the following weighted digraph G, where $e := (v_1, v_3)$ and $W := 2f(\mathcal{NG}(e; G)) + 2$.

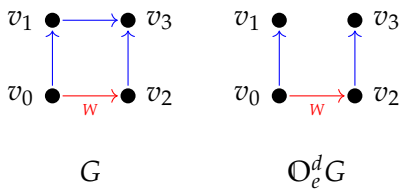


Figure 3.16: An example weighted digraph which illustrates that ${}^g\mathcal{H}_1$ is locally unstable to arbitrary edge deletions. Unlabelled edges have weight 1.

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 $W/2 > f(\mathcal{NG}(e; G))$. □

3.5.5.2 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 3.5.36. 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 3.5.37. An edge is separating if and only if there are no simple undirected circuits containing it.

Corollary 3.5.38. *Given a weighted digraph $G = (V, E, w)$ and a separating edge $e = (a, b) \in E$,*

$$ {}^g\mathcal{H}_1(G) \cong {}^g\mathcal{H}_1(\mathbb{O}_e^d G). \tag{3.156} $$

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 $\mathbb{O}_e^d G$. 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 $\mathbb{O}_e^d G$ is $\mathbb{O}_e^d G = G_1 \sqcup G_3$. 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 Theorem 3.4.18 and Theorem 3.4.21, we see

$$ {}^g\mathcal{H}_1(G) \cong {}^g\mathcal{H}_1(G_1) \oplus {}^g\mathcal{H}_1(G_2) \oplus {}^g\mathcal{H}_1(G_3) \cong {}^g\mathcal{H}_1(G_1) \oplus {}^g\mathcal{H}_1(G_3) \cong {}^g\mathcal{H}_1(\mathbb{O}_e^d G) \tag{3.157} $$

as required. □

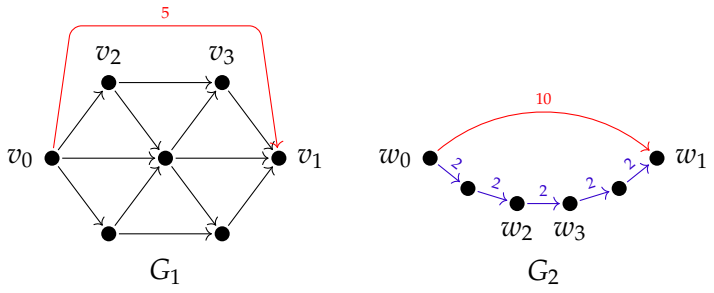


Figure 3.17: Some example weighted digraphs, used to interpret the consequences of the stability theorems obtained in Section 3.5.5. Unlabelled edges have weight 1.

3.5.5.3 Interpretation

Theorem 3.5.31 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 Figure 3.17. 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 Figure 3.17. 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.

3.5.6 Vertex deletion

Definition 3.5.39. Given a weighted digraph $G = (V, E, w)$ and a vertex $v \in V$, we define $\mathbb{O}_v^d G := (V \setminus \{v\}, E_v, w_v)$ 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 $\mathbb{O}_W^d G$ iteratively by choosing an arbitrary $w \in W$ then setting $\mathbb{O}_W^d G := \mathbb{O}_{W\setminus\{w\}}^d \mathbb{O}_w^d G$.

Remark 3.5.40. 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 3.4.18) allows us to delete isolated vertices.

Corollary 3.5.41. *Given a weighted digraph $G = (V, E, w)$ and an isolated vertex $v_i \in V$ (i.e. $\mathcal{N}(v) = \emptyset$), then*

$$\mathcal{B}^g\mathcal{H}_1(G) \cong \mathcal{B}^g\mathcal{H}_1(\mathbb{O}^d_{v_i}G). \tag{3.158}$$

However, in general, deleting a vertex from a digraph can drastically change its topology. This follows immediately from Theorem 3.5.16 and Theorem 3.5.35 since a local vertex deletion stability theorem would imply a local edge deletion stability theorem.

Corollary 3.5.42. *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*

$$d_B\big(\mathcal{B}^g\mathcal{H}_1(G), \mathcal{B}^g\mathcal{H}_1(\mathbb{O}^d_v G)\big) \leq f\big(\mathcal{NG}(v; G)\big). \tag{3.159}$$

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 3.5.43. *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

$$f(v) := \begin{cases} v_1 & \text{if } v = v_0, \\ v & \text{otherwise.} \end{cases} \tag{3.160}$$

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 3.5.43.1. *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 $i = v_0$, $j \neq v_0$. $\square$ *of Claim.*

Claim 3.5.43.2. *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 $t \geq 0$. $\square$ *of Claim.*

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

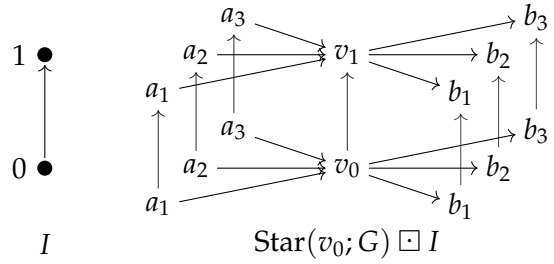


Figure 3.18: Schematic for the homotopy between $g \circ f$ (on the top row) and id (on the bottom row). The label on each node in $\mathrm{Star}(v_0; G) \boxdot I$ is its image under the homotopy F.

Claim 3.5.43.3. *(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

$$F(v,j) := \begin{cases} v_1 & \text{if } (v,j) = (v_0, 1), \\ v & \text{otherwise,} \end{cases} \tag{3.161}$$

which is shown in Figure 3.18. 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 $I = I_-$. □ *of Claim.*

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 3.5.16.

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 $w_1, w_2 \in W$.

3.5.7 Combining operations

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

$$d_b(\mathcal{B}^g\mathcal{H}_1(G_1), \mathcal{B}^g\mathcal{H}_1(G_2)) \leq \delta_1 \quad \text{and} \quad d_b(\mathcal{B}^g\mathcal{H}_1(G_2), \mathcal{B}^g\mathcal{H}_1(G_3)) \leq \delta_2. \tag{3.162}$$

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:

$$V(G_1) \underset{g}{\overset{f}{\leftrightarrows}} V(G_2) \underset{q}{\overset{p}{\leftrightarrows}} V(G_3). \tag{3.163}$$

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 Figure 3.19), to obtain $\mathbb{O}_\tau^d G_S$. In the final digraph, $\mathbb{O}_\tau^d G_S$, there is no alternative path between the endpoints of τ so Theorem 3.5.31 would yield an infinite bound

$$d_B\big(\mathcal{B}^g\mathcal{H}_1(G_S), \mathcal{B}^g\mathcal{H}_1(\mathbb{O}_\tau^d G_S)\big) < \infty. \tag{3.164}$$

However, the remaining child edges and vertices from the subdivision can be further deleted from $\mathbb{O}_\tau^d G_S$ to obtain $G_e := \mathbb{O}_e^d G$. Then, Corollary 3.5.38 and Corollary 3.5.41 imply that ${}^g\mathcal{H}_1(\mathbb{O}_\tau^d G_S) \cong {}^g\mathcal{H}_1(G_e)$. Combining Theorem 3.5.16 and Theorem 3.5.31, we can bound

$$\begin{aligned} d_B\big(\mathcal{B}^g\mathcal{H}_1(G_S), \mathcal{B}^g\mathcal{H}_1(\mathbb{O}_\tau^d G_S)\big) &= d_B\big(\mathcal{B}^g\mathcal{H}_1(G_S), \mathcal{B}^g\mathcal{H}_1(G_e)\big) && (3.165)\\ &\leq d_B\big(\mathcal{B}^g\mathcal{H}_1(G_S), \mathcal{B}^g\mathcal{H}_1(G)\big) + d_B\big(\mathcal{B}^g\mathcal{H}_1(G), \mathcal{B}^g\mathcal{H}_1(G_e)\big) && (3.166)\\ &\leq w(e) + d_e(\mathrm{st}(e), \mathrm{fn}(e)) && (3.167) \end{aligned}$$

where d_e is the shortest-path quasimetric in G_e.

Recall that the grounded interleaving used to prove edge deletion stability (Theorem 3.5.31) 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 3.5.44. *Given a weighted digraph $G = (V, E, w) \in \mathbf{WDgr}$, 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 $G_e := \mathbb{O}_e^d G$. 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.*

$$d_B\big(\mathcal{B}^g\mathcal{H}_1(G_S), \mathcal{B}^g\mathcal{H}_1(\mathbb{O}_\tau^d G_S)\big) \leq \max\left(d_e(a, b), \frac{w(e)}{2}\right). \tag{3.168}$$

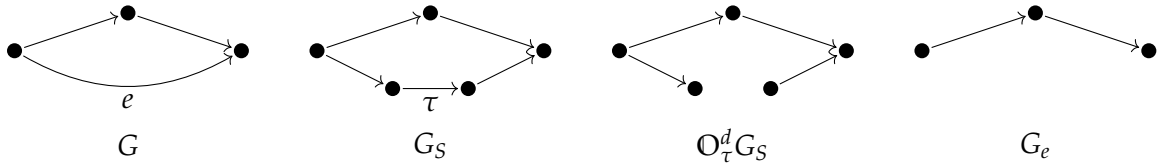


Figure 3.19: Schematic of the four weighted digraphs in Theorem 3.5.44, where e is the curved bottom edge and τ is as labelled.

Proof. As discussed above, it remains to prove

$$d_B\big(\mathcal{B}^g\mathcal{H}_1(G_S), \mathcal{B}^g\mathcal{H}_1(G_e)\big) \leq \max\left(d_e(a,b), \frac{w(e)}{2}\right) =: \delta. \tag{3.169}$$

Define the following vertex maps

$$f : V(G_e) \to V(G_S) \quad v \mapsto v; \tag{3.170}$$

$$g : V(G_S) \to V(G_e) \quad v \mapsto \begin{cases} v & \text{if } v \in V_{old}, \\ \mathrm{st}(e) & \text{if } v_{e,i} \in V_{new} \text{ and } CS(e)_i < 1/2, \\ \mathrm{fn}(e) & \text{if } v_{e,i} \in V_{new} \text{ and } CS(e)_i \geq 1/2. \end{cases} \tag{3.171}$$

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

Claim 3.5.44.1. *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, $(i,j) \in E(G_e^{t+\delta})$. $\square$ *of Claim.*

Claim 3.5.44.2. *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

$$(g(v_{e,i-1}), g(v_{e,i})) = (\mathrm{st}(e), \mathrm{st}(e)) \text{ or } (\mathrm{st}(e), \mathrm{fn}(e)) \text{ or } (\mathrm{fn}(e), \mathrm{fn}(e)). \tag{3.172}$$

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. $\square$ *of Claim.*

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 3.5.16. However, note that we require $2\delta \geq w(e)$ so that all edges in the image of the homotopy are edges of $G_S^{2\delta}$. $\square$

3.6 Examples

3.6.1 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 Figure 3.20). 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.

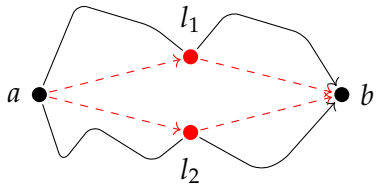


Figure 3.20: Illustration of the weighted digraph G, considered in Proposition 3.6.1. The top path p_1 has length l_1 and the bottom path p_2 has length l_2. The limiting death time of the sole feature corresponds to the limiting value of the earliest time that a long square appears of the form $av_1b - av_2b$ where v_1 is along p_1 and v_2 is along p_2 (as drawn in red dashed lines).

Proposition 3.6.1. *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 3.5.18). Then*

$$\lim_{n\to\infty} \mathcal{B}^g\mathcal{H}_1(\mathrm{IMS}_n(G)) = \left\{\left\{ \left[0, \frac{1}{2}l_1\right) \right\}\right\} \tag{3.173}$$

Proof. For brevity we denote $G_n := \mathrm{IMS}_n(G)$. By Corollary 3.4.3, 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 3.4.6 we see $\mathcal{D}(\mathfrak{R}(c^{(n)})) \leq \max(h_1^{(n)}, h_2^{(n)})$ where

$$h_i^{(n)} := \min\left\{ t \geq 0 \;\middle|\; \exists v_i \in V \text{ along } p_i^{(n)} \text{ such that } d(a, v_i) \leq t \text{ and } d(v_i, b) \leq t \right\}. \quad (3.174)$$

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 $\dim {}^g H_1(G_n, t_2) \neq 0$. In order to do so, we claim the inclusion chain map

$$\begin{array}{ccccc} 0 & \longrightarrow & C_1(G_n) & \xrightarrow{{}^g\partial_1} & C_0(G_n) \\ \downarrow{\scriptstyle j_2} & & \downarrow{\scriptstyle j_1} & & \downarrow{\scriptstyle j_0} \\ C_2(G_n^{t_2}) & \xrightarrow{{}^g\partial_2} & C_1(G_n \cup G_n^{t_2}) & \xrightarrow{{}^g\partial_1} & C_0(G_n \cup G_n^{t_2}) \end{array} \quad (3.175)$$

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 3.4.1, 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,

$$\frac{1}{2} l_1 \leq \mathcal{D}(\mathfrak{R}(c^{(n)})) \leq \max(h_1^{(n)}, h_2^{(n)}). \quad (3.176)$$

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

Note that description of Proposition 3.6.1 is not unique to this descriptor, indeed the same result holds for the standard pipeline. As discussed in Section 3.3.1, 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 3.6.2. *Given $G \in \mathbf{WDgr}$, let $\mathcal{H}_1$ denote the 'standard pipeline' with $C = \Omega$ and $F = F_d$, as used in Example 3.3.6. Then*

$$\lim_{n\to\infty} \mathcal{B}^g\mathcal{H}_1(\mathrm{IMS}_n(G)) = \lim_{n\to\infty} \mathcal{B}\mathcal{H}_1(\mathrm{IMS}_n(G)). \tag{3.177}$$

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 $i_* : H_1(G_n^t) \to {}^gH_1(G_n, t+\epsilon)$. 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 $j_* : {}^gH_1(G,t) \to H_1(G^{t+\epsilon})$.

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

$$d_B\left(\mathcal{B}^g\mathcal{H}_1(G_n), \mathcal{B}\mathcal{H}_1(G_n)\right) \leq \epsilon \tag{3.178}$$

for all $n \geq N$. □

3.6.2 Square motifs

Example 3.6.3. Further to the interpretation developed in Proposition 3.6.1, consider the four weighted digraphs in Figure 3.21. All edges are given unit weight and the barcodes are indicated under each digraph. Homology representatives for each of the features are given by

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

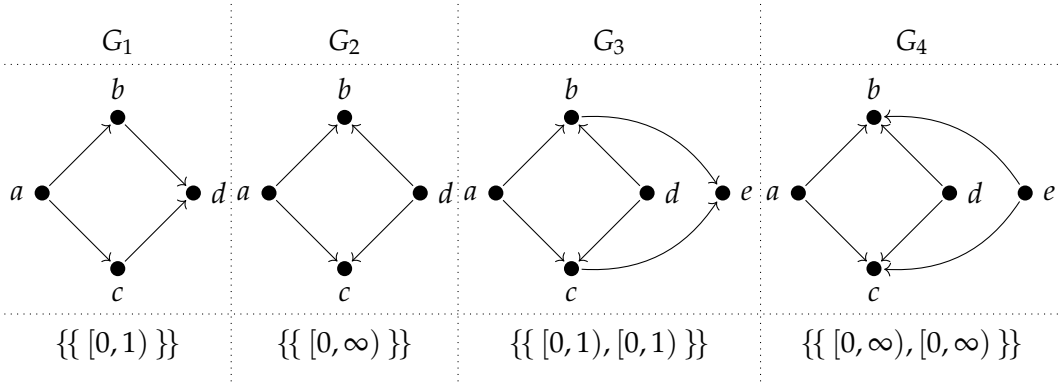


Figure 3.21: Interpreting ${}^{g}\mathcal{H}_1(G)$ via differences between directed paths in square motifs. Top row: name of weighted digraph; middle row: diagram where all weights are 1; bottom row: barcode of GRPPH.

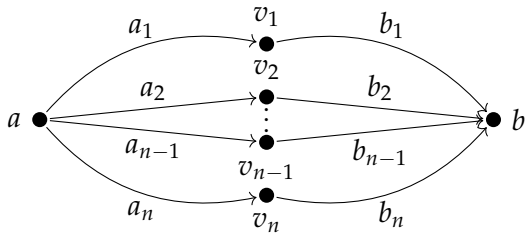


Figure 3.22: A weighted digraph with many paths (of 2 edges each) from source to sink for which we can compute $\mathcal{B}^{g}\mathcal{H}_1(G)$.

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 $[0, \infty)$.

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.

3.6.3 Multiple paths

Example 3.6.4. Consider Figure 3.22, 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 $\alpha_1 \leq \alpha_2 \leq \ldots \alpha_{n-1} \leq \alpha_n$.

Then, the barcode is

$$\mathcal{B}^g\mathcal{H}_1(G) = \{\{\,[0,\alpha_2),[0,\alpha_3),\ldots,[0,\alpha_{n-1}),[0,\alpha_n)\,\}\}\,. \tag{3.179}$$

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

3.6.4 Digraphs with identical quasimetrics

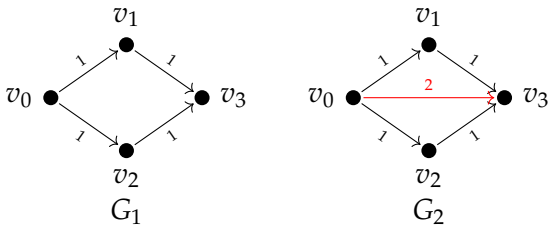


Figure 3.23: Two weighted digraphs with identical shortest-path quasimetric network but differing GRPPH which can be explained by the difference in circuit rank.

Example 3.6.5. Finally, consider the two weighted digraphs illustrated in Figure 3.23. 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,

$$\mathcal{B}^g\mathcal{H}_1(G_1) = \{\{\,[0,1)\,\}\} \quad \text{and} \quad \mathcal{B}^g\mathcal{H}_1(G_2) = \{\{\,[0,1),[0,2)\,\}\}\,. \tag{3.180}$$

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 $t = 2$.

3.6.5 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 $MG(X) := (X, X \times X \setminus \Delta_X, w)$ where

$$w((v_i,v_j)) := d(v_i,v_j) \quad \text{for all } (v_i,v_j) \in X \times X \setminus \Delta_X. \tag{3.181}$$

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

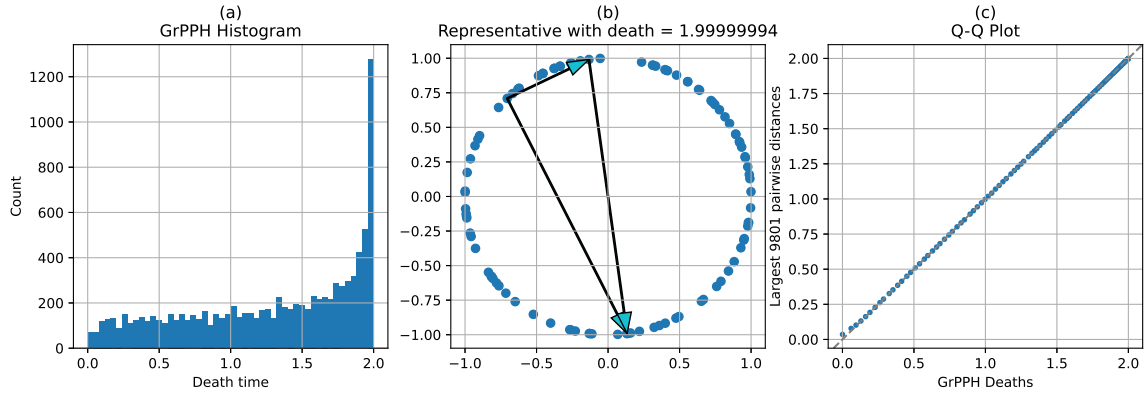


Figure 3.24: GRPPH of 100 points sampled on the unit circle. (a) A histogram of the death times of each of the features in ${}^{g}\mathcal{H}_1$. (b) Quantile-quantile plot of GRPPH deaths against the distribution of the 9801 largest pairwise distances. (c) The point sample along with a directed triangle which is a representative of the largest feature, with death time almost 2.

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

$$\#\text{edges} - \#\text{vertices} + \#\text{weak components} = n(n-1) - n + 1 = (n-1)^2 \tag{3.182}$$

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

$$\mathcal{D}(\mathfrak{R}(c)) = \max(d(v_0, v_1), d(v_1, v_2), d(v_0, v_2)). \tag{3.183}$$

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 [55, Example 3.11], all features must have death time at most $\operatorname{diam}(X)$. Therefore, the largest feature in ${}^{\mathcal{B}g}\mathcal{H}_1(MG(X))$ must have death time equal to $\operatorname{diam}(X)$.

In Figure 3.24, 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 Figure 3.24. 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 [32] (viewing the data as a directed network) or traditional, symmetric TDA methods.

Chapter 4

A notion of homotopy for directed graphs and their flag complexes

4.1 Introduction

Digraphs appear naturally in numerous application domains, including neuroscience [87], traffic network analysis [10] and social network analysis [83]. Frequently these digraphs support a dynamical system such as brain activity, traffic flow and (dis)information spread. A common hypothesis is that the structure of the digraph is critical in determining the evolution of the dynamical system. In order to assess this hypothesis, there is a need for interpretable and informative summaries of digraphs, that are amenable to statistical analysis.

A common approach is to associate a combinatorial object to the digraph built out of substructures in the digraph which are relevant to the application in hand. For example, one could study the set of all paths [54], all tournaments [51] or all directed cliques [75] in the digraph. The interest to us is that each of these combinatorial objects carries sufficient structure to build a chain complex, the homology of which can be used as an algebraic invariant of the underlying digraph. These invariants are most useful (such as in the context of persistence) when they are functorial, in the sense that there is a class of digraph maps which induce maps on homology.

In order to assess the discriminative power of these homologies and hence their utility in applications, some natural questions arise: When do two digraphs give rise to isomorphic homology? When do two digraph maps induce *the same* maps on homology?

In this work, we make progress towards these two questions in the case of the directed flag complex. This is a simplicial set in which the k-simplies are the directed $(k+1)$-cliques, i.e. $(k+1)$-tuples of distinct vertices $v_0 \dots v_k$ such that there is an edge $v_i \to v_j$ whenever

$i < j$. We focus on the directed flag complex because it has seen particular success in the field of neuroscience. The homology of the directed flag complex was proposed as a digraph invariant by Masulli and Villa [75] and, shortly thereafter, the Betti numbers of the complex associated to the activation graph of numerically simulated neocortical microcircuitry was shown to exhibit significant temporal patterns in response to stimuli [87, Figure 6]. Further applications may be enabled by the incorporation of persistence, and the presence of highly performant software for its computation [74], which can be used to capture the topological organisation of directed structures, across a range of scales [17].

In contrast, the directed flag complex has received comparatively little theoretical treatment. Here, we describe an equivalence relation $\simeq_{\mathrm{dFl}}$ for digraph maps such that if two maps are $\simeq_{\mathrm{dFl}}$-equivalent then they induce identical maps on the homology of the directed flag complex. In particular, this induces an equivalence relation $\simeq_{\mathrm{dFl}}$ on digraphs themselves, such that if two digraphs are $\simeq_{\mathrm{dFl}}$-equivalent then their directed flag complexes have isomorphic homology. To illustrate the utility of these relations in the context of persistence, we use the relation $\simeq_{\mathrm{dFl}}$ to show that the persistent homology of the directed flag complex applied to a filtration of digraphs is stable in certain circumstances.

4.1.1 Summary

Given a digraph G, we denote the associated directed flag complex by $\mathrm{dFl}(G)$. The directed flag complex is a particular example of a more general class of objects called ordered simplicial complexes (Definition 4.2.4). In this work, we primarily view an ordered simplicial complex as a special case of a regular path complex (Definition 4.2.3). The latter were first introduced in [54] as an invariant of directed graphs. To any regular path complex one can associate a chain complex and moreover this can be made into a functor $\Omega : \mathbf{WkRPC} \to \mathbf{Ch}$, where **WkRPC** is a category of regular path complex and **Ch** is the category of chain complexes of vector spaces over some background field.

The morphisms in **WkRPC** are called weak path morphisms (Definition 4.2.6) and are defined to be maps of the underlying vertex sets, subject to certain constraints. In Lemma 4.2.10, we show that a vertex map f defines a weak path morphism between ordered simplicial complexes if and only if

1. the image of a simplex is either contained in the target complex or is non-simplicial; and
2. if $v_0v_1v_2$ is a simplex in the domain and $f(v_0) = f(v_2)$ then $f(v_0) = f(v_1) = f(v_2)$.

We call the latter condition triangle-collapsing (Definition 4.2.7) since it ensures 2-simplices are either preserved in the image or are collapsed to degenerate lower-dimensional simplices, but *crucially* cannot be mapped to a reciprocal pair of 1-simplices. Taking these maps together with all ordered simplicial complexes yields a full subcategory of **WkRPC**, which we denote **TcOSC**.

Pre-composing with the functor which takes digraphs to their directed flag complex, we obtain a sequence of functors

$$\mathbf{TcDgr} \xrightarrow{\mathrm{dFl}} \mathbf{TcOSC} \xhookrightarrow{\iota} \mathbf{WkRPC} \xrightarrow{\Omega} \mathbf{Ch} \tag{4.1}$$

where **TcDgr** is a category of digraphs in which morphisms are defined via a similar 'triangle-collapsing' condition, so that dFl becomes a full and faithful functor. We show that this is a factorisation of the chain complex of $\mathrm{dFl}(G)$ (Lemma 4.2.25). This factorisation is the main perspective through which we construct our equivalence relation, $\simeq_{\mathrm{dFl}}$, by comparison with a similar relation for the well-studied category **WkRPC**.

In particular, there is a notion of homotopy, $\simeq$, between the morphisms of **WkRPC** such that if $f \simeq g$ then $\Omega(f)$ and $\Omega(g)$ are chain homotopic (see [53, Theorem 3.8]). The equivalence relation for digraph maps that we will construct is a pull-back of this equivalence relation, i.e.

$$f \simeq_{\mathrm{dFl}} g \iff (\iota \circ \mathrm{dFl})(f) \simeq (\iota \circ \mathrm{dFl})(g). \tag{4.2}$$

Our main contribution is an 'intrinsic' characterisation of this pull-back (Section 4.5.2).

To be more precise, [53] describes a system of one-step homotopies between morphisms in **WkRPC**. This is essentially a binary relation on the morphisms of **WkRPC**, which is then completed to the equivalence relation $\simeq$. As above, we can pull back this binary relation to one for the morphisms of **TcDgr**. Given two morphisms $f, g \in \mathbf{TcDgr}$, we can provide simple, edge-based conditions that are equivalent to the pair (f, g) belonging to this pull-back binary relation (Corollary 4.5.12). In particular, we relate (f, g) if

1. $x \rightrightarrows y \implies f(x) \rightrightarrows g(y)$; and
2. $x \to y$ and $f(x) = g(y) \implies f(x) = f(y) = g(x) = g(y)$,

where $\rightrightarrows$ indicates that either the vertices coincide or are joined by a directed edge. The relation $\simeq_{\mathrm{dFl}}$ can then be constructed by completing this to an equivalence relation.

To provide some intuition for this binary relation, one might imagine defining a one-step homotopy from f to g to be a sufficiently well-behaved digraph map out of an appropriate

'cylinder' digraph (in analogy to a homotopy of continuous maps). Indeed, there is a 'unit interval' digraph I on the vertices $\{0, 1\}$ and a product operation, $\times$, for digraphs. Supposing f and g are digraph maps $G \to H$, we will see that condition 1 is equivalent to f and g jointly defining a weak digraph map $F : G \times I \to H$ by

$$F(v, 0) = f(v) \quad \text{and} \quad F(v, 1) = g(v). \tag{4.3}$$

Condition 2 then ensures that certain directed 3-cliques within $G \times I$ are not mapped to a pair of reciprocal edges by F. However, we will also see that this is not, in general, equivalent to F defining a morphism in **TcDgr**.

As a first step, in Section 4.4, we study the pull-back binary relation for **TcOSC** along the inclusion functor $\iota : \mathbf{TcOSC} \hookrightarrow \mathbf{WkRPC}$. Again, we aim to describe the resulting relation 'intrinsically' with the category **TcOSC**. Corollary 4.4.9 achieves this by describing related morphisms in terms of a cylinder functor for this category. For an alternative view, **TcOSC** can also be fully and faithfully embedded in the category of simplicial sets, which itself has a notion of homotopy equivalence. In Section 4.4.3, we verify that the pull-back relation along this embedding coincides with the pull-back along ι.

Finally, in Section 4.6, we employ the equivalence relation $\simeq_{\mathrm{dFl}}$ to obtain stability results for the persistent homology of the directed flag complex of a filtration of digraphs. Typically, such stability results take the form of a bound on the interleaving distance. The relation $\simeq_{\mathrm{dFl}}$ allow us to construct interleavings of digraph filtrations up to $\simeq_{\mathrm{dFl}}$, facilitating the derivation of such bounds. We summarise this approach with Corollary 4.6.4. In Proposition 4.6.10, we apply this result to the shortest-path filtration of a weighted directed acyclic graph (DAG), showing that its persistent homology is stable to edge subdivision. However, persistent homology pipelines using the directed flag complex have some notable instabilities. We discuss some important examples in Section 4.6.3, namely edge subdivision of a non-DAG and the addition of an appendage edge.

4.1.2 Related work

There are numerous homotopy theories for both directed and undirected graphs. The relation we will develop is primarily derived from path homotopy theory, which was introduced by Grigor'yan et al. in a series of papers [52, 54, 55]. Initially developed for digraphs, path homotopy generalises a prior homotopy theory for undirected graphs [4, 5], viewing undirected graphs as a full subcategory of digraphs. This theory was further generalised to path complexes in later work [53].

When the digraph G is oriented, there is at most one directed clique on any subset of vertices. As such, $\mathrm{dFl}(G)$ becomes a simplicial complex and coincides with the definition of the directed flag complex given in [77]. In that work, the authors view directed graphs as a subcategory of pseudotopological spaces and show there is a weak homotopy equivalence between the realizations of G and the simplicial complex obtained by collapsing simplices in $\mathrm{dFl}(G)$ that share a common vertex set. That work draws on the broader discipline of Čech closure spaces. The Čech closure space perspective on discrete homotopy was first discussed in [88].

For undirected graphs, similar results were obtained via different methods in the following work [95]. The s-homotopy type of a graph, as defined in [12, 13], coincides with the simple homotopy type of its undirected flag complex [13, Theorem 2.10], and it was later shown that s-homotopy also coincides with $\mathfrak{J}$-homotopy for undirected graphs [39, Theorem 1.2]. The undirected flag complex of an undirected graph was generalised to semi-uniform spaces in [89]; these spaces come equipped with a homotopy theory to which the homology of the flag complex is invariant.

4.1.3 Notation

Given categories $\boldsymbol{\mathcal{C}}, \boldsymbol{\mathcal{D}}$, the category of functors from $\boldsymbol{\mathcal{C}}$ to $\boldsymbol{\mathcal{D}}$ is denoted $\mathbf{[\mathcal{C}, \mathcal{D}]}$. We use $\mathrm{Mor}_{\boldsymbol{\mathcal{C}}}(C_1, C_2)$ to denote the set of morphisms between two objects $C_1, C_2 \in C$ and use $\mathrm{Obj}(C)$ to denote the objects of $\boldsymbol{\mathcal{C}}$. Given functors $F, G : \boldsymbol{\mathcal{C}} \to \boldsymbol{\mathcal{D}}$, we denote a natural transformation by $\mu : F \Rightarrow G$. We denote components of μ by μ_C for each object $C \in \boldsymbol{\mathcal{C}}$, or optionally omit the notation when the component is clear from context.

4.2 Combinatorial and algebraic complexes

4.2.1 Path complexes

Definition 4.2.1. Let V be an arbitrary set, which we call the vertices.

1. An elementary k-path (on V) is any sequence of $(k+1)$ vertices, written $p = v_0 \dots v_k$, where $v_i \in V$.
2. Let $v_0 \dots \hat{v}_i \dots v_k$ denote the $(k-1)$-path obtained by removing the vertex v_i.
3. Given a function $f : V \to W$ and an elementary path $p = v_0 \dots v_k$, let $f(p)$ denote the elementary k-path $f(v_0) \dots f(v_k)$.
4. We say p is irregular if $v_i = v_{i+1}$ for some i, otherwise we say p is regular.
5. We say p is simplicial if each of the v_i are distinct.

6. If an elementary path p' can be obtained from another elementary path p by removing a subset of vertices, we say p' is a face of p.
7. If an elementary path p' can be obtained from another elementary path p by successively removing the initial or terminal vertex, we say p' is a sub-path of p.

Example 4.2.2. On the vertex set $V = \{a, b\}$, $p_1 := aab$ and $p_2 := aba$ are both elementary 2-paths. The path p_1 is irregular and thus non-simplicial, whilst p_2 is regular but still non-simplicial. The path $p_3 := aa$ is a sub-path and face of p_1 whilst p_3 is a face of p_2 but not a sub-path.

Next, we describe the notions of a path complex, first introduced in [54], and an ordered simplicial complex.

Definition 4.2.3. A path complex on V is a set P of elementary paths on V, such that

1. P contains all singletons, i.e. $v_0 \in P$ for every $v_0 \in V$;
2. P is closed under truncating paths at either end, i.e. if $v_0 \dots v_k \in P$ then $v_1 \dots v_k \in P$ and $v_0 \dots v_{k-1} \in P$.

We denote the base vertex set, $V(P) := V$. We call a path complex regular if all of its constituent paths are regular.

Definition 4.2.4. An ordered simplicial complex on V is a set K of *simplicial* paths on V, such that

1. P contains all singletons, i.e. $v_0 \in K$ for every $v_0 \in V$;
2. P is closed under taking ordered subsets, i.e. if $v_0 \dots v_k \in K$ then for every i, $v_0 \dots \hat{v_i} \dots v_k \in K$.

We denote the base vertex set, $V(K) := V$.

Remark 4.2.5. We emphasise that the ordering on the vertices of a path in a simplicial path is *not* inherited from some total order on V. Instead, each simplex carries a total order of its vertices and it is possible for a given subset of vertices to support multiple simplices (each with a different order).

For brevity, we refer to paths in an ordered simplicial complex as simplices and take 'simplicial complex' to mean an ordered simplicial complex. Clearly any simplicial complex is a regular path complex. Since we are primarily interested in simplicial complexes, we primarily focus on regular path complexes in this work. Next, we describe a range of categorical structures for these combinatorial complexes.

Definition 4.2.6. Given two path complexes, P_1 and P_2, a map $f : V(P_1) \to V(P_2)$ is a

1. weak path morphism if for any $p \in P_1$, $f(p)$ is either irregular or $f(p) \in P_2$;
2. strong path morphism if for any $p \in P_1$, $f(p) \in P_2$.

We denote the category of all path complexes with weak path morphisms as **WkPathC**, and with strong path morphisms as **StPathC**. We denote the full subcategories of all regular path complexes by **WkRPC** and **StRPC** respectively.

Definition 4.2.7. Given two ordered simplicial complexes, K_1 and K_2, a map $f : V(K_1) \to V(K_2)$ is a

1. weak simplicial morphism if for any $p \in K_1$, $f(p)$ is either non-simplicial or $f(p) \in K_2$;
2. triangle-collapsing simplicial morphism if f is weak simplicial and furthermore whenever there is a simplex $v_0v_1v_2 \in K_1$ such that $f(v_0) = f(v_2)$ then $f(v_0) = f(v_1) = f(v_2)$;
3. strong simplicial morphism if for any $p \in K_1$, $f(p) \in K_2$.

We denote the category of all simplicial complexes with weak simplicial morphisms as **WkOSC**, with triangle-collapsing simplicial morphisms as **TcOSC**, and with strong simplicial morphisms as **StOSC**.

Remark 4.2.8. Given a triangle-collapsing simplicial morphism $f : K_1 \to K_2$ if there is a simplex $p = v_0 \dots v_k$ such that $f(v_i) = f(v_j)$, then *all* the vertices v_m, for $i \leq m \leq j$, have the same image under f.

The relationship between strong simplicial and strong path morphism is obvious.

Lemma 4.2.9. *Given two simplicial complexes, K_1, K_2, a vertex map $f : V(K_1) \to V(K_2)$ is a strong simplicial morphism if and only if it is a strong path morphism. Therefore,* **StOSC** *is a full subcategory of* **StRPC**.

A weak simplicial morphism can fail to be a weak path morphism if it maps a simplex to a non-simplicial but regular path. For example, consider the following simplicial complexes

$$K_1 = \{0, 1, 2, 01, 02, 12, 012\}, \quad K_2 = \{0, 1, 01, 10\}. \tag{4.4}$$

Consider the vertex map $f : \{0, 1, 2\} \to \{0, 1\}$ which is the identity on $\{0, 1\}$ and maps $2 \mapsto 0$. It is easy to check that f is a weak simplicial morphism. However, it is not a weak path morphism because $f(012) = 010$ is regular but does not belong to K_1. Thankfully, precluding this scenario is necessary and sufficient to yield a weak path morphism.

Lemma 4.2.10. *A weak simplicial morphism $f : K_1 \to K_2$ is a weak path morphism if and only if it is a triangle-collapsing simplicial morphism. Therefore,* **TcOSC** *is the full subcategory of* **WkRPC**, *with objects restricted to all simplicial complexes.*

Proof. Suppose f is triangle-collapsing and take an arbitrary path $p = v_0 \dots v_k \in K_1$. Since f is weak simplicial there are two cases. In the first case $f(p) \in K_2$ and we are done. In the second case $f(p)$ is non-simplicial, i.e. there is some $i < j$ such that $f(v_i) = f(v_j)$. Since f is triangle collapsing, we must have $f(v_m) = f(v_i)$ for all $i \leq m \leq j$. In particular, $f(v_{i+1}) = f(v_i)$ and hence $f(p)$ is irregular. Therefore, f is a weak path morphism.

Now suppose f is not triangle-collapsing. Hence, there is some simplex $v_0v_1v_2 \in K_1$ with $f(v_0) = f(v_2)$ but $f(v_1) \neq f(v_0)$ and $f(v_1) \neq f(v_2)$. Therefore, $f(p)$ is non-simplicial, and hence not a simplex of K_2, but is also regular. Thus, f is not a weak path morphism. □

Definition 4.2.11. Given $k \geq 0$ and a path complex P, the k-skeleton of P, is the sub–path complex

$$\mathrm{Sk}_k(P) := \{p \in P \mid p \text{ is an elementary } l\text{-path for } l \leq k\}\,. \tag{4.5}$$

Remark 4.2.12. Any weak path morphism $f : P_1 \to P_2$ induces a weak path morphism $f : \mathrm{Sk}_k(P_1) \to \mathrm{Sk}_k(P_2)$. Hence, Sk_k is an endofunctor on **WkPathC**. This endofunctor restricts to an endofunctor on all the categories introduced in this section.

Definition 4.2.13. Given a path complex P, removing all irregular paths yields its maximal regular sub–path complex, which we call its regularisation. This operation constitutes a functor reg : **WkPathC** $\to$ **WkRPC** and reg : **StPathC** $\to$ **StRPC**.

4.2.2 Regular path complexes from digraphs

A directed graph, or digraph, is a pair $G = (V, E)$ where V is an arbitrary set and $E \subseteq (V \times V)$. We call V the set of vertices and E the set of (directed) edges. We denote the sets of vertices and edges by $V(G) := V$ and $E(G) := E$ respectively. An edge belonging to $\Delta_V := \{(v, v) \mid v \in V\}$ is called a self-loop. We call a digraph simple if it contains no self-loops, i.e. $E \subseteq (V \times V) \setminus \Delta_V$. **In this work, we assume that all digraphs are simple**.

We write $v_0 \to v_1$ to mean $(v_0, v_1) \in E$ and similarly $v_0 \not\to v_1$ to mean $(v_0, v_1) \notin E$; we write $v_0 \rightrightarrows v_1$ to mean either $v_0 = v_1$ or $v_0 \to v_1$. Given an edge, $e = (s, t) \in E$, we denote its endpoints by $\mathrm{st}(e) := s$ and $\mathrm{fn}(e) := t$. We say there is a reciprocal edge on $\{v_0, v_1\} \subseteq V$ if $v_0 \to v_1$ and $v_1 \to v_0$, and we write $v_0 \leftrightarrows v_1$. Finally, if G contains no reciprocal edges then we say G is oriented.

Definition 4.2.14. Given a directed graph G, a directed k-clique is a k-tuple of distinct vertices (written $v_0 \dots v_{k-1}$) so that there is an edge $v_i \to v_j$ whenever $i < j$. The directed flag complex of G, $\mathrm{dFl}(G)$, is the simplicial complex on V consisting of all directed cliques in G.

This construction can also be made functorial, but first we must define some categories of digraphs.

Definition 4.2.15. Given directed graphs $G_1 = (V_1, E_1)$ and $G_2 = (V_2, E_2)$, a map $f : V_1 \to V_2$ is a

1. weak digraph map if whenever $v \to w$ in G_1, either $f(v) = f(w)$ or $f(v) \to f(w)$ in G_2;
2. triangle-collapsing digraph map if f is a weak digraph map and furthermore whenever there is a directed 3-clique, $v_0v_1v_2$, such that $f(v_0) = f(v_2)$, then $f(v_0) = f(v_1) = f(v_2)$;
3. strong digraph map if whenever $v \to w$ in G_1 then $f(v) \to f(w)$ in G_2.

We denote the category of all simple digraphs with weak digraph maps as **WkDgr**, with triangle-collapsing digraph maps as **TcDgr** and with strong digraph maps as **StDgr**.

Remark 4.2.16. If H is oriented then any weak digraph map $f : G \to H$ is triangle-collapsing.

Lemma 4.2.17. *Given two digraphs G, H and a vertex map $f : V(G) \to V(H)$,*

1. *f is a weak digraph map $G \to H$ if and only if it is a weak simplicial morphism $K_G \to \mathrm{dFl}(H)$ for some simplicial complex $\mathrm{Sk}_1(\mathrm{dFl}(G)) \subseteq K_G \subseteq \mathrm{dFl}(G)$;*
2. *f is a strong digraph map $G \to H$ if and only if it is a strong simplicial morphism $K_G \to \mathrm{dFl}(H)$ for some simplicial complex $\mathrm{Sk}_1(\mathrm{dFl}(G)) \subseteq K_G \subseteq \mathrm{dFl}(G)$;*
3. *f is a triangle-collapsing digraph map $G \to H$ if and only if it is a triangle-collapsing simplicial morphism $\mathrm{dFl}(G) \to \mathrm{dFl}(H)$.*

In particular, dFl *is a full and faithful functor* **WkDgr** $\to$ **WkOSC**, **TcDgr** $\to$ **TcOSC** *and* **StDgr** $\to$ **StOSC**.

Proof. Suppose f induces a weak digraph map $G \to H$. Take any simplex $p = v_0 \dots v_k \in \mathrm{dFl}(G)$. Then for every $i \leq j$, there is an edge $v_i \to v_j$ in G. Since f is a weak digraph map $f(v_i) \rightrightarrows f(v_j)$ for every $i \leq j$. Either these are all edges, in which case $f(p) \in \mathrm{dFl}(G)$, or there is an equality in which case $f(p)$ is non-simplicial. Hence, f induces a weak simplicial

morphism $\mathrm{dFl}(G) \to \mathrm{dFl}(H)$ and so certainly induces a weak simplicial morphism $K_G \to \mathrm{dFl}(H)$ for any $\mathrm{Sk}_1(\mathrm{dFl}(G)) \subseteq K_G \subset \mathrm{dFl}(H)$.

Now suppose f induces a weak simplicial morphism $\mathrm{Sk}_1(\mathrm{dFl}(G)) \to \mathrm{dFl}(H)$. Take an edge $v \to w$ in G. Note there is a simplex $vw \in \mathrm{Sk}_1(\mathrm{dFl}(G))$. Therefore, either $f(vw) = f(v)f(w) \in \mathrm{dFl}(H)$, in which case $f(v) \to f(w)$ in H, or $f(vw)$ is non-simplicial, in which case $f(v) = f(w)$. Hence, f induces a weak digraph map $G \to H$.

A similar argument shows the second point and the third point follows immediately from definitions. □

Another way to construct a path complex from a digraph is by considering the collection of all its directed paths. This is the object studied by path homology [54].

Definition 4.2.18. Given a directed graph $G = (V, E)$, the allowed path complex, $\mathcal{A}(G)$, is the path complex on V such that given an arbitrary elementary path $p = v_0 \dots v_k$,

$$v_0 \dots v_k \in \mathcal{A}(G) \iff \text{for all } i,\ v_i \to v_{i+1} \text{ in } G. \tag{4.6}$$

We refer to the elements of $\mathcal{A}(G)$ as the allowed paths (in G).

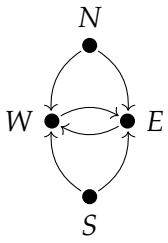


Figure 4.1: A simple example digraph G for which $\mathrm{dFl}(G) \neq \mathcal{A}(G)$

Example 4.2.19. Figure 4.1 shows an example digraph, G, for which $\mathcal{A}(G) \neq \mathrm{dFl}(G)$. First note that $EWE \in \mathcal{A}(G)$, but this path is not simplicial and hence does not belong to $\mathrm{dFl}(G)$. Also note that $\mathrm{dFl}(G)$ contains two simplices on the nodes $\{N, E, W\}$, namely NEW and NWE. We will see later, in Example 4.5.33, that $\mathrm{dFl}(G)$ has the homology of the 2-sphere whilst $\mathcal{A}(G)$ is contractible.

Note that $\mathcal{A}(G)$ is always a regular path complex but is not, in general, a simplicial complex. For example, consider any non-oriented digraph or, more generally, a digraph containing some path $a \to b \to c$ for which $a \not\to c$. The functoriality of this construction is simple: by a proof similar to that of Lemma 4.2.17 we obtain the following.

Lemma 4.2.20. *Given two digraphs G, H, and a vertex map $f : V(G) \to V(H)$,*

1. *f is a weak digraph map $G \to H$ if and only if it is a weak path morphism $P_G \to \mathcal{A}(H)$ for some path complex $\mathrm{Sk}_1(\mathcal{A}(G)) \subseteq P_G \subseteq \mathcal{A}(G)$;*
2. *f is a strong digraph map $G \to H$ if and only if it is a strong path morphism $P_G \to \mathcal{A}(H)$ for some path complex $\mathrm{Sk}_1(\mathcal{A}(G)) \subseteq P_G \subseteq \mathcal{A}(G)$.*

In particular, $\mathcal{A}$ is a full and faithful functor **WkDgr** $\to$ **WkRPC** *and* **StDgr** $\to$ **StRPC**.

In light of these results, we note that both $\mathrm{dFl}(G)$ and $\mathcal{A}(G)$ can be viewed as regular path complexes and **TcDgr** is the largest category on which both constructions are functorial. Therefore, we can view both constructions as functors **TcDgr** $\to$ **WkRPC** and then observe the following relation.

Lemma 4.2.21. *When viewed as functors* **TcDgr** $\to$ **WkRPC**, dFl *is a subfunctor of* $\mathcal{A}$.

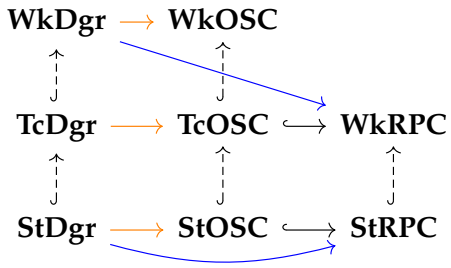
Proof. Given a digraph G, if $p = v_0 \dots v_k$ is a directed k-clique then p is certainly an allowed path. Therefore, $\mathrm{dFl}(G) \subset \mathcal{A}(G)$. Moreover, given a triangle-collapsing digraph map $f : G \to H$, there are induced maps $\mathrm{dFl}(f) : \mathrm{dFl}(G) \to \mathrm{dFl}(H)$ and $\mathcal{A}(f) : \mathcal{A}(G) \to \mathcal{A}(H)$ which trivially commute with the inclusions. □

We summarise this section with the diagram of functors shown in Figure 4.2. Each black arrow is an inclusion of subcategories. The vertical, dashed, black arrows are inclusions of wide subcategories. The horizontal, solid, black arrows are inclusions of full subcategories. Each orange arrow is the directed flag complex functor, dFl, and each blue arrow is the allowed path complex functor, $\mathcal{A}$. By Lemmas 4.2.17 and 4.2.20, the coloured arrows are full and faithful functors. The sub-diagram of black and orange arrows commutes, as does the sub-diagram of black and blue arrows. Given two paths through the diagram between the same categories, one containing an orange edge and one containing a blue edge, the path containing an orange edge yields a subfunctor of the other.

4.2.3 Chain complexes from regular path complexes

We now construct a chain complex which algebraically represents the combinatorial structure in a regular path complex. This construction is due to Grigor'yan et al. [54]. It is possible to repeat this construction for any path complex P, by first applying the regularisation functor, reg : **WkPathC** $\to$ **WkRPC**.

To begin, we choose an arbitrary field $\mathbb{F}$ and let **Ch** denote the category of non-negatively graded chain complexes of vector spaces over $\mathbb{F}$. Given a regular path complex P, generate



Wk : Weak maps/morphisms
Tc : Triangle-collapsing maps/morphisms
St : Strong maps/morphisms
Dgr : Digraphs
OSC : Ordered simplicial complexes
RPC : Regular path complexes
$\to$: Allowed path functor, $\mathcal{A}$
$\to$: Directed flag complex functor, dFl

Figure 4.2: The categories defined in this section, and the functors between them. The black arrows denote inclusions of subcategories; the dashed arrows are wide inclusions whilst the solid arrows are full inclusions. The orange arrows denote dFl and the blue arrows denote $\mathcal{A}$.

the following free $\mathbb{F}$-vector spaces:

$$\Lambda_k(P) := \mathbb{F}\left\langle \{p = v_0 \dots v_k \mid p \text{ is an elementary } k\text{-path on } V(P)\} \right\rangle, \tag{4.7}$$

$$\mathcal{R}_k(P) := \mathbb{F}\left\langle \{p = v_0 \dots v_k \mid p \text{ is a regular } k\text{-path on } V(P)\} \right\rangle, \tag{4.8}$$

$$\mathcal{I}_k(P) := \mathbb{F}\left\langle \{p = v_0 \dots v_k \mid p \text{ is an irregular } k\text{-path on } V(P)\} \right\rangle, \tag{4.9}$$

$$C_k(P) := \mathbb{F}\left\langle \{p = v_0 \dots v_k \mid p \in P\} \right\rangle. \tag{4.10}$$

We will not distinguish between paths and their corresponding generators in these free groups. We drop P from notation when it is clear from context.

Remark 4.2.22.
1. The vector spaces $\Lambda_k(P)$, $\mathcal{R}_k(P)$ and $\mathcal{I}_k(P)$ each depend only on $V(P)$.
2. There is always a direct sum decomposition $\Lambda_k = \mathcal{R}_k \oplus \mathcal{I}_k$.
3. Since we only consider regular path complexes, there is a subspace relation, $C_k(P) \subseteq \mathcal{R}_k(P)$.

We can define a non-regular boundary map amongst the $\Lambda_\bullet$ in the usual fashion: given a standard basis element $v_0 \dots v_k \in \Lambda_k$, define

$$\partial_k^{\text{nr}}(v_0 \dots v_k) := \sum_{i=0}^{k} (-1)^i v_0 \dots \hat{v_i} \dots v_k \tag{4.11}$$

and then extend linearly to a map $\partial_k^{\text{nr}} : \Lambda_k \to \Lambda_{k-1}$. As usual, $\partial_{k-1}^{\text{nr}} \circ \partial_k^{\text{nr}} = 0$ [54, Lemma 2.3].

Next, one can verify that $\partial_k(\mathcal{I}_k) \subseteq \mathcal{I}_{k-1}$ [54, Lemma 2.9] and hence ∂_k^{nr} passes to the quotient to define the regular boundary map $\partial_k : \mathcal{R}_k \to \mathcal{R}_{k-1}$. Intuitively, this boundary map can be computed by following the usual formula (4.11) and then dropping any summands

corresponding to irregular paths. The chain complex equation passes to the quotient to give $\partial_{k-1} \circ \partial_k = 0$.

Unfortunately, this boundary map does not restrict to a map $C_k \to C_{k-1}$ because, a priori, P is not a simplicial complex. Instead, we must pass to a further subspace

$$\Omega_k(P) := \{p \in C_k(P) \mid \partial_k(p) \in C_{k-1}(P)\}. \tag{4.12}$$

Notice that the chain complex equation implies that ∂_k restricts to a boundary map $\partial_k : \Omega_k \to \Omega_{k-1}$. Therefore, $\{\Omega_k, \partial_k\}_{k\geq 0}$ forms a chain complex, which we call the regular chain complex of P.

In general, elements of $\Omega_k(P)$ are *sums* of paths in P. Indeed, there may not even be a basis of $\Omega_k(P)$ for which each basis element is supported on a single path. Computing a basis for $\Omega_k(\mathcal{A}(G))$ is often non-trivial [38], especially for $k > 2$.

Theorem 4.2.23 ([55, Theorem 2.10])**.** *The regular chain complex of a regular path complex can be made into a functor* $\Omega : \mathbf{WkRPC} \to \mathbf{Ch}$.

Proof. Let $f : P_1 \to P_2$ be a weak path morphism of regular path complexes. We will define $f_\# : \Omega(P_1) \to \Omega(P_2)$ on each component. First we define maps $\mathcal{R}_k(f) : \mathcal{R}_k(P_1) \to \mathcal{R}_k(P_2)$ on the standard basis. Given an arbitrary path $p \in P_1$, let

$$\mathcal{R}_k(f)(p) := \begin{cases} f(p) & \text{if } f(p) \text{ is regular,} \\ 0 & \text{otherwise.} \end{cases} \tag{4.13}$$

One can verify that this yields a chain map and moreover, since f is a weak path morphism, $\mathcal{R}_k(f)(\Omega_k(P_1)) \subseteq \Omega_k(P_2)$. We let $f_\#$ denote the restriction of $\mathcal{R}_k(f)$ to $\Omega_k(P_1) \to \Omega_k(P_2)$. The construction of $\mathcal{R}_k(f)$ is clearly functorial and hence so is the restriction. □

Remark 4.2.24. There is a corresponding non-regular chain complex, which repeats the construction above but with the non-regular boundary map ∂^{nr}. This can distinguish between non-regular path complexes that share the same regularisation. However, this construction can only be made into a functor $\mathbf{StPathC} \to \mathbf{Ch}$. We refer the interested reader to [53].

Given a weak path morphism of regular path complexes $f : P_1 \to P_2$, we let $f_\# := \Omega(f)$ denote the chain map induced by Ω. We can compose Ω with the homology functor (in some degree k) to obtain a functor $H_k \circ \Omega : \mathbf{WkRPC} \to \mathbf{Vec}$. For brevity's sake, we denote $H_k(P) := H_k(\Omega(P))$ for a regular path complex P and $f_* := H_k(\Omega(f))$ for a weak path morphism f. We also use similar notational shortcuts when composing with the functor

which takes homology in every degree simultaneously, viewed as a graded vector space, $H : \mathbf{Ch} \to \mathbf{grVec}$. Finally, given an object or morphism in **WkPathC**, we use the same notation to denote these induced objects/morphisms, after first applying the reg functor.

In the case where P is in fact a simplicial complex, this construction simplifies substantially.

Lemma 4.2.25. *If P is an ordered simplicial complex then $\Omega_k(P) = C_k(P)$ for every k and moreover*

$$\partial_k(v_0 \dots v_k) := \sum_{i=0}^{k} (-1)^i v_0 \dots \hat{v}_i \dots v_k \tag{4.14}$$

for every $v_0 \dots v_k \in P$.

Proof. Given a path $p = v_0 \dots v_k \in P$, all the vertices are distinct because p is simplicial. Therefore, all the sub-paths $v_0 \dots \hat{v}_i \dots v_k$ are regular and thus equation (4.14) correctly computes the *regular* boundary map. Next, all the sub-paths belong to P because P is a simplicial complex and thus $\partial_k(p) \in C_{k-1}(P)$ and moreover $p \in \Omega_k(P)$. Since p was arbitrary, this shows that $C_k(P) \subseteq \Omega_k(P)$. □

If $P = \mathcal{A}(G)$ or $P = \mathrm{dFl}(G)$ for some digraph G, then $H_k(P)$ can be used as a functorial, algebraic invariant of G. For example, H_0 encodes the weakly connected components of G.

Proposition 4.2.26 ([54, Proposition 3.24])**.** *Given any digraph G, $H_0(\mathrm{dFl}(G)) = H_0(\mathcal{A}(G))$ and moreover* $\dim H_0(\mathrm{dFl}(G))$ *is the number of weakly connected components of G.*

4.3 Systems of one-step homotopies

All the homotopies presented herein are generated by a class of directional, one-step homotopies. The one-step homotopies determine a binary relation on the morphisms, which is subsequently completed to an equivalence relation by identifying 'zig-zags' of one-step homotopies. In this section, we present a small framework for developing these systems of one-step homotopies and relating them via functors.

Definition 4.3.1. Fix a locally small category $\mathcal{C}$.

1. A system of one-step homotopies $\mathcal{S}$ for $\mathcal{C}$ is a choice of digraph $\mathcal{S}(C_1, C_2)$ on the vertex set $\mathrm{Mor}_{\mathcal{C}}(C_1, C_2)$, for every pair of objects $C_1, C_2 \in C$.
2. We call an edge $f \to g$ in $\mathcal{S}(C_1, C_2)$ a one-step $\mathcal{S}$-homotopy from f to g.
3. If there is an edge $f \to g$ or $g \to f$ in $\mathcal{S}(C_1, C_2)$ then we say f and g are one-step $\mathcal{S}$-homotopic and write $f \simeq_{\mathcal{S},1} g$.

4. We say two morphisms f, g are $\mathcal{S}$-homotopic, and write $f \simeq_{\mathcal{S}} g$, if they belong to the same weak path component of $\mathcal{S}(C_1, C_2)$, i.e. if there is a sequence of morphisms

$$f = f_0, f_1, \ldots, f_{m-1}, f_m = g \tag{4.15}$$

such that $f_{i-1} \simeq_{\mathcal{S},1} f_i$ for every i. We refer to such a sequence of morphisms as a multi-step $\mathcal{S}$-homotopy from f to g.

For brevity, we drop the subscript $\mathcal{S}$ from the notation when it is clear from context.

Remark 4.3.2. By construction, $\simeq_{\mathcal{S}}$ is an equivalence relation on $\mathrm{Mor}_{\mathcal{C}}(C_1, C_2)$.

In many systems of one-step homotopies (such as the usual homotopy theory of topological spaces) each edge $f \to g$ is evidenced by the existence one or more 'homotopies', which are themselves morphisms in $\mathcal{C}$. For an arbitrary system, including the main system $\mathcal{S}_{\mathrm{dFl}}$ which we will later define, this may not necessarily be the case. However, we will shortly define systems which are associated to a choice of 'cylinder functor' wherein each edge $f \to g$ is indeed associated to at least one such 'homotopy' morphism in $\mathcal{C}$.

A system of one-step homotopies can be pulled back through a functor. Under sufficient conditions this pull-back can induce 'the same' equivalence relation on the morphisms.

Definition 4.3.3. Given a functor $\mu : \mathbf{C} \to \mathbf{D}$ and a system of one-step homotopies $\mathcal{S}$ for $\mathcal{D}$, the pull-back is the system of one-step homotopies $\mu^*\mathcal{S}$ for $\mathcal{C}$ given by

$$f \to g \text{ in } \mu^*\mathcal{S}(C_1, C_2) \iff \mu(f) \rightrightarrows \mu(g) \text{ in } \mathcal{S}(\mu(C_1), \mu(C_2)). \tag{4.16}$$

Remark 4.3.4. Given a pair of functors $\mu : \mathcal{C}_0 \to \mathcal{C}_1$, $\nu : \mathcal{C}_1 \to \mathcal{C}_2$ and a system of one-step homotopies $\mathcal{S}$ for $\mathcal{C}_2$, one can check that $(\nu \circ \mu)^*\mathcal{S} = \mu^*\nu^*\mathcal{S}$.

For every pair of objects, μ defines a weak digraph map $\mu^*\mathcal{S}(C_1, C_2) \to \mathcal{S}(\mu(C_1), \mu(C_2))$. Indeed, an alternative definition of $\mu^*\mathcal{S}$ is that $\mu^*\mathcal{S}(C_1, C_2)$ is the largest digraph on $\mathrm{Mor}_{\mathcal{C}}(C_1, C_2)$ such that μ induces such a weak digraph map. If μ is faithful then each of these vertex maps is injective and hence forms a strong digraph map. If μ is full and faithful then each map is bijective and hence forms an isomorphism of digraphs.

Lemma 4.3.5. *Given a functor $\mu : \mathbf{C} \to \mathbf{D}$ and a system of one-step homotopies $\mathcal{S}$ for $\mathcal{D}$ then*

$$f \simeq_{\mu^*\mathcal{S}} g \implies \mu(f) \simeq_{\mathcal{S}} \mu(g). \tag{4.17}$$

If μ is a full functor then

$$f \simeq_{\mu^*\mathcal{S}} g \iff \mu(f) \simeq_{\mathcal{S}} \mu(g). \tag{4.18}$$

Proof. If $f \simeq_{\mu^*\mathcal{S}} g$ then there is a sequence of morphisms

$$f = f_0, f_1, \ldots, f_{m-1}, f_m = g \tag{4.19}$$

such that $f_{i-1} \to f_i$ in $\mu^*\mathcal{S}(C_1, C_2)$ for every i. Hence,

$$\mu(f) = \mu(f_0), \mu(f_1), \ldots, \mu(f_{m-1}), \mu(f_m) = \mu(g) \tag{4.20}$$

is a sequence of morphisms, such that for every i either $\mu(f_{i-1}) = \mu(f_i)$ or $\mu(f_{i-1}) \to \mu(f_i)$ in $\mathcal{S}(\mu(C_1), \mu(C_2))$. After removing duplicate morphisms, we see $\mu(f) \simeq_{\mathcal{S}} \mu(g)$.

If μ is full then, by definition, μ is surjective as a vertex map $\mathrm{Mor}_{\boldsymbol{\mathcal{C}}}(C_1, C_2) \to \mathrm{Mor}_{\boldsymbol{\mathcal{D}}}(\mu(C_1), \mu(C_2))$. Hence, if $\mu(f) \simeq_{\mathcal{S}} \mu(g)$ then there is a sequence of morphisms as in equation (4.20) such that $\mu(f_{i-1}) \to \mu(f_i)$ in $\mathcal{S}(\mu(C_1), \mu(C_2))$ for every i. By the construction of the pull-back

$$f = f_0, f_1, \ldots, f_{m-1}, f_m = g \tag{4.21}$$

is a sequence of morphisms showing that $f \simeq_{\mu^*\mathcal{S}} g$. □

This equivalence relation for morphisms can induce an equivalence relation on the objects of $\boldsymbol{\mathcal{C}}$, in the standard way. However, we must first check that the relations between the morphisms compose transitively.

Definition 4.3.6. Let $\mathcal{S}$ be a system of one-step homotopies for a category $\boldsymbol{\mathcal{C}}$. We say two objects $C_1, C_2 \in \boldsymbol{\mathcal{C}}$ are $\mathcal{S}$-homotopy equivalent, and write $C_1 \simeq_{\mathcal{S}} C_2$, if there are a pair of morphisms $f : C_1 \to C_2$ and $g : C_2 \to C_1$ such that

$$g \circ f \simeq_{\mathcal{S}} \mathrm{id}_{C_1} \quad \text{and} \quad f \circ g \simeq_{\mathcal{S}} \mathrm{id}_{C_2}. \tag{4.22}$$

For brevity, we drop the subscript $\mathcal{S}$ from the notation $\simeq_{\mathcal{S}}$ when it is clear from context.

Definition 4.3.7. Let $\mathcal{S}$ be a system of one-step homotopies for a category $\boldsymbol{\mathcal{C}}$. We say $\mathcal{S}$ is transitive if given morphisms $g_1, g_2 \in \mathrm{Mor}_{\boldsymbol{\mathcal{C}}}(C_1, C_2)$, $f \in \mathrm{Mor}_{\boldsymbol{\mathcal{C}}}(C_0, C_1)$ and $h \in \mathrm{Mor}_{\boldsymbol{\mathcal{C}}}(C_2, C_3)$, such that $g_1 \simeq_{\mathcal{S}} g_2$, then $g_1 \circ f \simeq_{\mathcal{S}} g_2 \circ f$ and $h \circ g_1 \simeq_{\mathcal{S}} h \circ g_2$.

$$C_0 \xrightarrow{f} C_1 \underset{g_2}{\overset{g_1}{\rightrightarrows}} C_2 \xrightarrow{h} C_3 \tag{4.23}$$

Lemma 4.3.8. *Let $\mathcal{S}$ be a system of one-step homotopies for a category $\boldsymbol{\mathcal{C}}$. If $\mathcal{S}$ is transitive then $\simeq_{\mathcal{S}}$ is an equivalence relation on the objects of $\boldsymbol{\mathcal{C}}$.*

Proof. It is clear that $\simeq_{\mathcal{S}}$ is symmetric and reflexive, so it remains to show that $\simeq_{\mathcal{S}}$ is transitive. Suppose $C_0 \simeq_{\mathcal{S}} C_1$ and $C_1 \simeq_{\mathcal{S}} C_2$, then we have morphisms in the following arrangement

$$C_0 \underset{g_1}{\overset{f_1}{\rightleftarrows}} C_1 \underset{g_2}{\overset{f_2}{\rightleftarrows}} C_2 \tag{4.24}$$

such that each f_i, g_i pair satisfy the equations (4.22). Then

$$(g_1 \circ g_2) \circ (f_2 \circ f_1) = g_1 \circ (g_2 \circ f_2) \circ f_1 \simeq_{\mathcal{S}} g_1 \circ \mathrm{id}_{C_1} \circ f_1 = g_1 \circ f_1 \simeq_{\mathcal{S}} \mathrm{id}_{C_0}, \tag{4.25}$$

where in the first $\simeq_{\mathcal{S}}$ step we use the transitivity of $\mathcal{S}$. A similar proof shows that the opposite composition is $\mathcal{S}$-homotopic to id_{C_2}. Hence, we see $C_0 \simeq C_2$. □

A standard application of Lemmas 4.3.5 and 4.3.8 yields the following corollaries.

Corollary 4.3.9. *Given a full functor $\mu : \mathbf{C} \to \mathbf{D}$ and a system of one-step homotopies $\mathcal{S}$ for $\boldsymbol{\mathcal{D}}$, if $\mathcal{S}$ is transitive then the pull-back $\mu^*\mathcal{S}$ is transitive.*

Corollary 4.3.10. *Given a functor $\mu : \mathbf{C} \to \mathbf{D}$ and a system of one-step homotopies $\mathcal{S}$ for $\boldsymbol{\mathcal{D}}$,*

$$C_1 \simeq_{\mu^*\mathcal{S}} C_2 \implies \mu(C_1) \simeq_{\mathcal{S}} \mu(C_2). \tag{4.26}$$

If μ is a full functor then

$$C_1 \simeq_{\mu^*\mathcal{S}} C_2 \iff \mu(C_1) \simeq_{\mathcal{S}} \mu(C_2). \tag{4.27}$$

One way to construct a system of one-step homotopies is via a choice of cylinder functor (see [68, §I.2] for an introduction).

Definition 4.3.11. A cylinder functor for a category $\boldsymbol{\mathcal{C}}$ is a functor

$$\mathrm{Cyl} : \boldsymbol{\mathcal{C}} \to \boldsymbol{\mathcal{C}} \tag{4.28}$$

equipped with three natural transformations

$$\iota_0, \iota_1 : \mathrm{id}_{\boldsymbol{\mathcal{C}}} \Rightarrow \mathrm{Cyl}, \quad \text{and} \quad \rho : \mathrm{Cyl} \Rightarrow \mathrm{id}_{\boldsymbol{\mathcal{C}}}, \tag{4.29}$$

satisfying $\rho\iota_i = \mathrm{id}_c$ for each $i = 0, 1$.

With such a cylinder functor in hand, one can construct a system of one-step homotopies $\mathcal{S}$ for $\boldsymbol{\mathcal{C}}$ as follows. Given $f, g : C_1 \to C_2$, a one-step $\mathcal{S}$-homotopy from f to g is a morphism $F : \mathrm{Cyl}(C_1) \to C_2$ such that

$$F \circ \iota_0 = f \quad \text{and} \quad F \circ \iota_1 = g. \tag{4.30}$$

This determines the system $\mathcal{S}$ by including an arrow $f \to g$ in $\mathcal{S}(C_1, C_2)$ whenever such a one-step $\mathcal{S}$-homotopy exists. In a slight abuse of terminology, we use the term 'one-step $\mathcal{S}$-homotopy' to refer interchangeably between the edge in $\mathcal{S}(C_1, C_2)$ and a morphism that implies its existence. We refer to $\mathcal{S}$ as the system of one-step homotopies associated to Cyl.

Lemma 4.3.12 ([68, Lemma 2.3])**.** *The system of one-step homotopies associated to a cylinder functor is transitive.*

In this work, we opt for slightly more generality than cylinder functors in order to emphasise the important properties of the resultant systems. Moreover, in Section 4.5.2, we will see that it is not clear how to construct an appropriate cylinder functor in **TcDgr** without artificially inflating the category.

4.4 Homotopies for ordered simplicial complexes

In this section, we develop a system of one-step homotopies for ordered simplicial complexes, so that two homotopic simplicial morphisms induce identical maps on homology, through the functor $H \circ \Omega : \mathbf{TcOSC} \to \mathbf{grVec}$. In Section 4.4.1, we recall one such system for **WkRPC**, which was first defined in [53] in terms of a cylinder functor Cyl for **WkRPC**.

In Section 4.4.2, we pull back along the full inclusion $\mathbf{TcOSC} \hookrightarrow \mathbf{WkRPC}$, to obtain a system for ordered simplicial complexes. We then describe this pull-back in terms of a cylinder functor $\overline{\mathrm{Cyl}}$ for **TcOSC**. More precisely, in Theorem 4.4.8, we show that a vertex map F induces a one-step homotopy with respect to $\overline{\mathrm{Cyl}}$ if and only if it induces one with respect to Cyl.

For completeness, in Section 4.4.3 we instead consider embedding **TcOSC** in the category of simplicial sets. We repeat the process above for this inclusion, and show that the system of homotopies we obtain is identical to the one obtained in Section 4.4.2.

4.4.1 Homotopies for path complexes

Definition 4.4.1. Given a path complex P, the cylinder over P, $\mathrm{Cyl}(P)$, is a new path complex on $V(P) \times \{0, 1\}$ containing paths of the following forms

1. $(v_0, 0) \dots (v_k, 0)$ such that $v_0 \dots v_k \in P$;
2. $(v_0, 1) \dots (v_k, 1)$ such that $v_0 \dots v_k \in P$;
3. $(v_0, 0) \dots (v_i, 0)(v_i, 1) \dots (v_k, 1)$ such that $v_0 \dots v_k \in P$ and $0 \leq i \leq k$.

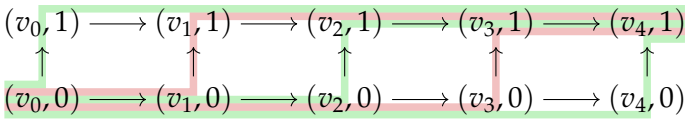


Figure 4.3: Visualisation of the lift $\mathfrak{L}(v_0v_1v_2v_3v_4)$. The lift is a sum of 5 different 5-paths, the green paths are added with coefficient $(+1)$ and the red paths are added with coefficient (-1).

This can be made into an endofunctor on **WkPathC** by sending a weak path morphism $f : P_1 \to P_2$ to the weak path morphism $\mathrm{Cyl}(f) : \mathrm{Cyl}(P_1) \to \mathrm{Cyl}(P_2)$, given by $(v, i) \mapsto (f(v), i)$. Moreover, there are weak path morphisms

$$\iota_i : P \to \mathrm{Cyl}(P), \quad v \mapsto (v, i), \tag{4.31}$$

$$\rho : \mathrm{Cyl}(P) \to P, \quad (v, i) \mapsto v \tag{4.32}$$

for $i = 0, 1$. These formulae determine natural transformations which, together with Cyl, constitute a cylinder functor for **WkPathC**. Since the cylinder of a regular path complex is again a regular path complex, this restricts to a cylinder functor on **WkRPC**. We let $\mathcal{S}_{\mathbf{WkPathC}}$ and $\mathcal{S}_{\mathbf{WkRPC}}$ denote the associated systems of one-step homotopies. Clearly, $\mathcal{S}_{\mathbf{WkRPC}} = \iota^* \mathcal{S}_{\mathbf{WkPathC}}$ where $\iota : \mathbf{WkRPC} \hookrightarrow \mathbf{WkPathC}$ is the inclusion of full subcategories.

As one might expect, the homology of a path complex is indeed invariant with respect to homotopy equivalence.

Theorem 4.4.2 ([53, Theorem 3.8])**.** *Suppose $f, g : P_1 \to P_2$ are $\mathcal{S}_{\mathbf{WkRPC}}$-homotopic weak path morphisms of regular path complexes, then there is a chain homotopy between $f_\#$ and $g_\#$. Therefore, f and g induce identical maps on homology, f_* and g_*. In particular, if $P_1 \simeq P_2$ then the homology groups $H(P_1)$ and $H(P_2)$ are isomorphic.*

We automatically inherit a similar result for any pull-back of $\mathcal{S}_{\mathbf{WkRPC}}$.

Corollary 4.4.3. *Given a functor $\mu : \mathcal{C} \to \mathbf{WkPathC}$, denote the pull-back $T := \mu^* \mathcal{S}_{\mathbf{WkPathC}}$. If $f \simeq_T g$, then $(H \circ \mu)(f)$ and $(H \circ \mu)(g)$ are identical maps on homology. In particular, if $C_1, C_2 \in \mathcal{C}$ are two objects such that $C_1 \simeq_T C_2$ then the homology groups $(H \circ \mu)(C_1)$ and $(H \circ \mu)(C_2)$ are isomorphic.*

Theorem 4.4.2 was shown in [53], we reproduce the proof here for clarity's seek and to make the chain homotopy explicit in the multi-step case. Given a regular path complex P, the lifting map is a linear map $\mathfrak{L} : \mathcal{R}_k(P) \to \mathcal{R}_{k+1}(\mathrm{Cyl}(P))$, which is given on the standard basis by

$$\mathfrak{L}(v_0 \dots v_k) := \sum_{i=0}^{k} (-1)^i \cdot (v_0, 0) \dots (v_i, 0)(v_i, 1) \dots (v_k, 1), \tag{4.33}$$

as visualised in Figure 4.3. Note that this clearly restricts to a linear map $\mathfrak{L} : C_k(P) \to C_{k+1}(\mathrm{Cyl}(P))$. The following result characterises how the lifting map interacts with the regular boundary map. For a proof, we refer the reader to [53].

Lemma 4.4.4 (Product Rule [53, Lemma 3.2, Theorem 3.8])**.** *Given a regular path complex P and $c \in \Omega_k(P)$,*

$$\partial\mathfrak{L}(v) + \mathfrak{L}(\partial c) = (\iota_1)_\#(v) - (\iota_0)_\#(c) \tag{4.34}$$

and hence the lifting map restricts to a linear map $\mathfrak{L} : \Omega_k(P) \to \Omega_{k+1}(\mathrm{Cyl}(P))$.

Proof of Theorem 4.4.2. To begin, we deal with the case where $f \simeq_1 g$. Suppose $F : \mathrm{Cyl}(P_1) \to P_2$ is a one-step homotopy from f to g. We define the chain homotopy $L_k : \Omega_k(P_1) \to \Omega_{k+1}(P_2)$ by $L_k(c) := F_\#(\mathfrak{L}(c))$. Using the fact that $F_\#$ is a chain map, along with the product rule, given any $c \in \Omega_k(P_1)$,

$$\begin{aligned}
(\partial L_k + L_{k-1}\partial)(c) &= \partial F_\#\mathfrak{L}(c) + F_\#\mathfrak{L}(\partial c) && (4.35)\\
&= F_\#\partial\mathfrak{L}(c) + F_\#\mathfrak{L}(\partial c) && (4.36)\\
&= F_\#(\partial\mathfrak{L}(c) + \mathfrak{L}(\partial c)) && (4.37)\\
&= F_\#((\iota_1)_\#(c) - (\iota_0)_\#(c)) && (4.38)\\
&= g_\#(c) - f_\#(c). && (4.39)
\end{aligned}$$

Hence, $\{L_k\}_{k\geq 0}$ constitutes a chain homotopy between $f_\#$ and $g_\#$, as required.

In the general case, suppose there is a finite sequence of weak path morphisms

$$f = f_0, f_1, \ldots, f_{m-1}, f_m = g \tag{4.40}$$

such that $f_{i-1} \simeq_1 f_i$ for $i = 1, \ldots, m$. For each i, there is a one-step homotopy $F_i : \mathrm{Cyl}(P_1) \to P_2$ between f_i and f_{i+1}. If F_i is from f_i to f_{i+1}, let $\alpha_i = 1$. If F_i is in the opposite direction, let $\alpha_i = -1$. By the argument above, F_i induces a chain homotopy $\{L_k^{(i)} : \Omega_k(P_1) \to \Omega_{k+1}(P_2)\}_{k\geq 0}$ that satisfies

$$\alpha_i \cdot (\partial L_k^{(i)} + L_{k-1}^{(i)}\partial)(c) = (f_{i+1})_\#(c) - (f_i)_\#(c) \tag{4.41}$$

for every $c \in \Omega_k(P_1)$. Therefore, if we define $L_k := \sum_{i=1}^m \alpha_i \cdot L_k^{(i)}$, then $\{L_k\}_{k\geq 0}$ is a chain homotopy between $f_\#$ and $g_\#$, as required. □

4.4.2 Ordered simplicial complexes as path complexes

Due to Lemma 4.2.10, there is an inclusion of full subcategories **TcOSC**, $\iota : \mathbf{TcOSC} \hookrightarrow \mathbf{WkRPC}$. Therefore, the system of one-step homotopies described in Section 4.4.1 restricts to a system for ordered simplicial complexes, by taking the pull-back $\iota^* \mathcal{S}_{\mathbf{WkRPC}}$. In this section, we seek to characterise this system intrinsically within **TcOSC**, i.e. without reference to path complexes or morphisms. First, we require the notion of the simplicial closure of a path complex.

Definition 4.4.5. A path complex, P, is called pre-simplicial if all of its constituent paths are simplicial. Given a pre-simplicial path complex P, its simplicial closure, $\overline{P}$ is the smallest simplicial complex containing K.

Remark 4.4.6.
1. A pre-simplicial path complex is necessarily regular.
2. The simplicial closure of a pre-simplicial path complex is well-defined because the set of all simplicial complexes is closed under arbitrary intersections.

It will be important to understand the structure of $\overline{\mathrm{Cyl}(K)}$, for a simplicial complex K, because a homotopy of path complexes has $\mathrm{Cyl}(K)$ as its domain. First we list the additional simplices in the closure of a cylinder.

Lemma 4.4.7. *Given any simplicial complex K, $\mathrm{Cyl}(K)$ is pre-simplicial. Moreover,*

$$\overline{\mathrm{Cyl}(K)} \setminus \mathrm{Cyl}(K) = \{(v_0, 0) \dots (v_i, 0)(v_{i+1}, 1)(v_k, 1) \mid v_0 \dots v_k \in K, 0 \leq i < k\}. \tag{4.42}$$

Proof. It is clear to see that all paths in the cylinder of K are simplicial. Then, for every $v_0 \dots v_k \in K$ and $0 \leq i < k$, $\mathrm{Cyl}(K)$ contains the path

$$(v_0, 0) \dots (v_i, 0)(v_{i+1}, 0)(v_{i+1}, 1) \dots (v_k, 1). \tag{4.43}$$

Therefore, in order to be closed under taking ordered subsets, $\overline{\mathrm{Cyl}(K)}$ must contain the path

$$(v_0, 0) \dots (v_i, 0)(v_{i+1}, 1) \dots (v_k, 1). \tag{4.44}$$

Finally, note that adding these paths to $\mathrm{Cyl}(K)$ yields a simplicial complex. □

With these, we are able to characterise the homotopies between morphisms in **TcOSC** without reference to path complexes or morphisms.

Theorem 4.4.8. *Given two triangle-collapsing simplicial morphisms $f, g : K_1 \to K_2$, a vertex map $F : V(K_1) \times \{0, 1\} \to V(K_2)$ induces a one-step $\mathcal{S}_{\mathbf{WkRPC}}$-homotopy from f to g if and only if F induces a triangle-collapsing simplicial morphism $\overline{\mathrm{Cyl}(K_1)} \to K_2$ such that*

$$F \circ \iota_0 = f \quad \text{and} \quad F \circ \iota_1 = g. \tag{4.45}$$

Proof. First, note that it suffices to show F induces a weak path morphism $\mathrm{Cyl}(K_1) \to K_2$ if and only if F induces a triangle-collapsing simplicial morphism $\overline{\mathrm{Cyl}(K_1)} \to K_2$, under the assumption that $F \circ \iota_0 = f$ and $F \circ \iota_1 = g$. Also note that f and g are weak path morphisms $K_1 \to K_2$, by Lemma 4.2.10.

One direction is straight-forward: if F induces a triangle-collapsing simplicial morphism $\overline{\mathrm{Cyl}(K_1)} \to K_2$ then, by Lemma 4.2.10, F induces a weak path morphism $\overline{\mathrm{Cyl}(K_1)} \to K_2$. Then, because $\mathrm{Cyl}(K_1) \subseteq \overline{\mathrm{Cyl}(K_1)}$, F must also induce a weak path morphism $\mathrm{Cyl}(K_1) \to K_2$.

For the opposite direction, suppose F induces a weak path morphism $\mathrm{Cyl}(K_1) \to K_2$. It remains to show that for $p \in \overline{\mathrm{Cyl}(K)} \setminus \mathrm{Cyl}(K)$, either $F(p) \in K_2$ or $F(p)$ is irregular. By Lemma 4.4.7, we can assume that p takes the form

$$(v_0, 0) \dots (v_i, 0)(v_{i+1}, 1) \dots (v_k, 1) \tag{4.46}$$

where $v_0 \dots v_k \in K_1$ and $0 \leq i < k$. Consider the cylinder path

$$p_0 := (v_0, 0) \dots (v_i, 0)(v_{i+1}, 0)(v_{i+1}, 1) \dots (v_k, 1) \in \mathrm{Cyl}(K_1). \tag{4.47}$$

Since F induces a weak path morphism $\mathrm{Cyl}(K_1) \to K_2$, either $F(p_0) \in K_2$ or $F(p_0)$ is irregular. If $F(p_0) \in K_2$ then we are done because p is a face of p_0 and K_2 is a simplicial complex. We deal with the remaining case, in which $F(p_0)$ is irregular. Consider the possible pairs of adjacent vertices in p_0 that have the same image under F. In most cases, this irregularity implies that $F(p)$ is also irregular. The remaining cases are $F(v_i, 0) = F(v_{i+1}, 0)$ and $F(v_{i+1}, 0) = F(v_{i+1}, 1)$.

Case 1: Suppose $F(v_i, 0) = F(v_{i+1}, 0)$. Then

$$F(p) = F\big((v_0, 0) \dots (v_{i-1}, 0)(v_i, 0)(v_{i+1}, 1)(v_{i+2}, 1) \dots (v_k, 1)\big) \tag{4.48}$$

$$= F\big((v_0, 0) \dots (v_{i-1}, 0)(v_{i+1}, 0)(v_{i+1}, 1)(v_{i+2}, 1) \dots (v_k, 1)\big) \in F(\mathrm{Cyl}(K)) \tag{4.49}$$

and hence either $F(p) \in K_2$ or $F(p)$ is irregular.

Case 2: Suppose $F(v_{i+1}, 0) = F(v_{i+1}, 1)$. Then

$$F(p) = F\big((v_0, 0) \dots (v_{i-1}, 0)(v_i, 0)(v_{i+1}, 1)(v_{i+2}, 1) \dots (v_k, 1)\big) \tag{4.50}$$

$$= F\big((v_0, 0) \dots (v_{i-1}, 0)(v_i, 0)(v_{i+1}, 0)(v_{i+2}, 1) \dots (v_k, 1)\big). \tag{4.51}$$

Iterating this argument, we either fall into a Case 1 (in which case $F(p) \in K_2$ or $F(p)$ is irregular), or we arrive at Case 2 with $i = k - 1$. This yields $F(v_k, 0) = F(v_k, 1)$ and

$$F(p) = F\big((v_0, 0) \dots (v_{i-1}, 0)(v_i, 0)(v_{i+1}, 1)(v_{i+2}, 1) \dots (v_k, 1)\big) \tag{4.52}$$

$$= F\big((v_0, 0) \dots (v_{k-1}, 0)(v_k, 1)\big). \tag{4.53}$$

From this we can conclude

$$F(p) = F\big((v_0, 0) \dots (v_{k-1}, 0)(v_k, 0)\big) = (F \circ \iota_0)(v_0 \dots v_k). \tag{4.54}$$

Then, since $F \circ \iota_0$ is a weak path morphism $K_1 \to K_2$, we obtain that either $F(p) \in K_2$ or $F(p)$ is irregular. □

The map $K \mapsto \overline{\mathrm{Cyl}(K)}$ is a functor $\overline{\mathrm{Cyl}} : \mathbf{TcOSC} \to \mathbf{TcOSC}$. Together with the usual natural transformations ι_i and ρ, this determines a cylinder functor for **TcOSC**. We denote the resulting system of one-step homotopies by $\mathcal{S}_{\mathbf{TcOSC}}$. To be explicit, in this system, a one-step $\mathcal{S}_{\mathbf{TcOSC}}$-homotopy from f to g is a triangle-collapsing simplicial morphism $F : \overline{\mathrm{Cyl}(K_1)} \to K_2$ such that

$$F \circ \iota_0 = f \quad \text{and} \quad F \circ \iota_0 = g. \tag{4.55}$$

Theorem 4.4.8 tells us that this is an intrinsic description of the pull-back system.

Corollary 4.4.9. *The pull-back system along the inclusion $\iota : \mathbf{TcOSC} \hookrightarrow \mathbf{WkRPC}$ is $\iota^* \mathcal{S}_{\mathbf{WkRPC}} = \mathcal{S}_{\mathbf{TcOSC}}$.*

We conclude this subsection with some simple technical results, which control the paths in $\overline{\mathrm{Cyl}(K)}$. These allow us to obtain an equivalent condition to F inducing a triangle-collapsing simplicial morphism $\overline{\mathrm{Cyl}(K_1)} \to K_2$, which depends only on the 1-skelton, $\mathrm{Sk}_1(K_1)$. This will be useful in the later sections.

Lemma 4.4.10. *Let K be a simplicial complex.*

1. *If $(v_0, j) \dots (v_k, j) \in \overline{\mathrm{Cyl}(K)}$ for some fixed j then $v_0 \dots v_k \in K$.*
2. *If $(v_0, j_0) \dots (v_k, j_k) \in \overline{\mathrm{Cyl}(K)}$ and $v_0 \dots v_k$ is simplicial then $v_0 \dots v_k \in K$.*
3. *If $(v_0, j_0) \dots (v_k, j_k) \in \overline{\mathrm{Cyl}(K)}$ and $v_i = v_j$ for $j > i$ then $j = i + 1$.*
4. *If $(v_0, j_0)(v_1, j_1)(v_2, j_2) \in \overline{\mathrm{Cyl}(K)}$ then $v_0 \neq v_2$.*

Proof. For the first point, suppose $p_1 := (v_0, j) \dots (v_k, j) \in \overline{\mathrm{Cyl}(K)}$ for some fixed j. Note that p_1 is not one of the simplices listed in Lemma 4.4.7, because the second coordinate is constant. Therefore, we have $p_1 \in \mathrm{Cyl}(K)$. Considering the three types of paths in $\mathrm{Cyl}(K)$ we see that $v_0 \dots v_k \in K$.

For the second point, suppose $p_2 := (v_0, j_0) \dots (v_k, j_k) \in \overline{\mathrm{Cyl}(K)}$ and $v_0 \dots v_k$ is simplicial. If the j_i are constant at 0 or 1 then p_2 is of the form above and hence $v_0 \dots v_k \in K$. Otherwise, note that we must have $p_2 \in \overline{\mathrm{Cyl}(K)} \setminus \mathrm{Cyl}(K)$ because the remaining paths

in $\mathrm{Cyl}(K)$ repeat vertices in the first coordinate. Hence, by Lemma 4.4.7, we see that $v_0 \dots v_k \in K$.

For the third point, suppose $p_3 := (v_0, j_0) \dots (v_k, j_k) \in \overline{\mathrm{Cyl}(K)}$. If $v_i = v_j$ for some $i < j$ then $v_0 \dots v_k$ is not simplicial. Considering the four types of paths in $\overline{\mathrm{Cyl}(K)}$, we see that p_3 must take the form

$$p_3 = (w_0, 0) \dots (w_i, 0)(w_i, 1)(w_{i+1}, 1) \dots (w_{k-1}, 1) \tag{4.56}$$

for some $w_0 \dots w_{k-1} \in K$. Since $w_0 \dots w_{k-1}$ is simplicial, the only repeated vertex is w_i, and hence we must have $j = i + 1$. The final point follows immediately from the third. □

Lemma 4.4.11. *Given two simplicial complexes K_1, K_2 and a weak simplicial morphism $F : \overline{\mathrm{Cyl}(K_1)} \to K_2$, denote $f := F \circ \iota_0$ and $g := F \circ \iota_1$. If both f and g are both triangle-collapsing simplicial morphisms $K_1 \to K_2$, then F is triangle collapsing if and only if*

$$xy \in K_1 \text{ and } f(x) = g(y) \quad \Longrightarrow \quad f(x) = f(y) = g(x) = g(y). \tag{4.57}$$

Proof. First, suppose F is triangle-collapsing. Then take $xy \in K_1$ such that $f(x) = g(y)$. This implies that $F(x, 0) = F(y, 1)$. Now note $(x, 0)(y, 0)(y, 1) \in \overline{\mathrm{Cyl}(K_1)}$ so $F(y, 0) = F(x, 0)$ and hence $f(y) = f(x)$. Likewise, $(x, 0)(x, 1)(y, 1) \in \overline{\mathrm{Cyl}(K_1)}$ so $F(x, 1) = F(x, 0)$ and hence $g(x) = f(x)$.

Now suppose F satisfies condition (4.57). Take a simplex $p = (v_0, j_0)(v_1, j_1)(v_2, j_2) \in \overline{\mathrm{Cyl}(K_1)}$ such that $F(v_0, j_0) = F(v_2, j_2)$. By Lemma 4.4.10.4, we may assume $v_0 \neq v_2$. If $p = \iota_j(p')$ for some $j \in \{0, 1\}$ and $p' \in K_1$ then $F(v_1, j_1) = F(v_0, j_0)$ because f and g are both triangle-collapsing. It remains to consider the case where $p = (v_0, 0)(v_1, j_1)(v_2, 1)$. Let us assume that $j_1 = 0$; the $j_1 = 1$ case admits a similar proof. First, note the condition $F(v_0, j_0) = F(v_2, j_2)$ becomes $f(v_0) = g(v_2)$. Since $j_1 = 0$ and p is simplicial, we must have $v_1 \neq v_0$. We split into two cases.

Case 1: Suppose $v_1 = v_2$. Then $(v_0, 0)(v_1, j_1) = (v_0, 0)(v_2, 0)$ is a face of $p \in \overline{\mathrm{Cyl}(K_1)}$, so we must have $(v_0, 0)(v_2, 0) \in \overline{\mathrm{Cyl}(K_1)}$. Therefore, by Lemma 4.4.10.1, we see $v_0 v_2 \in K_1$ and then condition (4.57) implies that

$$F(v_1, j_1) = f(v_1) = f(v_2) = f(v_0) = F(v_0, j_0). \tag{4.58}$$

Case 2: Suppose $v_1 \neq v_2$. In this case note that $v_0 v_1 v_2$ is simplicial and hence, by Lemma 4.4.10.2, $v_0 v_1 v_2 \in K_1$. Then, since K_1 is a simplicial complex, $v_0 v_2 \in K_1$ and hence condition (4.57) implies $g(v_2) = f(v_2)$. Now f is a triangle-collapsing simplicial morphism and $f(v_0) = g(v_2) = f(v_2)$ so $f(v_1) = f(v_0)$ which implies $F(v_1, j_1) = F(v_0, 0)$. □

4.4.3 Ordered simplicial complexes as simplicial sets

Those unfamiliar with path complexes may prefer to view an ordered simplicial complex as a simplicial set. For completeness, we consider this alternative perspective here. The remainder of this work is not dependent on this section.

First, we describe a full and faithful embedding $\sigma : \mathbf{TcOSC} \to \mathbf{sSet}$, into the category of simplicial sets. To each simplicial set, one can assign a chain complex, called its normalised Moore complex, via a functor $N : \mathbf{sSet} \to \mathbf{Ch}$ (see [48, §2]). From a homological point of view, these two perspectives are identical, in the sense that Ω and $N \circ \sigma$ are naturally isomorphic as functors $\mathbf{TcOSC} \to \mathbf{Ch}$. The category of simplicial sets is equipped with a system of one-step homotopies, via its 'left homotopies', which we denote $\mathcal{S}_{\mathbf{sSet}}$. In Corollary 4.4.21, we show that $\sigma^* \mathcal{S}_{\mathbf{sSet}} = \mathcal{S}_{\mathbf{TcOSC}}$ and hence these two perspectives yield identical systems of one-step homotopies. We begin with definitions of the simplex category and of simplicial sets, which facilitate the construction of σ.

Definition 4.4.12. Given an integer $n \geq 0$, the n^{th} ordinal is the ordered simplicial complex, $[n]$, on the vertex set $\{0, \ldots, n\}$ consisting of *all* strictly increasing sequences. Put another way, $[n]$ is the minimal simplicial complex containing the simplex $0 \ldots n$. The simplex category, $\mathbf{\Delta}$, is the full subcategory of **TcOSC**, with objects restricted to the ordinals.

Remark 4.4.13. Let G_n be the digraph on $\{0, \ldots, n\}$ with an edge $i \to j$ if and only if $i < j$. Then notice $\mathcal{A}(G_n) = \mathrm{dFl}(G_n) = [n]$. By Lemmas 4.2.10 and 4.2.20, a morphism $[m] \to [n]$ in $\mathbf{\Delta}$ is just a weak digraph map $G_m \to G_n$. This is equivalent to a (not necessarily strictly) increasing function $\{0, \ldots, m\} \to \{0, \ldots, n\}$.

Definition 4.4.14. A simplicial set is a contravariant functor from the simplex category into the category of sets, i.e. a functor $\mathbf{\Delta}^{\mathrm{op}} \to \mathbf{Set}$. A simplicial map is a morphism in the functor category $\mathbf{[\Delta^{op}, Set]}$, i.e. a natural transformation. This defines the category of simplicial sets, which we denote **sSet**.

Lemma 4.4.15. *There is a full and faithful embedding $\sigma : \mathbf{TcOSC} \to \mathbf{sSet}$.*

Proof. There is an inclusion of subcategories $\iota_{\mathbf{\Delta}} : \mathbf{\Delta} \to \mathbf{TcOSC}$. We define σ to be the restricted Yoneda embedding corresponding to this inclusion, i.e. $\sigma := \iota_{\mathbf{\Delta}}^* \circ Y$, where $\iota_{\mathbf{\Delta}}^*$ is the restriction functor $\iota_{\mathbf{\Delta}}^* : \mathbf{[TcOSC^{op}, Set]} \to \mathbf{[\Delta^{op}, Set]}$ and $Y : \mathbf{TcOSC} \to \mathbf{[TcOSC^{op}, Set]}$ is the Yoneda embedding. More concretely, given an ordinal $[n] \in \mathbf{\Delta}$,

$$\sigma(K)([n]) = \mathrm{Mor}_{\mathbf{TcOSC}}([n], K). \tag{4.59}$$

The proof that σ is full and faithful follows the usual proof that $\mathbf{\Delta}$ is dense in the category of small categories. Essentially, this amounts to showing that, given simplicial complexes K_1, K_2 and a natural transformation $\eta : \sigma(K_1) \Rightarrow \sigma(K_2)$, there is a unique vertex map $f : V(K_1) \to V(K_2)$ such that f induces a triangle-collapsing simplicial morphism $K_1 \to K_2$ and $\sigma(f) = \eta$. This vertex map is fully determined by the function

$$\eta_{[0]} : \mathrm{Mor}_{\mathbf{TcOSC}}([0], K_1) \to \mathrm{Mor}_{\mathbf{TcOSC}}([0], K_2), \tag{4.60}$$

which induces to a triangle-collapsing simplicial morphism precisely because η is natural. □

We now recall the homotopies in **sSet** and explore how they relate to the homotopies in **TcOSC**. First, we need the notion of product for a general category.

Definition 4.4.16. Given two objects $G, H \in \mathcal{C}$ in some category $\mathcal{C}$, the categorical product is an object $G \times_{\mathcal{C}} H$ together with morphisms $\pi_G : G \times_{\mathcal{C}} H \to G$ and $\pi_H : G \times_{\mathcal{C}} H \to H$ satisfying the following universal property:

(Universal Property of Products) Given $\phi_G : Y \to G$ and $\phi_H : Y \to H$, there is a unique morphism $f : Y \to G \times_{\mathcal{C}} H$ such that the following diagram commutes

$$\begin{array}{ccccc} & & Y & & \\ & \swarrow^{\phi_G} & \downarrow f & \searrow^{\phi_H} & \\ G & \xleftarrow{\pi_G} & G \times_{\mathcal{C}} H & \xrightarrow{\pi_H} & H \end{array} \tag{4.61}$$

Remark 4.4.17. The categorical product may not exist but, if it does, it is unique up to unique isomorphism.

Definition 4.4.18. Given a simplicial set $S \in \mathbf{sSet}$, the cylinder over S is the simplicial set

$$\mathrm{Cyl}(S) := S \times_{\mathbf{sSet}} \sigma([1]). \tag{4.62}$$

This construction can clearly be made into a functor $\mathrm{Cyl} : \mathbf{sSet} \to \mathbf{sSet}$. Since **sSet** is a category of functors into **Set** which is complete, this product exists and is computed 'pointwise'. That is, given an ordinal $[n] \in \mathbf{\Delta}$,

$$\mathrm{Cyl}(S)[n] = S[n] \times \sigma([1]) = S[n] \times \mathrm{Mor}_{\blacksquare}([n], [1]), \tag{4.63}$$

where $\times$ here denotes the usual Cartesian product of sets. This formula also extends to describe the structure morphisms, $\mathrm{Cyl}(S)(\theta)$, for each $\theta : [m] \to [n]$ in $\mathbf{\Delta}$. There are natural

transformations $e_0, e_1 : \mathrm{id}_{\mathbf{sSet}} \Rightarrow \mathrm{Cyl}$ and $\rho : \mathrm{Cyl} \Rightarrow \mathrm{id}_{\mathbf{sSet}}$. On each component $S \in \mathbf{sSet}$, these are given on the ordinals $[n] \in \mathbf{\Delta}$ by

$$(e_i)_S[n] : S[n] \to \mathrm{Cyl}(S)[n], \quad \tau \mapsto (\tau, c_i) \tag{4.64}$$

and

$$(\rho)_S[n] : \mathrm{Cyl}(S)[n] \to S[n], \quad (\tau, i) \mapsto \tau, \tag{4.65}$$

where $c_i \in \mathrm{Mor}_{\mathbf{TcOSC}}([n], [1])$ denotes the constant morphism which sends all vertices to i. These natural transformations compose as follows $\rho \circ e_o = \rho \circ e_1 = \mathrm{id}_{\mathbf{sSet}}$ which gives Cyl the structure of a cylinder functor. We denote the associated system of one-step homotopies by $\mathcal{S}_{\mathbf{sSet}}$. Again, to be explicit, given simplicial maps $f, g : S_1 \to S_2$ a one-step $\mathcal{S}_{\mathbf{sSet}}$-homotopy from f to g is a simplicial map $F : \mathrm{Cyl}(S_1) \to S_2$ such that

$$F \circ e_0 = f \quad \text{and} \quad F \circ e_1 = g. \tag{4.66}$$

We refer the interested reader to [49, §1.6] for a more thorough treatment of simplicial homotopies. As with $\mathcal{S}_{\mathbf{WkRPC}}$, homotopic simplicial maps induce identical maps on homology.

Theorem 4.4.19 ([48, Theorem 2.4])**.** *Suppose $f, g : S_1 \to S_2$ are $\mathcal{S}_{\mathbf{sSet}}$-homotopic simplicial maps then there is a chain homotopy between $N(f)$ and $N(g)$. Therefore, f and g induce identical maps on homology. In particular, if $S_1 \simeq S_2$ then the homology groups $(H \circ N)(S_1)$ and $(H \circ N)(S_2)$ are isomorphic.*

Lemma 4.4.20. *For any simplicial complex K, $\mathrm{Cyl}(\sigma(K)) = \sigma(\overline{\mathrm{Cyl}(K)})$ (up to a natural isomorphism). Moreover, as natural transformations $\sigma\iota_i = e_i\sigma$.*

Proof. Given an ordinal $[n]$,

$$\mathrm{Cyl}(\sigma(K))[n] = \mathrm{Mor}_{\mathbf{TcOSC}}([n], K) \times \mathrm{Mor}_{\mathbf{TcOSC}}([n], [1]) \tag{4.67}$$

and

$$\sigma(\overline{\mathrm{Cyl}(K)})[n] = \mathrm{Mor}_{\mathbf{TcOSC}}([n], \overline{\mathrm{Cyl}(K)}). \tag{4.68}$$

There are triangle-collapsing simplicial morphisms $\pi_K : \overline{\mathrm{Cyl}(K)} \to K$ and $\pi_{[1]} : \overline{\mathrm{Cyl}(K)} \to [1]$, given by projection to each of the coordinates of the vertex set. Composition with these projections induces a simplicial map $\phi : \sigma(\overline{\mathrm{Cyl}(K)}) \to \mathrm{Cyl}(\sigma(K))$. At each ordinal $[n]$, this map is given by

$$\phi_{[n]} : \mathrm{Mor}_{\mathbf{TcOSC}}([n], \overline{\mathrm{Cyl}(K)}) \to \mathrm{Mor}_{\mathbf{TcOSC}}([n], K) \times \mathrm{Mor}_{\mathbf{TcOSC}}([n], [1]) \tag{4.69}$$

$$p \mapsto (\pi_K \circ p, \pi_{[1]} \circ p). \tag{4.70}$$

By Lemma 4.4.7, each $\phi_{[n]}$ is a bijection and hence ϕ is an isomorphism of simplicial sets. Moreover, the formula is natural in K and hence this is a natural isomorphism between the two functors $\mathbf{TcOSC} \to \mathbf{sSet}$. Finally, the equation $\sigma\iota_i = e_i\sigma$ follows automatically from unwinding the definitions, modulo the natural isomorphism ϕ. □

Corollary 4.4.21. *Given two triangle-collapsing simplicial morphisms $f, g : K_1 \to K_2$,*

1. *if $F : \overline{\mathrm{Cyl}(K_1)} \to K_2$ is a one-step homotopy from f to g then $\sigma(F)$ is a one-step homotopy from $\sigma(f)$ to $\sigma(g)$;*
2. *if $F : \mathrm{Cyl}(\sigma(K_1)) \to \sigma(K_2)$ is a one-step homotopy from $\sigma(f)$ to $\sigma(g)$ then $\sigma^{-1}(F)$ is a one-step homotopy from f to g.*

In particular, the pull-back along σ is $\sigma^*\mathcal{S}_{\mathbf{sSet}} = \mathcal{S}_{\mathbf{TcOSC}}$.

Proof. The first point follows immediately from Lemma 4.4.20. For the second point, up to a unique isomorphism F is a morphism $\sigma(\overline{\mathrm{Cyl}(K_1)}) \to \sigma(K_2)$. By Lemma 4.4.15, σ is full and faithful, so there is a unique morphism $G : \overline{\mathrm{Cyl}(K_1)} \to \mathrm{Cyl}(K_2)$ in **TcOSC** such that $\sigma(G) = F$. Since σ is faithful, the equations

$$F \circ e_0(\sigma(K_1)) = \sigma(G) \circ \sigma(\iota_0) = \sigma(f) \quad \text{and} \quad F \circ e_1(\sigma(K_2)) = \sigma(G) \circ \sigma(\iota_1) = \sigma(g) \tag{4.71}$$

imply $G \circ \iota_0 = f$ and $G \circ \iota_1 = g$. □

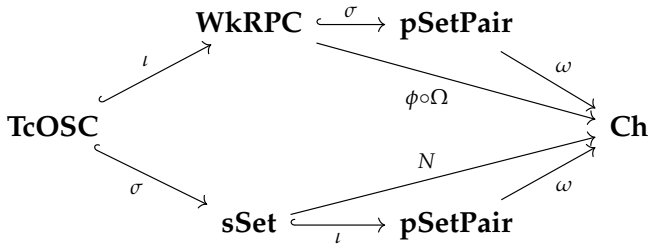
For the interested reader, these two perspectives can be unified by further embedding **sSet** and **WkRPC** in the category of path pairs of sets, **pSetPair**, which was introduced in [63].

Definition 4.4.22. A path set is a functor $\Pi^{\mathrm{op}} \to \mathbf{Set}$ where Π is the full subcategory of **WkRPC**, restricted to the objects $\{P_n\}_{n\geq 0}$ in which P_n is the minimal path complex containing the path $0 \dots n$. Any simplicial set restricts to a path set by composition with the inclusions $P_n \hookrightarrow [n]$. A path pair of sets is a pair (S, P) in which S is a simplicial set and P is a path subset of S. A morphism $(S_1, P_1) \to (S_2, P_2)$ in **pSetPair** is simply a simplicial map $f : S_1 \to S_2$ such that $f(P_1) \subseteq P_2$.

By [63, Proposition 9.6], there is a natural isomorphism ϕ such that the functors N and $\phi \circ \Omega$ factor through **pSetPair**, making the diagram of functors (4.73) commute. To be explicit, the functor $\iota : \mathbf{sSet} \to \mathbf{pSetPair}$ is given by the assignment $S \mapsto (S, S)$ whilst the functor $\sigma : \mathbf{WkRPC} \to \mathbf{pSetPair}$ can be defined as

$$\sigma(P) := \big(\mathrm{Mor}_{\mathbf{WkRPC}}(-, \mathcal{R}(P)),\ \mathrm{Mor}_{\mathbf{WkRPC}}(-, P)\big) \in [\mathbf{\Delta}^{\mathrm{op}}, \mathbf{Set}] \times [\Pi^{\mathrm{op}}, \mathbf{Set}], \tag{4.72}$$

where $\mathcal{R}(P)$ here denotes the path complex of all regular paths on $V(P)$. Note that $\iota\sigma \neq \sigma\iota$ because the simplicial set component of $\sigma\iota(K)$ is, in general, significantly larger than $\sigma(K)$.



(4.73)

The category **pSetPair** comes equipped with a cylinder functor [63, §9.1] which induces an associated system of one-step homotopies, $\mathcal{S}_{\mathbf{pSetPair}}$. This cylinder functor is derived from a box product for **pSetPair**. The cylinder functor for **WkRPC** can also be defined in terms of a box product for **pSetPair** and these two box products commute with σ by [63, Proposition 9.7]. Hence, $\sigma^*\mathcal{S}_{\mathbf{pSetPair}} = \mathcal{S}_{\mathbf{WkRPC}}$. Moreover, for a path pair of sets (S, P), the simplicial set component of the cylinder over (S, P) is $\mathrm{Cyl}(S)$ and hence $\mathcal{S}_{\mathbf{pSetPair}} = \mathcal{S}_{\mathbf{sSet}}$.

4.4.3.1 Other viewpoints

Since an ordered simplicial complex contains no 'degenerate' simplices, one may prefer to view an ordered simplicial complex as a semi-simplicial set (sometimes called a Δ-set). These objects are the combinatorial data underpinning what Hatcher calls a Δ-complex structure [57, §2.1]. One can factor the functor $\sigma : \mathbf{TcOSC} \to \mathbf{sSet}$ through the category of semi-simplicial sets, by composing with the functor which freely adds in all degeneracies to a semi-simplicial set (the functor G in [90, Lemma 1.1]). We choose to focus here on simplicial sets since they are better-studied in the literature and the two perspectives are compatible (see [90, §2]).

Finally, we note that there is a notion of homotopy theory for directed topological spaces, from the field of directed algebraic topology [44, §4.2]. To the authors' knowledge, this theory has not yet been connected to the various homotopy and homology theories of directed graphs. One could choose instead to view the directed flag complex as a directed topological space; this perspective requires further study.

4.5 Homotopies for digraphs

The over-arching goal for this section is to develop a system of one-step homotopies for **TcDgr**, to which the homology of the directed flag complex is invariant. In Section 4.5.1, we pull back along the allowed path complex functor $\mathcal{A} : \mathbf{WkDgr} \to \mathbf{WkRPC}$. We find that the

resulting system, $\mathcal{A}^*\mathcal{S}_{\mathbf{WkRPC}}$ coincides with the 'path homotopy' theory for digraphs, first introduced in [55]. In particular, the one-step homotopies can be characterised in terms of a cylinder functor for **WkDgr**. This coincidence is unsurprising since the path homology of digraphs was first defined as a functor which factors through **WkRPC** [54]. However, this connection motivates a similar study of the directed flag complex.

In Section 4.5.2, we repeat this process for the directed flag complex functor, dFl : **TcDgr** $\to$ **TcOSC**; this yields the desired system for **TcDgr**. Unlike in Section 4.5.1, there exist one-step $\mathcal{S}_{\mathbf{TcOSC}}$-homotopies which cannot be realised as the image of some morphism in **TcDgr**, through dFl. However, in Corollary 4.5.12, we obtain simple, edge-based conditions for the existence of these one-step $\mathcal{S}_{\mathbf{TcOSC}}$-homotopies and hence characterise $\mathrm{dFl}^*\,\mathcal{S}_{\mathbf{TcOSC}}$.

4.5.1 Pull-back along the allowed path functor

Consider the allowed path functor, $\mathcal{A}$: **WkDgr** $\to$ **WkRPC**. We can characterise the pull-back system $\mathcal{A}^*\mathcal{S}_{\mathbf{WkRPC}}$ intrinsically within the category **WkDgr**. First, in order to construct an appropriate cylinder functor, we must define sensible notions of intervals and products in **WkDgr**.

Definition 4.5.1. The standard unit interval is the digraph I on nodes $\{0,1\}$ whose only edge is $0 \to 1$.

Definition 4.5.2. Given two digraphs G, H, the box product, $G \boxdot H$ is a digraph on $V(G) \times V(H)$ such that

$$(x,y) \to (x',y') \iff (x = x \text{ and } y \to y') \text{ or } (x \to x' \text{ and } y = y'). \tag{4.74}$$

Taking the box product with the standard unit interval, gives us a 'cylinder' digraph. As one would hope, it is straightforward to show that $\mathcal{A}$ commutes with the relevant cylinder operations.

Lemma 4.5.3. *For any digraph G, $\mathcal{A}(G \boxdot I) = \mathrm{Cyl}(\mathcal{A}(G))$.*

Moreover, there are weak digraph maps $\iota_i : G \hookrightarrow G \boxdot I$ given by $\iota_i(v) := (v,i)$. Clearly $\mathcal{A}(\iota_i)$ is the component $(\iota_i)_G$ of the natural transformation ι_i, associated to the cylinder functor for **WkRPC**. This allows us to easily characterise when a vertex map constitutes a one-step homotopy.

Corollary 4.5.4. *Given two weak digraph maps $f, g : G \to H$, a vertex map $F : V(G) \times \{0,1\} \to V(H)$ is a one-step homotopy from $\mathcal{A}(f)$ to $\mathcal{A}(g)$ if and only if it induces a weak digraph map $F : G \boxdot I \to H$ such that*

$$F \circ \iota_0 = f \quad \textit{and} \quad F \circ \iota_1 = g. \tag{4.75}$$

Proof. The vertex map F is a weak digraph map $G \boxdot I \to H$ if and only if it is a weak path morphism $\mathcal{A}(G \boxdot I) \to \mathcal{A}(H)$, by Lemma 4.2.20. By Lemma 4.5.3, this is equivalent to F forming a weak path morphism $\mathrm{Cyl}(\mathcal{A}(G)) \to \mathcal{A}(H)$. □

We are led to the definition of a functor $\mathrm{Cyl} : \mathbf{WkDgr} \to \mathbf{WkDgr}$, given by $\mathrm{Cyl}(G) := G \boxdot I$. The inclusions ι_i are in fact natural transformation $\mathrm{id}_{\mathbf{WkDgr}} \Rightarrow \mathrm{Cyl}$ and the obvious projection map $\mathrm{Cyl}(G) \to G$ induces a natural transformation $\rho : \mathrm{Cyl} \Rightarrow \mathrm{id}_{\mathbf{WkDgr}}$. Together, this gives Cyl the structure of a cylinder functor and determines an associated system of one-step homotopies for **WkDgr**, which we denote $\mathcal{S}_{\mathcal{A}}$. For brevity, we use $\simeq_{\mathcal{A}}$, in lieu of $\simeq_{\mathcal{S}_{\mathcal{A}}}$, to denote the resulting equivalence relations. In this system, given two weak digraph maps $f, g : G \to H$, a one-step $\mathcal{S}_{\mathcal{A}}$-homotopy from f to g is a weak digraph map $F : G \boxdot I \to H$ such that

$$F \circ \iota_0 = f \quad \text{and} \quad F \circ \iota_1 = g. \tag{4.76}$$

Corollary 4.5.5. *The pull-back system along $\mathcal{A} : \mathbf{WkDgr} \to \mathbf{WkRPC}$ is $\mathcal{A}^* \mathcal{S}_{\mathbf{WkRPC}} = \mathcal{S}_{\mathcal{A}}$.*

A one-step $\mathcal{S}_{\mathcal{A}}$-homotopy is precisely the definition of a one-step digraph homotopy, in the sense introduced in [55]. Hence, $\mathcal{A}^* \mathcal{S}_{\mathbf{WkRPC}}$ coincides with the system introduced in [55]. It is also worth noting that this system appears in a broader framework of homotopy theories for digraphs, as the $(J_+, \boxplus)$-homotopy theory in [15]. To conclude, we note that this system enjoys a particularly simple characterisation of its one-step homotopies.

Corollary 4.5.6 ([55, §3])**.** *Given two weak digraph maps $f, g : G \to H$, there is a one-step $\mathcal{S}_{\mathcal{A}}$-homotopy from f to g if and only if for every $x \in V(G)$, $f(x) \overrightarrow{=} g(x)$.*

Remark 4.5.7. The category **WkDgr** admits a cofibration category structure in which the weak equivalences are those digraph maps which induce isomorphisms on path homology, $(H \circ \mathcal{A})(G)$ (see [19]). This structure can be refined so that the weak equivalences are those maps which induce an isomorphism on the entire second page of the magnitude path spectral sequence [60, Theorem 7.2].

4.5.2 Pull-back along the directed flag complex functor

We now consider the directed flag complex functor dFl : **TcDgr** $\to$ **TcOSC**; we seek an intrinsic characterisation of $\mathrm{dFl}^* \mathcal{S}_{\mathbf{TcOSC}}$. As before, we require a product structure for **TcDgr**. Note that, if $v_0 \to v_1$ is an edge in G then $(v_0, 0)(v_1, 1) \notin \mathrm{dFl}(G \boxdot I)$ and so $\overline{\mathrm{Cyl}(\mathrm{dFl}(G))} \not\subseteq \mathrm{dFl}(G \boxdot I)$. Hence, we must consider a larger product graph, to ensure that all the necessary directed cliques are present.

Definition 4.5.8. Given two digraphs G, H, the cross product, $G \times H$ is a digraph on $V(G) \times V(H)$ such that

$$(x, y) \to (x', y') \iff x \rightrightarrows x' \text{ and } y \rightrightarrows y' \text{ but not } (x = x' \text{ and } y = y'). \tag{4.77}$$

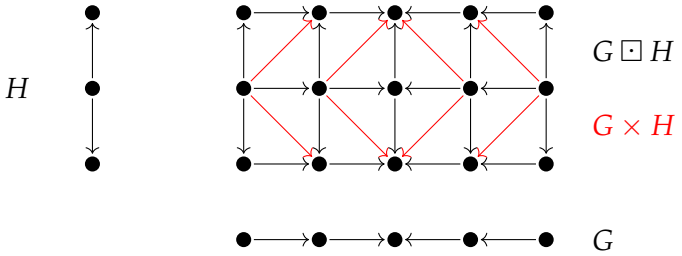


Figure 4.4: Illustration of the difference between $G \boxdot H$ and $G \times H$. The box product $G \boxdot H$ consists of just the black edges. The cross product $G \times H$ additionally contains the red edges.

For any digraphs G, H there is an inclusion $G \boxdot H \subseteq G \times H$, as illustrated in Figure 4.4. Unfortunately, dFl does not enjoy as simple a characterisation of its cylinders (c.f. Lemma 4.5.3 for $\mathcal{A}$). However, we can show that $G \times I$ is the minimal digraph whose directed flag complex contains $\overline{\mathrm{Cyl}(\mathrm{dFl}(G))}$. This is in analogy with $\overline{\mathrm{Cyl}(K)}$ being the minimal simplicial complex containing $\mathrm{Cyl}(K)$.

Lemma 4.5.9. *For any digraph G, $\mathrm{Cyl}(\mathrm{dFl}(G)) \subseteq \overline{\mathrm{Cyl}(\mathrm{dFl}(G))} \subseteq \mathrm{dFl}(G \times I)$. Moreover, $G \times I$ is the smallest such digraph, i.e.*

$$\mathrm{Cyl}(\mathrm{dFl}(G)) \subseteq \mathrm{dFl}(H) \implies G \times I \subset H. \tag{4.78}$$

Proof. To begin, note that

$$\overline{\mathrm{Cyl}(\mathrm{dFl}(G))} \subset \mathrm{dFl}(H) \iff \mathrm{Cyl}(\mathrm{dFl}(G)) \subseteq \mathrm{dFl}(H) \tag{4.79}$$

because $\mathrm{dFl}(H)$ is a simplicial complex. Therefore, it suffices to prove that $\mathrm{Cyl}(\mathrm{dFl}(G)) \subseteq \mathrm{dFl}(G \times I)$ and that $G \times I$ is the smallest such digraph.

Take a simplex $v_0 \dots v_k \in \mathrm{dFl}(G)$. Now for any $i < j$, $v_i \to v_j$ in G and certainly $0 = 0$, $1 = 1$ and $0 \to 1$ in I. Therefore, the paths $(v_0, 0) \dots (v_k, 0)$, $(v_0, 1) \dots (v_k, 1)$, and $(v_0, 0) \dots (v_i, 0)(v_i, 1) \dots (v_k, 1)$ are all cliques in $G \times I$. Hence, $\mathrm{Cyl}(\mathrm{dFl}(G)) \subseteq \mathrm{dFl}(G \times I)$.

Now suppose $\mathrm{Cyl}(\mathrm{dFl}(G)) \subseteq \mathrm{dFl}(H)$. Take a node $v \in V(G)$, then $(v, 0)(v, 1) \in \mathrm{Cyl}(\mathrm{dFl}(G))$, so there is an edge $(v, 0) \to (v, 1)$ in H. Take an edge $v \to w$ in G, then $(v, 0)(w, 0)(w, 1) \in \mathrm{Cyl}(\mathrm{dFl}(G))$ so there are edges $(v, 0) \to (w, 0)$ and $(v, 0) \to (w, 1)$ in H. Likewise, $(v, 0)(w, 1)(w, 1) \in \mathrm{Cyl}(\mathrm{dFl}(G))$ and hence there is an edge $(v, 1) \to (w, 1)$ in H. These are all the edges in $G \times I$ and hence $G \times I \subseteq H$. □

Despite this, it is typically a much stronger condition for a vertex map to induce a triangle-collapsing digraph map $G \times I \to H$ than to induce a triangle-collapsing simplicial morphism $\overline{\mathrm{Cyl}(\mathrm{dFl}(G))} \to \mathrm{dFl}(H)$. The main exception is when G is oriented.

Lemma 4.5.10. *Given a digraph G, the directed 3-cliques in $\mathrm{dFl}(G \times I) \setminus \overline{\mathrm{Cyl}(\mathrm{dFl}(G))}$ are*

$$\mathfrak{R} := \{(v_0, 0)(v_1, \alpha)(v_0, 1) \mid v_0 \circlearrowright v_1,\ \alpha = 0 \text{ or } 1\}. \tag{4.80}$$

Moreover, all simplices in $\mathrm{dFl}(G \times I) \setminus \overline{\mathrm{Cyl}(\mathrm{dFl}(G))}$ have at least one simplex from $\mathfrak{R}$ as a face. In particular, $\overline{\mathrm{Cyl}(\mathrm{dFl}(G))} = \mathrm{dFl}(G \times I)$ if and only if G is oriented.

Proof. Suppose there is a clique $p \in \mathrm{dFl}(G \times I) \setminus \overline{\mathrm{Cyl}(\mathrm{dFl}(G))}$. We can write

$$p = (v_0, 0) \dots (v_k, 0)(w_0, 1) \dots (w_m, 1) \tag{4.81}$$

where $v_0 \dots v_k$ and $w_0 \dots w_m$ are both simplices in $\mathrm{dFl}(G)$ and moreover $v_i \rightrightarrows w_j$ for every $0 \leq i \leq k$ and $0 \leq j \leq m$. Note that we must have $v_{i^*} = w_{j^*}$ for some (i^*, j^*) because otherwise $v_0 \dots v_k w_0 \dots w_m \in \mathrm{dFl}(G)$ and so $p \in \overline{\mathrm{Cyl}(\mathrm{dFl}(G))}$, by Lemma 4.4.7. Moreover, we can assume that $(i^*, j^*) \neq (k, 0)$ because if this is the only time that $v_i = w_j$ then $p \in \mathrm{Cyl}(\mathrm{dFl}(G))$. We split into cases depending on which coordinate differs.

Case 1: If $i^* \neq k$ then a face of p is

$$p' := (v_{i^*}, 0)(v_k, 0)(w_{j^*}, 1) = (v_{i^*}, 0)(v_k, 0)(v_{i^*}, 1). \tag{4.82}$$

This clique must be in $\mathrm{dFl}(G \times I)$ which implies that there is a reciprocal edge $v_{i^*} \circlearrowright v_k$.

Case 2: If $j^* \neq 0$ then a face of p is

$$p'' := (v_{i^*}, 0)(w_0, 0)(w_{j^*}, 1) = (v_{i^*}, 0)(w_0, 0)(v_{i^*}, 1). \tag{4.83}$$

This clique must be in $\mathrm{dFl}(G \times I)$ which implies that there is a reciprocal edge $v_{i^*} \circlearrowright w_0$. □

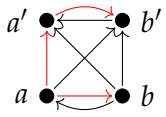


Figure 4.5: The digraph $G \times I$, as relabelled in the proof of Proposition 4.5.11. The tree, T, used to produce the basis for $\ker \partial_1$ is shown in red.

For general G, $\mathrm{dFl}(G \times I)$ may significantly differ from $\overline{\mathrm{Cyl}(\mathrm{dFl}(G))}$. Indeed, the two may not even be $\mathcal{S}_{\mathbf{TcOSC}}$-homotopy equivalent.

Proposition 4.5.11. *There exists a digraph G for which, as path complexes,* $\mathrm{Cyl}(\mathrm{dFl}(G)) \not\simeq \mathrm{dFl}(G \times I)$ *and* $\overline{\mathrm{Cyl}(\mathrm{dFl}(G))} \not\simeq \mathrm{dFl}(G \times I)$.

Proof. We first note that for any simplicial complex K, $\mathrm{Cyl}(K) \simeq K \simeq \overline{\mathrm{Cyl}(K)}$. The proof of this is relatively technical, so we delay these results to Propositions 4.A.4 and 4.A.5. Taking $K = \mathrm{dFl}(G)$, we see that it suffices to find G for which $\mathrm{dFl}(G) \not\simeq \mathrm{dFl}(G \times I)$.

Let G be the complete digraph on two nodes a, b, i.e. there is a reciprocal edge $a \circlearrowright b$ (see Figure 4.5). Since G contains no k-cliques for $k \geq 2$, a simple Euler characteristic argument shows that $\dim H_1(\mathrm{dFl}(G)) = 1$. Therefore, by Theorem 4.4.2, it suffices to show that $\dim H_1(\mathrm{dFl}(G \times I)) \neq 1$. In the sequel we show that in fact $\dim H_1(\mathrm{dFl}(G \times I)) = 0$. To reduce notation we change the names of the vertices in $G \times I$ as follows

$$(a,0) \mapsto a, \quad (b,0) \mapsto b, \quad (a,1) \mapsto a', \quad (b,1) \mapsto b'. \tag{4.84}$$

In the chain complex

$$\Omega_2(\mathrm{dFl}(G \times I)) \xrightarrow{\partial_2} \Omega_1(\mathrm{dFl}(G \times I)) \xrightarrow{\partial_1} \Omega_0(\mathrm{dFl}(G \times I)) \tag{4.85}$$

we can find a basis for $\ker \partial_1$ by first picking an oriented tree $T \subset G$ and then adding one cycle to the basis for each edge in $E(G) \setminus E(T)$. One such basis (corresponding to the tree shown in Figure 4.5) is as follows

$$c_1 := ab + ba \qquad = \partial_2(aba' + baa'), \tag{4.86}$$
$$c_2 := a'b' + b'a' \qquad = \partial_2(ab'a' + aa'b'), \tag{4.87}$$
$$c_3 := ab + ba' - aa' \qquad = \partial_2(aba'), \tag{4.88}$$
$$c_4 := aa' + a'b' - ab' \qquad = \partial_2(aa'b'), \tag{4.89}$$
$$c_5 := [aa' + a'b'] - [ab + bb'] = \partial_2(aa'b' - abb'). \tag{4.90}$$

Note that each of the 3-cliques on the right-hand side belongs to $\Omega_2(\mathrm{dFl}(G \times I))$. Therefore, every element of this basis for $\ker \partial_1$ is null-homologous. □

In light of this, we choose to characterise when there is a one-step $\mathcal{S}_{\mathbf{TcOSC}}$-homotopy from $\mathrm{dFl}(f)$ to $\mathrm{dFl}(g)$ in terms of edge-based conditions, akin to Corollary 4.5.6.

Corollary 4.5.12. *Given two triangle-collapsing digraph maps $f, g : G \to H$, a vertex map $F : V(G) \times \{0,1\} \to V(H)$ is a one-step $\mathcal{S}_{\mathbf{TcOSC}}$-homotopy from $\mathrm{dFl}(f)$ to $\mathrm{dFl}(g)$ if and only if the following conditions are all satisfied:*

1. $x \rightrightarrows y \implies f(x) \rightrightarrows g(y)$;
2. $x \to y$ *in* G *and* $f(x) = g(y) \implies f(x) = f(y) = g(x) = g(y)$;
3. $F \circ \iota_0 = f$ *and* $F \circ \iota_1 = g$.

Proof. First note that the third condition uniquely determines F and, furthermore, is necessary for F to induce a one-step $\mathcal{S}_{\mathbf{TcOSC}}$-homotopy from $\mathrm{dFl}(f)$ to $\mathrm{dFl}(g)$. Given this, the first condition becomes equivalent to F inducing a weak digraph map $G \times I \to H$. By Lemma 4.2.17, this is then equivalent to F inducing a weak simplicial morphism $\overline{\mathrm{Cyl}(\mathrm{dFl}(G))} \to \mathrm{dFl}(H)$. Then, Lemma 4.4.7 and Lemma 4.5.9, show us that

$$\mathrm{Sk}_1(\mathrm{dFl}(G \times I)) \subseteq \overline{\mathrm{Cyl}(\mathrm{dFl}(G))} \subseteq \mathrm{dFl}(G \times I). \tag{4.91}$$

Finally, Lemma 4.4.11 shows us that the second condition is equivalent to F inducing a triangle-collapsing simplicial morphism $\overline{\mathrm{Cyl}(\mathrm{dFl}(G))} \to \mathrm{dFl}(H)$. □

This begets the definition of a system of one-step homotopies for **TcDgr**.

Definition 4.5.13. For arbitrary digraphs G, H, we construct a digraph $\mathcal{S}_{\mathrm{dFl}}(G, H)$ on the vertex set $\mathrm{Mor}_{\mathbf{TcDgr}}(G, H)$ as follows. Given triangle-collapsing digraph maps $f, g : G \to H$, include an edge $f \to g$ if and only if

1. $x \rightrightarrows y \implies f(x) \rightrightarrows g(y)$; and
2. $x \to y$ and $f(x) = g(y) \implies f(x) = f(y) = g(x) = g(y)$.

Repeating this construction over all G, H yields a system of one-step homotopies, $\mathcal{S}_{\mathrm{dFl}}$, for **TcDgr**. For brevity, we use $\simeq_{\mathrm{dFl}}$, in lieu of $\simeq_{\mathcal{S}_{\mathrm{dFl}}}$, to denote the resulting equivalence relations.

Corollary 4.5.14. *The pull-back system along the directed flag complex functor* $\mathrm{dFl} : \mathbf{TcDgr} \to \mathbf{TcOSC}$ *is* $\mathrm{dFl}^* \mathcal{S}_{\mathbf{TcOSC}} = \mathcal{S}_{\mathrm{dFl}}$.

Comparing the conditions in Definition 4.5.13 to Corollary 4.5.4, we see that $\mathcal{S}_{\mathrm{dFl}}$ is a strictly smaller system.

Corollary 4.5.15. *For triangle-collapsing digraph maps $f, g : G \to H$, $f \simeq_{\mathrm{dFl}} g \implies f \simeq_{\mathcal{A}} g$*

This system is of a markedly different quality: we do not define the system in terms of a cylinder functor. However, when G is oriented, Lemma 4.5.10 makes it possible to find morphisms in **TcDgr** which induce one-step $\mathcal{S}_{\mathbf{TcOSC}}$-homotopies.

Corollary 4.5.16. *Let G be an oriented digraph and let $f, g : G \to H$ be triangle-collapsing digraph maps. A vertex map $F : V(G) \times \{0,1\} \to V(H)$ induces a one-step $\mathcal{S}_{\mathbf{TcOSC}}$-homotopy from* $\mathrm{dFl}(f)$ *to* $\mathrm{dFl}(g)$ *if and only if it induces a triangle-collapsing digraph map $F : G \times I \to H$, and*

$$F \circ \iota_0 = f \quad \textit{and} \quad F \circ \iota_1 = g. \tag{4.92}$$

Remark 4.5.17. Given triangle-collapsing digraph maps $f, g : G \to H$, if H is oriented then condition 2 of Definition 4.5.13 is automatically satisfied and hence a one-step $\mathcal{S}_{\mathrm{dFl}}$-homotopy reduces to a weak digraph map $F : G \times I \to H$ such that $F \circ \iota_0 = 0$ and $F \circ \iota_1 = 1$. This coincides with the notion of a $(J_+, \times)$-homotopy introduced in [15].

4.5.3 Comparing the two pull-backs

To illustrate the difference between the two systems, we now discuss a number of standard examples of $\simeq_{\mathcal{A}}$ equivalence that require extra conditions in order to pass over to the new system, $\mathcal{S}_{\mathrm{dFl}}$. These examples may help the practitioner to decide which homology theory is appropriate for a given application. We will primarily be concerned with when a digraph retracts onto one of its induced subgraphs.

Definition 4.5.18. Given a digraph $G = (V, E)$ an induced subgraph is a digraph consisting of some of the vertices in V but all the edges between those vertices, i.e. a digraph of the form $A = (V_A, E_A)$ where $V_A \subseteq V$ and $E_A = E \cap (V_A \times V_A)$.

Definition 4.5.19. Let $\{*\}$ denote the digraph on a single vertex. We say that a digraph G is $\mathcal{A}$-contractible (resp. dFl-contractible) if $G \simeq_{\mathcal{A}} \{*\}$ (resp. $G \simeq_{\mathrm{dFl}} \{*\}$).

Remark 4.5.20. Any dFl-contractible digraph is $\mathcal{A}$-contractible.

Definition 4.5.21. Given a digraph G and an induced subgraph $A \subseteq G$, we say a weak digraph map $r : G \to A$ is a retraction if $r\big|_A = \mathrm{id}_A$. Let $\iota : A \to G$ denote the inclusion digraph map. If $\iota \circ r \simeq_{\mathcal{A}} \mathrm{id}_G$ then we say r is an $\mathcal{A}$-deformation retraction. If r is triangle-collapsing and $\iota \circ r \simeq_{\mathrm{dFl}} \mathrm{id}_G$ then we say r is a dFl-deformation retraction.

Grigor'yan et al. obtained a particularly simple necessary condition for r to be an $\mathcal{A}$-deformation retract.

Proposition 4.5.22 ([55, Corollary 3.7])**.** *Given a digraph G and a retraction $r : G \to A$, suppose further that*

$$\forall x \in V(G) \quad x \rightrightarrows r(x) \quad \text{or} \quad \forall x \in V(G) \quad r(x) \rightrightarrows x, \tag{4.93}$$

where edges are required in G, then r is an $\mathcal{A}$-deformation retraction.

A similar result can be obtained for dFl-deformation retractions.

Proposition 4.5.23. *Given a digraph G, and a triangle-collapsing retraction $r : G \to A$, suppose further that*

$$\forall x \in V(G) \quad x \rightrightarrows r(x) \quad \text{and} \quad \forall (x, y) \in E(G) \quad x \to r(y) \tag{4.94}$$

or

$$\forall x \in V(G) \quad r(x) \rightrightarrows x \quad \text{and} \quad \forall (x, y) \in E(G) \quad r(x) \to y, \tag{4.95}$$

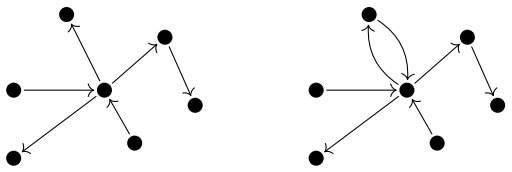
where edges are required in G, then r is a dFl*-deformation retraction.*

Proof. Let $\iota : A \to G$ denote the inclusion digraph map. Then $r \circ \iota = \mathrm{id}_A$ and $\iota \circ r$ is a digraph map $G \to G$, given by the same vertex map as r. Let us assume the first pair of conditions holds; the second pair admits a similar proof. We show that there is a one-step $\mathcal{S}_{\mathrm{dFl}}$-homotopy from id_G to $\iota \circ r$.

First note that both id_G and $\iota \circ r$ are triangle-collapsing digraph maps. Then, our assumptions on r imply that

$$x \rightrightarrows y \implies x \rightrightarrows r(y), \tag{4.96}$$

and hence the first condition of Definition 4.5.13 holds. For the second condition, take an edge $x \to y$ in G such that $\mathrm{id}_G(x) = (\iota \circ r)(y)$. Then our assumption on r implies that $x \to r(y) = x$. Therefore, the edge is a self-loop, which cannot exist. Hence, the second condition is vacuously satisfied. □



G_1 G_2

Figure 4.6: Examples of two pseudo-trees. The first example, G_1, is oriented and hence is both dFl-contractible and $\mathcal{A}$-contractible. The second example, G_2, is not oriented and hence is only $\mathcal{A}$-contractible.

Definition 4.5.24. We say a digraph G is a pseudo-tree if its underlying undirected graph (collapsing any reciprocal edges to a single edge) is a tree.

Proposition 4.5.25. *If G is a pseudo-tree then G is $\mathcal{A}$-contractible. Furthermore, a pseudo-tree is* dFl*-contractible if and only if it is oriented.*

Proof. For the first statement, we refer the reader to Example 3.10 of [55]. For the second, suppose G is not oriented. Then, since G is connected and not oriented, a standard induction argument shows $\#E(G) \geq \#V(G)$. Note that a pseudo-tree has no k-cliques for $k \geq 3$. Therefore, $\Omega(\mathrm{dFl}(G))$ has homology only in degrees 0 and 1. By considering the Euler characteristic, we obtain

$$\dim H_1(\mathrm{dFl}(G)) - \dim H_0(\mathrm{dFl}(G)) = \#E(G) - \#V(G) \geq 0 \tag{4.97}$$

Moreover, note that $\dim H_0(\mathrm{dFl}(G)) = 1$ since this counts the number of weakly connected components of G. Therefore, $H_1(\mathrm{dFl}(G))$ must be non-trivial and hence G is not dFl-contractible.

It remains to show that if G is oriented then G is dFl-contractible. We proceed by induction on the number of edges in G. If $\#E = 0$ then $G = \{*\}$, which is clearly dFl-contractible. Else, pick a leaf node v and let w be its unique neighbour. Let r be given by $r(v) := w$ and the identity elsewhere. Then we see that $G \setminus \{v\}$ is a dFl-deformation retract of G by Proposition 4.5.23. □



G_1 G_2

Figure 4.7: Two examples of star-like graphs. In both examples the centre node is the star centre. In the first example, there is no reciprocal edge and hence G_1 is both dFl-contractible and $\mathcal{A}$-contractible. In the second example, there is a reciprocal edge involving the star centre and hence G_2 is only $\mathcal{A}$-contractible.

Definition 4.5.26. We say a digraph G is star-like if there exists $v_0 \in V(G)$ such that for any other $v \in V(G)$, $v_0 \to v$. We say G is inverse star-like if there exists $v_0 \in V(G)$ such that for any other $v \in V(G)$, $v \to v_0$. We call v_0 a (inverted) star centre.

Corollary 4.5.27. *If G is star-like or inverse star-like then G is $\mathcal{A}$-contractible. If G has a (inverted) star centre v_0 and there are no reciprocal edges involving v_0, then G is* dFl*-contractible.*

Proof. For the first statement, we refer the reader to Example 3.11 of [55]. For the second statement, we deal with the case where G is star-like with star centre v_0, the inverse case admits a similar proof. Let G_0 be the induced subgraph on $\{v_0\}$ and let $r : G \to G_0$ be the unique digraph map. We show that r satisfies the conditions of Proposition 4.5.23.

Certainly r is triangle-collapsing. First, since v_0 is a star centre, for any $x \in V(G)$ we have $r(x) = v_0 \rightrightarrows x$. Second, given an edge $x \to y$, we cannot have $y = v_0$ because otherwise there would be a double edge $v_0 \circlearrowright x$. Since $y \neq v_0$, there must be an edge $v_0 = r(x) \to y$. □

Using these examples, we can show that the directed flag complex cannot be made into a functor from **WkDgr**. Indeed, even the subsequent chain complex cannot be made into such a functor.

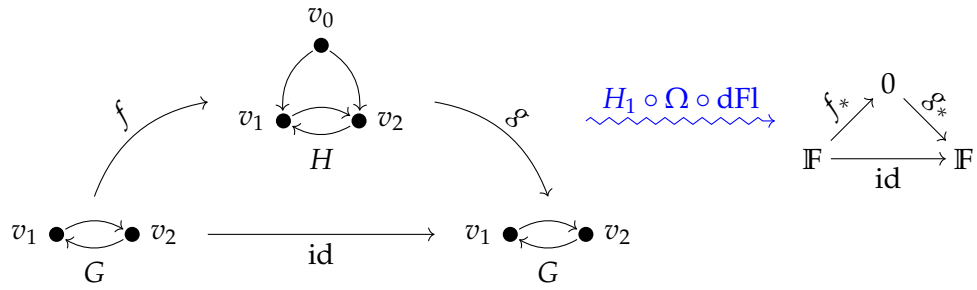


Figure 4.8: A commuting diagram in **Dgr** which illustrates why dFl cannot be made into a functor **Dgr** $\to$ **Ch**. For definitions of the digraph maps, see the proof of Proposition 4.5.28

Proposition 4.5.28. *The object map,* $\Omega \circ \mathrm{dFl} : \mathrm{Obj}(\mathbf{WkDgr}) \to \mathrm{Obj}(\mathbf{Ch})$*, which assigns to each digraph the chain complex associated to its directed flag complex, cannot be made into a functor* **WkDgr** $\to$ **Ch**.

Proof. Suppose, for contradiction, that such a functor exists. Consider the three digraphs illustrated in Figure 4.8. The weak digraph map f is given by the obvious inclusion, whilst g maps v_1 and v_2 to themselves and $v_0 \mapsto v_1$. Note that $g \circ f$ composes to the identity and hence the triangle of digraph maps commutes. Applying $H_1 \circ \Omega \circ \mathrm{dFl}$ to this diagram we find that $g_* \circ f_*$ is the identity on $H_1(\mathrm{dFl}(G))$. Using Lemma 4.2.25, we see that $\Omega(\mathrm{dFl}(G))$ is

$$0 \longrightarrow \mathbb{F}\langle v_1v_2, v_2v_1\rangle \xrightarrow{\partial_1} \mathbb{F}\langle v_1, v_2\rangle \tag{4.98}$$

and moreover $\partial_1(v_1v_2) = -\partial_1(v_2v_1) \neq 0$. Therefore, $H_1(\mathrm{dFl}(G)) \cong \mathbb{F}$. However, $H_1(\mathrm{dFl}(H))$ is the trivial vector space because H is star-like and hence dFl-contradctible. The identity on $H_1(\mathrm{dFl}(G))$ cannot factor through 0, so we obtain a contradiction. □

Definition 4.5.29. If there is a reciprocal edge $v_0 \leftrightarrows v_1$ then we say that w cones the reciprocal edge if

$$(w \to v_0 \text{ and } w \to v_1) \quad \text{or} \quad (v_0 \to w \text{ and } v_1 \to w). \tag{4.99}$$

Proposition 4.5.30. *Let G be a digraph, let $a \in V(G)$ and let the neighbouring vertices of a be denoted $b_0, \dots, b_n$. Then let H denote the subgraph of G on $V(G) \setminus \{a\}$. Suppose b_0 is such that, for all $i \in \{1, \dots, n\}$,*

$$a \to b_i \implies b_0 \to b_i; \tag{4.100}$$

$$a \leftarrow b_i \implies b_0 \leftarrow b_i. \tag{4.101}$$

Then there is a weak digraph map $r : G \to H$ given by $r\big|_H = \mathrm{id}_H$ and $r(a) = b_0$. Moreover, r is an $\mathcal{A}$-deformation retract. Furthermore, if there is not a reciprocal edge $a \leftrightarrows b_0$ and there does not exist $i \in \{1, \dots, n\}$ such that there is a reciprocal edge $b_0 \leftrightarrows b_i$ which is coned by a then r is a dFl*-deformation retract.*

Proof. For the proof that r is a weak digraph map and $\mathcal{A}$-deformation retract, see Example 3.15 of [55]. We prove the final statement, under the assumption that $a \to b_0$ (the $b_0 \to a$ case admits a similar proof). Note that the lack of a reciprocal edge means $b_0 \not\to a$.

We proceed by checking the conditions of Proposition 4.5.23. To begin, take arbitrary $x \in V(G)$ then either $x \neq a$, in which case $x = r(x)$, or $x = a$, in which case $x \to b_0 = r(x)$. Therefore, we see $x \rightrightarrows r(x)$. Next, take an edge $x \to y$. Either $y \neq a$, in which case $x \to y = r(y)$, or $y = a$ in which case $x = b_i$ for some i. We have seen that $b_0 \not\to a$, so we must have $i > 0$ and hence condition (4.101) implies that $x = b_i \to b_0 = r(y)$.

It remains to show that r is triangle-collapsing, so take a simplex $v_0v_1v_2 \in \mathrm{dFl}(G)$ and suppose that $r(v_0) = r(v_2)$ Since r is the identity on H we must have $\{v_0, v_2\} = \{a, b_0\}$ and $v_1 = b_i$ for some $i > 0$. Moreover, we know $b_0 \not\to a$ so $v_0v_1v_2$ must have the form ab_ib_0 for some $i > 0$. But then condition (4.100) implies $b_0 \to b_i$ and hence there is a reciprocal edge $b_i \leftrightarrows b_0$ which is coned by a, contradicting our assumption. □

Proposition 4.5.31. *Let G_n be a complete digraph on n nodes. Then G_n is $\mathcal{A}$-contractible for every $n \geq 1$ but G_n is* dFl*-contractible if and only if $n = 1$.*

Proof. The first statement follows immediately from Corollary 4.5.27 because G_n is star-like. For the second, it is clear that G_1 is contractible. Given $n > 1$, note that $\dim \Omega_n(\mathrm{dFl}(G_n)) = 0$ because there can be no n-simplices on a vertex set of size n. Therefore, it suffices to find a cycle, i.e. an element $c_{n-1} \in \Omega_{n-1}(\mathrm{dFl}(G_n))$ such that $\partial c_{n-1} = 0$. We construct these

inductively via a suspension-style argument, starting with the base case $n = 2$. We denote the nodes of G_n by $x_1, \ldots, x_n$. When $n = 2$, we can take $c_1 := x_0x_1 + x_1x_0$.

Now suppose $c_{n-1} \in \Omega_{n-1}(\mathrm{dFl}(G_n))$ is a cycle, such that $\partial c_{n-1} = 0$. Using Lemma 4.2.25, we can write c_{n-1} as a sum of $(n-1)$-simplices,

$$c_{n-1} = \sum_{v_0 \ldots v_{n-1} \in \mathrm{dFl}(G_n)} \alpha^{v_0 \ldots v_{n-1}} v_0 \ldots v_{n-1} \tag{4.102}$$

for some coefficients $\alpha^{v_0 \ldots v_{n-1}} \in \mathbb{F}$. Note that every $(n-1)$-simplex in $\mathrm{dFl}(G_n)$ is also an $(n-1)$-simplex of $\mathrm{dFl}(G_{n+1})$. Hence, suppressing notation for the inclusion $G_n \hookrightarrow G_{n+1}$, we can view c_{n-1} as an element of $\Omega_{n-1}(\mathrm{dFl}(G_{n+1}))$. Moreover, for every $(n-1)$-simplex $v_0 \ldots v_{n-1} \in \mathrm{dFl}(G_n)$, there are n-simplices $v_0 \ldots v_{n-1}x_{n+1}, x_{n+1}v_0 \ldots v_{n-1} \in \mathrm{dFl}(G_{n+1})$. We form two sums

$$c_{n,s} := \sum_{v_0 \ldots v_{n-1} \in \mathrm{dFl}(G_n)} \alpha^{v_0 \ldots v_{n-1}} (x_{n+1} v_0 \ldots v_{n-1}), \tag{4.103}$$

$$c_{n,t} := \sum_{v_0 \ldots v_{n-1} \in \mathrm{dFl}(G_n)} \alpha^{v_0 \ldots v_{n-1}} (v_0 \ldots v_{n-1} x_{n+1}), \tag{4.104}$$

such that $c_{n,s}, c_{n,t} \in \Omega_n(\mathrm{dFl}(G_{n+1}))$. A standard calculation, using the fact that $\partial c_{n-1} = 0$, shows that $\partial(c_{n,s}) = c_{n-1}$ and $\partial(c_{n,t}) = (-1)^n c_{n-1}$. Therefore, we can set $c_n := c_{n,s} + (-1)^{n+1} c_{n,t}$ to complete the inductive step. □

The complex $\mathrm{dFl}(G_n)$ is otherwise known as the complex of injective words on the alphabet $\{1, \ldots, n\}$. The author thanks Professor Richard Hepworth for this insight, as well as the following citations and proof via the Euler characteristic. Since $\mathrm{dFl}(G_n)$ is a simplicial set, it can be realised as a cell complex. It was first shown by Farmer that this cell complex has the homology of a wedge of $(!n)$-many $(n-1)$-spheres, where $!n$ is the number of derangements on n symbols [45]. A new proof of this was more recently given by Kerz [69, Theorem 1]. This equivalence on homology was strengthened by Björner and Wachs to a homotopy equivalence, by showing that the complex is shellable [11, Theorem 6.1].

Proposition 4.5.32 ([45]). *Let G_n denote a complete digraph on n nodes. If $n > 1$ then*

$$\dim H_k(\mathrm{dFl}(G_n)) = \begin{cases} 1 & \text{if } k = 0, \\ !n & \text{if } k = n-1, \\ 0 & \text{otherwise,} \end{cases} \tag{4.105}$$

where $!n$ is the number of derangements of $\{1, \ldots, n\}$.

We note that, for the statement above, it suffices to know that $\dim H_k(\mathrm{dFl}(G_n)) = 0$ for $0 < k < n-1$. Then, since G_n is connected, Proposition 4.2.26 gives $\dim H_0(\mathrm{dFl}(G_n)) = 1$. We can compute the Euler characteristic of $\Omega(\mathrm{dFl}(G_n))$ in two ways to obtain

$$1 + (-1)^{n-1} \dim H_{n-1}(\mathrm{dFl}(G_n)) = \sum_{k=0}^{n-1} (-1)^k \binom{n}{k+1} (k+1)! \tag{4.106}$$

which rearranges to give

$$\dim H_{n-1}(\mathrm{dFl}(G_n)) = \sum_{k=0}^{n} (-1)^k \binom{n}{k} (n-k)!. \tag{4.107}$$

The latter expression is a well-known expression for $!n$, as computed via the inclusion-exclusion principle.

To further elucidate the connection to derangements, denote the set of derangements by $D_n := \{\sigma = (v_1, \ldots, v_n) \in \mathrm{dFl}(G_n) \mid \sigma \text{ is a derangement}\}$ and note that $\mathrm{dFl}(G_n) \setminus D_n$ is still an ordered simplicial complex. Libman showed that the cell complex realization of this simplicial set is contractible [72, Theorem A], showing that the $(!n)$-many $(n-1)$-spheres can be 'parameterised' by the derangements. We refer the interested reader to [20] for further treatment of the topology of classes of directed flag complexes which are similarly generated by permutations.

Example 4.5.33. Finally, we revisit the digraph, G, in Figure 4.1 from Example 4.2.19. First, we can show G is $\mathcal{A}$-contractible. Using Proposition 4.5.30, we can successively retract nodes N and S onto the subgraph spanned by $\{x_0, x_1\}$. This is a complete digraph; hence G is $\mathcal{A}$-contractible. For the directed flag complex, the maximum non-trivial homology is in dimension 2 because there are no directed 4-cliques. One can then verify computationally (e.g. via Flagser [74]) that $H(\mathrm{dFl}(G))$ is the homology of the 2-sphere. Alternatively, viewing $\mathrm{dFl}(G)$ as a semi-simplicial set, one can recognise that $\mathrm{dFl}(G)$ coincides with the semi-simplicial set underlying a standard Δ-complex structure on S^2 (see [57, §2.1] for a definition).

4.6 Stability of persistent directed flag complex homology

In this section, we study the stability of persistent homology pipelines which use the directed flag complex. We do not study the stability of path homology, since it has already been well-treated [26, 32, 100].

Closely related to the directed flag complex is the ordered tuples complex, as defined by Turner [96]. This complex can also be viewed as a path complex and can be made into a

functor on **WkDgr**. This allows Turner to obtain stability theorems for persistent ordered tuples homology, with respect to the correspondence distance [96, Theorem 21]. In contrast, Turner showed that ordered set homology (i.e. directed flag complex homology) does not enjoy such a general stability result [96, §5.2].

4.6.1 The interleaving distance

Let $\mathbf{R}$ denote $\mathbb{R}$ equipped with the $\leq$ poset structure, viewed as a category. Given a category $\mathcal{C}$, a persistent $\mathcal{C}$-object is a functor $M : \mathbf{R} \to \mathcal{C}$; the morphisms $M(s \leq t)$ are called the structure maps of M. Given a persistent **TcDgr**-object, $M \in$ **[R, TcDgr]**, its persistent (directed flag complex) homology is $(H \circ \mathrm{dFl})(M)$. The resulting object is a persistent finite-dimensional graded vector space, commonly known as a persistence module. Any persistence module can be described by a complete, discrete invariant, called its barcode (for more information, see [28, 36]). The barcode can be used as a statistical descriptor of the initial persistent object M. For more background on persistent homology and its statistical analysis, we refer the reader to [27].

In order to be useful for applications, it is important that this construction is stable. That is, when $M, N \in$ **[R, TcDgr]** differ by some small perturbation to the input data, it is desirable to obtain bounds on

$$d((H \circ \mathrm{dFl})(M), (H \circ \mathrm{dFl})(N)), \tag{4.108}$$

for some appropriate choice of metric, d. We can use the structure of $\mathbf{R}$ to put an extended pseudometric on **[R, $\mathcal{C}$]**, the category of persistent $\mathcal{C}$-objects.

Definition 4.6.1. Fix a category $\mathcal{C}$ and $\delta \geq 0$ and two persistent objects $M, N : \mathbf{R} \to \mathcal{C}$.

1. The δ-shift of M, $M[\delta]$, is the persistent $\mathcal{C}$-object given by

$$M[\delta](t) := M(t + \delta) \quad \text{and} \quad M[\delta](s \leq t) := M(s + \delta \leq t + \delta). \tag{4.109}$$

2. The δ-transition morphism, $\mathcal{T}(M, \delta)$, is a natural transformation $M \Rightarrow M[\delta]$, given by $\mathcal{T}(M, \delta)(t) := M(t \leq t + \delta)$.
3. A δ-interleaving between M and N is a pair of natural transformations

$$f : M \Rightarrow N[\delta] \quad \text{and} \quad g : N \Rightarrow M[\delta], \tag{4.110}$$

such that $g[\delta] \circ f = \mathcal{T}(M, 2\delta)$ and $f[\delta] \circ g = \mathcal{T}(N, 2\delta)$. If such an interleaving exists, we say M and N are δ-interleaved (in $\mathcal{C}$).

4. Finally, the interleaving distance between M and N is

$$d_I(M,N) := \{\delta \geq 0 \mid M \text{ and } N \text{ are } \delta\text{-interleaved}\} \in [0,\infty]. \tag{4.111}$$

The interleaving distance is an extended pseudo-metric on persistent $\mathcal{C}$-objects [16, Theorem 3.3]. Moreover, the interleaving distance behaves well with respect to functors.

Lemma 4.6.2. *Any functor $F : \mathcal{C} \to \mathcal{D}$ induces a functor $F : [\mathbf{R}, \mathcal{C}] \to [\mathbf{R}, \mathcal{D}]$. Endowing these functor categories with the interleaving distance, F is a 1-Lipschitz map, i.e.*

$$d_I(F(M), F(N)) \leq d_I(M,N) \tag{4.112}$$

for all $M, N \in [\mathbf{R}, \mathcal{C}]$.

Proof. The functor is defined in the obvious way: given $M \in [\mathbf{R}, \mathcal{C}]$, we define $F(M)(t) := F(M(t))$ and $F(M)(s \leq t) := F(M(s \leq t))$. If f, g constitute a δ-interleaving between $M, N \in [\mathbf{R}, \mathcal{C}]$ then $F(f), F(g)$ constitute a δ-interleaving between $F(M)$ and $F(N)$. □

4.6.2 Stability theorems

Since $\mathcal{S}_{\text{dFl}}$ is a pull-back of $\mathcal{S}_{\mathbf{WkPathC}}$, it is a transitive system. Consequently, the composition operator for morphisms in **TcDgr** passes to the quotient to define a composition operator between $\simeq_{\text{dFl}}$ equivalence classes of triangle-collapsing digraph maps. This allows us to define a new category.

Definition 4.6.3. The naive dFl-homotopy category, $\mathbf{h}_{\text{dFl}}\mathbf{TcDgr}$ is the category whose objects are simple digraphs and morphisms are $\simeq_{\text{dFl}}$ equivalence classes of triangle-collapsing digraph maps. We use the notation $[\bullet]_{\text{dFl}}$ to denote objects and morphisms in this category.

Note that the objects of $\mathbf{h}_{\text{dFl}}\mathbf{TcDgr}$ are the same as **TcDgr**, but the morphisms are collapsed, according to the $\simeq_{\text{dFl}}$ equivalence relation. Corollary 4.4.3 automatically yields the following factorisation and hence a stability theorem.

Corollary 4.6.4. *The directed flag complex homology functor $H \circ \text{dFl} : \mathbf{TcDgr} \to \mathbf{grVec}$ factors through $\mathbf{h}_{\text{dFl}}\mathbf{TcDgr}$. Therefore, given $M, N \in [\mathbf{R}, \mathbf{TcDgr}]$,*

$$d_I((H \circ \text{dFl})(M), (H \circ \text{dFl})(N)) \leq d_I([M]_{\text{dFl}}, [N]_{\text{dFl}}) \leq d_I(M,N). \tag{4.113}$$

More explicitly, if there are natural transformations

$$f : M \Rightarrow N[\delta] \quad \textit{and} \quad g : N \Rightarrow M[\delta], \tag{4.114}$$

such that $g[\delta] \circ f \simeq_{\text{dFl}} \mathcal{T}(M, 2\delta)(t)$ and $f[\delta] \circ g \simeq_{\text{dFl}} \mathcal{T}(N, 2\delta)(t)$ for every $t \in \mathbf{R}$ then,

$$d_I\big((H \circ \text{dFl})(M), (H \circ \text{dFl})(N)\big) \leq \delta. \tag{4.115}$$

Essentially, this tells us that it suffices to δ-interleave M and N, up to $\simeq_{\mathrm{dFl}}$ equivalence. In general, this is a significantly easier task, especially when M and N are on different vertex sets. Note that we only require a homotopy at each $t \in \mathbf{R}$; the homotopies need not be natural, i.e. they need not commute with the structure maps of M and N. In the sequel, we construct explicit examples of such interleavings.

A filtration of digraphs on V is a family of digraphs $\{G_t\}_{t\in\mathbb{R}}$ such that $V(G_t) = V$ is constant and if $s \leq t$ then $E(G_s) \subseteq E(G_t)$. Connecting these graphs via the natural inclusions, we can view a filtration as a functor $G_\bullet : \mathbf{R} \to \mathbf{StDgr}$. Equivalently, we can view such a filtration as a function $\mathrm{ET}(G_\bullet) : V \times V \to \mathbb{R} \cup \{\pm\infty\}$, which records the entrance time of each edge:

$$\mathrm{ET}(G_\bullet)(v, w) := \inf\{t \in \mathbb{R} \mid (v, w) \in E(G_t)\}. \tag{4.116}$$

Proposition 4.6.5. *Let $G_\bullet, G'_\bullet$ be two filtrations of digraphs on a common vertex set V. Then*

$$d_I\big((H \circ \mathrm{dFl})(G_\bullet), (H \circ \mathrm{dFl})(G'_\bullet)\big) \leq \|\mathrm{ET}(G_\bullet) - \mathrm{ET}(G'_\bullet)\|_\infty. \tag{4.117}$$

Proof. Take arbitrary $\epsilon > 0$ and denote $\delta := \|\mathrm{ET}(G_\bullet) - \mathrm{ET}(G'_\bullet)\|_\infty$. Using the identity vertex map in both directions, it is straightforward to show that $G_\bullet$ and $G'_\bullet$ are $(\delta + \epsilon)$-interleaved. Lemma 4.6.2 applied to the functor $H \circ \mathrm{dFl}$ then implies $d_I\big((H \circ \mathrm{dFl})(G_\bullet), (H \circ \mathrm{dFl})(G'_\bullet)\big) \leq \delta + \epsilon$. Since ϵ was arbitrary, we can take $\epsilon \to 0$. □

Note that, in this proof, we construct an interleaving directly in **TcDgr**, rather than in the homotopy category $\mathbf{h}_{\mathrm{dFl}}\mathbf{TcDgr}$. In the following, we will see an example where we must use $\mathcal{S}_{\mathrm{dFl}}$.

Definition 4.6.6. A weighted digraph is a triple $G = (V, E, w)$ where (V, E) is a digraph and $w : E \to (0, \infty)$ is a positively-valued function, which we call the weighting. We use $w(G) := w$ to refer to the weighting.

Definition 4.6.7. Fix a weighted digraph G, with weighting $w := w(G)$.

1. A path p is a sequence of edges $e_0, \ldots, e_k$ such that, writing $e_i = (s_i, t_i)$, then $s_{i+1} = t_i$ for every i. We say p is a path from s_0 to t_k and write $p : s_0 \rightsquigarrow t_k$.
2. The length of such a path is the sum of its weights, $\ell(p) := \sum_{i=0}^{k} w(e_i)$.
3. The shortest-path filtration of G is a filtration of digraphs on $V(G)$, $\mathrm{SP}(G) := \{\mathrm{SP}(G)_t\}_{t\in\mathbb{R}}$, where

$$(i, j) \in E(\mathrm{SP}(G)_t) \iff \text{there is a path } p : i \rightsquigarrow j \text{ such that } \ell(p) \leq t. \tag{4.118}$$

4. We say G is directed acyclic, or a DAG, if whenever there is a (non-trivial) path $p : v_0 \rightsquigarrow v_1$ then there is not a path $v_1 \rightsquigarrow v_0$.

Remark 4.6.8. If G is a DAG then $\mathrm{SP}(G)_t$ is oriented for every $t \in \mathbb{R}$.

Given a weighted digraph, one can obtain a new weighted digraph by subdividing each of its edges.

Definition 4.6.9. Fix a weighted digraph G, with weighting $w := w(G)$.

1. A subdivision is a map from a subset of edges, $F \subseteq E(G)$, into the formal disjoint union of the interiors of the standard simplices, $S : F \to \sqcup_{d\in\mathbb{N}} \mathrm{int}(\Delta^d)$.
2. Given such a subdivision and an edge $e \in F$, let $d(e)$ denote the dimension such that $S(e) \in \mathrm{int}(\Delta^{d(e)})$. Further, denote the components of $S(e)$ by $S(e)_0, \ldots, S(e)_{d(e)}$.
3. The subdivision of G by S is a new weighted digraph $\mathbb{O}^s_S G := (V_S, E_S, w_S)$ given by

$$V_S := V(G) \sqcup \bigsqcup_{e\in F} \{v_{e,1}, \ldots, v_{e,d(e)}\}, \tag{4.119}$$

$$E_S := \big(E(G) \setminus F\big) \sqcup \bigsqcup_{e\in F} \{\tau_{e,0}, \ldots, \tau_{e,d(e)}\}, \tag{4.120}$$

$$w_S(\tau) := \begin{cases} w(\tau) & \text{if } \tau \in E(G), \\ S(e)_i \cdot w(e) & \text{if } \tau = \tau_{e,i}, \end{cases} \tag{4.121}$$

where $\tau_{e,i} = (v_{e,i}, v_{e,i+1})$, and we denote $v_{e,0} := \mathrm{st}(e)$ and $v_{e,d(e)+1} := \mathrm{fn}(e)$.

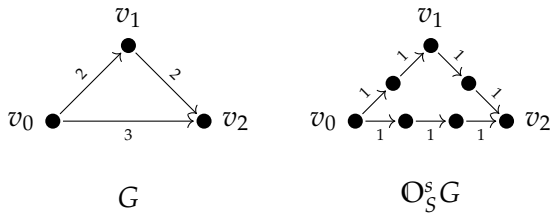


Figure 4.9: Two weighted digraphs, where weights are denoted by edge annotations. The second digraph, $\mathbb{O}^s_S G$, is the result of applying the subdivision S to the first digraph, G, where $S(v_0, v_2) = (1/3, 1/3, 1/3)$ and $S(v_0, v_1) = S(v_1, v_2) = (1/2, 1/2)$.

Proposition 4.6.10. *Let G be a weighted DAG and $S : F \to \sqcup_{d\in\mathbb{N}} \mathrm{int}(\Delta^d)$ a subdivision. Then*

$$d_I\big((H \circ \mathrm{dFl})(\mathrm{SP}(G)), (H \circ \mathrm{dFl})(\mathrm{SP}(\mathbb{O}^s_S G))\big) \leq \max_{e\in F} w(e). \tag{4.122}$$

Proof. Denote $\delta := \max_{e\in F} w(e)$. By Corollary 4.6.4, it suffices to show that $[\mathrm{SP}(G)]_{\mathrm{dFl}}$ and $[\mathrm{SP}(\mathbb{O}^s_S G)]_{\mathrm{dFl}}$ are δ-interleaved in $\mathbf{h}_{\mathrm{dFl}}\mathbf{TcDgr}$. We construct this interleaving via two vertex

maps, as was done in [26, Theorem 5.16]. Let $f : V(G) \to V(\mathbb{O}_S^s G)$ be given by the obvious inclusion. Let $g : V(\mathbb{O}_S^s G) \to V(G)$ be given by the formula

$$g(v) := \begin{cases} v & \text{if } v \in V(G), \\ \mathrm{st}(e) & \text{if } v = v_{e,i} \text{ and } \sum_{j<i} S(e)_j < 1/2, \\ \mathrm{fn}(e) & \text{if } v = v_{e,i} \text{ and } \sum_{j<i} S(e)_j \geq 1/2. \end{cases} \tag{4.123}$$

It is easy to see that f does not increase the lengths of paths and hence induces a digraph map $f : \mathrm{SP}(G)_t \to \mathrm{SP}(\mathbb{O}_S^s G)_t$ at every $t \in \mathbb{R}$. That is, f induces a natural transformation $\mathrm{SP}(G) \Rightarrow \mathrm{SP}(\mathbb{O}_S^s G)$. Since $\mathrm{SP}(\mathbb{O}_S^s G)$ is a filtration, f also induces a natural transformation $\mathrm{SP}(G) \Rightarrow \mathrm{SP}(\mathbb{O}_S^s G)[\delta]$. Note that since f is injective, it is certainly triangle-collapsing.

In the opposite direction, g can increase the length of paths. In particular, the start of the path may increase by $\delta/2$ and so too may the end of the path. Therefore, g induces a digraph map $g : \mathrm{SP}(\mathbb{O}_S^s G)_t \to \mathrm{SP}(G)_{t+\delta}$ at every $t \in \mathbb{R}$. That is, g induces a natural transformation $\mathrm{SP}(\mathbb{O}_S^s G) \Rightarrow \mathrm{SP}(G)[\delta]$. For more details, see the proof of [26, Theorem 5.16]. Next, the only time we observe $g(x) = g(y)$ with $x \neq y$ is if $x = v_{e,i}$ and $y = v_{e,k}$ for some $e \in F$ and $i < k$. Moreover, x and y must appear in the same half of a subdivided edge. If there is a triangle $xvy \in \mathrm{dFl}(\mathrm{SP}(\mathbb{O}_S^s G)_t)$ then v must appear between x and y in the subdivision of e (since G is directed acyclic). That is, we must have $v = v_{e,j}$ for some $i < j < k$, in which case $g(v) = g(x)$. Hence, g is triangle-collapsing.

It remains to show that these transformations constitute a δ-interleaving in $\mathbf{h}_{\mathrm{dFl}}\mathbf{TcDgr}$. Notice that, as vertex maps $g \circ f = \mathrm{id}_{V(G)}$ and hence $g \circ f = \mathcal{T}(\mathrm{SP}(G), 2\delta)$. In general, it is not the case that $f \circ g = \mathcal{T}(\mathrm{SP}(\mathbb{O}_S^s G), 2\delta)$. Instead, we will show $f \circ g \simeq_{\mathrm{dFl}} \mathcal{T}(\mathrm{SP}(\mathbb{O}_S^s G), 2\delta)$ at every $t \in \mathbb{R}$.

To prove this dFl-homotopy equivalence, we go via an interim digraph map $h : V(\mathbb{O}_S^s G) \to V(\mathbb{O}_S^s G)$, given by

$$h(v) := \begin{cases} v & \text{if } v \in V(G), \\ \mathrm{st}(e) & \text{if } v = v_{e,i} \text{ and } \sum_{j<i} S(e)_j < 1/2, \\ v & \text{if } v = v_{e,i} \text{ and } \sum_{j<i} S(e)_j \geq 1/2. \end{cases} \tag{4.124}$$

We first note that this is indeed a digraph map $h : \mathrm{SP}(\mathbb{O}_S^s G)_t \to \mathrm{SP}(\mathbb{O}_S^s G)_{t+2\delta}$ because h can only lengthen paths at their beginning, and by at most $\delta/2$. Moreover, h is triangle-collapsing at every $t \in \mathbb{R}$, by the same argument as g.

Claim 4.6.10.1. *If $x \rightrightarrows y$ in $\mathrm{SP}(\mathbb{O}_S^s G)_t$ then $h(x) \rightrightarrows y$ in $\mathrm{SP}(\mathbb{O}_S^s G)_{t+2\delta}$.*

Proof of Claim. Take a vertex $x \in V(\mathbb{O}_S^s G)$, then either $h(x) = x$ or $x = v_{e,i}$ for some $e \in F$ and $h(x) = \mathrm{st}(e)$, in which case there is a path of length at most $\delta/2$ in $\mathbb{O}_S^s G$ from $h(x)$ to

x. Therefore, $h(x) \to x$ in $\mathrm{SP}(\mathbb{O}_S^s G)_{t+2\delta}$ for every $t \in \mathbb{R}$. Suppose now $x \to y$ is an edge in $\mathrm{SP}(\mathbb{O}_S^s G)_t$, i.e. there is a path $x \rightsquigarrow y$ in $\mathbb{O}_S^s G$ of length at most t. Again, either $h(x) = x$ or there is a path $h(x) \rightsquigarrow x$ of length at most $\delta/2$. Therefore, there is a path $h(x) \rightsquigarrow y$ of length at most $t + \delta/2$. □ *of Claim.*

Claim 4.6.10.2. *There is no edge $x \to y$ in $\mathrm{SP}(\mathbb{O}_S^s G)_t$ such that $h(x) = y$.*

Proof of Claim. Suppose for contradiction there is an edge $x \to y$ is an edge in $\mathrm{SP}(\mathbb{O}_S^s G)_t$ such that $h(x) = y$. As we have noted before, there is always a path $h(x) \rightsquigarrow x$ in $\mathbb{O}_S^s G$, hence there is a composite path $h(x) \rightsquigarrow x \rightsquigarrow y = h(x)$. Since $x \neq y = h(x)$, this violates the assumption that G is directed acyclic. □ *of Claim.*

Hence, there is a one-step $\mathcal{S}_{\mathrm{dFl}}$-homotopy from h to $\mathrm{id}_{V(\mathbb{O}_S^s G)}$, viewed as digraph maps $\mathrm{SP}(\mathbb{O}_S^s G)_t \to \mathrm{SP}(\mathbb{O}_S^s G)_{t+2\delta}$.

Claim 4.6.10.3. *If $x \rightrightarrows y$ in $\mathrm{SP}(\mathbb{O}_S^s G)_t$ then $h(x) \rightrightarrows (f \circ g)(y)$ in $\mathrm{SP}(\mathbb{O}_S^s G)_{t+2\delta}$.*

Proof of Claim. Take a vertex $x \in V(\mathbb{O}_S^s G)$, then either $h(x) = (f \circ g)(x)$ or $x = v_{e,i}$ for some $e \in F$ and $h(x) = x$ and $(f \circ g)(x) = \mathrm{fn}(e)$, in which case there is a path of length at most $\delta/2$ in $\mathbb{O}_S^s G$ from $h(x)$ to $(f \circ g)(x)$. Therefore, $h(x) \to (f \circ g)(x)$ in $\mathrm{SP}(\mathbb{O}_S^s G)_{t+2\delta}$ for every $t \in \mathbb{R}$. Suppose now $x \to y$ is an edge in $\mathrm{SP}(\mathbb{O}_S^s G)_t$, i.e. there is a path $x \rightsquigarrow y$ in $\mathbb{O}_S^s G$ of length at most t. Again, either $h(x) = (f \circ g)(x)$ or there is a path $h(x) \rightsquigarrow (f \circ g)(x)$ of length at most $\delta/2$. Since $(f \circ g)$ increases path lengths by at most δ, there is a path $(f \circ g)(x) \rightsquigarrow (f \circ g)(y)$ of length at most $t + \delta$. Composing these paths yields a path $h(x) \rightsquigarrow (f \circ g)(y)$ of length at most $t + 3\delta/2$. □ *of Claim.*

Claim 4.6.10.4. *If there is an edge $x \to y$ in $\mathrm{SP}(\mathbb{O}_S^s G)_t$ such that $h(x) = (f \circ g)(y)$ then $h(x) = h(y) = (f \circ g)(x) = (f \circ g)(y)$.*

Proof of Claim. First note that for any node $v \in V(\mathbb{O}_S^s G)$, $(f \circ g)(v) = h(v)$ if and only if $v \in V(G)$ or v appears in the first half of its subdivided edge (i.e. $v = v_{e,i}$ for some $e \in F$ and $\sum_{j<i} S(e)_j \geq 1/2$)

As we saw in the previous claims, there must be a path $h(x) \rightsquigarrow x$ and hence there is a path $(f \circ g)(y) \rightsquigarrow x \rightsquigarrow y$. Since G is directed acyclic this means that there cannot be a path $y \rightsquigarrow (f \circ g)(y)$. Hence, y cannot be a new node in the second half of a subdivided edge. Therefore, $h(y) = (f \circ g)(y)$. Finally, note that $h(x) = (f \circ g)(y)$ is a node in $V(G)$, so x also cannot be a new node in the second half of a subdivided edge since these are mapped to $V(\mathbb{O}_S^s G) \setminus V(G)$ by h. Again, this implies $h(x) = (f \circ g)(x)$. □ *of Claim.*

Hence, viewed as digraph maps $\mathrm{SP}(\mathbb{O}_S^s G)_t \to \mathrm{SP}(\mathbb{O}_S^s G)_{t+2\delta}$, there is a one-step $\mathcal{S}_{\mathrm{dFl}}$-homotopy from h to $(f \circ g)$. □

4.6.3 Instabilities

Despite these stability guarantees, there are many seemingly innocuous edits one can make to a weighted digraph that drastically alter the persistent homology of its directed flag complex. As a first example, as a partial converse to Proposition 4.6.10, we exhibit a non-DAG which is not stable to subdivision.

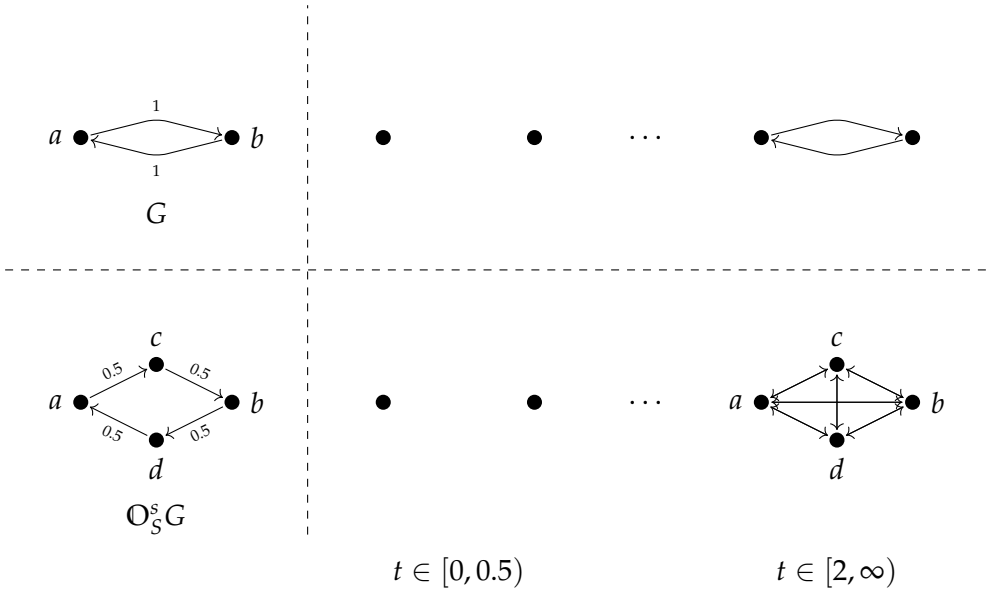


Figure 4.10: A weighted digraph which, upon subdividing, yields a persistence module at ∞ interleaving distance.

Proposition 4.6.11. *There exists a weighted digraph G and a subdivision $S : E(G) \to \sqcup_{d \in \mathbb{N}} \Delta^d$ such that*

$$d_I\big((H \circ \mathrm{dFl})(\mathrm{SP}(G)), (H \circ \mathrm{dFl})(\mathrm{SP}(\mathbb{O}_S^s G))\big) = \infty. \tag{4.125}$$

Proof. Consider the two weighted digraphs illustrated in Figure 4.10. The second weighted digraph is obtained from the first via the subdivision $S(a,b) = S(b,a) = (1/2, 1/2)$. The first module, $(H \circ \mathrm{dFl})(\mathrm{SP}(G))$, has non-trivial homology in degree 1 for $t \in [1, \infty)$. For proof of this, recall the computations done in the proof of Proposition 4.5.28. In contrast, $\mathrm{SP}(\mathbb{O}_S^s G)_t$ is the complete digraph on 4 nodes, for all $t \in [2, \infty)$. Via a choice of basis for $\ker \partial_1$ or through explicit computation (i.e. Flagser [74]), one can verify that this digraph has trivial homology in degree 1. Rank constraints imply that there is no ϵ-interleaving between these modules for any $\epsilon \geq 0$. □

An arguably worse stability is that adding a single edge can also lead to a change in the persistent homology that is unbounded.

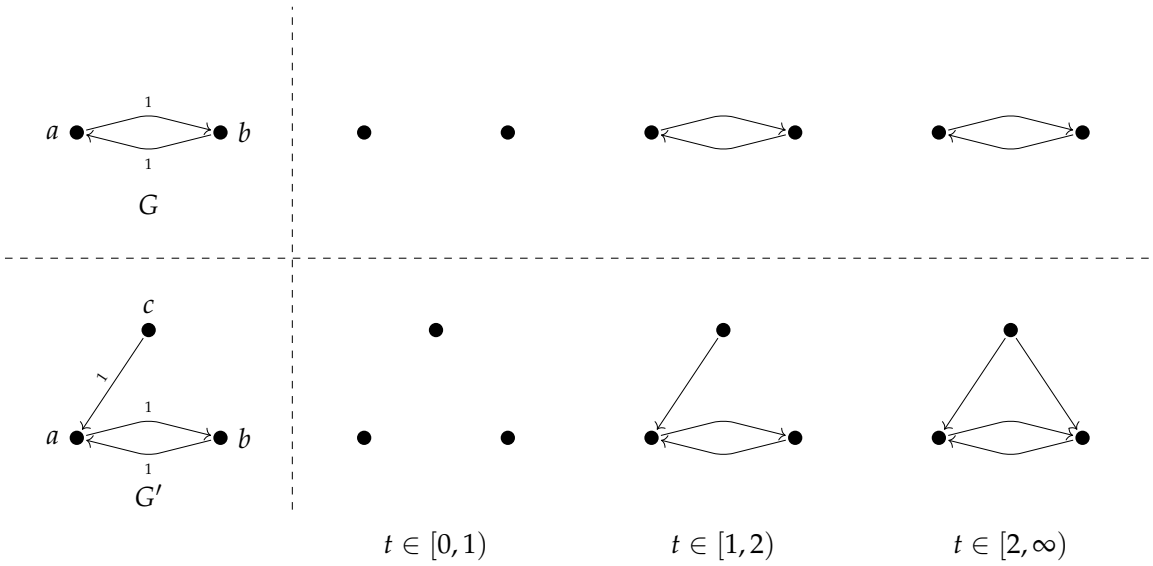


Figure 4.11: A weighted digraph which, upon adding a single appendage, yields a persistence module at ∞ interleaving distance.

Proposition 4.6.12. *There exists a weighted digraph G such that adding an appendage edge yields a weighed digraph G' such that*

$$d_I\big((H \circ \mathrm{dFl})(\mathrm{SP}(G)), (H \circ \mathrm{dFl})(\mathrm{SP}(G'))\big) = \infty. \tag{4.126}$$

Proof. Consider the two weighted digraphs illustrated in Figure 4.11. The first module, $(H \circ \mathrm{dFl})(\mathrm{SP}(G))$, has non-trivial homology in degree 1 for $t \in [1, \infty)$. In contrast, the second module, $(H \circ \mathrm{dFl})(\mathrm{SP}(G'))$, has trivial homology in degree 1 for $t \in [2, \infty)$. For proof of this, recall the computations done in the proof of Proposition 4.5.28. Rank constraints imply that there is no ϵ-interleaving between these modules for any $\epsilon \geq 0$. □

In contrast, the module $(H \circ \mathcal{A})(\mathrm{SP}(G))$ is invariant to adding appendages, because a 0-interleaving at the homotopy level can be constructed via Proposition 4.5.30.

4.A Mapping cylinders

Definition 4.A.1. Given a weak path morphism $f : P_1 \to P_2$, the mapping cylinder of f, $\mathrm{MCyl}(f)$, is a path complex on $(V(P_1) \times \{0\}) \cup (V(P_2) \times \{1\})$ containing paths of the following form

1. $(x_0, 0) \dots (x_k, 0)$ such that $x_0 \dots x_k \in P_1$;
2. $(y_0, 1) \dots (y_k, 1)$ such that $y_0 \dots y_k \in P_2$;
3. $(x_0, 0) \dots (x_i, 0)(f(x_i), 1) \dots (f(x_k), 1)$ such that $x_0 \dots x_k \in P_1$ and $0 \leq i \leq k$.

Remark 4.A.2. Let $f : P_1 \to P_2$ be a weak path morphism of regular path complexes. The mapping cylinder, $\mathrm{MCyl}(f)$ is itself a regular path complex if and only if f is also a strong path morphism.

Note that for any path complex P, $\mathrm{MCyl}(\mathrm{id}_P) = \mathrm{Cyl}(P)$. We verify that mapping cylinders satisfy a universal property, analogous to the usual one in the category of topological spaces.

Proposition 4.A.3. *Given a weak path morphism $f : P_1 \to P_2$, let $\iota_1 : P_1 \to \mathrm{Cyl}(P_1)$ denote the natural inclusion $x \mapsto (x, 1)$. Then $\mathrm{MCyl}(f)$ is the pushout of f and ι_1 in* **WkPathC**.

Proof. First, define two vertex maps $\phi_1 : P_2 \to \mathrm{MCyl}(f)$ and $\phi_2 : \mathrm{Cyl}(P_1) \to \mathrm{MCyl}(f)$ by

$$\phi_1(y) := (y, 1) \quad \text{and} \quad \phi_2(x, i) := \begin{cases} (x, 0) & \text{if } i = 0, \\ (f(x), 1) & \text{if } i = 1. \end{cases} \tag{4.127}$$

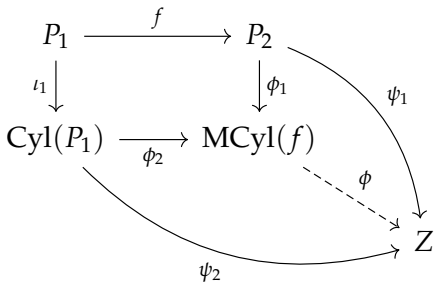
These maps clearly describe weak path morphisms which make the following diagram commute

$$\begin{array}{ccc} P_1 & \xrightarrow{f} & P_2 \\ {\scriptstyle \iota_1}\downarrow & & \downarrow{\scriptstyle \phi_1} \\ \mathrm{Cyl}(P_1) & \xrightarrow[\phi_2]{} & \mathrm{MCyl}(f) \end{array} \tag{4.128}$$

Given another commuting diagram

$$\begin{array}{ccc} P_1 & \xrightarrow{f} & P_2 \\ {\scriptstyle \iota_1}\downarrow & & \downarrow{\scriptstyle \psi_1} \\ \mathrm{Cyl}(P_1) & \xrightarrow[\psi_2]{} & Z \end{array} \tag{4.129}$$

in **WkDgr**, define $\phi : \mathrm{MCyl}(f) \to Z$ by $\phi(x, 0) := \psi_2(x, 0)$ and $\phi(y, 1) := \psi_1(y)$. As vertex maps, ϕ is clearly the unique map that makes the following diagram commute



(4.130)

It remains to verify that ϕ is a weak path morphism. Given a path $p \in \mathrm{MCyl}(f)$, we split into three cases:

Case 1: Suppose $p = (x_0, 0) \dots (x_k, 0)$ where $x_0 \dots x_k \in P_1$. Let p' denote the same path but viewed as an element of $\mathrm{Cyl}(P_1)$ so that $\phi_2(p') = p$. Then, $\phi(p) = \psi_2(p')$ must be either irregular or a path in Z since ψ_2 is a weak path morphism.

Case 2: Suppose $p = (y_1, 1) \dots (y_k, 1)$ where $p' := y_0 \dots y_k \in P_2$. Then $p = \phi_1(p')$ so $\phi(p) = \psi_1(p')$ is either irregular or a path in Z since ψ_1 is a weak path morphism.

Case 3: Suppose $p = (x_0, 0) \dots (x_i, 0)(f(x_i), 1) \dots (f(x_k), 1)$ where $x_0 \dots x_k \in P_1$ and $0 \leq i \leq n$. Then let $p' := (x_0, 0) \dots (x_i, 0)(x_i, 1) \dots (x_k, 1) \in \mathrm{Cyl}(P_1)$, and observe

$$\begin{aligned} \phi(p) &= \psi_2(x_0, 0) \dots \psi_2(x_i, 0)\psi_1(f(x_i)) \dots \psi_1(f(x_k)) && (4.131)\\ &= \psi_2(x_0, 0) \dots \psi_2(x_i, 0)\psi_2(x_i, 1) \dots \psi_2(x_k, 1) && (4.132)\\ &= \psi_2(p') && (4.133) \end{aligned}$$

where the second equality holds because diagram (4.129) commutes. Now, since ψ_2 is a weak path morphism, we see that $\phi(p)$ must be irregular or a path in Z. □

Proposition 4.A.4. *Given a weak path morphism $f : P_1 \to P_2$, $\mathrm{MCyl}(f) \simeq P_2$.*

Proof. Let $\gamma : P_2 \to \mathrm{MCyl}(f)$ denote the natural inclusion $y \mapsto (y, 1)$ which is certainly a weak path morphism. Let $\rho : \mathrm{MCyl}(f) \to P_2$ be given by $(x, 0) \mapsto f(x)$ and $(y, 1) \mapsto y$. Again this is a weak path morphism because paths of the form $(x_0, 0) \dots (x_i, 0)(f(x_i), 1) \dots (f(x_k), 1) \in \mathrm{MCyl}(f)$ have an irregular image. Note that $\rho \circ \gamma = \mathrm{id}_{P_2}$. It remains to show that $\gamma \circ \rho \simeq \mathrm{id}_{\mathrm{MCyl}(f)}$; we construct the homotopy explicitly.

We shall denote vertices of $\mathrm{Cyl}(\mathrm{MCyl}(f))$ by (v, α, β) where α is the index for the inner mapping cylinder and β is the index for outer cylinder. Let $F : \mathrm{Cyl}(\mathrm{MCyl}(f)) \to \mathrm{MCyl}(f)$ be given by

$$F(v, \alpha, 0) := (v, \alpha) \quad \text{and} \quad F(v, \alpha, 1) := \begin{cases} (f(v), 1) & \text{if } \alpha = 0, \\ (v, 1) & \text{if } \alpha = 1. \end{cases} \tag{4.134}$$

There are inclusion $\iota_i : \mathrm{MCyl}(f) \hookrightarrow \mathrm{Cyl}(\mathrm{MCyl}(f))$, for $i = 0, 1$, given by $\iota_i(v, \alpha) = (v, \alpha, i)$. Note that $F \circ \iota_0 = \mathrm{id}_{\mathrm{MCyl}(f)}$ and $F \circ \iota_1 = \gamma \circ \rho$. It remains to show that F is a weak path morphism.

Take an arbitrary path $p \in \mathrm{MCyl}(f)$, and denote the vertices $p = (x_0, \alpha_0) \dots (x_k, \alpha_k)$. Now consider the path

$$p_j := (x_0, \alpha_0, 0) \dots (x_j, \alpha_j, 0)(x_j, \alpha_j, 1) \dots (x_k, \alpha_k, 1) \in \mathrm{Cyl}(\mathrm{MCyl}(f)) \tag{4.135}$$

for some $j \in [0,k]$. Note that the α_i are either constant at 0 or 1, or there is some index m such that $\alpha_i = 0$ for $i \leq m$ and $\alpha_i = 1$ for $i > m$. We split into cases.

Case 1: Suppose $\alpha_i = 0$ for every i. In this case, note that $x_0 \dots x_k \in P_1$. Then,

$$F(p_j) = (x_0, 0) \dots (x_j, 0)(f(x_j), 1) \dots (f(x_k), 1) \tag{4.136}$$

which is a path in $\mathrm{MCyl}(f)$.

Case 2: Suppose $\alpha_i = 1$ for every i. In this case, $x_0 \dots x_k \in P_2$ and

$$F(p_j) = (x_0, 1) \dots (x_j, 1)(x_j, 1) \dots (x_k, 1) \tag{4.137}$$

which is clearly irregular.

Case 3: Suppose $\alpha_i = 0$ for $i \leq m$ and $\alpha_i = 1$ for $i > m$ where $m < j$. In this case note that $\alpha_j = 1$ and hence

$$F(x_j, \alpha_j, 0) = (x_j, 1) = F(x_j, \alpha_j, 1). \tag{4.138}$$

Therefore, $F(p_j)$ must be irregular.

Case 4: Suppose $\alpha_i = 0$ for $i \leq m$ and $\alpha_i = 1$ for $i > m$ such that $m \geq j$. In this case, note that p_j contains adjacent vertices $(x_m, 0, 1)(x_{m+1}, 1, 1)$. Moreover, since p is a path in $\mathrm{MCyl}(f)$ we must have $x_{m+1} = f(x_m)$. Mapping through F we see

$$F(x_m, 0, 1) = (f(x_m), 1) = (x_{m+1}, 1) = F(x_{m+1}, 1, 1) \tag{4.139}$$

and thus $F(p_j)$ is irregular.

All other paths $q \in \mathrm{Cyl}(\mathrm{MCyl}(f))$ are of the form $\iota_i(q')$ for some $i \in \{0,1\}$ and $q' \in \mathrm{MCyl}(f)$. If $i = 0$ then $F(q) = (F \circ \iota_0)(q') = q'$; if $i = 1$ then $F(q) = (F \circ \iota_1)(q') = (\gamma \circ \rho)(q')$. Since γ and ρ are weak path morphisms, we see either $F(q) \in \mathrm{MCyl}(f)$ or $F(q)$ is irregular. □

A similar result holds for triangle-collapsing simplicial morphisms. It is possible to show this either by adapting the proof above, or by appealing to the connection with simplicial sets.

Proposition 4.A.5. *Given a triangle-collapsing simplicial morphism $f : K_1 \to K_2$, $\overline{\mathrm{MCyl}(f)} \simeq K_2$.*

Proof. By Corollary 4.4.21, the system of one-step homotopies for **TcOSC** is a pull-back of the standard system for simplicial sets, $\mathcal{S}_{\mathbf{TcOSC}} = \sigma^* \mathcal{S}_{\mathbf{sSet}}$. Using Proposition 4.A.3, we can also show $\overline{\mathrm{MCyl}(f)}$ is a pushout in **TcOSC** and hence $\sigma(\overline{\mathrm{MCyl}(f)})$ is the usual mapping cylinder for $\sigma(f)$. The result then follows from the homotopy equivalence $\sigma(\overline{\mathrm{MCyl}(f)}) \simeq \sigma(K_2)$ in simplicial sets [91, Proposition 2.68]. □

4.B Grounded pipelines

4.B.1 Background

This work was initially motivated by the stability analysis of grounded pipelines for weighted digraphs (Chapter 3 in this thesis). To recall, these pipelines require two components:

1. a filtration map, $F : \mathrm{WDgr} \to \mathrm{Obj}(\mathbf{[R, Dgr]})$, which assigns a filtration of digraphs to each weighted digraph; and
2. a digraph chain complex, i.e. a functor $C : \mathbf{InclDgr} \to \mathbf{Ch}$, where **InclDgr** is the wide subcategory of **WkDgr** restricted to morphisms which are inclusions.

Given these complexes, one obtains the following commutative diagram for any weighted digraph G, in which we use $F^t G$ as shorthand for $F(G)(t)$:

$$\begin{array}{ccccccccc} \cdots C_3(F^tG) & \xrightarrow{\partial_3} & C_2(F^tG) & \xrightarrow{\iota_\# \circ \partial_2} & C_1(G \cup F^tG) & \xrightarrow{\partial_1} & C_0(G \cup F^tG) \cdots \\ & & \searrow^{\partial_2} & & \uparrow_{\iota_\#} & & \uparrow_{\iota_\#} \\ & & & C_1(F^tG) & \xrightarrow[\partial_1]{} & C_0(F^tG) & \end{array} \tag{4.140}$$

The top row of this diagram is a chain complex, which we denote ${}^gC_F(G)(t)$. By Lemma 3.3.10, the functoriality of C ensures that this can be made into a persistent chain complex ${}^gC_F(G) \in \mathbf{[R, Ch]}$. The persistent homology of this can then be used as a topological summary of G.

The initial study (Chapter 3 in this thesis) considered this pipeline with $F = \mathrm{SP}$ and $C = \Omega \circ \mathcal{A}$. **In this section, we take $F = \mathrm{SP}$ and $C = \Omega \circ \mathrm{dFl}$ and study the resulting pipeline.** The assignment $G \mapsto {}^gC_F(G)$ can be made functorial, but we must restrict to morphisms between weighted digraphs that induce triangle-collapsing morphisms $\mathrm{SP}(G)_t \to \mathrm{SP}(H)_t$.

Definition 4.B.1. Given a weighted digraph G, the shortest-path quasimetric is a quasimetric on $V(G)$ given for $i, j \in V(G)$ by

$$d_G(i,j) := \inf\{t \geq 0 \mid \text{there is a path } p : i \rightsquigarrow j \text{ with } \ell(p) \leq t\}. \tag{4.141}$$

Given two weighted digraphs G, H, a vertex map $f : V(G) \to V(H)$ is called a contraction if

$$d_H(f(i), f(j)) \leq d_G(i,j) \tag{4.142}$$

for all $i, j \in V(G)$.

Note that if f is a contraction then it induces a weak digraph map $\mathrm{SP}(G)_t \to \mathrm{SP}(H)_t$ for every $t \in \mathbb{R}$.

Definition 4.B.2. A digraph map $f : G \to H$ is path-collapsing if whenever there are paths $i \rightsquigarrow j \rightsquigarrow k$ and $f(i) = f(k)$ then $f(i) = f(j) = f(k)$.

Note that any path-collapsing digraph map is a triangle-collapsing map. Moreover, if $f : G \to H$ is path-collapsing and a contraction then it induces triangle-collapsing digraphs maps $\mathrm{SP}(G)_t \to \mathrm{SP}(H)_t$ for every $t \in \mathbb{R}$. With these observations, we can deduce the functoriality of this pipeline, via Theorem 3.3.13.

Definition 4.B.3. Let **ContPcWDgr** denote the category of all simple, weighted digraphs where morphisms are vertex maps which are both contractions and path-collapsing digraph maps.

Lemma 4.B.4. *The assignment* $G \mapsto {}^{g}C_F(G)$ *can be made into a functor* **ContPcWDgr** $\to$ ***[*R*, Ch]****.*

Definition 4.B.5. Grounded persistent directed flag homology is the functor ${}^{g}\widehat{\mathcal{H}}_1$: **ContPcWDgr** $\to$ **[R, Vec]** given by ${}^{g}\widehat{\mathcal{H}}_1 := H_1 \circ {}^{g}C_F$.

In the rest of this section we study the stability of ${}^{g}\widehat{\mathcal{H}}_1$. Namely, we wish to understand what alterations can be made to a weighted digraph G without dramatically changing ${}^{g}\widehat{\mathcal{H}}_1(G)$, as measured by the interleaving distance.

4.B.2 Stability theorems

We aim to state a general stability theorem for ${}^{g}\widehat{\mathcal{H}}_1$, along the lines of Theorem 3.5.8. First, we require relative version of the homotopies in $\mathcal{S}_{\mathrm{dFl}}$ such that the induced chain homotopy is trivial on the fixed component. We keep the presentation fairly close to Chapter 3 in order to emphasise the differences.

Definition 4.B.6. Given two triangle-collapsing digraph maps $f, g : G \to H$ and a subset $A \subseteq V(G)$, if $f \simeq_{\mathrm{dFl},1} g$ and moreover $f(v) = g(v)$ for every $v \in A$, then we say that f and g are one-step $\mathcal{S}_{\mathrm{dFl}}$-homotopic relative A. This determines a system of one-step homotopies, which we denote $\mathcal{S}_{\mathrm{dFl},A}$. We denote the resulting equivalence relation by $f \simeq_{\mathrm{dFl}} g$ (rel A) and say that f and g are $\mathcal{S}_{\mathrm{dFl}}$-homotopic relative A.

By essentially the same proof as Lemma 3.2.37, we obtain the following behaviour of the induced chain homotopy between maps that are $\mathcal{S}_{\mathrm{dFl}}$-homotopic relative A.

Lemma 4.B.7. *Suppose $f, g : G \to H$ are triangle-collapsing digraph maps that are $\mathcal{S}_{\mathrm{dFl}}$-homotopic relative A, for some $A \subseteq V(G)$. Then,* $\mathrm{dFl}(f)$ *and* $\mathrm{dFl}(G)$ *are $\mathcal{S}_{\mathbf{WkRPC}}$-homotopic so Theorem 4.4.2 induces a chain homotopy between $\Omega(\mathrm{dFl}(f))$ and $\Omega(\mathrm{dFl}(g))$; denote its components by*

$$L_k : \Omega_k(\mathrm{dFl}(G)) \to \Omega_{k+1}(\mathrm{dFl}(H)). \tag{4.143}$$

Let G_A denote the induced subgraph of G on the vertices in A. Then $L_k(\Omega_k(\mathrm{dFl}(G_A))) = 0$.

We now introduce δ-shifting vertex maps which will be our main method for constructing interleavings.

Definition 4.B.8 (Previously Definition 3.5.1)**.** Given two weighted digraphs G, H, a δ-shifting vertex map is a vertex map $f : V(G) \to V(H)$ such that f induces weak digraph maps $\mathrm{SP}(G)_t \to \mathrm{SP}(H)_{t+\delta}$ and $G \cup \mathrm{SP}(G)_t \to H \cup \mathrm{SP}(H)_{t+\delta}$, for every $t \in \mathbb{R}$.

Definition 4.B.9 (c.f. Definition 3.5.5)**.** Let G, H be two weighted digraphs and let $f : V(G) \to V(H)$ and $g : V(H) \to V(G)$ be two vertex maps between them,

1. Construct the following sets:

$$V_{\mathrm{fix}}(g, f) := \{v \in V(G) \mid (g \circ f)(v) = v\}, \tag{4.144}$$
$$E_{\mathrm{fix}}(g, f) := \{e \in E(G) \mid (g \circ f)(e) = e\}, \tag{4.145}$$
$$V_{\mathrm{diff}}(g, f) := \{v \in V(G) \mid (g \circ f)(v) \neq v\}, \tag{4.146}$$
$$E_{\mathrm{diff}}(g, f) := \{e \in E(G) \mid (g \circ f)(e) \neq e\}. \tag{4.147}$$

   Moreover, denote $G_{\mathrm{diff}}(g, f) := (V(G), E_{\mathrm{diff}}(g, f))$.
2. If both $\mathrm{id} : G_{\mathrm{diff}}(g, f) \to \mathrm{SP}(G)_{2\delta}$ and $g \circ f : G_{\mathrm{diff}}(g, f) \to \mathrm{SP}(G)_{2\delta}$ are triangle-collapsing digraph maps and furthermore they are $\mathcal{S}_{\mathrm{dFl}}$-homotopic relative $V_{\mathrm{fix}}(g, f)$ then we say that (g, f) has grounded dFl-codistortion $\leq \delta$.
3. If f and g are both path-collapsing, δ-shifting vertex maps and the pairs (g, f) and (f, g) both have grounded dFl-codistortion $\leq \delta$ then we say they form a δ-grounded dFl-interleaving.

With these definitions and a proof essentially identical to that of Theorem 3.5.8, we obtain our main stability theorem for ${}^{g}\widehat{\mathcal{H}}_1$.

Theorem 4.B.10. *Given two weighted digraphs G, H, if there is a δ-grounded* dFl*-interleaving between them then*

$$d_I({}^{g}\widehat{\mathcal{H}}_1(G), {}^{g}\widehat{\mathcal{H}}_1(H)) \leq \delta. \tag{4.148}$$

In Chapter 3, the analogous theorem was used to study the stability of the pipeline to various perturbations of weighted digraphs. With this new theory, the results of that work go through for ${}^{g}\widehat{\mathcal{H}}_1$ as soon as one has made the necessary constraints to ensure that the δ-shifting vertex maps are path-collapsing and that the one-step homotopies belong to $\mathcal{S}_{\mathrm{dFl},V_{\mathrm{fix}}(g,f)}$.

To this end, note that if G is a digraph and H is a DAG then any weak digraph map $G \to H$ is necessarily path-collapsing. Therefore, denoting the category of weighted DAGs with contractions for morphisms by **ContWDag**, we have an inclusion of categories **ContWDag** $\subseteq$ **ContPcWDgr**. As such, most of the stability theorems obtained in Chapter 3 apply automatically once we restrict to weighted DAGs. To summarise these results, we present Table 4.1, which is a reproduction of Table 3.1 but with additional annotations to denote when the result applies unrestricted to ${}^{g}\widehat{\mathcal{H}}_1$, and when the result must first be restricted to DAGs.

Operation	**Locally Stable**	**Non-locally Stable**	**Locally Unstable**	**Isomorphism**
Weight perturbation	Thm 3.5.11$^{\checkmark}$			
Edge subdivision	Thm 3.5.16$^{\restriction}$			
Edge collapse	Thm 3.5.25$^{\blacklozenge\restriction}$		Thm 3.5.28$^{\checkmark}$	
Edge deletion	Cor. 3.5.34$^{\blacklozenge\checkmark}$	Thm 3.5.31$^{\checkmark}$	Thm 3.5.35$^{\checkmark}$	Thm 3.5.38$^{\blacklozenge\restriction}$
Vertex deletion		Thm 3.5.43$^{\restriction}$	Cor. 3.5.42$^{\checkmark}$	Cor. 3.5.41$^{\blacklozenge\checkmark}$

Table 4.1: Stability and instability theorems for ${}^{g}\widehat{\mathcal{H}}_1$, under various digraph operations. $\blacklozenge$ Denotes a theorem which only applies to a subset of such operations. $\checkmark$ Denotes a theorem which applies to the directed flag complex unrestricted. $\restriction$ Denotes a theorem which applies to the directed flag complex after restricting to weighted digraphs G, such that G and $\mathbb{O}_\theta^T G$ are both DAGs.

Bibliography

[1] Atte Aalto, Lauri Viitasaari, Pauliina Ilmonen, Laurent Mombaerts, and Jorge Gonçalves. "Gene regulatory network inference from sparsely sampled noisy data". In: *Nature Communications* 11.1 (2020), p. 3493. ISSN: 2041-1723. DOI: 10.1038/s41467-020-17217-1.

[2] Noga Alon and Joel H. Spencer. *The probabilistic method*. Fourth. Wiley series in discrete mathematics and optimization. Hoboken, New Jersey: Wiley, 2016. ISBN: 9781119061953.

[3] Yasuhiko Asao. "Magnitude homology and path homology". In: *Bulletin of the London Mathematical Society* 55.1 (2023), pp. 375–398. DOI: 10.1112/blms.12734.

[4] Eric Babson, Hélène Barcelo, Mark de Longueville, and Reinhard Laubenbacher. "Homotopy theory of graphs". In: *Journal of Algebraic Combinatorics* 24.1 (Aug. 2006), pp. 31–44. ISSN: 1572-9192. DOI: 10.1007/s10801-006-9100-0.

[5] Hélène Barcelo, Xenia Kramer, Reinhard Laubenbacher, and Christopher Weaver. "Foundations of a Connectivity Theory for Simplicial Complexes". In: *Advances in Applied Mathematics* 26.2 (2001), pp. 97–128. ISSN: 0196-8858. DOI: 10.1006/aama.2000.0710.

[6] Ulrich Bauer. "Ripser: efficient computation of Vietoris–Rips persistence barcodes". In: *Journal of Applied and Computational Topology* 5.3 (Sept. 2021), pp. 391–423. ISSN: 2367-1734. DOI: 10.1007/s41468-021-00071-5.

[7] Ulrich Bauer, Michael Kerber, Jan Reininghaus, and Hubert Wagner. "Phat – Persistent Homology Algorithms Toolbox". In: *Journal of Symbolic Computation* 78 (2017). Algorithms and Software for Computational Topology, pp. 76–90. ISSN: 0747-7171. DOI: 10.1016/j.jsc.2016.03.008.

[8] Ulrich Bauer and Michael Lesnick. "Induced matchings and the algebraic stability of persistence barcodes". In: *Journal of Computational Geometry* 6.2 (2015), pp. 162–191. DOI: 10.20382/jocg.v6i2a9.

[9] Katherine Benjamin, Lamisah Mukta, Gabriel Moryoussef, Christopher Uren, Heather A Harrington, Ulrike Tillmann, and Agnese Barbensi. "Homology of homologous knotted proteins". In: *Journal of the Royal Society Interface* 20.201 (2023), p. 20220727. DOI: 10.1098/rsif.2022.0727.

[10] Alyson Bittner, Brittany Terese Fasy, Maia Grudzien, Sayonita Ghosh Hajra, Jici Huang, Kristine Pelatt, Courtney Thatcher, Altansuren Tumurbaatar, and Carola Wenk. "Comparing directed and weighted road maps". In: *Research in Computational Topology*. Cham: Springer, 2018, pp. 57–70. DOI: 10.1007/978-3-319-89593-2_4.

[11] Anders Björner and Michelle Wachs. "On lexicographically shellable posets". In: *Transactions of the American Mathematical Society* 277.1 (1983), pp. 323–341. DOI: 10.1090/S0002-9947-1983-0690055-6.

[12] Romain Boulet, Etienne Fieux, and Bertrand Jouve. "s-homotopy for finite graphs". In: *Electronic Notes in Discrete Mathematics* 31 (2008). The International Conference on Topological and Geometric Graph Theory, pp. 123–127. ISSN: 1571-0653. DOI: 10.1016/j.endm.2008.06.025.

[13] Romain Boulet, Etienne Fieux, and Bertrand Jouve. "Simplicial simple-homotopy of flag complexes in terms of graphs". In: *European Journal of Combinatorics* 31.1 (Jan. 2010), pp. 161–176. ISSN: 0195-6698. DOI: 10.1016/j.ejc.2009.05.003.

[14] Emma L. Brown, Thierry L. Lefebvre, Paul W. Sweeney, Bernadette J. Stolz, Janek Gröhl, Lina Hacker, Ziqiang Huang, Dominique-Laurent Couturier, Heather A. Harrington, Helen M. Byrne, and Sarah E. Bohndiek. "Quantification of vascular networks in photoacoustic mesoscopy". In: *Photoacoustics* 26 (2022), p. 100357. ISSN: 2213-5979. DOI: 10.1016/j.pacs.2022.100357.

[15] Peter Bubenik and Nikola Milićević. "Homotopy, homology, and persistent homology using closure spaces". In: *Journal of Applied and Computational Topology* 8.3 (Sept. 2024), pp. 579–641. ISSN: 2367-1734. DOI: 10.1007/s41468-024-00183-8.

[16] Peter Bubenik and Jonathan A. Scott. "Categorification of Persistent Homology". In: *Discrete & Computational Geometry* 51.3 (2014), pp. 600–627. ISSN: 1432-0444. DOI: 10.1007/s00454-014-9573-x.

[17] Luigi Caputi, Anna Pidnebesna, and Jaroslav Hlinka. "Promises and pitfalls of topological data analysis for brain connectivity analysis". In: *NeuroImage* 238 (2021), p. 118245. ISSN: 1053-8119. DOI: 10.1016/j.neuroimage.2021.118245.

[18] Luigi Caputi and Henri Riihimäki. "Hochschild homology, and a persistent approach via connectivity digraphs". In: *Journal of Applied and Computational Topology* (Mar. 2023). ISSN: 2367-1734. DOI: 10.1007/s41468-023-00118-9.

[19] Daniel Carranza, Brandon Doherty, Chris Kapulkin, Morgan Opie, Maru Sarazola, and Liang Ze Wong. *Cofibration category of digraphs for path homology*. 2022. arXiv: 2212.12568 [math.CO].

[20] Wojtek Chacholski, Ran Levi, and Roy Meshulam. "On the topology of complexes of injective words". In: *Journal of Applied and Computational Topology* 4.1 (Mar. 2020), pp. 29–44. ISSN: 2367-1734. DOI: 10.1007/s41468-019-00039-6.

[21] Thomas Chaplin. *First Betti number of the path homology of random directed graphs - Code and Data Repository*. Version 3612203. Aug. 12, 2021. URL: https://github.com/tomchaplin/phrg-code.

[22] Thomas Chaplin. *First Betti number of the path homology of random directed graphs - Code and Data Repository*. Oct. 2022. DOI: 10.17605/OSF.IO/ZVUMB. URL: https://osf.io/zvumb.

[23] Thomas Chaplin. *GrPPHATI*. Sept. 15, 2023. URL: https://github.com/tomchaplin/grpphati.

[24] Thomas Chaplin. "First Betti number of the path homology of random directed graphs". In: *Journal of Applied and Computational Topology* 8.6 (Oct. 2024), pp. 1503–1549. ISSN: 2367-1734. DOI: 10.1007/s41468-022-00108-3.

[25] Thomas Chaplin, Heather A. Harrington, and Ulrike Tillmann. *A notion of homotopy for directed graphs and their flag complexes*. 2024. arXiv: 2411.04572 [math.AT].

[26] Thomas Chaplin, Heather A. Harrington, and Ulrike Tillmann. "Grounded Persistent Path Homology: A Stable, Topological Descriptor for Weighted Digraphs". In: *Foundations of Computational Mathematics* (Aug. 2024). ISSN: 1615-3383. DOI: 10.1007/s10208-024-09679-2.

[27] Frédéric Chazal and Bertrand Michel. "An Introduction to Topological Data Analysis: Fundamental and Practical Aspects for Data Scientists". In: *Frontiers in Artificial Intelligence* 4 (2021). ISSN: 2624-8212. DOI: 10.3389/frai.2021.667963.

[28] Frédéric Chazal, Vin de Silva, Marc Glisse, and Steve Oudot. *The Structure and Stability of Persistence Modules*. Cham: Springer International Publishing, 2016. ISBN: 978-3-319-42545-0. DOI: 10.1007/978-3-319-42545-0_2.

[29] Mauricio Che, Fernando Galaz-García, Luis Guijarro, Ingrid Membrillo Solis, and Motiejus Valiunas. "Basic metric geometry of the bottleneck distance". In: *Proceedings of the American Mathematical Society* 152 (2024). DOI: 10.1090/proc/16776.

[30] Louis H. Y. Chen, Larry Goldstein, and Qi-Man Shao. "Fundamentals of Stein's Method". In: *Normal Approximation by Stein's Method*. Berlin, Heidelberg: Springer Berlin Heidelberg, 2011, pp. 13–44. ISBN: 978-3-642-15007-4. DOI: 10.1007/978-3-642-15007-4_2.

[31] Samir Chowdhury and Facundo Mémoli. "A functorial Dowker theorem and persistent homology of asymmetric networks". In: *Journal of Applied and Computational Topology* 2.1 (Oct. 2018), pp. 115–175. ISSN: 2367-1734. DOI: 10.1007/s41468-018-0020-6.

[32] Samir Chowdhury and Facundo Mémoli. "Persistent Path Homology of Directed Networks". In: *Proceedings of the 2018 Annual ACM-SIAM Symposium on Discrete Algorithms (SODA)*, pp. 1152–1169. DOI: 10.1137/1.9781611975031.75.

[33] David Cohen-Steiner, Herbert Edelsbrunner, and John Harer. "Stability of persistence diagrams". In: *Proceedings of the Twenty-First Annual Symposium on Computational Geometry*. SCG '05. Pisa, Italy: Association for Computing Machinery, 2005, pp. 263–271. ISBN: 1581139918. DOI: 10.1145/1064092.1064133.

[34] David Cohen-Steiner, Herbert Edelsbrunner, John Harer, and Yuriy Mileyko. "Lipschitz Functions Have Lp-Stable Persistence". In: *Foundations of Computational Mathematics* 10.2 (Apr. 2010), pp. 127–139. ISSN: 1615-3383. DOI: 10.1007/s10208-010-9060-6.

[35] David Cohen-Steiner, Herbert Edelsbrunner, John Harer, and Dmitriy Morozov. "Persistent homology for kernels, images, and cokernels". In: *Proceedings of the twentieth annual ACM-SIAM symposium on Discrete algorithms*. SIAM. 2009, pp. 1011–1020. DOI: 10.1137/1.9781611973068.110.

[36] William Crawley-Boevey. "Decomposition of pointwise finite-dimensional persistence modules". In: *Journal of Algebra and Its Applications* 14.05 (2015), p. 1550066. DOI: 10.1142/S0219498815500668.

[37] Sebastiano Cultrera di Montesano, Ondřej Draganov, Herbert Edelsbrunner, and Morteza Saghafian. *Chromatic Alpha Complexes*. Dec. 2022. arXiv: 2212.03128 [math.AT].

[38] Tamal K. Dey, Tianqi Li, and Yusu Wang. "An Efficient Algorithm for 1-Dimensional (Persistent) Path Homology". In: *36th International Symposium on Computational Geometry (SoCG 2020)*. Ed. by Sergio Cabello and Danny Z. Chen. Vol. 164. Leibniz International Proceedings in Informatics (LIPIcs). Dagstuhl, Germany: Schloss Dagstuhl–Leibniz-Zentrum für Informatik, 2020, 36:1–36:15. ISBN: 978-3-95977-143-6. DOI: 10.4230/LIPIcs.SoCG.2020.36.

[39] Anton Dochtermann and Takahiro Matsushita. *Simple homotopy of flag simplicial complexes and contractible contractions of graphs*. 2023. arXiv: 2312.10976 [math.CO].

[40] Barbara Giunti, Jānis Lazovskis, and Bastian Rieck. *DONUT: Database of Original & Non-Theoretical Uses of Topology*. https://donut.topology.rocks. 2022.

[41] Meyer Dwass. "Modified Randomization Tests for Nonparametric Hypotheses". In: *The Annals of Mathematical Statistics* 28.1 (1957), pp. 181–187. DOI: 10.1214/aoms/1177707045.

[42] Herbert Edelsbrunner, John Harer, et al. "Persistent homology-a survey". In: *Surveys on Discrete and Computational Geometry: Twenty Years Later*. Ed. by Jacob E. Goodman, János Pach, and Richard Pollack. Vol. 453. Contemporary mathematics. Providence, RI: American Mathematical Society, 2008, pp. 257–282. ISBN: 978-0-8218-4239-3. DOI: 10.1090/conm/453.

[43] Paul Erdős and Alfréd Rényi. "On the evolution of random graphs". In: *Publ. Math. Inst. Hung. Acad. Sci* 5.1 (1960), pp. 17–60.

[44] Lisbeth Fajstrup, Eric Goubault, Emmanuel Haucourt, Samuel Mimram, and Martin Raussen. *Directed algebraic topology and concurrency*. Springer, 2016. ISBN: 978-3-319-15397-1. DOI: 10.1007/978-3-319-15398-8.

[45] Frank D Farmer. "Cellular homology for posets". In: *Math. Japon* 23.6 (1979), pp. 607–613.

[46] Xinbo Gao, Bing Xiao, Dacheng Tao, and Xuelong Li. "A survey of graph edit distance". In: *Pattern Analysis and Applications* 13.1 (2010), pp. 113–129. ISSN: 1433-755X. DOI: 10.1007/s10044-008-0141-y.

[47] Barbara Giunti. *TDA-Applications*. URL: https://www.zotero.org/groups/2425412/tda-applications (visited on 09/07/2022).

[48] Paul G. Goerss and John F. Jardine. "Classical results and constructions". In: *Simplicial Homotopy Theory*. Basel: Birkhäuser Basel, 2009, pp. 139–194. ISBN: 978-3-0346-0189-4. DOI: 10.1007/978-3-0346-0189-4_3.

[49] Paul G. Goerss and John F. Jardine. "Simplicial sets". In: *Simplicial Homotopy Theory*. Basel: Birkhäuser Basel, 2009, pp. 1–64. ISBN: 978-3-0346-0189-4. DOI: 10.1007/978-3-0346-0189-4_1.

[50] Christian Goodbrake, David Beers, Travis B. Thompson, Heather A. Harrington, and Alain Goriely. "Brain chains as topological signatures for Alzheimer's disease". In: *Journal of Applied and Computational Topology* (Apr. 2024). ISSN: 2367-1734. DOI: 10.1007/s41468-024-00168-7.

[51] Dejan Govc, Ran Levi, and Jason P. Smith. "Complexes of tournaments, directionality filtrations and persistent homology". In: *Journal of Applied and Computational Topology* 5.2 (2021), pp. 313–337. ISSN: 2367-1734. DOI: 10.1007/s41468-021-00068-0.

[52] A. A. Grigor'yan, Yong Lin, Yu. V. Muranov, and Shing-Tung Yau. "Path Complexes and their Homologies". In: *Journal of Mathematical Sciences* 248.5 (Aug. 2020), pp. 564–599. ISSN: 1573-8795. DOI: 10.1007/s10958-020-04897-9.

[53] Alexander Grigor'yan, Rolando Jimenez, Yuri Muranov, and Shing-Tung Yau. "Homology of path complexes and hypergraphs". In: *Topology and its Applications* 267 (2019), p. 106877. ISSN: 0166-8641. DOI: 10.1016/j.topol.2019.106877.

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

[55] Alexander Grigor'yan, Yong Lin, Yuri Muranov, and Shing-Tung Yau. "Homotopy theory for digraphs". In: *Pure and Applied Mathematics Quarterly* 10.4 (Mar. 4, 2015), pp. 619–674. DOI: 10.4310/PAMQ.2014.v10.n4.a2.

[56] Alexander Grigor'yan, Rolando Jimenez, Yuri Muranov, and Shing-Tung Yau. "On the path homology theory of digraphs and Eilenberg–Steenrod axioms". In: *Homology, Homotopy and Applications* 20.2 (2018), pp. 179–205. DOI: 10.4310/HHA.2018.v20.n2.a9.

[57] Allen Hatcher. *Algebraic topology*. Cambridge, 2002. ISBN: 9780521791601.

[58] Alec Helm, Ann S. Blevins, and Danielle S. Bassett. *The growing topology of the C. elegans connectome*. bioRxiv. 2021. DOI: 10.1101/2020.12.31.424985.

[59] R Hepworth and S Willerton. "Categorifying the magnitude of a graph". In: *Homology, Homotopy and Applications* 19.2 (2017), pp. 31–60. DOI: 10.4310/HHA.2017.v19.n2.a3.

[60] Richard Hepworth and Emily Roff. *Bigraded path homology and the magnitude-path spectral sequence*. 2024. arXiv: 2404.06689 [math.AT].

[61] Piers J Ingram, Michael PH Stumpf, and Jaroslav Stark. "Network motifs: structure does not determine function". In: *BMC genomics* 7.1 (2006), pp. 1–12. DOI: 10.1186/1471-2164-7-108.

[62] Mohd Sabri Ismail, Mohd Salmi Md Noorani, Munira Ismail, Fatimah Abdul Razak, and Mohd Almie Alias. "Early warning signals of financial crises using persistent homology". In: *Physica A: Statistical Mechanics and its Applications* 586 (2022), p. 126459. ISSN: 0378-4371. DOI: 10.1016/j.physa.2021.126459.

[63] Sergei O. Ivanov and Fedor Pavutnitskiy. "Simplicial approach to path homology of quivers, marked categories, groups and algebras". In: *Journal of the London Mathematical Society* 109.1 (2024), e12812. DOI: 10.1112/jlms.12812.

[64] Matthew Kahle. "Topology of random clique complexes". In: *Discrete Mathematics* 309.6 (2009), pp. 1658–1671. ISSN: 0012-365X. DOI: 10.1016/j.disc.2008.02.037.

[65] Matthew Kahle. "Sharp vanishing thresholds for cohomology of random flag complexes". In: *Annals of Mathematics* (2014), pp. 1085–1107. DOI: 10.4007/annals.2014.179.3.5.

[66] Matthew Kahle. "Topology of random simplicial complexes: a survey". In: *Algebraic Topology: Applications and New Directions*. Ed. by Ulrike Tillmann, Søren Galatius, and Dev Sinha. Vol. 620. Contemporary Mathematics. Amer. Mathematical Society, 2014, pp. 201–222. DOI: 10.1090/conm/620/12367.

[67] Matthew Kahle and Elizabeth Meckes. "Limit theorems for Betti numbers of random simplicial complexes". In: *Homology, Homotopy and Applications* 15.1 (2013), pp. 343–374. DOI: 10.4310/HHA.2013.v15.n1.a17.

[68] K Heiner Kamps and Timothy Porter. *Abstract homotopy and simple homotopy theory*. World Scientific, 1997. ISBN: 981-02-1602-5.

[69] Moritz C Kerz. "The complex of words and Nakaoka stability". In: *Homology, Homotopy and Applications* 7.1 (2005), pp. 77–85. DOI: 10.4310/HHA.2005.v7.n1.a5.

[70] Jure Leskovec and Andrej Krevl. *SNAP Datasets: Stanford Large Network Dataset Collection*. http://snap.stanford.edu/data. June 2014.

[71] Lu Li, Connor Thompson, Gregory Henselman-Petrusek, Chad Giusti, and Lori Ziegelmeier. "Minimal Cycle Representatives in Persistent Homology using Linear Programming: an Empirical Study with User's Guide". In: *Frontiers in artificial intelligence* 4 (2021), p. 73. DOI: 10.3389/frai.2021.681117.

[72] Assaf Libman. *The complex of injective words of permutations which are not derangements is contractible*. 2023. arXiv: 2309.06808 [math.CO].

[73] Yong Lin, Shiquan Ren, Chong Wang, and Jie Wu. "Weighted Path homology of Weighted Digraphs and Persistence". In: (Oct. 2019). arXiv: 1910.09891 [math.AT].

[74] Daniel Lütgehetmann, Dejan Govc, Jason P. Smith, and Ran Levi. "Computing Persistent Homology of Directed Flag Complexes". In: *Algorithms* 13.1 (2020). ISSN: 1999-4893. DOI: 10.3390/a13010019.

[75] Paolo Masulli and Alessandro E. P. Villa. "The topology of the directed clique complex as a network invariant". In: *SpringerPlus* 5.1 (2016), p. 388. ISSN: 2193-1801. DOI: 10.1186/s40064-016-2022-y.

[76] John D. Medaglia. "Graph Theoretic Analysis of Resting State Functional MR Imaging." In: *Neuroimaging clinics of North America* 27 (4 Nov. 2017), pp. 593–607. DOI: 10.1016/j.nic.2017.06.008.

[77] Nikola Milićević and Nicholas A. Scoville. *The directed Vietoris-Rips complex and homotopy and singular homology groups of finite digraphs*. 2024. arXiv: 2409.01370 [math.AT].

[78] Dmitriy Morozov and Arnur Nigmetov. "Towards lockfree persistent homology". In: *Proceedings of the 32nd ACM Symposium on Parallelism in Algorithms and Architectures*. 2020, pp. 555–557. DOI: 10.1145/3350755.3400244.

[79] David E Muñoz, Elizabeth Munch, Brittany Terese Fasy, and Firas A Khasawneh. "The Walk-Length Filtration for Persistent Homology on Weighted Directed Graphs". In: *Computational Geometry: Young Researchers Forum 2024 – book of abstracts*. 2024, pp. 33–36.

[80] Vidit Nanda. *Computational Algebraic Topology - Lecture Notes*. 2021. URL: https://people.maths.ox.ac.uk/nanda/cat/TDANotes.pdf (visited on 07/11/2022).

[81] John T. Nardini, Bernadette J. Stolz, Kevin B. Flores, Heather A. Harrington, and Helen M. Byrne. "Topological data analysis distinguishes parameter regimes in the Anderson-Chaplain model of angiogenesis". In: *PLOS Computational Biology* 17.6 (June 2021), pp. 1–29. DOI: 10.1371/journal.pcbi.1009094.

[82] Abhinav Natarajan, Thomas Chaplin, Adam Brown, and Maria-Jose Jimenez. *Morse Theory for Chromatic Delaunay Triangulations*. May 2024. arXiv: 2405.19303 [math.AT].

[83] Marcia Oliveira and Joao Gama. "An overview of social network analysis". In: *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery* 2.2 (2012), pp. 99–115. DOI: 10.1002/widm.1048.

[84] Metin Öner and İpek Deveci Kocakoç. "JMASM 49: A compilation of some popular goodness of fit tests for normal distribution: Their algorithms and MATLAB codes (MATLAB)". In: *Journal of Modern Applied Statistical Methods* 16.2 (2017), p. 30. DOI: 10.22237/jmasm/1509496200.

[85] Nina Otter, Mason A. Porter, Ulrike Tillmann, Peter Grindrod, and Heather A. Harrington. "A roadmap for the computation of persistent homology." In: *EPJ data science* 6 (1 2017), p. 17. DOI: 10.1140/epjds/s13688-017-0109-5.

[86] Dale Purves, George J Augustine, David Fitzpatrick, WC Hall, AS LaMantia, JO McNamara, and L White, eds. *Neuroscience*. International sixth edition. New York: Sinauer Associates, 2018. ISBN: 9781605358703.

[87] Michael W Reimann, Max Nolte, Martina Scolamiero, Katharine Turner, Rodrigo Perin, Giuseppe Chindemi, Paweł D łotko, Ran Levi, Kathryn Hess, and Henry Markram. "Cliques of neurons bound into cavities provide a missing link between structure and function". In: *Frontiers in computational neuroscience* 11 (2017), p. 48. DOI: 10.3389/fncom.2017.00048.

[88] Antonio Rieser. "Čech closure spaces: A unified framework for discrete and continuous homotopy". In: *Topology and its Applications* 296 (2021), p. 107613. ISSN: 0166-8641. DOI: 10.1016/j.topol.2021.107613.

[89] Antonio Rieser. *Vietoris-Rips Homology Theory for Semi-Uniform Spaces*. 2021. arXiv: 2008.05739 [math.AT].

[90] Colin Patrick Rourke and Brian Joseph Sanderson. "Δ-sets I: Homotopy theory". In: *The Quarterly Journal of Mathematics* 22.3 (Sept. 1971), pp. 321–338. ISSN: 0033-5606. DOI: 10.1093/qmath/22.3.321.

[91] Christian Rüschoff. *Lecture notes: simplicial sets*. 2017. URL: https://www.mathi.uni-heidelberg.de/~rueschoff/ss17sset/sset.pdf (visited on 10/01/2024).

[92] Ann E. Sizemore, Chad Giusti, Ari Kahn, Jean M. Vettel, Richard F. Betzel, and Danielle S. Bassett. "Cliques and cavities in the human connectome". In: *Journal of Computational Neuroscience* 44.1 (2018), pp. 115–145. ISSN: 1573-6873. DOI: 10.1007/s10827-017-0672-6.

[93] Bernadette J. Stolz, Jakob Kaeppler, Bostjan Markelc, Franziska Braun, Florian Lipsmeier, Ruth J. Muschel, Helen M. Byrne, and Heather A. Harrington. "Multiscale topology characterizes dynamic tumor vascular networks". In: *Science Advances* 8.23 (2022), eabm2456. DOI: 10.1126/sciadv.abm2456.

[94] Paul W. Sweeney, Angela d'Esposito, Simon Walker-Samuel, and Rebecca J. Shipley. "Modelling the transport of fluid through heterogeneous, whole tumours in silico". In: *PLOS Computational Biology* 15.6 (June 2019), e1006751. DOI: 10.1371/journal.pcbi.1006751.

[95] Jonathan Treviño-Marroquín. *Graphs and Their Vietoris-Rips Complexes Have the Same Pseudotopological Weak Homotopy Type*. 2024. arXiv: 2406.00149 [math.AT].

[96] Katharine Turner. "Rips filtrations for quasimetric spaces and asymmetric functions with stability results". In: *Algebraic & Geometric Topology* 19.3 (2019), pp. 1135–1170. DOI: 10.2140/agt.2019.19.1135.

[97] Yu Wu, Gabriel Shindnes, Vaibhav Karve, Derrek Yager, Daniel B Work, Arnab Chakraborty, and Richard B Sowers. "Congestion barcodes: Exploring the topology of urban congestion using persistent homology". In: *2017 IEEE 20th International Conference on Intelligent Transportation Systems (ITSC)*. IEEE. 2017, pp. 1–6. DOI: 10.1109/ITSC.2017.8317777.

[98] Jiacheng Xing, Jiaying Jia, Xin Wu, and Liqun Kuang. "A Spatiotemporal Brain Network Analysis of Alzheimer's Disease Based on Persistent Homology." In: *Frontiers in aging neuroscience* 14 (2022), p. 788571. DOI: 10.3389/fnagi.2022.788571.

[99] M. Yutin. *Performant Path Homology*. Version 89eb473. URL: https://github.com/SteveHuntsmanBAESystems/PerformantPathHomology (visited on 07/20/2020).

[100] Shen Zhang. *Stability of Persistent Path Diagrams*. 2024. arXiv: 2406.11998 [math.AT].