

Metric and Topological Approaches to Network Data Analysis

Dissertation

**Presented in Partial Fulfillment of the Requirements for the Degree Doctor
of Philosophy in the Graduate School of The Ohio State University**

By

Samir Chowdhury, B.S., M.S.

Graduate Program in Mathematics

The Ohio State University

2019

Dissertation Committee:

Facundo Mémoli, Advisor

Matthew Kahle

David Sivakoff

© Copyright by
Samir Chowdhury
2019

Abstract

Network data, which shows the relationships between entities in complex systems, is becoming available at an ever-increasing rate. In particular, advances in data acquisition and computational power have shifted the bottleneck in analyzing weighted and directed network datasets towards the domain of available mathematical methods. Thus there is a pressing need to develop mathematical foundations for analyzing such datasets.

In this thesis, we present methods for applying one of the flagship tools of topological data analysis—*persistent homology*—to weighted, directed network datasets. We ground these methods in a network distance $d_{\mathcal{N}}$ that had appeared in a restricted context in earlier literature, and is now fully developed in this thesis. This development independently provides *metric* methods for network data analysis, including and invoking methods from optimal transport.

In our framework, a network dataset is represented as a set of points X (equipped with the minimalistic structure of a first countable topological space) and a (continuous) edge weight function $\omega_X : X \times X \rightarrow \mathbb{R}$. With this terminology, a finite network dataset is viewed as a finite sample from some infinite underlying process—a compact network. This perspective is especially appropriate for data streams that are so large that they are “essentially infinite”, or are perhaps being generated continuously in time.

We show that the space of all compact networks is the completion of the space of all finite networks. We develop the notion of isomorphism in this space, and explore a range of different geodesics that exist in this space. We develop sampling theorems and explain their use in obtaining probabilistic convergence guarantees. Several persistent homology methods—notably including persistent path homology—are also developed. By virtue of the sampling theorems, we are able to define these methods even for infinite networks.

Our theoretical contributions are complemented by software packages that we developed in the course of producing this thesis. We illustrate the theory and implementations via experiments on simulated and real-world data.

In my family, I found boundless sources of love and encouragement. My grandparents, aunts, and uncles provided me with all the support I could ask for. My mother gave me insight from her own life as an academic, my father taught me the quadratic formula and my first lessons in the sciences, and my brother lifted the weight of responsibility from my shoulders so I could pursue my own path. It is to them that I dedicate this work.

Acknowledgments

My thesis advisor, Facundo Mémoli, has invested incredible amounts of time and energy into my education, and I can only hope to pay it forward. Even during the final stages of writing this thesis, I was amazed at how his vision had materialized as connections between all the research that we had done together, and how neatly the different concepts rolled out and interlocked, like the pieces of a puzzle. I am happy and grateful that I was able to work under his guidance for the past five years.

I am lucky to have had many mentors over the years; special thanks go to Henry Adams, Jose Perea, Chad Giusti, Matthew Kahle, and David Sivakoff, who have all contributed crucially to the development of my career. I would like to thank Neil Falkner, whose instruction was critical for my education, and also Boris Hasselblatt, Richard Weiss, and Genevieve Walsh, who first led me into mathematics.

Many academics took the time to give me advice and direction over the years. From these conversations, I have often carried away a new insight into a complex problem, or a deeper realization about the academic world at large. Javier Arsuaga, Pablo Cámara, Chao Chen, Paweł Dłotko, Moon Duchin, Greg Henselman, Kathryn Hess, Steve Huntsman, Sara Kališnik, Katherine Kinnaird, Sanjeevi Krishnan, Melissa McGuirl, Amit Patel, Xaq Pitkow, Vanessa Robins, Manish Saggar, Santiago Segarra, Elchanan Solomon, Justin Solomon, Mimi Tsuruga, Mariel Vázquez, Bei Wang, Sunny Xiao, Lori Ziegelmeier—thank you all, you make this community a wonderful place to be in.

Our research group—and more broadly, the TGDA group at OSU—comprised a wonderful community of grad students and postdocs who contributed to my way of thinking about problems. Tom Needham and Ben Schweinhart were both exceptional in that regard. Tamal Dey gave the seminar talk that first got me interested in applied topology, and Anastasios Sidiropoulos got me thinking about directed graphs.

My friends in Columbus made sure that my time here was rich and rewarding. Thanks to the 91 crew for the endless laughs, to Sunita and Sabrina for many therapeutic conversations, and to the Rahmans for making me part of their family. Thanks to Tia for contributing to my career, in so many ways known and unknown. Danke schön to Natalie for her warmth, wit, and wisdom. My fellow grad students were a constant source of support and entertainment; I was lucky to have Marissa as a study buddy, Katie for keeping my harebrained ideas about the TAGGS seminar in check, and Evan and Kevin for many welcome distractions. Osama, Hanbaek, and I were there for each other through all the highs and lows. I could not ask for it to be any other way.

Vita

2013-present	PhD student, Mathematics, The Ohio State University
2016	M.S. Mathematics, The Ohio State University
2013	B.S. Engineering Science, Mathematics, Tufts University

Publications

Research Publications

S. Chowdhury, B. Dai, F. Mémoli. “The importance of forgetting: Limiting memory improves recovery of topological characteristics from neural data”. *PloS ONE*, 2018.

S. Chowdhury, F. Mémoli. “A functorial Dowker theorem and persistent homology of asymmetric networks”. *Journal of Applied and Computational Topology*, 2018.

S. Chowdhury, F. Mémoli. “Persistent path homology of directed networks”. *Proceedings of the Twenty-Ninth Annual ACM-SIAM Symposium on Discrete Algorithms*, 2018.

S. Chowdhury, F. Mémoli. “Explicit geodesics in Gromov-Hausdorff space ”. *Electronic Research Announcements in Mathematical Sciences*, 2018.

S. Chowdhury, F. Mémoli, Z. Smith. “Improved error bounds for tree representations of metric spaces”. *Advances in Neural Information Processing Systems*, 2016.

Z. Smith, S. Chowdhury, F. Mémoli. “Hierarchical representations of network data with optimal distortion bounds”. *50th Asilomar Conference on Signals, Systems and Computers*, 2016.

S. Chowdhury, F. Mémoli. “Persistent homology of directed networks”. *50th Asilomar Conference on Signals, Systems and Computers*, 2016.

S. Chowdhury, F. Mémoli. “Distances between directed networks and applications”. *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2016.

S. Chowdhury, F. Mémoli. “Metric structures on networks and applications”. *53rd Annual Allerton Conference on Communication, Control, and Computing (Allerton)*, 2015.

Fields of Study

Major Field: Mathematics

Table of Contents

	Page
Abstract	ii
Dedication	iii
Acknowledgments	iv
Vita	v
List of Tables	x
List of Figures	xii
1. Introduction	1
1.1 Organization of this thesis	3
1.2 Networks and $d_{\mathcal{N}}$: Definition, reformulations, and first results	4
1.3 Network models: the cycle networks and the SBM networks	10
1.3.1 The directed circles	12
1.3.2 The finite cycle networks	15
1.3.3 The network stochastic block model	16
1.4 Persistent homology on networks: Simplicial constructions	17
1.4.1 Background on persistent homology	18
1.4.2 Related literature on persistent homology of directed networks	20
1.4.3 The Vietoris-Rips filtration of a network	21
1.4.4 The Dowker filtration of a network	22
1.4.5 A Functorial Dowker Theorem	24
1.4.6 Dowker persistence diagrams and asymmetry	27
1.5 Persistent path homology of networks	35
1.5.1 Path homology of digraphs	35
1.5.2 The persistent path homology of a network	38

1.5.3	An application: Characterizing the diagrams of cycle networks	41
1.6	The case of compact networks	42
1.6.1	ε -systems and finite sampling	42
1.6.2	Weak isomorphism and $d_{\mathcal{N}}$	44
1.6.3	An additional axiom coupling weight function with topology	47
1.6.4	Skeletons, motifs, and motif reconstruction	49
1.7	Diagrams of compact networks and convergence results	53
1.8	Completeness, compactness, and geodesics	54
1.8.1	Completeness of $(\mathcal{CN}/\cong^w, d_{\mathcal{N}})$	54
1.8.2	Precompact families in \mathcal{CN}/\cong^w	55
1.8.3	Geodesics: existence and explicit examples	55
1.9	Measure networks and the $d_{\mathcal{N},p}$ distances	61
1.9.1	The structure of measure networks	63
1.9.2	Couplings and the distortion functional	65
1.9.3	Interval representation and continuity of distortion	66
1.9.4	Optimality of couplings in the network setting	68
1.9.5	The Network Gromov-Wasserstein distance	69
1.9.6	The Network Gromov-Prokhorov distance	72
1.9.7	Lower bounds and measure network invariants	72
1.10	Computational aspects	78
1.10.1	Software packages	79
1.10.2	Simulated hippocampal networks	79
1.10.3	Clustering SBMs and migration networks	82
2.	Metric structures of $d_{\mathcal{N}}$ and $d_{\mathcal{N},p}$	89
2.1	Proofs from §1.2	89
2.2	ε -systems and finite sampling	91
2.3	Proofs from §1.6.2	96
2.4	Skeletons and motif reconstruction	102
2.4.1	The skeleton of a compact network	102
2.4.2	Reconstruction via motifs and skeletons	106
2.5	Completeness, compactness, and geodesics	110
2.5.1	The completion of \mathcal{CN}/\cong^w	110
2.5.2	Precompact families in \mathcal{CN}/\cong^w	114
2.5.3	Geodesics in \mathcal{CN}/\cong^w	115
2.6	Lower bounds on $d_{\mathcal{N}}$	118
2.6.1	Quantitative stability of invariants of networks	121
2.6.2	Proofs involving lower bounds for $d_{\mathcal{N}}$	123
2.7	Proofs from §1.9	127

3.	Persistent Homology on Networks	138
3.1	Background on persistence and interleaving	138
3.1.1	Homology, persistence, and tameness	138
3.1.2	Interleaving distance and stability of persistence vector spaces. .	140
3.2	Simplicial constructions	142
3.2.1	Stability of Vietoris-Rips and Dowker constructions	142
3.2.2	The Functorial Dowker Theorem and equivalence of diagrams .	143
3.2.3	The equivalence between the finite FDT and the simplicial FNTs	149
3.2.4	Dowker persistence diagrams of cycle networks	156
3.3	Persistent path homology	160
3.3.1	Digraph maps and functoriality	160
3.3.2	Homotopy of digraphs	163
3.3.3	The Persistent Path Homology of a Network	164
3.3.4	PPH and Dowker persistence	165
3.4	Diagrams of compact networks	170
4.	Algorithms, computation, and experiments	172
4.1	The complexity of computing $d_{\mathcal{N}}$	172
4.2	Computing lower bounds for $d_{\mathcal{N}}$	173
4.2.1	An algorithm for computing minimum matchings	174
4.2.2	Computational example: randomly generated networks	174
4.2.3	Computational example: simulated hippocampal networks	176
4.3	Lower bounds for $d_{\mathcal{N},p}$	178
4.3.1	Numerical stability of entropic regularization	179
4.4	Complexity of PPH and algorithmic aspects	184
4.4.1	The modified algorithm	189
4.5	More experiments using Dowker persistence	191
4.5.1	U.S. economy input-output accounts	191
4.5.2	U.S. migration	195
4.5.3	Global migration	205
	Bibliography	211

List of Tables

Table	Page
1.1 Left: The five classes of SBM networks corresponding to the experiment in §1.10.3. N refers to the number of communities, v refers to the vector that was used to compute a table of means via $G_5(v)$, and n_i is the number of nodes in each community. Right: $G_5(v)$ for $v = [0, 25, 50, 75, 100]$	85
1.2 Two-community SBM networks as described in §1.10.3.	87
4.1 The first two columns contain sample 0-dimensional persistence intervals, as produced by Javaplex. We have added the labels in column 3, and the common δ -sinks in column 4.	193
4.2 Quantitative estimates on migrant flow, following the interpretation presented in §4.5.2. In each row, we list a simplex of the form $[s_i, s_j]$ (resp. $[s_i, s_j, s_l]$ for 2-simplices) and any possible δ -sinks s_k . We hypothesize that s_k receives at least $(1 - \delta)(\text{influx}(s_k))$ migrants from each of s_i, s_j (resp. s_i, s_j, s_l)—these lower bounds are presented in the third column. The fourth column contains the true migration numbers. Notice that the [FL,GA] simplex appears to show the greatest error between the lower bound and the true migration. Following the interpretation suggested earlier in §4.5.2, this indicates that Florida and Georgia appear to have strong coherence of preference, relative to the other pairs of states spanning 1-simplices in this table.	202
4.3 Quantitative estimates on migrant flow, following the interpretation presented in §4.5.2. The entries in this table follow the same rules as those of Table 4.2. Notice that the [OR,WA] and [AZ,CA] simplices show the greatest error between the lower bound and the true migration. Following the interpretation in §4.5.2, this suggests that these two pairs of states exhibit stronger coherence of preference than the other pairs of states forming 1-simplices in this table.	203

- 4.4 Short 0-dimensional Dowker persistence intervals capture regions which receive most of their incoming migrants from a single source. Each interval $[0, \delta)$ corresponds to a 0-simplex which becomes subsumed into a 1-simplex at resolution δ . We list these 1-simplices in the second column, and their δ sinks in the third column. The definition of a δ -sink enables us to produce a lower bound on the migration into each sink, which we provide in the fourth column. We also list the true migration numbers in the fifth column, and the reader can consult §4.5.2 for our explanation of the error between the true migration and the lower bounds on migration. 208
- 4.5 Representative 1-cycles for several intervals in the 1-dimensional persistence barcode for the global migration dataset. Each of the first four cycles has the special property that it becomes a boundary due to a single sink at the right endpoint of its associated persistence interval. This permits us to obtain a lower bound on the migration into this sink from each of the regions in the cycle. The last row contains a cycle without this special property. The font colors of the persistence intervals correspond to the colors of the highlighted 1-dimensional bars in Figure 4.16. 210

List of Figures

Figure	Page
1.1 The two networks on the left have different cardinalities, but computing correspondences shows that $d_{\mathcal{N}}(X, Y) = 1$. Similarly one computes $d_{\mathcal{N}}(X, Z) = 0$, and thus $d_{\mathcal{N}}(Y, Z) = 1$ by triangle inequality. On the other hand, the bijection given by the arrows shows $\hat{d}_{\mathcal{N}}(Y, Z) = 1$. Applying Proposition 12 then recovers $d_{\mathcal{N}}(X, Y) = 1$.	10
1.2 The directed circle $(\vec{\mathbb{S}}^1, \omega_{\vec{\mathbb{S}}^1})$, the directed circle on 6 nodes $(\vec{\mathbb{S}}_6^1, \omega_{\vec{\mathbb{S}}_6^1})$, and the directed circle with reversibility ρ , for some $\rho \in [1, \infty)$. Traveling in a clockwise direction is possibly only in the directed circle with reversibility ρ , but this incurs a penalty modulated by ρ .	15
1.3 A cycle network on 6 nodes, along with its weight matrix. Note that the weights are highly asymmetric.	16
1.4 A network SBM on 50 nodes, split into 5 communities, along with the matrices of means and variances. The deepest blue corresponds to values ≈ 1 , and the deepest yellow corresponds to values ≈ 29 .	17
1.5 A schematic of some of the simplicial constructions on directed networks. \mathcal{F} is the collection of filtered simplicial complexes. Dgm is the collection of persistence diagrams. We study the Rips (\mathfrak{R}) and Dowker ($\mathfrak{D}^{si}, \mathfrak{D}^{so}$) filtrations, each of which takes a network as input and produces a filtered simplicial complex. s and t denote the network transformations of symmetrization (replacing a pair of weights between two nodes by the maximum weight) and transposition (swapping the weights between pairs of nodes). \mathfrak{R} is insensitive to both s and t . But $\mathfrak{D}^{si} \circ t = \mathfrak{D}^{so}$, $\mathfrak{D}^{so} \circ t = \mathfrak{D}^{si}$, and in general, \mathfrak{D}^{si} and \mathfrak{D}^{so} are <i>not</i> invariant under t (Theorem 50).	17
1.6 Computing the Dowker sink and source complexes of a network (X, ω_X) . Observe that the sink and source complexes are different in the range $1 \leq \delta < 2$.	23

1.7	The first column contains illustrations of cycle networks G_3, G_4, G_5 and G_6 . The second column contains the corresponding Dowker persistence barcodes, in dimensions 0 and 1. Note that the persistent intervals in the 1-dimensional barcodes agree with the result in Theorem 40. The third column contains the Rips persistence barcodes of each of the cycle networks. Note that for $n = 3, 4$, there are no persistent intervals in dimension 1. On the other hand, for $n = 6$, there are two persistent intervals in dimension 1.	31
1.8	(Y, ω_Y) is the (a, c) -swap of (X, ω_X) .	32
1.9	Dowker persistence barcodes of networks (X, ω_X) and (Y, ω_Y) from Figure 1.8.	33
1.10	Rips persistence barcodes of networks (X, ω_X) and (Y, ω_Y) from Figure 1.8. Note that the Rips diagrams indicate no persistent homology in dimensions higher than 0, in contrast with the Dowker diagrams in Figure 1.9.	33
1.11	A two-node digraph on the vertex set $Y = \{a, b\}$.	36
1.12	Two types of square digraphs.	37
1.13	Working over $\mathbb{Z}/2\mathbb{Z}$ coefficients, we find that $\text{Dgm}_1^{\Xi}(\mathcal{X})$ and $\text{Dgm}_1^{\mathfrak{D}}(\mathcal{Y})$ are trivial, whereas $\text{Dgm}_1^{\mathfrak{D}}(\mathcal{X}) = \text{Dgm}_1^{\Xi}(\mathcal{Y}) = \{(1, 2)\} = \{(1, 2)\}$.	40
1.14	Left: $\mathfrak{G}_{\square_3}^{\delta}$ is (digraph) homotopy equivalent to a point at $\delta = 1$, as can be seen by collapsing points along the orange lines. Right: $\mathfrak{D}_{\delta, \square_3}^{\text{si}}$ becomes contractible at $\delta = \sqrt{2}$, but has nontrivial homology in dimension 2 that persists across the interval $[1, \sqrt{2}]$.	41
1.15	Relaxing the requirements on the maps of this “tripod structure” is a natural way to weaken the notion of strong isomorphism.	45
1.16	Note that Remark 69 <i>does not</i> fully characterize weak isomorphism, even for finite networks: All three networks above, with the given weight matrices, are Type I weakly isomorphic since C maps surjectively onto A and B . But there are no surjective, weight preserving maps $A \rightarrow B$ or $B \rightarrow A$.	46

1.17 Left: Z represents a terminal object in $\mathfrak{p}(X)$, and f, g are weight preserving surjections $X \rightarrow Z$. Here $\varphi \in \text{Aut}(Z)$ is such that $g = \varphi \circ f$. Right: Here we show more of the poset structure of $\mathfrak{p}(X)$. In this case we have $X \succeq V \succeq Y \dots \succeq Z$.	51
1.18 Interpolating between the skeleton and blow-up constructions.	52
1.19 Illustrations of the finite networks we consider in this paper. Notice that the edge weights are asymmetric. The numbers in each node correspond to probability masses; for each network, these masses sum to 1.	62
1.20 The $d_{\mathcal{N},p}$ distance between the two one-node networks is simply $\frac{1}{2} \alpha - \alpha' $. In Example 109 we give an explicit formula for computing $d_{\mathcal{N},p}$ between an arbitrary network and a one-node network.	70
1.21 Networks at $d_{\mathcal{N},p}$ -distance zero which are not strongly isomorphic.	71
1.22 Bottom right: Sample place cell spiking pattern matrix. The x -axis corresponds to the number of time steps, and the y -axis corresponds to the number of place cells. Black dots represent spikes. Clockwise from bottom middle: Sample distribution of place field centers in 4, 3, 0, 1, and 2-hole arenas.	81
1.23 Single linkage dendrogram corresponding to the distance matrix obtained by computing bottleneck distances between 1-dimensional Dowker persistence diagrams of our database of hippocampal networks (§1.10.2). Note that the 4, 3, and 2-hole arenas are well separated into clusters at threshold 0.1.	83
1.24 Single linkage dendrogram corresponding to the distance matrix obtained by computing bottleneck distances between 1-dimensional Rips persistence diagrams of our database of hippocampal networks (§1.10.2). Notice that the hierarchical clustering fails to capture the correct arena types.	84

1.25 Left: TLB dissimilarity matrix for SBM community networks in §1.10.3. Classes 1 and 3 are similar, even though networks in Class 3 have twice as many nodes as those in Class 1. Classes 2 and 5 are most dissimilar because of the large difference in their edge weights. Class 4 has a different number of communities than the others, and is dissimilar to Classes 1 and 3 even though all their edge weights are in comparable ranges. Right: TLB dissimilarity matrix for two-community SBM networks in §1.10.3. The near-zero values on the diagonal are a result of using the adaptive λ -search described in Chapter 4.	86
1.26 Result of applying the TLB to the migration networks in §1.10.3. Left: Dissimilarity matrix. Nodes 1-5 correspond to female migration from 1960- 2000, and nodes 6-10 correspond to male migration from 1960-2000. Right: Single linkage dendrogram. Notice that overall migration patterns change in time, but within a time period, migration patterns are grouped according to gender.	88
2.1 The trace map erases data between pairs of nodes.	119
2.2 The out map applied to each node yields the greatest weight of an arrow <i>leaving</i> the node, and the in map returns the greatest weight <i>entering</i> the node.	120
2.3 Lower-bounding $d_{\mathcal{N}}$ by using global spectra (cf. Example 148).	122
3.1 Directed d -cubes that are all homotopy equivalent.	163
4.1 Left: Lower bound matrix arising from matching local spectra on the database of community networks. Right: Corresponding single linkage dendrogram. The labels indicate the number of communities and the total number of nodes. Results correspond to using local spectra as described in Proposition 180.	175
4.2 Left: Lower bound matrix arising from matching global spectra on the database of community networks. Right: Corresponding single linkage dendrogram. The labels indicate the number of communities and the total number of nodes. Results correspond to using global spectra as signatures.	176
4.3 Single linkage dendrogram based on local spectrum lower bound of Propo- sition 180 corresponding to hippocampal networks with place field radii $0.2L$, $0.1L$, and $0.05L$ (clockwise from top left).	178

4.4	Sinkhorn computations for $d_{\mathcal{N},2}(\mathcal{X}, \mathcal{Y})$ are carried out in the “stable region” for K , and the end result is rescaled to recover $d_{\mathcal{N},2}(X, Y)$	183
4.5	Left: The rows and columns of M_p are initially arranged so that the domain and codomain vectors are in increasing and decreasing allow time, respectively. If there are no domain (codomain) vectors having a particular allow time, then the corresponding vertical (horizontal) strip is omitted. Right: After converting to column echelon form, the domain vectors of $M_{p,G}$ need not be in the original ordering. But the codomain vectors are still arranged in decreasing allow time.	189
4.6	0 and 1-dimensional Dowker persistence barcodes for US economic sector data, obtained by the process described in §4.5.1. The long 1-dimensional persistence intervals that are colored in red are examined in §4.5.1 and Figures 4.9,4.12.	193
4.7	Investment patterns at $\delta = 0.75$	195
4.8	Investment patterns at $\delta = 0.94$	195
4.9	Here we illustrate the representative nodes for one of the 1-dimensional persistence intervals in Figure 4.6. This 1-cycle [PC,CH] + [CH,PL] + [PL,WO] - [WO,FM] + [FM,PM] - [PM,PC] persists on the interval [0.75, 0.95]. At $\delta = 0.94$, we observe that this 1-cycle has joined the homology equivalence class of the shorter 1-cycle illustrated on the right. Unidirectional arrows represent an asymmetric flow of investment. A full description of the meaning of each arrow is provided in §4.5.1.	195
4.10	Investment patterns at $\delta = 0.81$	196
4.11	Investment patterns at $\delta = 0.99$	196
4.12	Representative nodes for another 1-dimensional persistence interval in Figure 4.6. A full description of this cycle is provided in §4.5.1.	196

- 4.13 An example of a 1-cycle becoming a 1-boundary due to a single mutual sink r , as described in the interpretation of 1-cycles in §4.5.2. The figure on the left shows a connected component of $\mathfrak{D}_{\delta,S}^{\text{si}}$, consisting of $[s_1, s_2]$, $[s_2, s_3]$, $[s_3, s_4]$. The arrows are meant to suggest that r will eventually become a δ -sink for each of these 1-simplices, for some large enough δ . The progression of these simplices for increasing values of δ are shown from left to right. In the leftmost figure, r is not a δ -sink for any of the three 1-simplices. Note that r has become a δ -sink for $[s_3, s_4]$ in the middle figure. Finally, in the rightmost figure, r has become a δ -sink for each of the three 1-simplices. . . 201
- 4.14 0 and 1-dimensional Dowker persistence barcodes for U.S. migration data . 203
- 4.15 U.S. map with representative cycles of the persistence intervals that were highlighted in Figure 4.14. The cycle on the left appears at $\delta_1 = 0.87$, and the cycle on the right appears at $\delta_2 = 0.90$. The red lines indicate the 1-simplices that participate in each cycle. Each red line is decorated with an arrowhead $s_i \rightarrow s_j$ if and only if s_j is a sink for the simplex $[s_i, s_j]$. The blue arrows point towards all possible alternative δ -sinks, and are interpreted as follows: Tennessee is a 0.90-sink for the Kentucky-Georgia simplex, West Virginia is a 0.90-sink for the Ohio-Florida simplex, and Alabama is a 0.90-sink for the Georgia-Florida simplex. 204
- 4.16 Dowker persistence barcodes for global migration dataset. 208
- 4.17 **Top:** Two cycles corresponding to the left endpoints of the (Djibouti-Somalia-Uganda-Eritrea-Ethiopia) and (Kiribati-Papua New Guinea-Australia-United Kingdom-Tuvalu) persistence intervals listed in Table 4.5. The δ values are 0.73, 0.77, respectively. **Bottom:** Two cycles corresponding to the left endpoints of the (China-Thailand-Philippines) and (China-Indonesia-Malaysia) persistence intervals listed in Table 4.5. The δ values are 0.77, 0.75, respectively. **Meaning of arrows:** In each cycle, an arrow $s_i \rightarrow s_j$ means that $\omega_S(s_i, s_j) \leq \delta$, i.e. that s_j is a sink for the simplex $[s_i, s_j]$. We can verify separately that for $\delta = 0.77$, the Kiribati-Papua New Guinea simplex has the Solomon Islands as a δ -sink, and that the Philippines-Thailand simplex has Taiwan as a δ -sink. Similarly, the China-Malaysia simplex has Singapore as a δ -sink, for $\delta = 0.75$ 209

Background definitions and conventions

We write \mathbb{N} to denote the natural numbers $1, 2, 3, \dots$, \mathbb{Z}_+ to denote $\{0\} \cup \mathbb{N}$, and \mathbb{R}_+ to denote the nonnegative reals. The power set of a set S is denoted $\text{pow}(S)$. The cardinality of S is denoted $|S|$ or $\text{card}(S)$. The empty set is denoted \emptyset . The identity map on S is denoted id_S .

A simplicial complex K consists of a vertex set $V(K)$ and a collection of nonempty subsets $\sigma \in \text{pow}(V(K))$ such that whenever $\sigma \in K$ and $\tau \subseteq \sigma$, we also have $\tau \in K$. This is typically known as an abstract simplicial complex. Any such complex has a *geometric realization* which can be realized as follows. For convenience, assume $n := |V(K)| < \infty$. Enumerate $V(K)$ as $\{v_1, v_2, \dots, v_n\}$. Define $f : V(K) \rightarrow \mathbb{R}^n$ by $v_i \mapsto e_i$, the i th standard basis vector in \mathbb{R}^n . This is the vector of all zeros with a 1 in the i th coordinate. Given a simplex $\sigma \in K$, the corresponding geometric simplex is the convex hull of $\{f(v) : v \in \sigma\}$, i.e. the collection of all points $\sum_{v \in \sigma} t_v f(v)$ such that $\sum_{v \in \sigma} t_v = 1$ and each $t_v \geq 0$. The geometric realization of K , denoted $|K|$, is the collection of all geometric simplices corresponding to simplices in K .

A metric space is a set X and a metric function $d_X : X \times X \rightarrow [0, \infty)$ such that for all $x, x', x'' \in X$, the following holds:

- (Nondegeneracy) $d_X(x, x') = 0$ if and only if $x = x'$.
- (Symmetry) $d_X(x, x') = d_X(x', x)$.
- (Triangle inequality) $d_X(x, x'') \leq d_X(x, x') + d_X(x', x'')$.

A metric space (X, d_X) is *totally bounded* if for any $\varepsilon > 0$, there exists a finite subset $S \subseteq X$ such that $d_X(x, S) < \varepsilon$ for all $x \in S$. Here $d_X(x, S) := \min_{s \in S} d_X(x, s)$.

Given a set X , a *topology* on X is a subset $\tau_X \subseteq \text{pow}(X)$ such that:

- Both \emptyset and X belong to τ_X
- Any arbitrary union of elements of τ_X belongs to τ_X (i.e. τ_X is closed under taking arbitrary unions)
- A finite intersection of elements of τ_X belongs to τ_X (i.e. τ_X is closed under taking finite intersections).

The elements of τ_X are referred to as the *open sets* of X .

An *open cover* of a topological space X is a collection of open sets $\{U_i \subseteq X : i \in I\}$ indexed by some set I such that each U_i is nonempty, and $\bigcup_{i \in I} U_i = X$.

A *base* or *basis* for τ_X is a subcollection of τ_X such that every open set in X can be written as a union of elements in the subcollection. A *local base* at a point $x \in X$ is a collection of open sets containing x such that every open set containing x contains some element in this collection.

There are always two topologies that one can place on a set X : the *discrete topology* $\tau_X = \text{pow}(X)$, and the *trivial topology* $\tau_X = \{\emptyset, X\}$.

A *point cloud* is a discrete subset of d -dimensional Euclidean space for some $d \in \mathbb{N}$.

Given two topological spaces X and Y , two continuous maps $f, g : X \rightarrow Y$ are said to be *homotopic* if there exists a continuous map $F : X \times [0, 1] \rightarrow Y$ such that $F|_{X \times \{0\}} = f$ and $F|_{X \times \{1\}} = g$. X and Y are said to be *homotopy equivalent* if there exist maps $f : X \rightarrow Y$ and $g : Y \rightarrow X$ such that $g \circ f \simeq \text{id}_X$ and $f \circ g \simeq \text{id}_Y$. In this case, f and g are said to be *homotopy inverses*.

The indicator function of a set S is denoted $\mathbb{1}_S$. We denote measure spaces via the triple (X, \mathcal{F}, μ) , where X is a set, \mathcal{F} is a σ -field on X , and μ is the measure on \mathcal{F} . Given a measure space (X, \mathcal{F}, μ) , we write $L^0 = L^0(\mu)$ to denote the collection of \mathcal{F} -measurable functions $f : X \rightarrow \mathbb{R}$. For all $p \in (0, \infty)$ and all $f \in L^0$, we define $\|f\|_p := (\int |f|^p d\mu)^{1/p}$. For $p = \infty$, $\|f\|_\infty := \inf\{M \in [0, \infty] : \mu(|f| > M) = 0\}$. Then for any $p \in (0, \infty]$, $L^p = L^p(\mu) := \{f \in L^0 : \|f\|_p < \infty\}$.

Given a measurable real-valued function $f : X \rightarrow \mathbb{R}$ and $t \in \mathbb{R}$, we will occasionally write $\{f \leq t\}$ to denote the set $\{x \in X : f(x) \leq t\}$.

Lebesgue measure on the reals will be denoted by λ . We write λ_I to denote the Lebesgue measure on the unit interval $I = [0, 1]$.

Suppose we have a measure space (X, \mathcal{F}, μ) , a measurable space (Y, \mathcal{G}) , and a measurable function $f : X \rightarrow Y$. The *pushforward* or *image measure* of f is defined to be the measure $f_*\mu$ on \mathcal{G} given by writing $f_*\mu(A) := \mu(f^{-1}[A])$ for all $A \in \mathcal{G}$.

A particular case where we deal with pushforward measures is the following: given a product space $\mathcal{X} = X_1 \times X_2 \times \dots \times X_n$ and a measure μ on \mathcal{X} , the canonical projection maps $\pi_i : \mathcal{X} \rightarrow X_i$, for $i = 1, \dots, n$, define pushforward measures that we denote $(\pi_i)_*\mu$. If each X_i is itself a measure space with measure μ_i , then we say that μ has *marginals* μ_i , for $i = 1, \dots, n$, if $(\pi_i)_*\mu = \mu_i$ for each i . We also consider projection maps of the form $(\pi_i, \pi_j, \pi_k) : \mathcal{X} \rightarrow X_i \times X_j \times X_k$ for $i, j, k \in \{1, \dots, n\}$, and denote the corresponding pushforward by $(\pi_i, \pi_j, \pi_k)_*\mu$. Notice that we can take further projections of the form $(\pi_i, \pi_j)^{ijk} : X_i \times X_j \times X_k \rightarrow X_i \times X_j$, and the images of these projections are precisely those given by projections of the form $(\pi_i, \pi_j) : \mathcal{X} \rightarrow X_i \times X_j$.

Remark 1. Let $\mathcal{X} = X_1 \times X_2 \times \dots \times X_n$ be a product space with a measure μ as above, and suppose each X_i is equipped with a measure μ_i such that μ has marginals μ_i . Let $i, j, k \in \{1, \dots, n\}$. Then the measure $(\pi_i, \pi_j, \pi_k)_*\mu$ on $X_i \times X_j \times X_k$ has marginals $(\pi_i, \pi_k)_*\mu$ and $(\pi_j)_*\mu$ on $X_i \times X_k$ and X_j , respectively. To see this, let the projections

$X_i \times X_j \times X_k \rightarrow X_i \times X_k$ and $X_i \times X_j \times X_k \rightarrow X_j$ be denoted $(\pi_i, \pi_k)^{ijk}$ and $(\pi_j)^{ijk}$, respectively. Let $B \subseteq X_j$ be measurable. Then

$$(\pi_j)_*^{ijk}(B) = (\pi_i, \pi_j, \pi_k)_*\mu(X_i \times B \times X_k) = \mu(X_1 \times \dots \times X_{j-1} \times B \times X_{j+1} \times \dots \times X_n) \\ = (\pi_j)_*(B).$$

Next let $A \subseteq X_i$ and $C \subseteq X_k$ be measurable. Then

$$(\pi_i, \pi_k)_*^{ijk}(A \times C) = (\pi_i, \pi_j, \pi_k)_*\mu(A \times X_j \times C) \\ = \mu(X_1 \times \dots \times X_{i-1} \times A \times X_{i+1} \times \dots \times X_{k-1} \times C \times X_{k+1} \times \dots \times X_n) \\ = (\pi_i, \pi_k)_*(A \times C).$$

Chapter 1: Introduction

Let X be a set, and let $\omega_X : X \times X \rightarrow \mathbb{R}$ be any function. The pair (X, ω_X) (with some additional constraints, cf. Definition 1) is what we refer to as a *network*, and is the central object of study in this thesis. Networks arise in many guises, and in many contexts [90], so it is necessary to explain the motivation behind the preceding definition. Perhaps the most common type of network is an undirected graph $G = (V, E)$ consisting of a vertex set V and edge set E . Letting n be the number of elements in V , the adjacency matrix of G is an $n \times n$ binary matrix with a 1 in entry (i, j) if $v_i \rightarrow v_j$, and 0 otherwise. In particular, the adjacency matrix is symmetric. More generally, a directed graph relaxes the symmetry condition by allowing for individual edges $v_i \rightarrow v_j$ and $v_j \rightarrow v_i$. Finally, a weighted, directed graph allows for real-valued weights on the edges. The pair (X, ω_X) encapsulates this idea. In this sense, the networks we study in this thesis are weighted, directed graphs with self-loops¹. We denote the collection of *all* networks (cf. Definition 1) by \mathcal{N} .

In addition to graphs, \mathcal{N} contains many other classes of objects, including metric spaces, directed metric spaces, Riemannian manifolds, and Finsler manifolds. As will be shown throughout this work, the perspective of viewing networks as generalized metric spaces (as opposed to combinatorial objects) enables the import of many techniques—both new and old—from the theory of metric spaces into network analysis. This ranges from classical results of Gromov on reconstructions of metric spaces, to more modern tools such as persistent homology and applied optimal transport.

The question which initially motivated this thesis is as follows. In an unpublished manuscript first appearing in 2012, Grigor’yan, Lin, Muranov, and Yau defined a homology theory for digraphs (i.e. directed graphs) called *path homology* that generalized the standard notion of simplicial homology [62]. By this time, the theory of *persistent homology*—a multiresolution version of simplicial homology used for data analysis—had already become quite popular. So a natural question was: would it be possible to produce a *persistent* version of path homology? More specifically, would it be possible to produce a persistent path homology theory with satisfactory algorithms and implementations as well as nice theoretical properties?

In this thesis, we provide a positive answer to the preceding question. More generally, we develop a framework with machinery in place for extending many metric data analysis

¹ While multigraphs (graphs allowing for multiple edges between a pair of vertices) and hypergraphs (multiple nodes sharing a “hyperedge”) are often useful in practice, we do not consider them in this work.

techniques (including variants of persistent homology) to the setting of directed networks while preserving nice theoretical properties. In particular, we generalize the two most common extant *simplicial* persistent homology methods—the Vietoris-Rips and Čech methods [47]—and fit them to the directed network setting. Our contributions in these directions have already been published [40, 37].

The crucial component of this framework is the examination of a (pseudo)metric $d_{\mathcal{N}}$ on the collection of all networks, which first appeared in a slightly restricted version in [22]. In this setting, $d_{\mathcal{N}}$ was used to provide theoretical guarantees for the robustness of certain hierarchical clustering methods on directed networks. This network distance is structurally analogous to the Gromov-Hausdorff distance d_{GH} on \mathcal{CM} —the collection of all compact metric spaces [17]. An important remark is that $d_{\mathcal{N}}$ (or even d_{GH}) is NP-hard to compute.

With the development of various data analysis methods relying on $d_{\mathcal{N}}$, it became increasingly necessary to understand and develop the core theoretical properties of $d_{\mathcal{N}}$ itself. Keeping this goal in mind, we provide a comprehensive analysis of $d_{\mathcal{N}}$ in this thesis (see also [35, 39, 34]). Because its structural counterpart for metric spaces (d_{GH}) is already well-studied, we were interested in extending the desirable results for d_{GH} to the setting of $d_{\mathcal{N}}$. As we show here, a wide range of results does extend to the $d_{\mathcal{N}}$ setting. The crux of this development is in realizing that d_{GH} , by nature of its definition, enforces enough structure on \mathcal{CM} that only certain abstract consequences of metric properties such as symmetry and triangle inequality become necessary when proving results on \mathcal{CM} . These consequences, when assumed as properties of networks, are quite natural. Most importantly, they allow us to prove strong results about $d_{\mathcal{N}}$. We expect that in general, when proving results about d_{GH} , verifying these assumptions (cf. Definitions 1, 2, and also 28) would allow one to prove the results on the much broader setting of $d_{\mathcal{N}}$ with little additional work.

As a natural extension of these ideas, we develop families of metrics $d_{\mathcal{N},p}$ for $p = [1, \infty]$ that are L^p versions of $d_{\mathcal{N}}$ (cf. [38]). These are structurally analogous to the Gromov-Wasserstein distances developed in [85, 116, 86, 117]. We currently have ongoing work that leverages the formulation and foundational work on $d_{\mathcal{N},p}$.

When studying a metric, an important item to clarify is the “curvature” of the metric [17]. Knowledge of curvature has important practical consequences; for example, it gives theoretical guarantees on the existence of means [96]. Following the extensive study of the (Alexandrov) curvature of the space of metric measure spaces in [117], we became interested in testing similar ideas in the settings of $d_{\mathcal{N}}$ and d_{GH} . Interestingly, even the restricted metric d_{GH} does not admit any curvature bounds (in the Alexandrov sense). The deeper reason behind this is the existence of “wild” geodesics in the space of compact metric spaces. In [36], we produced explicit constructions of infinite families of these wild-type geodesics in \mathcal{CM} ; these results are reproduced here.

The $d_{\mathcal{N}}$ distance at the core of this work is a pseudometric, and it is of natural importance to understand its zero sets thoroughly. This is related to the classic question “Can one hear the shape of a drum”: to understand the behavior of our methods, we need to know which networks are perceived by $d_{\mathcal{N}}$ to be the same. We provide a full treatment of

this question. We provide an independent definition of “weakly isomorphic networks” and prove that these are precisely the networks at 0 $d_{\mathcal{N}}$ -distance. Weakly isomorphic networks essentially live on a fiber over a core subnetwork that we call a *skeleton*. In particular, when comparing two networks having different sizes, one may traverse the fibers, pick out two representatives having the same number of nodes, and compute a simplified (but still NP-hard) distance $\hat{d}_{\mathcal{N}}$ to obtain the original distance $d_{\mathcal{N}}$.

We continually develop *network invariants* that serve as polynomial-time proxies for $d_{\mathcal{N}}$. The persistent homology methods are all examples of such invariants. One of the other invariants we consider, the *motif sets*, are based on the “curvature class” invariants defined by Gromov for metric spaces [65]. In the case of compact metric spaces, the motif sets form a *full* invariant. We are able to recover this result for a subcollection of \mathcal{CN} satisfying an additional topological condition that we call *coherence*, or more specifically, Axiom A2 (cf. Definition 28). In other words (some readers may be more familiar with this terminology), the map from weak isomorphism classes of (a subcollection of) networks to motif sets is *injective*. The proof of this result is quite short for metric spaces, but requires a long sequence of verifications for networks. The interesting part, however, is that the proof is made possible via Axiom A2, which is an abstract consequence of the triangle inequality (cf. Remark 74). In particular, any finite network satisfies Axiom A2, even if it violates the triangle inequality by an arbitrary margin. Actually, to be more accurate, given any finite network, one may traverse the fiber of weakly isomorphic networks down to its skeleton, and this skeleton will satisfy Axiom A2.

In addition to the theoretical results outlined above, we present practical implementations on both real and simulated datasets. These implementations are available as the PersNet, PPH, and GWnets software packages, and are written in a combination of Matlab, Python, and C++.

Finally, we note that this thesis combines narratives and results from [35, 39, 34, 36, 38, 37, 40]. All of these papers have been developed jointly with Facundo Mémoli. While each of these papers can be read in a self-contained manner, we have made the effort to consolidate the landscape developed in those papers and distill it into this thesis.

1.1 Organization of this thesis

Chapter 1 is written as an extended abstract that contains definitions, statements of main results, and examples. It provides a reasonably comprehensive overview of the entire thesis. Proofs of the main results, as well as auxiliary results, definitions, and examples, are provided in the later chapters.

Chapter 1 is organized as follows. The first set of definitions involving $d_{\mathcal{N}}$ is presented in §1.2. In §1.3 we present some network models that serve as important examples throughout this work. Then we switch to persistent homology and persistent path homology in §1.4–1.5. Afterwards we return to $d_{\mathcal{N}}$ in §1.6. Results on convergence follow in §1.7,

which also contains important guarantees on when persistent homology can be meaningfully applied to infinite networks. In §1.8 we discuss the existence of convex-combination and wild-type geodesics in \mathcal{N} . Development of the $d_{\mathcal{N},p}$ distances is carried out in §1.9. Finally, in §1.10, we discuss computational complexity, algorithms, and the results of implementing our methods on one particular dataset.

Chapter 2 is devoted to proofs of the results about $d_{\mathcal{N}}$ stated in §1, along with auxiliary results. In particular, §2.6 discusses lower bounds for computing $d_{\mathcal{N}}$. Chapter 3 contains proofs of statements involving persistent homology. Finally we address computational aspects and additional experiments in Chapter 4.

1.2 Networks and $d_{\mathcal{N}}$: Definition, reformulations, and first results

Proofs from this section can be found in §2.1.

From a historical perspective, our definition of a network is based on a definition that appeared in [20, 22]. Here a network was defined to be a pair (X, A_X) , where X is a finite set of nodes and $A_X : X \times X \rightarrow \mathbb{R}_+$ is a metric function without the symmetry or triangle inequality properties.

A natural generalization is to instead consider pairs (X, ω_X) where X is any arbitrary set and ω_X is any function from $X \times X$ to \mathbb{R} . While $d_{\mathcal{N}}$ can still be defined on such networks, this definition is much less amenable to results. In such a setting, only ω_X has any structure, and we are essentially restricted to studying functions $X \times X \rightarrow \mathbb{R}$. It turns out that the following definition is much better for our purposes.

Definition 1 (Network). A *network* is a pair (X, ω_X) , where X is a first countable topological space and $\omega_X : X \times X \rightarrow \mathbb{R}$ is continuous with respect to the product topology. Here X and ω_X are referred to as a *node set* and a *weight function*, respectively. The collection of all networks is denoted \mathcal{N} . A *subnetwork* of (X, ω_X) is any subset of X (equipped with the subspace topology) along with the appropriate restriction of ω_X .

Recall that a space is first countable if each point in the space has a countable local basis (see [114, p. 7] for more details). First countability is a technical condition guaranteeing that when the underlying topological space of a network is compact, it is also sequentially compact. Notice that these conditions are automatically satisfied in the finite setting, as ω_X is then trivially continuous.

A further observation is that the “correct” restriction of \mathcal{N} to work with is the collection of *compact* networks:

Definition 2 (Compact and finite networks). A *compact network* is a network (X, ω_X) where X is compact. The collection of compact networks is denoted \mathcal{CN} . This includes the collection of *finite* networks, which we denote by \mathcal{FN} .

For data analysis purposes, we expect to only ever work with finite networks. However, datasets are often viewed as being sampled from some “infinite” object (as is the case

for very large datasets), and compact networks turn out to be the appropriate model for our purposes. In particular, whereas \mathcal{FN} is not complete, we have the natural inclusion $\mathcal{FN} \subseteq \mathcal{CN}$, and the latter is a complete pseudometric space (see §1.6).

Letting \mathcal{FM} , \mathcal{CM} , and \mathcal{M} denote the spaces of finite, compact, and arbitrary metric spaces, respectively, we also note the containments

$$\mathcal{FM} \subsetneq \mathcal{FN}, \mathcal{CM} \subsetneq \mathcal{CN}, \text{ and } \mathcal{M} \subsetneq \mathcal{N}.$$

We now proceed to define $d_{\mathcal{N}}$, starting with some auxiliary definitions.

Definition 3 (Correspondence). Let $(X, \omega_X), (Y, \omega_Y) \in \mathcal{N}$. A *correspondence between X and Y* is a relation $R \subseteq X \times Y$ such that $\pi_X(R) = X$ and $\pi_Y(R) = Y$, where π_X and π_Y are the canonical projections of $X \times Y$ onto X and Y , respectively. The collection of all correspondences between X and Y will be denoted $\mathcal{R}(X, Y)$, abbreviated to \mathcal{R} when the context is clear.

Example 2 (1-point correspondence). Let X be a set, and let $\{p\}$ be the set with one point. Then there is a unique correspondence $R = \{(x, p) : x \in X\}$ between X and $\{p\}$.

Example 3 (Diagonal correspondence). Let $X = \{x_1, \dots, x_n\}$ and $Y = \{y_1, \dots, y_n\}$ be two enumerated sets with the same cardinality. A useful correspondence is the *diagonal correspondence*, defined as $\Delta := \{(x_i, y_i) : 1 \leq i \leq n\}$. When X and Y are infinite sets with the same cardinality, and $\varphi : X \rightarrow Y$ is a given bijection, then we write the diagonal correspondence as $\Delta := \{(x, \varphi(x)) : x \in X\}$.

Definition 4 (Distortion of a correspondence). Let $(X, \omega_X), (Y, \omega_Y) \in \mathcal{N}$ and let $R \in \mathcal{R}(X, Y)$. The *distortion* of R is given by:

$$\text{dis}(R) := \sup_{(x,y),(x',y') \in R} |\omega_X(x, x') - \omega_Y(y, y')|.$$

Remark 4 (Composition of correspondences). Let $(X, \omega_X), (Y, \omega_Y), (Z, \omega_Z) \in \mathcal{N}$, and let $R \in \mathcal{R}(X, Y)$, $S \in \mathcal{R}(Y, Z)$. Then we define:

$$R \circ S := \{(x, z) \in X \times Z \mid \exists y, (x, y) \in R, (y, z) \in S\}.$$

In the proof of Theorem 72, we verify that $R \circ S \in \mathcal{R}(X, Z)$, and that $\text{dis}(R \circ S) \leq \text{dis}(R) + \text{dis}(S)$.

Following prior work in [22], we make the following definition.

Definition 5 (The first network distance). Let $(X, \omega_X), (Y, \omega_Y) \in \mathcal{N}$. We define the *network distance* between X and Y as follows:

$$d_{\mathcal{N}}((X, \omega_X), (Y, \omega_Y)) := \frac{1}{2} \inf_{R \in \mathcal{R}} \text{dis}(R).$$

When the context is clear, we will often write $d_{\mathcal{N}}(X, Y)$ to denote $d_{\mathcal{N}}((X, \omega_X), (Y, \omega_Y))$. We define the collection of *optimal correspondences* \mathcal{R}^{opt} between X and Y to be the collection $\{R \in \mathcal{R}(X, Y) : \text{dis}(R) = 2d_{\mathcal{N}}(X, Y)\}$. This set is always nonempty when $X, Y \in \mathcal{FN}$, but may be empty in general.

Remark 5. We remark that when restricted to the special case of networks that are also metric spaces, the network distance $d_{\mathcal{N}}$ agrees with the Gromov-Hausdorff distance. Details on the Gromov-Hausdorff distance can be found in [17].

Remark 6. The intuition behind the preceding definition of network distance may be better understood by examining the case of a finite network. Given a finite set X and two edge weight functions ω_X, ω'_X defined on it, we can use the ℓ^∞ distance as a measure of network similarity between (X, ω_X) and (X, ω'_X) :

$$\|\omega_X - \omega'_X\|_{\ell^\infty(X \times X)} := \max_{x, x' \in X} |\omega_X(x, x') - \omega'_X(x, x')|.$$

A generalization of the ℓ^∞ distance is required when dealing with networks having different sizes: Given two sets X and Y , we need to decide how to match up points of X with points of Y . Any such matching will yield a subset $R \subseteq X \times Y$ such that $\pi_X(R) = X$ and $\pi_Y(R) = Y$, where π_X and π_Y are the projection maps from $X \times Y$ to X and Y , respectively. This is precisely a correspondence, as defined above. A valid notion of network similarity may then be obtained as the distortion incurred by choosing an optimal correspondence—this is precisely the idea behind the definition of the network distance above.

We will eventually verify that $d_{\mathcal{N}}$ as defined above is a pseudometric (Theorem 72), which will justify calling $d_{\mathcal{N}}$ a “network distance”. Because $d_{\mathcal{N}}$ is a pseudometric, it is important to understand its zero sets. To this end, we first develop the notion of *strong isomorphism* of networks. The definition follows below.

Definition 6 (Weight preserving maps). Let $(X, \omega_X), (Y, \omega_Y) \in \mathcal{N}$. A map $\varphi : X \rightarrow Y$ is *weight preserving* if:

$$\omega_X(x, x') = \omega_Y(\varphi(x), \varphi(x')) \text{ for all } x, x' \in X.$$

Definition 7 (Strong isomorphism). Let $(X, \omega_X), (Y, \omega_Y) \in \mathcal{N}$. To say (X, ω_X) and (Y, ω_Y) are *strongly isomorphic* means that there exists a weight preserving bijection $\varphi : X \rightarrow Y$. We will denote a strong isomorphism between networks by $X \cong^s Y$. Note that this notion is exactly the usual notion of isomorphism between weighted graphs.

Given two strongly isomorphic networks, i.e. networks $(X, \omega_X), (Y, \omega_Y)$ and a weight preserving bijection $\varphi : X \rightarrow Y$, it is easy to use the diagonal correspondence (Example 3) to verify that $d_{\mathcal{N}}(X, Y) = 0$. However, it is easy to see that the reverse implication is not true in general. Using the one-point correspondence (Example 2), one can see that $d_{\mathcal{N}}(N_1(1), N_2(\mathbb{1}_{2 \times 2})) = 0$. Here $\mathbb{1}_{n \times n}$ denotes the all-ones matrix of size $n \times n$ for any $n \in \mathbb{N}$. However, these two networks are not strongly isomorphic, because they do not even have the same cardinality. Thus to understand the zero sets of $d_{\mathcal{N}}$, we need to search for a different, perhaps weaker notion of isomorphism. This will be further explored in Section 1.6.

Now we state another reformulation of $d_{\mathcal{N}}$ which will be especially useful when proving results about persistent homology.

Definition 8 (Distortion of a map between two networks). Given any $(X, \omega_X), (Y, \omega_Y) \in \mathcal{N}$ and a map $\varphi : (X, \omega_X) \rightarrow (Y, \omega_Y)$, the *distortion* of φ is defined as:

$$\text{dis}(\varphi) := \sup_{x, x' \in X} |\omega_X(x, x') - \omega_Y(\varphi(x), \varphi(x'))|.$$

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

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

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

Proposition 7. Let $(X, \omega_X), (Y, \omega_Y) \in \mathcal{N}$. Then,

$$d_{\mathcal{N}}(X, Y) = \frac{1}{2} \inf \{ \sup \{ \text{dis}(\varphi), \text{dis}(\psi), C_{X,Y}(\varphi, \psi), C_{Y,X}(\psi, \varphi) \} : \varphi : X \rightarrow Y, \psi : Y \rightarrow X \text{ any maps} \}.$$

Remark 8. Proposition 7 is analogous to a result of Kalton and Ostrovskii [75, Theorem 2.1] where—instead of $d_{\mathcal{N}}$ —one has the Gromov-Hausdorff distance between metric spaces. An important remark is that in the Kalton-Ostrovskii formulation, there is only one co-distortion term. When Proposition 7 is applied to metric spaces, the two co-distortion terms become equal by symmetry, and thus the Kalton-Ostrovskii formulation is recovered. But *a priori*, the lack of symmetry in the network setting requires us to consider both terms. We thank Pascal Wild for pointing this out to us in an early manuscript.

The second network distance

Even though the definition of $d_{\mathcal{N}}$ is very general, in some restricted settings it may be convenient to consider a network distance that is easier to formulate. For example, in computational purposes it suffices to assume that we are computing distances between finite networks. Also, a reduction in computational cost is obtained if we restrict ourselves to computing distortions of bijections instead of general correspondences. The next definition arises from such considerations.

Definition 9 (The second network distance). Let $(X, \omega_X), (Y, \omega_Y) \in \mathcal{N}$ be such that $\text{card}(X) = \text{card}(Y)$. Then define:

$$\widehat{d}_{\mathcal{N}}(X, Y) := \frac{1}{2} \inf_{\varphi} \sup_{x, x' \in X} |\omega_X(x, x') - \omega_Y(\varphi(x), \varphi(x'))|,$$

where $\varphi : X \rightarrow Y$ ranges over all bijections from X to Y .

Notice that $\widehat{d}_{\mathcal{N}}(X, Y) = 0$ if and only if $X \cong^s Y$. Also, $\widehat{d}_{\mathcal{N}}$ satisfies symmetry and triangle inequality. It turns out via Example 9 that $d_{\mathcal{N}}$ and $\widehat{d}_{\mathcal{N}}$ agree on networks over two nodes. However, the two notions do not agree in general. In particular, a minimal example where $d_{\mathcal{N}} \neq \widehat{d}_{\mathcal{N}}$ occurs for three node networks, as we show in Remark 10.

Example 9 (Networks with two nodes). Let $(X, \omega_X), (Y, \omega_Y) \in \mathcal{FN}$ where $X = \{x_1, x_2\}$ and $Y = \{y_1, y_2\}$. Then we claim $d_{\mathcal{N}}(X, Y) = \widehat{d}_{\mathcal{N}}(X, Y)$. Furthermore, if $X = N_2 \left(\begin{pmatrix} \alpha & \delta \\ \beta & \gamma \end{pmatrix} \right)$ and $Y = N_2 \left(\begin{pmatrix} \alpha' & \delta' \\ \beta' & \gamma' \end{pmatrix} \right)$, then we have the explicit formula:

$$d_{\mathcal{N}}(X, Y) = \frac{1}{2} \min (\Gamma_1, \Gamma_2), \text{ where}$$

$$\Gamma_1 = \max (|\alpha - \alpha'|, |\beta - \beta'|, |\delta - \delta'|, |\gamma - \gamma'|),$$

$$\Gamma_2 = \max (|\alpha - \gamma'|, |\gamma - \alpha'|, |\delta - \beta'|, |\beta - \delta'|).$$

Details for this calculation are in §2.1.

Remark 10 (A three-node example where $d_{\mathcal{N}} \neq \widehat{d}_{\mathcal{N}}$). Assume (X, ω_X) and (Y, ω_Y) are two networks with the same cardinality. Then

$$d_{\mathcal{N}}(X, Y) \leq \widehat{d}_{\mathcal{N}}(X, Y).$$

The inequality holds because each bijection induces a correspondence, and we are minimizing over all correspondences to obtain $d_{\mathcal{N}}$. However, the inequality may be strict, as demonstrated by the following example. Let $X = \{x_1, \dots, x_3\}$ and let $Y = \{y_1, \dots, y_3\}$. Define $\omega_X(x_1, x_1) = \omega_X(x_3, x_3) = \omega_X(x_1, x_3) = 1, \omega_X = 0$ elsewhere, and define $\omega_Y(y_3, y_3) = 1, \omega_Y = 0$ elsewhere. In terms of matrices, $X = N_3(\Sigma_X)$ and $Y = N_3(\Sigma_Y)$, where

$$\Sigma_X = \begin{pmatrix} 1 & 0 & 1 \\ 0 & 0 & 0 \\ 0 & 0 & 1 \end{pmatrix} \text{ and } \Sigma_Y = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 1 \end{pmatrix}.$$

Define $\Gamma(x, x', y, y') = |\omega_X(x, x') - \omega_Y(y, y')|$ for $x, x' \in X, y, y' \in Y$. Let φ be any bijection. Then we have:

$$\begin{aligned} \max_{x, x' \in X} \Gamma(x, x', \varphi(x), \varphi(x')) &= \max \{ \Gamma(x_1, x_3, \varphi(x_1), \varphi(x_3)), \Gamma(x_1, x_1, \varphi(x_1), \varphi(x_1)), \\ &\quad \Gamma(x_3, x_3, \varphi(x_3), \varphi(x_3)), \Gamma(\varphi^{-1}(y_3), \varphi^{-1}(y_3), y_3, y_3) \} \\ &= 1. \end{aligned}$$

So $\widehat{d}_{\mathcal{N}}(X, Y) = \frac{1}{2}$. On the other hand, consider the correspondence

$$R = \{(x_1, y_3), (x_2, y_2), (x_3, y_3), (x_2, y_1)\}.$$

Then $\max_{(x, y), (x', y') \in R} |\omega_X(x, x') - \omega_Y(y, y')| = 0$. Thus $d_{\mathcal{N}}(X, Y) = 0 < \widehat{d}_{\mathcal{N}}(X, Y)$.

Example 11 (Networks with three nodes). Let $(X, \omega_X), (Y, \omega_Y) \in \mathcal{FN}$, where we write $X = \{x_1, x_2, x_3\}$ and $Y = \{y_1, y_2, y_3\}$. Because we do not necessarily have $d_{\mathcal{N}} = \widehat{d}_{\mathcal{N}}$ on three node networks by Remark 10, the computation of $d_{\mathcal{N}}$ becomes more difficult than in the two node case presented in Example 9. A certain reduction is still possible, which we present next. Consider the following list \mathcal{L} of matrices representing correspondences, where a 1 in position (i, j) means that (x_i, y_j) belongs to the correspondence.

$$\begin{array}{ccccc}
\left(\begin{smallmatrix} 1 & & \\ & 1 & \\ & & 1 \end{smallmatrix} \right) & \left(\begin{smallmatrix} 1 & & \\ & 1 & \\ & & 1 \end{smallmatrix} \right) & \left(\begin{smallmatrix} 1 & & \\ & 1 & \\ & & 1 \end{smallmatrix} \right) & \left(\begin{smallmatrix} 1 & 1 & \\ & 1 & \\ & & 1 \end{smallmatrix} \right) & \left(\begin{smallmatrix} 1 & & \\ & 1 & \\ & & 1 \end{smallmatrix} \right) \\
\hline
\left(\begin{smallmatrix} 1 & 1 & \\ & 1 & \\ & & 1 \end{smallmatrix} \right) & \left(\begin{smallmatrix} 1 & & \\ & 1 & \\ & & 1 \end{smallmatrix} \right) & \left(\begin{smallmatrix} 1 & 1 & \\ & 1 & \\ & & 1 \end{smallmatrix} \right) & \left(\begin{smallmatrix} 1 & 1 & 1 \\ & 1 & \\ & & 1 \end{smallmatrix} \right) & \left(\begin{smallmatrix} 1 & 1 & \\ & 1 & \\ & & 1 \end{smallmatrix} \right) \\
\hline
\left(\begin{smallmatrix} 1 & 1 & 1 \\ & 1 & \\ & & 1 \end{smallmatrix} \right) & \left(\begin{smallmatrix} 1 & 1 & \\ & 1 & \\ & & 1 \end{smallmatrix} \right) & \left(\begin{smallmatrix} 1 & 1 & 1 \\ & 1 & \\ & & 1 \end{smallmatrix} \right) & \left(\begin{smallmatrix} 1 & 1 & 1 \\ & 1 & \\ & & 1 \end{smallmatrix} \right) & \left(\begin{smallmatrix} 1 & 1 & 1 \\ & 1 & \\ & & 1 \end{smallmatrix} \right)
\end{array}$$

Now let $R \in \mathcal{R}(X, Y)$ be any correspondence. Then R contains a correspondence $S \in \mathcal{R}(X, Y)$ such that the matrix form of S is listed in \mathcal{L} . Thus $\text{dis}(R) \geq \text{dis}(S)$, since we are maximizing over a larger set. It follows that $d_{\mathcal{N}}(X, Y)$ is obtained by taking $\arg \min \frac{1}{2} \text{dis}(S)$ over all correspondences $S \in \mathcal{R}(X, Y)$ with matrix forms listed in \mathcal{L} .

For an example of this calculation, let S denote the correspondence

$$S := \{(x_1, y_1), (x_2, y_2), (x_3, y_3)\}$$

represented by the matrix $\left(\begin{smallmatrix} 1 & 1 & \\ & 1 & \\ & & 1 \end{smallmatrix} \right)$. Then $\text{dis}(S)$ is the maximum among the following:

$$\begin{array}{lll}
|\omega_X(x_1, x_1) - \omega_Y(y_1, y_1)| & |\omega_X(x_1, x_2) - \omega_Y(y_1, y_2)| & |\omega_X(x_1, x_3) - \omega_Y(y_1, y_3)| \\
|\omega_X(x_2, x_1) - \omega_Y(y_2, y_1)| & |\omega_X(x_2, x_2) - \omega_Y(y_2, y_2)| & |\omega_X(x_2, x_3) - \omega_Y(y_2, y_3)| \\
|\omega_X(x_3, x_1) - \omega_Y(y_3, y_1)| & |\omega_X(x_3, x_2) - \omega_Y(y_3, y_2)| & |\omega_X(x_3, x_3) - \omega_Y(y_3, y_3)|.
\end{array}$$

The following proposition provides an explicit connection between $d_{\mathcal{N}}$ and $\widehat{d}_{\mathcal{N}}$. We have not defined the $A \cong_I^w B$ notation, but for now it can be interpreted as $d_{\mathcal{N}}(A, B) = 0$ (cf. Definition 25). An illustration is also provided in Figure 1.1.

Proposition 12. *Let $(X, \omega_X), (Y, \omega_Y) \in \mathcal{N}$. Then,*

$$\begin{aligned}
d_{\mathcal{N}}(X, Y) = \inf \{ \widehat{d}_{\mathcal{N}}(X', Y') : X', Y' \in \mathcal{N}, X' \cong_I^w X, Y' \cong_I^w Y, \\
\text{and } \text{card}(X') = \text{card}(Y') \}.
\end{aligned}$$

The moral of the preceding proposition is that networks at 0 $d_{\mathcal{N}}$ -distance live on a fiber above their equivalence class (where equivalence is with respect to $d_{\mathcal{N}}$), and the distance between two networks can always be computed by computing $\widehat{d}_{\mathcal{N}}$ between two representatives having the same number of points. While this notion of picking representatives with the same number of points may seem somewhat mysterious, we refer the reader to Proposition 79 and Figure 1.18 for explicit details on how this process is carried out.

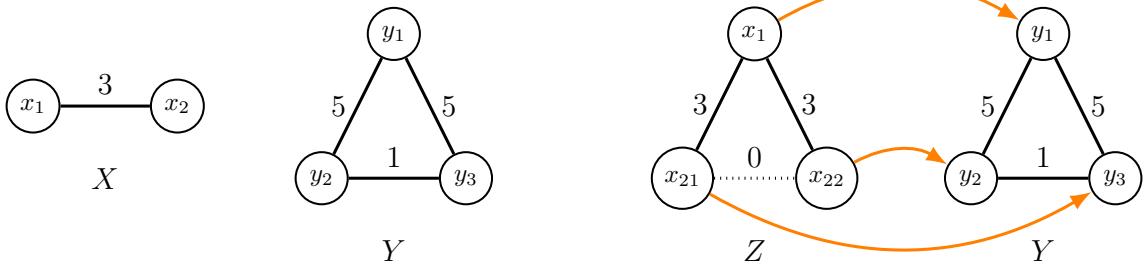


Figure 1.1: The two networks on the left have different cardinalities, but computing correspondences shows that $d_{\mathcal{N}}(X, Y) = 1$. Similarly one computes $d_{\mathcal{N}}(X, Z) = 0$, and thus $d_{\mathcal{N}}(Y, Z) = 1$ by triangle inequality. On the other hand, the bijection given by the arrows shows $\hat{d}_{\mathcal{N}}(Y, Z) = 1$. Applying Proposition 12 then recovers $d_{\mathcal{N}}(X, Y) = 1$.

Remark 13 (Computational aspects of $d_{\mathcal{N}}$ and $\hat{d}_{\mathcal{N}}$). Even though $\hat{d}_{\mathcal{N}}$ has a simpler formulation than $d_{\mathcal{N}}$, computing $\hat{d}_{\mathcal{N}}$ still turns out to be an NP-hard problem, as we discuss in §1.10. Moreover, we show in Theorem 178 that computing $d_{\mathcal{N}}$ is at least as hard as computing $\hat{d}_{\mathcal{N}}$.

Instead of trying to compute $d_{\mathcal{N}}$, we will focus on finding network invariants that can be computed easily. Finding invariants and stability results guaranteeing their validity as proxies for $d_{\mathcal{N}}$ is an overarching goal of this work.

1.3 Network models: the cycle networks and the SBM networks

A *dissimilarity* network is a network (X, A_X) where A_X is a map from $X \times X$ to \mathbb{R}_+ , and $A_X(x, x') = 0$ if and only if $x = x'$. Neither symmetry nor triangle inequality is assumed. We denote the collection of all such networks as $\mathcal{FN}^{\text{dis}}$, $\mathcal{CN}^{\text{dis}}$, and \mathcal{N}^{dis} for the finite, compact, and general settings, respectively.

Example 14. Finite metric spaces and finite ultrametric spaces constitute obvious examples of dissimilarity networks. Recall that, in an ultrametric space (X, d_X) , we have the strong triangle inequality $d_X(x, x') \leq \max \{d_X(x, x''), d_X(x'', x')\}$ for all $x, x', x'' \in X$. More interesting classes of dissimilarity networks arise by relaxing the symmetry and triangle inequality conditions of metric spaces.

Definition 10 (Finite reversibility). The *reversibility* ρ_X of a dissimilarity network (X, A_X) is defined to be the following quantity:

$$\rho_X := \sup_{x \neq x' \in X} \frac{A_X(x, x')}{A_X(x', x)}.$$

(X, A_X) is said to have *finite reversibility* if $\rho_X < \infty$. Notice that $\rho_X \geq 1$, with equality if and only if A_X is symmetric.

Dissimilarity networks satisfying the symmetry condition, but not the triangle inequality, have a long history dating back to Fréchet's thesis [55] and continuing with work by Pitcher and Chittenden [101], Niemytzki [91], Galvin and Shore [57, 58], and many others, as summarized in [67]. One of the interesting directions in this line of work was the development of a “local triangle inequality” and related metrization theorems [91], which has been continued more recently in [122].

Dissimilarity networks satisfying the triangle inequality, but not symmetry, include the special class of objects called *directed metric spaces*, which we define below.

Definition 11. Let (X, A_X) be a dissimilarity network. Given any $x \in X$ and $r \in \mathbb{R}_+$, the *forward-open ball* of radius r centered at x is

$$B^+(x, r) := \{x' \in X : A_X(x, x') < r\}.$$

The *forward-open topology induced by A_X* is the topology on X generated by the collection $\{B^+(x, r) : x \in X, r > 0\}$. The idea of forward open balls is prevalent in the study of Finsler geometry; see [6, p. 149] for details.

Definition 12 (Directed metric spaces). A *directed metric space* or *quasi-metric space* is a dissimilarity network (X, ν_X) such that X is equipped with the forward-open topology induced by ν_X and $\nu_X : X \times X \rightarrow \mathbb{R}_+$ satisfies:

$$\nu_X(x, x'') \leq \nu_X(x, x') + \nu_X(x', x'') \text{ for all } x, x', x'' \in X.$$

The function ν_X is called a directed metric or quasi-metric on X . Notice that compact directed metric spaces constitute a subfamily of $\mathcal{CN}^{\text{dis}}$.

Directed metric spaces with finite reversibility were studied in [108], and constitute important examples of networks that are strictly non-metric. More specifically, the authors of [108] extended notions of Hausdorff distance and Gromov-Hausdorff distance to the setting of directed metric spaces with finite reversibility, and our network distance $d_{\mathcal{N}}$ subsumes this theory while extending it to even more general settings.

Remark 15 (Finsler metrics). An interesting class of directed metric spaces arises from studying Finsler manifolds. A *Finsler manifold* (M, F) is a smooth, connected manifold M equipped with an asymmetric norm F (called a *Finsler function*) defined on each tangent space of M [6]. A Finsler function induces a directed metric $d_F : M \times M \rightarrow \mathbb{R}_+$ as follows: for each $x, x' \in M$,

$$d_F(x, x') := \inf \left\{ \int_a^b F(\gamma(t), \dot{\gamma}(t)) dt : \gamma : [a, b] \rightarrow M \text{ a smooth curve joining } x \text{ and } x' \right\}.$$

Finsler metric spaces have received interest in the applied literature. In [104], the authors prove that Finsler metric spaces with *reversible geodesics* (i.e. the reverse curve $\gamma'(t) := \gamma(1 - t)$ of any geodesic $\gamma : [0, 1] \rightarrow M$ is also a geodesic) is a *weighted quasi-metric* [104, p. 2]. Such objects have been shown to be essential in biological sequence comparison [115].

1.3.1 The directed circles

In this section, we explicitly construct an infinite network in \mathcal{N}^{dis} , and a family of infinite networks in $\mathcal{CN}^{\text{dis}}$.

The general directed circle

First we construct an asymmetric network in \mathcal{N}^{dis} . To motivate this construction, recall from the classification of topological 1-manifolds that *any connected, closed topological 1-manifold is homeomorphic to the circle \mathbb{S}^1* . So as a first construction of a quasi-metric space, it is reasonable to adopt \mathbb{S}^1 as our model and endow it with a quasi-metric weight function.

First define the set $\vec{\mathbb{S}}^1 := \{e^{i\theta} \in \mathbb{C} : \theta \in [0, 2\pi)\}$. For any $\alpha, \beta \in [0, 2\pi)$, define $\vec{d}(\alpha, \beta) := \beta - \alpha \bmod 2\pi$, with the convention $\vec{d}(\alpha, \beta) \in [0, 2\pi)$. Then $\vec{d}(\alpha, \beta)$ is the counterclockwise geodesic distance along the unit circle from $e^{i\alpha}$ to $e^{i\beta}$. As such, it satisfies the triangle inequality and vanishes on a pair $(e^{i\theta_1}, e^{i\theta_2})$ if and only if $\theta_1 = \theta_2$. Next for each $e^{i\theta_1}, e^{i\theta_2} \in \vec{\mathbb{S}}^1$, define

$$\omega_{\vec{\mathbb{S}}^1}(e^{i\theta_1}, e^{i\theta_2}) := \vec{d}(\theta_1, \theta_2).$$

To finish the construction, we specify $\vec{\mathbb{S}}^1$ to have the discrete topology. Clearly this is first countable and makes $\omega_{\vec{\mathbb{S}}^1}$ continuous, but the resulting network will not be compact. Hence it is natural to ask if there exists a coarser topology that we can place on $\vec{\mathbb{S}}^1$.

We claim that a coarser topology does not work to make $(\vec{\mathbb{S}}^1, \omega_{\vec{\mathbb{S}}^1})$ fit the framework of \mathcal{N} . To see why, let $\alpha \in [0, 2\pi)$. Suppose $\omega_{\vec{\mathbb{S}}^1}$ is continuous with respect to some topology on $\vec{\mathbb{S}}^1$, to be determined. Fix $0 < \varepsilon \ll 2\pi$, and define $V := \omega_{\vec{\mathbb{S}}^1}^{-1}((-\varepsilon, \varepsilon))$. Then V is open in the product topology, and in particular contains $(e^{i\alpha}, e^{i\alpha})$. Since V is a union of open rectangles, there exists an open set $U \subseteq \vec{\mathbb{S}}^1$ such that $(e^{i\alpha}, e^{i\alpha}) \in U \times U \subseteq V$. Suppose towards a contradiction that $U \neq \{e^{i\alpha}\}$. Then there exists $e^{i\beta} \in U$, for some $\beta \neq \alpha$. Then $\omega_{\vec{\mathbb{S}}^1}(e^{i\alpha}, e^{i\beta}) \in (0, \varepsilon)$. But by the definition of $\omega_{\vec{\mathbb{S}}^1}$, we must have $\omega_{\vec{\mathbb{S}}^1}(e^{i\beta}, e^{i\alpha}) \in [2\pi - \varepsilon, 2\pi)$, which is a contradiction to having $\omega_{\vec{\mathbb{S}}^1}(U, U) \subseteq (-\varepsilon, \varepsilon)$.

Definition 13. We define the *directed unit circle* to be $(\vec{\mathbb{S}}^1, \omega_{\vec{\mathbb{S}}^1})$ with the discrete topology.

The directed circles with finite reversibility

Now we define a family of directed circles parametrized by reversibility. Unlike the construction in §1.3.1, these directed networks belong to the family $\mathcal{CN}^{\text{dis}}$. An illustration is provided in Figure 1.2.

Recall from §1.3.1 that for $\alpha, \beta \in [0, 2\pi)$, we wrote $\vec{d}(\alpha, \beta)$ to denote the counter-clockwise geodesic distance along the unit circle from $e^{i\alpha}$ to $e^{i\beta}$. Fix $\rho \geq 1$. For each $e^{i\theta_1}, e^{i\theta_2} \in \vec{\mathbb{S}}^1$, define

$$\omega_{\vec{\mathbb{S}}^1, \rho}(e^{i\theta_1}, e^{i\theta_2}) := \min \left(\vec{d}(\theta_1, \theta_2), \rho \vec{d}(\theta_2, \theta_1) \right).$$

In particular, $\omega_{\vec{\mathbb{S}}^1, \rho}$ has reversibility ρ (cf. Definition 10).

Finally, we equip $\vec{\mathbb{S}}^1$ with the standard subspace topology generated by the open balls in \mathbb{C} . In this case, $\vec{\mathbb{S}}^1$ is compact and first countable. It remains to check that $\omega_{\vec{\mathbb{S}}^1, \rho}$ is continuous.

Proposition 16. $\omega_{\vec{\mathbb{S}}^1, \rho} : \vec{\mathbb{S}}^1 \times \vec{\mathbb{S}}^1 \rightarrow \mathbb{R}$ is continuous.

Proof of Proposition 16. It suffices to show that the preimages of basic open sets under $\omega_{\vec{\mathbb{S}}^1, \rho}$ are open. Let (a, b) be an open interval in \mathbb{R} . Let $(e^{i\alpha}, e^{i\beta}) \in \omega_{\vec{\mathbb{S}}^1, \rho}^{-1}[(a, b)]$, where $\alpha, \beta \in [0, 2\pi)$. There are three cases: (1) $\alpha < \beta$, (2) $\beta < \alpha$, or (3) $\alpha = \beta$.

Suppose first that $\alpha < \beta$. There are two subcases: either $\omega_{\vec{\mathbb{S}}^1, \rho}(e^{i\alpha}, e^{i\beta}) = \vec{d}(\alpha, \beta)$, or $= \rho \vec{d}(\beta, \alpha)$.

Fix $r > 0$ to be determined later, but small enough so that $B(\alpha, r) \cap B(\beta, r) = \emptyset$. Let $\gamma \in B(\alpha, r)$, $\delta \in B(\beta, r)$. Then $\vec{d}(\gamma, \delta) \in B(\vec{d}(\alpha, \beta), 2r)$. Also,

$$\left| \rho \vec{d}(\gamma, \delta) - \rho \vec{d}(\alpha, \beta) \right| = \rho \left| \vec{d}(\gamma, \delta) - \vec{d}(\alpha, \beta) \right| < 2r\rho.$$

Now r can be made arbitrarily small, so that for any $\gamma \in B(\alpha, r)$ and any $\delta \in B(\beta, r)$, we have $\omega_{\vec{\mathbb{S}}^1, \rho}(e^{i\gamma}, e^{i\delta}) \in (a, b)$. It follows that $(e^{i\alpha}, e^{i\beta})$ is contained in an open set contained inside $\omega_{\vec{\mathbb{S}}^1, \rho}^{-1}[(a, b)]$. An analogous proof shows this to be true for the $\beta < \alpha$ case.

Next suppose $\alpha = \beta$. Fix $0 < r < b/(2\rho)$. We need to show $\omega_{\vec{\mathbb{S}}^1, \rho}(B(\alpha, r), B(\alpha, r)) \subseteq (a, b)$. Note that $0 \in (a, b)$. Let $\gamma, \delta \in B(\alpha, r)$. There are three subcases. If $\gamma = \delta$, then $\omega_{\vec{\mathbb{S}}^1, \rho}(e^{i\gamma}, e^{i\delta}) = 0 \in (a, b)$. If $\vec{d}(\gamma, \delta) < 2r$, then $\omega_{\vec{\mathbb{S}}^1, \rho}(e^{i\gamma}, e^{i\delta}) < 2r < b$. Finally, suppose $\vec{d}(\gamma, \delta) \geq 2r$. Then we must have $\vec{d}(\delta, \gamma) < 2r$, so $\omega_{\vec{\mathbb{S}}^1, \rho}(e^{i\gamma}, e^{i\delta}) \leq \rho \vec{d}(\delta, \gamma) < 2r\rho < b$. Thus for any $\gamma, \delta \in B(\alpha, r)$, we have $\omega_{\vec{\mathbb{S}}^1, \rho}(e^{i\gamma}, e^{i\delta}) \in (a, b)$.

It follows that $\omega_{\vec{\mathbb{S}}^1, \rho}^{-1}[(a, b)]$ is open. This proves the claim. \square

We summarize the preceding observations in the following:

Definition 14. Let $\rho \in [1, \infty)$. We define the *directed unit circle with reversibility ρ* to be $(\vec{\mathbb{S}}^1, \omega_{\vec{\mathbb{S}}^1, \rho})$. This is a compact, asymmetric network in $\mathcal{CN}^{\text{dis}}$.

This asymmetric network provides us with concrete examples of ε -approximations (Definition 22), for any $\varepsilon > 0$. To see this, fix any $n \in \mathbb{N}$, and consider the *directed circle network on n nodes with reversibility ρ* obtained by writing

$$\vec{\mathbb{S}}_n^1 := \left\{ e^{\frac{2\pi ik}{n}} \in \mathbb{C} : k \in \{0, 1, \dots, n-1\} \right\},$$

and defining $\omega_{n,\rho}$ to be the restriction of $\omega_{\vec{\mathbb{S}}_n^1, \rho}$ on this set. The pair $(\vec{\mathbb{S}}_n^1, \omega_{n,\rho})$ is the network thus obtained. An illustration of $\vec{\mathbb{S}}^1$ and $\vec{\mathbb{S}}_n^1$ for $n = 6$ is provided in Figure 1.2.

Theorem 17. *As $n \rightarrow \infty$, the sequence of finite dissimilarity networks $(\vec{\mathbb{S}}_n^1, \omega_{n,\rho})$ limits to the dissimilarity network $(\vec{\mathbb{S}}^1, \omega_{\vec{\mathbb{S}}^1, \rho})$ in the sense of $d_{\mathcal{N}}$.*

Proof of Theorem 17. Let $\varepsilon > 0$, and let $n \in \mathbb{N}$ be such that $2\pi/n < \varepsilon$. It suffices to show that $d_{\mathcal{N}}((\vec{\mathbb{S}}^1, \omega_{\vec{\mathbb{S}}^1, \rho}), (\vec{\mathbb{S}}_n^1, \omega_{n,\rho})) < \varepsilon$. Define a correspondence between $\vec{\mathbb{S}}^1$ and $\vec{\mathbb{S}}_n^1$ as follows:

$$R := \left\{ (e^{i\theta}, e^{\frac{2\pi ik}{n}}) : \theta \in \left(\frac{2\pi k}{n} - \frac{\varepsilon}{\rho}, \frac{2\pi k}{n} + \varepsilon\right), k \in \{0, 1, 2, \dots, n-1\} \right\}.$$

Essentially this is the same as taking ε -balls around each $e^{\frac{2\pi ik}{n}}$, except that the reversibility parameter skews one side of the ε -ball. Next let $0 \leq \theta_1 \leq \theta_2 < 2\pi$, and let $j, k \in \{0, 1, \dots, n-1\}$ be such that $\theta_1 \in (\frac{2\pi j}{n} - \frac{\varepsilon}{\rho}, \frac{2\pi j}{n} + \varepsilon)$ and $\theta_2 \in (\frac{2\pi k}{n} - \frac{\varepsilon}{\rho}, \frac{2\pi k}{n} + \varepsilon)$.

Suppose first that $k = j$. Then

$$\min(\vec{d}(\theta_1, \theta_2), \vec{d}(\theta_2, \theta_1)) \leq \varepsilon + \frac{\varepsilon}{\rho}, \text{ and } \max(\vec{d}(\theta_1, \theta_2), \vec{d}(\theta_2, \theta_1)) \leq \rho(\varepsilon + \frac{\varepsilon}{\rho}) = \rho\varepsilon + \varepsilon.$$

As $\varepsilon \rightarrow 0$, this quantity tends to zero, so $\omega_{\vec{\mathbb{S}}_n^1, \rho}(e^{i\theta_1}, e^{i\theta_2}) \rightarrow 0$. The other cases follow from similar observations; the key idea is that as $\varepsilon \rightarrow 0$, the $\omega_{\vec{\mathbb{S}}_n^1, \rho}$ value between any two points on a “skewed ε -ball” also tends to 0. \square

Remark 18. Finite reversibility is critical when defining directed circles on n nodes. Without this condition, correspondences as above lead to terms like the following:

$$\max \left(|\omega_{\vec{\mathbb{S}}^1}(e^{i\theta_1}, e^{i\theta_2}) - \omega_{\vec{\mathbb{S}}^1}(e^{\frac{2\pi ik}{n}}, e^{\frac{2\pi ik}{n}})|, |\omega_{\vec{\mathbb{S}}^1}(e^{i\theta_2}, e^{i\theta_1}) - \omega_{\vec{\mathbb{S}}^1}(e^{\frac{2\pi ik}{n}}, e^{\frac{2\pi ik}{n}})| \right) \approx 2\pi.$$

The problem here is that as $\omega_{\vec{\mathbb{S}}^1}(e^{i\theta_1}, e^{i\theta_2}) \rightarrow 0$, we adversely have $\omega_{\vec{\mathbb{S}}^1}(e^{i\theta_2}, e^{i\theta_1}) \rightarrow 2\pi$. Indeed, one of our later results (cf. §1.8) shows that \mathcal{CN} is complete. Because $(\vec{\mathbb{S}}^1, \omega_{\vec{\mathbb{S}}^1}) \notin \mathcal{CN}$, it follows that there cannot a sequence of finite networks converging to $(\vec{\mathbb{S}}^1, \omega_{\vec{\mathbb{S}}^1})$.

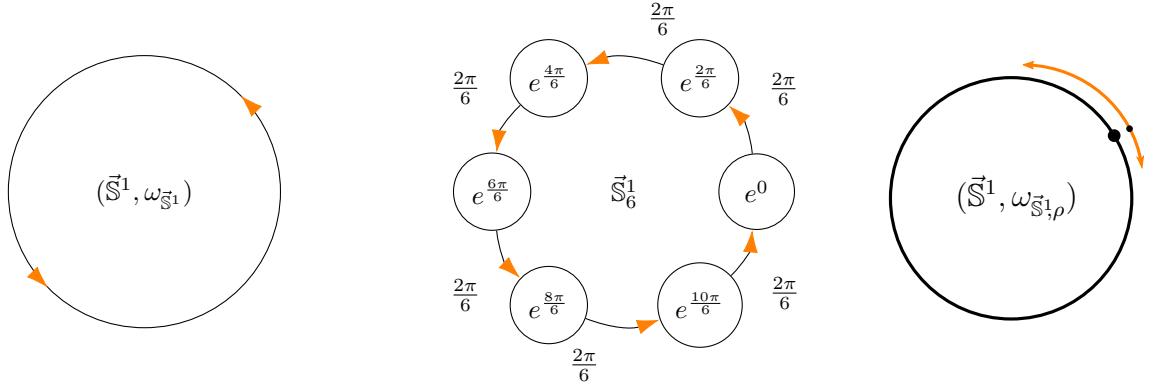


Figure 1.2: The directed circle $(\vec{S}^1, \omega_{\vec{S}^1})$, the directed circle on 6 nodes $(\vec{S}_6^1, \omega_{\vec{S}_6^1})$, and the directed circle with reversibility ρ , for some $\rho \in [1, \infty)$. Traveling in a clockwise direction is possibly only in the directed circle with reversibility ρ , but this incurs a penalty modulated by ρ .

Remark 19 (Directed circle with finite reversibility—forward-open topology version). Instead of using the subspace topology generated by the standard topology on \mathbb{C} , we can also endow $(\vec{S}^1, \omega_{\vec{S}^1, \rho})$ with the forward-open topology generated by $\omega_{\vec{S}^1, \rho}$. The open balls in this topology are precisely the open balls in the subspace topology induced by the standard topology, the only adjustment being the “center” of each ball. The directed metric space $(\vec{S}^1, \omega_{\vec{S}^1, \rho})$ equipped with the forward-open topology is another example of a compact, asymmetric network in $\mathcal{CN}^{\text{dis}}$.

1.3.2 The finite cycle networks

For each $n \in \mathbb{N}$, let (X_n, E_n, W_{E_n}) denote the weighted graph with vertex set $X_n := \{x_1, x_2, \dots, x_n\}$, edge set $E_n := \{(x_1, x_2), (x_2, x_3), \dots, (x_{n-1}, x_n), (x_n, x_1)\}$, and edge weights $W_{E_n} : E_n \rightarrow \mathbb{R}$ given by writing $W_{E_n}(e) = 1$ for each $e \in E_n$. Next let $\omega_{G_n} : X_n \times X_n \rightarrow \mathbb{R}$ denote the shortest path distance induced on $X_n \times X_n$ by W_{E_n} . Then we write $G_n := (X_n, \omega_{G_n})$ to denote the network with node set X_n and weights given by ω_{G_n} . Note that $\omega_{G_n}(x, x) = 0$ for each $x \in X_n$. See Figure 1.3 for an example.

We say that G_n is the *cycle network of length n*. Cycle networks are highly asymmetric because for every consecutive pair of nodes (x_i, x_{i+1}) in a graph G_n , where $1 \leq i \bmod(n) \leq n$, we have $\omega_{G_n}(x_i, x_{i+1}) = 1$, whereas $\omega_{G_n}(x_{i+1}, x_i) = \text{diam}(G_n) = n - 1$, which is much larger than 1 when n is large.

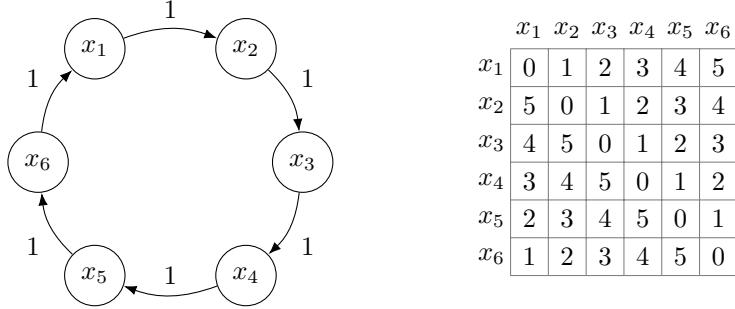


Figure 1.3: A cycle network on 6 nodes, along with its weight matrix. Note that the weights are highly asymmetric.

1.3.3 The network stochastic block model

We now describe a generative model for random networks, based on the popular stochastic block model for sampling random graphs [2]. The current network SBM model we describe is a composition of Gaussian distributions. However, the construction can be adjusted easily to work with other distributions.

Fix a number of *communities* $N \in \mathbb{N}$. For $1 \leq i, j \leq N$, fix a mean μ_{ij} and a variance σ_{ij}^2 . This collection $\mathcal{G} := \{\mathcal{N}(\mu_{ij}, \sigma_{ij}^2) : 1 \leq i, j \leq N\}$ of N^2 independent Gaussian distributions comprise the network SBM.

To sample a random network (X, ω_X) of n nodes from this SBM, start by fixing $n_i \in \mathbb{N}, 1 \leq i \leq N$ such that $\sum_i n_i = n$. For $1 \leq i \leq N$, let X_i be a set with n_i points. Define $X := \cup_{i=1}^n X_i$. Next sample each node weight as $\omega_X(x, x') \sim \mathcal{N}(\mu_{ij}, \sigma_{ij}^2)$, where $x \in X_i$ and $x' \in X_j$. An illustration of this process is provided in Figure 1.4.

A minimalistic way to obtain a measure network from the preceding construction is to equip the pair (X, ω_X) with the uniform measure μ_X that assigns a mass of $1/n$ to each point.

The justification for defining a network SBM this ways comes from the understanding of ε -systems that we develop in §1.6.1.

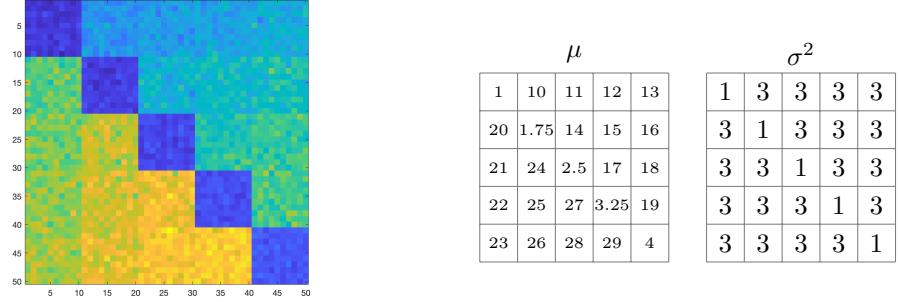


Figure 1.4: A network SBM on 50 nodes, split into 5 communities, along with the matrices of means and variances. The deepest blue corresponds to values ≈ 1 , and the deepest yellow corresponds to values ≈ 29 .

1.4 Persistent homology on networks: Simplicial constructions

Prior to developing the notion of persistent path homology, we first explored related ideas in the simplicial setting, especially with regards to capturing information from directed networks. We explain these contributions in the current section. An overview is presented in Figure 1.5.

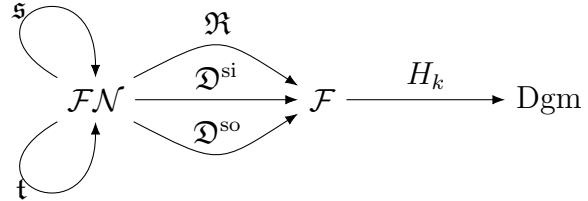


Figure 1.5: A schematic of some of the simplicial constructions on directed networks. \mathcal{F} is the collection of filtered simplicial complexes. Dgm is the collection of persistence diagrams. We study the Rips (\mathfrak{R}) and Dowker (\mathfrak{D}^{si} , \mathfrak{D}^{so}) filtrations, each of which takes a network as input and produces a filtered simplicial complex. \mathfrak{s} and \mathfrak{t} denote the network transformations of symmetrization (replacing a pair of weights between two nodes by the maximum weight) and transposition (swapping the weights between pairs of nodes). \mathfrak{R} is insensitive to both \mathfrak{s} and \mathfrak{t} . But $\mathfrak{D}^{\text{si}} \circ \mathfrak{t} = \mathfrak{D}^{\text{so}}$, $\mathfrak{D}^{\text{so}} \circ \mathfrak{t} = \mathfrak{D}^{\text{si}}$, and in general, \mathfrak{D}^{si} and \mathfrak{D}^{so} are *not* invariant under \mathfrak{t} (Theorem 50).

1.4.1 Background on persistent homology

Homology is a classical construction which assigns a k -dimensional signature to a topological space, where k ranges over the nonnegative integers. When the space is a simplicial complex, the resulting homology theory is called simplicial homology. In practice, simplicial homology is readily computable via matrix operations.

Datasets “in the wild” are typically discrete objects. More specifically, our measurement and recording technologies are discrete, and therefore the datasets that we curate from a process are necessarily discrete. A priori, these datasets are equipped with the uninteresting discrete topology. However, there are several well-studied [47] methods for imposing an artificial topology on a discrete dataset. Here are two examples.

Example 20 (Vietoris-Rips complexes). Let (X, d_X) be a metric space. Given a scale parameter $\delta \geq 0$, the *Vietoris-Rips complex at scale δ* is

$$\text{VR}_\delta(X) := \{\sigma \subseteq X : \sigma \text{ finite, nonempty}, \max_{x, x' \in \sigma} d_X(x, x') \leq \delta\}.$$

Example 21 (Čech complexes). Let (X, d_X) be a metric space. Given a scale parameter $\delta \geq 0$, the *Čech complex at scale δ* is

$$\check{C}_\delta(X) := \{\sigma \subseteq X : \sigma \text{ finite, nonempty}, \exists p \in X \text{ such that } \max_{x \in \sigma} d_X(x, p) \leq \delta\}.$$

Both of these constructions are accompanied by interesting theorems. For Vietoris-Rips complexes, the following results are known (see [3] for more details).

Theorem 22 (Hausmann’s theorem, [69]). *Let M be a compact Riemannian manifold. Then for sufficiently small $r > 0$, we have the homotopy equivalence $M \simeq \text{VR}_r(M)$.*

Theorem 23 (Latschev’s theorem, [79]). *Let M be a compact Riemannian manifold. Then for sufficiently small $r, \eta > 0$, we have $M \simeq \text{VR}_r(X)$ whenever X is a metric space with $d_{GH}(X, M) < \eta$. In particular, X can be a sufficiently dense, finite sample of M .*

The Čech complex of a metric space X at scale δ coincides with the *nerve* simplicial complex when we take a cover of X by δ -balls:

Definition 15 (Nerve of a cover). Let X be a topological space, and let $\mathcal{A} = \{A_i\}_{i \in I}$ be an open cover of X indexed by I . The *nerve* of \mathcal{A} is the simplicial complex $\mathcal{N}(\mathcal{A}) := \{\sigma \in \text{pow}(I) : \sigma \text{ is finite, nonempty, and } \cap_{i \in \sigma} A_i \neq \emptyset\}$.

Theorem 24 (Nerve theorem [68] Corollary 4G.3). *Let X be a paracompact space (every open cover admits a locally finite open refinement), and let \mathcal{A} be an open cover such that every nonempty, finite intersection of sets in \mathcal{A} is contractible. Then $X \simeq |\mathcal{N}(\mathcal{A})|$.*

Returning to the topic of imposing an artificial topology on a dataset via one of these constructions, the following question comes to mind: what is the “correct” scale parameter to use when defining either the Vietoris-Rips or the Čech complexes? The theory of persistent homology (PH) enables the user to bypass this consideration and instead view the homological signatures at a range of scale parameters, along with information about how signatures from one resolution “include” into the signatures at another resolution [56, 103, 49, 123]. The essential idea is to fix a method for “topologizing” a dataset (e.g. the Vietoris-Rips or Čech constructions), choose a collection of scale parameters $0 \leq \delta_0 < \delta_1 < \dots < \delta_n$ (e.g. choose all the scales at which new simplices are added), and then apply the (simplicial) homology functor with coefficients in a field. The nested simplicial complexes and their inclusion maps

$$\dots \hookrightarrow K_{\delta_i}(X) \hookrightarrow K_{\delta_{i+1}}(X) \hookrightarrow K_{\delta_{i+2}}(X) \hookrightarrow \dots$$

form a sequence of vector spaces with linear maps:

$$\dots \rightarrow H_\bullet(K_{\delta_i}(X)) \rightarrow H_\bullet(K_{\delta_{i+1}}(X)) \rightarrow H_\bullet(K_{\delta_{i+2}}(X)) \rightarrow \dots$$

Here the H_\bullet denotes homology in a given dimension \bullet ranging over \mathbb{Z}_+ . The entire collection $\{K_\delta(X) \hookrightarrow K_{\delta'}(X)\}_{\delta \leq \delta'}$ is known as a *simplicial filtration* or a *filtered simplicial complex*, and the collection of vector spaces with linear maps is a *persistent vector space*. More precisely, we have the following definition:

Definition 16. A *persistent vector space* \mathcal{V} is a family $\{V^\delta \xrightarrow{\nu_{\delta,\delta'}} V^{\delta'}\}_{\delta \leq \delta' \in \mathbb{R}}$ of vector spaces and linear maps such that: (1) $\nu_{\delta,\delta}$ is the identity map for any $\delta \in \mathbb{R}$, and (2) $\nu_{\delta,\delta''} = \nu_{\delta',\delta''} \circ \nu_{\delta,\delta'}$ whenever $\delta \leq \delta' \leq \delta''$.

A classification result in [21, §5.2] shows that at least in “nice” settings, a certain object called a *persistence diagram/barcode* ([26]) is a full invariant of a persistent vector space. When it is well-defined, a persistence diagram is essentially a list of the topological signatures of the dataset along with the ranges of scale parameters along which each signature *persists*. In typical data analysis use cases, the barcode or diagram is simply the output of applying persistent homology to a dataset.

Persistent homology computations are matrix operations and hence computable, and there are numerous software packages currently available for efficient PH computations. PH computation is theoretically justified by certain *stability theorems*, one of which is the following (an early version for Vietoris-Rips filtrations on finite metric spaces appeared in [25]):

Theorem 25 (PH stability for metric spaces, [27] Theorem 5.2). *Let X, Y be two totally bounded metric spaces, and let $k \in \mathbb{Z}_+$. Let Dgm_k^\bullet denote the k -dimensional persistence diagram of either the Vietoris-Rips or Čech filtrations. Then we have:*

$$d_B(\text{Dgm}_k^\bullet(X), \text{Dgm}_k^\bullet(Y)) \leq 2d_{GH}(X, Y).$$

Here d_B is a certain metric on persistence diagrams called the *bottleneck distance*. It is essentially a matching metric that can be computed via the Hungarian algorithm. Stability results of this form show that PH outputs change in a controlled way when the input dataset is perturbed. This provides the theoretical justification for using PH in data analysis.

Finally, we introduce another definition that will be of use to us: the *interleaving distance* d_I is an extended pseudometric between persistent vector spaces (cf. §3.1.2). Even when persistent vector spaces do not have well-defined persistence diagrams, we can refer to the interleaving distance between the persistent vector spaces.

1.4.2 Related literature on persistent homology of directed networks

The overview of persistent homology described in the previous section is limited to methods that accept metric spaces as input datasets. Some of the extant “network” PH literature considers graph datasets that actually satisfy metric properties [80, 76], and these fit into the metric pipeline described above. Some more general approaches for obtaining persistence diagrams from networks are followed in [71, 23, 60, 99]. In all these cases, the networks are required to be symmetric.

It was pointed out in [27] that Vietoris-Rips and Čech complexes could be defined for dissimilarity spaces satisfying the symmetry property (but relaxing the other properties of a metric space). Another key contribution of this paper was in showing that persistence diagrams were well defined not just for finite metric spaces, but also totally bounded metric spaces (in particular, *compact* metric spaces). A notable contribution for finite directed networks was made in [118], where Turner considered several generalizations of Vietoris-Rips complexes to the directed setting and proved their stability with respect to the finite version of d_N (while d_N was referred to as a “correspondence distance” in [118], we adhere to the d_N formulation that had already appeared in [22]). In particular, Turner considered *ordered tuple complexes*, which are morally quite different from extant simplicial constructions. These OT complexes, also known as directed Rips/flag complexes, had been used in the non-persistent setting in [102]. An efficient implementation of persistent homology using directed flag complexes has also been developed recently by Lütgehetmann [83].

As stated earlier, one of the primary goals motivating this thesis was to develop the notion of persistent path homology (PPH). Independently, we also studied a particular simplicial construction called the *Dowker complex*. This complex had already appeared in the symmetric setting in [27] with a different motivation, but in [37] we studied its behavior thoroughly in the setting of directed networks, where it proved to be quite powerful. In particular, we developed several experiments where Dowker persistence performed significantly than its directed Vietoris-Rips counterpart. Regardless, PPH still appears to be truly sensitive to asymmetry in more ways than even the Dowker complex, and thus seems to be a natural candidate when studying directed networks.

Interestingly, based on our work in [37] and [40], the Dowker complex is a directed analogue of the Čech complex, and there is evidence to suggest that PPH is the Čech

analogue of the directed Rips/flag/OT complex. In this way, our work complements the contributions of [118].

We make a final remark to situate our work in the existing literature. In a 2018 update to [118], Turner asks if any of the directed generalizations of the Vietoris-Rips complex hold in the setting of infinite networks (more precisely, the question is about infinite “set-function” pairs, which are just networks in our terminology). The answer is “yes”: as we had already shown in [34], by the framework we develop for infinite networks, and in particular by our notion of ε -systems, all of these directed generalizations of the Vietoris-Rips and Čech complex constructions are well-defined for compact (in particular, infinite) networks.

1.4.3 The Vietoris-Rips filtration of a network

Following the definition for metric spaces, we define the Vietoris-Rips complex for a network $(X, \omega_X) \in \mathcal{N}$ as follows:

$$\mathfrak{R}_X^\delta := \{\sigma \text{ finite } \in \text{pow}(X) : \max_{x, x' \in \sigma} \omega_X(x, x') \leq \delta\}.$$

To any network (X, ω_X) , we may associate the *Vietoris-Rips filtration* $\{\mathfrak{R}_X^\delta \hookrightarrow \mathfrak{R}_X^{\delta'}\}_{\delta \leq \delta'}$. We denote the k -dimensional persistence vector space associated to this filtration by $\mathbf{PVec}_k^{\mathfrak{R}}(X)$.

It is not at all clear that the corresponding persistence diagram $\text{Dgm}_k^{\mathfrak{R}}(X)$ is well-defined in general, although it is well-defined when $(X, \omega_X) \in \mathcal{FN}$ (the finite case is easy to see). The fact that the persistence diagram is well-defined when $(X, \omega_X) \in \mathcal{CN}$ is presented in Theorem 86, and is a consequence of the machinery we develop for $d_{\mathcal{N}}$.

The Vietoris-Rips persistence diagram, when defined, is stable to small perturbations of the input data:

Proposition 26. *Let $(X, \omega_X), (Y, \omega_Y) \in \mathcal{CN}$, and let $k \in \mathbb{Z}_+$. Then*

$$d_1(\mathbf{PVec}_k^{\mathfrak{R}}(X), \mathbf{PVec}_k^{\mathfrak{R}}(Y)) \leq 2d_{\mathcal{N}}(X, Y).$$

We omit the proof because it is similar to that of Proposition 29, which we will prove in detail. We also remark that we obtain stability for persistence diagrams with respect to the bottleneck distance in Corollary 177. More specifically, Corollary 177 states that we have:

$$d_B(\text{Dgm}_k^{\mathfrak{R}}(X), \text{Dgm}_k^{\mathfrak{R}}(Y)) \leq 2d_{\mathcal{N}}(X, Y).$$

Remark 27. The preceding result serves a dual purpose: (1) it shows that the Vietoris-Rips persistence diagram is robust to noise in input data, and (2) it shows that instead of computing the network distance between two networks, one can compute the bottleneck distance between their Vietoris-Rips persistence diagrams as a suitable proxy. The advantage to computing bottleneck distance is that it can be done in polynomial time (see [52]), whereas computing $d_{\mathcal{N}}$ is NP-hard in general. We remind the reader that the problem of computing $d_{\mathcal{N}}$ includes the problem of computing the Gromov-Hausdorff distance between finite

metric spaces, which is an NP-hard problem [105]. We remark that the idea of computing Vietoris-Rips persistence diagrams to compare finite metric spaces first appeared in [25], and moreover, that Proposition 26 is an extension of Theorem 3.1 in [25].

The Vietoris-Rips filtration in the setting of symmetric networks has been used in [71, 23, 60, 99], albeit without addressing stability results.

We now introduce a definition that will help us gauge the performance of various PH methods on directed networks.

Definition 17 (Symmetrization and Transposition). Define the *max-symmetrization* map $\mathfrak{s} : \mathcal{N} \rightarrow \mathcal{N}$ by $(X, \omega_X) \mapsto (X, \widehat{\omega}_X)$, where for any network (X, ω_X) , we define $\widehat{\omega}_X : X \times X \rightarrow \mathbb{R}$ as follows:

$$\widehat{\omega}_X(x, x') := \max(\omega_X(x, x'), \omega_X(x', x)), \text{ for } x, x' \in X.$$

Also define the *transposition* map $\mathfrak{t} : \mathcal{N} \rightarrow \mathcal{N}$ by $(X, \omega_X) \mapsto (X, \omega_X^\top)$, where for any $(X, \omega_X) \in \mathcal{N}$, we define $\omega_X^\top(x, x') := \omega_X(x', x)$ for $x, x' \in X$. For convenience, we denote $X^\top := \mathfrak{t}(X)$ for any network X .

Remark 28 (Vietoris-Rips is insensitive to asymmetry). A critical weakness of the Vietoris-Rips complex construction is that it is not sensitive to asymmetry. To see this, consider the symmetrization map \mathfrak{s} defined in Definition 17, and let $(X, \omega_X) \in \mathcal{FN}$. Now for any $\sigma \in \text{pow}(X)$, we have $\max_{x, x' \in \sigma} \omega_X(x, x') = \max_{x, x' \in \sigma} \widehat{\omega}_X(x, x')$. It follows that for each $\delta \geq 0$, the Rips complexes of (X, ω_X) and $(X, \widehat{\omega}_X) = \mathfrak{s}(X, \omega_X)$ are equal, i.e. $\mathfrak{R} = \mathfrak{R} \circ \mathfrak{s}$. Thus the Rips persistence diagrams of the original and max-symmetrized networks are equal.

1.4.4 The Dowker filtration of a network

The Dowker complexes of a network, as defined below, comprise a natural generalization of the Čech complex for metric spaces. In the network setting, the lack of symmetry causes the Čech complex to decouple into a *sink* complex and a *source* complex. For historical reasons that we explain in §1.4.5, these complexes are called the Dowker complexes.

Given $(X, \omega_X) \in \mathcal{N}$, and for any $\delta \in \mathbb{R}$, consider the following relation on X :

$$R_{\delta, X} := \{(x, x') : \omega_X(x, x') \leq \delta\}. \quad (1.1)$$

Then $R_{\delta, X} \subseteq X \times X$, and for any $\delta' \geq \delta$, we have $R_{\delta, X} \subseteq R_{\delta', X}$. Using $R_{\delta, X}$, we build a simplicial complex $\mathfrak{D}_{\delta}^{\text{si}}$ as follows:

$$\mathfrak{D}_{\delta, X}^{\text{si}} := \{\sigma = [x_0, \dots, x_n] : \text{there exists } x' \in X \text{ such that } (x_i, x') \in R_{\delta, X} \text{ for each } x_i\}. \quad (1.2)$$

If $\sigma \in \mathfrak{D}_{\delta, X}^{\text{si}}$, it is clear that any face of σ also belongs to $\mathfrak{D}_{\delta, X}^{\text{si}}$. We call $\mathfrak{D}_{\delta, X}^{\text{si}}$ the *Dowker δ -sink simplicial complex* associated to X , and refer to x' as a δ -sink for σ (where σ and x' should be clear from context).

Since $R_{\delta,X}$ is an increasing sequence of sets, it follows that $\mathfrak{D}_{\delta,X}^{\text{si}}$ is an increasing sequence of simplicial complexes. In particular, for $\delta' \geq \delta$, there is a natural inclusion map $\mathfrak{D}_{\delta,X}^{\text{si}} \hookrightarrow \mathfrak{D}_{\delta',X}^{\text{si}}$. We write $\mathfrak{D}_X^{\text{si}}$ to denote the filtration $\{\mathfrak{D}_{\delta,X}^{\text{si}} \hookrightarrow \mathfrak{D}_{\delta',X}^{\text{si}}\}_{\delta \leq \delta'}$ associated to X . We call this the *Dowker sink filtration* on X . The corresponding persistent vector space is denoted $\text{PVec}_k^{\text{si}}(X)$. When it is defined, we will denote the k -dimensional persistence diagram arising from this filtration by $\text{Dgm}_k^{\text{si}}(X)$. Once again, we point the reader to Theorem 86, where we show that this diagram is well-defined when $(X, \omega_X) \in \mathcal{CN}$. The case $(X, \omega_X) \in \mathcal{FN}$ is well-defined for easy reasons, and the reader may keep the finite case in mind for now.

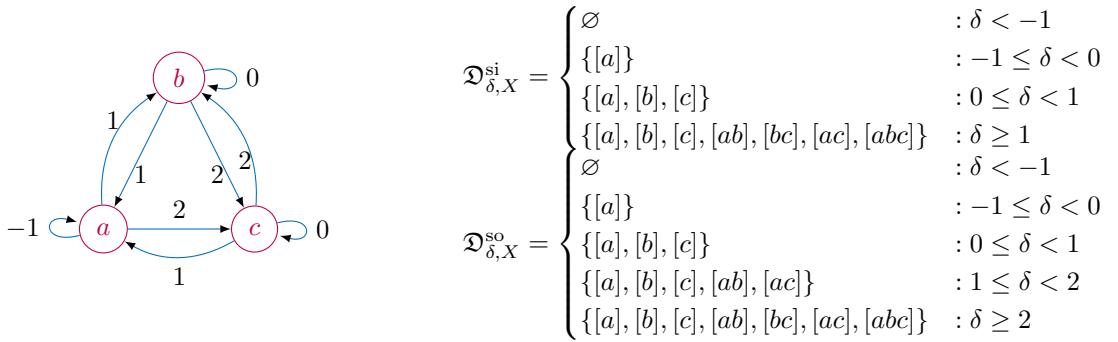


Figure 1.6: Computing the Dowker sink and source complexes of a network (X, ω_X) . Observe that the sink and source complexes are different in the range $1 \leq \delta < 2$.

Practitioners of persistent homology might recall that there are *two* Dowker complexes [59, p. 73]. One of these is the sink complex defined above. We define its dual below:

$$\mathfrak{D}_{\delta,X}^{\text{so}} := \{\sigma = [x_0, \dots, x_n] : \text{there exists } x' \in X \text{ such that } (x', x_i) \in R_{\delta,X} \text{ for each } x_i\}. \quad (1.3)$$

We call $\mathfrak{D}_{\delta,X}^{\text{so}}$ the *Dowker δ -source simplicial complex* associated to X . The filtration $\{\mathfrak{D}_{\delta,X}^{\text{so}} \hookrightarrow \mathfrak{D}_{\delta',X}^{\text{so}}\}_{\delta \leq \delta'}$ associated to X is called the *Dowker source filtration*, denoted $\mathfrak{D}_X^{\text{so}}$. We denote the k -dimensional persistence diagram (when defined) arising from this filtration by $\text{Dgm}_k^{\text{so}}(X)$. Notice that any construction using $\mathfrak{D}_{\delta,X}^{\text{si}}$ can also be repeated using $\mathfrak{D}_{\delta,X}^{\text{so}}$, so we focus on the case of the sink complexes and restate results in terms of source complexes where necessary. A subtle point to note here is that each of these Dowker complexes can be used to construct a persistence diagram. A folklore result in the literature about persistent homology of metric spaces, known as *Dowker duality*, is that the two persistence diagrams arising this way are equal [27, Remark 4.8]:

$$\text{Dgm}_k^{\text{si}}(X) = \text{Dgm}_k^{\text{so}}(X) \text{ for any } k \in \mathbb{Z}_+,$$

Thus it makes sense to talk about “the” Dowker diagram associated to X . In particular, in §1.4.5 we describe a stronger result—a functorial Dowker theorem—from which the duality follows easily in the general setting of networks.

The sink and source filtrations are not equal in general; this is illustrated in Figure 1.6.

As in the case of the Rips filtration, both the Dowker sink and source filtrations are stable.

Proposition 29. *Let $(X, \omega_X), (Y, \omega_Y) \in \mathcal{CN}$. Then*

$$d_I(\mathbf{PVec}_k^\bullet(X), \mathbf{PVec}_k^\bullet(Y)) \leq 2d_N(X, Y).$$

Here \mathbf{PVec}^\bullet refers to either of $\mathbf{PVec}^{\text{si}}$ and $\mathbf{PVec}^{\text{so}}$.

Once again, we obtain stability for persistence diagrams with respect to the bottleneck distance in Corollary 177. More specifically, Corollary 177 states that we have:

$$d_B(\mathbf{Dgm}_k^\bullet(X), \mathbf{Dgm}_k^\bullet(Y)) \leq 2d_N(X, Y).$$

Remark 30. The preceding result shows that the Dowker persistence diagram is robust to noise in input data, and that the bottleneck distance between Dowker persistence diagrams arising from two networks can be used as a proxy for computing the actual network distance. Note the analogy with Remark 27.

Both the Dowker and Rips filtrations are valid methods for computing persistent homology of networks, by virtue of their stability results (Propositions 26 and 29). However, we present the Dowker filtration as an appropriate method for capturing directionality information in directed networks. In §1.4.6 we discuss this particular feature of the Dowker filtration in full detail.

Remark 31 (Symmetric networks). In the setting of symmetric networks, the Dowker sink and source simplicial filtrations coincide, and so we automatically obtain $\mathbf{Dgm}_k^{\text{so}}(X) = \mathbf{Dgm}_k^{\text{si}}(X)$ for any $k \in \mathbb{Z}_+$ and any $(X, \omega_X) \in \mathcal{CN}$.

Remark 32 (The metric space setting and relation to witness complexes). When restricted to the setting of metric spaces, the Dowker complex resembles a construction called the witness complex [44]. In particular, a version of the Dowker complex for metric spaces, constructed in terms of *landmarks* and *witnesses*, was discussed in [27], along with stability results. When restricted to the special networks that are pseudo-metric spaces, our definitions and results agree with those presented in [27].

1.4.5 A Functorial Dowker Theorem

We now abstract some of the definitions presented above. Let X, Y be two totally ordered sets, and let $R \subseteq X \times Y$ be a nonempty relation. Then one defines two simplicial

complexes E_R and F_R as follows. A finite subset $\sigma \subseteq X$ belongs to E_R whenever there exists $y \in Y$ such that $(x, y) \in R$ for each $x \in \sigma$. Similarly a finite subset $\tau \subseteq Y$ belongs to F_R whenever there exists $x \in X$ such that $(x, y) \in R$ for each $y \in \tau$. These constructions can be traced back to [46], who proved the following result that we refer to as *Dowker's theorem*:

Theorem 33 (Dowker's theorem; Theorem 1a, [46]). *Let X, Y be two totally ordered sets, let $R \subseteq X \times Y$ be a nonempty relation, and let E_R, F_R be as above. Then for each $k \in \mathbb{Z}_+$,*

$$H_k(E_R) \cong H_k(F_R).$$

There is also a strong form of Dowker's theorem that Björner proves via the classical *nerve theorem* [11, Theorems 10.6, 10.9]. Below we write $|X|$ to denote the geometric realization of a simplicial complex X .

Theorem 34 (The strong form of Dowker's theorem; Theorem 10.9, [11]). *Under the assumptions of Theorem 33, we in fact have $|E_R| \simeq |F_R|$.*

The Functorial Dowker Theorem is the following generalization of the strong form of Dowker's theorem: instead of a single nonempty relation $R \subseteq X \times Y$, consider any pair of nested, nonempty relations $R \subseteq R' \subseteq X \times Y$. Then there exist homotopy equivalences between the geometric realizations of the corresponding complexes that commute with the canonical inclusions, up to homotopy. We formalize this statement below.

Theorem 35 (The Functorial Dowker Theorem (FDT)). *Let X, Y be two totally ordered sets, let $R \subseteq R' \subseteq X \times Y$ be two nonempty relations, and let $E_R, F_R, E_{R'}, F_{R'}$ be their associated simplicial complexes. Then there exist homotopy equivalences $\Gamma_{|E_R|} : |F_R| \rightarrow |E_R|$ and $\Gamma_{|E_{R'}|} : |F_{R'}| \rightarrow |E_{R'}|$ such that the following diagram commutes up to homotopy:*

$$\begin{array}{ccc} |F_R| & \xrightarrow{|\iota_E|} & |F_{R'}| \\ \Gamma_{|E_R|} \downarrow \simeq & & \simeq \downarrow \Gamma_{|E_{R'}|} \\ |E_R| & \xrightarrow{|\iota_F|} & |E_{R'}| \end{array}$$

In other words, we have $|\iota_F| \circ \Gamma_{|E_R|} \simeq \Gamma_{|E_{R'}|} \circ |\iota_E|$, where ι_E, ι_F are the canonical inclusions.

From Theorem 35 we automatically obtain Theorem 34 (the strong form of Dowker's theorem) as an immediate corollary. The strong form does not appear in Dowker's original paper [46], but Björner has given a proof using the nerve theorem [11, Theorems 10.6, 10.9]. Moreover, Björner writes in a remark following [11, Theorem 10.9] that the nerve theorem and the strong form of Dowker's theorem are equivalent, in the sense that one

implies the other. We were not able to find an elementary proof of the strong form of Dowker's theorem in the existing literature. However, such an elementary proof is provided by our proof of Theorem 35 (given in Section 3.2.2), which we obtained by extending ideas in Dowker's original proof of Theorem 33.²

Whereas the Functorial Dowker Theorem and our elementary proof are of independent interest, it has been suggested in [27, Remark 4.8] that such a functorial version of Dowker's theorem could also be proved using a functorial nerve theorem [28, Lemma 3.4]. Despite being an interesting possibility, we were not able to find a detailed proof of this claim in the literature. In addition, Björner's remark regarding the equivalence between the nerve theorem and the strong form of Dowker's theorem suggests the following question:

Question 1. *Are the Functorial Nerve Theorem (FNT) of [28] and the Functorial Dowker Theorem (FDT, Theorem 35) equivalent?*

This question is of fundamental importance because the Nerve Theorem is a crucial tool in the applied topology literature and its functorial generalizations are equally important in persistent homology. In general, the answer is *no*, and moreover, one (of the FNT and FDT) is not stronger than the other. The FNT of [28] is stated for paracompact spaces, which are more general than the simplicial complexes of the FDT. However, the FNT of [28] is stated for spaces with *finitely-indexed* covers, so the associated nerve complexes are necessarily finite. All the complexes involved in the statement of the FDT are allowed to be infinite, so the FDT is more general than the FNT in this sense.

To clarify these connections, we formulate a simplicial Functorial Nerve Theorem (Theorem 38) and prove it via a finite formulation of the FDT (Theorem 36). In turn, we show that the simplicial FNT implies the finite FDT, thus proving the equivalence of these formulations (Theorem 39).

We begin with a weaker formulation of Theorem 35 and some simplicial Functorial Nerve Theorems.

Theorem 36 (The finite FDT). *Let X, Y be two totally ordered sets, and without loss of generality, suppose X is finite. Let $R \subseteq R' \subseteq X \times Y$ be two nonempty relations, and let $E_R, F_R, E_{R'}, F_{R'}$ be their associated simplicial complexes (as in Theorem 35). Then there exist homotopy equivalences $\Gamma_{|E_R|} : |F_R| \rightarrow |E_R|$ and $\Gamma_{|E_{R'}|} : |F_{R'}| \rightarrow |E_{R'}|$ that commute up to homotopy with the canonical inclusions.*

The finite FDT (Theorem 36) is an immediate consequence of the general FDT (Theorem 35).

Recall the definition of the nerve complex. Let $\mathcal{A} = \{A_i\}_{i \in I}$ be a family of nonempty sets indexed by I . The *nerve* of \mathcal{A} is the simplicial complex $\mathcal{N}(\mathcal{A}) := \{\sigma \in \text{pow}(I) : \sigma \text{ is finite, nonempty, and } \cap_{i \in \sigma} A_i \neq \emptyset\}$.

²A thread with ideas towards the proof of Theorem 34 was discussed in [1, last accessed 4.24.2017], but the proposed strategy was incomplete. We have inserted an addendum in [1] proposing a complete proof with a slightly different construction.

Definition 18 (Covers of simplices and subcomplexes). Let Σ be a simplicial complex. Then a collection of subcomplexes $\mathcal{A}_\Sigma = \{\Sigma_i\}_{i \in I}$ is said to be a *cover of subcomplexes* for Σ if $\Sigma = \cup_{i \in I} \Sigma_i$. Furthermore, \mathcal{A}_Σ is said to be a *cover of simplices* if each $\Sigma_i \in \mathcal{A}_\Sigma$ has the property that $\Sigma_i = \text{pow}(V(\Sigma_i))$. In this case, each Σ_i has precisely one top-dimensional simplex, consisting of the vertex set $V(\Sigma_i)$.

We present two *simplicial* formulations of the Functorial Nerve Theorem that turn out to be equivalent; the statements differ in that one is about covers of simplices and the other is about covers of subcomplexes.

Theorem 37 (Functorial Nerve I). *Let $\Sigma \subseteq \Sigma'$ be two simplicial complexes, and let $\mathcal{A}_\Sigma = \{\Sigma_i\}_{i \in I}$, $\mathcal{A}_{\Sigma'} = \{\Sigma'_i\}_{i \in I'}$ be finite covers of simplices for Σ and Σ' such that $I \subseteq I'$ and $\Sigma_i \subseteq \Sigma'_i$ for each $i \in I$. In particular, $\text{card}(I') < \infty$. Suppose that for each finite subset $\sigma \subseteq I'$, the intersection $\cap_{i \in \sigma} \Sigma'_i$ is either empty or contractible (and likewise for $\cap_{i \in \sigma} \Sigma_i$). Then $|\Sigma| \simeq |\mathcal{N}(\mathcal{A}_\Sigma)|$ and $|\Sigma'| \simeq |\mathcal{N}(\mathcal{A}_{\Sigma'})|$, via maps that commute up to homotopy with the canonical inclusions.*

Theorem 38 (Functorial Nerve II). *The statement of Theorem 37 holds even if \mathcal{A}_Σ and $\mathcal{A}_{\Sigma'}$ are covers of subcomplexes. Explicitly, the statement is as follows. Let $\Sigma \subseteq \Sigma'$ be two simplicial complexes, and let $\mathcal{A}_\Sigma = \{\Sigma_i\}_{i \in I}$, $\mathcal{A}_{\Sigma'} = \{\Sigma'_i\}_{i \in I'}$ be finite covers of subcomplexes for Σ and Σ' such that $I \subseteq I'$ and $\Sigma_i \subseteq \Sigma'_i$ for each $i \in I$. In particular, $\text{card}(I') < \infty$. Suppose that for each finite subset $\sigma \subseteq I'$, the intersection $\cap_{i \in \sigma} \Sigma'_i$ is either empty or contractible (and likewise for $\cap_{i \in \sigma} \Sigma_i$). Then $|\Sigma| \simeq |\mathcal{N}(\mathcal{A}_\Sigma)|$ and $|\Sigma'| \simeq |\mathcal{N}(\mathcal{A}_{\Sigma'})|$, via maps that commute up to homotopy with the canonical inclusions.*

The following result summarizes our answer to Question 1.

Theorem 39 (Equivalence). *The finite FDT, the FNT I, and the FNT II are all equivalent. Moreover, all of these results are implied by the FDT, as below:*



1.4.6 Dowker persistence diagrams and asymmetry

From the very definition of the Rips complex at any given resolution, one can see that the Rips complex is blind to asymmetry in the input data (Remark 28). In this section, we argue that either of the Dowker source and sink complexes is sensitive to asymmetry. Thus when analyzing datasets containing asymmetric information, one may wish to use the Dowker filtration instead of the Rips filtration. In particular, this property suggests that the Dowker persistence diagram is a stronger invariant for directed networks than the Rips persistence diagram.

In this section, we consider the *cycle networks* from §1.3, for which the Dowker persistence diagrams capture meaningful structure, whereas the Rips persistence diagrams do not.

We then probe the question “What happens to the Dowker or Rips persistence diagram of a network upon reversal of one (or more) edges?” Intuitively, if either of these persistence diagrams captures asymmetry, we would see a change in the diagram after applying this reversal operation to an edge.

To provide further evidence that Dowker persistence is sensitive to asymmetry, we computed both the Rips and Dowker persistence diagrams, in dimensions 0 and 1, of cycle networks G_n , for values of n between 3 and 6. Computations were carried out using Javaplex in Matlab with \mathbb{Z}_2 coefficients. The results are presented in Figure 1.7. Based on our computations, we were able to conjecture and prove the result in Theorem 40, which gives a precise characterization of the 1-dimensional Dowker persistence diagram of a cycle network G_n , for any n . Furthermore, the 1-dimensional Dowker persistence barcode for any G_n contains only one persistent interval, which agrees with our intuition that there is only one nontrivial loop in G_n . On the other hand, for large n , the 1-dimensional Rips persistence barcodes contain more than one persistent interval. This can be seen in the Rips persistence barcode of G_6 , presented in Figure 1.7. Moreover, for $n = 3, 4$, the 1-dimensional Rips persistence barcode does not contain any persistent interval at all. This suggests that Dowker persistence diagrams/barcodes are an appropriate method for analyzing cycle networks, and perhaps asymmetric networks in general.

The following theorem contains the characterization result for 1-dimensional Dowker persistence diagrams of cycle networks.

Theorem 40. *Let $G_n = (X_n, \omega_{G_n})$ be a cycle network for some $n \in \mathbb{N}$, $n \geq 3$. Then we obtain:*

$$\text{Dgm}_1^{\mathfrak{D}}(G_n) = \{(1, \lceil n/2 \rceil) \in \mathbb{R}^2\}.$$

Thus $\text{Dgm}_1^{\mathfrak{D}}(G_n)$ consists of precisely the point $(1, \lceil n/2 \rceil) \in \mathbb{R}^2$ with multiplicity 1.

Remark 41. From our experimental results (cf. Figure 1.7), it appears that the 1-dimensional Rips persistence diagram of a cycle network does not admit a characterization as simple as that given by Theorem 40 for the 1-dimensional Dowker persistence diagram. Moreover, the Rips complexes $\mathfrak{R}_{G_n}^{\delta}$, $\delta \in \mathbb{R}$, $n \in \mathbb{N}$ correspond to certain types of *independence complexes* that appear independently in the literature, and whose homotopy types remain open [53, Question 5.3]. On a related note, we point the reader to [3] for a complete characterization of the homotopy types of Rips complexes of points on the circle (equipped with the restriction of the arc length metric).

To elaborate on the connection to [53], we write H_n^k to denote the undirected graph with vertex set $\{1, \dots, n\}$, and edges given by pairs (i, j) where $1 \leq i < j \leq n$ and either $j - i < k$ or $(n + i) - j < k$. Next we write $\text{Ind}(H_n^k)$ to denote the *independence complex* of H_n^k , which is the simplicial complex consisting of subsets $\sigma \subseteq \{1, 2, \dots, n\}$ such that no two elements of σ are connected by an edge in H_n^k . Then we have $\text{Ind}(H_n^k) = \mathfrak{R}_{G_n}^{n-k}$

for each $k, n \in \mathbb{N}$ such that $k < n$. To gain intuition for this equality, fix a basepoint 1, and consider the values of $j \in \mathbb{N}$ for which the simplex $[1, j]$ belongs to $\text{Ind}(H_n^k)$ and to $\mathfrak{R}_{G_n}^{n-k}$, respectively. In either case, we have $k + 1 \leq j \leq n - k + 1$. Using the rotational symmetry of the points, one then obtains the remaining 1-simplices. Rips complexes are determined by their 1-skeleton, so this suffices to construct $\mathfrak{R}_{G_n}^{n-k}$. Analogously, $\text{Ind}(H_n^k)$ is determined by the edges in H_n^k , and hence also by its 1-skeleton. In [53, Question 5.3], the author writes that the homotopy type of $\text{Ind}(H_n^k)$ is still unsolved. Characterizing the persistence diagrams $\text{Dgm}_k^{\mathfrak{R}}(G_n)$ thus seems to be a useful future step, both in providing a computational suggestion for the homotopy type of $\text{Ind}(H_n^k)$, and also in providing a valuable example in the study of persistence of directed networks.

Remark 42. Theorem 40 has the following implication for data analysis: nontrivial 1-dimensional homology in the Dowker persistence diagram of an asymmetric network suggests the presence of directed cycles in the underlying data. Of course, it is not necessarily true that nontrivial 1-dimensional persistence can occur *only* in the presence of a directed circle.

Remark 43. Our motivation for studying cycle networks is that they constitute directed analogues of circles, and we were interested in seeing if the 1-dimensional Dowker persistence diagram would be able to capture this analogy. Theorem 40 shows that this is indeed the case: we get a single nontrivial 1-dimensional persistence interval, which is what we would expect when computing the persistent homology of a circle in the metric space setting.

While we had initially proved Theorem 40 using elementary methods, Henry Adams observed that Dowker complexes on cycle networks can be precisely related to nerve complexes built over arcs on the circle. Using the techniques developed in [3] and [4], one obtains the following results:

Theorem 44 (Even dimension). *Fix $n \in \mathbb{N}$, $n \geq 3$. If $l \in \mathbb{N}$ is such that n is divisible by $(l + 1)$, and $k := \frac{nl}{l+1}$ is such that $0 \leq k \leq n - 2$, then $\text{Dgm}_{2l}^{\mathfrak{D}}(G_n)$ consists of precisely the point $(\frac{nl}{l+1}, \frac{nl}{l+1} + 1)$ with multiplicity $\frac{n}{l+1} - 1$. If l or k do not satisfy the conditions above, then $\text{Dgm}_{2l}^{\mathfrak{D}}(G_n)$ is trivial.*

As a special case, we know that if n is odd, then $\text{Dgm}_2^{\mathfrak{D}}(G_n)$ is trivial. If n is even, then $\text{Dgm}_2^{\mathfrak{D}}(G_n)$ consists of the point $(\frac{n}{2}, \frac{n}{2} + 1)$ with multiplicity $\frac{n}{2} - 1$.

Theorem 45 (Odd dimension). *Fix $n \in \mathbb{N}$, $n \geq 3$. Then for $l \in \mathbb{N}$, define $M_l := \left\{ m \in \mathbb{N} : \frac{nl}{l+1} < m < \frac{n(l+1)}{l+2} \right\}$. If M_l is empty, then $\text{Dgm}_{2l+1}^{\mathfrak{D}}(G_n)$ is trivial. Otherwise, we have:*

$$\text{Dgm}_{2l+1}^{\mathfrak{D}}(G_n) = \left\{ \left(a_l, \left\lceil \frac{n(l+1)}{l+2} \right\rceil \right) \right\},$$

where $a_l := \min \{m \in M_l\}$. We use set notation (instead of multisets) to mean that the multiplicity is 1.

In particular, for $l = 0$, we have $\frac{nl}{l+1} = 0$ and $\frac{n(l+1)}{l+2} = \frac{n}{2} \geq 3/2$, so $1 \in M_l$. Thus we have $\text{Dgm}_1^{\mathfrak{D}}(G_n) = \left\{ \left(1, \lceil \frac{n}{2} \rceil \right) \right\}$, and so Theorem 45 recovers Theorem 40 as a special case.

Sensitivity to network transformations

We now make a definition:

Definition 19 (Pair swaps). Let (X, ω_X) be a network. For any $z, z' \in X$, define the (z, z') -swap of (X, ω_X) to be the network $S_X(z, z') := (X^{z,z'}, \omega_X^{z,z'})$ defined as follows:

$$X^{z,z'} := X,$$

$$\text{For any } x, x' \in X^{z,z'}, \quad \omega_X^{z,z'}(x, x') := \begin{cases} \omega_X(x', x) & : x = z, x' = z' \\ \omega_X(x', x) & : x' = z, x = z' \\ \omega_X(x, x') & : \text{otherwise.} \end{cases}$$

We then pose the following question:

Given a network (X, ω_X) and an (x, x') -swap $S_X(x, x')$ for some $x, x' \in X$, how do the Rips or Dowker persistence diagrams of $S_X(x, x')$ differ from those of (X, ω_X) ?

This situation is illustrated in Figure 1.8. Example 49 shows an example where the Dowker persistence diagram captures the variation in a network that occurs after a pair swap, whereas the Rips persistence diagram fails to capture this difference. Furthermore, Remark 47 shows that Rips persistence diagrams always fail to do so.

We also consider the extreme situation where all the directions of the edges of a network are reversed, i.e. the network obtained by applying the pair swap operation to each pair of nodes. We would intuitively expect that the persistence diagrams would not change. The following discussion shows that the Rips and Dowker persistence diagrams are invariant under taking the transpose of a network.

Proposition 46. Recall the transposition map t and the shorthand notation $X^\top = t(X)$ from Definition 17. Let $k \in \mathbb{Z}_+$. Then $\text{Dgm}_k^{\text{si}}(X) = \text{Dgm}_k^{\text{so}}(X^\top)$, and therefore $\text{Dgm}_k^{\mathfrak{D}}(X) = \text{Dgm}_k^{\mathfrak{D}}(X^\top)$ by Theorem 35.

Remark 47 (Pair swaps and their effect). Let $(X, \omega_X) \in \mathcal{CN}$, let $z, z' \in X$, and let $\sigma \in \text{pow}(X)$. Then we have:

$$\max_{x, x' \in \sigma} \omega_X(x, x') = \max_{x, x' \in \sigma} \omega_X^{z,z'}(x, x').$$

Using this observation, one then repeats the arguments used in the proof of Proposition 46 to show that:

$$\text{Dgm}_k^{\mathfrak{R}}(X) = \text{Dgm}_k^{\mathfrak{R}}(S_X(z, z')), \text{ for each } k \in \mathbb{Z}_+.$$

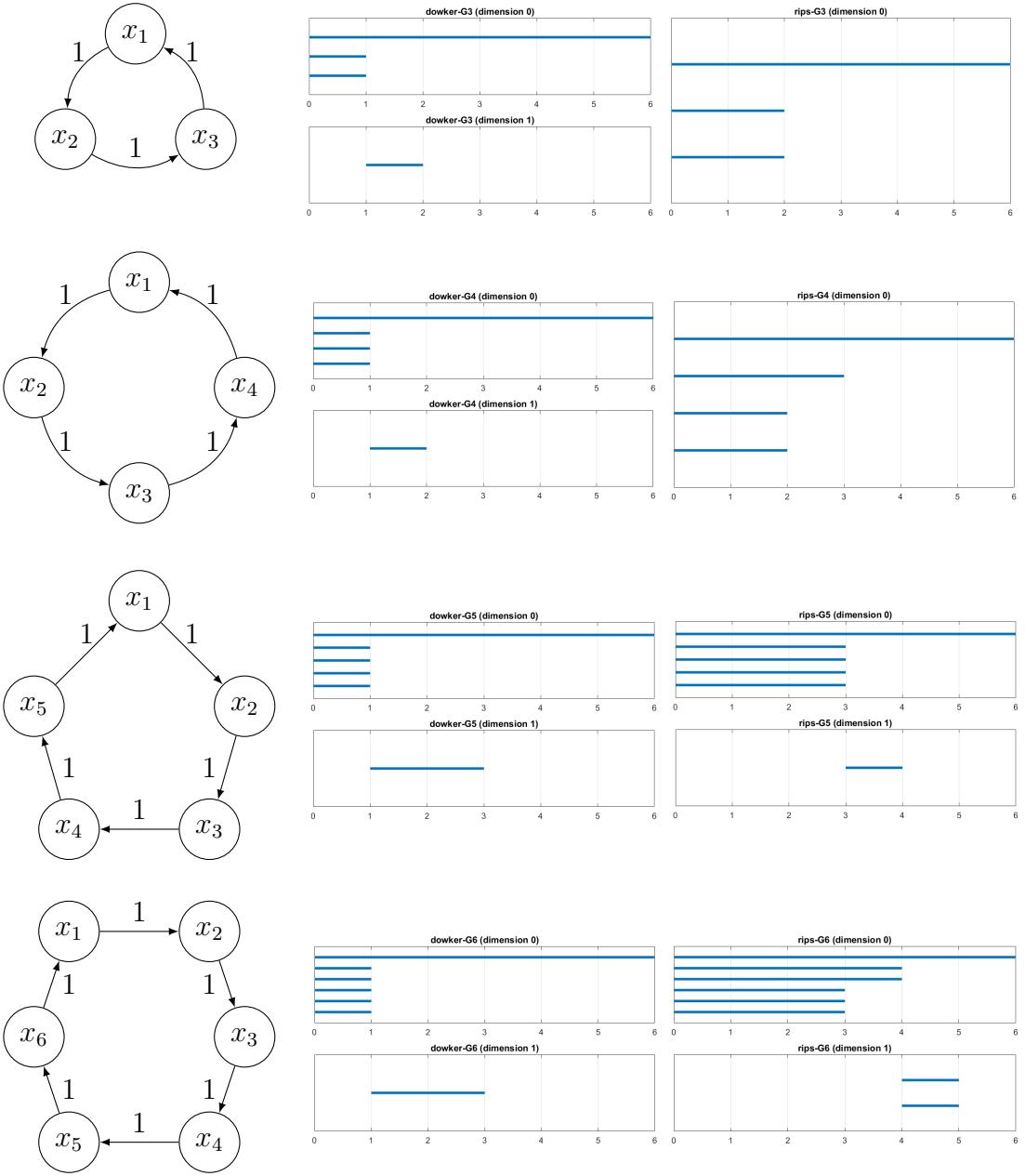


Figure 1.7: The first column contains illustrations of cycle networks G_3, G_4, G_5 and G_6 . The second column contains the corresponding Dowker persistence barcodes, in dimensions 0 and 1. Note that the persistent intervals in the 1-dimensional barcodes agree with the result in Theorem 40. The third column contains the Rips persistence barcodes of each of the cycle networks. Note that for $n = 3, 4$, there are no persistent intervals in dimension 1. On the other hand, for $n = 6$, there are two persistent intervals in dimension 1.

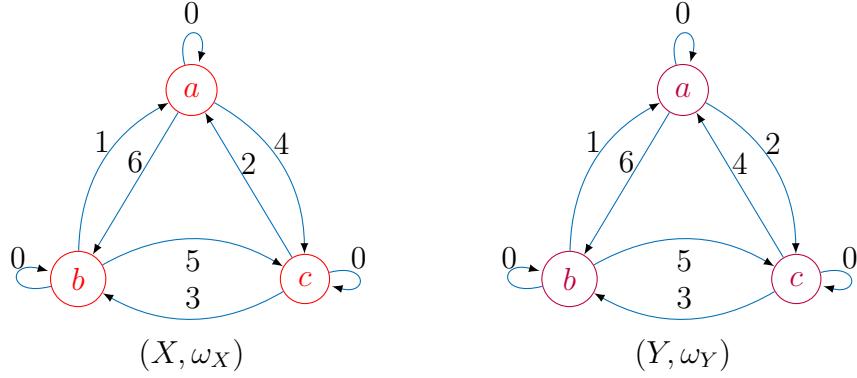


Figure 1.8: (Y, ω_Y) is the (a, c) -swap of (X, ω_X) .

This encodes the intuitive fact that Rips persistence diagrams are blind to pair swaps. Moreover, successively applying the pair swap operation over all pairs produces the transpose of the original network, and so it follows that $\text{Dgm}_k^{\mathfrak{R}}(X) = \text{Dgm}_k^{\mathfrak{R}}(X^\top)$.

On the other hand, k -dimensional Dowker persistence diagrams are not necessarily invariant to pair swaps when $k \geq 1$. Indeed, Example 49 below constructs a space X for which there exist points $z, z' \in X$ such that

$$\text{Dgm}_1^{\mathfrak{D}}(X) \neq \text{Dgm}_1^{\mathfrak{D}}(S_X(z, z')).$$

However, 0-dimensional Dowker persistence diagrams are still invariant to pair swaps:

Proposition 48. *Let $(X, \omega_X) \in \mathcal{CN}$, let z, z' be any two points in Z , and let $\sigma \in \text{pow}(X)$. Then we have:*

$$\text{Dgm}_0^{\mathfrak{D}}(X) = \text{Dgm}_0^{\mathfrak{D}}(S_X(z, z')).$$

Example 49. Consider the three node dissimilarity networks (X, ω_X) and (Y, ω_Y) in Figure 1.8. Note that (Y, ω_Y) coincides with $S_X(a, c)$. We present both the Dowker and Rips persistence barcodes obtained from these networks. Note that the Dowker persistence barcode is sensitive to the difference between (X, ω_X) and (Y, ω_Y) , whereas the Rips barcode is blind to this difference.

To show how the Dowker complex is constructed, we also list the Dowker sink complexes of the networks in Figure 1.8, and also the corresponding homology dimensions across a range of resolutions. Note that when we write $[a, b](a)$, we mean that a is a sink

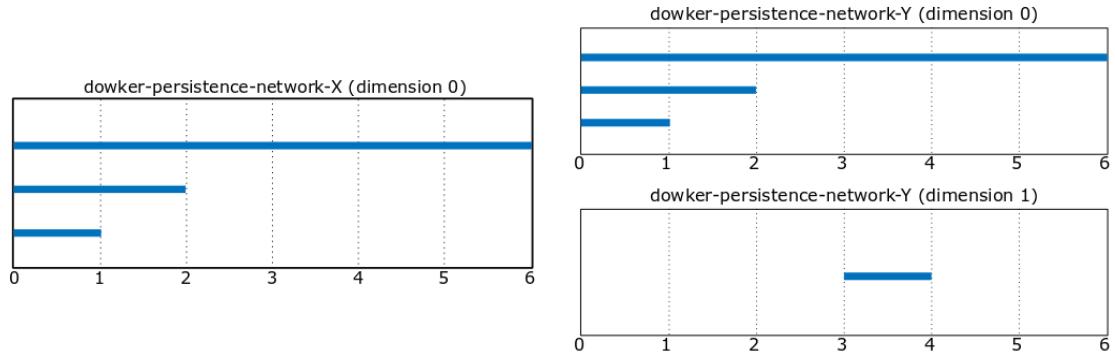


Figure 1.9: Dowker persistence barcodes of networks (X, ω_X) and (Y, ω_Y) from Figure 1.8.

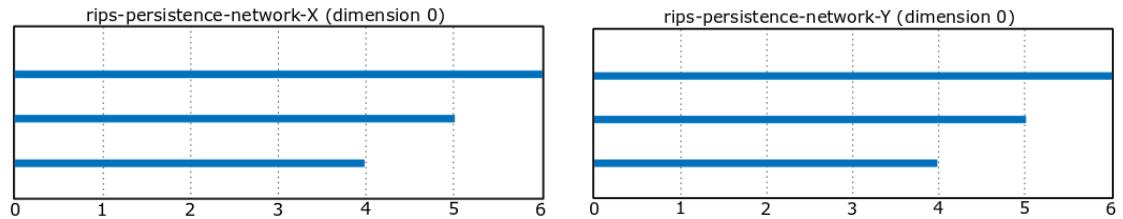


Figure 1.10: Rips persistence barcodes of networks (X, ω_X) and (Y, ω_Y) from Figure 1.8. Note that the Rips diagrams indicate no persistent homology in dimensions higher than 0, in contrast with the Dowker diagrams in Figure 1.9.

corresponding to the simplex $[a, b]$.

$$\begin{array}{lll}
\mathfrak{D}_{0,X}^{\text{si}} = \{[a], [b], [c]\} & \dim(H_1(\mathfrak{D}_{0,X}^{\text{si}})) = 0 \\
\mathfrak{D}_{1,X}^{\text{si}} = \{[a], [b], [c], [a, b](a)\} & \dim(H_1(\mathfrak{D}_{1,X}^{\text{si}})) = 0 \\
\mathfrak{D}_{2,X}^{\text{si}} = \{[a], [b], [c], [a, b](a), [a, c](a), [b, c](a), [a, b, c](a)\} & \dim(H_1(\mathfrak{D}_{2,X}^{\text{si}})) = 0 \\
\mathfrak{D}_{3,X}^{\text{si}} = \{[a], [b], [c], [a, b](a), [a, c](a), [b, c](a), [a, b, c](a)\} & \dim(H_1(\mathfrak{D}_{3,X}^{\text{si}})) = 0 \\
\\
\mathfrak{D}_{0,Y}^{\text{si}} = \{[a], [b], [c]\} & \dim(H_1(\mathfrak{D}_{0,Y}^{\text{si}})) = 0 \\
\mathfrak{D}_{1,Y}^{\text{si}} = \{[a], [b], [c], [a, b](a)\} & \dim(H_1(\mathfrak{D}_{1,Y}^{\text{si}})) = 0 \\
\mathfrak{D}_{2,Y}^{\text{si}} = \{[a], [b], [c], [a, b](a), [a, c](c)\} & \dim(H_1(\mathfrak{D}_{2,Y}^{\text{si}})) = 0 \\
\mathfrak{D}_{3,Y}^{\text{si}} = \{[a], [b], [c], [a, b](a), [a, c](c), [b, c](b)\} & \dim(H_1(\mathfrak{D}_{3,Y}^{\text{si}})) = 1 \\
\mathfrak{D}_{4,Y}^{\text{si}} = \{[a], [b], [c], [a, b](a), [a, c](a), [b, c](a), [a, b, c](a)\} & \dim(H_1(\mathfrak{D}_{4,Y}^{\text{si}})) = 0
\end{array}$$

Note that for $\delta \in [3, 4)$, $\dim(H_1(\mathfrak{D}_{\delta,Y}^{\text{si}})) = 1$, whereas $\dim(H_1(\mathfrak{D}_{\delta,X}^{\text{si}})) = 0$ for each $\delta \in \mathbb{R}$.

Based on the discussion in Remark 47, Proposition 48, and Example 49, we conclude the following:

Moral: *Unlike Rips persistence diagrams, Dowker persistence diagrams are truly sensitive to asymmetry.*

We summarize some of these results:

Theorem 50. *Recall the symmetrization and transposition maps \mathfrak{s} and \mathfrak{t} from Definition 17. Then:*

1. $\mathfrak{R} \circ \mathfrak{s} = \mathfrak{R}$,
2. $\mathfrak{D}^{\text{so}} \circ \mathfrak{t} = \mathfrak{D}^{\text{si}}$, and
3. $\mathfrak{D}^{\text{si}} \circ \mathfrak{t} = \mathfrak{D}^{\text{so}}$.

Also, there exist $(X, \omega_X), (Y, \omega_Y) \in \mathcal{FN}$ such that $(\mathfrak{D}^{\text{si}} \circ \mathfrak{s})(X) \neq \mathfrak{D}^{\text{si}}(X)$, and $(\mathfrak{D}^{\text{so}} \circ \mathfrak{s})(Y) \neq \mathfrak{D}^{\text{so}}(Y)$.

Proof. These follow from Example 49, Remark 28, and Proposition 46. \square

1.5 Persistent path homology of networks

To define PPH, we first summarize and condense some concepts that appeared in [62]. We also point the reader to Section 3.1.1 for the necessary background on chain complexes and associated constructions.

Remark 51 (Reconstructing networks from path filtrations). We make an observation to explain the moral difference between path homology and the simplicial constructions we considered above. In the setting of metric spaces with the VR filtration, it is always possible to recover full metric information from the filtered space. Specifically, given a pair (x, x') , we can recover $d_X(x, x')$ simply as the filtration value of the simplex $[x, x']$. If the metric space is *geodesic*, then $d_X(x, x')$ can also be recovered from the Čech filtration—the simplex $[x, x']$ is witnessed by their midpoint, so $d_X(x, x')$ is twice the filtration value of $[x, x']$ (thanks to F. Mémoli for this observation).

In the setting of networks, however, edge weight information cannot typically be recovered from filtration values. The issue lies in the equality $[a, b] = -[b, a]$ for (oriented) simplicial complexes. Even in the case of the VR filtration, we know that the filtration value of $[a, b]$ corresponds to either $\omega_X(a, b)$ or $\omega_X(b, a)$, but it is impossible to know which.

Path homology, in contrast, allows us to recover full metric information. The key idea is that the equality $[a, b] = -[b, a]$ is removed, and $[a, b], [b, a]$ are linearly independent at the chain complex level. Thus the filtration value of $[a, b]$ recovers $\omega_X(a, b)$, and that of $[b, a]$ recovers $\omega_X(b, a)$.

1.5.1 Path homology of digraphs

Elementary paths on a set

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

$$\partial_p^{\text{nr}}([x_0, \dots, x_p]) := \sum_{i=0}^p (-1)^i [x_0, \dots, \widehat{x}_i, \dots, x_p],$$

for each elementary p -path $[x_0, \dots, x_p] \in \Lambda_p$. Here \widehat{x}_i denotes omission of x_i from the sequence. The maps ∂_p^{nr} are referred to as the *non-regular boundary maps*. For $p = -1$, one defines $\partial_{-1}^{\text{nr}} : \Lambda_{-1} \rightarrow \Lambda_{-2}$ to be the zero map. Then $\partial_{p+1}^{\text{nr}} \circ \partial_p^{\text{nr}} = 0$ for any integer $p \geq -1$ [64, Lemma 2.2]. It follows that $(\Lambda_p, \partial_p^{\text{nr}})_{p \in \mathbb{Z}_+}$ is a chain complex.

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

Example 52 (Paths on a double edge). We will soon explain the interaction between paths on a set and the edges on a digraph. First consider a digraph on a vertex set $Y = \{a, b\}$ as in Figure 1.11. Notice that there is a legitimate “path” on this digraph of the form aba , obtained by following the directions of the edges. But notice that applying ∂_2^{nr} to the 2-path aba yields $\partial_2^{\text{nr}}(aba) = ba - aa + ab$, and aa is not a valid path on this particular digraph (self-loops are disallowed). To handle situations like this, one needs to consider *regular paths*, which are explained in the next section.



Figure 1.11: A two-node digraph on the vertex set $Y = \{a, b\}$.

Regular paths on a set

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

$$\begin{aligned}\mathcal{R}_p &= \mathcal{R}_p(X, \mathbb{K}) := \mathbb{K}[\{[x_0, \dots, x_p] : [x_0, \dots, x_p] \text{ is regular}\}] \\ \mathcal{I}_p &= \mathcal{I}_p(X, \mathbb{K}) := \mathbb{K}[\{[x_0, \dots, x_p] : [x_0, \dots, x_p] \text{ is irregular}\}].\end{aligned}$$

One can further verify that $\partial_p^{\text{nr}}(\mathcal{I}_p) \subseteq \mathcal{I}_{p-1}$ [64, Lemma 2.6], and so ∂_p^{nr} is well-defined on Λ_p/\mathcal{I}_p . Since $\mathcal{R}_p \cong \Lambda_p/\mathcal{I}_p$ via a natural linear isomorphism, one can define $\partial_p : \mathcal{R}_p \rightarrow \mathcal{R}_{p-1}$ as the pullback of ∂_p^{nr} via this isomorphism [64, Definition 2.7]. Then ∂_p is referred to as the *regular boundary map* in dimension p , where $p \in \mathbb{Z}_+$. Now we obtain a new chain complex $(\mathcal{R}_p, \partial_p)_{p \in \mathbb{Z}_+}$.

Example 53 (Regular paths on a double edge). Consider again the digraph in Figure 1.11. Applying the regular boundary map to the 2-path aba yields $\partial_2(aba) = ba + ab$. This example illustrates the following general principle: *Irregular paths arising from an application of ∂_\bullet are treated as zeros*.

Allowed paths on digraphs

We now expand on the notion of paths on a set to discuss paths on a digraph. We follow the intuition developed in Examples 52 and 53.

Let $G = (X, E)$ be a digraph (possibly infinite) without self-loops. For each $p \in \mathbb{Z}_+$, one defines an elementary p -path $[x_0, \dots, x_p]$ on X to be *allowed* if $(x_i, x_{i+1}) \in E$ for each $0 \leq i \leq p-1$. For each $p \in \mathbb{Z}_+$, the free vector space on the collection of allowed p -paths

on (X, E) is denoted $\mathcal{A}_p = \mathcal{A}_p(G) = \mathcal{A}_p(X, E, \mathbb{K})$, and is called the *space of allowed p -paths*. One further defines $\mathcal{A}_{-1} := \mathbb{K}$ and $\mathcal{A}_{-2} := \{0\}$.



Figure 1.12: Two types of square digraphs.

∂ -invariant paths and path homology

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

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

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

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

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

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

Example 54 (Paths on squares). We illustrate the construction of Ω_\bullet for the digraphs in Figure 1.12.

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

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

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

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

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

To compare this with a simplicial approach, consider the directed clique complex homology studied in [102, 84, 118]. Given a digraph $G = (X, E)$, the directed clique complex is defined to be the *ordered* simplicial complex [88, p. 76] given by writing:

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

Here we use parentheses to denote ordered simplices. For the squares in Figure 1.12, we have:

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

and so their simplicial homologies are equal.

Remark 55 (The challenge of finding a natural basis for Ω_\bullet). The digraph G_N in Example 54 is a minimal example showing that it is nontrivial to compute bases for the vector spaces Ω_\bullet . Specifically, while it is trivial to read off bases for the allowed paths \mathcal{A}_\bullet from a digraph, one needs to consider *linear combinations* of allowed paths in a systematic manner to obtain bases for the ∂ -invariant paths.

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

We derive a procedure for systematically computing bases for Ω_\bullet in §4.4.

1.5.2 The persistent path homology of a network

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

$$E_X^\delta := \{(x, x') \in X \times X : x \neq x', \omega_X(x, x') \leq \delta\}.$$

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

The functoriality of the path homology construction (Appendix 3.3.1, Proposition 172) enables us to obtain a persistent vector space from a digraph filtration. Thus we make the following definition:

Definition 20. Let $\mathfrak{G} = \{\mathfrak{G}^\delta \hookrightarrow \mathfrak{G}^{\delta'}\}_{\delta \leq \delta' \in \mathbb{R}}$ be a digraph filtration. Then for each $p \in \mathbb{Z}_+$, we define the p -dimensional persistent path homology of \mathfrak{G} to be the following persistent vector space:

$$\mathbf{PVec}_p^{\Xi}(\mathfrak{G}) := \{H_p^{\Xi}(\mathfrak{G}^\delta) \xrightarrow{(\iota_{\delta, \delta'})^\#} H_p^{\Xi}(\mathfrak{G}^{\delta'})\}_{\delta \leq \delta' \in \mathbb{R}}.$$

When it is defined, the diagram associated to $\mathbf{PVec}_p^{\Xi}(\mathfrak{G})$ is denoted $\text{Dgm}_p^{\Xi}(\mathfrak{G})$.

In particular, by Theorem 86, the path persistence diagram in dimension p is defined for any $(X, \omega_X) \in \mathcal{CN}$. We write $\text{Dgm}_p^{\Xi}(X)$ to denote this diagram.

Persistent path homology is stable to perturbations of input data, and hence amenable to data analysis:

Theorem 56 (Stability). *Let $(X, \omega_X), (Y, \omega_Y) \in \mathcal{CN}$. Let $p \in \mathbb{Z}_+$. Then,*

$$d_I(\mathbf{PVec}_p^{\Xi}(X), \mathbf{PVec}_p^{\Xi}(Y)) \leq 2d_N(X, Y).$$

We note that by Corollary 177, we have:

$$d_B(\text{Dgm}_k^{\Xi}(X), \text{Dgm}_k^{\Xi}(Y)) \leq 2d_N(X, Y).$$

Remark 57. While the preceding stability result is analogous to those for Vietoris-Rips and Dowker persistence, the proofs in this setting require results on the homotopy of digraphs that were recently developed in [63] (cf. Section 3.3.2).

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

Example 58 (PPH vs Dowker for metric n -cubes). In the setting of metric spaces, PPH is generally different from Dowker persistence in dimensions ≥ 2 . To see this, consider \mathbb{R}^n equipped with the Euclidean distance for $n \geq 3$. Define

$$\square_n := \{(i_1, i_2, \dots, i_n) : i_j \in \{0, 1\} \ \forall 1 \leq j \leq n\}.$$

Then $\mathfrak{G}_{\square_n}^\delta$ has no edges for $\delta < 1$, and for $\delta = 1$, it has precisely an edge between any two points of \square_n that differ on a single coordinate. But at $\delta = 1$, $\mathfrak{G}_{\square_n}^\delta$ is homotopy equivalent to $\mathfrak{G}_{\square_{n-1}}^\delta$: the homotopy equivalence is given by collapsing points that differ exactly on the n th coordinate (see Figure 1.14). Proceeding recursively, we see that $\mathfrak{G}_{\square_1}^\delta$ is contractible at $\delta = 1$. However, $\mathfrak{D}^{\text{si}}(\square_n)$ is not contractible at $\delta = 1$. Moreover, an explicit verification for the $n = 3$ case shows that $\text{Dgm}_2^{\mathfrak{D}}(\square_3)$ consists of the point $(1, \sqrt{2})$ with multiplicity 7. Thus $\text{Dgm}_2^{\mathfrak{D}}(\square_3) \neq \text{Dgm}_2^{\Xi}(\square_3)$.

Theorem 59. *Let $\mathcal{X} = (X, A_X) \in \mathcal{CN}$ be a symmetric network, and fix $\mathbb{K} = \mathbb{Z}/p\mathbb{Z}$ for some prime p . Then $\text{Dgm}_1^{\Xi}(\mathcal{X}) = \text{Dgm}_1^{\mathfrak{D}}(\mathcal{X})$.*

The preceding result shows that on metric spaces, PPH agrees with Dowker persistence in dimension 1. The converse implication is not true: in §1.5.3, we provide a family of highly asymmetric networks for which PPH agrees with Dowker persistence in dimension 1. On the other hand, the examples in Figure 1.13 show that equality in dimension 1 does not necessarily hold for asymmetric networks. Moreover, it turns out that the four-point configurations illustrated in Figure 1.13 can be used to give another partial characterization of the networks for which PPH and Dowker persistence do agree in dimension 1. We present this statement next.

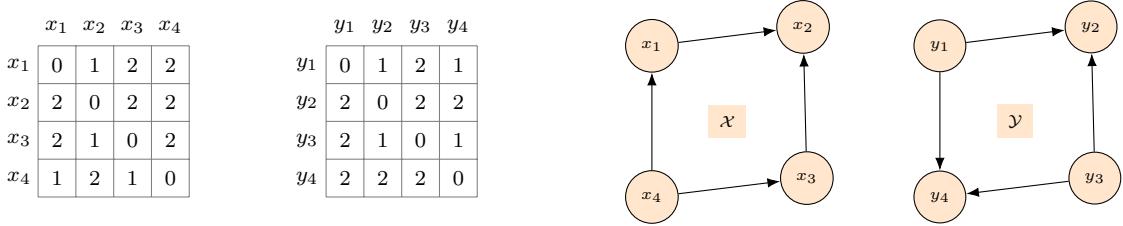
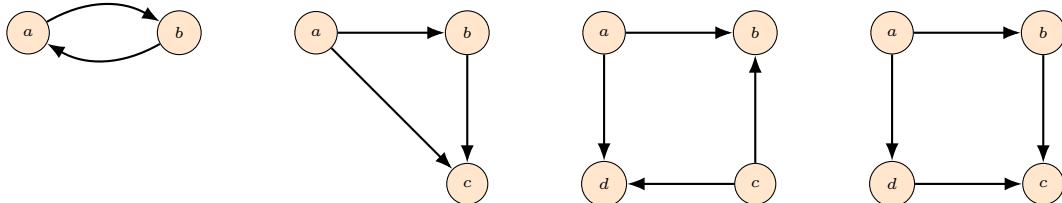


Figure 1.13: Working over $\mathbb{Z}/2\mathbb{Z}$ coefficients, we find that $\text{Dgm}_1^{\Xi}(\mathcal{X})$ and $\text{Dgm}_1^{\mathfrak{D}}(\mathcal{Y})$ are trivial, whereas $\text{Dgm}_1^{\mathfrak{D}}(\mathcal{X}) = \text{Dgm}_1^{\Xi}(\mathcal{Y}) = \{(1, 2)\} = \{(1, 2)\}$.

Definition 21 (Squares, triangles, and double edges). Let \mathfrak{G} be a finite digraph. Then we define the following local configurations of edges between distinct nodes a, b, c, d :

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



Finally, we define a network (X, A_X) to be *square-free* if \mathfrak{G}_X^δ does not contain a four-point subset whose induced subgraph is a short or long square, for any $\delta \in \mathbb{R}$. An important

observation is that to be a square, the subgraph induced by a four-point subset cannot just include one of the configurations pictured above; it must exclude diagonal edges as well.

Remark 60. We thank Guilherme Vituri for pointing out the need to exclude the $b \rightarrow d$ edge in the definition of a long square.

Theorem 61. Let $\mathcal{X} = (X, A_X) \in \mathcal{CN}$ be a square-free network, and fix $\mathbb{K} = \mathbb{Z}/p\mathbb{Z}$ for some prime p . Then $\text{Dgm}_1^{\Xi}(\mathcal{X}) = \text{Dgm}_1^{\mathfrak{D}}(\mathcal{X})$.

Remark 62. The proofs of Theorems 59 and 61 both require an argument where simplices are paired up—this requires us to use $\mathbb{Z}/p\mathbb{Z}$ coefficients in both theorem statements.

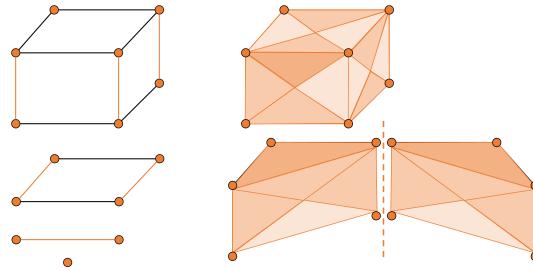


Figure 1.14: **Left:** $\mathfrak{G}_{\square,3}^{\delta}$ is (digraph) homotopy equivalent to a point at $\delta = 1$, as can be seen by collapsing points along the orange lines. **Right:** $\mathfrak{D}_{\delta,\square,3}^{\text{si}}$ becomes contractible at $\delta = \sqrt{2}$, but has nontrivial homology in dimension 2 that persists across the interval $[1, \sqrt{2}]$.

1.5.3 An application: Characterizing the diagrams of cycle networks

Notice that the cycle networks G_n defined in §1.3 are square-free. If $x_1, x_2, \dots, x_k \in X_n$ appear in G_n in this clockwise order, we write $x_1 \preceq x_2 \preceq \dots \preceq x_k$. If $a \prec b \prec c \prec d \prec a$ are four nodes on a cycle network, then for any $\delta \in \mathbb{R}$ such that we have an edge $a \rightarrow d$, we automatically have an edge $a \rightarrow c$. Thus the subgraph induced by $\{a, b, c, d\}$ cannot be either a long or a short square.

Cycle networks constitute an interesting family of examples with surprising connections to existing literature [4] [3]. In particular, their Dowker persistence diagrams can be fully characterized by results in [4], [3], and [37]. More specifically, given any $n \geq 3$, we know that $\text{Dgm}_1^{\mathfrak{D}}(G_n)$ consists of the point $(1, \lceil n/2 \rceil)$ with multiplicity 1. In this sense, a cycle network is a directed analogue of the circle.

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

Theorem 63. Let G_n be a cycle network for some integer $n \geq 3$. Fix a field $\mathbb{K} = \mathbb{Z}/p\mathbb{Z}$ for some prime p . Then $\text{Dgm}_1^{\Xi}(G_n) = \{(1, \lceil n/2 \rceil)\}$.

1.6 The case of compact networks

Having surveyed the constructions of persistent homology on networks, we return to $d_{\mathcal{N}}$ and discuss some more results.

1.6.1 ε -systems and finite sampling

Proofs from this section, along with an auxiliary lemma, are provided in §2.2

In this section, we develop the notion of ε -systems of networks. These are related to ε -nets for metric spaces. The key idea is that instead of trying to replicate the notion of an ε -net in a network setting (where even the notion of an open ball does not make sense), it generalizes a particular consequence of an ε -net to the network setting. This enables us to obtain the important result that any compact network can indeed be approximated up to arbitrary precision by a finite network. This result in turn is instrumental for guaranteeing that compact networks have well-defined persistence diagrams.

Proofs from this section are provided in §2.2. We start with a collection of statements about compact networks, and end with a result about ε -systems in networks equipped with a Borel probability measure (see §1.9.1 for more on these *measure networks*).

Definition 22 (ε -approximations). Let $\varepsilon > 0$. A network $(X, \omega_X) \in \mathcal{N}$ is said to be ε -approximable by $(Y, \omega_Y) \in \mathcal{N}$ if $d_{\mathcal{N}}(X, Y) < \varepsilon$. In this case, Y is said to be an ε -approximation of X . Typically, we will be interested in the case where X is infinite and Y is finite, i.e. in ε -approximating infinite networks by finite networks.

Definition 23 (ε -systems). Let $\varepsilon > 0$. For any network (X, ω_X) , an ε -system on X is a finite open cover $\mathcal{U} = \{U_1, \dots, U_n\}, n \in \mathbb{N}$, of X such that for any $1 \leq i, j \leq n$, we have $\omega_X(U_i, U_j) \subseteq B(r_{ij}, \varepsilon)$ for some $r_{ij} \in \mathbb{R}$.

In some cases, we will be interested in the situation where X is a finite union of connected components $\{X_1, \dots, X_n\}, n \in \mathbb{N}$. By a refined ε -system, we will mean an ε -system such that each element of the ε -system is contained in precisely one connected component of X .

Theorem 64 (\exists of refined ε -systems). Any compact network (X, ω_X) has a refined ε -system for any $\varepsilon > 0$. In particular, by picking a representative from each element of \mathcal{U} , we get a finite network X' such that $d_{\mathcal{N}}(X, X') < \varepsilon$.

Remark 65. When considering a compact metric space (X, d_X) , the preceding theorem relates to the well-known notion of taking finite ε -nets in a metric space. Recall that for $\varepsilon > 0$, a subset $S \subseteq X$ is an ε -net if for any point $x \in X$, we have $B(x, \varepsilon) \cap S \neq \emptyset$. Such an ε -net satisfies the nice property that $d_{\text{GH}}(X, S) < \varepsilon$ [17, 7.3.11]. In particular, one can find a finite ε -net of (X, d_X) for any $\varepsilon > 0$ by compactness.

We do not make quantitative estimates on the cardinality of the ε -approximation produced in Theorem 64. In the setting of compact metric spaces, the size of an ε -net relates

to the rich theory of metric entropy developed by Kolmogorov and Tihomirov [51, Chapter 17].

The preceding result shows that refined ε -systems always exist; this result relies crucially on the assumption that the network is compact. The proof of the theorem uses the continuity of $\omega_X : X \times X \rightarrow \mathbb{R}$ and the compactness of $X \times X$. In the setting of compact subsets of Euclidean space or compact metric spaces, ε -systems are easy to construct: we can just take a cover by ε -balls, and then extract a finite subcover by invoking compactness. The strength of Theorem 64 lies in proving the existence of ε -systems even when symmetry and triangle inequality (key requirements needed to guarantee the standard properties of ε -balls) *are not assumed*. The next result shows that by sampling points from all the elements of an ε -system, one obtains a finite, quantitatively good approximation to the underlying network.

Theorem 66 (ε -systems and d_N). *Let (X, ω_X) be a compact network, let $\varepsilon > 0$, and let \mathcal{U} be an ε -system on X . Suppose X' is any finite subset of X that has nonempty intersection with each element in \mathcal{U} . Then there exists a correspondence $R' \in \mathcal{R}(X, X')$ such that $\text{dis}(R') < 4\varepsilon$, and for each $(x, x') \in R'$ we have $\{x, x'\} \in U$ for some $U \in \mathcal{U}$. In particular, it follows that*

$$d_N((X, \omega_X), (X', \omega_X|_{X' \times X'})) < 2\varepsilon.$$

The first statement in the preceding theorem asserts that we can choose a “well-behaved” correspondence that associates to each point in X a point in X' that belongs to the same element in the ε -system. We omit the proof, as it follows essentially from the proof technique for the next result.

By virtue of Theorem 64, one can always approximate a compact network up to any given precision. The next theorem implies that a sampled network limits to the underlying compact network as the sample gets more and more dense.

Theorem 67 (Limit of dense sampling). *Let (X, ω_X) be a compact network, and let $S = \{s_1, s_2, \dots\}$ be a countable dense subset of X with a fixed enumeration. For each $n \in \mathbb{N}$, let X_n be the finite network with node set $\{s_1, \dots, s_n\}$ and weight function $\omega_X|_{X_n \times X_n}$. Then we have:*

$$d_N(X, X_n) \downarrow 0 \text{ as } n \rightarrow \infty.$$

We now briefly venture into *measure networks*, which are Polish spaces X equipped with a Borel probability measure μ_X and an essentially bounded, measurable weight function $\omega_X : X \times X \rightarrow \mathbb{R}$ (more in §1.9.1). For such a network, it makes sense to ask about an “optimal” ε -system, as in the next definition.

Definition 24. Let (X, ω_X, μ_X) be a measure network. Let \mathcal{U} be any ε -system on X . We define the *minimal mass function*

$$\mathfrak{m}(\mathcal{U}) := \min \{\mu_X(U) : U \in \mathcal{U}, \mu_X(U) > 0\}.$$

Note that \mathfrak{m} returns the minimal non-zero mass of an element in \mathcal{U} .

Next let $\varepsilon > 0$. Define a function $\mathfrak{M}_\varepsilon : \mathcal{CN} \rightarrow (0, 1]$ as follows:

$$\mathfrak{M}_\varepsilon(X) := \sup \{ \mathfrak{m}(\mathcal{U}) : \mathcal{U} \text{ a refined } \varepsilon\text{-system on } X \}.$$

Since \mathcal{U} covers X , we know that the total mass of \mathcal{U} is 1. Thus the set of elements \mathcal{U} with positive mass is nonempty, and so $\mathfrak{m}(\mathcal{U})$ is strictly positive. It follows that $\mathfrak{M}_\varepsilon(X)$ is strictly positive. More is true when μ_X is fully supported on X : given any ε -system \mathcal{U} on X and any $U \in \mathcal{U}$, we automatically have $\mu_X(U) > 0$. To see this, suppose $\mu_X(U) = 0$. Then $U \cap \text{supp}(\mu_X) = \emptyset$, which is a contradiction because $\text{supp}(\mu_X) = X$ and $U \cap X \neq \emptyset$ by our convention for an open cover (i.e. that empty elements are excluded, see §).

In the preceding definition, for a given $\varepsilon > 0$, the function $\mathfrak{M}_\varepsilon(X)$ considers the collection of all *refined* ε -systems on X , and then maximizes the minimal mass of any element in such an ε -system. For an example, consider the setting of Euclidean space \mathbb{R}^d : ε -systems can be constructed using ε -balls, and the mass of an ε -ball scales as ε^d . The functions in Definition 24 are crucial to the next result, which shows that as we sample points from a distribution on a network, the sampled subnetwork converges almost surely to the support of the distribution.

Theorem 68 (Probabilistic network approximation). *Let (X, ω_X) be a network equipped with a Borel probability measure μ_X . For each $i \in \mathbb{N}$, let $x_i : \Omega \rightarrow X$ be an independent random variable defined on some probability space $(\Omega, \mathcal{F}, \mathbb{P})$ with distribution μ_X . For each $n \in \mathbb{N}$, let $\mathbb{X}_n = \{x_1, x_2, \dots, x_n\}$. Let $\varepsilon > 0$. Then we have:*

$$\mathbb{P}\left(\{\omega \in \Omega : d_{\mathcal{N}}(\text{supp}(\mu_X), \mathbb{X}_n(\omega)) \geq \varepsilon\}\right) \leq \frac{(1 - \mathfrak{M}_{\varepsilon/2}(\text{supp}(\mu_X)))^n}{\mathfrak{M}_{\varepsilon/2}(\text{supp}(\mu_X))},$$

where $\mathbb{X}_n(\omega)$ is the subnetwork induced by $\{x_1(\omega), \dots, x_n(\omega)\}$. In particular, the subnetwork \mathbb{X}_n converges almost surely to X in the $d_{\mathcal{N}}$ -sense.

As noted before, the mass of an ε -ball in d -dimensional Euclidean space scales as ε^d . Thus in the setting of Euclidean space \mathbb{R}^d , the quantity on the right would scale as $\varepsilon^{-d}(1 - \varepsilon^d)^n$.

1.6.2 Weak isomorphism and $d_{\mathcal{N}}$

Proofs from this section can be found in §2.3

We now focus on the structure of the zero sets of $d_{\mathcal{N}}$. To proceed in this direction, first notice that a strong isomorphism between two networks (X, ω_X) and (Y, ω_Y) , given by a bijection $f : X \rightarrow Y$, is equivalent to the following condition: there exists a set Z and bijective maps $\varphi_X : Z \rightarrow X, \varphi_Y : Z \rightarrow Y$ such that $\omega_X(\varphi_X(z), \varphi_X(z')) = \omega_Y(\varphi_Y(z), \varphi_Y(z'))$ for each $z, z' \in Z$. To see this, simply let $Z = \{(x, f(x)) : x \in X\}$ and let φ_X, φ_Y be the projection maps on the first and second coordinates, respectively. Based on this observation, we make the next definition.

Definition 25. Let (X, ω_X) and $(Y, \omega_Y) \in \mathcal{N}$. We define X and Y to be *Type I weakly isomorphic*, denoted $X \cong_{\text{I}}^w Y$, if there exists a set Z and surjective maps $\phi_X : Z \rightarrow X$ and $\phi_Y : Z \rightarrow Y$ such that $\omega_X(\phi_X(z), \phi_X(z')) = \omega_Y(\phi_Y(z), \phi_Y(z'))$ for each $z, z' \in Z$.

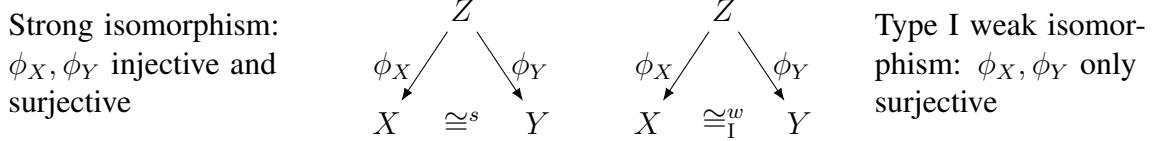


Figure 1.15: Relaxing the requirements on the maps of this “tripod structure” is a natural way to weaken the notion of strong isomorphism.

Notice that Type I weak isomorphism is in fact a *relaxation* of the notion of strong isomorphism. Indeed, if in addition to being surjective, we require the maps ϕ_X and ϕ_Y to be injective, then the strong notion of isomorphism is recovered. In this case, the map $\phi_Y \circ \phi_X^{-1} : X \rightarrow Y$ would be a weight preserving bijection between the networks X and Y . The relaxation of strong isomorphism to a Type I weak isomorphism is illustrated in Figure 1.15. Also observe that the relaxation is *strict*. For example, the networks $X = N_1(1)$ and $Y = N_2(\mathbb{1}_{2 \times 2})$, are weakly but not strongly isomorphic via the map that sends both nodes of Y to the single node of X .

We remark that when dealing with infinite networks, it will turn out that an even weaker notion of isomorphism is required. We define this weakening next.

Definition 26. Let (X, ω_X) and $(Y, \omega_Y) \in \mathcal{N}$. We define X and Y to be *Type II weakly isomorphic*, denoted $X \cong_{\text{II}}^w Y$, if for each $\varepsilon > 0$, there exists a set Z_ε and surjective maps $\phi_X^\varepsilon : Z_\varepsilon \rightarrow X$ and $\phi_Y^\varepsilon : Z_\varepsilon \rightarrow Y$ such that

$$|\omega_X(\phi_X^\varepsilon(z), \phi_X^\varepsilon(z')) - \omega_Y(\phi_Y^\varepsilon(z), \phi_Y^\varepsilon(z'))| < \varepsilon \text{ for all } z, z' \in Z_\varepsilon. \quad (1.4)$$

Remark 69 (Type I isomorphism is stronger than Type II). Let $(X, \omega_X), (Y, \omega_Y) \in \mathcal{CN}$ and suppose $\varphi : X \rightarrow Y$ is a surjective map such that $\omega_X(x, x') = \omega_Y(\varphi(x'), \varphi(x'))$ for all $x, x' \in X$. Then X and Y are Type I weakly isomorphic and hence Type II weakly isomorphic, i.e. $X \cong_{\text{II}}^w Y$. This result follows from Definition 25 by: (1) choosing $Z = X$, and (2) letting ϕ_X be the identity map, and (3) letting $\phi_Y = \varphi$. The converse implication, i.e. that Type I weak isomorphism implies the existence of a surjective map as above, is not true: an example is shown in Figure 1.16.

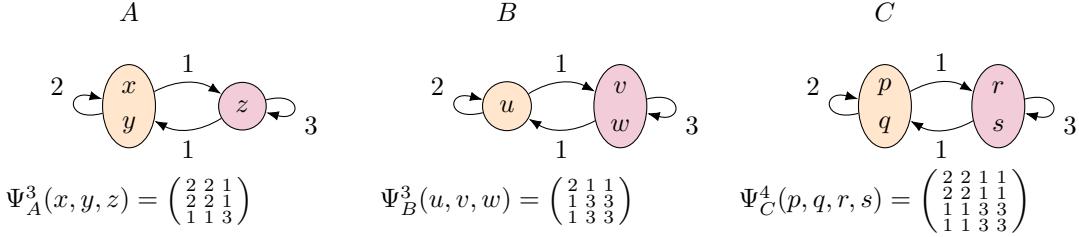


Figure 1.16: Note that Remark 69 does not fully characterize weak isomorphism, even for finite networks: All three networks above, with the given weight matrices, are Type I weakly isomorphic since C maps surjectively onto A and B . But there are no surjective, weight preserving maps $A \rightarrow B$ or $B \rightarrow A$.

It is easy to see that strong isomorphism induces an equivalence class on \mathcal{N} . The same is true for both types of weak isomorphism, and we record this result in the following proposition.

Proposition 70. *Weak isomorphism of Types I and II both induce equivalence relations on \mathcal{N} .*

In the setting of \mathcal{FN} , it is not difficult to show that the two types of weak isomorphism coincide. This is the content of the next proposition. By virtue of this result, there is no ambiguity in dropping the “Type I/II” modifier when saying that two finite networks are weakly isomorphic.

Proposition 71. *Let $X, Y \in \mathcal{FN}$ be finite networks. Then X and Y are Type I weakly isomorphic if and only if they are Type II weakly isomorphic.*

Type I weak isomorphisms will play a vital role in the content of this paper, but for now, we focus on Type II weak isomorphism. The next theorem justifies calling $d_{\mathcal{N}}$ a network distance, and shows that $d_{\mathcal{N}}$ is compatible with Type II weak isomorphism.

Theorem 72. *$d_{\mathcal{N}}$ is a metric on \mathcal{N} modulo Type II weak isomorphism.*

The proof is in §2.1. For finite networks, we immediately obtain:

The restriction of $d_{\mathcal{N}}$ to \mathcal{FN} yields a metric modulo Type I weak isomorphism.

The proof of Proposition 71 will follow from the proof of Theorem 72. In fact, an even stronger result is true: weak isomorphism of Types I and II coincide for compact networks as well.

Theorem 73 (Weak isomorphism in \mathcal{CN}). *Let $X, Y \in \mathcal{CN}$. Then X and Y are Type II weakly isomorphic if and only if X and Y are Type I weakly isomorphic, i.e. there exists a set V and surjections $\varphi_X : V \rightarrow X$, $\varphi_Y : V \rightarrow Y$ such that:*

$$\omega_X(\varphi_X(v), \varphi_X(v')) = \omega_Y(\varphi_Y(v), \varphi_Y(v')) \quad \text{for all } v, v' \in V.$$

1.6.3 An additional axiom coupling weight function with topology

We now explore some additional constraints on the coupling between the topology on a network and its weight function. Using these constraints, we are able to prove that weakly isomorphic networks have, in a particular sense, a strongly isomorphic core. Moreover, weak isomorphism on the whole is guaranteed by having a certain equality of substructures called *motifs*. In particular, this generalizes an observation of Gromov about reconstruction via motif sets in metric spaces [65, 3.27 $\frac{1}{2}$] to the setting of *directed* metric spaces.

First we present a definition that will be used later, and will also help us understand the topological constraints we later impose.

Definition 27 (An equivalence relation and a quotient space). Let $(X, \omega_X) \in \mathcal{N}$. Define the equivalence relation \sim as follows:

$$x \sim x' \text{ iff } \omega_X(x, z) = \omega_X(x', z) \text{ and } \omega_X(z, x) = \omega_X(z, x') \text{ for all } z \in X.$$

Next define $\sigma : X \rightarrow X/\sim$ to be the canonical map sending any $x \in X$ to its equivalence class $[x] \in X/\sim$. Also define $\omega_{X/\sim}([x], [x']) := \omega_X(x, x')$ for $[x], [x'] \in X/\sim$. To check that this map is well-defined, let $a, a' \in X$ be such that $a \sim x$ and $a' \sim x'$. Then,

$$\omega_X(a, a') = \omega_X(x, a') = \omega_X(x, x'),$$

where the first equality holds because $a \sim x$, and the second equality holds because $a' \sim x'$. We equip X/\sim with the quotient topology, i.e. a set is open in X/\sim if and only if its preimage under σ is open in X . Then σ is a surjective, continuous map.

Recall that we often write $x_n \rightarrow x$ to mean that a sequence $(x_n)_{n \in \mathbb{N}}$ in a topological space X is converging to $x \in X$, i.e. any open set containing x contains all but finitely many of the x_n terms. We also often write “ $(x_n)_{n \in \mathbb{N}}$ is eventually inside $A \subseteq X$ ” to mean that $x_n \in A$ for all but finitely many n . Also recall that given a subspace $Z \subseteq X$ equipped with the subspace topology, we say that a particular topological property (e.g. convergence or openness) holds *relative Z* or *rel Z* if it holds in the set Z equipped with the subspace topology. Throughout this section, we use the “relative” terminology extensively as a bookkeeping device to keep track of the subspace with respect to which some topological property holds.

Definition 28. Let $(X, \omega_X) \in \mathcal{N}$. We say that X has a *coherent* topology if the following axioms are satisfied for any subnetwork Z of X equipped with the subspace topology:

A1 (Open sets in a first countable space) A set $A \subseteq Z$ is open rel Z if and only if for any sequence $(x_n)_{n \in \mathbb{N}}$ in Z converging rel Z to a point $x \in A$, there exists $N \in \mathbb{N}$ such that $x_n \in A$ for all $n \geq N$.

A2 (Topological triangle inequality) A sequence $(x_n)_{n \in \mathbb{N}}$ in Z converges rel Z to a point $x \in Z$ if and only if $\omega_X(x_n, \bullet)|_Z \xrightarrow{\text{unif.}} \omega_X(x, \bullet)|_Z$ and $\omega_X(\bullet, x_n)|_Z \xrightarrow{\text{unif.}} \omega_X(\bullet, x)|_Z$.

Axiom **A1** is a characterization of open sets in first countable spaces; we mention it explicitly for easy reference. Axiom **A2** gives a characterization of convergence (and hence of the open sets, via **A1**) in terms of the given weight function. Note that **A2** does not discount the possibility of a sequence converging to non-unique limits, does not force a space to be Hausdorff, and does not force convergent sequences to be Cauchy. The name *topological triangle inequality* is explained in the next remark.

Remark 74 (The “topological triangle inequality”). Consider a metric space (X, d_X) . One key property of such a space is that whenever $d_X(x, x')$ is small, we also have $d_X(x, \bullet) \approx d_X(x', \bullet)$ by the triangle inequality. Said differently, if we have a sequence $(x_n)_n$ and $x_n \rightarrow x$, then $|d_X(x_n, z) - d_X(x, z)| \leq d_X(x, x_n) \rightarrow 0$ for any $z \in X$. Conversely, if $|d_X(x_n, z) - d_X(x, z)| \rightarrow 0$ for all $z \in X$, then by letting $z = x$, we immediately obtain $d_X(x_n, x) \rightarrow 0$.

Axiom **A2** abstracts away this consequence of the triangle inequality into its network formulation. However, there is more subtlety in the definition. First consider the relation \sim . Informally, if we relax the definition of \sim and require “approximate equality” instead of strict equality, we say that $x \sim^\varepsilon x'$ if

$$\omega_X(x, x) \approx \omega_X(x, x') \approx \omega_X(x', x) \approx \omega_X(x', x') \quad (1.5)$$

$$\omega_X(x, z) \approx \omega_X(x', z) \text{ and } \omega_X(z, x) \approx \omega_X(z, x') \text{ for all } z \in X. \quad (1.6)$$

Here the ε decoration on \sim is incorporated into the \approx notation in the obvious way. As we observed earlier, in a metric space, the triangle inequality ensures that (1.5) implies (1.6). More generally, let $x \in X$, let Z be a small ε -ball containing x , and suppose $(x_n)_n$ is a sequence in Z . If Z is small, then (1.5) holds for any (x_n, x) pair in Z and forces (1.6) to hold, not just in Z , but in all of X . This type of local-to-global inference is a consequence of the triangle inequality.

In a network (X, ω_X) , **A2** captures this type of local-to-global inference in a weak sense. Suppose $Z \subseteq X$, $\{x_n\}_n \subseteq Z$, $x \in Z$, and $x_n \rightarrow x$ rel Z . Note that (1.5) does not force (1.6) to hold even in Z , and so we explicitly assume $x_n \rightarrow x$ rel Z , which implicitly assumes (1.6) restricted to Z .

By a fact about convergence in a relative topology, $(x_n)_n$ in Z converges to $x \in Z$ rel Z if and only if it converges rel X . So $x_n \rightarrow x$ rel Z automatically forces $x_n \rightarrow x$ rel X . Thus by **A2**, we know that (1.6) holds, not just in Z , but in all of X . Because **A2** generalizes the triangle inequality in some sense, and relies on properties of the subspace topology, we interpret it as a topological triangle inequality.

Remark 75 (Heredity of coherence). An alternative formulation of a coherent topology—without invoking the “any subnetwork Z of X ” terminology—would be to say that X satisfies **A2**, and that **A2** is *hereditary*, meaning that any subspace also satisfies **A2**. Note that first countability is hereditary, so any subspace of X automatically satisfies **A1**.

One of the reasons for discussing coherent topologies is that it enables us to prove that isometric maps are continuous. This also justifies Axiom **A2** as a fundamental property that we should expect networks to have.

Proposition 76. Let $(X, \omega_X), (Y, \omega_Y)$ be networks with coherent topologies. Suppose $f : X \rightarrow Y$ is a weight-preserving map and $f(X)$ is a subnetwork of Y with the subspace topology. Then f is continuous.

The proof of this result is in §2.4.

Remark 77 (Relation to Kuratowski embedding). In the setting of a metric space (X, d_X) , the map $X \rightarrow C_b(X)$ given by $x \mapsto d_X(x, \bullet)$ is an isometry known as the *Kuratowski embedding*. Here $C_b(X)$ is the space of bounded, continuous functions on X equipped with the uniform norm. Since this is an isometry, we know that $x_n \rightarrow x$ in X iff $d_X(x_n, \bullet) \xrightarrow{\text{unif.}} d_X(x, \bullet)$ in $C_b(X)$.

In the setting of a general network (X, ω_X) , we do not start with a notion of convergence of the form $x_n \rightarrow x$. However, by continuity of ω_X , we are able to use the language of convergence in $C_b(X)$. The intuition behind Axiom A2 is to use convergence in $C_b(X)$ to induce a notion of convergence in X , with the appropriate adjustments needed for the asymmetry of ω_X .

We use the name “coherent” because it was used in the context of describing the coupling between a metric-like function and its topology as far back as in [101].

Remark 78 (Examples of coherent topologies). Let (X, d_X) be a compact metric space. Axioms A1-A2 hold in X by properties of the metric topology and the triangle inequality. Let (Z, d_Z) denote a metric subspace equipped with the restriction of d_X . Any subspace of a first countable space is first countable, so Z is first countable and thus satisfies A1. Axiom A2 holds for Z by the triangle inequality of d_Z . Thus the metric topology on (X, d_X) is coherent.

The network $N_2 \left(\begin{smallmatrix} \alpha & \beta \\ \gamma & \delta \end{smallmatrix} \right)$ where $\alpha, \beta, \gamma, \delta$ are all distinct is a minimal example of an asymmetric network with a coherent topology. In general, for a topology on a finite network to be coherent, it needs to be coarser than the discrete topology. Consider the network $N_2 \left(\begin{smallmatrix} 1 & 1 \\ 1 & 1 \end{smallmatrix} \right)$ on node set $\{p, q\}$. If we assume that the constant sequence (p, p, \dots) converges to q in the sense of Axiom A2, then $\{q\}$ cannot be open for Axiom A1 to be satisfied. However, the trivial topology $\{\emptyset, \{p, q\}\}$ is coherent. More generally, the discrete topology on the skeleton $\text{sk}(X)$ of any finite network X (essentially X / \sim , but defined more precisely in §1.6.4) is coherent.

The directed network with finite reversibility $(\vec{\mathbb{S}}^1, \omega_{\vec{\mathbb{S}}^1, \rho})$ described in §1.3 is a compact, asymmetric network with a coherent topology.

1.6.4 Skeletons, motifs, and motif reconstruction

Proofs from this section can be found in §2.4.

In this section, we provide further details on the structure of the fiber of weakly isomorphic networks. We begin by defining a motif set, which is the network analogue of Gromov’s curvature classes [65, 3.27]. Informally, for each $n \in \mathbb{N}$, the n -motif set is the

collection of $n \times n$ weight matrices obtained from n -tuples of points in X , possibly with repetition. This is made precise next, after introducing some notation. For a sequence $(x_i)_{i=1}^n$ of nodes in a network X , we will denote the associated weight matrix by $((\omega_X(x_i, x_j)))_{i,j=1}^n$. Entry (i, j) of this matrix is simply $\omega_X(x_i, x_j)$.

Definition 29 (Motif set). For each $n \in \mathbb{N}$ and each $(X, \omega_X) \in \mathcal{CN}$, define $\Psi_X^n : X^n \rightarrow \mathbb{R}^{n \times n}$ to be the map $(x_1, \dots, x_n) \mapsto ((\omega_X(x_i, x_j)))_{i,j=1}^n$, where the $(())$ notation refers to the square matrix associated with the sequence. Note that Ψ_X^n is simply a map that sends each sequence of length n to its corresponding weight matrix. Let $\mathcal{C}(\mathbb{R}^{n \times n})$ denote the closed subsets of $\mathbb{R}^{n \times n}$. Then let $M_n : \mathcal{CN} \rightarrow \mathcal{C}(\mathbb{R}^{n \times n})$ denote the map defined by

$$(X, \omega_X) \mapsto \{\Psi_X^n(x_1, \dots, x_n) : x_1, \dots, x_n \in X\}.$$

We refer to $M_n(X)$ as the n -motif set of X . The interpretation is that $M_n(X)$ is a bag containing all the motifs of X that one can form by looking at all subnetworks of size n (with repetitions). Notice that the image of M_n is closed in $\mathbb{R}^{n \times n}$ because each coordinate is the continuous image of the compact set $X \times X$ under ω_X , hence the image of M_n is compact in $\mathbb{R}^{n \times n}$ and hence closed.

It is easy to come up with examples of networks that share the same motif sets, but are not strongly isomorphic. However, as we later show in Theorem 84, weak isomorphism of compact, separable, and coherent networks is precisely characterized by equality of motif sets. Another crucial object for this result is the notion of a skeleton, which we define next.

Definition 30 (Automorphisms). Let $(X, \omega_X) \in \mathcal{CN}$. We define the *automorphisms* of X to be the collection

$$\text{Aut}(X) := \{\varphi : X \rightarrow X : \varphi \text{ a weight preserving bijection}\}.$$

Definition 31 (Poset of weak isomorphism). Let $(X, \omega_X) \in \mathcal{CN}$. Define a set $\mathfrak{p}(X)$ as follows:

$$\mathfrak{p}(X) := \{(Y, \omega_Y) \in \mathcal{CN} : \text{there exists a surjective, weight preserving map } \varphi : X \rightarrow Y\}.$$

Next we define a partial order \preceq on $\mathfrak{p}(X)$ as follows: for any $(Y, \omega_Y), (Z, \omega_Z) \in \mathfrak{p}(X)$,

$$(Y, \omega_Y) \preceq (Z, \omega_Z) \iff \text{there exists a surjective, weight preserving map } \varphi : Z \rightarrow Y.$$

Then the set $\mathfrak{p}(X)$ equipped with \preceq is called the *poset of weak isomorphism* of X .

Definition 32 (Terminal networks in \mathcal{CN}). Let $(X, \omega_X) \in \mathcal{CN}$. A compact network $Z \in \mathfrak{p}(X)$ is *terminal* if:

1. For each $Y \in \mathfrak{p}(X)$, there exists a weight preserving surjection $\varphi : Y \rightarrow Z$.
2. Let $Y \in \mathfrak{p}(X)$. If $f : Y \rightarrow Z$ and $g : Y \rightarrow Z$ are weight preserving surjections, then there exists $\varphi \in \text{Aut}(Z)$ such that $g = \varphi \circ f$.

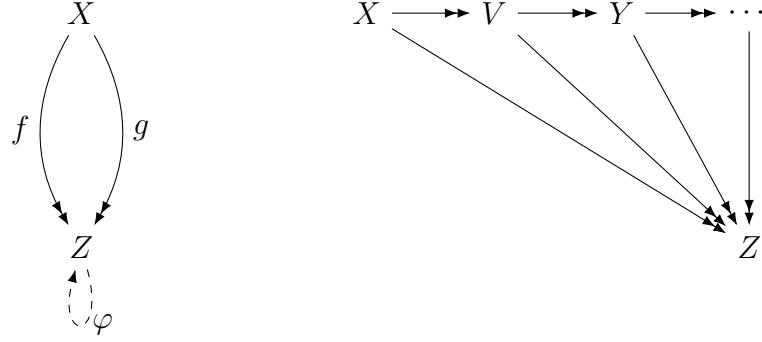


Figure 1.17: **Left:** Z represents a terminal object in $\mathfrak{p}(X)$, and f, g are weight preserving surjections $X \rightarrow Z$. Here $\varphi \in \text{Aut}(Z)$ is such that $g = \varphi \circ f$. **Right:** Here we show more of the poset structure of $\mathfrak{p}(X)$. In this case we have $X \succeq V \succeq Y \dots \succeq Z$.

One of our main results (Theorem 84) shows that two *weakly isomorphic networks* have *strongly isomorphic skeleta*.

A terminal network captures the idea of a minimal substructure of a network. One may ask if anything interesting can be said about superstructures of a network. This motivates the following construction of a “blow-up” network. We provide an illustration in Figure 1.18.

Definition 33. Let (X, ω_X) be any network. Let $\mathbf{k} = (k_x)_{x \in X}$ be a choice of an index set k_x for each node $x \in X$. Consider the network $X[\mathbf{k}]$ with node set $\bigcup_{x \in X} \{(x, i) : i \in k_x\}$ and weights ω given as follows: for $x, x' \in X$ and for $i \in k_x, i' \in k_{x'}$,

$$\omega((x, i), (x', i')) := \omega_X(x, x').$$

The topology on $X[\mathbf{k}]$ is given as follows: the open sets are of the form $\bigcup_{x \in U} \{(x, i) : i \in k_x\}$, where U is open in X . By construction, $X[\mathbf{k}]$ is first countable with respect to this topology. We will call any such $X[\mathbf{k}]$ a *blow-up network* of X .

In a blow-up network of X , each node $x \in X$ is replaced by another network, indexed by k_x . All internal weights of this network are constant and all outgoing weights are preserved from the original network. If X is compact, then so is $X[\mathbf{k}]$.

We also observe that X is weakly isomorphic to any of its blow-ups $Y = X[\mathbf{k}]$. To see this, let $Z = X[\mathbf{k}]$, let $\phi_Y : Z \rightarrow Y$ be the map sending each (x, i) to (x, i) , and let $\phi_X : Z \rightarrow X$ be the map sending each (x, i) to x . Then ϕ_X, ϕ_Y are surjective, weight preserving maps from Z onto X and Y respectively. By Remark 69, we obtain $X \cong^w Y$.

The construction of blow-up networks provides a different perspective on Proposition 12:

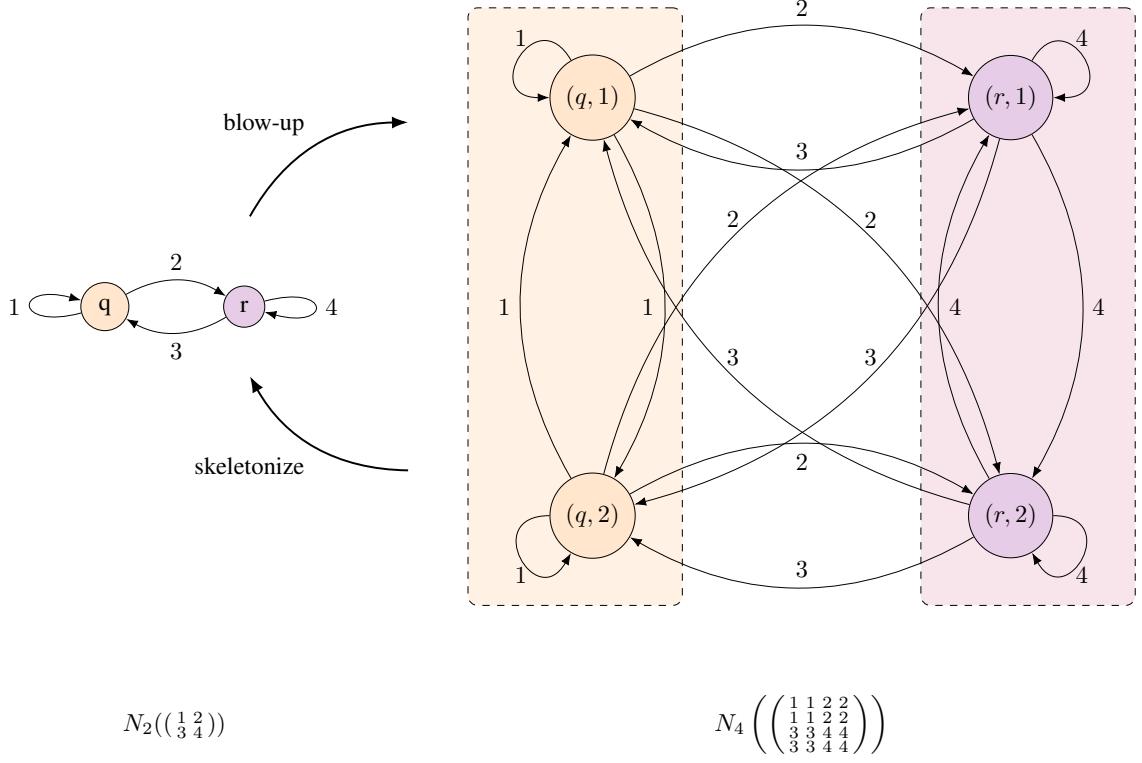


Figure 1.18: Interpolating between the skeleton and blow-up constructions.

Proposition 79 (see also Proposition 12). *Let $(X, \omega_X), (Y, \omega_Y) \in \mathcal{N}$. Then $X \cong^w Y$ if and only if there exist blow-ups X', Y' such that $X' \cong^s Y'$.*

We now define the skeleton of a compact network. Observe that when X is compact, X/\sim is the continuous image of a compact space and so is compact. In general, first countability of a topological space is *not* preserved under a surjective continuous map, but it is preserved when the surjective, continuous map is also open [114, p. 27]. The following proposition gives a sufficient condition on X which will ensure that X/\sim is first countable.

Proposition 80. *Suppose $(X, \omega_X) \in \mathcal{N}$ has a coherent topology. Then the map $\sigma : X \rightarrow X/\sim$ is an open map, i.e. it maps open sets to open sets.*

Definition 34 (The skeleton of a compact network). Suppose $(X, \omega_X) \in \mathcal{CN}$ has a coherent topology. The *skeleton* of X is defined to be $(\text{sk}(X), \omega_{\text{sk}(X)}) \in \mathcal{CN}$, where $\text{sk}(X) := X/\sim$, and

$$\omega_{\text{sk}(X)}([x], [x']) := \omega_X(x, x') \text{ for all } [x], [x'] \in \text{sk}(X).$$

Observe that $\text{sk}(X)$ is compact because X is compact, and first countable by Proposition 80 and the fact that the image of first countable space under an open, surjective, and continuous map is also first countable [114, p. 27]. Furthermore, $\omega_{\text{sk}(X)}$ is well defined by the definition of \sim .

The following proposition shows that skeletons inherit the property of coherence.

Proposition 81. *Let (X, ω_X) be a compact network with a coherent topology. The quotient topology on $(\text{sk}(X), \omega_{\text{sk}(X)})$ is also coherent.*

In addition to coherence, the skeleton has the following useful property.

Proposition 82. *Let (X, ω_X) be a compact network with a coherent topology. Then its skeleton $(\text{sk}(X), \omega_{\text{sk}(X)})$ is Hausdorff.*

Theorem 83 (Skeletons are terminal). *Let $(X, \omega_X) \in \mathcal{CN}$ be such that the topology on X is coherent. Then $(\text{sk}(X), \omega_{\text{sk}(X)}) \in \mathcal{CN}$ is terminal in $\mathfrak{p}(X)$.*

Recall that a topological space is *separable* if it contains a countable dense subset.

Theorem 84. *Suppose $(X, \omega_X), (Y, \omega_Y)$ are separable, compact networks with coherent topologies. Then the following are equivalent:*

1. $X \cong^w Y$.
2. $M_n(X) = M_n(Y)$ for all $n \in \mathbb{N}$.
3. $\text{sk}(X) \cong^s \text{sk}(Y)$.

1.7 Diagrams of compact networks and convergence results

Our aim in this work is to describe the convergence of persistent homology methods applied to network data. When dealing with finite networks, the vector spaces resulting from applying a persistent homology method will necessarily be finite dimensional. However, our setting is that of infinite (more specifically, *compact*) networks, and so we need additional machinery to ensure that our methods output well-defined persistent vector spaces. The following definition and theorem are provided in full detail in [26].

Definition 35 (§2.1, [26]). A persistent vector space $\mathcal{V} = \{V^\delta \xrightarrow{\nu_{\delta, \delta'}} V^{\delta'}\}_{\delta \leq \delta' \in \mathbb{R}}$ is *q-tame* if $\nu_{\delta, \delta'}$ has finite rank whenever $\delta < \delta'$.

Theorem 85 ([26], also [27] Theorem 2.3). *Any q-tame persistent vector space \mathcal{V} has a well-defined persistence diagram $\text{Dgm}(\mathcal{V})$. If \mathcal{U}, \mathcal{V} are ε -interleaved q-tame persistent vector spaces, then $d_B(\text{Dgm}(\mathcal{U}), \text{Dgm}(\mathcal{V})) \leq \varepsilon$.*

As a consequence of developing the notion of ε -systems, we are able to prove the following:

Theorem 86. *Let $(X, \omega_X) \in \mathcal{CN}$, $k \in \mathbb{Z}_+$. Then the persistence vector spaces associated to the Vietoris-Rips, Dowker, and PPH constructions are all q -tame.*

The metric space analogue of Theorem 86 for VR and Čech complexes appeared in [27, Proposition 5.1]; the same proof structure works in the setting of networks after applying our results on approximation via ε -systems.

For this next result, we again refer the reader to §1.9.1 for additional details on measure networks.

Theorem 87 (Convergence). *Let (X, ω_X) be a measure network equipped with a Borel probability measure μ_X . For each $i \in \mathbb{N}$, let $x_i : \Omega \rightarrow X$ be an independent random variable defined on some probability space $(\Omega, \mathcal{F}, \mathbb{P})$ with distribution μ_X . For each $n \in \mathbb{N}$, let $\mathbb{X}_n = \{x_1, x_2, \dots, x_n\}$. Let $\varepsilon > 0$. Then we have:*

$$\mathbb{P}\left(\{\omega \in \Omega : d_B(Dgm^\bullet(\text{supp}(\mu_X)), Dgm^\bullet(\mathbb{X}_n(\omega))) \geq \varepsilon\}\right) \leq \frac{(1 - \mathfrak{M}_{\varepsilon/4}(\text{supp}(\mu_X)))^n}{\mathfrak{M}_{\varepsilon/4}(\text{supp}(\mu_X))},$$

where $\mathbb{X}_n(\omega)$ is the subnetwork induced by $\{x_1(\omega), \dots, x_n(\omega)\}$ and Dgm^\bullet is either of the Vietoris-Rips, Dowker, or PPH diagrams. In particular, either of these three persistent vector spaces of the subnetwork \mathbb{X}_n converges almost surely to that of $\text{supp}(\mu_X)$ in bottleneck distance.

1.8 Completeness, compactness, and geodesics

Proofs from this section are provided in §2.5.

1.8.1 Completeness of $(\mathcal{CN}/\cong^w, d_{\mathcal{N}})$

The following important result further justifies working in \mathcal{CN} :

Theorem 88. *The completion of $(\mathcal{FN}/\cong^w, d_{\mathcal{N}})$ is $(\mathcal{CN}/\cong^w, d_{\mathcal{N}})$.*

The result of Theorem 88 can be summarized as follows:

The limit of a convergent sequence of finite networks is a compact topological space with a continuous weight function.

Completeness of \mathcal{CN}/\cong^w gives us a first useful criterion for convergence of networks. Ideally, we would also want a criterion for convergence along the lines of sequential compactness. In the setting of compact metric spaces, Gromov's Precompactness Theorem implies that the topology induced by the Gromov-Hausdorff distance admits many *pre-compact families* of compact metric spaces (i.e. collections whose closure is compact)

[65, 17, 98]. Any sequence in such a precompact family has a subsequence converging to some limit point of the family. In the next section, we extend these results to the setting of networks. Namely, we show that there are many families of compact networks that are precompact under the metric topology induced by $d_{\mathcal{N}}$.

1.8.2 Precompact families in \mathcal{CN}/\cong^w

We begin this section with some definitions.

Definition 36 (Diameter for networks, [35]). For any network (X, ω_X) , define $\text{diam}(X) := \sup_{x, x' \in X} |\omega_X(x, x')|$. For compact networks, the sup is replaced by max.

Definition 37. A family \mathcal{F} of weak isomorphism classes of compact networks is *uniformly approximable* if: (1) there exists $D \geq 0$ such that for every $[X] \in \mathcal{F}$, we have $\text{diam}(X) \leq D$, and (2) for every $\varepsilon > 0$, there exists $N(\varepsilon) \in \mathbb{N}$ such that for each $[X] \in \mathcal{F}$, there exists a finite network Y satisfying $\text{card}(Y) \leq N(\varepsilon)$ and $d_{\mathcal{N}}(X, Y) < \varepsilon$.

Remark 89. The preceding definition is an analogue of the definition of *uniformly totally bounded* families of compact metric spaces [17, Definition 7.4.13], which is used in formulating the precompactness result in the metric space setting. A family of compact metric spaces is said to be uniformly totally bounded if there exists $D \in \mathbb{R}_+$ such that each space has diameter bounded above by D , and for any $\varepsilon > 0$ there exists $N_\varepsilon \in \mathbb{N}$ such that each space in the family has an ε -net with cardinality bounded above by N_ε . Recall that given a metric space (X, d_X) and $\varepsilon > 0$, a subset $S \subseteq X$ is an ε -net if for any point $x \in X$, we have $B(x, \varepsilon) \cap S \neq \emptyset$. Such an ε -net satisfies the nice property that $d_{\text{GH}}(X, S) < \varepsilon$ [17, 7.3.11]. Thus an ε -net is an ε -approximation of the underlying metric space in the Gromov-Hausdorff distance.

Theorem 90. Let \mathfrak{F} be a uniformly approximable family in \mathcal{CN}/\cong^w . Then \mathfrak{F} is precompact, i.e. any sequence in \mathfrak{F} contains a subsequence that converges in \mathcal{CN}/\cong^w .

1.8.3 Geodesics: existence and explicit examples

Thus far, we have motivated our discussion of compact networks by viewing them as limiting objects of finite networks. By the results of the preceding section, we know that $(\mathcal{CN}/\cong^w, d_{\mathcal{N}})$ is complete and obeys a well-behaved compactness criterion. In this section, we prove that this metric space is also *geodesic*, i.e. any two compact networks can be joined by a rectifiable curve with length equal to the distance between the two networks.

Geodesic spaces can have a variety of practical implications. For example, geodesic spaces that are also complete and locally compact are *proper* (i.e. any closed, bounded subset is compact), by virtue of the Hopf-Rinow theorem [17, §2.5.3]. Any probability measure with finite second moment supported on such a space has a *barycenter* [92, Lemma 3.2], i.e. a “center of mass”. Conceivably, such a result can be applied to a compact,

geodesically convex region of $(\mathcal{CN}/\cong^w, d_{\mathcal{N}})$ to compute an “average” network from a collection of networks. Such a result is of interest in statistical inference, e.g. when one wishes to represent a noisy collection of networks by a single network. Similar results on barycenters of geodesic spaces can be found in [61, 82]. We leave a treatment of this topic from a probabilistic framework as future work, and only use this vignette to motivate the results in this section.

We begin with some definitions.

Definition 38 (Curves and geodesics). A *curve* on \mathcal{N} joining (X, ω_X) to (Y, ω_Y) is any continuous map $\gamma : [0, 1] \rightarrow \mathcal{N}$ such that $\gamma(0) = (X, \omega_X)$ and $\gamma(1) = (Y, \omega_Y)$. We will write *a curve on \mathcal{FN}* (resp. *a curve on \mathcal{CN}*) to mean that the image of γ is contained in \mathcal{FN} (resp. \mathcal{CN}). Such a curve is called a *geodesic* [16, §I.1] between X and Y if for all $s, t \in [0, 1]$ one has:

$$d_{\mathcal{N}}(\gamma(t), \gamma(s)) = |t - s| \cdot d_{\mathcal{N}}(X, Y).$$

A metric space is called a *geodesic space* if any two points can be connected by a geodesic.

The following theorem is a useful result about geodesics:

Theorem 91 ([17], Theorem 2.4.16). *Let (X, d_X) be a complete metric space. If for any $x, x' \in X$ there exists a midpoint z such that $d_X(x, z) = d_X(z, y) = \frac{1}{2}d_X(x, y)$, then X is geodesic.*

As a first step towards showing that \mathcal{CN}/\cong^w is geodesic, we show that the collection of finite networks forms a geodesic space.

Theorem 92. *The metric space $(\mathcal{FN}/\cong^w, d_{\mathcal{N}})$ is a geodesic space. More specifically, let $[X], [Y] \in (\mathcal{FN}/\cong^w, d_{\mathcal{N}})$. Then, for any $R \in \mathcal{R}^{\text{opt}}(X, Y)$, we can construct a geodesic $\gamma_R : [0, 1] \rightarrow \mathcal{FN}/\cong^w$ between $[X]$ and $[Y]$ as follows:*

$$\gamma_R(0) := [(X, \omega_X)], \gamma_R(1) := [(Y, \omega_Y)], \text{ and } \gamma_R(t) := [(R, \omega_{\gamma_R(t)})] \text{ for } t \in (0, 1),$$

where for each $(x, y), (x', y') \in R$ and $t \in (0, 1)$,

$$\omega_{\gamma_R(t)}((x, y), (x', y')) := (1 - t) \cdot \omega_X(x, x') + t \cdot \omega_Y(y, y').$$

A key step in the proof of the preceding theorem is to choose an optimal correspondence between two finite networks. This may not be possible, in general, for compact networks. However, using the additional results on precompactness and completeness of \mathcal{CN}/\cong^w , we are able to obtain the desired geodesic structure in Theorem 93. The proof is similar to the one used by the authors of [73] to prove that the metric space of isometry classes of compact metric spaces endowed with the Gromov-Hausdorff distance is geodesic.

Theorem 93. *The complete metric space $(\mathcal{CN}/\cong^w, d_{\mathcal{N}})$ is geodesic.*

Remark 94. Consider the collection of compact metric spaces endowed with the Gromov-Hausdorff distance. This collection can be viewed as a subspace of $(\mathcal{CN}/\cong^w, d_{\mathcal{N}})$. It is known (via a proof relying on Theorem 91) that this restricted metric space is geodesic [73]. Furthermore, it was proved in [36] that an optimal correspondence always exists in this setting, and that such a correspondence can be used to construct explicit geodesics instead of resorting to Theorem 91. The key technique used in [36] was to take a convergent sequence of increasingly-optimal correspondences, use a result about compact metric spaces called Blaschke's theorem [17, Theorem 7.3.8] to show that the limiting object is closed, and then use metric properties such as the Hausdorff distance to guarantee that this limiting object is indeed a correspondence. A priori, such techniques cannot be readily adapted to the network setting, and while one can obtain a convergent sequence of increasingly-optimal correspondences, the obstruction lies in showing that the limiting object is indeed a correspondence. Thus in our proof of Theorem 93, we resort to the indirect route of using Theorem 91.

While the existence of geodesics is satisfactory, it turns out that in some sense, there are “too many” geodesics in \mathcal{CN} . This problem is already apparent in \mathcal{FM} . As shown in [36], \mathcal{FM} contains both branching and non-unique geodesics. The simultaneous presence of both these types of geodesics precludes the placement of curvature bounds (in the sense of Alexandrov curvature, which is a commonly used notion of curvature in metric geometry) on even \mathcal{FM} . The issue lies in the definition of $d_{\text{GH}}/d_{\mathcal{N}}$: the l^∞ structure of these metrics is what enables the existence of these exotic geodesics. We reproduce these results in the next few sections. First we show a quick lemma.

Lemma 95. *Let (Z, d_Z) be a metric space. Let $S, T \in \mathbb{R}$, with $S < T$, and $\gamma : [S, T] \rightarrow Z$ be a curve such that*

$$d_Z(\gamma(s), \gamma(t)) \leq \frac{|s - t|}{|S - T|} \cdot d_Z(\gamma(S), \gamma(T)), \text{ for all } s, t \in [S, T].$$

Then, in fact,

$$d_Z(\gamma(s), \gamma(t)) = \frac{|s - t|}{|S - T|} \cdot d_Z(\gamma(S), \gamma(T)), \text{ for all } s, t \in [S, T].$$

Proof of Lemma 95. Suppose the inequality is strict. Suppose also that $s \leq t$. Then by the triangle inequality, we obtain:

$$\begin{aligned} d_Z(\gamma(S), \gamma(T)) &\leq d_Z(\gamma(S), \gamma(s)) + d_Z(\gamma(s), \gamma(t)) + d_Z(\gamma(t), \gamma(T)) \\ &< \frac{(s - S) + (t - s) + (T - t)}{T - S} \cdot d_Z(\gamma(S), \gamma(T)). \end{aligned}$$

This is a contradiction. Similarly we get a contradiction for the case $t < s$. This proves the lemma. \square

Deviant geodesics

For any $n \in \mathbb{N}$, let Δ_n denote the n -point discrete space, often called the n -point *unit simplex*. Fix $n \in \mathbb{N}$, $n \geq 2$. We will construct an infinite family of *deviant geodesics* between Δ_1 and Δ_n , named as such because they deviate from the straight-line geodesics given by Theorem 92. As a preliminary step, we describe the straight-line geodesic between Δ_1 and Δ_n of the form given by Theorem 92. Let $\{p\}$ and $\{x_1, \dots, x_n\}$ denote the underlying sets of Δ_1 and Δ_n . There is a unique correspondence $R := \{(p, x_1), \dots, (p, x_n)\}$ between these two sets. According to the setup in Theorem 92, the straight-line geodesic between Δ_1 and Δ_n is then given by the metric spaces $(R, d_{\gamma_R(t)})$, for $t \in (0, 1)$. Here $d_{\gamma_R(t)}((p, x_i), (p, x_j)) = t \cdot d_{\Delta_n}(x_i, x_j) = t$ for each $t \in (0, 1)$ and each $1 \leq i, j \leq n$. This corresponds to the all- t matrix with 0s on the diagonal. Finally, we note that the unique correspondence R necessarily has distortion 1. Thus $d_{\text{GH}}(\Delta_1, \Delta_n) = \frac{1}{2}$.

Now we give the parameters for the construction of a certain family of deviant geodesics between Δ_1 and Δ_n . For any $\alpha \in (0, 1]$ and $t \in [0, 1]$, define

$$f(\alpha, t) := \begin{cases} t\alpha & : 0 \leq t \leq \frac{1}{2} \\ \alpha - t\alpha & : \frac{1}{2} < t \leq 1 \end{cases}$$

Next let m be a positive integer such that $1 \leq m \leq n$, and fix a set

$$X_{n+m} := \{x_1, x_2, x_3, \dots, x_{n+m}\}.$$

Fix $\alpha_1, \dots, \alpha_m \in (0, 1]$. For each $0 \leq t \leq 1$, define the matrix $\delta_t := ((d_{ij}^t))_{i,j=1}^{n+m}$ by:

$$\text{For } 1 \leq i, j \leq n+m, \quad d_{ij}^t := \begin{cases} 0 & : i = j \\ f(\alpha_i, t) & : j - i = n \\ f(\alpha_j, t) & : i - j = n \\ t & : \text{otherwise.} \end{cases}$$

This is a block matrix $\begin{pmatrix} A & B \\ B^T & C \end{pmatrix}$ where A is the $n \times n$ all- t matrix with 0s on the diagonal, C is an $m \times m$ all- t matrix with 0s on the diagonal, and B is the $n \times m$ all- t matrix with $f(\alpha_1, t), f(\alpha_2, t), \dots, f(\alpha_m, t)$ on the diagonal.

We first claim that δ_t is the distance matrix of a pseudometric space. Symmetry is clear. We now check the triangle inequality. In the cases $1 \leq i, j, k \leq n$ and $n+1 \leq i, j, k \leq n+m$, the points x_i, x_j, x_k form the vertices of an equilateral triangle with side length t . Suppose $1 \leq i, j \leq n$ and $n+1 \leq k \leq n+m$. Then the triple x_i, x_j, x_k forms an isosceles triangle with equal longest sides of length t , and a possibly shorter side of length $f(\alpha_i, t)$ (if $|k-i| = n$), $f(\alpha_j, t)$ (if $|k-j| = n$), or just a third equal side with length t in the remaining cases. The case $1 \leq i \leq n, n+1 \leq j, k \leq n+m$ is similar. This verifies the triangle inequality. Also note that δ_t is the distance matrix of a bona fide metric space for $t \in (0, 1)$. For $t = 1$, we identify the points x_i and x_{i-n} , for $n+1 \leq i \leq n+m$, to obtain Δ_n , and for $t = 0$, we identify all points together to obtain Δ_1 . This allows us to define

geodesics between Δ_1 and Δ_n as follows. Let $\vec{\alpha}$ denote the vector $(\alpha_1, \dots, \alpha_m)$. We define a map $\gamma_{\vec{\alpha}} : [0, 1] \rightarrow M$ by writing:

$$\gamma_{\vec{\alpha}}(t) := (X_{n+m}, \delta_t) \quad t \in [0, 1],$$

where we can take quotients at the endpoints as described above.

We now verify that these curves are indeed geodesics. There are three cases: $s, t \in [0, \frac{1}{2}]$, $s, t \in (\frac{1}{2}, 1]$, and $s \in [0, \frac{1}{2}], t \in (\frac{1}{2}, 1]$. By using the diagonal correspondence diag , we check case-by-case that $\text{dis}(\text{diag}) \leq |t - s|$. Thus for any $s, t \in [0, 1]$, we have $d_{\text{GH}}(\gamma_{\vec{\alpha}}(s), \gamma_{\vec{\alpha}}(t)) \leq \frac{1}{2}|t - s| = |t - s| \cdot d_{\text{GH}}(\Delta_1, \Delta_n)$. It follows by Lemma 95 that $\gamma_{\vec{\alpha}}$ is a geodesic between Δ_1 and Δ_n . Furthermore, since $\vec{\alpha} \in (0, 1]^m$ was arbitrary, this holds for any such $\vec{\alpha}$. Thus we have an infinite family of geodesics $\gamma_{\vec{\alpha}} : [0, 1] \rightarrow M$ from Δ_1 to Δ_n .

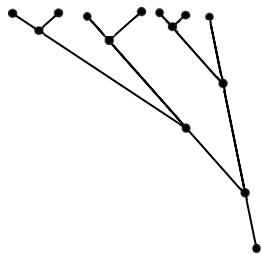
A priori, some of these geodesics may intersect at points other than the endpoints. By this we mean that there may exist $t \in (0, 1)$ and $\vec{\alpha} \neq \vec{\beta} \in (0, 1]^m$ such that $[\gamma_{\vec{\alpha}}(t)] = [\gamma_{\vec{\beta}}(t)]$ in M/\sim . This is related to the branching phenomena that we describe in the next section. For now, we give an infinite subfamily of geodesics that do not intersect each other anywhere except at the endpoints. Recall that the *separation* of a finite metric space (X, d_X) is the smallest positive distance in X , which we denote by $\text{sep}(X)$. If $\text{sep}(X) < \text{sep}(Y)$ for two finite metric spaces X and Y , then $d_{\text{GH}}(X, Y) > 0$.

Let \prec denote the following relation on $(0, 1]^m$: for $\vec{\alpha}, \vec{\beta} \in (0, 1]^m$, set $\vec{\alpha} \prec \vec{\beta}$ if $\alpha_i < \beta_i$ for each $1 \leq i \leq m$. Next let $\vec{\alpha}, \vec{\beta} \in (0, 1]^m$ be such that $\vec{\alpha} \prec \vec{\beta}$. Then $\gamma_{\vec{\beta}}$ is a geodesic from Δ_1 to Δ_n which is distinct (i.e. non-isometric) from $\gamma_{\vec{\alpha}}$ everywhere except at its endpoints. This is because the condition $\vec{\alpha} \prec \vec{\beta}$ guarantees that for each $t \in (0, 1)$, $\text{sep}(\gamma_{\vec{\alpha}}(t)) < \text{sep}(\gamma_{\vec{\beta}}(t))$. Hence $d_{\text{GH}}(\gamma_{\vec{\alpha}}(t), \gamma_{\vec{\beta}}(t)) > 0$ for all $t \in (0, 1)$.

Finally, let $\vec{\alpha} \in (0, 1)^m$, and let $\vec{1}$ denote the all-ones vector of length m . For $\eta \in [0, 1]$, define $\vec{\beta}(\eta) := (1 - \eta)\vec{\alpha} + \eta\vec{1}$. Then by the observations about the relation \prec , $\{\gamma_{\vec{\beta}(\eta)} : \eta \in [0, 1]\}$ is an infinite family of geodesics from Δ_1 to Δ_n that do not intersect each other anywhere except at the endpoints.

Note that one could choose the diameter of Δ_n to be arbitrarily small and still obtain deviant geodesics via the construction above.

Branching geodesics



The structure of d_{GH} permits *branching geodesics*, as illustrated on the right. We use the notation $(a)^+$ for any $a \in \mathbb{R}$ to denote $\max(0, a)$. As above, fix $n \in \mathbb{N}$, $n \geq 2$, and consider the straight-line geodesic between Δ_1 and Δ_n described at the beginning of Section 1.8.3. Throughout this section, we denote this geodesic by $\gamma : [0, 1] \rightarrow M$. We will construct an infinite family of geodesics which branch off from γ . For convenience, we will overload notation and write, for each $t \in [0, 1]$, the distance matrix of $\gamma(t)$ as $\gamma(t)$. Recall from above that $\gamma(t)$ is a symmetric $n \times n$ matrix with the following form:

$$\begin{pmatrix} 0 & t & t & \dots & t \\ 0 & t & \dots & t & \\ \ddots & \dots & \vdots & & \\ & & & & 0 \end{pmatrix}$$

For $t > a_1$, we have $d_{\text{GH}}(\gamma(t), \gamma^{(a_1)}(t)) > 0$, because any correspondence between $\gamma(t), \gamma^{(a_1)}(t)$ has distortion at least $t - a_1$. Thus $\gamma^{(a_1)}$ branches off from γ at a_1 .

The construction of $\gamma^{(a_1)}(t)$ above is a special case of a *one-point metric extension*. Such a construction involves appending an extra row and column to the distance matrix of the starting space; explicit conditions for the entries of the new row and column are stated in [97, Lemma 5.1.22]. In particular, $\gamma^{(a_1)}(t)$ above satisfies these conditions.

Procedurally, the $\gamma^{(a_1)}(t)$ construction can be generalized as follows. Let (\bullet) denote any finite subsequence of $(a_i)_{i \in \mathbb{N}}$. We also allow (\bullet) to be the empty subsequence. Let a_j denote the terminal element in this subsequence. Then for any a_k , $k > j$, we can construct $\gamma^{(\bullet, a_k)}$ as follows:

1. Take the rightmost column of $\gamma^{(\bullet)}(t)$, replace the only 0 by $(t - a_k)^+$, append a 0 at the bottom.
2. Append this column on the right to a copy of $\gamma^{(\bullet)}(t)$.
3. Append the transpose of another copy of this column to the bottom of the newly constructed matrix to make it symmetric.

The objects produced by this construction satisfy the one-point metric extension conditions [97, Lemma 5.1.22], and hence are distance matrices of pseudometric spaces. By taking the appropriate quotients, we obtain valid distance matrices. Symmetry is satisfied by definition, and the triangle inequality is satisfied because any triple of points forms an isosceles triangle with longest sides equal. We write $\Gamma^{(\bullet)}(t)$ to denote the matrix obtained from $\gamma^{(\bullet)}(t)$ after taking quotients. As an example, we obtain the following matrices after taking quotients for $\gamma^{(a_1)}(t)$ above, for $0 \leq t \leq a_1$ (below left) and for $a_1 < t \leq 1$ (below right):

$$\left(\begin{array}{cccc} 0 & t & \dots & t \\ 0 & \dots & t & \\ \ddots & \vdots & & \\ & & & 0 \end{array} \right) \quad \left| \quad \left(\begin{array}{ccccc} 0 & t & \dots & t & t \\ 0 & \dots & t & t & \\ \ddots & \vdots & \vdots & \vdots & \\ 0 & (t - a_1) & & & 0 \end{array} \right) \right.$$

Now let $(a_{i_j})_{j=1}^k$ be any finite subsequence of $(a_i)_{i \in \mathbb{N}}$. For notational convenience, we write $(b_i)_i$ instead of $(a_{i_j})_{j=1}^k$. $\Gamma^{(b_i)_i}$ is a curve in M ; we need to check that it is moreover a geodesic.

Let $s \leq t \in [0, 1]$. Then $\Gamma^{(b_i)_i}(s)$ and $\Gamma^{(b_i)_i}(t)$ are square matrices with $n+p$ and $n+q$ columns, respectively, for nonnegative integers p and q . It is possible that the matrix grows in size between s and t , so we have $q \geq p$. Denote the underlying point set by $\{x_1, x_2, \dots, x_{n+p}, \dots, x_{n+q}\}$. Then define:

$$A := \{(x_i, x_i) : 1 \leq i \leq n+p\}, \quad B := \{(x_{n+p}, x_j) : n+p < j \leq n+q\}, \quad R := A \cup B.$$

Here B is possibly empty. Note that R is a correspondence between $\Gamma^{(b_i)_i}(s)$ and $\Gamma^{(b_i)_i}(t)$, and by direct calculation we have $\text{dis}(R) \leq |t-s|$. Hence we have

$$d_{\text{GH}}(\Gamma^{(b_i)_i}(s), \Gamma^{(b_i)_i}(t)) \leq \frac{1}{2} \cdot |t-s| = |t-s| \cdot d_{\text{GH}}(\Delta_1, \Delta_n).$$

An application of Lemma 95 now shows that $\Gamma^{(b_i)_i}$ is a geodesic.

The finite subsequence $(b_i)_i$ of $(a_i)_{i \in \mathbb{N}}$ was arbitrary. Thus we have an infinite family of geodesics which branch off from γ . Since the increasing sequence $(a_i)_{i \in \mathbb{N}} \in (0, 1)^{\mathbb{N}}$ was arbitrary, the branching could occur at arbitrarily many points along γ .

Remark 96. The existence of branching geodesics shows that $(M/\sim, d_{\text{GH}})$ is *not* an Alexandrov space with curvature bounded below [17, Chapter 10]. Moreover, the existence of deviant (i.e. non-unique) geodesics shows that $(M/\sim, d_{\text{GH}})$ cannot have curvature bounded from above, i.e. $(M/\sim, d_{\text{GH}})$ is not a CAT(k) space for any $k > 0$ [16, Proposition 2.11].

Thus far, we have produced a comprehensive treatment of $d_{\mathcal{N}}$, \mathcal{CN} , and persistent homology methods on \mathcal{CN} that are stable with respect to $d_{\mathcal{N}}$. This concludes the original objective stated at the beginning of §1.

We now deviate slightly and consider the case of *measure networks*, which comprise the network analogue of metric measure spaces. There is some indirect connection between persistent homology methods and measure networks (cf. the convergence results in §1.7), but currently there is no direct analogue to “Gromov-Hausdorff stability of persistent homology” for measure networks. However, the notion of measure networks enables the import of methods other than persistent homology to the setting of network data analysis. Thus we devote the next section to laying the foundations for this topic.

1.9 Measure networks and the $d_{\mathcal{N},p}$ distances

The intuitive idea behind $d_{\mathcal{N}}$ is to search for the best possible alignment of edges (according to weights) while simultaneously aligning nodes. One crucial observation about this setup is that $d_{\mathcal{N}}$ is very sensitive to outliers: the l^∞ structure of $d_{\mathcal{N}}$ is sensitive to even a single outlier. In other words, $d_{\mathcal{N}}$ is not equipped with a method for handling the significance of nodes. Techniques based on optimal transport (OT) provide an elegant solution

to this problem by endowing a network with a probability measure. The user adjusts the measure to signify important network substructures and to smooth out the effect of outliers. This approach was adopted in [70] to compare various real-world network datasets modeled as *metric measure (mm) spaces*—metric spaces equipped with a probability measure. This work was based in turn on the formulation of the *Gromov-Wasserstein (GW) distance* between mm spaces presented in [85, 86].

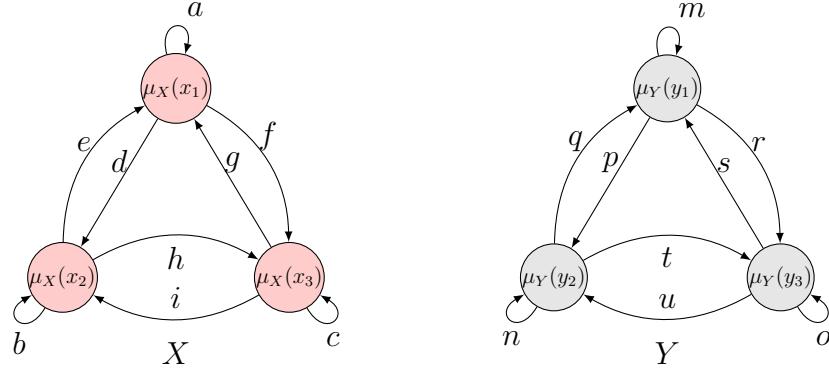


Figure 1.19: Illustrations of the finite networks we consider in this paper. Notice that the edge weights are asymmetric. The numbers in each node correspond to probability masses; for each network, these masses sum to 1.

An alternative definition of the GW distance due to Sturm (the *transportation* formulation) appeared in [116], although this formulation is less amenable to practical computations than the one in [85] (the *distortion* formulation). Both the transportation and distortion formulations were studied carefully in [85, 86, 117]. It was further observed by Sturm in [117] that the definition of the (distortion) GW distance can be extended to *gauged measure spaces* of the form (X, d_X, μ_X) . Here X is a Polish space, d_X is a symmetric L^2 function on $X \times X$ (that does not necessarily satisfy the triangle inequality), and μ_X is a Borel probability measure on X . These results are particularly important in the context of the current paper.

Exact computation of GW distances amounts to solving a nonconvex quadratic program. Towards this end, the computational techniques presented in [85, 86] included both readily-computable lower bounds and an alternate minimization scheme for reaching a local minimum of the GW objection function. This alternate minimization scheme involved solving successive linear optimization problems, and was used for the computations in [70].

From now on, we reserve the term *network* for network datasets that cannot necessarily be represented as metric spaces, unless qualified otherwise. An illustration is provided in

Figure 1.19. Already in [70], it was observed that numerical computation of GW distances between networks worked well for network comparison even when the underlying datasets failed to be metric. This observation was further developed in [100], where the focus from the outset was to define generalized discrepancies between matrices that are not necessarily metric.

On the computational front, the authors of [100] directly attacked the nonconvex optimization problem by considering an *entropy-regularized* form of the GW distance (ERGW) following [111], and using a projected gradient descent algorithm based on results in [10, 111]. This approach was also used (for a generalized GW distance) on graph-structured datasets in [119]. It was pointed out in [119] that the gradient descent approach for the ERGW problem occasionally requires a large amount of regularization to obtain convergence, and that this could possibly lead to over-regularized solutions. A different approach, developed in [85, 86], considers the use of lower bounds on the GW distance as opposed to solving the full GW optimization problem. This is a practical approach for many use cases, in which it may be sufficient to simply obtain lower bounds for the GW distance.

In the current section, we use the GW distance formulation to define and develop a metric structure on the space of measure networks. Additionally, by following the approaches used in [85, 86], we are able to produce quantitatively stable network invariants that produce polynomial-time lower bounds on this network GW distance.

1.9.1 The structure of measure networks

Let X be a Polish space with Borel σ -field denoted by writing $\text{Borel}(X)$, and let μ_X be a Borel probability measure on $\text{Borel}(X)$. We will write $\text{Prob}(X)$ to denote the collection of Borel probability measures supported on X . For each $1 \leq p < \infty$, denote by $L^p(\mu_X)$ the space of μ_X -measurable functions $f : X \rightarrow \mathbb{R}$ such that $|f|^p$ is μ_X -integrable. For $p = \infty$, denote by $L^\infty(\mu_X)$ the space of essentially bounded μ_X -measurable functions, i.e. functions that are bounded except on a set of measure zero. Formally, these spaces are equivalence classes of functions, where functions are equivalent if they agree μ_X -a.e.

We write $\mu_X \otimes \mu_X$ (equivalently $\mu_X^{\otimes 2}$) to denote the product measure on $\text{Borel}(X) \otimes \text{Borel}(X)$ (equivalently $\text{Borel}(X)^{\otimes 2}$). Next let $\omega_X \in L^\infty(\mu_X^{\otimes 2})$. Then ω_X is essentially bounded. Since μ_X is finite, it follows that $|\omega_X|^p$ is integrable for any $1 \leq p < \infty$.

By a *measure network*, we mean a triple (X, ω_X, μ_X) . The naming convention arises from the case when X is finite; in such a case, we can view the pair (X, ω_X) as a complete directed graph with asymmetric real-valued edge weights. Accordingly, the points of X are called *nodes*, pairs of nodes are called *edges*, and ω_X is called the *edge weight function* of X . The collection of all measure networks will be denoted \mathcal{N}_m .

Remark 97. Sturm has studied symmetric, L^2 versions of measure networks (called *gauged measure spaces*) in [117], and we point to his work as an excellent reference on the geometry of such spaces. Our motivation comes from studying networks, hence the difference in our naming conventions.

The information contained in a network should be preserved when we relabel the nodes in a compatible way; we formalize this idea by the following notion of *strong isomorphism* of measure networks.

Definition 39 (Strong isomorphism). To say $(X, \omega_X, \mu_X), (Y, \omega_Y, \mu_Y) \in \mathcal{N}_m$ are *strongly isomorphic* means that there exists a Borel measurable bijection $\varphi : \text{supp}(X) \rightarrow \text{supp}(Y)$ (with Borel measurable inverse φ^{-1}) such that

- $\omega_X(x, x') = \omega_Y(\varphi(x), \varphi(x'))$ for all $x, x' \in \text{supp}(X)$, and
- $\varphi_*\mu_X = \mu_Y$.

We will denote a strong isomorphism between measure networks by $X \cong^s Y$.

Example 98. Networks with one or two nodes will be very instructive in providing examples and counterexamples, so we introduce them now with some special terminology.

- By $N_1(a)$ we will refer to the network with one node $X = \{p\}$, a weight $\omega_X(p, p) = a$, and the Dirac measure $\delta_p = \mathbb{1}_p$.
- By $N_2((\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}), \alpha, \beta)$ we will mean a two-node network with node set $X = \{p, q\}$, and weights and measures given as follows:

$$\begin{aligned}\omega_X(p, p) &= a & \mu_X(\{p\}) &= \alpha \\ \omega_X(p, q) &= b & \mu_X(\{q\}) &= \beta \\ \omega_X(q, p) &= c \\ \omega_X(q, q) &= d\end{aligned}$$

- Given a k -by- k matrix $\Sigma \in \mathbb{R}^{k \times k}$ and a $k \times 1$ vector $v \in \mathbb{R}_+^k$ with sum 1, we automatically obtain a network on k nodes that we denote as $N_k(\Sigma, v)$. Notice that $N_k(\Sigma, v) \cong_1 N_\ell(\Sigma', v')$ if and only if $k = \ell$ and there exists a permutation matrix P of size k such that $\Sigma' = P \Sigma P^T$ and $PA = A'$.

Notation. Even though μ_X takes sets as its argument, we will often omit the curly braces and use $\mu_X(p, q, r)$ to mean $\mu_X(\{p, q, r\})$.

We wish to define a notion of distance on \mathcal{N}_m that is compatible with isomorphism. A natural analog is the Gromov-Wasserstein distance defined between metric measure spaces [85]. To adapt that definition for our needs, we first recall the definition of a measure coupling.

1.9.2 Couplings and the distortion functional

Let $(X, \omega_X, \mu_X), (Y, \omega_Y, \mu_Y)$ be two measure networks. A *coupling* between these two networks is a probability measure μ on $X \times Y$ with marginals μ_X and μ_Y , respectively. Stated differently, couplings satisfy the following property:

$$\mu(A \times Y) = \mu_X(A) \text{ and } \mu(X \times B) = \mu_Y(B), \text{ for all } A \in \text{Borel}(X) \text{ and } B \in \text{Borel}(Y).$$

The collection of all couplings between (X, ω_X, μ_X) and (Y, ω_Y, μ_Y) will be denoted $\mathcal{C}(\mu_X, \mu_Y)$, abbreviated to \mathcal{C} when the context is clear.

In the case where we have a coupling μ between two measures ν, ν' on the same network (X, ω_X) , the quantity $\mu(A \times B)$ is interpreted as the amount of mass transported from A to B when interpolating between the two distributions ν and ν' . In this special case, a coupling is also referred to as a *transport plan*.

Here we also recall that the product σ -field on $X \times Y$, denoted $\text{Borel}(X) \otimes \text{Borel}(Y)$, is defined as the σ -field generated by the measurable rectangles $A \times B$, where $A \in \text{Borel}(X)$ and $B \in \text{Borel}(Y)$. Because our spaces are all Polish, we always have $\text{Borel}(X \times Y) = \text{Borel}(X) \otimes \text{Borel}(Y)$ [13, Lemma 6.4.2].

The product measure $\mu_X \otimes \mu_Y$ is defined on the measurable rectangles by writing

$$\mu_X \otimes \mu_Y(A \times B) := \mu_X(A)\mu_Y(B), \text{ for all } A \in \text{Borel}(X) \text{ and for all } B \in \text{Borel}(Y).$$

By a consequence of Fubini's theorem and the π - λ theorem, the property above uniquely defines the product measure $\mu_X \otimes \mu_Y$ among measures on $\text{Borel}(X \times Y)$.

Example 99 (Product coupling). Let $(X, \omega_X, \mu_X), (Y, \omega_Y, \mu_Y) \in \mathcal{N}_m$. The set $\mathcal{C}(\mu_X, \mu_Y)$ is always nonempty, because the *product measure* $\mu := \mu_X \otimes \mu_Y$ is always a coupling between μ_X and μ_Y .

Example 100 (1-point coupling). Let X be a set, and let $Y = \{p\}$ be the set with one point. Then for any probability measure μ_X on X there is a unique coupling $\mu = \mu_X \otimes \delta_p$ between μ_X and δ_p . To see this, first we check that μ as defined above is a coupling. Let $A \in \text{Borel}(X)$. Then $\mu(A \times Y) = \mu_X(A)\delta_p(Y) = \mu_X(A)$, and similarly $\mu(X \times \{p\}) = \mu_X(X)\delta_p(\{p\}) = \delta_p(\{p\})$. Thus $\mu \in \mathcal{C}(X, Y)$. For uniqueness, let ν be another coupling. It suffices to show that ν agrees with μ on the measurable rectangles. Let $A \in \text{Borel}(X)$, and observe that

$$\nu(A \times \{p\}) = (\pi_X)_*\nu(A) = \mu_X(A) = \mu_X(A)\delta_p(\{p\}) = \mu(A \times \{p\}).$$

On the other hand, $\nu(A \times \emptyset) \leq \nu(X \times \emptyset) = (\pi_Y)_*\nu(\emptyset) = 0 = \mu_X(A)\delta_p(\emptyset) = \mu(A \times \emptyset)$.

Thus ν satisfies the property $\nu(A \times B) = \mu_X(A)\delta_p(B)$. Thus by uniqueness of the product measure, $\nu = \mu_X \otimes \delta_p$. Finally, note that we can endow X and Y with weight functions ω_X and ω_Y , thus adapting this example to the case of networks.

Example 101 (Diagonal coupling). Let $(X, \omega_X, \mu_X) \in \mathcal{N}_m$. The *diagonal coupling* between μ_X and itself is defined by writing

$$\Delta(A \times B) := \int_{X \times X} \mathbb{1}_{A \times B}(x, x') d\mu_X(x) d\delta_x(x') \quad \text{for all } A \in \text{Borel}(X), B \in \text{Borel}(Y).$$

To see that this is a coupling, let $A \in \text{Borel}(X)$. Then,

$$\Delta(A \times X) = \int_{X \times X} \mathbb{1}_{A \times X}(x, x') d\mu_X(x) d\delta_x(x') = \int_X \mathbb{1}_A(x) d\mu_X(x) = \mu_X(A),$$

and similarly $\Delta(X \times A) = \mu_X(A)$. Thus $\Delta \in \mathcal{C}(\mu_X, \mu_X)$.

Now we turn to the notion of the *distortion* of a coupling. Let $(X, \omega_X, \mu_X), (Y, \omega_Y, \mu_Y)$ be two measure networks. For convenience, we define the function

$$\Omega_{X,Y} : (X \times Y)^2 \rightarrow \mathbb{R} \text{ by writing } (x, y, x', y') \mapsto \omega_X(x, x') - \omega_Y(y, y').$$

Next let $\mu \in \mathcal{C}(\mu_X, \mu_Y)$, and consider the probability space $(X \times Y)^2$ equipped with the product measure $\mu \otimes \mu$. For each $p \in [1, \infty)$ the p -distortion of μ is defined as:

$$\begin{aligned} \text{dis}_p(\mu) &= \left(\int_{X \times Y} \int_{X \times Y} |\omega_X(x, x') - \omega_Y(y, y')|^p d\mu(x, y) d\mu(x', y') \right)^{1/p} \\ &= \|\Omega_{X,Y}\|_{L^p(\mu \otimes \mu)}. \end{aligned}$$

For $p = \infty$, the p -distortion is defined as:

$$\text{dis}_p(\mu) := \sup\{|\omega_X(x, x') - \omega_Y(y, y')| : (x, y), (x', y') \in \text{supp}(\mu)\}.$$

When the context is clear, we will often write $\|f\|_p$ to denote $\|f\|_{L^p(\mu \otimes \mu)}$.

1.9.3 Interval representation and continuity of distortion

We now record some standard results about Polish spaces (see also [117, §1.3]). Recall that for a measure space (X, \mathcal{F}, μ) , an *atom* is an element $A \in \mathcal{F}$ such that $0 < \mu(A) < \infty$ and for every $B \in \mathcal{F}$ such that $B \subseteq A$, we have $\mu(B) = 0$ or $\mu(B) = \mu(A)$. In our network setting, the atoms are singletons. To see this, let $(X, \omega_X, \mu_X) \in \mathcal{N}_m$. The underlying measurable space consists of the Polish space X and its Borel σ -field. Because the topology on X is just the metric topology for a suitable metric, we can use standard techniques involving intersections of elements of covers to show that any atom is necessarily a singleton. Next, since μ_X is a finite measure, $\text{Borel}(X)$ can have at most countably many atoms. In particular, μ_X can be decomposed as the sum of a countable number of atomic (Dirac) measures and a nonatomic measure [74]:

$$\mu_X = \sum_{i=1}^{\infty} c_i \delta_{x_i} + \mu'_X, \quad x_i \in X, c_i \in [0, 1] \text{ for each } i \in \mathbb{N}.$$

In what follows, we follow the presentation in [117]. Since X is Polish, it can be viewed as a *standard Borel space* [113] and therefore as the pushforward of Lebesgue measure on the unit interval I . More specifically, let $C_0 = 0$, write $C_i = \sum_{j=1}^i c_j$ for $i \in \mathbb{N} \cup \{\infty\}$, $I' = [C_\infty, 1]$, and $X' = \text{supp}(\mu'_X)$. Now X' is a standard Borel space equipped with a nonatomic measure, so by [113, Theorem 3.4.23], there is a Borel isomorphism $\rho' : I' \rightarrow X'$ such that $\mu'_X = \rho'_* \lambda_{I'}$, where $\lambda_{I'}$ denotes Lebesgue measure restricted to I' . Define the representation map $\rho : I \rightarrow X$ as follows:

$$\rho([C_{i-1}, C_i)) := \{x_i\} \text{ for all } i \in \mathbb{N}, \quad \rho|_{[C_\infty, 1]} := \rho'.$$

The map ρ' is not necessarily unique, and therefore neither is ρ . Any such map ρ is called a *parametrization* of X . In particular, we have $\mu_X = \rho_* \lambda_I$.

The benefit of this construction is that it allows us to represent the underlying measurable space of a network via the unit interval I . Moreover, by taking the pullback of ω_X via ρ , we obtain a network $(I, \rho^* \omega_X, \lambda_I)$. As we will see in the next section, this permits the strategy of proving results over I and transporting them back to X using ρ .

Remark 102 (A 0-distortion coupling between a space and its interval representation). Let $(X, \omega_X, \mu_X) \in \mathcal{N}_m$, and let $(I, \rho^* \omega_X, \lambda_I)$ be an interval representation of X for some parametrization ρ . Consider the map $(\rho, \text{id}) : I \rightarrow X \times I$ given by $i \mapsto (\rho(i), i)$. Define $\mu := (\rho, \text{id})_* \lambda_I$. Let $A \in \text{Borel}(X)$ and $B \in \text{Borel}(I)$. Then $\mu(A \times I) = \lambda_I(\{j \in I : \rho(j) \in A\}) = \mu_X(A)$. Also, $\mu(X \times B) = \lambda_I(\{j \in B : \rho(j) \in X\}) = \lambda_I(B)$. Thus μ is a coupling between μ_X and λ_I . Moreover, for any $A \in \text{Borel}(X)$ and any $B \in \text{Borel}(I)$, if for each $j \in B$ we have $\rho(j) \notin A$, then we have $\mu(A \times B) = 0$. In particular, $\mu(A \times B) = \mu((A \cap \rho(B)) \times B)$. Also, given $(x, i) \in X \times I$, we have that $\rho(i) \neq x$ implies $(x, i) \notin \text{supp}(\mu)$.

Let $1 \leq p < \infty$. For convenience, define $\omega_I := \rho^* \omega_X$. An explicit computation of $\text{dis}_p(\mu)$ shows:

$$\begin{aligned} \text{dis}_p(\mu)^p &= \int_{X \times I} \int_{X \times I} |\omega_X(x, x') - \omega_I(i, i')|^p d\mu(x, i) d\mu(x', i') \\ &= \int_I \int_I |\omega_X(\rho(i), \rho(i')) - \omega_I(i, i')|^p d\lambda_I(i) d\lambda_I(i') \\ &= 0. \end{aligned}$$

For $p = \infty$, we have:

$$\begin{aligned} \text{dis}_\infty(\mu) &= \sup\{|\omega_X(x, x') - \omega_I(i, i')| : (x, i), (x', i') \in \text{supp}(\mu)\} \\ &= \sup\{|\omega_X(\rho(i), \rho(i')) - \omega_I(i, i')| : i, i' \in \text{supp}(\lambda)\} \\ &= 0. \end{aligned}$$

1.9.4 Optimality of couplings in the network setting

We now collect some results about probability spaces. Let X be a Polish space. A subset $P \subseteq \text{Prob}(X)$ is said to be *tight* if for all $\varepsilon > 0$, there is a compact subset $K_\varepsilon \subseteq X$ such that $\mu_X(X \setminus K_\varepsilon) \leq \varepsilon$ for all $\mu_X \in P$.

A sequence $(\mu_n)_{n \in \mathbb{N}} \in \text{Prob}(X)^{\mathbb{N}}$ is said to *converge narrowly* to $\mu_X \in \text{Prob}(X)$ if

$$\lim_{n \rightarrow \infty} \int_X f d\mu_n = \int_X f d\mu_X \quad \text{for all } f \in C_b(X),$$

the space of continuous, bounded, real-valued functions on X . Narrow convergence is induced by a distance [5, Remark 5.1.1], hence the convergent sequences in $\text{Prob}(X)$ completely determine a topology on $\text{Prob}(X)$. This topology on $\text{Prob}(X)$ is called the *narrow topology*. In some references [117], narrow convergence (resp. narrow topology) is called *weak convergence* (resp. *weak topology*).

A further consequence of having a metric on $\text{Prob}(X)$ [5, Remark 5.1.1] is that singletons are closed. This simple fact will be used below.

Theorem 103 (Prokhorov, [5] Theorem 5.1.3). *Let X be a Polish space. Then $P \subseteq \text{Prob}(X)$ is tight if and only if it is relatively compact, i.e. its closure is compact in $\text{Prob}(X)$.*

Lemma 104 (Lemma 4.4, [121]). *Let X, Y be two Polish spaces, and let $P_X \subseteq \text{Prob}(X)$, $P_Y \subseteq \text{Prob}(Y)$ be tight in their respective spaces. Then the set $\mathcal{C}(P_X, P_Y) \subseteq \text{Prob}(X \times Y)$ of couplings with marginals in P_X and P_Y is tight in $\text{Prob}(X \times Y)$.*

Lemma 105 (Compactness of couplings; Lemma 1.2, [117]). *Let X, Y be two Polish spaces. Let $\mu_X \in \text{Prob}(X)$, $\mu_Y \in \text{Prob}(Y)$. Then $\mathcal{C}(X, Y)$ is compact in $\text{Prob}(X \times Y)$.*

Proof. The singletons $\{\mu_X\}$, $\{\mu_Y\}$ are closed and of course compact in $\text{Prob}(X)$, $\text{Prob}(Y)$. Hence by Prokhorov's theorem, they are tight. Now consider $\mathcal{C}(X, Y) \subseteq \text{Prob}(X \times Y)$. Since this is obtained by intersecting the preimages of the continuous projections onto the marginals μ_X and μ_Y , we know that it is closed. Furthermore, $\mathcal{C}(X, Y)$ is tight by Lemma 104. Then by another application of Prokhorov's theorem, it is compact. \square

The following lemma appeared for the L^2 case in [117].

Lemma 106 (Continuity of the distortion functional on intervals). *Let $1 \leq p < \infty$, and let (I, σ_X, λ_I) , $(I, \sigma_Y, \lambda_I) \in \mathcal{N}_m$. The distortion functional dis_p is continuous on $\mathcal{C}(\lambda_I, \lambda_I) \subseteq \text{Prob}(I \times I)$. For $p = \infty$, dis_∞ is lower semicontinuous.*

The next lemma is standard.

Lemma 107 (Gluing lemma, Lemma 1.4 in [117], also Lemma 7.6 in [120]). *Let $\mu_1, \mu_2, \dots, \mu_k$ be probability measures supported on Polish spaces X_1, \dots, X_k . For each $i \in \{1, \dots, k-1\}$, let $\mu_{i,i+1} \in \mathcal{C}(\mu_i, \mu_{i+1})$. Then there exists $\mu \in \text{Prob}(X_1 \times X_2 \times \dots \times X_k)$ with marginals $\mu_{i,i+1}$ on $X_i \times X_{i+1}$ for each $i \in \{1, \dots, k-1\}$.*

1.9.5 The Network Gromov-Wasserstein distance

For each $p \in [1, \infty]$, we define:

$$d_{\mathcal{N},p}(X, Y) := \frac{1}{2} \inf_{\mu \in \mathcal{C}(\mu_X, \mu_Y)} \text{dis}_p(\mu) \quad \text{for each } (X, \omega_X, \mu_X), (Y, \omega_Y, \mu_Y) \in \mathcal{N}_m.$$

As we will see below, $d_{\mathcal{N},p}$ is a legitimate pseudometric on \mathcal{N}_m . The structure of $d_{\mathcal{N},p}$ is analogous to a formulation of the *Gromov-Wasserstein distance* between metric measure spaces [86, 117].

Remark 108 (Boundedness of $d_{\mathcal{N},p}$). Recall from Example 99 that for any $X, Y \in \mathcal{N}_m$, $\mathcal{C}(\mu_X, \mu_Y)$ always contains the product coupling, and is thus nonempty. A consequence is that $d_{\mathcal{N},p}(X, Y)$ is bounded for any $p \in [1, \infty]$. Indeed, by taking the product coupling $\mu := \mu_X \otimes \mu_Y$ we have

$$d_{\mathcal{N},p}(X, Y) \leq \frac{1}{2} \text{dis}_p(\mu).$$

Suppose first that $p \in [1, \infty)$. Applying Minkowski's inequality, we obtain:

$$\begin{aligned} \text{dis}_p(\mu) &= \|\omega_X - \omega_Y\|_{L^p(\mu \otimes \mu)} \\ &\leq \|\omega_X\|_{L^p(\mu \otimes \mu)} + \|\omega_Y\|_{L^p(\mu \otimes \mu)} \\ &= \left(\int_{X \times Y} \int_{X \times Y} |\omega_X(x, x')|^p d\mu(x, y) d\mu(x', y') \right)^{1/p} \\ &\quad + \left(\int_{X \times Y} \int_{X \times Y} |\omega_Y(y, y')|^p d\mu(x, y) d\mu(x', y') \right)^{1/p} \\ &= \left(\int_X \int_X |\omega_X(x, x')|^p d\mu_X(x) d\mu_X(x') \right)^{1/p} \\ &\quad + \left(\int_Y \int_Y |\omega_Y(y, y')|^p d\mu_Y(y) d\mu_Y(y') \right)^{1/p} \\ &= \|\omega_X\|_{L^p(\mu_X \otimes \mu_X)} + \|\omega_Y\|_{L^p(\mu_Y \otimes \mu_Y)} < \infty. \end{aligned}$$

The case $p = \infty$ case is analogous, except that integrals are replaced by taking essential suprema as needed.

In some simple cases, we obtain explicit formulas for computing $d_{\mathcal{N},p}$.

Example 109 (Easy examples of $d_{\mathcal{N},p}$). Let $a, b \in \mathbb{R}$ and consider the networks $N_1(a)$ and $N_1(b)$. The unique coupling between the two networks is the product measure $\mu = \delta_x \otimes \delta_y$, where we understand x, y to be the nodes of the two networks. Then for any $p \in [1, \infty]$, we obtain:

$$\text{dis}_p(\mu) = |\omega_{N_1(a)}(x, x) - \omega_{N_1(b)}(y, y)| = |a - b|.$$

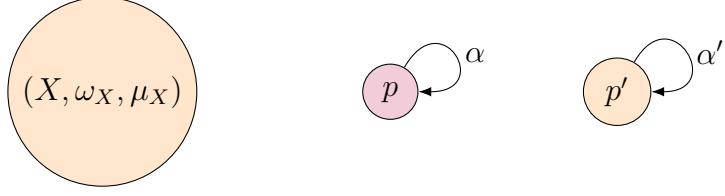


Figure 1.20: The $d_{N,p}$ distance between the two one-node networks is simply $\frac{1}{2}|\alpha - \alpha'|$. In Example 109 we give an explicit formula for computing $d_{N,p}$ between an arbitrary network and a one-node network.

Thus $d_{N,p}(N_1(a), N_1(b)) = \frac{1}{2}|a - b|$.

Let $(X, \omega_X, \mu_X) \in \mathcal{N}_m$ be any network and let $N_1(a) = (\{y\}, a)$ be a network with one node. Once again, there is a unique coupling $\mu = \mu_X \otimes \delta_y$ between the two networks. For any $p \in [1, \infty)$, we obtain:

$$d_{N,p}(X, N_1(a)) = \frac{1}{2} \text{dis}_p(\mu) = \frac{1}{2} \left(\int_X \int_X |\omega_X(x, x') - a|^p d\mu_X(x) d\mu_X(x') \right)^{1/p}.$$

For $p = \infty$, we have $d_{N,p}(X, N_1(a)) = \sup\{\frac{1}{2}|\omega_X(x, x') - a| : x, x' \in \text{supp}(\mu_X)\}$.

Remark 110. $d_{N,p}$ is not necessarily a metric modulo strong isomorphism. Let $X = \{x_1, x_2, x_3\}$ and $Y = \{y_1, y_2, y_3\}$. Consider a coupling μ given as:

$$\mu = \begin{matrix} & y_1 & y_2 & y_3 \\ x_1 & 1/3 & 0 & 0 \\ x_2 & 1/6 & 0 & 0 \\ x_3 & 0 & 1/6 & 1/3 \end{matrix}.$$

Next equip X and Y with edge weights $\{e, f, g, h\}$ as in Figure 1.21.

Comparing the edge weights, it is clear that X and Y are not strongly isomorphic. However, $d_{N,p}(X, Y) = 0$ for all $p \in [1, \infty]$. To see this, define:

$$G = \{(x_1, y_1), (x_2, y_1), (x_3, y_2), (x_3, y_3)\}$$

Then G contains all the points with positive μ -measure. Given any two points $(x, y), (x', y') \in G$, we observe that $|\omega_X(x, x') - \omega_Y(y, y')| = 0$. Thus for any $p \in [1, \infty]$, $\text{dis}_p(\mu) = 0$, and so $d_{N,p}(X, Y) = 0$.

The definition of $d_{N,p}$ is sensible in the sense that it captures the notion of a distance:

Theorem 111. For each $p \in [1, \infty]$, $d_{N,p}$ is a pseudometric on \mathcal{N}_m .

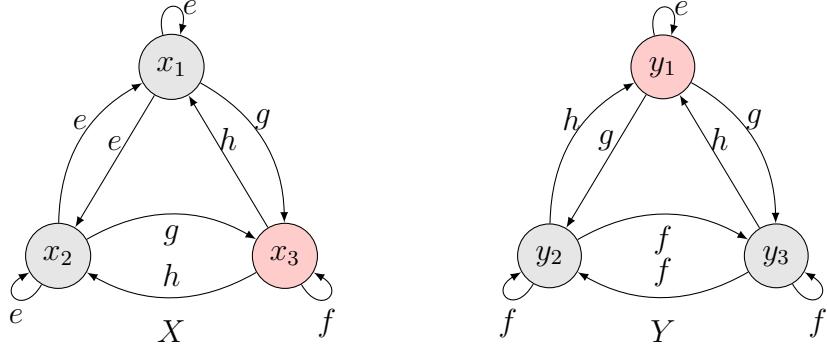


Figure 1.21: Networks at $d_{\mathcal{N},p}$ -distance zero which are not strongly isomorphic.

By the next result, this infimum is actually attained. Hence we may write:

$$d_{\mathcal{N},p}(X, Y) := \frac{1}{2} \min_{\mu \in \mathcal{C}(X, Y)} \text{dis}_p(\mu).$$

Definition 40 (Optimal couplings). Let $(X, \omega_X, \mu_X), (Y, \omega_Y, \mu_Y) \in \mathcal{N}_m$, and let $p \in [1, \infty)$. A coupling $\mu \in \mathcal{C}(\mu_X, \mu_Y)$ is *optimal* if $\text{dis}_p(\mu) = 2d_{\mathcal{N},p}(X, Y)$.

The next result stands out in contrast to the case for $d_{\mathcal{N}}$: whereas we do not have results about optimality of $d_{\mathcal{N}}$, the following result comes relatively easily by virtue of Prokhorov's lemma.

Theorem 112. Let (X, ω_X, μ_X) and (Y, ω_Y, μ_Y) be two measure networks, and let $p \in [1, \infty]$. Then there exists an optimal coupling, i.e. a minimizer for $\text{dis}_p(\cdot)$ in $\mathcal{C}(\mu_X, \mu_Y)$.

It remains to discuss the precise pseudometric structure of $d_{\mathcal{N},p}$. The following definition is a relaxation of strong isomorphism.

Definition 41 (Weak isomorphism). $(X, \omega_X, \mu_X), (Y, \omega_Y, \mu_Y) \in \mathcal{N}$ are *weakly isomorphic*, denoted $X \cong^w Y$, if there exists a Borel probability space (Z, μ_Z) with measurable maps $f : Z \rightarrow X$ and $g : Z \rightarrow Y$ such that

- $f_*\mu_Z = \mu_X$, $g_*\mu_Z = \mu_Y$, and
- $\|f^*\omega_X - g^*\omega_Y\|_\infty = 0$.

Here $f^*\omega_X : Z \times Z \rightarrow \mathbb{R}$ is the pullback weight function given by the map $(z, z') \mapsto \omega_X(f(z), f(z'))$. The map $g^*\omega_Y$ is defined analogously. For the definition to make sense, we need to check that $f^*\omega_X$ is measurable. Let $(a, b) \in \text{Borel}(\mathbb{R})$. Then $B := \{\omega_X \in$

$(a, b)\}$ is measurable because ω_X is measurable. Because f is measurable, we know that $(f, f) : Z \times Z \rightarrow X \times X$ is measurable. Thus $A := (f, f)^{-1}(B)$ is measurable. Now we write:

$$\begin{aligned} A &= \{(z, z') \in Z^2 : ((f(z), f(z')) \in B\} \\ &= \{(z, z') \in Z^2 : \omega_X(f(z), f(z')) \in (a, b)\} \\ &= (f^*\omega_X)^{-1}(a, b). \end{aligned}$$

Thus $f^*\omega_X$ is measurable. Similarly, we verify that $g^*\omega_Y$ is measurable.

Theorem 113 (Pseudometric structure of $d_{\mathcal{N},p}$). *Let $(X, \omega_X, \mu_X), (Y, \omega_Y, \mu_Y) \in \mathcal{N}$, and let $p \in [1, \infty]$. Then $d_{\mathcal{N},p}(X, Y) = 0$ if and only if $X \cong^w Y$.*

Remark 114. Theorem 113 is in the same spirit as related results for gauged measure spaces [117] and for networks under $d_{\mathcal{N}}$, as discussed earlier. The “tripod structure” $X \leftarrow Z \rightarrow Y$ described above is much more difficult to obtain in the setting of $d_{\mathcal{N}}$.

In the next section we follow a brief diversion to study a *Gromov-Prokhorov* distance between measure networks. While it is not the main focus of the current paper, it turns out to be useful for the notion of *interleaving stability* that we define in §1.9.7.

1.9.6 The Network Gromov-Prokhorov distance

Let $\alpha \in [0, \infty)$. For any $(X, \omega_X, \mu_X), (Y, \omega_Y, \mu_Y) \in \mathcal{N}_m$, we write $\mathcal{C} := \mathcal{C}(\mu_X, \mu_Y)$ and define:

$$\begin{aligned} d_{\mathcal{N},\alpha}^{\mathcal{GP}}(X, Y) &:= \frac{1}{2} \inf_{\mu \in \mathcal{C}} \inf \{\varepsilon > 0 : \\ &\quad \mu \otimes \mu (\{x, y, x', y' \in (X \times Y)^2 : |\omega_X(x, x') - \omega_Y(y, y')| \geq \varepsilon\}) \leq \alpha \varepsilon\}. \end{aligned}$$

Theorem 115. *For each $\alpha \in [0, \infty)$, $d_{\mathcal{N},\alpha}^{\mathcal{GP}}$ is a pseudometric on \mathcal{N}_m .*

Lemma 116 (Relation between Gromov-Prokhorov and Gromov-Wasserstein). *Consider $(X, \omega_X, \mu_X), (Y, \omega_Y, \mu_Y) \in \mathcal{N}_m$. We always have:*

$$d_{\mathcal{N},0}^{\mathcal{GP}}(X, Y) = d_{\mathcal{N},\infty}(X, Y).$$

1.9.7 Lower bounds and measure network invariants

Let (V, d_V) denote a pseudometric space. By a *(pseudo)metric-valued network invariant*, we mean a function $\iota : \mathcal{N}_m \rightarrow V$ such that $X \cong Y$ implies $d_V(\iota(X), \iota(Y)) = 0$. We are also interested in \mathbb{R} -parametrized network invariants, which are functions $\iota : \mathcal{N}_m \times \mathbb{R} \rightarrow V$ such that $X \cong Y$ implies $d_V(\iota(X), \iota(Y)) = 0$. This is a bona fide generalization of the non-parametrized setting, because any map $\iota : \mathcal{N}_m \rightarrow V$ can be viewed as being parametrized by a constant object $\{0\}$.

There are two notions of stability that we are interested in.

Definition 42 (Lipschitz stability). Let $p \in [1, \infty]$. A *Lipschitz-stable network invariant* is an invariant $\iota_p : \mathcal{N}_m \rightarrow V$ for which there exists a Lipschitz constant $L(\iota_p) > 0$ such that

$$d_V(\iota_p(X), \iota_p(Y)) \leq L(\iota_p)d_{\mathcal{N},p}(X, Y) \text{ for all } X, Y \in \mathcal{N}_m.$$

Definition 43 (Interleaving stability). Let $p \in [1, \infty]$. An *interleaving-stable network invariant* is an \mathbb{R} -parametrized invariant $\iota_p : \mathcal{N}_m \times \mathbb{R} \rightarrow V$ for which there exists an interleaving constant $\alpha \in \mathbb{R}$ and a symmetric interleaving function $\varepsilon : \mathcal{N}_m \times \mathcal{N}_m \rightarrow \mathbb{R}$ such that

$$\iota_p(X, t) \leq \iota_p(Y, t + \varepsilon_{XY}) + \alpha\varepsilon_{XY} \leq \iota_p(X, t + 2\varepsilon_{XY}) + 2\alpha\varepsilon_{XY} \text{ for all } t \in \mathbb{R} \text{ and } X, Y \in \mathcal{N}_m.$$

Here $\varepsilon_{XY} := \varepsilon(X, Y)$. In Example 117 below, we give some invariants that are interleaving stable.

Example 117 (A map that ignores/emphasizes large edge weights). Let $t \in \mathbb{R}$. For each $p \in [1, \infty]$, the p th t -sublevel set map for the weight function, denoted $\text{sub}_{p,t}^w : \mathcal{N}_m \rightarrow \mathbb{R}_+$, is given as:

$$\begin{aligned} \text{sub}_{p,t}^w(X, \omega_X, \mu_X) &= \left(\int_{\{\omega_X \leq t\}} |\omega_X(x, x')|^p d(\mu_X \otimes \mu_X)(x, x') \right)^{1/p} \quad \text{for } p \in [1, \infty), \\ \text{sub}_{p,t}^w(X, \omega_X, \mu_X) &= \sup\{|\omega_X(x, x')| : x, x' \in \text{supp}(\mu_X), \omega_X(x, x') \leq t\} \quad \text{for } p = \infty. \end{aligned}$$

This map de-emphasizes large edge weights in a measure network. Analogously, one can consider integrating over the set $\{\omega_X \geq t\}$. In this case, the larger edge weights are emphasized. The corresponding superlevel set invariant is denoted $\text{sup}_{p,t}^w$.

Theorem 118 (Interleaving stability of the sublevel/superlevel set weight invariants). Let $p \in [1, \infty]$. The sub_p^w invariant is interleaving-stable with interleaving constant $\alpha = 1$ and interleaving function $d_{\mathcal{N},\infty}$. The sup_p^w invariant is interleaving-stable with interleaving constant $\alpha = -1$ and interleaving function $-d_{\mathcal{N},\infty}$.

We now define a family of local invariants that incorporate data from the networks at a much finer scale. Computing these local invariants amounts to solving an optimal transport (OT) problem, which is a linear programming (LP) task.

Example 119 (A generalized eccentricity function). Let (X, ω_X, μ_X) be a measure network. Then consider the $\text{ecc}_{p,X}^{\text{out}} : X \rightarrow \mathbb{R}_+$ map

$$\text{ecc}_{p,X}^{\text{out}}(s) := \left(\int_X |\omega_X(s, x)|^p d\mu_X(x) \right)^{1/p} = \|\omega_X(s, \cdot)\|_{L^p(\mu_X)}.$$

The $p = \infty$ version is defined analogously, with the integral replaced by a supremum over the support. We can also replace $\omega_X(s, \cdot)$ above with $\omega_X(\cdot, s)$ to obtain another map $\text{ecc}_{p,X}^{\text{in}}$. In general, the two maps will not agree due to the asymmetry of the network. This invariant is an asymmetric generalization of the p -eccentricity function for metric measure spaces [86, Definition 5.3]

Example 120 (A joint eccentricity function). Let (X, ω_X, μ_X) and (Y, ω_Y, μ_Y) be two measure networks, and let $p \in [1, \infty]$. Define the (*outer*) *joint eccentricity function* $\text{ecc}_{p,X,Y}^{\text{out}} : X \times Y \rightarrow \mathbb{R}_+$ of X and Y as follows: for each $(s, t) \in X \times Y$,

$$\text{ecc}_{p,X,Y}^{\text{out}}(s, t) := \inf_{\mu \in \mathcal{C}(\mu_X, \mu_Y)} \|\omega_X(s, \cdot) - \omega_Y(t, \cdot)\|_{L^p(\mu)}.$$

For $p \in [1, \infty)$, this invariant has the following form:

$$\text{ecc}_{p,X,Y}^{\text{out}}(s, t) := \left(\inf_{\mu \in \mathcal{C}(\mu_X, \mu_Y)} \int_{X \times Y} |\omega_X(s, x') - \omega_Y(t, y')|^p d\mu(x', y') \right)^{1/p}.$$

One obtains the *inner joint eccentricity function* by using the term $\omega_X(\cdot, s) - \omega_Y(\cdot, t)$ above, and we denote this by $\text{ecc}_{p,X,Y}^{\text{in}}$.

Theorem 121 (Stability of local \mathbb{R} -valued invariants). *The eccentricity and joint eccentricity invariants are both Lipschitz stable, with Lipschitz constant 2. Formally, for any $(X, \omega_X, \mu_X), (Y, \omega_Y, \mu_Y) \in \mathcal{N}_m$, we have:*

$$\inf_{\mu \in \mathcal{C}(\mu_X, \mu_Y)} \|\text{ecc}_{p,X}^{\text{out}} - \text{ecc}_{p,Y}^{\text{out}}\|_{L^p(\mu)} \leq 2d_{\mathcal{N},p}(X, Y), \quad (\text{eccentricity bound})$$

$$\inf_{\mu \in \mathcal{C}(\mu_X, \mu_Y)} \|\text{ecc}_{p,X,Y}^{\text{out}}\|_{L^p(\mu)} \leq 2d_{\mathcal{N},p}(X, Y). \quad (\text{joint eccentricity bound})$$

Moreover, the joint eccentricity invariant provides a stronger bound than the eccentricity bound, i.e.

$$\inf_{\mu \in \mathcal{C}(\mu_X, \mu_Y)} \|\text{ecc}_{p,X}^{\text{out}} - \text{ecc}_{p,Y}^{\text{out}}\|_{L^p(\mu)} \leq \inf_{\mu \in \mathcal{C}(\mu_X, \mu_Y)} \|\text{ecc}_{p,X,Y}^{\text{out}}\|_{L^p(\mu)} \leq 2d_{\mathcal{N},p}(X, Y).$$

Finally, the analogous bounds hold in the case of the inner eccentricity and inner joint eccentricity functions.

Remark 122. The analogous bounds in the setting of metric measure spaces were provided in [85], where the eccentricity and joint eccentricity bounds were called the First and Third Lower Bounds, respectively. The TLB later appeared in [107].

Having described the form of the local network invariants, we now leverage a particularly useful fact about optimal transport over the real line. For probability measures over \mathbb{R} , the method for constructing an optimal coupling is known, and this gives a simple formula for computing the OT cost in terms of the cumulative distribution functions of the measures [120, Remark 2.19]. Later we obtain lower bounds based on distributions over \mathbb{R} that can be computed easily and remain stable with respect to the local invariants described above.

Remark 123. The structure of the joint eccentricity bound (i.e. the TLB) in Theorem 121 shows that a priori, it involves solving an ensemble of OT problems, one for each pair $(x, y) \in X \times Y$, and a final OT problem once $\text{ecc}_{p,X,Y}^{\text{out}}$ is computed.

Example 124 (Pushforward via ω_X). Recall that given any (X, ω_X, μ_X) , the corresponding pushforward of $\mu_X \otimes \mu_X$ via ω_X is given as follows: for any generator of $\text{Borel}(\mathbb{R})$ of the form $(a, b) \subseteq \mathbb{R}$,

$$\begin{aligned} (\omega_X)_*(\mu_X \otimes \mu_X)(a, b) &:= (\mu_X \otimes \mu_X)(\{\omega_X \in (a, b)\}) \\ &= \int_X \int_X \mathbb{1}_{\{\omega_X \in (a, b)\}}(x, x') d\mu_X(x) d\mu_X(x'). \end{aligned}$$

For convenience, we define $\nu_X := (\omega_X)_*(\mu_X^{\otimes 2})$. This distribution is completely determined by its cumulative distribution function, which we denote by F_{ω_X} . This is a function $\mathbb{R} \rightarrow [0, 1]$ given by:

$$F_{\omega_X}(t) := (\mu_X \otimes \mu_X)(\{\omega_X \leq t\}) = \int_X \int_X \mathbb{1}_{\{\omega_X \leq t\}}(x, x') d\mu_X(x) d\mu_X(x').$$

The distribution-valued invariant above is a global invariant. The corresponding local versions are below.

Example 125 (Pushforward via a single coordinate of ω_X). Let (X, ω_X, μ_X) and $x \in X$ be given. Then we can define *local* distribution-valued invariants as follows: for any generator of $\text{Borel}(\mathbb{R})$ of the form $(a, b) \subseteq \mathbb{R}$,

$$\begin{aligned} (\omega_X(x, \cdot))_* \mu_X(a, b) &:= \mu_X(\{x' \in X : \omega_X(x, x') \in (a, b)\}) \\ (\omega_X(\cdot, x))_* \mu_X(a, b) &:= \mu_X(\{x' \in X : \omega_X(x', x) \in (a, b)\}). \end{aligned}$$

We adopt the following shorthand:

$$\lambda_X(x) := (\omega_X(x, \cdot))_* \mu_X, \quad \rho_X(x) := (\omega_X(\cdot, x))_* \mu_X.$$

Here we write λ and ρ to refer to the “left” and “right” arguments, respectively. The corresponding distribution functions are defined as follows: for any $t \in \mathbb{R}$,

$$\begin{aligned} F_{\omega_X(x, \cdot)}(t) &:= \mu_X(\{\omega_X(x, \cdot) \leq t\}) = \int_X \mathbb{1}_{\{\omega_X(x, \cdot) \leq t\}}(x') d\mu_X(x') \\ F_{\omega_X(\cdot, x)}(t) &:= \mu_X(\{\omega_X(\cdot, x) \leq t\}) = \int_X \mathbb{1}_{\{\omega_X(\cdot, x) \leq t\}}(x') d\mu_X(x'). \end{aligned}$$

It is interesting to note that we get such a pair of distributions for each $x \in X$. Thus we can add yet another layer to this construction, via the maps $\mathcal{N}_m \rightarrow \text{pow}(\text{Prob}(\mathbb{R}))$ defined by writing

$$\begin{aligned} (X, \omega_X, \mu_X) &\mapsto \{\lambda_X(x) : x \in X\}, \text{ and} \\ (X, \omega_X, \mu_X) &\mapsto \{\rho_X(x) : x \in X\} \text{ for each } (X, \omega_X, \mu_X) \in \mathcal{N}_m. \end{aligned}$$

Assume for now that we equip $\text{Prob}(\mathbb{R})$ with the Wasserstein metric. Write $\mathbb{X} := \{\lambda_X(x)\}_{x \in X}$, let $d_{\mathbb{X}}$ denote the Wasserstein metric, and let $\mu_{\mathbb{X}} := (\lambda_X)_* \mu_X$. More specifically, for any

$A \in \text{Borel}(\mathbb{X})$, we have $\mu_{\mathbb{X}}(A) = \mu_X(\{x \in X : \lambda_X(x) \in A\})$. This yields a metric measure space $(\mathbb{X}, d_{\mathbb{X}}, \mu_{\mathbb{X}})$. So even though we do not start off with a metric space, the operation of passing into distributions over \mathbb{R} forces a metric structure on (X, ω_X, μ_X) .

Next let $(Y, \omega_Y, \mu_Y) \in \mathcal{N}$, and suppose $(\mathbb{Y}, d_{\mathbb{Y}}, \mu_{\mathbb{Y}})$ are defined as above. Since $\mathbb{X}, \mathbb{Y} \subseteq \text{Prob}(\mathbb{R})$, we know that $\mu_{\mathbb{X}}, \mu_{\mathbb{Y}}$ are both distributions on $\text{Prob}(\mathbb{R})$. Thus we can compare them via the p -Wasserstein distance as follows, for $p \in [1, \infty)$:

$$d_{W,p}(\mu_{\mathbb{X}}, \mu_{\mathbb{Y}}) = \inf_{\mu \in \mathcal{C}(\mu_{\mathbb{X}}, \mu_{\mathbb{Y}})} \left(\int_{\text{Prob}(\mathbb{R})^2} d_{W,p}(\lambda_X(x), \lambda_Y(y))^p d\mu(\lambda_X(x), \lambda_Y(y)) \right)^{1/p}$$

By the change of variables formula, this quantity coincides with one that we show below to be a lower bound for $2d_{\mathcal{N},p}(X, Y)$ (cf. Inequality (1.14) of Theorem 127).

Example 126 (Pushforward via eccentricity). Let $(X, \omega_X, \mu_X), (Y, \omega_Y, \mu_Y) \in \mathcal{N}$, and let $(a, b) \in \text{Borel}(\mathbb{R})$. Recall the outer and inner eccentricity functions $\text{ecc}_{p,X}^{\text{out}}$ and $\text{ecc}_{p,X}^{\text{in}}$ from Example 119. These functions induce distributions as follows:

$$\begin{aligned} (\text{ecc}_{p,X}^{\text{out}})_* \mu_X(a, b) &= \mu_X(\{x \in X : \text{ecc}_{p,X}^{\text{out}}(x) \in (a, b)\}), \\ (\text{ecc}_{p,X}^{\text{in}})_* \mu_X(a, b) &= \mu_X(\{x \in X : \text{ecc}_{p,X}^{\text{in}}(x) \in (a, b)\}). \end{aligned}$$

Next let $\mu \in \mathcal{C}(\mu_X, \mu_Y)$ and recall the joint outer/inner eccentricity functions $\text{ecc}_{p,X,Y}^{\text{out}}$ and $\text{ecc}_{p,X,Y}^{\text{in}}$ from Example 120. These functions induce distributions as below:

$$\begin{aligned} (\text{ecc}_{p,X,Y}^{\text{out}})_* \mu(a, b) &= \mu(\{(x, y) \in X \times Y : \text{ecc}_{p,X,Y}^{\text{out}}(x, y) \in (a, b)\}), \\ (\text{ecc}_{p,X,Y}^{\text{in}})_* \mu(a, b) &= \mu(\{(x, y) \in X \times Y : \text{ecc}_{p,X,Y}^{\text{in}}(x, y) \in (a, b)\}). \end{aligned}$$

Theorem 127 (Stability of the ω_X and eccentricity-pushforward distributions). *Suppose $(X, \omega_X, \mu_X), (Y, \omega_Y, \mu_Y) \in \mathcal{N}_m$. Then we have the following statements about Lipschitz*

stability, for $p \in [1, \infty)$:

$$2d_{N,p}(X, Y) \geq \inf_{\mu \in \mathcal{C}(\mu_X \otimes \mu_X, \mu_Y \otimes \mu_Y)} \left(\int_{X^2 \times Y^2} |\omega_X(x, x') - \omega_Y(y, y')|^p d\mu(x, x', y, y') \right)^{1/p} \quad (1.7)$$

$$\geq \inf_{\nu \in \mathcal{C}(\nu_X, \nu_Y)} \left(\int_{\mathbb{R}^2} |a - b|^p d\nu(a, b) \right)^{1/p}. \quad (1.8)$$

$$2d_{N,p}(X, Y) \geq \inf_{\mu \in \mathcal{C}(\mu_X, \mu_Y)} \left(\int_{X \times Y} |\text{ecc}_{p,X}^{\text{out}}(x) - \text{ecc}_{p,Y}^{\text{out}}(y)|^p d\mu(x, y) \right)^{1/p} \quad (1.9)$$

$$\geq \inf_{\gamma \in \mathcal{C}((\text{ecc}_{p,X}^{\text{out}})_* \mu_X, (\text{ecc}_{p,Y}^{\text{out}})_* \mu_Y)} \left(\int_{\mathbb{R}^2} |a - b|^p d\gamma(a, b) \right)^{1/p}. \quad (1.10)$$

$$2d_{N,p}(X, Y) \geq \inf_{\mu \in \mathcal{C}(\mu_X, \mu_Y)} \left(\int_{X \times Y} |\text{ecc}_{p,X}^{\text{in}}(x) - \text{ecc}_{p,Y}^{\text{in}}(y)|^p d\mu(x, y) \right)^{1/p} \quad (1.11)$$

$$\geq \inf_{\gamma \in \mathcal{C}((\text{ecc}_{p,X}^{\text{in}})_* \mu_X, (\text{ecc}_{p,Y}^{\text{in}})_* \mu_Y)} \left(\int_{\mathbb{R}^2} |a - b|^p d\gamma(a, b) \right)^{1/p}. \quad (1.12)$$

$$2d_{N,p}(X, Y) \geq \inf_{\mu \in \mathcal{C}(\mu_X, \mu_Y)} \left(\int_{X \times Y} \inf_{\gamma \in \mathcal{C}(\mu_X, \mu_Y)} \int_{X \times Y} |\omega_X(x, x') - \omega_Y(y, y')|^p d\gamma(x', y') d\mu(x, y) \right)^{1/p} \quad (1.13)$$

$$\geq \inf_{\mu \in \mathcal{C}(\mu_X, \mu_Y)} \left(\int_{X \times Y} \inf_{\gamma \in \mathcal{C}(\lambda_X(x), \lambda_Y(y))} \int_{\mathbb{R}^2} |a - b|^p d\gamma(a, b) d\mu(x, y) \right)^{1/p}. \quad (1.14)$$

$$2d_{N,p}(X, Y) \geq \inf_{\mu \in \mathcal{C}(\mu_X, \mu_Y)} \left(\int_{X \times Y} \inf_{\gamma \in \mathcal{C}(\mu_X, \mu_Y)} \int_{X \times Y} |\omega_X(x, x') - \omega_Y(y, y')|^p d\gamma(x, y) d\mu(x', y') \right)^{1/p} \quad (1.15)$$

$$\geq \inf_{\mu \in \mathcal{C}(\mu_X, \mu_Y)} \left(\int_{X \times Y} \inf_{\gamma \in \mathcal{C}(\rho_X(x'), \rho_Y(y'))} \int_{\mathbb{R}^2} |a - b|^p d\gamma(a, b) d\mu(x', y') \right)^{1/p}. \quad (1.16)$$

Here recall that $\nu_X = (\omega_X)_*(\mu_X^{\otimes 2})$, $\nu_Y = (\omega_Y)_*(\mu_Y^{\otimes 2})$, $\lambda_X = (\omega_X(x, \cdot))_* \mu_X$, $\lambda_Y = (\omega_Y(y, \cdot))_* \mu_Y$, $\rho_X = (\omega_X(\cdot, x))_* \mu_X$, and $\rho_Y = (\omega_Y(\cdot, y))_* \mu_Y$. Inequalities (1.7)-(1.8) appeared as the Second Lower Bound and its relaxation in [85]. Inequalities (1.9), (1.11), (1.13), and (1.15) are the eccentricity bounds in Theorem 121. Inequalities (1.10), (1.12), (1.14), and (1.16) are their relaxations. In the symmetric case, these outer/inner pairs of inequalities coincide; they appeared as the First and Third Lower Bounds and their relaxations in [85].

In Inequality (1.8), both ν_X and ν_Y are probability distributions on \mathbb{R} , and the right hand side is precisely the p -Wasserstein distance between ν_X and ν_Y . Analogous statements hold for Inequalities (1.14) and (1.16).

To connect with the nomenclature introduced in [85], we give names to the inequalities in Theorem 127: (1.7)-(1.8) are the SLB inequalities, (1.9)-(1.12) are the FLB inequalities, and (1.13)-(1.16) are the TLB inequalities. These abbreviations stand for First, Second, and Third Lower Bound. Note that due to the asymmetry of the network setting, we get twice as many FLB and TLB inequalities as we would for the metric measure setting.

Next we describe the formula for computing OT over \mathbb{R} (see [120, Remark 2.19]). Let measure spaces (X, μ_X) , (Y, μ_Y) and measurable functions $f : X \rightarrow \mathbb{R}$, $g : Y \rightarrow \mathbb{R}$ be given. Then let $F, G : \mathbb{R} \rightarrow [0, 1]$ denote the cumulative distribution functions of f and g :

$$F(t) := \mu_X(f \leq t), \quad G(t) := \mu_Y(g \leq t).$$

The generalized inverses $F^{-1} : [0, 1] \rightarrow \mathbb{R}$, $G^{-1} : [0, 1] \rightarrow \mathbb{R}$ are given as:

$$F^{-1}(t) := \inf\{u \in \mathbb{R} : F(u) \geq t\}, \quad G^{-1}(t) := \inf\{u \in \mathbb{R} : G(u) \geq t\}$$

Then for $p \geq 1$, we have:

$$\inf_{\mu \in \mathcal{C}(f_* \mu_X, g_* \mu_Y)} \int_{\mathbb{R} \times \mathbb{R}} |a - b|^p d\mu(a, b) = \int_0^1 |F^{-1}(t) - G^{-1}(t)|^p dt \quad (1.17)$$

For $p = 1$, we even have:

$$\inf_{\mu \in \mathcal{C}(f_* \mu_X, g_* \mu_Y)} \int_{\mathbb{R} \times \mathbb{R}} |a - b| d\mu(a, b) = \int_{\mathbb{R}} |F(t) - G(t)| dt \quad (1.18)$$

These formulae are easily adapted to obtain closed form solutions for the lower bounds given by Inequalities (1.10) and (1.12) and for the inner OT problems in (1.14), (1.16) of Theorem 127.

1.10 Computational aspects

In this section, we present two sets of experiments. One of these uses Dowker persistence, the other uses the lower bounds on $d_{\mathcal{N},p}$. Further experiments and details about algorithms are presented in Chapter 4.

Throughout this thesis, we have presented numerous quantities that can potentially be computed. It turns out that $d_{\mathcal{N}}$ is mostly intractable—it is NP-hard, and is difficult to compute for networks having more than four or five nodes. The $d_{\mathcal{N},p}$ distances are still NP-hard, but because they relax the purely combinatorial structure of $d_{\mathcal{N}}$, they can be approximately computed via quadratic optimization approaches. For both the $d_{\mathcal{N}}$ and $d_{\mathcal{N},p}$ distances, we provide invariants and associated lower bounds that are computed as linear programs, hence become much more computationally tractable.

On the side of persistent homology computations, it turns out that Dowker (and also Vietoris-Rips) persistence is remarkably easy to implement—the only task is compute the

simplicial filtration, after which any out-of-the-box persistent homology package can take over (we use `Javaplex` for the latter).

Computing PPH, however, is more tricky, a priori. As we point out in Remark 55, we are not supplied with a basis for path homology computations, and computing this basis requires significant additional preprocessing, at least for a naive implementation of PPH. One of our contributions is in showing that the basis computation and persistence computation rely on the same matrix operations, so both steps can be combined and carried out together.

1.10.1 Software packages

For $d_{\mathcal{N}}$, related lower bounds, and Vietoris-Rips/Dowker persistence computations, we developed the `PersNet` software package jointly with Facundo Mémoli—see <https://github.com/fmemoli/PersNet>.

For computations related to $d_{\mathcal{N},p}$, we released the `GWnets` package <https://github.com/samirchowdhury/GWnets>.

For PPH computation (up to dimension 1), we have implementations in C++, Python 2.7, and Matlab—see <https://github.com/samirchowdhury>. The C++ implementation is the fastest of the three.

1.10.2 Simulated hippocampal networks

In the neuroscience literature, it has been shown that as an animal explores a given *environment* or *arena*, specific “place cells” in the hippocampus show increased activity at specific spatial regions, called “place fields” [93]. Each place cell shows a *spike* in activity when the animal enters the place field linked to this place cell, accompanied by a drop in activity as the animal moves far away from this place field. To understand how the brain processes this data, a natural question to ask is the following: Is the time series data of the place cell activity, referred to as “spike trains”, enough to detect the structure of the arena?

Approaches based on homology [41] and persistent homology [43] have shown positive results in this direction. In [43], the authors simulated the trajectory of a rat in an arena containing “holes.” A simplicial complex was then built as follows: whenever $n + 1$ place cells with overlapping place fields fired together, an n -simplex was added. This yielded a filtered simplicial complex indexed by a time parameter. By computing persistence, it was then shown that the number of persistent bars in the 1-dimensional barcode of this filtered simplicial complex would accurately represent the number of holes in the arena.

We repeated this experiment with the following change in methodology: we simulated the movement of an animal, and corresponding hippocampal activity, in arenas with a variety of obstacles. We then induced a directed network from each set of hippocampal activity data, and computed the associated 1-dimensional Dowker persistence diagrams. We were interested in seeing if the bottleneck distances between diagrams arising from similar arenas would differ significantly from the bottleneck distance between diagrams arising from

different arenas. To further exemplify our methods, we repeated our analysis after computing the 1-dimensional Rips persistence diagrams from the hippocampal activity networks.

In our experiment, there were five arenas. The first was a square of side length $L = 10$, with four circular ‘‘holes’’ or ‘‘forbidden zones’’ of radius $0.2L$ that the trajectory could not intersect. The other four arenas were those obtained by removing the forbidden zones one at a time. In what follows, we refer to the arenas of each type as *4-hole*, *3-hole*, *2-hole*, *1-hole*, and *0-hole arenas*. For each arena, a random-walk trajectory of 5000 steps was generated, where the animal could move along a square grid with 20 points in each direction. The grid was obtained as a discretization of the box $[0, L] \times [0, L]$, and each step had length $0.05L$. The animal could move in each direction with equal probability. If one or more of these moves took the animal outside the arena (a disallowed move), then the probabilities were redistributed uniformly among the allowed moves. Each trajectory was tested to ensure that it covered the entire arena, excluding the forbidden zones. Formally, we write the time steps as a set $T := \{1, 2, \dots, 5000\}$, and denote the trajectory as a map $\text{traj} : T \rightarrow [0, L]^2$.

For each of the five arenas, 20 trials were conducted, producing a total of 100 trials. For each trial l_k , an integer n_k was chosen uniformly at random from the interval $[150, 200]$. Then n_k place fields of radius $0.05L$ were scattered uniformly at random inside the corresponding arena for each l_k . An illustration of the place field distribution is provided in Figure 1.22. A spike on a place field was recorded whenever the trajectory would intersect it. So for each $1 \leq i \leq n_k$, the spiking pattern of cell x_i , corresponding to place field PF_i , was recorded via a function $r_i : T \rightarrow \{0, 1\}$ given by:

$$r_i(t) = \begin{cases} 1 & : \text{if } \text{traj}(t) \text{ intersects } \text{PF}_i, \\ 0 & : \text{otherwise} \end{cases} \quad t \in T.$$

The matrix corresponding to r_i is called the *raster* of cell x_i . A sample raster is illustrated in Figure 1.22. For each trial l_k , the corresponding network (X_k, ω_{X_k}) was constructed as follows: X_k consisted of n_k nodes representing place fields, and for each $1 \leq i, j \leq n_k$, the weight $\omega_{X_k}(x_i, x_j)$ was given by:

$$\omega_{X_k}(x_i, x_j) := 1 - \frac{N_{i,j}(5)}{\sum_{i=1}^{n_k} N_{i,j}(5)},$$

where $N_{i,j}(5) = \text{card}(\{(s, t) \in T^2 : t \in [2, 5000], t - s \in [1, 5], r_j(t) = 1, r_i(s) = 1\})$.

In words, $N_{i,j}(5)$ counts the pairs of times (s, t) , $s < t$, such that cell x_j spikes (at a time t) after cell x_i spikes (at a time s), and the delay between the two spikes is fewer than 5 time steps. The idea is that if cell x_j frequently fires within a short span of time after cell x_i fires, then place fields PF_i and PF_j are likely to be in close proximity to each other. The column sum of the matrix corresponding to ω_{X_k} is normalized to 1, and so $\omega_{X_k}^\top$ can be interpreted as the transition matrix of a Markov process.

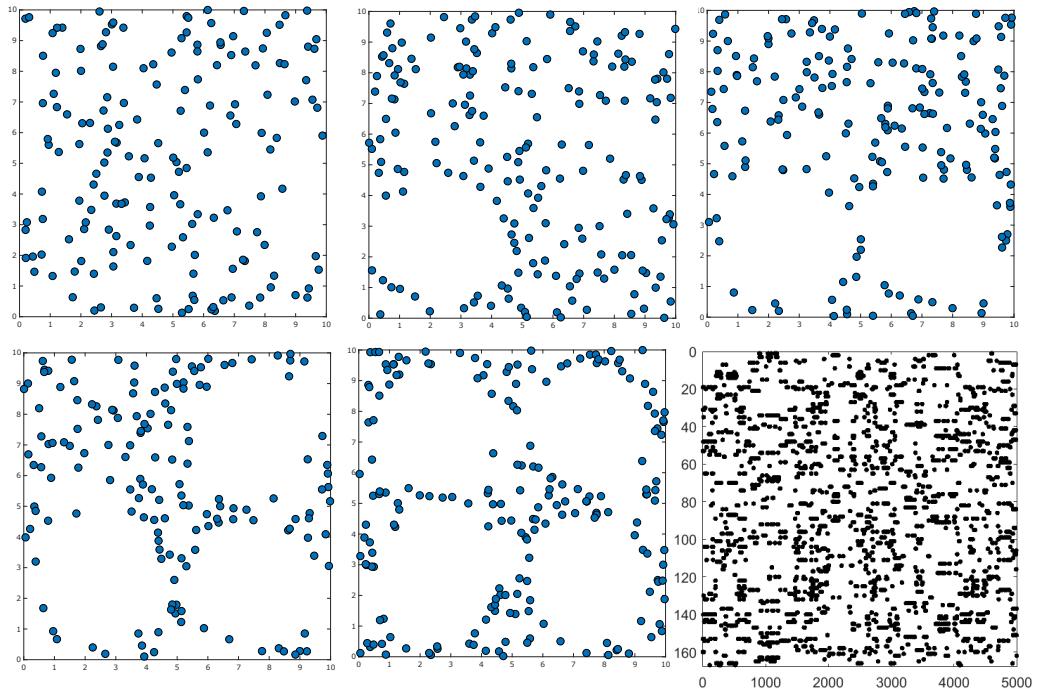


Figure 1.22: **Bottom right:** Sample place cell spiking pattern matrix. The x -axis corresponds to the number of time steps, and the y -axis corresponds to the number of place cells. Black dots represent spikes. **Clockwise from bottom middle:** Sample distribution of place field centers in 4, 3, 0, 1, and 2-hole arenas.

Next, we computed the 1-dimensional Dowker persistence diagrams of each of the 100 networks. Note that $\text{Dgm}_1^{\mathfrak{D}}(\omega_X) = \text{Dgm}_1^{\mathfrak{D}}(\omega_X^\top)$ by Proposition 46, so we are actually obtaining the 1-dimensional Dowker persistence diagrams of transition matrices of Markov processes. We then computed a 100×100 matrix consisting of the bottleneck distances between all the 1-dimensional persistence diagrams. The single linkage dendrogram generated from this bottleneck distance matrix is shown in Figure 1.23. The labels are in the format `env-<nh>-<nn>`, where `nh` is the number of holes in the arena/environment, and `nn` is the number of place fields. Note that with some exceptions, networks corresponding to the same arena are clustered together. We conclude that the Dowker persistence diagram succeeded in capturing the intrinsic differences between the five classes of networks arising from the five different arenas, even when the networks had different sizes.

We then computed the Rips persistence diagrams of each network, and computed the 100×100 bottleneck distance matrix associated to the collection of 1-dimensional diagrams. The single linkage dendrogram generated from this matrix is given in Figure 1.24. Notice that the Rips dendrogram does not do a satisfactory job of classifying arenas correctly.

Remark 128. We note that an alternative method of comparing the networks obtained from our simulations would have been to compute the pairwise network distances, and plot the results in a dendrogram. But $d_{\mathcal{N}}$ is NP-hard to compute—this follows from the fact that computing $d_{\mathcal{N}}$ includes the problem of computing Gromov-Hausdorff distance between finite metric spaces, which is NP-hard [105]. So instead, we are computing the bottleneck distances between 1-dimensional Dowker persistence diagrams, as suggested by Remark 30.

Remark 129. It is possible to compare the current approach with the one taken in [43] on a common dataset. We performed this comparison in [32] for a similar experiment, but with a stochastic firing model for the place cells. Interestingly, it turns out that the network approach with Dowker persistence performs better than the approach in [43], as indicated by computing 1-nearest neighbor classification error rates on the bottleneck distance matrices. A possible explanation is that preprocessing the spiking data into a network automatically incorporates a form of error correction, where the errors consist of stochastic firing between cells that are non-adjacent. On the other hand, such errors are allowed to accumulate over time in the approach taken in [43]. For an alternative error-correction approach, see [33].

1.10.3 Clustering SBMs and migration networks

We now describe the specifics of an experiment on clustering a collection of network SBMs and migration networks. We perform clustering with respect to the TLB, i.e. the bounds given by Inequalities (1.14) and (1.16).

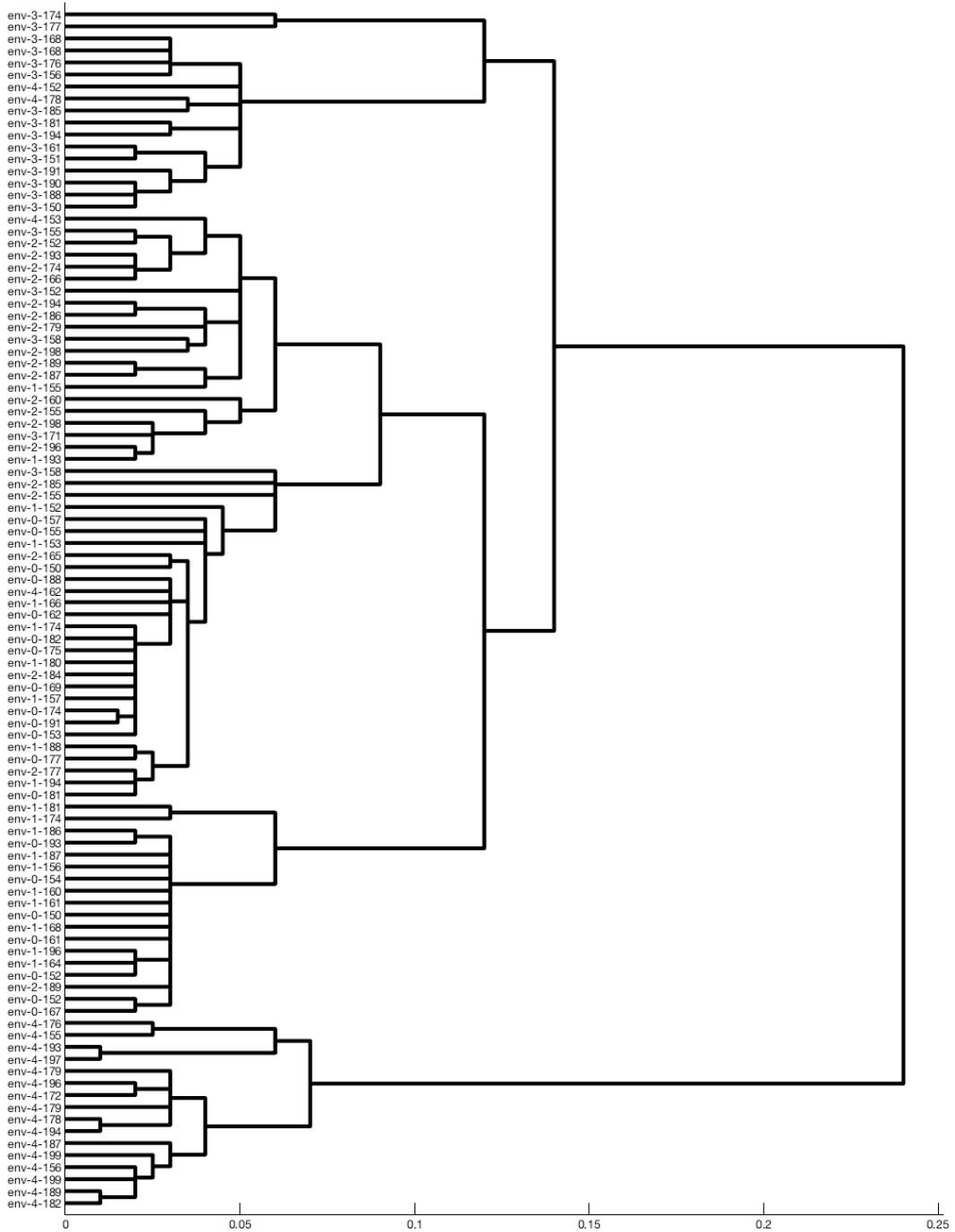


Figure 1.23: Single linkage dendrogram corresponding to the distance matrix obtained by computing bottleneck distances between 1-dimensional Dowker persistence diagrams of our database of hippocampal networks (§1.10.2). Note that the 4, 3, and 2-hole arenas are well separated into clusters at threshold 0.1.

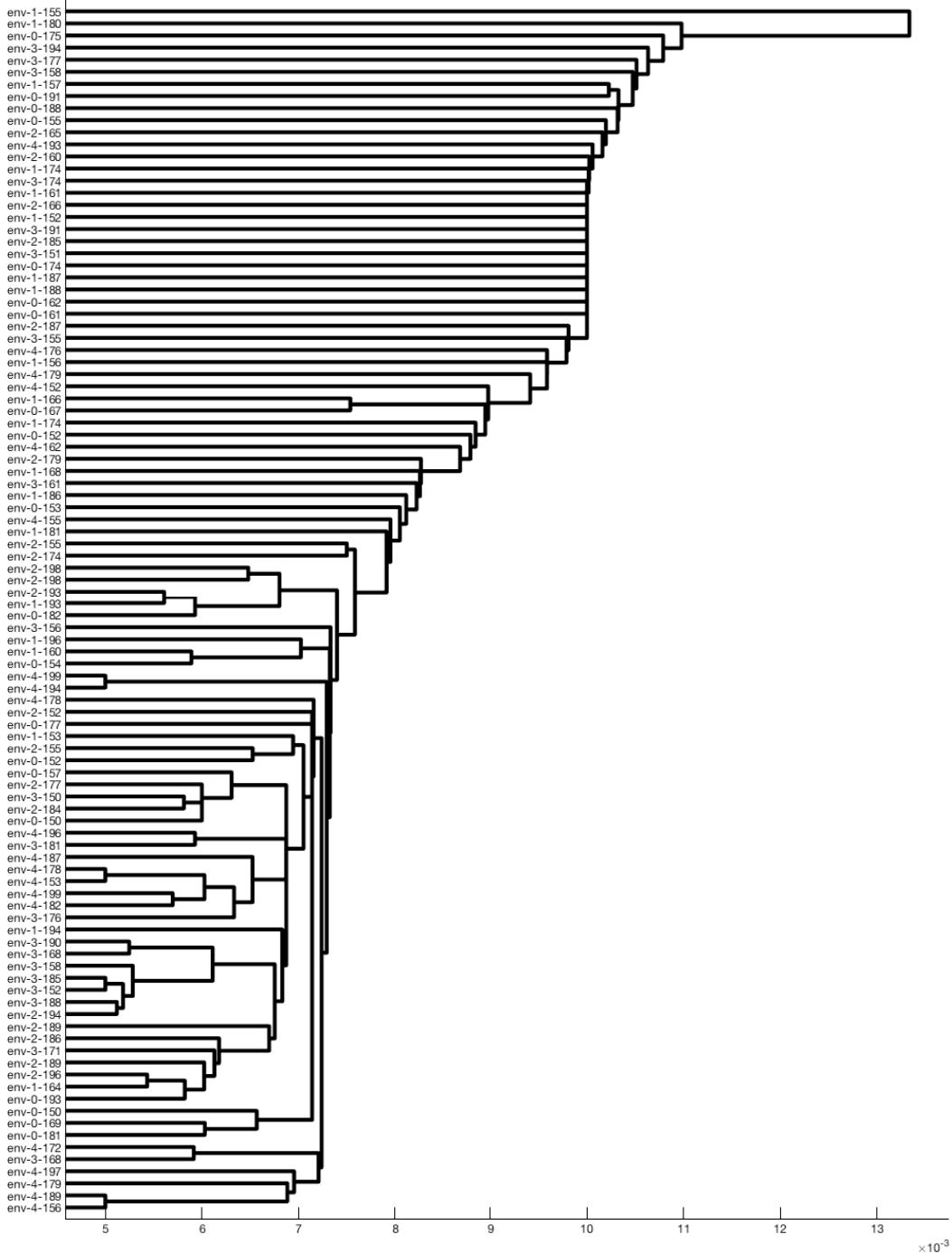


Figure 1.24: Single linkage dendrogram corresponding to the distance matrix obtained by computing bottleneck distances between 1-dimensional Rips persistence diagrams of our database of hippocampal networks (§1.10.2). Notice that the hierarchical clustering fails to capture the correct arena types.

Class #	N	v	n_i	Sample cycle network of means
1	5	[0,25,50,75,100]	10	0 25 50 75 100
2	5	[0,50,100,150, 200]	10	100 0 25 50 75
3	5	[0,25,50,75,100]	20	75 100 0 25 50
4	2	[0,100]	25	50 75 100 0 25
5	5	[-100,-50,0,50,100]	10	25 50 75 100 0

Table 1.1: **Left:** The five classes of SBM networks corresponding to the experiment in §1.10.3. N refers to the number of communities, v refers to the vector that was used to compute a table of means via $G_5(v)$, and n_i is the number of nodes in each community. **Right:** $G_5(v)$ for $v = [0, 25, 50, 75, 100]$.

Experiment: SBMs from cycle networks.

Let $N \in \mathbb{N}$, and let $v = [v_1, \dots, v_N]$ be an $N \times 1$ vector. Define the right-shift operator ρ by $\rho([v_1, \dots, v_N]) = [v_N, v_1, \dots, v_{N-1}]$. The *cycle network* $G_N(v)$ is defined to be the N -node network whose weight matrix is given by $[v^T, \rho(v)^T, (\rho^2(v))^T, \dots, (\rho^{N-1}(v))^T]^T$. This is analogous to the cycle networks defined earlier in §1.3.

In our network SBM generation procedure, we started with a vector of means v and generate $G_N(v)$ for certain choices of N . This gave us the N^2 choices of means to be used for each network SBM. To keep the experiment simple, we fixed the matrix of variances to be the $N \times N$ matrix whose entries are all 5s. We made 5 choices of v , and sampled 10 networks for each choice. The objective was then to see how well the TLB could split the collection of 50 networks into 5 classes corresponding to the 5 different community structures. The different parameters used in our experiments are listed in Table 1.1. The TLB was computed essentially according to the scheme presented in Algorithm 4, except that instead of comparing the 2-Wasserstein distance between pushforward distributions over \mathbb{R} , we used Sinkhorn iterations to approximate the solution to each “inner” OT problem as described in Theorem 127 Inequalities (1.13) and (1.15).

In this experiment, we were interested in understanding the behavior of the TLB on different community structures. Class 1 is our reference; compared to this reference, class 2 differs in its edge weights, class 3 differs in the number of nodes in each community, class 4 differs in the number of communities, and class 5 differs by having a larger proportion of negative edge weights. The TLB results in Figure 1.25 show that classes 1 and 3 are treated as being very similar, whereas the other classes are all mutually well-separated. One interesting suggestion arising from this experiment is that the TLB can be used for network simplification: given a family of networks which are all at low TLB distance to each other, it may be reasonable to retain only the smallest network in the family as the “minimal representative” network.

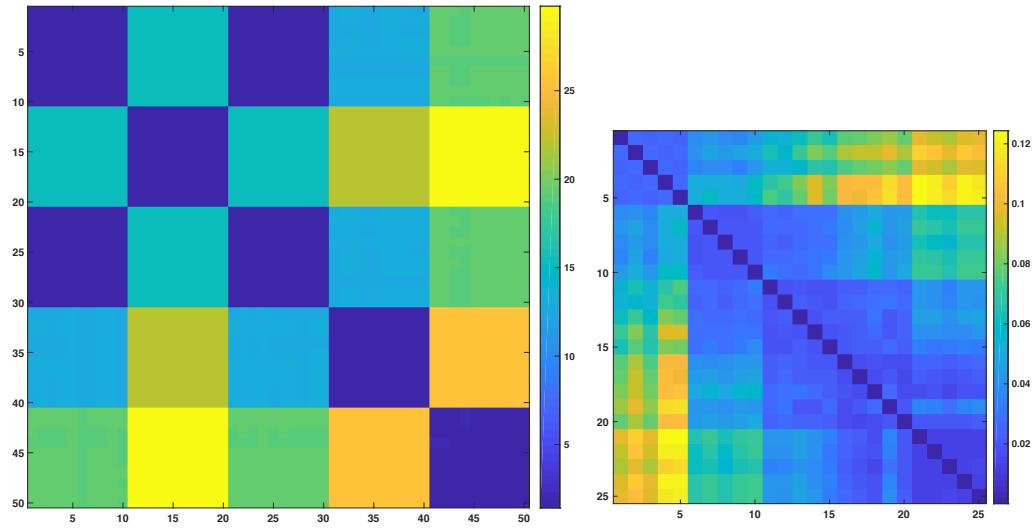


Figure 1.25: **Left:** TLB dissimilarity matrix for SBM community networks in §1.10.3. Classes 1 and 3 are similar, even though networks in Class 3 have twice as many nodes as those in Class 1. Classes 2 and 5 are most dissimilar because of the large difference in their edge weights. Class 4 has a different number of communities than the others, and is dissimilar to Classes 1 and 3 even though all their edge weights are in comparable ranges. **Right:** TLB dissimilarity matrix for two-community SBM networks in §1.10.3. The near-zero values on the diagonal are a result of using the adaptive λ -search described in Chapter 4.

Class #	N	v	n_i
1	2	[0,0]	10
2	2	[0,5]	10
3	2	[0,10]	10
4	2	[0,15]	10
5	2	[0,20]	10

Table 1.2: Two-community SBM networks as described in §1.10.3.

Experiment: Two-community SBMs with sliding means

Having understood the interaction of the TLB with network community structure, we next investigated how the TLB behaves with respect to edge weights. In our second experiment, we used a 2×1 means vector v , and varied v as $[0, 0], [0, 5], \dots, [0, 20]$ (see Table 1.2). The SBM means were then given by $G_2(v)$ for the various choices of v . The variances were fixed to be the all 5s matrix. The edge weight histograms of the resulting SBM networks then looked like samples from two Gaussian distributions, with one of the Gaussians sliding away from the other. Finally, we normalized each network by its largest weight in absolute value, so that its normalized edge weights were in $[-1, 1]$.

The purpose of this experiment was to test the performance of the TLB on SBMs coming from a mixture of Gaussians. Note that normalization ensures that simpler invariants such as the size invariant would likely fail in this setting. The TLB still performs reasonably well in this setting, as illustrated by the dissimilarity matrix in Figure 1.25. The computations were carried out as in the setting of §1.10.3.

Experiment: Real migration networks

For an experiment involving real-world networks, we compared global bilateral migration networks produced by the World Bank [66, 94]. The data consists of 10 networks, each having 225 nodes corresponding to countries/administrative regions. The (i, j) -th entry in each network is the number of people living in region i who were born in region j . The 10 networks comprise such data for male and female populations in 1960, 1970, 1980, 1990, and 2000. When extracting the data, we removed the entries corresponding to refugee populations, the Channel Islands, the Isle of Man, Serbia, Montenegro, and Kosovo, because the data corresponding to these regions was incomplete/inconsistent across the database.

The TLB computations were carried out as in Algorithm 4. In particular, we used Equation (1.17) to obtain the Wasserstein distance between pushforward distributions over \mathbb{R} . The result of applying the TLB to this dataset is illustrated in Figure 1.26. To better understand the dissimilarity matrix, we also computed its single linkage dendrogram. The

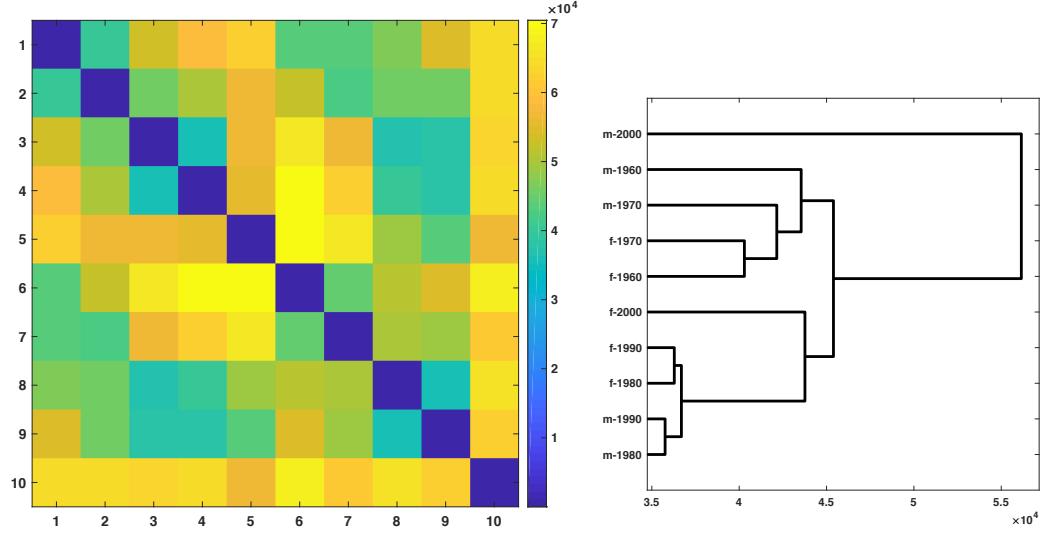


Figure 1.26: Result of applying the TLB to the migration networks in §1.10.3. **Left:** Dissimilarity matrix. Nodes 1-5 correspond to female migration from 1960-2000, and nodes 6-10 correspond to male migration from 1960-2000. **Right:** Single linkage dendrogram. Notice that overall migration patterns change in time, but within a time period, migration patterns are grouped according to gender.

dendrogram suggests that between 1980 and 1990, both male and female populations had quite similar migration patterns. Within these years, however, migration patterns were more closely tied to gender. This effect is more pronounced between 1960 and 1970, where we see somewhat greater divergence between migration patterns based on gender. Male migration in 2000 is especially divergent, with the greatest dissimilarity to all the other datasets.

The labels in the dissimilarity matrix are as follows: 1-5 correspond to “f-1960” through “f-2000”, and 6-10 correspond to “m-1960” through “m-2000”. The color gradient in the dissimilarity matrix suggests that within each gender, migration patterns change in a way that is parametrized by time. This of course reflects the shifts in global technological and economical forces which make migration attractive and/or necessary with time.

Chapter 2: Metric structures of $d_{\mathcal{N}}$ and $d_{\mathcal{N},p}$

In this chapter, we supply the proofs of the results on the metric structure of \mathcal{N} that were stated in §1. Along the way, we occasionally state additional definitions and results.

2.1 Proofs from §1.2

Proof of Proposition 7. First we show that:

$$d_{\mathcal{N}}(X, Y) \geq \frac{1}{2} \inf \{ \max(\text{dis}(\varphi), \text{dis}(\psi), C_{X,Y}(\varphi, \psi), C_{Y,X}(\psi, \varphi)) : \varphi : X \rightarrow Y, \psi : Y \rightarrow X \text{ any maps} \}.$$

Let $\varepsilon > d_{\mathcal{N}}(X, Y)$, and let R be a correspondence such that $\text{dis}(R) < 2\varepsilon$. We define maps $\varphi : X \rightarrow Y$ and $\psi : Y \rightarrow X$ as follows: for each $x \in X$, set $\varphi(x) = y$ for some y such that $(x, y) \in R$. Similarly, for each $y \in Y$, set $\psi(y) = x$ for some x such that $(x, y) \in R$.

Let $x \in X, y \in Y$. Then we have

$$|\omega_X(x, \psi(y)) - \omega_Y(\varphi(x), y)| < 2\varepsilon \quad \text{and} \quad |\omega_X(\psi(y), x) - \omega_Y(y, \varphi(x))| < 2\varepsilon.$$

Since $x \in X, y \in Y$ were arbitrary, it follows that $C_{X,Y}(\varphi, \psi) \leq 2\varepsilon$ and $C_{Y,X}(\psi, \varphi) \leq 2\varepsilon$.

Also for any $x, x' \in X$, we have $(x, \varphi(x)), (x', \varphi(x')) \in R$, and so

$$|\omega_X(x, x') - \omega_Y(\varphi(x), \varphi(x'))| < 2\varepsilon.$$

Thus $\text{dis}(\varphi) \leq 2\varepsilon$, and similarly $\text{dis}(\psi) \leq 2\varepsilon$. This proves the “ \geq ” case.

Next we wish to show the “ \leq ” case. Suppose φ, ψ are given, and

$$\frac{1}{2} \max(\text{dis}(\varphi), \text{dis}(\psi), C_{X,Y}(\varphi, \psi), C_{Y,X}(\psi, \varphi)) < \varepsilon,$$

for some $\varepsilon > 0$.

Let $R_X = \{(x, \varphi(x)) : x \in X\}$ and let $R_Y = \{(\psi(y), y) : y \in Y\}$. Then $R = R_X \cup R_Y$ is a correspondence. We wish to show that for any $z = (a, b), z' = (a', b') \in R$,

$$|\omega_X(a, a') - \omega_Y(b, b')| < 2\varepsilon.$$

This will show that $\text{dis}(R) \leq 2\varepsilon$, and so $d_{\mathcal{N}}(X, Y) \leq \varepsilon$.

To see this, let $z, z' \in R$. Note that there are four cases: (1) $z, z' \in R_X$, (2) $z, z' \in R_Y$, (3) $z \in R_X, z' \in R_Y$, and (4) $z \in R_Y, z' \in R_X$. In the first two cases, the desired inequality follows because $\text{dis}(\varphi), \text{dis}(\psi) < 2\epsilon$. The inequality follows in cases (3) and (4) because $C_{X,Y}(\varphi, \psi) < 2\epsilon$ and $C_{Y,X}(\psi, \varphi) < 2\epsilon$, respectively. Thus $d_N(X, Y) \leq \epsilon$. \square

Proof of Example 9. We start with some notation: for $x, x' \in X, y, y' \in Y$, let

$$\Gamma(x, x', y, y') = |\omega_X(x, x') - \omega_Y(y, y')|.$$

Let $\varphi : X \rightarrow Y$ be a bijection. Note that $R_\varphi := \{(x, \varphi(x)) : x \in X\}$ is a correspondence, and this holds for any bijection (actually any surjection) φ . Since we minimize over all correspondences for d_N , we conclude $d_N(X, Y) \leq \hat{d}_N(X, Y)$.

For the reverse inequality, we represent all the elements of $\mathcal{R}(X, Y)$ as 2-by-2 binary matrices R , where a 1 in position ij means $(x_i, y_j) \in R$. Denote the matrix representation of each $R \in \mathcal{R}(X, Y)$ by $\text{mat}(R)$, and the collection of such matrices as $\text{mat}(\mathcal{R})$. Then we have:

$$\text{mat}(\mathcal{R}) = \left\{ \begin{pmatrix} 1 & a \\ b & 1 \end{pmatrix} : a, b \in \{0, 1\} \right\} \cup \left\{ \begin{pmatrix} a & 1 \\ 1 & b \end{pmatrix} : a, b \in \{0, 1\} \right\}$$

Let $A = \{(x_1, y_1), (x_2, y_2)\}$ (in matrix notation, this is $\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$) and let $B = \{(x_1, y_2), (x_2, y_1)\}$ (in matrix notation, this is $\begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$). Let $R \in \mathcal{R}(X, Y)$. Note that either $A \subseteq R$ or $B \subseteq R$. Suppose that $A \subseteq R$. Then we have:

$$\max_{(x,y),(x',y') \in A} \Gamma(x, x', y, y') \leq \max_{(x,y),(x',y') \in R} \Gamma(x, x', y, y')$$

Let $\Omega(A)$ denote the quantity on the left hand side. A similar result holds in the case $B \subseteq R$:

$$\max_{(x,y),(x',y') \in B} \Gamma(x, x', y, y') \leq \max_{(x,y),(x',y') \in R} \Gamma(x, x', y, y')$$

Let $\Omega(B)$ denote the quantity on the left hand side. Since either $A \subseteq R$ or $B \subseteq R$, we have

$$\min \{\Omega(A), \Omega(B)\} \leq \min_{R \in \mathcal{R}} \max_{(x,y),(x',y') \in R} \Gamma(x, x', y, y')$$

We may identify A with the bijection given by $x_1 \mapsto y_1$ and $x_2 \mapsto y_2$. Similarly we may identify B with the bijection sending $x_1 \mapsto y_2$, $x_2 \mapsto y_1$. Thus we have

$$\min_{\varphi} \max_{x, x' \in X} \Gamma(x, x', \varphi(x), \varphi(x')) \leq \min_{R \in \mathcal{R}} \max_{(x,y),(x',y') \in R} \Gamma(x, x', y, y').$$

So we have $\hat{d}_N(X, Y) \leq d_N(X, Y)$. Thus $\hat{d}_N = d_N$.

Next, let $\{p, q\}$ and $\{p', q'\}$ denote the vertex sets of X and Y . Consider the bijection φ given by $p \mapsto p'$, $q \mapsto q'$ and the bijection ψ given by $p \mapsto q'$, $q \mapsto p'$. Note that the weight matrix is determined by setting $\omega_X(p, p) = \alpha$, $\omega_X(p, q) = \delta$, $\omega_X(q, p) = \beta$, and $\omega_X(q, q) = \gamma$, and similarly for Y . Then we get

$$\begin{aligned} \text{dis}(\varphi) &= \max(|\alpha - \alpha'|, |\beta - \beta'|, |\gamma - \gamma'|, |\delta - \delta'|) \\ \text{dis}(\psi) &= \max(|\alpha - \gamma'|, |\gamma - \alpha'|, |\delta - \beta'|, |\beta - \delta'|). \end{aligned}$$

The formula follows immediately. \square

Proof of Proposition 12. We begin with an observation. Given $X, Y \in \mathcal{N}$, let $X', Y' \in \mathcal{N}$ be such that $X \cong_{\text{II}}^w X'$, $Y \cong_{\text{II}}^w Y'$, and $\text{card}(X') = \text{card}(Y')$. Then we have:

$$d_{\mathcal{N}}(X, Y) \leq d_{\mathcal{N}}(X, X') + d_{\mathcal{N}}(X', Y') + d_{\mathcal{N}}(Y', Y) = d_{\mathcal{N}}(X', Y') \leq \widehat{d}_{\mathcal{N}}(X', Y'),$$

where the last inequality follows from Remark 10.

Next let $\eta > d_{\mathcal{N}}(X, Y)$, and let $R \in \mathscr{R}(X, Y)$ be such that $\text{dis}(R) < 2\eta$. We wish to find networks X' and Y' such that $\widehat{d}_{\mathcal{N}}(X', Y') < \eta$. Write $Z = X \times Y$, and write $f : Z \rightarrow X$ and $g : Z \rightarrow Y$ to denote the (surjective) projection maps $(x, y) \mapsto x$ and $(x, y) \mapsto y$. Notice that we may write $R = \{(f(z), g(z)) : z \in R \subseteq Z\}$. In particular, by the definition of a correspondence, the restrictions of f, g to R are still surjective.

Define two weight functions $f^*\omega, g^*\omega : R \times R \rightarrow \mathbb{R}$ by $f^*\omega(z, z') = \omega_X(f(z), f(z'))$ and $g^*\omega(z, z') = \omega_Y(g(z), g(z'))$. Let $(U, \omega_U) = (R, f^*\omega)$ and let $(V, \omega_V) = (R, g^*\omega)$. Note that $d_{\mathcal{N}}(X, U) = 0$ by Remark 69, because $\text{card}(U) \geq \text{card}(X)$ and for all $z, z' \in U$, we have $\omega_U(z, z') = f^*\omega(z, z') = \omega_X(f(z), f(z'))$ for the surjective map f . Similarly $d_{\mathcal{N}}(Y, V) = 0$.

Next let $\varphi : U \rightarrow V$ be the bijection $z \mapsto z$. Then we have:

$$\begin{aligned} \sup_{z, z' \in U} |\omega_U(z, z') - \omega_V(\varphi(z), \varphi(z'))| &= \sup_{z, z' \in U} |\omega_U(z, z') - \omega_V(z, z')| \\ &= \sup_{z, z' \in R} |\omega_X(f(z), f(z')) - \omega_Y(g(z), g(z'))| \\ &= \sup_{(x, y), (x', y') \in R} |\omega_X(x, x') - \omega_Y(y, y')| \\ &= \text{dis}(R). \text{ In particular,} \\ \inf_{\varphi: U \rightarrow V \text{ bijection}} \text{dis}(\varphi) &\leq \text{dis}(R). \end{aligned}$$

So there exist networks U, V with the same node set (and thus the same cardinality) such that $\widehat{d}_{\mathcal{N}}(U, V) \leq \frac{1}{2} \text{dis}(R) < \eta$. We have already shown that $d_{\mathcal{N}}(X, Y) \leq \widehat{d}_{\mathcal{N}}(U, V)$. Since $\eta > d_{\mathcal{N}}(X, Y)$ was arbitrary, it follows that we have:

$$d_{\mathcal{N}}(X, Y) = \inf \left\{ \widehat{d}_{\mathcal{N}}(X', Y') : X' \cong_{\text{II}}^w X, Y' \cong_{\text{II}}^w Y, \text{ and } \text{card}(X') = \text{card}(Y') \right\}. \quad \square$$

2.2 ε -systems and finite sampling

In this section, we present proofs of results stated in §1.6.1.

Proof of Theorem 64. Once an ε -system has been found, the refinement can be produced by standard methods. So we focus on proving the existence of an ε -system. The idea is to find a cover of X by open sets G_1, \dots, G_q and representatives $x_i \in G_i$ for each $1 \leq i \leq q$ such that whenever we have $(x, x') \in G_i \times G_j$, we know by continuity of ω_X

that $|\omega_X(x, x') - \omega_X(x_i, x_j)| < \varepsilon$. Then we define a correspondence that associates each $x \in G_i$ to x_i , for $1 \leq i \leq q$. Such a correspondence has distortion bounded above by ε .

Let $\varepsilon > 0$. Let \mathcal{B} be a base for the topology on X .

Let $\{B(r, \varepsilon/4) : r \in \mathbb{R}\}$ be an open cover for \mathbb{R} . Then by continuity of ω_X , we get that

$$\{\omega_X^{-1}[B(r, \varepsilon/4)] : r \in \mathbb{R}\}$$

is an open cover for $X \times X$. Each open set in this cover can be written as a union of open rectangles $U \times V$, for $U, V \in \mathcal{B}$. Thus the following set is an open cover of $X \times X$:

$$\mathcal{U} := \{U \times V : U, V \in \mathcal{B}, U \times V \subseteq \omega_X^{-1}[B(r, \varepsilon/4)], r \in \mathbb{R}\}.$$

Claim 1. There exists a finite open cover $\mathcal{G} = \{G_1, \dots, G_q\}$ of X such that for any $1 \leq i, j \leq q$, we have $G_i \times G_j \subseteq U \times V$ for some $U \times V \in \mathcal{U}$.

Proof of Claim 1. The proof of the claim proceeds by a repeated application of the Tube Lemma [89, Lemma 26.8]. Since $X \times X$ is compact, we take a finite subcover:

$$\mathcal{U}^f := \{U_1 \times V_1, \dots, U_n \times V_n\}, \text{ for some } n \in \mathbb{N}.$$

Let $x \in X$. Then we define:

$$\mathcal{U}_x^f := \{U \times V \in \mathcal{U}^f : x \in U\},$$

and write

$$\mathcal{U}_x^f = \left\{ U_{i_1}^x \times V_{i_1}^x, \dots, U_{i_{m(x)}}^x \times V_{i_{m(x)}}^x \right\}.$$

Here $m(x)$ is an integer depending on x , and $\{i_1, \dots, i_{m(x)}\}$ is a subset of $\{1, \dots, n\}$.

Since \mathcal{U}^f is an open cover of $X \times X$, we know that \mathcal{U}_x^f is an open cover of $\{x\} \times X$. Next define:

$$A_x := \bigcap_{k=1}^{m(x)} U_{i_k}^x.$$

Then A_x is open and contains x . In the literature [89, p. 167], the set $A_x \times X$ is called a *tube* around $\{x\} \times X$. Notice that $A_x \times X \subseteq \mathcal{U}_x^f$. Since x was arbitrary in the preceding construction, we define \mathcal{U}_x^f and A_x for each $x \in X$. Then note that $\{A_x : x \in X\}$ is an open cover of X . Using compactness of X , we choose $\{s_1, \dots, s_p\} \subseteq X$, $p \in \mathbb{N}$, such that $\{A_{s_1}, \dots, A_{s_p}\}$ is a finite subcover of X .

Once again let $x \in X$, and let \mathcal{U}_x^f and A_x be defined as above. Define the following:

$$B_x := \{A_x \times V_{i_k}^x : 1 \leq k \leq m(x)\}.$$

Since $x \in A_x$ and $X \subseteq \bigcup_{k=1}^{m(x)} V_{i_k}^x$, it follows that B_x is a cover of $\{x\} \times X$. Furthermore, since $\{A_{s_1}, \dots, A_{s_p}\}$ is a cover of X , it follows that the finite collection $\{B_{s_1}, \dots, B_{s_p}\}$ is a cover of $X \times X$.

Let $z \in X$. Since $X \subseteq \cup_{k=1}^{m(x)} V_{i_k}^x$, we pick $V_{i_k}^x$ for $1 \leq k \leq m(x)$ such that $z \in V_{i_k}^x$. Since x was arbitrary, such a choice exists for each $x \in X$. Therefore, we define:

$$C_z := \{V \in \mathcal{B} : z \in V, A_{s_i} \times V \in B_{s_i} \text{ for some } 1 \leq i \leq p\}.$$

Since each B_{s_i} is finite and there are finitely many B_{s_i} , we know that C_z is a finite collection. Next define:

$$D_z := \bigcap_{V \in C_z} V.$$

Then D_z is open and contains z . Notice that $X \times D_z$ is a tube around $X \times \{z\}$. Next, using the fact that $\{A_{s_i} : 1 \leq i \leq p\}$ is an open cover of X , pick $A_{s_{i(z)}}$ such that $z \in A_{s_{i(z)}}$. Here $1 \leq i(z) \leq p$ is some integer depending on z . Then define

$$G_z := D_z \cap A_{s_{i(z)}}.$$

Then G_z is open and contains z . Since z was arbitrary, we define G_z for each $z \in X$. Then $\{G_z : z \in X\}$ is an open cover of X , and we take a finite subcover:

$$\mathcal{G} := \{G_1, \dots, G_q\}, q \in \mathbb{N}.$$

Finally, we need to show that for any choice of $1 \leq i, j \leq q$, we have $G_i \times G_j \subseteq U \times V$ for some $U \times V \in \mathcal{U}$. Let $1 \leq i, j \leq q$. Note that we can write $G_i = G_w$ and $G_j = G_y$ for some $w, y \in X$. By the definition of G_w , we then have the following for some index $i(w)$ depending on w :

$$G_w \subseteq A_{s_{i(w)}} \subseteq U^{s_{i(w)}} \text{ for some } U^{s_{i(w)}} \times V^{s_{i(w)}} \in \mathcal{U}_{s_{i(w)}}^f, 1 \leq i(w) \leq p.$$

Note that the second containment holds by definition of $A_{s_{i(w)}}$. Since $\mathcal{U}_{s_{i(w)}}^f$ is a cover of $\{s_{i(w)}\} \times X$, we choose $V^{s_{i(w)}}$ to contain y . Then observe that $A_{s_{i(w)}} \times V^{s_{i(w)}} \in B_{s_{i(w)}}$. Then $V^{s_{i(w)}} \in C_y$, and so we have:

$$G_y \subseteq D_y \subseteq V^{s_{i(w)}}.$$

It follows that $G_i \times G_j = G_w \times G_y \subseteq U^{s_{i(w)}} \times V^{s_{i(w)}} \in \mathcal{U}$. ■

Now we fix $\mathcal{G} = \{G_1, \dots, G_q\}$ as in Claim 1. Before defining X' , we perform a disjointification step. Define:

$$\tilde{G}_1 := G_1, \tilde{G}_2 := G_2 \setminus \tilde{G}_1, \tilde{G}_3 := G_3 \setminus (\tilde{G}_1 \cup \tilde{G}_2), \dots, \tilde{G}_q := G_q \setminus \left(\cup_{k=1}^{q-1} \tilde{G}_k \right).$$

Finally we define X' as follows: pick a representative $x_i \in \tilde{G}_i$ for each $1 \leq i \leq q$. Let $X' = \{x_i : 1 \leq i \leq q\}$. Define a correspondence between X and X' as follows:

$$R := \{(x, x_i) : x \in \tilde{G}_i, 1 \leq i \leq q\}.$$

Let $(x, x_i), (x', x_j) \in R$. Then we have $(x, x'), (x_i, x_j) \in \tilde{G}_i \times \tilde{G}_j \subseteq G_i \times G_j$. By the preceding work, we know that $G_i \times G_j \subseteq U \times V$, for some $U \times V \in \mathcal{U}$. Therefore $\omega_X(x, x'), \omega_X(x_i, x_j) \in B(r, \varepsilon/4)$ for some $r \in \mathbb{R}$. It follows that:

$$|\omega_X(x, x') - \omega_X(x_i, x_j)| < \varepsilon/2.$$

Since $(x, x_i), (x', x_j) \in R$ were arbitrary, we have $\text{dis}(R) < \varepsilon/2$. Hence $d_{\mathcal{N}}(X, X') < \varepsilon$. \square

Proof of Theorem 67. The first part of this proof is similar to that of Theorem 64. Let $\varepsilon > 0$. Let \mathcal{B} be a base for the topology on X . Then $\{\omega_X^{-1}[B(r, \varepsilon/8)] : r \in \mathbb{R}\}$ is an open cover for $X \times X$. Each open set in this cover can be written as a union of open rectangles $U \times V$, for $U, V \in \mathcal{B}$. Thus the following set is an open cover of $X \times X$:

$$\mathcal{U} := \{U \times V : U, V \in \mathcal{B}, U \times V \subseteq \omega_X^{-1}[B(r, \varepsilon/8)], r \in \mathbb{R}\}.$$

By applying Claim 1 from the proof of Theorem 64, we obtain a finite open cover $\mathcal{G} = \{G_1, \dots, G_q\}$ of X such that for any $1 \leq i, j \leq q$, we have $G_i \times G_j \subseteq U \times V$ for some $U \times V \in \mathcal{U}$. For convenience, we assume that each G_i is nonempty.

Now let $1 \leq i \leq q$. Then $G_i \cap S \neq \emptyset$, because S is dense in X . Choose $p(i) \in \mathbb{N}$ such that $s_{p(i)} \in G_i$. We repeat this process for each $1 \leq i \leq q$, and then define

$$n := \max \{p(1), p(2), \dots, p(q)\}.$$

Now define X_n to be the network with node set $\{s_1, s_2, \dots, s_n\}$ and weight function given by the appropriate restriction of ω_X . Also define S_n to be the network with node set $\{s_{p(1)}, s_{p(2)}, \dots, s_{p(q)}\}$ and weight function given by the restriction of ω_X .

Claim 2. Let A be a subset of X equipped with the weight function $\omega_X|_{A \times A}$. Then $d_{\mathcal{N}}(S_n, A) < \varepsilon/2$.

Proof of Claim 2. We begin with $\mathcal{G} = \{G_1, \dots, G_q\}$. Notice that each G_i contains $s_{p(i)}$. To avoid ambiguity in our construction, we will need to ensure that G_i does not contain $s_{p(j)}$ for $i \neq j$. So our first step is to obtain a cover of A by disjoint sets while ensuring that each $s_{p(i)} \in S_n$ belongs to exactly one element of the new cover. We define:

$$\begin{aligned} G_1^* &:= G_1 \setminus S_n, \quad G_2^* := G_2 \setminus S_n, \quad G_3^* := G_3 \setminus S_n, \dots, \quad G_q^* := G_q \setminus S_n, \text{ and} \\ \tilde{G}_1 &:= G_1^* \cup \{s_{p(1)}\}, \quad \tilde{G}_2 := (G_2^* \setminus \tilde{G}_1) \cup \{s_{p(2)}\}, \quad \tilde{G}_3 := \left(G_3^* \setminus (\tilde{G}_1 \cup \tilde{G}_2)\right) \cup \{s_{p(3)}\}, \dots, \\ \tilde{G}_q &:= \left(G_q^* \setminus \left(\bigcup_{k=1}^{q-1} \tilde{G}_k\right)\right) \cup \{s_{p(q)}\}. \end{aligned}$$

Notice that $\{\tilde{G}_i : 1 \leq i \leq q\}$ is a cover for A , and for each $1 \leq i \leq q$, \tilde{G}_i contains $s_{p(j)}$ if and only if $i = j$. Now we define a correspondence between A and S_n as follows:

$$R := \{(x, s_{p(i)}) : x \in A \cap \tilde{G}_i, 1 \leq i \leq q\}.$$

Next let $(x, s_{p(i)}), (x', s_{p(j)}) \in R$. Then we have $(x, x'), (s_{p(i)}, s_{p(j)}) \in \tilde{G}_i \times \tilde{G}_j \subseteq G_i \times G_j \subseteq U \times V$ for some $U \times V \in \mathcal{U}$. Therefore $\omega_X(x, x')$ and $\omega_X(s_{p(i)}, s_{p(j)})$ both belong to $B(r, \varepsilon/8)$ for some $r \in \mathbb{R}$. Thus we have:

$$|\omega_X(x, x') - \omega_X(s_{p(i)}, s_{p(j)})| < \varepsilon/4.$$

It follows that $\text{dis}(R) < \varepsilon/4$, and so $d_N(A, S_n) < \varepsilon/2$. ■

Finally, we note that $d_N(X, X_n) \leq d_N(X, S_n) + d_N(S_n, X_n) < \varepsilon/2 + \varepsilon/2 = \varepsilon$, by Claim 2. Since $\varepsilon > 0$ was arbitrary, it follows that $d_N(X, X_n) \rightarrow 0$.

For the final statement in the theorem, let $m \geq n$ and observe that $S_n \subseteq X_n \subseteq X_m$. Thus whenever we have $d_N(X, X_n) < \varepsilon$, we also have $d_N(X, X_m) < \varepsilon$. It follows that:

$$d_N(X, X_m) \leq d_N(X, X_n) \text{ for any } m, n \in \mathbb{N}, m \geq n.$$
□

Next we proceed to Theorem 68. We first prove the following useful lemma:

Lemma 130. *Assume the setup of (X, ω_X) , μ_X , $(\Omega, \mathcal{F}, \mathbb{P})$, and \mathbb{X}_n for each $n \in \mathbb{N}$ as in Theorem 68. Fix $\varepsilon > 0$, and let $\mathcal{U} = \{U_1, \dots, U_m\}$ be a refined ε -system on $\text{supp}(\mu_X)$. For each $1 \leq i \leq m$ and each $n \in \mathbb{N}$, define the following event:*

$$A_i := \bigcap_{k=1}^n \{\omega \in \Omega : x_k(\omega) \notin U_i\} \subseteq \Omega.$$

Then we have $\mathbb{P}(\bigcup_{k=1}^m A_k) \leq \frac{1}{m(\mathcal{U})}(1 - \mathfrak{m}(\mathcal{U}))^n$.

Proof of Lemma 130. Here we are considering the probability that at least one of the U_i has empty intersection with \mathbb{X}_n . By independence, $\mathbb{P}(A_i) = (1 - \mu_X(U_i))^n$. Then we have:

$$\mathbb{P}\left(\bigcup_{k=1}^m A_k\right) \leq \sum_{k=1}^m \mathbb{P}(A_k) = \sum_{k=1}^m (1 - \mu_X(U_k))^n \leq m \cdot \max_{1 \leq k \leq m} (1 - \mu_X(U_k))^n \leq \frac{(1 - \mathfrak{m}(\mathcal{U}))^n}{m(\mathcal{U})}.$$

Here the first inequality follows by subadditivity of measure, and the last inequality follows because the total mass $\mu_X(\text{supp}(\mu_X)) = 1$ is an upper bound for $m \cdot \mathfrak{m}(\mathcal{U})$. Note also that each $U \in \mathcal{U}$ has nonzero mass, by the observation in Definition 24. □

Proof of Theorem 68. By endowing $\text{supp}(\mu_X)$ with the restriction of ω_X to $\text{supp}(\mu_X) \times \text{supp}(\mu_X)$ it may itself be viewed as a network with full support, so for notational convenience, we assume $X = \text{supp}(\mu_X)$.

First observe that $\mathfrak{M}_{\varepsilon/2}(X) \in (0, 1]$. Let $r \in (0, \mathfrak{M}_{\varepsilon/2}(X))$, and let \mathcal{U}_r be an $\varepsilon/2$ -system on X such that $\mathfrak{m}(\mathcal{U}_r) \in (r, \mathfrak{M}_{\varepsilon/2}(X)]$. For convenience, write $m := |\mathcal{U}_r|$, and also write $\mathcal{U}_r = \{U_1, \dots, U_m\}$. For each $1 \leq i \leq m$, define A_i as in the statement of Lemma 130. Then by Lemma 130, the probability that at least one U_i has empty intersection with \mathbb{X}_n is bounded as $\mathbb{P}(\bigcup_{k=1}^m A_k) \leq \frac{1}{m(\mathcal{U}_r)}(1 - \mathfrak{m}(\mathcal{U}_r))^n$. On the other hand, if U_i has nonempty

intersection with \mathbb{X}_n for each $1 \leq i \leq m$, then by Theorem 66, we obtain $d_{\mathcal{N}}(X, \mathbb{X}_n) < \varepsilon$. For each $n \in \mathbb{N}$, define: $B_n := \{\omega \in \Omega : d_{\mathcal{N}}(X, \mathbb{X}_n(\omega)) \geq \varepsilon\}$. Then we have:

$$\mathbb{P}(B_n) \leq \mathbb{P}\left(\bigcup_{k=1}^m A_k\right) \leq \frac{(1 - \mathfrak{m}(\mathcal{U}_r))^n}{\mathfrak{m}(\mathcal{U}_r)}.$$

Since $r \in (0, \mathfrak{M}_{\varepsilon/2}(X))$ was arbitrary, letting r approach $\mathfrak{M}_{\varepsilon/2}(X)$ shows that $\mathbb{P}(B_n) \leq \frac{(1 - \mathfrak{M}_{\varepsilon/2}(X))^n}{\mathfrak{M}_{\varepsilon/2}(X)}$. We have by Definition 24 that $\mathfrak{M}_{\varepsilon/2}(X)$ is strictly positive. Thus the term on the right side of the inequality is an element of a convergent geometric series, so

$$\sum_{n=1}^{\infty} \mathbb{P}(B_n) \leq \frac{1}{\mathfrak{M}_{\varepsilon/2}(X)} \sum_{n=1}^{\infty} (1 - \mathfrak{M}_{\varepsilon/2}(X))^n < \infty.$$

By the Borel-Cantelli lemma, we have $\mathbb{P}(\limsup_{n \rightarrow \infty} B_n) = 0$. The result follows. \square

2.3 Proofs from §1.6.2

Proof of Proposition 70. The case for Type I weak isomorphism is similar to that of Type II, so we omit it. For Type II weak isomorphism, the reflexive and symmetric properties are easy to see, so we only provide details for verifying transitivity. Let $A, B, C \in \mathcal{N}$ be such that $A \cong_{\text{II}}^w B$ and $B \cong_{\text{II}}^w C$. Let $\varepsilon > 0$, and let P, S be sets with surjective maps $\varphi_A : P \rightarrow A, \varphi_B : P \rightarrow B, \psi_B : S \rightarrow B, \psi_C : S \rightarrow C$ such that:

$$\begin{aligned} |\omega_A(\varphi_A(p), \varphi_A(p')) - \omega_B(\varphi_B(p), \varphi_B(p'))| &< \varepsilon/2 \quad \text{for each } p, p' \in P, \text{ and} \\ |\omega_B(\psi_B(s), \psi_B(s')) - \omega_C(\psi_C(s), \psi_C(s'))| &< \varepsilon/2 \quad \text{for each } s, s' \in S. \end{aligned}$$

Next define $T := \{(p, s) \in P \times S : \varphi_B(p) = \psi_B(s)\}$.

Claim 3. The projection maps $\pi_P : T \rightarrow P$ and $\pi_S : T \rightarrow S$ are surjective.

Proof. Let $p \in P$. Then $\varphi_B(p) \in B$, and since $\psi_B : S \rightarrow B$ is surjective, there exists $s \in S$ such that $\psi_B(s) = \varphi_B(p)$. Thus $(p, s) \in T$, and $\pi_P(p, s) = p$. This suffices to show that $\pi_P : T \rightarrow P$ is a surjection. The case for $\pi_S : T \rightarrow S$ is similar. \blacksquare

It follows from the preceding claim that $\varphi_A \circ \pi_P : T \rightarrow A$ and $\psi_C \circ \pi_S : T \rightarrow C$ are surjective. Next let $(p, s), (p', s') \in T$. Then,

$$\begin{aligned} &|\omega_A(\varphi_A(\pi_P(p, s)), \varphi_A(\pi_P(p', s'))) - \omega_C(\psi_C(\pi_S(p, s)), \psi_C(\pi_S(p', s')))| \\ &= |\omega_A(\varphi_A(p), \varphi_A(p')) - \omega_C(\psi_C(s), \psi_C(s'))| \\ &= |\omega_A(\varphi_A(p), \varphi_A(p')) - \omega_B(\varphi_B(p), \varphi_B(p')) + \omega_B(\varphi_B(p), \varphi_B(p')) - \omega_C(\psi_C(s), \psi_C(s'))| \\ &= |\omega_A(\varphi_A(p), \varphi_A(p')) - \omega_B(\varphi_B(p), \varphi_B(p')) + \omega_B(\psi_B(s), \psi_B(s')) - \omega_C(\psi_C(s), \psi_C(s'))| \\ &< \varepsilon/2 + \varepsilon/2 = \varepsilon. \end{aligned}$$

Since $\varepsilon > 0$ was arbitrary, it follows that $A \cong_{\text{II}}^w C$. \square

Proof of Theorem 72. It is clear that $d_{\mathcal{N}}(X, Y) \geq 0$. To show $d_{\mathcal{N}}(X, X) = 0$, consider the correspondence $R = \{(x, x) : x \in X\}$. Then for any $(x, x), (x', x') \in R$, we have $|\omega_X(x, x') - \omega_X(x, x')| = 0$. Thus $\text{dis}(R) = 0$ and $d_{\mathcal{N}}(X, X) = 0$.

Next we show symmetry, i.e. $d_{\mathcal{N}}(X, Y) \leq d_{\mathcal{N}}(Y, X)$ and $d_{\mathcal{N}}(Y, X) \leq d_{\mathcal{N}}(X, Y)$. The two cases are similar, so we just show the second inequality. Let $\eta > d_{\mathcal{N}}(X, Y)$. Let $R \in \mathcal{R}(X, Y)$ be such that $\text{dis}(R) < 2\eta$. Then define $\tilde{R} = \{(y, x) : (x, y) \in R\}$. Note that $\tilde{R} \in \mathcal{R}(Y, X)$. We have:

$$\begin{aligned}\text{dis}(\tilde{R}) &= \sup_{(y, x), (y', x') \in \tilde{R}} |\omega_Y(y, y') - \omega_X(x, x')| \\ &= \sup_{(x, y), (x', y') \in R} |\omega_Y(y, y') - \omega_X(x, x')| \\ &= \sup_{(x, y), (x', y') \in R} |\omega_X(x, x') - \omega_Y(y, y')| = \text{dis}(R).\end{aligned}$$

So $\text{dis}(R) = \text{dis}(\tilde{R})$. Then $d_{\mathcal{N}}(Y, X) = \frac{1}{2} \inf_{S \in \mathcal{R}(Y, X)} \text{dis}(S) \leq \frac{1}{2} \text{dis}(\tilde{R}) < \eta$. This shows $d_{\mathcal{N}}(Y, X) \leq d_{\mathcal{N}}(X, Y)$. The reverse inequality follows by a similar argument.

Next we prove the triangle inequality. Let $R \in \mathcal{R}(X, Y)$, $S \in \mathcal{R}(Y, Z)$, and let

$$R \circ S = \{(x, z) \in X \times Z \mid \exists y, (x, y) \in R, (y, z) \in S\}$$

First we claim that $R \circ S \in \mathcal{R}(X, Z)$. This is equivalent to checking that for each $x \in X$, there exists z such that $(x, z) \in R \circ S$, and for each $z \in Z$, there exists x such that $(x, z) \in R \circ S$. The proofs of these two conditions are similar, so we just prove the former. Let $x \in X$. Let $y \in Y$ be such that $(x, y) \in R$. Then there exists $z \in Z$ such that $(y, z) \in S$. Then $(x, z) \in R \circ S$.

Next we claim that $\text{dis}(R \circ S) \leq \text{dis}(R) + \text{dis}(S)$. Let $(x, z), (x', z') \in R \circ S$. Let $y \in Y$ be such that $(x, y) \in R$ and $(y, z) \in S$. Let $y' \in Y$ be such that $(x', y') \in R$, $(y', z') \in S$. Then we have:

$$\begin{aligned}|\omega_X(x, x') - \omega_Z(z, z')| &= |\omega_X(x, x') - \omega_Y(y, y') + \omega_Y(y, y') - \omega_Z(z, z')| \\ &\leq |\omega_X(x, x') - \omega_Y(y, y')| + |\omega_Y(y, y') - \omega_Z(z, z')| \\ &\leq \text{dis}(R) + \text{dis}(S).\end{aligned}$$

This holds for any $(x, z), (x', z') \in R \circ S$, and proves the claim.

Now let $\eta_1 > d_{\mathcal{N}}(X, Y)$, let $\eta_2 > d_{\mathcal{N}}(Y, Z)$, and let $R \in \mathcal{R}(X, Y)$, $S \in \mathcal{R}(Y, Z)$ be such that $\text{dis}(R) < 2\eta_1$ and $\text{dis}(S) < 2\eta_2$. Then we have:

$$d_{\mathcal{N}}(X, Z) \leq \frac{1}{2} \text{dis}(R \circ S) \leq \frac{1}{2} \text{dis}(R) + \frac{1}{2} \text{dis}(S) < 2\eta_1 + 2\eta_2.$$

This shows that $d_{\mathcal{N}}(X, Z) \leq d_{\mathcal{N}}(X, Y) + d_{\mathcal{N}}(Y, Z)$, and proves the triangle inequality.

Finally, we claim that $X \cong_{\Pi}^w Y$ if and only if $d_{\mathcal{N}}(X, Y) = 0$. Suppose $d_{\mathcal{N}}(X, Y) = 0$. Let $\varepsilon > 0$, and let $R(\varepsilon) \in \mathcal{R}(X, Y)$ be such that $\text{dis}(R(\varepsilon)) < \varepsilon$. Then for any $z = (x, y), z' = (x', y') \in R(\varepsilon)$, we have $|\omega_X(x, x') - \omega_Y(y, y')| < \varepsilon$. But this is equivalent

to writing $|\omega_X(\pi_X(z), \pi_X(z')) - \omega_Y(\pi_Y(z), \pi_Y(z'))| < \varepsilon$, where $\pi_X : R(\varepsilon) \rightarrow X$ and $\pi_Y : R(\varepsilon) \rightarrow Y$ are the canonical projection maps. This holds for each $\varepsilon > 0$. Thus $X \cong_{\text{II}}^w Y$.

Conversely, suppose $X \cong_{\text{II}}^w Y$, and for each $\varepsilon > 0$ let $Z(\varepsilon)$ be a set with surjective maps $\phi_X^\varepsilon : Z(\varepsilon) \rightarrow X$, $\phi_Y^\varepsilon : Z(\varepsilon) \rightarrow Y$ such that $|\omega_X(\phi_X^\varepsilon(z), \phi_X^\varepsilon(z')) - \omega_Y(\phi_Y^\varepsilon(z), \phi_Y^\varepsilon(z'))| < \varepsilon$ for all $z, z' \in Z(\varepsilon)$. For each $\varepsilon > 0$, let $R(\varepsilon) = \{(\phi_X^\varepsilon(z), \phi_Y^\varepsilon(z)) : z \in Z(\varepsilon)\}$. Then $R(\varepsilon) \in \mathcal{R}(X, Y)$ for each $\varepsilon > 0$, and $\text{dis}(R(\varepsilon)) = \sup_{z, z' \in Z} |\omega_X(\phi_X^\varepsilon(z), \phi_X^\varepsilon(z')) - \omega_Y(\phi_Y^\varepsilon(z), \phi_Y^\varepsilon(z'))| < \varepsilon$.

We conclude that $d_{\mathcal{N}}(X, Y) = 0$. Thus $d_{\mathcal{N}}$ is a metric modulo Type II weak isomorphism. \square

Proof of Theorem 73. By the definition of \cong_1^w , it is clear that if $X \cong_1^w Y$, then $d_{\mathcal{N}}(X, Y) = 0$, i.e. $X \cong_{\text{II}}^w Y$ (cf. Theorem 72).

Conversely, suppose $d_{\mathcal{N}}(X, Y) = 0$. Our strategy is to obtain a set $Z \subseteq X \times Y$ with canonical projection maps $\pi_X : Z \rightarrow X$, $\pi_Y : Z \rightarrow Y$ and surjections $\psi_X : X \rightarrow \pi_X(Z)$, $\psi_Y : Y \rightarrow \pi_Y(Z)$ as in the following diagram:

$$\begin{array}{ccccccc} & X & & Z & & Y & \\ \text{id}_X \swarrow & \searrow \psi_X & \pi_X \swarrow & \searrow \pi_Y & \swarrow \psi_Y & \searrow \text{id}_Y & \\ X & \cong_1^w & \pi_X(Z) & \cong_1^w & \pi_Y(Z) & \cong_1^w & Y \end{array}$$

Furthermore, we will require:

$$\omega_X(\pi_X(z), \pi_X(z')) = \omega_Y(\pi_Y(z), \pi_Y(z')) \quad \text{for all } z, z' \in Z, \quad (2.1)$$

$$\omega_X(x, x') = \omega_X(\psi_X(x), \psi_X(x')) \quad \text{for all } x, x' \in X, \quad (2.2)$$

$$\omega_Y(y, y') = \omega_Y(\psi_Y(y), \psi_Y(y')) \quad \text{for all } y, y' \in Y. \quad (2.3)$$

As a consequence, we will obtain a chain of Type I weak isomorphisms

$$X \cong_1^w \pi_X(Z) \cong_1^w \pi_Y(Z) \cong_1^w Y.$$

Since Type I weak isomorphism is an equivalence relation (Proposition 70), it will follow that X and Y are Type I weakly isomorphic.

By applying Theorem 64, we choose sequences of finite subnetworks $\{X_n \subseteq X : n \in \mathbb{N}\}$ and $\{Y_n \subseteq Y : n \in \mathbb{N}\}$ such that $d_{\mathcal{N}}(X_n, X) < 1/n$ and $d_{\mathcal{N}}(Y_n, Y) < 1/n$ for each $n \in \mathbb{N}$. By the triangle inequality, $d_{\mathcal{N}}(X_n, Y_n) < 2/n$ for each n .

For each $n \in \mathbb{N}$, let $T_n \in \mathcal{R}(X_n, X)$, $P_n \in \mathcal{R}(Y, Y_n)$ be such that $\text{dis}(T_n) < 2/n$ and $\text{dis}(P_n) < 2/n$. Define $\alpha_n := 4/n - \text{dis}(T_n) - \text{dis}(P_n)$, and notice that $\alpha_n \rightarrow 0$ as $n \rightarrow \infty$. Since $d_{\mathcal{N}}(X, Y) = 0$ by assumption, for each $n \in \mathbb{N}$ we let $S_n \in \mathcal{R}(X, Y)$ be such that $\text{dis}(S_n) < \alpha_n$. Then,

$$\text{dis}(T_n \circ S_n \circ P_n) \leq \text{dis}(T_n) + \text{dis}(S_n) + \text{dis}(P_n) < 4/n. \quad (\text{cf. Remark 4})$$

Then for each $n \in \mathbb{N}$, we define $R_n := T_n \circ S_n \circ P_n \in \mathcal{R}(X_n, Y_n)$. By Remark 4, we know that R_n has the following expression:

$$R_n = \{(x_n, y_n) \in X_n \times Y_n : \text{there exist } \tilde{x} \in X, \tilde{y} \in Y \text{ such that } (x_n, \tilde{x}) \in T_n, \\ (\tilde{x}, \tilde{y}) \in S_n, (\tilde{y}, y_n) \in P_n\}.$$

Next define:

$$\mathcal{S} := \{(\tilde{x}_n, \tilde{y}_n)_{n \in \mathbb{N}} \in (X \times Y)^{\mathbb{N}} : (\tilde{x}_n, \tilde{y}_n) \in S_n \text{ for each } n \in \mathbb{N}\}.$$

Since X, Y are first countable and compact, the product $X \times Y$ is also first countable and compact, hence sequentially compact. Any sequence in a sequentially compact space has a convergent subsequence, so for convenience, we replace each sequence in \mathcal{S} by a convergent subsequence. Next define:

$$Z := \{(x, y) \in X \times Y : (x, y) \text{ a limit point of some } (\tilde{x}_n, \tilde{y}_n)_{n \in \mathbb{N}} \in \mathcal{S}\}.$$

Claim 4. Z is a closed subspace of $X \times Y$. Hence it is compact and sequentially compact.

The second statement in the claim follows from the first: assuming that Z is a closed subspace of the compact space $X \times Y$, we obtain that Z is compact. Any subspace of a first countable space is first countable, so Z is also first countable. Next, observe that $\pi_X(Z)$ equipped with the subspace topology is compact, because it is a continuous image of a compact space. It is also first countable because it is a subspace of the first countable space X . Furthermore, the restriction of ω_X to $\pi_X(Z)$ is continuous. Thus $\pi_X(Z)$ equipped with the restriction of ω_X is a compact network, and by similar reasoning, we get that $\pi_Y(Z)$ equipped with the restriction of ω_Y is also a compact network.

Proof of Claim 4. We will show that $Z \subseteq X \times Y$ contains all its limit points. Let $(x, y) \in X \times Y$ be a limit point of Z . Let $\{U_n \subseteq X : n \in \mathbb{N}, (x, y) \in U_n\}$ be a countable neighborhood base of (x, y) . For each $n \in \mathbb{N}$, the finite intersection $V_n := \cap_{i=1}^n U_i$ is an open neighborhood of (x, y) , and thus contains a point $(x_n, y_n) \in Z$ that is distinct from (x, y) (by the definition of a limit point). Pick such an (x_n, y_n) for each $n \in \mathbb{N}$. Then $(x_n, y_n)_{n \in \mathbb{N}}$ is a sequence in Z converging to (x, y) such that $(x_n, y_n) \in V_n$ for each $n \in \mathbb{N}$.

For each $n \in \mathbb{N}$, note that because $(x_n, y_n) \in Z$ and V_n is an open neighborhood of (x_n, y_n) , there exists a sequence in \mathcal{S} converging to (x_n, y_n) for which all but finitely many terms are contained in V_n . So for each $n \in \mathbb{N}$, let $(\tilde{x}_n, \tilde{y}_n) \in S_n$ be such that $(\tilde{x}_n, \tilde{y}_n) \in V_n$. Then the sequence $(\tilde{x}_n, \tilde{y}_n)_{n \in \mathbb{N}} \in \mathcal{S}$ converges to (x, y) . Thus $(x, y) \in Z$. Since (x, y) was an arbitrary limit point of Z , it follows that Z is closed. ■

Proof of Equation 2.1. We now prove Equation 2.1. Let $z = (x, y)$, $z' = (x', y') \in Z$, and let $(\tilde{x}_n, \tilde{y}_n)_{n \in \mathbb{N}}, (\tilde{x}'_n, \tilde{y}'_n)_{n \in \mathbb{N}}$ be elements of \mathcal{S} that converge to $(x, y), (x', y')$ respectively.

We wish to show $|\omega_X(x, x') - \omega_Y(y, y')| = 0$. Let $\varepsilon > 0$, and observe that:

$$\begin{aligned} & |\omega_X(x, x') - \omega_Y(y, y')| \\ &= |\omega_X(x, x') - \omega_X(\tilde{x}_n, \tilde{x}'_n) + \omega_X(\tilde{x}_n, \tilde{x}'_n) - \omega_Y(\tilde{y}_n, \tilde{y}'_n) + \omega_Y(\tilde{y}_n, \tilde{y}'_n) - \omega_Y(y, y')| \\ &\leq |\omega_X(x, x') - \omega_X(\tilde{x}_n, \tilde{x}'_n)| + |\omega_X(\tilde{x}_n, \tilde{x}'_n) - \omega_Y(\tilde{y}_n, \tilde{y}'_n)| + |\omega_Y(\tilde{y}_n, \tilde{y}'_n) - \omega_Y(y, y')|. \end{aligned}$$

Claim 5. Suppose we are given sequences $(\tilde{x}_n, \tilde{y}_n)_{n \in \mathbb{N}}, (\tilde{x}'_n, \tilde{y}'_n)_{n \in \mathbb{N}}$ in Z converging to (x, y) and (x', y') in Z , respectively. Then there exists $N \in \mathbb{N}$ such that for all $n \geq N$, we have:

$$|\omega_X(x, x') - \omega_X(\tilde{x}_n, \tilde{x}'_n)| < \varepsilon/4, \quad |\omega_Y(\tilde{y}_n, \tilde{y}'_n) - \omega_Y(y, y')| < \varepsilon/4.$$

Proof of Claim 5. Write $a := \omega_X(x, x'), b := \omega_Y(y, y')$. Since ω_X, ω_Y are continuous, we know that $\omega_X^{-1}[B(a, \varepsilon/4)]$ and $\omega_Y^{-1}[B(b, \varepsilon/4)]$ are open neighborhoods of (x, x') and (y, y') . Since each open set in the product space $X \times X$ is a union of open rectangles of the form $A \times A'$ for A, A' open subsets of X , we choose an open set $A \times A' \subseteq \omega_X^{-1}[B(a, \varepsilon/4)]$ such that $(x, x') \in A \times A'$. Similarly, we choose an open set $B \times B' \subseteq \omega_Y^{-1}[B(b, \varepsilon/4)]$ such that $(y, y') \in B \times B'$. Then $A \times B, A' \times B'$ are open neighborhoods of $(x, y), (x', y')$ respectively. Since $(\tilde{x}_n, \tilde{y}_n)_{n \in \mathbb{N}}$ and $(\tilde{x}'_n, \tilde{y}'_n)_{n \in \mathbb{N}}$ converge to (x, y) and (x', y') , respectively, we choose $N \in \mathbb{N}$ such that for all $n \geq N$, we have $(\tilde{x}_n, \tilde{y}_n) \in A \times B$ and $(\tilde{x}'_n, \tilde{y}'_n) \in A' \times B'$. The claim now follows. ■

Now choose $N \in \mathbb{N}$ such that the property in Claim 5 is satisfied, as well as the additional property that $8/N < \varepsilon/4$. Then for any $n \geq N$, we have:

$$|\omega_X(x, x') - \omega_Y(y, y')| \leq \varepsilon/4 + |\omega_X(\tilde{x}_n, \tilde{x}'_n) - \omega_Y(\tilde{y}_n, \tilde{y}'_n)| + \varepsilon/4.$$

Separately note that for each $n \in \mathbb{N}$, having $(\tilde{x}_n, \tilde{y}_n), (\tilde{x}'_n, \tilde{y}'_n) \in S_n$ implies that there exist (x_n, y_n) and $(x'_n, y'_n) \in R_n$ such that $(x_n, \tilde{x}_n), (x'_n, \tilde{x}'_n) \in T_n$ and $(\tilde{y}_n, y_n), (\tilde{y}'_n, y'_n) \in P_n$. Thus we can bound the middle term above as follows:

$$\begin{aligned} & |\omega_X(\tilde{x}_n, \tilde{x}'_n) - \omega_Y(\tilde{y}_n, \tilde{y}'_n)| \\ &= |\omega_X(\tilde{x}_n, \tilde{x}'_n) - \omega_X(x_n, x'_n) + \omega_X(x_n, x'_n) - \omega_Y(y_n, y'_n) + \omega_Y(y_n, y'_n) - \omega_Y(\tilde{y}_n, \tilde{y}'_n)| \\ &\leq |\omega_X(\tilde{x}_n, \tilde{x}'_n) - \omega_X(x_n, x'_n)| + |\omega_X(x_n, x'_n) - \omega_Y(y_n, y'_n)| + |\omega_Y(y_n, y'_n) - \omega_Y(\tilde{y}_n, \tilde{y}'_n)| \\ &\leq \text{dis}(T_n) + \text{dis}(R_n) + \text{dis}(P_n) < 8/n \leq 8/N < \varepsilon/4. \end{aligned}$$

The preceding calculations show that:

$$|\omega_X(x, x') - \omega_Y(y, y')| < \varepsilon.$$

Since $\varepsilon > 0$ was arbitrary, it follows that $\omega_X(x, x') = \omega_Y(y, y')$. This proves Equation 2.1.

It remains to define surjective maps $\psi_X : X \rightarrow \pi_X(Z), \psi_Y : Y \rightarrow \pi_Y(Z)$ and to verify Equations 2.2 and 2.3. Both cases are similar, so we only show the details of constructing ψ_X and verifying Equation 2.2.

Construction of ψ_X . Let $x \in X$. Suppose first that $x \in \pi_X(Z)$. Then we simply define $\psi_X(x) = x$. We also make the following observation, to be used later: for each $n \in \mathbb{N}$, letting $y \in Y$ be such that $(x, y) \in S_n$, there exists $x_n \in X_n$ and $y_n \in Y_n$ such that $(x_n, x) \in T_n$ and $(y, y_n) \in P_n$.

Next suppose $x \in X \setminus \pi_X(Z)$. For each $n \in \mathbb{N}$, let $x_n \in X_n$ be such that $(x_n, x) \in T_n$, and let $\tilde{x}_n \in X$ be such that $(x_n, \tilde{x}_n) \in T_n$. Also for each $n \in \mathbb{N}$, let $\tilde{y}_n \in Y$ be such that $(\tilde{x}_n, \tilde{y}_n) \in S_n$. Then for each $n \in \mathbb{N}$, let $y_n \in Y_n$ be such that $(\tilde{y}_n, y_n) \in P_n$. Then by sequential compactness of $X \times Y$, the sequence $(\tilde{x}_n, \tilde{y}_n)_{n \in \mathbb{N}}$ has a convergent subsequence which belongs to \mathcal{S} and converges to a point $(\tilde{x}, \tilde{y}) \in Z$. In particular, we obtain a sequence $(\tilde{x}_n)_{n \in \mathbb{N}}$ converging to a point \tilde{x} , such that (x_n, x) and $(x_n, \tilde{x}_n) \in T_n$ for each $n \in \mathbb{N}$. Define $\psi_X(x) = \tilde{x}$.

Since $x \in X$ was arbitrary, this construction defines $\psi_X : X \rightarrow \pi_X(Z)$. Note that ψ_X is simply the identity on $\pi_X(Z)$, hence is surjective.

Proof of Equation 2.2. Now we verify Equation 2.2. Let $\varepsilon > 0$. There are three cases to check:

Case 1: $x, x' \in \pi_X(Z)$ In this case, we have:

$$|\omega_X(x, x') - \omega_X(\psi_X(x), \psi_X(x'))| = \omega_X(x, x') - \omega_X(x, x') = 0.$$

Case 2: $x, x' \in X \setminus \pi_X(Z)$ By continuity of ω_X , we obtain an open neighborhood $U := \omega_X^{-1}[B(\omega_X(\psi_X(x), \psi_X(x')), \varepsilon/2)]$ of (x, x') . By the definition of ψ_X on $X \setminus \pi_X(Z)$, we obtain sequences $(\tilde{x}_n, \tilde{y}_n)_{n \in \mathbb{N}}$ and $(\tilde{x}'_n, \tilde{y}'_n)_{n \in \mathbb{N}}$ in \mathcal{S} converging to $(\psi_X(x), \tilde{y})$ and $(\psi_X(x'), \tilde{y}')$ for some $\tilde{y}, \tilde{y}' \in Y$. By applying Claim 5, we obtain $N \in \mathbb{N}$ such that for all $n \geq N$, we have $(\tilde{x}_n, \tilde{x}'_n) \in U$. Note that we also obtain sequences $(x_n)_{n \in \mathbb{N}}$ and $(x'_n)_{n \in \mathbb{N}}$ such that $(x_n, x), (x_n, \tilde{x}_n) \in T_n$ and $(x'_n, x'), (x'_n, \tilde{x}'_n) \in T_n$. Choose N large enough so that it satisfies the property above and also that $4/N < \varepsilon/2$. Then for any $n \geq N$,

$$\begin{aligned} & |\omega_X(x, x') - \omega_X(\psi_X(x), \psi_X(x'))| \\ &= |\omega_X(x, x') - \omega_X(x_n, x'_n) + \omega_X(x_n, x'_n) - \omega_X(\tilde{x}_n, \tilde{x}'_n) + \omega_X(\tilde{x}_n, \tilde{x}'_n) - \omega_X(\psi_X(x), \psi_X(x'))| \\ &\leq \text{dis}(T_n) + \text{dis}(T_n) + \varepsilon/2 < 4/n + \varepsilon/2 \leq 4/N + \varepsilon/2 < \varepsilon. \end{aligned}$$

Case 3: $x \in \pi_X(Z), x' \in X \setminus \pi_X(Z)$ By the definition of ψ_X on $X \setminus \pi_X(Z)$, we obtain:

(1) a sequence $(\tilde{x}'_n)_{n \in \mathbb{N}}$ converging to $\psi_X(x')$, and (2) another sequence $(x'_n)_{n \in \mathbb{N}}$ such that (x'_n, x') and (x'_n, \tilde{x}'_n) both belong to T_n , for each $n \in \mathbb{N}$. By the definition of ψ_X on $\pi_X(Z)$, we obtain a sequence $(x_n)_{n \in \mathbb{N}}$ such that $(x_n, x) \in T_n$ for each $n \in \mathbb{N}$.

Let $U := \omega_X^{-1}[B(\omega_X(x, \psi_X(x')), \varepsilon/2)]$. Since $(\tilde{x}'_n)_{n \in \mathbb{N}}$ converges to $\psi_X(x')$, we know that all but finitely many terms of the sequence $(x, \tilde{x}'_n)_{n \in \mathbb{N}}$ belong to U . So

we choose N large enough so that for each $n \geq N$, we have:

$$\begin{aligned} & |\omega_X(x, x') - \omega_X(x, \psi_X(x'))| \\ &= |\omega_X(x, x') - \omega_X(x_n, x'_n) + \omega_X(x_n, x'_n) - \omega_X(x, \tilde{x}'_n) + \omega_X(x, \tilde{x}'_n) - \omega_X(x, \psi_X(x'))| \\ &\leq \text{dis}(T_n) + \text{dis}(T_n) + \varepsilon/2 < 4/n + \varepsilon/2 \leq 4/N + \varepsilon/2 < \varepsilon. \end{aligned}$$

Since $\varepsilon > 0$ was arbitrary, Equation 2.2 follows. The construction of ψ_Y and proof for Equation 2.3 are similar. This concludes the proof of the theorem. \square

As a consequence of Theorem 73, we see that weak isomorphisms of Types I and II coincide in the setting of \mathcal{CN} . Thus we recover a desirable notion of equivalence in the setting of compact networks.

2.4 Skeletons and motif reconstruction

In this section, we prove the results stated in §1.6.4. We also state and prove certain auxiliary results and a definition, cf. Propositions 132, 133, Definition 44, and Theorem 134.

Proposition 76. *Let $(X, \omega_X), (Y, \omega_Y)$ be networks with coherent topologies. Suppose $f : X \rightarrow Y$ is a weight-preserving map and $f(X)$ is a subnetwork of Y with the subspace topology. Then f is continuous.*

Proof. Let V' be an open subset of Y , and write $V := V' \cap f(X)$. Then V is open rel $f(X)$. We need to show that $U := f^{-1}(V') = f^{-1}(V)$ is open. Let $x \in U$, and suppose $(x_n)_n$ is a sequence in X converging to x . Then $f(x_n) \rightarrow f(x)$ rel $f(X)$. To see this, note that

$$\begin{aligned} \|\omega_Y(f(x_n), \bullet)|_{f(X)} - \omega_Y(f(x), \bullet)|_{f(X)}\| &= \|\omega_Y(f(x_n), f(\bullet))|_X - \omega_Y(f(x), f(\bullet))|_X\| \\ &= \|\omega_X(x_n, \bullet) - \omega_X(x, \bullet)\|, \end{aligned}$$

and the latter converges to 0 uniformly by Axiom A2 for X . Similarly, $\|\omega_Y(f(x_n)), f(\bullet)|_{f(X)} - \omega_Y(f(x), f(\bullet))|_{f(X)}\|$ converges to 0 uniformly. Thus by Axiom A2 for $f(X)$, we have $f(x_n) \rightarrow f(x)$ rel $f(X)$. But then there must exist $N \in \mathbb{N}$ such that $f(x_n) \in V$ for all $n \geq N$. Then $x_n \in U$ for all $n \geq N$. Thus U is open rel X by A1. This concludes the proof. \square

2.4.1 The skeleton of a compact network

We now prove that the skeleton of a compact network is terminal in the sense of Definition 32.

Proposition 80. *Suppose $(X, \omega_X) \in \mathcal{N}$ has a coherent topology. Then the map $\sigma : X \rightarrow X/\sim$ is an open map, i.e. it maps open sets to open sets.*

Proof of Proposition 80. Let $U \subseteq X$ be open. We need to show $\sigma^{-1}(\sigma(U))$ is open. For convenience, define $V := \sigma^{-1}(\sigma(U))$. Let $v \in V$. Then $\sigma(v) = [v] = [x]$ for some $x \in U$.

Let $(v_n)_{n \in \mathbb{N}}$ be any sequence in X such that $v_n \rightarrow v$ rel X . We first show that $v_n \rightarrow x$ rel X . We know $\omega_X(v_n, \bullet) \xrightarrow{\text{unif.}} \omega_X(v, \bullet)$ and $\omega_X(\bullet, v_n) \xrightarrow{\text{unif.}} \omega_X(\bullet, v)$ by Axiom A2. But $\omega_X(v, \bullet) = \omega_X(x, \bullet)$ and $\omega_X(\bullet, v) = \omega_X(\bullet, x)$, because $x \sim v$. By A2, we then have $v_n \rightarrow x$ rel X . But then there exists $N \in \mathbb{N}$ such that $v_n \in U \subseteq V$ for all $n \geq N$. This shows that any sequence (v_n) in X converging rel X to an arbitrary point $v \in V$ must eventually be in V . Thus V is open rel X , by Axiom A1. This concludes the proof. \square

The following lemma summarizes useful facts about weight preserving maps and the relation \sim .

Lemma 131. *Let $(X, \omega_X), (Y, \omega_Y) \in \mathcal{N}$, and let $f : X \rightarrow Y$ be a weight preserving surjection. Then,*

1. *f preserves equivalence classes of \sim , i.e. $x \sim x'$ for $x, x' \in X$ iff $f(x) \sim f(x')$.*
2. *f preserves weights between equivalence classes, i.e. $\omega_{X/\sim}([x], [x']) = \omega_{Y/\sim}([f(x)], [f(x')])$ for any $[x], [x'] \in X/\sim$.*

Proof of Lemma 131. For the first assertion, let $x \sim x'$ for some $x, x' \in X$. We wish to show $f(x) \sim f(x')$. Let $y \in Y$, and write $y = f(z)$ for some $z \in X$. Then,

$$\omega_Y(f(x), y) = \omega_Y(f(x), f(z)) = \omega_X(x, z) = \omega_X(x', z) = \omega_Y(f(x'), f(z)) = \omega_Y(f(x'), y).$$

Similarly we have $\omega_Y(y, f(x)) = \omega_Y(y, f(x'))$ for any $y \in Y$. Thus $f(x) \sim f(x')$.

Conversely suppose $f(x) \sim f(x')$. Let $z \in X$. Then,

$$\omega_X(x, z) = \omega_Y(f(x), f(z)) = \omega_Y(f(x'), f(z)) = \omega_X(x', z),$$

and similarly we get $\omega_X(z, x) = \omega_X(z, x')$. Thus $x \sim x'$. This proves the first assertion.

The second assertion holds by definition:

$$\omega_{Y/\sim}([f(x)], [f(x')]) = \omega_Y(f(x), f(x')) = \omega_X(x, x') = \omega_{X/\sim}([x], [x']). \quad \square$$

Proposition 81. *Let (X, ω_X) be a compact network with a coherent topology. The quotient topology on $(\text{sk}(X), \omega_{\text{sk}(X)})$ is also coherent.*

Proof of Proposition 81. Let Z be any subnetwork of $\text{sk}(X)$. Axiom A1 holds for any first countable space, and we have already shown that $\text{sk}(X)$ is first countable. Any subspace of a first countable space is first countable, so Z satisfies A1.

Next we verify Axiom A2. We begin with the “if” statement. Let $[x] \in Z$ and let $([x_n])_n$ be some sequence in Z . Suppose we have

$$\omega_{\text{sk}(X)}([x_n], [\bullet])|_Z \xrightarrow{\text{unif.}} \omega_{\text{sk}(X)}([x], [\bullet])|_Z, \quad \omega_{\text{sk}(X)}([\bullet], [x_n])|_Z \xrightarrow{\text{unif.}} \omega_{\text{sk}(X)}([\bullet], [x])|_Z.$$

Then we also have the following:

$$\omega_X(x_n, \bullet)|_{\sigma^{-1}(Z)} \xrightarrow{\text{unif.}} \omega_X(x, \bullet)|_{\sigma^{-1}(Z)}, \quad \omega_X(\bullet, x_n)|_{\sigma^{-1}(Z)} \xrightarrow{\text{unif.}} \omega_X(\bullet, x)|_{\sigma^{-1}(Z)}.$$

Since X is coherent and $\sigma^{-1}(Z)$ is a subnetwork, it follows by Axiom A2 that $x_n \rightarrow x$ rel $\sigma^{-1}(Z)$.

Let $V \subseteq Z$ be an open set rel Z containing $[x]$. We wish to show $[x_n] \rightarrow [x]$ rel Z , so it suffices to show that V contains all but finitely many of the $[x_n]$ terms. Since Z has the subspace topology, we know that $V = Z \cap V'$ for some open set $V' \subseteq \text{sk}(X) = \sigma(X)$. Write $U' := \sigma^{-1}(V')$. By continuity of σ , U' is open. Write $U := \sigma^{-1}(Z) \cap U'$. Then U is open rel $\sigma^{-1}(Z)$. Since $x_n \rightarrow x$ rel $\sigma^{-1}(Z)$, all but finitely many of the x_n terms belong to U . Thus all but finitely many of the $[x_n]$ terms belong to V . Thus $[x_n] \rightarrow [x]$ rel Z .

Now we show the “only if” statement. First we invoke the Axiom of Choice to pick a representative from each equivalence class of X/\sim . We denote this collection of representatives by Y and give it the subspace topology. Define $\tau := \sigma|_Y$. Then $\tau : Y \rightarrow \text{sk}(X)$ is a bijection given by $x \mapsto [x]$. By the discussion following Definition 28, we know that Y is coherent.

Let $([x_n])_n$ be a sequence in Z converging rel Z to some $[x] \in Z$. First we show $x_n \rightarrow x$ rel Y . Let $A \subseteq Y$ be an open set rel Y containing x . Then $\tau(A)$ is an open set rel $\tau(Y) = \text{sk}(X)$ containing $[x]$ (Proposition 80). In particular, $\tau(A) \cap Z$ is open rel Z . Thus $([x_n])_n$ is eventually inside $\tau(A) \cap Z$, in particular $\tau(A)$, by the definition of convergence rel Z . Because τ is a bijection, we have that $(x_n)_n = (\tau^{-1}([x_n])_n$ is eventually inside A . Thus any open set rel Y containing x also contains all but finitely many terms of $(x_n)_n$. It follows by the definition of convergence that $x_n \rightarrow x$ rel Y .

Since Y is coherent, it follows by Axiom A2 that we have $\omega_X(x_n, \bullet)|_Y \xrightarrow{\text{unif.}} \omega_X(x, \bullet)|_Y$ and $\omega_X(\bullet, x_n)|_Y \xrightarrow{\text{unif.}} \omega_X(\bullet, x)|_Y$. By the definition of \sim , we then have:

$$\omega_{\text{sk}(X)}([x_n], [\bullet]) = \omega_X(x_n, \bullet)|_Y \xrightarrow{\text{unif.}} \omega_X(x, \bullet)|_Y = \omega_{\text{sk}(X)}([x], [\bullet]).$$

Similarly we have $\omega_{\text{sk}(X)}([\bullet], [x_n]) \xrightarrow{\text{unif.}} \omega_{\text{sk}(X)}([\bullet], [x])$. This shows the “only if” statement.

This verifies Axiom A2 for Z . Since $Z \subseteq \text{sk}(X)$ was arbitrary, this concludes the proof. \square

Proposition 82. *Let (X, ω_X) be a compact network with a coherent topology. Then its skeleton $(\text{sk}(X), \omega_{\text{sk}(X)})$ is Hausdorff.*

Proof of Proposition 82. Let $[x] \neq [x'] \in \text{sk}(X)$. By first countability, we take a countable open neighborhood base $\{U_n : n \in \mathbb{N}\}$ of $[x]$ such that $U_1 \supseteq U_2 \supseteq U_3 \dots$ (if necessary, we replace U_n by $\cap_{i=1}^n U_i$). Similarly, we take a countable open neighborhood base $\{V_n : n \in \mathbb{N}\}$ of $[x']$ such that $V_1 \supseteq V_2 \supseteq V_3 \dots$. To show that $\text{sk}(X)$ is Hausdorff, it suffices to show that there exists $n \in \mathbb{N}$ such that $U_n \cap V_n = \emptyset$.

Towards a contradiction, suppose $U_n \cap V_n \neq \emptyset$ for each $n \in \mathbb{N}$. For each $n \in \mathbb{N}$, let $[y_n] \in U_n \cap V_n$. Any open set containing $[x]$ contains U_N for some $N \in \mathbb{N}$, and thus contains $[y_n]$ for all $n \geq N$. Thus $[y_n] \rightarrow [x]$ rel $\text{sk}(X)$. Similarly, $[y_n] \rightarrow [x']$ rel $\text{sk}(X)$. Because $\text{sk}(X)$ has a coherent topology (Proposition 81) and thus satisfies Axiom A2, we then have:

$$\begin{aligned}\omega_{\text{sk}(X)}([x'], [\bullet]) &= \text{unif lim}_n \omega_{\text{sk}(X)}([y_n], [\bullet]) = \omega_{\text{sk}(X)}([x], [\bullet]), \\ \omega_{\text{sk}(X)}([\bullet], [x']) &= \text{unif lim}_n \omega_{\text{sk}(X)}([\bullet], [y_n]) = \omega_{\text{sk}(X)}([\bullet], [x]).\end{aligned}$$

But then $x \sim x'$ and so $[x] = [x']$, a contradiction. \square

We are now ready to prove that skeletons are terminal, in the sense of Definition 32 (also recall Definitions 30 and 31).

Theorem 83 (Skeletons are terminal). *Let $(X, \omega_X) \in \mathcal{CN}$ be such that the topology on X is coherent. Then $(\text{sk}(X), \omega_{\text{sk}(X)}) \in \mathcal{CN}$ is terminal in $\mathfrak{p}(X)$.*

Proof of Theorem 83. Let $Y \in \mathfrak{p}(X)$. Let $f : X \rightarrow Y$ be a weight preserving surjection. We first prove that there exists a weight preserving surjection $g : Y \rightarrow \text{sk}(X)$.

Since f is surjective, for each $y \in Y$ we can write $y = f(x_y)$ for some $x_y \in X$. Then define $g : Y \rightarrow \text{sk}(X)$ by $g(y) := [x_y]$.

To see that g is surjective, let $[x] \in \text{sk}(X)$. Write $y = f(x)$. Then there exists $x_y \in X$ such that $f(x_y) = y$ and $g(y) = [x_y]$. Since f preserves equivalence classes (Lemma 131) and $f(x_y) = f(x)$, we have $x \sim x_y$. Thus $[x_y] = [x]$, and so $g(y) = [x]$.

To see that g preserves weights, let $y, y' \in Y$. Then,

$$\omega_Y(y, y') = \omega_Y(f(x_y), f(x_{y'})) = \omega_X(x_y, x_{y'}) = \omega_{\text{sk}(X)}([x_y], [x_{y'}]) = \omega_{\text{sk}(X)}(g(y), g(y')).$$

This proves that the skeleton satisfies the first condition for being terminal.

Next suppose $g : Y \rightarrow \text{sk}(X)$ and $h : Y \rightarrow \text{sk}(X)$ are two weight preserving surjections. We wish to show $h = \psi \circ g$ for some $\psi \in \text{Aut}(\text{sk}(X))$.

For each $[x] \in \text{sk}(X)$, we use the surjectivity of g to pick $y_x \in Y$ such that $g(y_x) = [x]$. Then we define $\psi : \text{sk}(X) \rightarrow \text{sk}(X)$ by $\psi([x]) = \psi(g(y_x)) := h(y_x)$.

To see that ψ is surjective, let $[x] \in \text{sk}(X)$. Since h is surjective, there exists $y'_x \in Y$ such that $h(y'_x) = [x]$. Write $[u] = g(y'_x)$. We have already chosen y_u such that $g(y_u) = [u]$. Since g preserves equivalence classes (Lemma 131), it follows that $y'_x \sim y_u$. Then,

$$\psi([u]) = \psi(g(y_u)) = h(y_u) = h(y'_x) = [x],$$

where the second-to-last equality holds because h preserves equivalence classes (Lemma 131).

To see that ψ is injective, let $[x], [x'] \in \text{sk}(X)$ be such that $\psi([x]) = h(y_x) = h(y_{x'}) = \psi([x'])$. Since h preserves equivalence classes (Lemma 131), we have $y_x \sim y_{x'}$. Next,

$g(y_x) = [x]$ and $g(y_{x'}) = [x']$ by the choices we made earlier. Since $y_x \sim y_{x'}$ and g preserves clusters, we have $g(y_x) \sim g(y_{x'})$. Thus $[x] = [x']$.

Next we wish to show that ψ preserves weights. Let $[x], [x'] \in \text{sk}(X)$. Then,

$$\begin{aligned}\omega_{\text{sk}(X)}(\psi([x]), \psi([x'])) &= \omega_{\text{sk}(X)}(h(y_x), h(y_{x'})) = \omega_Y(y_x, y_{x'}) = \omega_{\text{sk}(X)}(g(y_x), g(y_{x'})) \\ &= \omega_{\text{sk}(X)}([x], [x']).\end{aligned}$$

Thus ψ is a bijective, weight preserving automorphism of $\text{sk}(X)$. Finally we wish to show that $h = \psi \circ g$. Let $y \in Y$, and write $g(y) = [x]$ for some $x \in X$. Since g preserves equivalence classes (Lemma 131), we have $y \sim y_x$, where $g(y_x) = [x]$. Then,

$$\psi(g(y)) = \psi([x]) = \psi(g(y_x)) = h(y_x) = h(y),$$

where the last equality holds because h preserves equivalence classes (Lemma 131). Thus for each $y \in Y$, we have $h(y) = \psi(g(y))$. This shows that the skeleton satisfies the second condition for being terminal. We conclude the proof. \square

2.4.2 Reconstruction via motifs and skeletons

Our goal in this section is to prove that weak isomorphism, equality of motif sets, and strong isomorphism between skeleta are equivalent in the setting of compact networks with coherent topologies. However, we need to preface this theorem by proving some preparatory results.

Proposition 132. *Let $(X, \omega_X), (Y, \omega_Y)$ be compact networks such that $M_n(X) = M_n(Y)$ for all $n \in \mathbb{N}$. Suppose X contains a countable subset S_X . Then there exists a weight-preserving map $f : S_X \rightarrow Y$.*

Proof of Proposition 132. We proceed via a diagonal argument. Write $S_X = \{x_1, x_2, \dots, x_n, \dots\}$. For each $n \in \mathbb{N}$, let $f_n : S_X \rightarrow Y$ be a map that preserves weights on $\{x_1, \dots, x_n\}$. Such a map exists by the assumption that $M_n(X) = M_n(Y)$.

Since Y is first countable and compact, hence sequentially compact, the sequence $(f_n(x_1))_n$ has a convergent subsequence; we write this as $(f_{1,n}(x_1))_n$. Since f_k is weight-preserving on $\{x_1, x_2\}$ for $k \geq 2$, we know that $f_{1,n}$ is weight-preserving on $\{x_1, x_2\}$ for $n \geq 2$. Using sequential compactness again, we have that $(f_{1,n}(x_2))_n$ has a convergent subsequence $(f_{2,n}(x_2))_n$. This sequence converges at both x_1 and x_2 , and $f_{2,n}$ is weight-preserving on $\{x_1, x_2\}$ for $n \geq 2$. Proceeding in this way, we obtain the diagonal sequence $(f_{n,n})_n$ which converges pointwise on S_X . Furthermore, for any $n \in \mathbb{N}$, $f_{k,k}$ is weight-preserving on $\{x_1, \dots, x_n\}$ for $k \geq n$.

Next define $f : S_X \rightarrow Y$ by setting $f(x) := \lim_n f_{n,n}(x)$ for each $x \in S_X$. It remains to show that f is weight-preserving. Let $x_n, x_m \in S_X$, and let $k \geq \max(m, n)$. Then $\omega_X(x_n, x_m) = \omega_Y(f_{k,k}(x_n), f_{k,k}(x_m))$. Using (sequential) continuity of ω_Y , we then have:

$$\omega_Y(f(x_n), f(x_m)) = \omega_Y(\lim_k f_{k,k}(x_n), \lim_k f_{k,k}(x_m)) = \lim_k \omega_Y(f_{k,k}(x_n), f_{k,k}(x_m)) = \omega_X(x_n, x_m).$$

In the second equality above, we used the fact that a sequence converges in the product topology iff the components converge. Since $x_n, x_m \in S_X$ were arbitrary, this concludes the proof. \square

Proposition 133. *Let $(X, \omega_X), (Y, \omega_Y)$ be compact networks. Suppose $f : S_X \rightarrow Y$ is a weight-preserving function defined on a countable dense subset $S_X \subseteq X$. Then f extends to a weight-preserving map on X .*

Proof of Proposition 133. Let $x \in X \setminus S_X$. By first countability, we take a countable neighborhood base $\{U_n : n \in \mathbb{N}\}$ of x such that $U_1 \supseteq U_2 \supseteq U_3 \dots$ (if necessary, we replace U_n by $\cap_{i=1}^n U_i$). For each $n \in \mathbb{N}$, let $x_n \in U_n \cap S_X$. Then $x_n \rightarrow x$. To see this, let U be any open set containing x . Then $U_n \subseteq U$ for some $n \in \mathbb{N}$, and so $x_k \in U_n \subseteq U$ for all $k \geq n$.

Because Y is compact and first countable, hence sequentially compact, the sequence $(f(x_n))_n$ has a convergent subsequence; let y be its limit. Define $f(x) = y$. Extend f to all of X this way.

We need to verify that f is weight-preserving. Let $x, x' \in X$. Invoking the definition of f , let $(x_n)_n, (x'_n)_n$ be sequences in S_X converging to x, x' such that $f(x_n) \rightarrow f(x)$ and $f(x'_n) \rightarrow f(x')$. By sequential continuity and the standard result that a sequence converges in the product topology iff the components converge, we have

$$\lim_n \omega_Y(f(x_n), f(x'_n)) = \omega_Y(f(x), f(x')); \quad \lim_n \omega_X(x_n, x'_n) = \omega_X(x, x').$$

Let $\varepsilon > 0$. By the previous observation, fix $N \in \mathbb{N}$ such that for all $n \geq N$, we have $|\omega_Y(f(x_n), f(x'_n)) - \omega_Y(f(x), f(x'))| < \varepsilon$ and $|\omega_X(x_n, x'_n) - \omega_X(x, x')| < \varepsilon$. Then,

$$\begin{aligned} |\omega_X(x, x') - \omega_Y(f(x), f(x'))| &= |\omega_X(x, x') - \omega_X(x_n, x'_n) + \omega_X(x_n, x'_n) - \omega_Y(f(x), f(x'))| \\ &\leq |\omega_X(x, x') - \omega_X(x_n, x'_n)| + |\omega_X(x_n, x'_n) - \omega_Y(f(x), f(x'))| < 2\varepsilon. \end{aligned}$$

Thus $\omega_X(x, x') = \omega_Y(f(x), f(x'))$. Since $x, x' \in X$ were arbitrary, this concludes the proof. \square

The next result generalizes the result that an isometric embedding of a compact metric space into itself is automatically surjective [17, Theorem 1.6.14]. However, before presenting the theorem we first discuss an auxiliary construction that is used in its proof.

Definition 44 (The canonical pseudometric of a network). Let (X, ω_X) be any network. For any subset $A \subseteq X$, define $\Gamma_A : X \times X \rightarrow \mathbb{R}_+$ by

$$\Gamma_A(x, x') := \max \left(\sup_{a \in A} |\omega_X(x, a) - \omega_X(x', a)|, \sup_{a \in A} |\omega_X(a, x) - \omega_X(a, x')| \right).$$

Then Γ_A satisfies symmetry, triangle inequality, and $\Gamma_A(x, x) = 0$ for all $x \in X$. Thus Γ_A is a pseudometric on X . Moreover, Γ_A is a bona fide metric on $\text{sk}(A)$. The construction is “canonical” because it does not rely on any coupling between the topology of X and ω_X : even the continuity of ω_X is not necessary for this construction.

Next, for any $E \subseteq X$ and any $y \in X$, define $\Gamma_A(y, E) := \inf_{y' \in E} \Gamma_A(y, y')$. Then $\Gamma_A(\bullet, E)$ behaves as a proxy for the “distance to a set” function, where the set is fixed to be E .

Theorem 134. *Let (X, ω_X) be a compact network with a coherent, Hausdorff topology. Suppose $f : X \rightarrow X$ is a weight-preserving map. Then f is surjective.*

Proof of Theorem 134. Towards a contradiction, suppose $f(X) \neq X$. By Proposition 76, f is continuous. Define $X_0 := X$, and $X_n := f(X_{n-1})$ for each $n \in \mathbb{N}$. The continuous image of a compact space is compact, and compact subspaces of a Hausdorff space are closed. Thus we obtain a decreasing sequence of nonempty compact sets $X_0 \supseteq X_1 \supseteq X_2 \supseteq \dots$. Then $Z := \cap_{n \in \mathbb{N}} X_n$ is nonempty and compact, hence closed.

We now break up the proof up into several claims.

Claim 6. $f(Z) = Z$.

To see this, first note that $f(\cap_{n \in \mathbb{N}} X_n) \subseteq \cap_{n \in \mathbb{N}} f(X_n) \subseteq Z$. Next let $v \in Z$. For each $n \in \mathbb{N}$, let $u_n \in X_n$ be such that $f(u_n) = v$. Since singletons in a Hausdorff space are closed, we know that $\{v\}$ is closed. By continuity, it follows that $f^{-1}(\{v\})$ is closed.

By sequential compactness, the sequence $(u_n)_n$ has a convergent subsequence that converges to some limit u . Since each $u_n \in f^{-1}(\{v\})$ and a closed set contains its limit points, we then have $u \in f^{-1}(\{v\})$. Thus $f(u) = v$, and $v \in f(Z)$. Hence $Z = f(Z)$. This proves the claim.

Let $x \in X_0 \setminus X_1$. Define $x_0 := x$, and for each $n \in \mathbb{N}$, define $x_n := f(x_{n-1})$. Then $(x_n)_n$ is a sequence in the sequentially compact space X , and so it has a convergent subsequence $(x_{n_k})_k$. Let z be the limit of this subsequence.

Claim 7. $z \in Z$.

To see this, suppose towards a contradiction that $z \notin Z$. Then there exists $N \in \mathbb{N}$ such that $z \notin X_N$. Since X_N is closed, we have that $X \setminus X_N$ is open. By the definition of convergence, $X \setminus X_N$ contains all but finitely many terms of the sequence $(x_{n_k})_k$. But each x_{n_k} belongs to X_{n_k} , which is a subset of X_N for sufficiently large k . Thus infinitely many terms of the sequence $(x_{n_k})_k$ belong to X_N , a contradiction. Hence $z \in Z$.

Now we invoke the Γ_\bullet construction as in Definition 44.

Claim 8. For any $E \subseteq X$ and any $y \in E$,

$$\Gamma_E(y, Z) = \Gamma_{f(E)}(f(y), f(Z)).$$

To see this claim, fix $y \in E$. Let $v \in f(Z)$. Then $v = f(y')$ for some $y' \in Z$, and $\Gamma_{f(E)}(f(y), v) = \Gamma_E(y, y')$. To see the latter assertion, let $u \in f(E)$; then $u = f(y'')$ for some $y'' \in E$. Because f is weight-preserving, we then have:

$$|\omega_X(f(y), u) - \omega_X(v, u)| = |\omega_X(f(y), f(y'')) - \omega_X(f(y'), f(y''))| = |\omega_X(y, y'') - \omega_X(y', y'')|,$$

$$|\omega_X(u, f(y)) - \omega_X(u, v)| = |\omega_X(f(y''), f(y)) - \omega_X(f(y''), f(y'))| = |\omega_X(y'', y) - \omega_X(y'', y')|.$$

The preceding equalities show that for each $v \in f(Z)$, there exists $y' \in Z$ such that $\Gamma_{f(E)}(f(y), v) = \Gamma_E(y, y')$. Conversely, for any $y' \in Z$, we have $\Gamma_{f(E)}(f(y), f(y')) = \Gamma_E(y, y')$. It follows that $\Gamma_{f(E)}(f(y), f(Z)) = \Gamma_E(y, Z)$.

Claim 9. $\Gamma_X(x, Z) = 0$.

To see this, assume towards a contradiction that $\Gamma_X(x, Z) = \varepsilon > 0$ (Γ_X is positive by definition). Since $f(Z) = Z$, we have by the preceding claim that $\Gamma_X(x, Z) = \Gamma_{f(X)}(f(x), Z) = \dots = \Gamma_{f^n(X)}(f^n(x), Z)$ for each $n \in \mathbb{N}$. In particular, for any $k \in \mathbb{N}$,

$$\varepsilon = \Gamma_{f^{n_k}(X)}(f^{n_k}(x), Z) \leq \Gamma_{f^{n_k}(X)}(f^{n_k}(x), z) \leq \Gamma_X(f^{n_k}(x), z).$$

Here the first inequality follows because the left hand side includes an infimum over $z \in Z$, and the second inequality holds because the right hand side includes a supremum over a larger set.

Since $x_{n_k} \rightarrow z$ rel X , we have by Axiom A2 that

$$\|\omega_X(x_{n_k}, \bullet) - \omega_X(z, \bullet)\| \xrightarrow{\text{unif.}} 0, \quad \|\omega_X(\bullet, x_{n_k}) - \omega_X(\bullet, z)\| \xrightarrow{\text{unif.}} 0.$$

Thus for large enough k , we have:

$$\sup_{y \in X} |\omega_X(x_{n_k}, y) - \omega_X(z, y)| < \varepsilon, \quad \sup_{y \in X_{n_k}} |\omega_X(y, x_{n_k}) - \omega_X(y, z)| < \varepsilon.$$

Thus $\Gamma_X(f^{n_k}(x), z) < \varepsilon$, which is a contradiction. This proves the claim.

Recall that by assumption, $x \notin Z$. For each $n \in \mathbb{N}$, let $z_n \in Z$ be such that $\Gamma_X(x, z_n) < 1/n$. Then for each $x' \in X$, we have

$$\begin{aligned} \max(|\omega_X(x, x') - \omega_X(z_n, x')|, |\omega_X(x', x) - \omega_X(x', z_n)|) &< 1/n, \text{ i.e.} \\ \max(\|\omega_X(x, \bullet) - \omega_X(z_n, \bullet)\|, \|\omega_X(\bullet, x) - \omega_X(\bullet, z_n)\|) &< 1/n. \end{aligned}$$

Thus the sequence $(z_n)_n$ converges to x , by Axiom A2. Hence any open set containing x also contains infinitely many points of Z that are distinct from x . Thus x is a limit point of the closed set Z , and so $x \in Z$. This is a contradiction. \square

Theorem 84. Suppose $(X, \omega_X), (Y, \omega_Y)$ are separable, compact networks with coherent topologies. Then the following are equivalent:

1. $X \cong^w Y$.
2. $M_n(X) = M_n(Y)$ for all $n \in \mathbb{N}$.
3. $\text{sk}(X) \cong^s \text{sk}(Y)$.

Proof of Theorem 84. (2) follows from (1) by the stability of motif sets (Theorem 149). (1) follows from (3) by the triangle inequality of $d_{\mathcal{N}}$. We need to show that (2) implies (3).

First observe that $\text{sk}(X)$, being a continuous image of the separable space X , is separable, and likewise for $\text{sk}(Y)$. Let S_X, S_Y denote countable dense subsets of $\text{sk}(X)$ and $\text{sk}(Y)$. Next, because $d_{\mathcal{N}}(X, \text{sk}(X)) = 0$, an application of Theorem 149 shows that $M_n(X) = M_n(\text{sk}(X))$ for each $n \in \mathbb{N}$. The analogous result holds for $\text{sk}(Y)$. Thus $M_n(\text{sk}(X)) = M_n(\text{sk}(Y))$ for each $n \in \mathbb{N}$. Since X and Y have coherent topologies, so do $\text{sk}(X)$ and $\text{sk}(Y)$, by Proposition 81. By Propositions 132 and 133, there exist weight-preserving maps $\varphi : \text{sk}(X) \rightarrow \text{sk}(Y)$ and $\psi : \text{sk}(Y) \rightarrow \text{sk}(X)$. Define $X^{(1)} := \psi(\text{sk}(Y))$ and $Y^{(1)} := \varphi(\text{sk}(X))$. Also define φ_1 and ψ_1 to be the restrictions of φ and ψ to $X^{(1)}$ and $Y^{(1)}$, respectively. Finally define $X^{(2)} := \psi_1(Y^{(1)})$ and $Y^{(2)} := \varphi_1(X^{(1)})$. Then we have the following diagram.

$$\begin{array}{ccccc} \text{sk}(X) & \supseteq & X^{(1)} & \supseteq & X^{(2)} \\ \varphi \swarrow \quad \searrow \psi & & \varphi_1 \swarrow \quad \searrow \psi_1 & & \\ \text{sk}(Y) & \supseteq & Y^{(1)} & \supseteq & Y^{(2)} \end{array}$$

Now $\psi \circ \varphi$ is a weight-preserving map from $\text{sk}(X)$ into itself. Furthermore, it is continuous by Proposition 76. Since $\text{sk}(X)$ is Hausdorff (Proposition 82), an application of Theorem 134 now shows that $\psi \circ \varphi : \text{sk}(X) \rightarrow \text{sk}(X)$ is surjective. It follows from Definition 32 that $\psi \circ \varphi$ is an automorphism of $\text{sk}(X)$, hence a bijection. It follows that φ is injective. The dual argument for $\varphi \circ \psi$ shows that ψ is also injective.

Since $\psi \circ \varphi(\text{sk}(X)) = X^{(2)} = \text{sk}(X)$ and $X^{(2)} \subseteq X^{(1)} \subseteq \text{sk}(X)$, we must have $X^{(1)} = \text{sk}(X)$. Similarly, $Y^{(1)} = \text{sk}(Y)$. Thus $\varphi : \text{sk}(X) \rightarrow \text{sk}(Y)$ and $\psi : \text{sk}(Y) \rightarrow \text{sk}(X)$ are weight-preserving bijections. In particular, we have $\text{sk}(X) \cong^s \text{sk}(Y)$. This concludes the proof. \square

2.5 Completeness, compactness, and geodesics

2.5.1 The completion of \mathcal{CN}/\cong^w

A very natural question regarding \mathcal{CN}/\cong^w is if it is complete. This indeed turns out to be the case, and its proof is the content of the current section.

Lemma 135. *Let $X_1, \dots, X_n \in \mathcal{FN}$, and for each $i = 1, \dots, n-1$, let $R_i \in \mathcal{R}(X_i, X_{i+1})$. Define*

$$\begin{aligned} R &:= R_1 \circ R_2 \circ \dots \circ R_n \\ &= \{(x_1, x_n) \in X_1 \times X_n \mid \exists (x_i)_{i=2}^{n-1}, (x_i, x_{i+1}) \in R_i \text{ for all } i\}. \end{aligned}$$

Then $\text{dis}(R) \leq \sum_{i=1}^n \text{dis}(R_i)$.

Proof. We proceed by induction, beginning with the base case $n = 2$. For convenience, write $X := X_1$, $Y := X_2$, and $Z := X_3$. Let $(x, z), (x', z') \in R_1 \circ R_2$. Let $y \in Y$ be such that $(x, y) \in R_1$ and $(y, z) \in R_2$. Let $y' \in Y$ be such that $(x', y') \in R_1$, $(y', z') \in R_2$. Then we have:

$$\begin{aligned} |\omega_X(x, x') - \omega_Z(z, z')| &= |\omega_X(x, x') - \omega_Y(y, y') + \omega_Y(y, y') - \omega_Z(z, z')| \\ &\leq |\omega_X(x, x') - \omega_Y(y, y')| + |\omega_Y(y, y') - \omega_Z(z, z')| \\ &\leq \text{dis}(R) + \text{dis}(S). \end{aligned}$$

This holds for any $(x, z), (x', z') \in R \circ S$, and proves the claim.

Suppose that the result holds for $n = N \in \mathbb{N}$. Write $R' = R_1 \circ \dots \circ R_N$ and $R = R' \circ R_{N+1}$. Since R' is itself a correspondence, applying the base case yields:

$$\begin{aligned} \text{dis}(R) &\leq \text{dis}(R') + \text{dis}(R_{N+1}) \\ &\leq \sum_{i=1}^N \text{dis}(R_i) + \text{dis}(R_{N+1}) \quad \text{by induction} \\ &= \sum_{i=1}^{N+1} \text{dis}(R_i). \end{aligned}$$

This proves the lemma. \square

Theorem 88. *The completion of $(\mathcal{FN}/\cong^w, d_{\mathcal{N}})$ is $(\mathcal{CN}/\cong^w, d_{\mathcal{N}})$.*

Proof. Let $([X_i])_{i \in \mathbb{N}}$ be a Cauchy sequence in \mathcal{FN}/\cong^w . First we wish to show this sequence converges in \mathcal{CN}/\cong^w . Note that $(X_i)_{i \in \mathbb{N}}$ is a Cauchy sequence in \mathcal{FN} , since the distance between two equivalence classes is given by the distance between any representatives. To show $(X_i)_i$ converges, it suffices to show that a subsequence of $(X_i)_i$ converges, so without loss of generality, suppose $d_{\mathcal{N}}(X_i, X_{i+1}) < 2^{-i}$ for each i . Then for each i , there exists $R_i \in \mathcal{R}(X_i, X_{i+1})$ such that $\text{dis}(R_i) \leq 2^{-i+1}$. Fix such a sequence $(R_i)_{i \in \mathbb{N}}$. For $j > i$, define

$$R_{ij} := R_i \circ R_{i+1} \circ R_{i+2} \circ \dots \circ R_{j-1}.$$

By Lemma 135, $\text{dis}(R_{ij}) \leq \text{dis}(R_i) + \text{dis}(R_{i+1}) + \dots + \text{dis}(R_{j-1}) \leq 2^{-i+2}$. Next define:

$$\overline{X} := \{(x_j) : (x_j, x_{j+1}) \in R_j \text{ for all } j \in \mathbb{N}\}.$$

To see $\overline{X} \neq \emptyset$, let $x_1 \in X_1$, and use the (nonempty) correspondences to pick a sequence (x_1, x_2, x_3, \dots) . By construction, $(x_i) \in \overline{X}$.

Define $\omega_{\overline{X}}((x_j), (x'_j)) = \limsup_{j \rightarrow \infty} \omega_{X_j}(x_j, x'_j)$. We claim that $\omega_{\overline{X}}$ is bounded, and thus is a real-valued weight function. To see this, let $(x_j), (x'_j) \in \overline{X}$. Let $j \in \mathbb{N}$. Then we

have:

$$\begin{aligned}
|\omega_{X_j}(x_j, x'_j)| &= |\omega_{X_j}(x_j, x'_j) - \omega_{X_{j-1}}(x_{j-1}, x'_{j-1}) + \omega_{X_{j-1}}(x_{j-1}, x'_{j-1}) - \dots \\
&\quad - \omega_{X_1}(x_1, x'_1) + \omega_{X_1}(x_1, x'_1)| \\
&\leq |\omega_{X_1}(x_1, x'_1)| + \text{dis}(R_1) + \text{dis}(R_2) + \dots + \text{dis}(R_{j-1}) \\
&\leq |\omega_{X_1}(x_1, x'_1)| + 2
\end{aligned}$$

But j was arbitrary. Thus we obtain:

$$|\omega_{\bar{X}}((x_j), (x'_j))| = \limsup_{j \rightarrow \infty} \omega_{X_j}(x_j, x'_j) \leq |\omega_{X_1}(x_1, x'_1)| + 2 < \infty.$$

Claim 10. $(\bar{X}, \omega_{\bar{X}}) \in \mathcal{CN}$. More specifically, \bar{X} is a first countable compact topological space, and $\omega_{\bar{X}}$ is continuous with respect to the product topology on $\bar{X} \times \bar{X}$.

Proof of Claim 10. We equip $\prod_{i \in \mathbb{N}} X_i$ with the product topology. First note that the countable product $\prod_{i \in \mathbb{N}} X_i$ of first countable spaces is first countable. Any subspace of a first countable space is first countable, so $\bar{X} \subseteq \prod_{i \in \mathbb{N}} X_i$ is first countable. By Tychonoff's theorem, $\prod_{i \in \mathbb{N}} X_i$ is compact. So to show that \bar{X} is compact, we only need to show that it is closed.

If $\bar{X} = \prod_{i \in \mathbb{N}} X_i$, we would automatically know that \bar{X} is compact. Suppose not, and let $(x_i)_{i \in \mathbb{N}} \in (\prod_{i \in \mathbb{N}} X_i) \setminus \bar{X}$. Then there exists $N \in \mathbb{N}$ such that $(x_N, x_{N+1}) \notin R_N$. Define:

$$U := X_1 \times X_2 \times \dots \times \{x_N\} \times \{x_{N+1}\} \times X_{N+2} \times \dots$$

Since X_i has the discrete topology for each $i \in \mathbb{N}$, it follows that $\{x_N\}$ and $\{x_{N+1}\}$ are open. Hence U is an open neighborhood of $(x_i)_{i \in \mathbb{N}}$ and is disjoint from $\prod_{i \in \mathbb{N}} X_i$. It follows that $(\prod_{i \in \mathbb{N}} X_i) \setminus \bar{X}$ is open, hence \bar{X} is closed and thus compact.

It remains to show that $\omega_{\bar{X}}$ is continuous. We will show that preimages of open sets in \mathbb{R} under $\omega_{\bar{X}}$ are open. Let $(a, b) \subseteq \mathbb{R}$, and suppose $\omega_{\bar{X}}^{-1}[(a, b)]$ is nonempty (otherwise, there is nothing to show). Let $(x_i)_{i \in \mathbb{N}}, (x'_i)_{i \in \mathbb{N}} \in \bar{X} \times \bar{X}$ be such that

$$\alpha := \omega_{\bar{X}}((x_i)_i, (x'_i)_i) \in (a, b).$$

Write $r' := \min(|\alpha - a|, |b - \alpha|)$, and define $r := \frac{1}{2}r'$.

Let $N \in \mathbb{N}$ be such that $2^{-N+3} < r$. Consider the following open sets:

$$\begin{aligned}
U &:= \{x_1\} \times \{x_2\} \times \dots \times \{x_N\} \times X_{N+1} \times X_{N+2} \times \dots \subseteq \prod_{i \in \mathbb{N}} X_i, \\
V &:= \{x'_1\} \times \{x'_2\} \times \dots \times \{x'_N\} \times X_{N+1} \times X_{N+2} \times \dots \subseteq \prod_{i \in \mathbb{N}} X_i.
\end{aligned}$$

Next write $A := \bar{X} \cap U$ and $B := \bar{X} \cap V$. Then A and B are open with respect to the subspace topology on \bar{X} . Thus $A \times B$ is open in $\bar{X} \times \bar{X}$. Note that $(x_i)_{i \in \mathbb{N}} \in A$

and $(x'_i)_{i \in \mathbb{N}} \in B$. We wish to show that $A \times B \subseteq \omega_{\overline{X}}^{-1}[(a, b)]$, so it suffices to show that $\omega_{\overline{X}}(A, B) \subseteq (a, b)$.

Let $(z_i)_{i \in \mathbb{N}} \in A$ and $(z'_i)_{i \in \mathbb{N}} \in B$. Notice that $z_i = x_i$ and $z'_i = x'_i$ for each $i \leq N$. So for $n \leq N$, we have $|\omega_{X_n}(z_n, z'_n) - \omega_{X_n}(x_n, x'_n)| = 0$.

Next let $n \in \mathbb{N}$, and note that:

$$\begin{aligned} & |\omega_{X_{N+n}}(z_{N+n}, z'_{N+n}) - \omega_{X_{N+n}}(x_{N+n}, x'_{N+n})| \\ &= |\omega_{X_{N+n}}(z_{N+n}, z'_{N+n}) - \omega_{X_N}(z_N, z'_N) + \omega_{X_N}(z_N, z'_N) - \omega_{X_{N+n}}(x_{N+n}, x'_{N+n})| \\ &= |\omega_{X_{N+n}}(z_{N+n}, z'_{N+n}) - \omega_{X_N}(z_N, z'_N) + \omega_{X_N}(x_N, x'_N) - \omega_{X_{N+n}}(x_{N+n}, x'_{N+n})| \\ &\leq \text{dis}(R_{N,N+n}) + \text{dis}(R_{N,N+n}) \leq 2^{-N+2} + 2^{-N+2} = 2^{-N+3} < r. \end{aligned}$$

Here the second to last inequality follows from Lemma 135. The preceding calculation holds for arbitrary $n \in \mathbb{N}$. It follows that:

$$\limsup_{i \rightarrow \infty} \omega_{X_i}(x_i, x'_i) - \limsup_{i \rightarrow \infty} \omega_{X_i}(z_i, z'_i) \leq \limsup_{i \rightarrow \infty} (\omega_{X_i}(x_i, x'_i) - \omega_{X_i}(z_i, z'_i)) < r,$$

and similarly $\limsup_{i \rightarrow \infty} \omega_{X_i}(z_i, z'_i) - \limsup_{i \rightarrow \infty} \omega_{X_i}(x_i, x'_i) < r$. Thus we have $\omega_{\overline{X}}((z_i)_i, (z'_i)_i) \in (a, b)$. This proves continuity of $\omega_{\overline{X}}$. \square

Next we claim that $X_i \xrightarrow{d_{\mathcal{N}}} \overline{X}$ as $i \rightarrow \infty$. Fix $i \in \mathbb{N}$. We wish to construct a correspondence $S \in \mathcal{R}(X_i, \overline{X})$. Let $y \in X_i$. We write $x_i = y$ and pick $x_1, x_2, \dots, x_{i-1}, x_{i+1}, \dots$ such that $(x_j, x_{j+1}) \in R_j$ for each $j \in \mathbb{N}$. We denote this sequence by $(x_j)^{x_i=y}$, and note that by construction, it lies in \overline{X} . Conversely, for any $(x_j) \in \overline{X}$, we simply pick its i th coordinate x_i as a corresponding element in X_i . We define:

$$\begin{aligned} S &:= A \cup B, \text{ where} \\ A &:= \{(y, (x_j)^{x_i=y}) : y \in X_i\} \\ B &:= \{(x_i, (x_k)) : (x_k) \in \overline{X}\} \end{aligned}$$

Then $S \in \mathcal{R}(X_i, \overline{X})$. We claim that $\text{dis}(S) \leq 2^{-i+2}$. Let $z = (y, (x_k)), z' = (y', (x'_k)) \in B$. Let $n \in \mathbb{N}, n \geq i$. Then we have:

$$\begin{aligned} |\omega_{X_i}(y, y') - \omega_{X_n}(x_n, x'_n)| &= |\omega_{X_i}(y, y') - \omega_{X_{i+1}}(x_{i+1}, x'_{i+1}) + \omega_{X_{i+1}}(x_{i+1}, x'_{i+1}) - \dots \\ &\quad + \omega_{X_{n-1}}(x_{n-1}, x'_{n-1}) + \omega_{X_n}(x_n, x'_n)| \\ &\leq \text{dis}(R_i) + \text{dis}(R_{i+1}) + \dots + \text{dis}(R_{n-1}) \\ &\leq 2^{-i+1} + 2^{-i} + \dots + 2^{-n+2} \\ &\leq 2^{-i+2}. \end{aligned}$$

This holds for arbitrary $n \geq i$. It follows that we have:

$$|\omega_{X_i}(y, y') - \omega_{\overline{X}}((x_k), (x'_k))| \leq 2^{-i+2}.$$

Similar inequalities hold for $z, z' \in A$, and for $z \in A, z' \in B$. Thus $\text{dis}(S) \leq 2^{-i+2}$. It follows that $d_{\mathcal{N}}(X_i, \overline{X}) \leq 2^{-i+1}$. Thus the sequence $([X_i])_i$ converges to $[\overline{X}] \in \mathcal{CN}/\cong^w$.

Finally, we need to check that $(\mathcal{CN}/\cong^w, d_{\mathcal{N}})$ is complete. Let $([Y_n])_n$ be a Cauchy sequence in \mathcal{CN}/\cong^w . For each n , let $[X_n] \in \mathcal{FN}/\cong^w$ be such that $d_{\mathcal{N}}([X_n], [Y_n]) < \frac{1}{n}$. Let $\varepsilon > 0$. Then for sufficiently large m and n , we have:

$$d_{\mathcal{N}}([X_n], [X_m]) \leq d_{\mathcal{N}}([X_n], [Y_n]) + d_{\mathcal{N}}([Y_n], [Y_m]) + d_{\mathcal{N}}([Y_m], [X_m]) < \varepsilon.$$

Thus $([X_n])_n$ is a Cauchy sequence in \mathcal{FN}/\cong^w . By applying what we have shown above, this sequence converges to some $[\overline{X}] \in \mathcal{CN}/\cong^w$. By applying the triangle inequality, we see that the sequence $([Y_n])_n$ also converges to $[\overline{X}]$. This shows completeness, and concludes the proof. \square

The result of Theorem 88 can be summarized as follows:

The limit of a convergent sequence of finite networks is a compact topological space with a continuous weight function.

Remark 136. The technique of composed correspondences used in the preceding proof can also be used to show that the collection of isometry classes of compact metric spaces endowed with the Gromov-Hausdorff distance is a complete metric space. Standard proofs of this fact [98, §10] do not use correspondences, relying instead on a method of endowing metrics on disjoint unions of spaces and then computing Hausdorff distances.

Remark 137. In the proof of Theorem 88, note that the construction of the limit is dependent upon the initial choice of optimal correspondences. However, all such limits obtained from different choices of optimal correspondences belong to the same weak isomorphism class.

2.5.2 Precompact families in \mathcal{CN}/\cong^w

We now prove Theorem 90. Our proof is modeled on the proof of an analogous result for compact metric spaces proposed by Gromov [65]. We use one fact proved in a different section (Proposition 145): for compact networks X, Y such that $d_{\mathcal{N}}(X, Y) < \varepsilon$, we have $\text{diam}(X) \leq \text{diam}(Y) + 2\varepsilon$.

Proof of Theorem 90. Let $D \geq 0$ be such that $\text{diam}(X) \leq D$ for each $[X] \in \mathfrak{F}$. It suffices to prove that \mathfrak{F} is totally bounded, because Theorem 88 gives completeness, and these two properties together imply precompactness. Let $\varepsilon > 0$. We need to find a finite family $\mathcal{G} \subseteq \mathcal{CN}/\cong^w$ such that for every $[F] \in \mathfrak{F}$, there exists $[G] \in \mathcal{G}$ with $d_{\mathcal{N}}(F, G) < \varepsilon$. Define:

$$\mathcal{A} := \{A \in \mathcal{FN} : \text{card}(A) \leq N(\varepsilon/2), d_{\mathcal{N}}(A, F) < \varepsilon/2 \text{ for some } [F] \in \mathfrak{F}\}.$$

Each element of \mathcal{A} is an $n \times n$ matrix, where $1 \leq n \leq N(\varepsilon/2)$. For each $A \in \mathcal{A}$, there exists $[F] \in \mathfrak{F}$ with $d_{\mathcal{N}}(A, F) < \varepsilon/2$, and by Proposition 145, we have $\text{diam}(A) \leq$

$\text{diam}(F) + 2(\varepsilon/2) \leq D + \varepsilon$. Thus the matrices in \mathcal{A} have entries in $[-D - \varepsilon, D + \varepsilon]$. Let $N \gg 1$ be such that:

$$\frac{2D + 2\varepsilon}{N} < \frac{\varepsilon}{4},$$

and write the refinement of $[-D - \varepsilon, D + \varepsilon]$ into N pieces as:

$$W := \left\{ -D - \varepsilon + k \left(\frac{2D + 2\varepsilon}{N} \right) : 0 \leq k \leq N \right\}.$$

Write $\mathcal{A} = \bigsqcup_{i=1}^{N(\varepsilon/2)} \mathcal{A}_i$, where each \mathcal{A}_i consists of the $i \times i$ matrices of \mathcal{A} . For each i define:

$$\mathcal{G}_i := \{(G_{pq})_{1 \leq p, q \leq i} : G_{pq} \in W\}, \text{ the } i \times i \text{ matrices with entries in } W.$$

Let $\mathcal{G} = \bigsqcup_{i=1}^{N(\varepsilon/2)} \mathcal{G}_i$ and note that this is a finite collection. Furthermore, for each $A_i \in \mathcal{A}_i$, there exists $G_i \in \mathcal{G}_i$ such that

$$\|A_i - G_i\|_\infty < \frac{\varepsilon}{4}.$$

Taking the diagonal correspondence between A_i and G_i , it follows that $d_N(A_i, G_i) < \varepsilon/2$. Hence for any $[F] \in \mathfrak{F}$, there exists $A \in \mathcal{A}$ and $G \in \mathcal{G}$ such that

$$d_N(F, G) \leq d_N(F, A) + d_N(A, G) < \varepsilon/2 + \varepsilon/2 = \varepsilon.$$

This shows that \mathfrak{F} is totally bounded, and concludes the proof. \square

2.5.3 Geodesics in \mathcal{CN}/\cong^w

We now prove our results about the geodesic structures of \mathcal{FN}/\cong^w and \mathcal{CN}/\cong^w .

Proof of Theorem 92. Let $[X], [Y] \in \mathcal{FN}/\cong^w$. We will show the existence of a curve $\gamma : [0, 1] \rightarrow \mathcal{FN}$ such that $\gamma(0) = (X, \omega_X)$, $\gamma(1) = (Y, \omega_Y)$, and for all $s, t \in [0, 1]$,

$$d_N(\gamma(s), \gamma(t)) = |t - s| \cdot d_N(X, Y).$$

Note that this yields $d_N([\gamma(s)], [\gamma(t)]) = |t - s| \cdot d_N([X], [Y])$ for all $s, t \in [0, 1]$, which is what we need to show.

Let $R \in \mathcal{R}^{\text{opt}}(X, Y)$, i.e. let R be a correspondence such that $\text{dis}(R) = 2d_N(X, Y)$. For each $t \in (0, 1)$ define $\gamma(t) := (R, \omega_{\gamma(t)})$, where

$$\omega_{\gamma(t)}((x, y), (x', y')) := (1 - t) \cdot \omega_X(x, x') + t \cdot \omega_Y(y, y') \quad \text{for all } (x, y), (x', y') \in R.$$

Also define $\gamma(0) = (X, \omega_X)$ and $\gamma(1) = (Y, \omega_Y)$.

Claim 11. For any $s, t \in [0, 1]$,

$$d_N(\gamma(s), \gamma(t)) \leq |t - s| \cdot d_N(X, Y).$$

Suppose for now that Claim 11 holds. We further claim that this implies, for all $s, t \in [0, 1]$,

$$d_{\mathcal{N}}(\gamma(s), \gamma(t)) = |t - s| \cdot d_{\mathcal{N}}(X, Y).$$

To see this, assume towards a contradiction that there exist $s_0 < t_0$ such that :

$$d_{\mathcal{N}}(\gamma(s_0), \gamma(t_0)) < |t_0 - s_0| \cdot d_{\mathcal{N}}(X, Y).$$

$$\begin{aligned} \text{Then } d_{\mathcal{N}}(X, Y) &\leq d_{\mathcal{N}}(X, \gamma(s_0)) + d_{\mathcal{N}}(\gamma(s_0), \gamma(t_0)) + d_{\mathcal{N}}(\gamma(t_0), Y) \\ &< |s_0 - 0| \cdot d_{\mathcal{N}}(X, Y) + |t_0 - s_0| \cdot d_{\mathcal{N}}(X, Y) + |1 - t_0| \cdot d_{\mathcal{N}}(X, Y) \\ &= d_{\mathcal{N}}(X, Y), \text{ a contradiction.} \end{aligned}$$

Thus it suffices to show Claim 11. There are three cases: (i) $s, t \in (0, 1)$, (ii) $s = 0, t \in (0, 1)$, and (iii) $s \in (0, 1), t = 1$. The latter two cases are similar, so we just prove (i) and (ii). For (i), fix $s, t \in (0, 1)$. Notice that $\Delta := \text{diag}(R \times R) := \{(r, r) : r \in R\}$ is a correspondence in $\mathcal{R}(R, R)$. Then we obtain:

$$\begin{aligned} \text{dis}(\Delta) &= \max_{(a,a),(b,b) \in \Delta} |\omega_{\gamma(t)}(a, b) - \omega_{\gamma(s)}(a, b)| \\ &= \max_{(x,y),(x',y') \in R} |\omega_{\gamma(t)}((x, y), (x', y')) - \omega_{\gamma(s)}((x, y), (x', y'))| \\ &= \max_{(x,y),(x',y') \in R} |(1-t)\omega_X(x, x') + t \cdot \omega_Y(y, y') - (1-s)\omega_X(x, x') - s \cdot \omega_Y(y, y')| \\ &= \max_{(x,y),(x',y') \in R} |(s-t)\omega_X(x, x') - (s-t)\omega_Y(y, y')| \\ &= |t - s| \cdot \max_{(x,y),(x',y') \in R} |\omega_X(x, x') - \omega_Y(y, y')| \\ &\leq 2|t - s| \cdot d_{\mathcal{N}}(X, Y). \end{aligned}$$

Finally $d_{\mathcal{N}}(\gamma(t), \gamma(s)) \leq \frac{1}{2} \text{dis}(\Delta) \leq |t - s| \cdot d_{\mathcal{N}}(X, Y)$.

For (ii), fix $s = 0, t \in (0, 1)$. Define $R_X = \{(x, (x, y)) : (x, y) \in R\}$. Then R_X is a correspondence in $\mathcal{R}(X, R)$.

$$\begin{aligned} \text{dis}(R_X) &= \max_{(x,(x,y)),(x',(x',y')) \in R_X} |\omega_X(x, x') - (1-t) \cdot \omega_X(x, x') - t \cdot \omega_Y(y, y')| \\ &= \max_{(x,(x,y)),(x',(x',y')) \in R_X} t \cdot |\omega_X(x, x') - \omega_Y(y, y')| \\ &= t \text{dis}(R) = 2t \cdot d_{\mathcal{N}}(X, Y). \end{aligned}$$

Thus $d_{\mathcal{N}}(X, \gamma(t)) \leq t \cdot d_{\mathcal{N}}(X, Y)$. The proof for case (iii), i.e. that $d_{\mathcal{N}}(\gamma(s), Y) \leq |1 - s| \cdot d_{\mathcal{N}}(X, Y)$, is similar. This proves Claim 11, and the result follows. \square

Proof of Theorem 93. Let $[X], [Y] \in \mathcal{CN}/\cong^w$. It suffices to find a geodesic between X and Y , because the distance between any two equivalence classes is given by the distance between any two representatives, and hence we will obtain a geodesic between $[X]$ and $[Y]$.

Let $(X_n)_n, (Y_n)_n$ be sequences in \mathcal{FN} such that $d_{\mathcal{N}}(X_n, X) < \frac{1}{n}$ and $d_{\mathcal{N}}(Y_n, Y) < \frac{1}{n}$ for each n . For each n , let R_n be an optimal correspondence between X_n and Y_n , endowed with the weight function

$$\omega_n((x, y), (a, b)) = \frac{1}{2}\omega_{X_n}(x, a) + \frac{1}{2}\omega_{X_n}(y, b).$$

By the proof of Theorem 92, the network (R_n, ω_n) is a midpoint of X_n and Y_n .

Claim 12. The collection $\{R_n : n \in \mathbb{N}\}$ is precompact.

Assume for now that Claim 12 is true. Then we can pick a sequence (R_n) that converges to some $R \in \mathcal{CN}$. Then we obtain:

$$\begin{aligned} d_{\mathcal{N}}(X, R) &\leq d_{\mathcal{N}}(X, X_n) + d_{\mathcal{N}}(X_n, R_n) + d_{\mathcal{N}}(R_n, R) \\ &= d_{\mathcal{N}}(X, X_n) + \frac{1}{2}d_{\mathcal{N}}(X_n, Y_n) + d_{\mathcal{N}}(R_n, R) \rightarrow \frac{1}{2}d_{\mathcal{N}}(X, Y). \end{aligned}$$

Similarly $d_{\mathcal{N}}(R, Y) \leq \frac{1}{2}d_{\mathcal{N}}(X, Y)$. Furthermore, equality holds in both inequalities, because we would get a contradiction otherwise. Thus R is a midpoint of X and Y , and moreover, $[R]$ is a midpoint of $[X]$ and $[Y]$. The result now follows by an application of Theorem 91.

It remains to prove Claim 12. By Theorem 90, it suffices to show that $\{R_n\}$ is uniformly approximable.

Since $d_{\mathcal{N}}(X_n, X) \rightarrow 0$ and $d_{\mathcal{N}}(Y_n, Y) \rightarrow 0$, we can choose $D > 0$ large enough so that $\text{diam}(X_n) \leq \frac{D}{2}$ and $\text{diam}(Y_n) \leq \frac{D}{2}$ for all n . Then $\text{diam}(R_n) \leq D$ for all n .

Let $\varepsilon > 0$. Fix N large enough so that $\frac{1}{N} < \frac{\varepsilon}{2}$, and write $N(\varepsilon) = \max_{n \leq N} \text{card}(R_n)$. We wish to show that every R_n is ε -approximable by a finite network with cardinality up to $N(\varepsilon)$. For any $n \leq N$, we know R_n approximates itself, and $\text{card}(R_n) \leq N(\varepsilon)$. Next let $n > N$. It will suffice to show that R_n is ε -approximable by R_N .

Let S, T be optimal correspondences between X_n, X_N and Y_n, Y_N respectively. Note that $d_{\mathcal{N}}(X_N, X_n) \leq d_{\mathcal{N}}(X_N, X) + d_{\mathcal{N}}(X, X_n) \leq \frac{1}{N} + \frac{1}{N} = \frac{2}{N}$, and similarly $d_{\mathcal{N}}(Y_N, Y_n) \leq \frac{2}{N}$. Thus $\text{dis}(S) \leq \frac{4}{N}$ and $\text{dis}(T) \leq \frac{4}{N}$. Next write

$$Q = \{(x, y, x', y') \in R_N \times R_n : (x, x') \in S, (y, y') \in T\}.$$

Observe that since S and T are correspondences, Q is a correspondence between R_N and R_n . Next we calculate $\text{dis}(Q)$:

$$\begin{aligned} \text{dis}(Q) &= \max_{\substack{(x, y, x', y'), \\ (a, b, a', b') \in Q}} |\omega_N((x, y), (a, b)) - \omega_n((x', y'), (a', b'))| \\ &= \max_{\substack{(x, y, x', y'), \\ (a, b, a', b') \in Q}} |\frac{1}{2}\omega_{X_N}(x, a) + \frac{1}{2}\omega_{Y_N}(y, b) - \frac{1}{2}\omega_{X_n}(x', a') - \frac{1}{2}\omega_{Y_n}(y', b')| \\ &\leq \frac{1}{2} \max_{(x, x'), (a, a') \in S} |\omega_{X_N}(x, a) - \omega_{X_n}(x', a')| + \frac{1}{2} \max_{(y, y'), (b, b') \in T} |\omega_{Y_N}(y, b) - \omega_{Y_n}(y', b')| \\ &= \frac{1}{2} \text{dis}(S) + \frac{1}{2} \text{dis}(T) \leq \frac{4}{N}. \end{aligned}$$

Thus $d_{\mathcal{N}}(R_N, R_n) \leq \frac{2}{N} < \varepsilon$. This shows that any R_n can be ε -approximated by a network having up to $N(\varepsilon)$ points. Thus $\{R_n\}$ is uniformly approximable, hence precompact. Thus Claim 12 and the result follow. \square

Remark 138 (Branching and deviant geodesics). It is important to note that there exist geodesics in \mathcal{CN}/\cong^w that deviate from the straight-line form given by Theorem 92. Even in the setting of compact metric spaces, there exist infinite families of branching and deviant geodesics [36].

2.6 Lower bounds on $d_{\mathcal{N}}$

At this point, we have computed $d_{\mathcal{N}}$ between several examples of networks, as in Example 9 and Remark 10. We also asserted in Remark 13 that $d_{\mathcal{N}}$ is in general difficult to compute. The solution we propose is to compute *quantitatively stable invariants* of networks, and compare the invariants instead of comparing the networks directly. In this section, we restrict our attention to computing invariants of compact networks, which satisfy the useful property that the images of the weight functions are compact.

Intuitively, the invariants that we associate to two strongly isomorphic networks should be the same. We define an \mathbb{R} -invariant of networks to be a map $\iota : \mathcal{CN} \rightarrow \mathbb{R}$ such that for any $X, Y \in \mathcal{CN}$, if $X \cong^s Y$ then $\iota(X) = \iota(Y)$. Any \mathbb{R} -invariant is an example of a *pseudometric* (and in particular, a *metric*) *space valued invariant*, which we define next. Recall that a pseudometric space (V, d_V) is a metric space where we allow $d_V(v, v') = 0$ even if $v \neq v'$.

Definition 45. Let (V, d_V) be any metric or pseudometric space. A *V-valued invariant* is any map $\iota : \mathcal{CN} \rightarrow V$ such that $\iota(X, \omega_X) = \iota(Y, \omega_Y)$ whenever $X \cong^s Y$.

Recall that $\text{pow}(\mathbb{R})$, the nonempty elements of the power set of \mathbb{R} , is a pseudometric space when endowed with the Hausdorff distance [17, Proposition 7.3.3].

In what follows, we will construct several maps and claim that they are pseudometric space valued invariants; this claim will be substantiated in Proposition 144. We will eventually prove that our proposed invariants are *quantitatively stable*. This notion is made precise in §2.6.1.

Example 139. Define the *diameter* map to be the map

$$\text{diam} : \mathcal{CN} \rightarrow \mathbb{R} \text{ given by } (X, \omega_X) \mapsto \max_{x, x' \in X} |\omega_X(x, x')|.$$

Then diam is an \mathbb{R} -invariant. Observe that the maximum is achieved for $(X, \omega_X) \in \mathcal{CN}$ because X (hence $X \times X$) is compact and $\omega_X : X \times X \rightarrow \mathbb{R}$ is continuous.

Example 140. Define the *spectrum* map

$$\text{spec} : \mathcal{CN} \rightarrow \text{pow}(\mathbb{R}) \text{ by } (X, \omega_X) \mapsto \{\omega_X(x, x') : x, x' \in X\}.$$

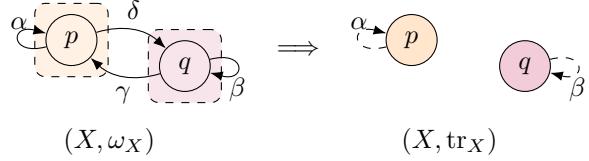


Figure 2.1: The trace map erases data between pairs of nodes.

The spectrum also has two local variants. Define the out-local spectrum of X by $x \mapsto \text{spec}_X^{\text{out}}(x) := \{\omega_X(x, x'), x' \in X\}$. Notice that $\text{spec}(X) = \bigcup_{x \in X} \text{spec}_X^{\text{out}}(x)$ for any network X , thus justifying the claim that this construction localizes spec. Similarly, we define the in-spectrum of X as the map $x \mapsto \text{spec}_X^{\text{in}}(x) := \{\omega_X(x', x) : x' \in X\}$. Notice that one still has $\text{spec}(X) = \bigcup_{x \in X} \text{spec}_X^{\text{in}}(x)$ for any network X . Finally, we observe that the two local versions of spec do not necessarily coincide in an asymmetric network.

The spectrum is closely related to the *multisets* used by Boutin and Kemper [15] to produce invariants of weighted undirected graphs. For an undirected graph G , they considered the collection of all subgraphs with three nodes, along with the edge weights for each subgraph (compare to our notion of spectrum). Then they proved that the distribution of edge weights of these subgraphs is an invariant when G belongs to a certain class of graphs.

Example 141. Define the *trace* map $\text{tr} : \mathcal{CN} \rightarrow \text{pow}(\mathbb{R})$ by $(X, \omega_X) \mapsto \text{tr}(X) := \{\omega_X(x, x) : x \in X\}$. This also defines an associated map $x \mapsto \text{tr}_X(x) := \omega_X(x, x)$. An example is provided in Figure 2.1: in this case, we have $(X, \text{tr}_X) = (\{p, q\}, (\alpha, \beta))$.

Example 142 (The out and in maps). Let $(X, \omega_X) \in \mathcal{CN}$, and let $x \in X$. Now define $\text{out} : \mathcal{CN} \rightarrow \text{pow}(\mathbb{R})$ and $\text{in} : \mathcal{CN} \rightarrow \text{pow}(\mathbb{R})$ by

$$\begin{aligned}\text{out}(X) &= \left\{ \max_{x' \in X} |\omega_X(x, x')| : x \in X \right\} && \text{for all } (X, \omega_X) \in \mathcal{CN} \\ \text{in}(X) &= \left\{ \max_{x' \in X} |\omega_X(x', x)| : x \in X \right\} && \text{for all } (X, \omega_X) \in \mathcal{CN}.\end{aligned}$$

For each $x \in X$, $\max_{x' \in X} |\omega_X(x, x')|$ and $\max_{x' \in X} |\omega_X(x', x)|$ are achieved because $\{x\} \times X$ and $X \times \{x\}$ are compact. We also define the associated maps out_X and in_X by writing, for any $(X, \omega_X) \in \mathcal{CN}$ and any $x \in X$,

$$\text{out}_X(x) = \max_{x' \in X} |\omega_X(x, x')| \quad \text{in}_X(x) = \max_{x' \in X} |\omega_X(x', x)|.$$

To see how these maps operate on a network, let $X = \{p, q, r\}$ and consider the weight matrix $\Sigma = \begin{pmatrix} 1 & 2 & 3 \\ 0 & 0 & 4 \\ 0 & 0 & 5 \end{pmatrix}$. The network corresponding to this matrix is shown in Figure 2.2. We

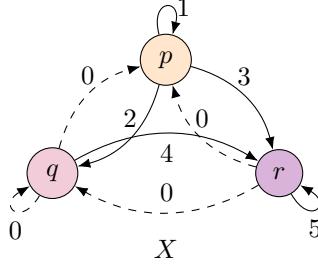


Figure 2.2: The out map applied to each node yields the greatest weight of an arrow *leaving* the node, and the in map returns the greatest weight *entering* the node.

ascertain the following directly from the matrix:

$$\begin{aligned} \text{out}_X(p) &= 3 & \text{in}_X(p) &= 1 \\ \text{out}_X(q) &= 4 & \text{in}_X(q) &= 2 \\ \text{out}_X(r) &= 5 & \text{in}_X(r) &= 5. \end{aligned}$$

So the out map returns the maximum (absolute) value in each row, and the in map pulls out the maximum (absolute) value in each column of the weight matrix. As in the preceding example, we may use the Hausdorff distance to compare the images of networks under the out and in maps.

Constructions similar to out and in have been used by Jon Kleinberg to study the problem of searching the World Wide Web for user-specified queries [77]. In Kleinberg's model, for a search query σ , *hubs* are pages that point to highly σ -relevant pages (compare to out_{\bullet}), and *authorities* are pages that are pointed to by pages that have a high σ -relevance (compare to in_{\bullet}). Good hubs determine good authorities, and good authorities turn out to be good search results.

Example 143 (m^{out} and m^{in}). Define the maps $m^{\text{out}} : \mathcal{CN} \rightarrow \mathbb{R}$ and $m^{\text{in}} : \mathcal{CN} \rightarrow \mathbb{R}$ by

$$\begin{aligned} m^{\text{out}}((X, \omega_X)) &= \min_{x \in X} \text{out}_X(x) && \text{for all } (X, \omega_X) \in \mathcal{CN} \\ m^{\text{in}}((X, \omega_X)) &= \min_{x \in X} \text{in}_X(x) && \text{for all } (X, \omega_X) \in \mathcal{CN}. \end{aligned}$$

Then both m^{in} and m^{out} are \mathbb{R} -invariants. We take the minimum when defining m^{out} , m^{in} because for any network (X, ω_X) , we have $\max_{x \in X} \text{out}_X(x) = \max_{x \in X} \text{in}_X(x) = \text{diam}(X)$. Also observe that the minima are achieved above because X is compact.

Proposition 144. *The maps out , in , tr , spec , and spec^\bullet are $\text{pow}(\mathbb{R})$ -invariants. Similarly, diam , m^{out} , and m^{in} are \mathbb{R} -invariants.*

Next we see that the motif sets defined in §1.6.4 are also invariants.

Definition 46 (Motif sets are metric space valued invariants). Our use of motif sets is motivated by the following observation, which appeared in [87, Section 5]. For any $n \in \mathbb{N}$, let $\mathcal{C}(\mathbb{R}^{n \times n})$ denote the set of closed subsets of $\mathbb{R}^{n \times n}$. Under the Hausdorff distance induced by the ℓ^∞ metric on $\mathbb{R}^{n \times n}$, this set becomes a valid metric space [17, Proposition 7.3.3]. The motif sets defined in Definition 29 define a metric space valued invariant as follows: for each $n \in \mathbb{N}$, let $M_n : \mathcal{CN} \rightarrow \mathcal{C}(\mathbb{R}^{n \times n})$ be the map $X \mapsto M_n(X)$. We call this the *motif set invariant*. So for $(X, \omega_X), (Y, \omega_Y) \in \mathcal{CN}$, for each $n \in \mathbb{N}$, we let $(Z, d_Z) = (\mathbb{R}^{n \times n}, \ell^\infty)$ and consider the following distance between the n -motif sets of X and Y :

$$d_n(M_n(X), M_n(Y)) := d_{\mathcal{H}}^Z(M_n(X), M_n(Y)).$$

Since $d_{\mathcal{H}}$ is a proper distance between closed subsets, $d_n(M_n(X), M_n(Y)) = 0$ if and only if $M_n(X) = M_n(Y)$.

2.6.1 Quantitative stability of invariants of networks

Let (V, d_V) be a given pseudometric space. The V -valued invariant $\iota : \mathcal{CN} \rightarrow V$ is said to be *quantitatively stable* if there exists a constant $L > 0$ such that

$$d_V(\iota(X), \iota(Y)) \leq L \cdot d_{\mathcal{N}_1}(X, Y)$$

for all networks X and Y . The least constant L such that the above holds for all $X, Y \in \mathcal{CN}$ is the Lipschitz constant of ι and is denoted $\mathbf{L}(\iota)$.

Note that by identifying a non-constant quantitatively stable V -valued invariant ι , we immediately obtain a lower bound for the $d_{\mathcal{N}}$ distance between any two compact networks (X, ω_X) and (Y, ω_Y) . Furthermore, given a finite family $\iota_\alpha : \mathcal{CN} \rightarrow V$, $\alpha \in A$, of non-constant quantitatively stable invariants, we may obtain the following lower bound for the distance between compact networks X and Y :

$$d_{\mathcal{N}}(X, Y) \geq \left(\max_{\alpha \in A} \mathbf{L}(\iota_\alpha) \right)^{-1} \max_{\alpha \in A} d_V(\iota_\alpha(X), \iota_\alpha(Y)).$$

It is often the case that computing $d_V(\iota(X), \iota(Y))$ is substantially simpler than computing the $d_{\mathcal{N}}$ distance between X and Y (which leads to a possibly NP-hard problem). The invariants described in the previous section are quantitatively stable.

Proposition 145. *The invariants diam , tr , out , in , m^{out} , and m^{in} are quantitatively stable, with Lipschitz constant $\mathbf{L} = 2$.*

Example 146. Proposition 145 provides simple lower bounds for the $d_{\mathcal{N}}$ distance between compact networks. One application is the following: for all networks X and Y , we have $d_{\mathcal{N}}(X, Y) \geq \frac{1}{2} |\text{diam}(X) - \text{diam}(Y)|$. For example, consider the weight matrices

$$\Sigma := \begin{pmatrix} 0 & 5 & 2 \\ 3 & 1 & 4 \\ 1 & 4 & 3 \end{pmatrix} \text{ and } \Sigma' := \begin{pmatrix} 3 & 4 & 2 \\ 3 & 1 & 5 \\ 3 & 3 & 4 \end{pmatrix}.$$

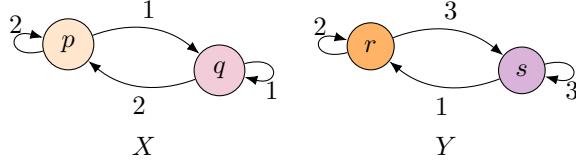


Figure 2.3: Lower-bounding $d_{\mathcal{N}}$ by using global spectra (cf. Example 148).

Let $X = N_3(\Sigma)$ and $Y = N_3(\Sigma')$. By comparing the diagonals, we can easily see that $X \not\cong^s Y$, but let us see how the invariants we proposed can help. Note that $\text{diam}(X) = \text{diam}(Y) = 5$, so the lower bound provided by diameter ($\frac{1}{2}|5 - 5| = 0$) does not help in telling the networks apart. However, $\text{tr}(X) = \{0, 1, 3\}$ and $\text{tr}(Y) = \{3, 1, 4\}$, and Proposition 145 then yields

$$d_{\mathcal{N}}(X, Y) \geq \frac{1}{2} d_{\mathcal{H}}^{\mathbb{R}}(\{0, 1, 3\}, \{1, 3, 4\}) = \frac{1}{2}.$$

Consider now the out and in maps. Note that one has $\text{out}(X) = \{5, 4\}$, $\text{out}(Y) = \{4, 5\}$, $\text{in}(X) = \{3, 5, 4\}$, and $\text{in}(Y) = \{3, 4, 5\}$. Then $d_{\mathcal{H}}^{\mathbb{R}}(\text{out}(X), \text{out}(Y)) = 0$, and $d_{\mathcal{H}}^{\mathbb{R}}(\text{in}(X), \text{in}(Y)) = 0$. Thus in both cases, we obtain $d_{\mathcal{N}}(X, Y) \geq 0$. So in this particular example, the out and in maps are not useful for obtaining a lower bound to $d_{\mathcal{N}}(X, Y)$ via Proposition 145.

Now we state a proposition regarding the stability of global and local spectrum invariants. These will be of particular interest for computational purposes as we explain in §4.2.

Proposition 147. *Let spec^\bullet refer to either the out or in version of local spectrum. Then, for all $(X, \omega_X), (Y, \omega_Y) \in \mathcal{CN}$ we have*

$$\begin{aligned} d_{\mathcal{N}}(X, Y) &\geq \frac{1}{2} \inf_{R \in \mathcal{R}} \sup_{(x, y) \in R} d_{\mathcal{H}}^{\mathbb{R}}(\text{spec}_X^\bullet(x), \text{spec}_Y^\bullet(y)) \\ &\geq \frac{1}{2} d_{\mathcal{H}}^{\mathbb{R}}(\text{spec}(X), \text{spec}(Y)). \end{aligned}$$

As a corollary, we get $\mathbf{L}(\text{spec}^\bullet) = \mathbf{L}(\text{spec}) = 2$.

Example 148 (An application of Proposition 147). Consider the networks in Figure 2.3. By Proposition 147, we may calculate a lower bound for $d_{\mathcal{N}}(X, Y)$ by simply computing the Hausdorff distance between $\text{spec}(X)$ and $\text{spec}(Y)$, and dividing by 2. In this example, $\text{spec}(X) = \{1, 2\}$ and $\text{spec}(Y) = \{1, 2, 3\}$. Thus $d_{\mathcal{H}}^{\mathbb{R}}(\text{spec}(X), \text{spec}(Y)) = 1$, and $d_{\mathcal{N}}(X, Y) \geq \frac{1}{2}$.

Computing the lower bound involving local spectra requires solving a bottleneck linear assignment problem over the set of all correspondences between X and Y . This can be solved in polynomial time; details are provided in §4.2. The second lower bound stipulates computing the Hausdorff distance on \mathbb{R} between the (global) spectra of X and Y – a computation which can be carried out in (smaller) polynomial time as well.

To conclude this section, we state a theorem asserting that motif sets form a family of quantitatively stable invariants.

Theorem 149. *For each $n \in \mathbb{N}$, M_n is a stable invariant with $L(M_n) = 2$.*

2.6.2 Proofs involving lower bounds for $d_{\mathcal{N}}$

Proof of Proposition 144. All of these cases are easy to check, so we will just record the proof for spec . Suppose (X, ω_X) and (Y, ω_Y) are strongly isomorphic via φ . Let $x \in X$. Let $\alpha \in \text{spec}_X(x)$. Then there exists $x' \in X$ such that $\alpha = \omega_X(x, x')$. But since $\omega_X(x, x') = \omega_Y(\varphi(x), \varphi(x'))$, we also have $\alpha \in \text{spec}_Y(\varphi(x))$. Thus $\text{spec}_X(x) \subseteq \text{spec}_Y(\varphi(x))$. The reverse containment is similar. Thus for any $x \in X$, $\text{spec}_X(x) = \text{spec}_Y(\varphi(x))$. Since $\text{spec}(X) = \bigcup_{x \in X} \text{spec}_X(x)$, it follows that $\text{spec}(X) = \text{spec}(Y)$. \square

Lemma 150. *Let $(X, \omega_X), (Y, \omega_Y) \in \mathcal{CN}$. Let f represent any of the maps tr, out , and in , and let f_X (resp. f_Y) represent the corresponding map $\text{tr}_X, \text{out}_X, \text{in}_X$ (resp. $\text{tr}_Y, \text{out}_Y, \text{in}_Y$). Then we obtain:*

$$d_{\mathcal{H}}^{\mathbb{R}}(f(X), f(Y)) = \inf_{R \in \mathcal{R}(X, Y)} \sup_{(x, y) \in R} |f_X(x) - f_Y(y)|.$$

Proof of Lemma 150. Observe that $f(X) = \{f_X(x) : x \in X\} = f_X(X)$, so we need to show

$$d_{\mathcal{H}}^{\mathbb{R}}(f_X(X), f_Y(Y)) = \inf_{R \in \mathcal{R}(X, Y)} \sup_{(x, y) \in R} |f_X(x) - f_Y(y)|.$$

Recall that by the definition of Hausdorff distance on \mathbb{R} , we have

$$d_{\mathcal{H}}^{\mathbb{R}}(f_X(X), f_Y(Y)) = \max \left\{ \sup_{x \in X} \inf_{y \in Y} |f_X(x) - f_Y(y)|, \sup_{y \in Y} \inf_{x \in X} |f_X(x) - f_Y(y)| \right\}.$$

Let $a \in X$ and let $R \in \mathcal{R}(X, Y)$. Then there exists $b \in Y$ such that $(a, b) \in R$. Then we have:

$$\begin{aligned} |f_X(a) - f_Y(b)| &\leq \sup_{(x, y) \in R} |f_X(x) - f_Y(y)|, \text{ and so} \\ \inf_{b \in Y} |f_X(a) - f_Y(b)| &\leq \sup_{(x, y) \in R} |f_X(x) - f_Y(y)|. \end{aligned}$$

This holds for all $a \in X$. Then,

$$\sup_{a \in X} \inf_{b \in Y} |f_X(a) - f_Y(b)| \leq \sup_{(x, y) \in R} |f_X(x) - f_Y(y)|.$$

This holds for all $R \in \mathcal{R}(X, Y)$. So we have

$$\sup_{a \in X} \inf_{b \in Y} |f_X(a) - f_Y(b)| \leq \inf_{R \in \mathcal{R}} \sup_{(x,y) \in R} |f_X(x) - f_Y(y)|.$$

By a similar argument, we also have

$$\sup_{b \in Y} \inf_{a \in X} |f_X(a) - f_Y(b)| \leq \inf_{R \in \mathcal{R}} \sup_{(x,y) \in R} |f_X(x) - f_Y(y)|.$$

$$\text{Thus } d_{\mathcal{H}}^{\mathbb{R}}(f_X(X), f_Y(Y)) \leq \inf_{R \in \mathcal{R}} \sup_{(x,y) \in R} |f_X(x) - f_Y(y)|.$$

Now we show the reverse inequality. Let $x \in X$, and let $\eta > d_{\mathcal{H}}^{\mathbb{R}}(f_X(X), f_Y(Y))$. Then there exists $y \in Y$ such that $|f_X(x) - f_Y(y)| < \eta$. Define $\varphi(x) = y$, and extend φ to all of X in this way. Let $y \in Y$. Then there exists $x \in X$ such that $|f_X(x) - f_Y(y)| < \eta$. Define $\psi(y) = x$, and extend ψ to all of Y in this way. Let $R = \{(x, \varphi(x)) : x \in X\} \cup \{(\psi(y), y) : y \in Y\}$. Then for each $(a, b) \in R$, we have $|f_X(a) - f_Y(b)| < \eta$. Thus we have $\inf_{R \in \mathcal{R}} \sup_{(x,y) \in R} |f_X(x) - f_Y(y)| < \eta$. Since $\eta > d_{\mathcal{H}}^{\mathbb{R}}(f_X(X), f_Y(Y))$ was arbitrary, it follows that

$$\inf_{R \in \mathcal{R}(X, Y)} \sup_{(x,y) \in R} |f_X(x) - f_Y(y)| \leq d_{\mathcal{H}}^{\mathbb{R}}(f_X(X), f_Y(Y)). \quad \square$$

Proof of Proposition 145. Let $\eta > d_{\mathcal{N}}(X, Y)$. We break this proof into three parts.

The diam case. Recall that diam is an \mathbb{R} -valued invariant, so we wish to show $|\text{diam}(X) - \text{diam}(Y)| \leq 2d_{\mathcal{N}}(X, Y)$. Let $R \in \mathcal{R}(X, Y)$ be such that for any $(a, b), (a', b') \in R$, we have $|\omega_X(a, a') - \omega_Y(b, b')| < 2\eta$.

Let $x, x' \in X$ such that $|\omega_X(x, x')| = \text{diam}(X)$, and let y, y' be such that $(x, y), (x', y') \in R$. Then we have:

$$\begin{aligned} |\omega_X(x, x') - \omega_Y(y, y')| &< 2\eta \\ |\omega_X(x, x') - \omega_Y(y, y')| + |\omega_Y(y, y')| &< 2\eta + |\omega_Y(y, y')| \\ |\omega_X(x, x')| &< \text{diam}(Y) + 2\eta. \end{aligned}$$

Thus $\text{diam}(X) < \text{diam}(Y) + 2\eta$

Similarly, we get $\text{diam}(Y) < \text{diam}(X) + 2\eta$. It follows that $|\text{diam}(X) - \text{diam}(Y)| < 2\eta$. Since $\eta > d_{\mathcal{N}}(X, Y)$ was arbitrary, it follows that:

$$|\text{diam}(X) - \text{diam}(Y)| \leq 2d_{\mathcal{N}}(X, Y).$$

For tightness, consider the networks $X = N_1(1)$ and $Y = N_1(2)$. By direct computation, we have that $d_{\mathcal{N}}(X, Y) = \frac{1}{2}$. On the other hand, $\text{diam}(X) = 1$ and $\text{diam}(Y) = 2$ so that $|\text{diam}(X) - \text{diam}(Y)| = 1 = 2d_{\mathcal{N}}(X, Y)$.

The cases tr, out, and in. First we show $\mathbf{L}(\text{tr}) = 2$. By Lemma 150, it suffices to show

$$\inf_{R \in \mathcal{R}(X, Y)} \sup_{(x,y) \in R} |\text{tr}_X(x) - \text{tr}_Y(y)| < 2\eta.$$

Let $R \in \mathcal{R}(X, Y)$ be such that for any $(a, b), (a', b') \in R$, we have $|\omega_X(a, a') - \omega_Y(b, b')| < 2\eta$. Then we obtain $|\omega_X(a, a) - \omega_Y(b, b)| < 2\eta$. Thus $|\text{tr}_X(a) - \text{tr}_Y(b)| < 2\eta$. Since $(a, b) \in R$ was arbitrary, it follows that $\sup_{(a,b) \in R} |\text{tr}_X(a) - \text{tr}_X(b)| < 2\eta$. It follows that $\inf_{R \in \mathcal{R}} \sup_{(a,b) \in R} |\text{tr}_X(a) - \text{tr}_X(b)| < 2\eta$. The result now follows because $\eta > d_N(X, Y)$ was arbitrary. The proofs for out and in are similar, so we just show the former. By Lemma 150, it suffices to show

$$\inf_{R \in \mathcal{R}(X, Y)} \sup_{(x,y) \in R} |\text{out}_X(x) - \text{out}_Y(y)| < 2\eta.$$

Recall that $\text{out}_X(x) = \max_{x' \in X} |\omega_X(x, x')|$. Let $R \in \mathcal{R}(X, Y)$ be such that $|\omega_X(x, x') - \omega_Y(y, y')| < 2\eta$ for any $(x, y), (x', y') \in R$. By triangle inequality, it follows that $|\omega_X(x, x')| < |\omega_Y(y, y')| + 2\eta$. In particular, for $(x', y') \in R$ such that $|\omega_X(x, x')| = \text{out}_X(x)$, we have $\text{out}_X(x) < |\omega_Y(y, y')| + 2\eta$. Hence $\text{out}_X(x) < \text{out}_Y(y) + 2\eta$. Similarly, $\text{out}_Y(y) < \text{out}_X(x) + 2\eta$. Thus we have $|\text{out}_X(x) - \text{out}_Y(y)| < 2\eta$. This holds for all $(x, y) \in R$, so we have:

$$\sup_{(x,y) \in R} |\text{out}_X(x) - \text{out}_Y(y)| < 2\eta.$$

Minimizing over all correspondences, we get:

$$\inf_{R \in \mathcal{R}} \sup_{(a,b) \in R} |\text{out}_X(a) - \text{out}_Y(b)| < 2\eta.$$

The result follows because $\eta > d_N(X, Y)$ was arbitrary.

Finally, we need to show that our bounds for the Lipschitz constant are tight. Let $X = N_1(1)$ and let $Y = N_1(2)$. Then $d_N(X, Y) = \frac{1}{2}$. We also have $d_{\mathcal{H}}^{\mathbb{R}}(\text{tr}(X), \text{tr}(Y)) = |1 - 2| = 1$, and similarly $d_{\mathcal{H}}^{\mathbb{R}}(\text{out}(X), \text{out}(Y)) = d_{\mathcal{H}}^{\mathbb{R}}(\text{in}(X), \text{in}(Y)) = 1$.

The cases m^{out} and m^{in} . The two cases are similar, so let's just prove $L(m^{\text{out}}) = 2$. Since m^{out} is an \mathbb{R} -invariant, we wish to show $|m^{\text{out}}(X) - m^{\text{out}}(Y)| < 2\eta$. It suffices to show:

$$|m^{\text{out}}(X) - m^{\text{out}}(Y)| \leq d_{\mathcal{H}}^{\mathbb{R}}(\text{out}(X), \text{out}(Y)),$$

because we have already shown

$$d_{\mathcal{H}}^{\mathbb{R}}(\text{out}(X), \text{out}(Y)) = \inf_{R \in \mathcal{R}(X, Y)} \sup_{(x,y) \in R} |\text{out}_X(x) - \text{out}_Y(y)| < 2\eta.$$

Here we have used Lemma 150 for the first equality above.

Let $\varepsilon > d_{\mathcal{H}}^{\mathbb{R}}(\text{out}(X), \text{out}(Y))$. Then for any $x \in X$, there exists $y \in Y$ such that:

$$|\text{out}_X(x) - \text{out}_Y(y)| < \varepsilon.$$

Let $a \in X$ be such that $m^{\text{out}}(X) = \text{out}_X(a)$. Then we have:

$$|\text{out}_X(a) - \text{out}_Y(y)| < \varepsilon,$$

for some $y \in Y$. In particular, we have:

$$m^{\text{out}}(Y) \leq \text{out}_Y(y) < \varepsilon + \text{out}_X(a) = \varepsilon + m^{\text{out}}(X).$$

Similarly, we obtain:

$$m^{\text{out}}(X) < \varepsilon + m^{\text{out}}(Y).$$

Thus we have $|m^{\text{out}}(X) - m^{\text{out}}(Y)| < \varepsilon$. Since $\varepsilon > d_{\mathcal{H}}^{\mathbb{R}}(\text{out}(X), \text{out}(Y))$ was arbitrary, we have:

$$|m^{\text{out}}(X) - m^{\text{out}}(Y)| \leq d_{\mathcal{H}}^{\mathbb{R}}(\text{out}(X), \text{out}(Y)).$$

The inequality now follows by Lemma 150 and our proof in the case of the out map.

For tightness, note that $|m^{\text{out}}(N_1(1)) - m^{\text{out}}(N_1(2))| = |1 - 2| = 1 = 2 \cdot \frac{1}{2} = 2d_{\mathcal{N}}(N_1(1), N_1(2))$. The same example works for the m^{in} case. \square

Proof of Proposition 147. (First inequality.) Let $X, Y \in \mathcal{CN}$ and let $\eta > d_{\mathcal{N}}(X, Y)$. Let $R \in \mathcal{R}(X, Y)$ be such that $\sup_{(x,y),(x',y') \in R} |\omega_X(x, x') - \omega_Y(y, y')| < 2\eta$. Let $(x, y) \in R$, and let $\alpha \in \text{spec}_X(x)$. Then there exists $x' \in X$ such that $\omega_X(x, x') = \alpha$. Let $y' \in Y$ be such that $(x', y') \in R$. Let $\beta = \omega_Y(y, y')$. Note $\beta \in \text{spec}_Y(y)$. Also note that $|\alpha - \beta| < 2\eta$. By a symmetric argument, for each $\beta \in \text{spec}_Y(y)$, there exists $\alpha \in \text{spec}_X(x)$ such that $|\alpha - \beta| < 2\eta$. So $d_{\mathcal{H}}^{\mathbb{R}}(\text{spec}_X(x), \text{spec}_Y(y)) < 2\eta$. This is true for any $(x, y) \in R$, and so we have $\sup_{(x,y) \in R} d_{\mathcal{H}}^{\mathbb{R}}(\text{spec}_X(x), \text{spec}_Y(y)) \leq 2\eta$. Then we have:

$$\inf_{R \in \mathcal{R}} \sup_{(x,y) \in R} d_{\mathcal{H}}^{\mathbb{R}}(\text{spec}_X(x), \text{spec}_Y(y)) \leq 2\eta.$$

Since $\eta > d_{\mathcal{N}}(X, Y)$ was arbitrary, the first inequality follows.

(Second inequality.) Let $R \in \mathcal{R}(X, Y)$. Let $\eta(R) = \sup_{(x,y) \in R} d_{\mathcal{H}}^{\mathbb{R}}(\text{spec}_X(x), \text{spec}_Y(y))$. Let $\alpha \in \text{spec}(X)$. Then $\alpha \in \text{spec}_X(x)$ for some $x \in X$. Let $y \in Y$ such that $(x, y) \in R$. Then there exists $\beta \in \text{spec}_Y(y)$ such that $|\alpha - \beta| \leq d_{\mathcal{H}}^{\mathbb{R}}(\text{spec}_X(x), \text{spec}_Y(y))$, and in particular, $|\alpha - \beta| \leq \eta(R)$. In other words, for each $\alpha \in \text{spec}(X)$, there exists $\beta \in \text{spec}(Y)$ such that $|\alpha - \beta| \leq \eta(R)$. By a symmetric argument, for each $\beta \in \text{spec}(Y)$, there exists $\alpha \in \text{spec}(X)$ such that $|\alpha - \beta| \leq \eta(R)$. Thus $d_{\mathcal{H}}^{\mathbb{R}}(\text{spec}(X), \text{spec}(Y)) \leq \eta(R)$. This holds for any $R \in \mathcal{R}$. Thus we have

$$d_{\mathcal{H}}^{\mathbb{R}}(\text{spec}(X), \text{spec}(Y)) \leq \inf_{R \in \mathcal{R}} \sup_{(x,y) \in R} d_{\mathcal{H}}^{\mathbb{R}}(\text{spec}_X(x), \text{spec}_Y(y)).$$

This proves the second inequality. \square

Proof of Theorem 149. Let $n \in \mathbb{N}$. We wish to show $d_n(\text{M}_n(X), \text{M}_n(Y)) \leq 2d_{\mathcal{N}}(X, Y)$. Let $R \in \mathcal{R}(X, Y)$. Let $(x_i) \in X^n$, and let $(y_i) \in Y^n$ be such that for each i , we have $(x_i, y_i) \in R$. Then for all $j, k \in \{1, \dots, n\}$, $|\omega_X(x_i, x_j) - \omega_Y(y_i, y_j)| \leq \text{dis}(R)$.

Thus $\inf_{(y_i) \in Y^n} |\omega_X(x_i, x_j) - \omega_Y(y_i, y_j)| \leq \text{dis}(R)$. This is true for any $(x_i) \in X^n$. Thus we get:

$$\sup_{(x_i) \in X^n} \inf_{(y_i) \in Y^n} |\omega_X(x_i, x_j) - \omega_Y(y_i, y_j)| \leq \text{dis}(R).$$

By a symmetric argument, we get $\sup_{(y_i) \in Y^n} \inf_{(x_i) \in X^n} |\omega_X(x_i, x_j) - \omega_Y(y_i, y_j)| \leq \text{dis}(R)$. Thus $d_n(\text{M}_n(X), \text{M}_n(Y)) \leq \text{dis}(R)$. This holds for any $R \in \mathcal{R}(X, Y)$. Thus we have

$$d_n(\text{M}_n(X), \text{M}_n(Y)) \leq \inf_{R \in \mathcal{R}(X, Y)} \text{dis}(R) = 2d_N(X, Y).$$

For tightness, let $X = N_1(1)$ and let $Y = N_1(2)$. Then $d_N(X, Y) = \frac{1}{2}$, so we wish to show $d_n(\text{M}_n(X), \text{M}_n(Y)) = 1$ for each $n \in \mathbb{N}$. Let $n \in \mathbb{N}$. Let $\mathbb{1}_{n \times n}$ denote the $n \times n$ matrix with 1 in each entry. Then $\text{M}_n(X) = \{\mathbb{1}_{n \times n}\}$ and $\text{M}_n(Y) = \{2 \cdot \mathbb{1}_{n \times n}\}$. Thus $d_n(\text{M}_n(X), \text{M}_n(Y)) = 1$. Since n was arbitrary, we conclude that equality holds for each $n \in \mathbb{N}$. \square

2.7 Proofs from §1.9

Proof of Lemma 106. First suppose $p \in [1, \infty)$. We will construct a sequence of continuous functionals that converges uniformly to dis_p . Since the uniform limit of continuous functions is continuous, this will show that dis_p is continuous.

Continuous functions are dense in $L^p(\lambda_I^{\otimes 2})$ (see e.g. [14]), so for each $n \in \mathbb{N}$, we pick $\sigma_X^n, \sigma_Y^n \in L^p(\lambda_I^{\otimes 2})$ such that

$$\|\sigma_X - \sigma_X^n\|_{L^p(\lambda_I^{\otimes 2})} \leq 1/n, \quad \|\sigma_Y - \sigma_Y^n\|_{L^p(\lambda_I^{\otimes 2})} \leq 1/n.$$

For each $n \in \mathbb{N}$, define the functional $\text{dis}_p^n : \mathcal{C}(\lambda_I, \lambda_I) \rightarrow \mathbb{R}_+$ as follows:

$$\text{dis}_p^n(\nu) := \left(\int_{I \times I} \int_{I \times I} |\sigma_X^n(i, i') - \sigma_Y^n(j, j')|^p d\nu(i, j) d\nu(i', j') \right)^{1/p}.$$

Because $|\sigma_X^n - \sigma_Y^n|^p$ is continuous and hence bounded on the compact cube $I^2 \times I^2$, we know that $|\sigma_X^n - \sigma_Y^n|^p \in C_b(I^2 \times I^2)$.

We claim that dis_p^n is continuous. Since the narrow topology on $\text{Prob}(I \times I)$ is induced by a distance [5, Remark 5.1.1], it suffices to show sequential continuity. Let $\nu \in \mathcal{C}(\lambda_I, \lambda_I)$, and let $(\nu_m)_{m \in \mathbb{N}}$ be a sequence in $\mathcal{C}(\lambda_I, \lambda_I)$ converging narrowly to ν . Then we have

$$\begin{aligned} \lim_{m \rightarrow \infty} \text{dis}_p^n(\nu_m) &= \lim_{m \rightarrow \infty} \left(\int_{I^2 \times I^2} |\sigma_X^n - \sigma_Y^n|^p d\nu_m \otimes d\nu_m \right)^{1/p} \\ &= \left(\int_{I^2 \times I^2} |\sigma_X^n - \sigma_Y^n|^p d\nu \otimes d\nu \right)^{1/p} \\ &= \text{dis}_p^n(\nu). \end{aligned}$$

Here the second equality follows from the definition of convergence in the narrow topology and the fact that the integrand is bounded and continuous. This shows sequential continuity (hence continuity) of dis_p^n .

Finally, we show that $(\text{dis}_p^n)_{n \in \mathbb{N}}$ converges to dis_p uniformly. Let $\mu \in \mathcal{C}(\lambda_I, \lambda_I)$. Then,

$$\begin{aligned}
|\text{dis}_p(\mu) - \text{dis}_p^n(\mu)| &= \left| \|\sigma_X - \sigma_Y\|_{L^p(\mu^{\otimes 2})} - \|\sigma_X^n - \sigma_Y^n\|_{L^p(\mu^{\otimes 2})} \right| \\
&\leq \|\sigma_X - \sigma_Y - \sigma_X^n + \sigma_Y^n\|_{L^p(\mu^{\otimes 2})} \\
&\leq \|\sigma_X - \sigma_X^n\|_{L^p(\mu^{\otimes 2})} + \|\sigma_Y - \sigma_Y^n\|_{L^p(\mu^{\otimes 2})} \\
&= \left(\int_{I \times I} \int_{I \times I} |\sigma_X(i, i') - \sigma_X^n(i, i')|^p d\mu(i, j) d\mu(i', j') \right)^{1/p} \dots \\
&\quad + \left(\int_{I \times I} \int_{I \times I} |\sigma_Y(j, j') - \sigma_Y^n(j, j')|^p d\mu(i, j) d\mu(i', j') \right)^{1/p} \\
&= \left(\int_I \int_I |\sigma_X(i, i') - \sigma_X^n(i, i')|^p d\lambda_I(i) d\lambda_I(i') \right)^{1/p} \dots \\
&\quad + \left(\int_I \int_I |\sigma_Y(j, j') - \sigma_Y^n(j, j')|^p d\lambda_I(j) d\lambda_I(j') \right)^{1/p} \\
&= \|\sigma_X - \sigma_X^n\|_{L^p(\lambda_I^{\otimes 2})} + \|\sigma_Y - \sigma_Y^n\|_{L^p(\lambda_I^{\otimes 2})} \\
&\leq 2/n.
\end{aligned}$$

But $\mu \in \mathcal{C}(\lambda_I, \lambda_I)$ was arbitrary. This shows that dis_p is the uniform limit of continuous functions, hence is continuous. Here the first and second inequalities followed from Minkowski's inequality.

Now suppose $p = \infty$. Let $\mu \in \mathcal{C}(\lambda_I, \lambda_I)$ be arbitrary. Recall that because we are working over probability spaces, Jensen's inequality can be used to show that for any $1 \leq q \leq r < \infty$, we have $\text{dis}_q(\mu) \leq \text{dis}_r(\mu)$. Moreover, we have $\lim_{q \rightarrow \infty} \text{dis}_q(\mu) = \text{dis}_\infty(\mu)$. The supremum of a family of continuous functions is lower semicontinuous. In our case, $\text{dis}_\infty = \sup\{\text{dis}_q : q \in [1, \infty)\}$, and we have shown above that all the functions in this family are continuous. Hence dis_∞ is lower semicontinuous. \square

Proof of Theorem 111. Let $(X, \omega_X, \mu_X), (Y, \omega_Y, \mu_Y), (Z, \omega_Y, \mu_Y) \in \mathcal{N}_m$. It is clear that $d_{\mathcal{N}, p}(X, Y) \geq 0$. To show $d_{\mathcal{N}, p}(X, X) = 0$, consider the diagonal coupling Δ (see Example 101). For $p \in [1, \infty)$, we have:

$$\begin{aligned}
\text{dis}_p(\Delta) &= \left(\int_{X \times X} \int_{X \times X} |\omega_X(x, x') - \omega_X(z, z')|^p d\Delta(x, z) d\Delta(x', z') \right)^{1/p} \\
&= \left(\int_X \int_X |\omega_X(x, x') - \omega_X(x, x')|^p d\mu_X(x) d\mu_X(x') \right)^{1/p} \\
&= 0.
\end{aligned}$$

For $p = \infty$, we have:

$$\begin{aligned}\text{dis}_p(\Delta) &= \sup\{|\omega_X(x, x') - \omega_X(z, z')| : (x, z), (x', z') \in \text{supp}(\Delta)\} \\ &= \sup\{|\omega_X(x, x') - \omega_X(x, x')| : x, x' \in \text{supp}(\mu_X)\} \\ &= 0.\end{aligned}$$

Thus $d_{\mathcal{N},p}(X, X) = 0$ for any $p \in [1, \infty]$. For symmetry, notice that for any $\mu \in \mathcal{C}(\mu_X, \mu_Y)$, we can define $\tilde{\mu} \in \mathcal{C}(\mu_Y, \mu_X)$ by $\tilde{\mu}(y, x) = \mu(x, y)$. Then $\text{dis}_p(\mu) = \text{dis}_p(\tilde{\mu})$, and this will show $d_{\mathcal{N},p}(X, Y) = d_{\mathcal{N},p}(Y, X)$.

Finally, we need to check the triangle inequality. Let $\varepsilon > 0$, and let $\mu_{12} \in \mathcal{C}(\mu_X, \mu_Y)$ and $\mu_{23} \in \mathcal{C}(\mu_Y, \mu_Z)$ be couplings such that $2d_{\mathcal{N},p}(X, Y) \geq \text{dis}_p(\mu_{12}) - \varepsilon$ and $2d_{\mathcal{N},p}(Y, Z) \geq \text{dis}_p(\mu_{23}) - \varepsilon$. Invoking Lemma 107, we obtain a probability measure $\mu \in \text{Prob}(X \times Y \times Z)$ with marginals μ_{12}, μ_{23} , and a marginal μ_{13} that is a coupling between μ_X and μ_Z . This coupling is not necessarily optimal. For $p \in [1, \infty)$ we have:

$$\begin{aligned}2d_{\mathcal{N},p}(X, Z) &\leq \text{dis}_p(\mu_{13}) \\ &= \left(\int_{X \times Z} \int_{X \times Z} |\omega_X(x, x') - \omega_Z(z, z')|^p d\mu_{13}(x, z) d\mu_{13}(x', z') \right)^{1/p} \\ &= \left(\int_{X \times Y \times Z} \int_{X \times Y \times Z} |\omega_X(x, x') - \omega_Z(z, z')|^p d\mu(x, y, z) d\mu(x', y', z') \right)^{1/p} \\ &= \|\omega_X - \omega_Y + \omega_Y - \omega_Z\|_{L^p(\mu \otimes \mu)} \\ &\leq \|\omega_X - \omega_Y\|_{L^p(\mu \otimes \mu)} + \|\omega_Y - \omega_Z\|_{L^p(\mu \otimes \mu)} \\ &= \left(\int_{X \times Y} \int_{X \times Y} |\omega_X(x, x') - \omega_Y(y, y')|^p d\mu_{12}(x, y) d\mu_{12}(x', y') \right)^{1/p} \dots \\ &\quad + \left(\int_{Y \times Z} \int_{Y \times Z} |\omega_Y(y, y') - \omega_Z(z, z')|^p d\mu_{23}(y, z) d\mu_{23}(y', z') \right)^{1/p} \\ &\leq 2d_{\mathcal{N},p}(X, Y) + 2d_{\mathcal{N},p}(Y, Z) + 2\varepsilon.\end{aligned}$$

The second inequality above follows from Minkowski's inequality. Letting $\varepsilon \rightarrow 0$ now proves the triangle inequality in the case $p \in [1, \infty)$.

For $p = \infty$ we have:

$$\begin{aligned}2d_{\mathcal{N},p}(X, Z) &\leq \text{dis}_p(\mu_{13}) \\ &= \sup\{|\omega_X(x, x') - \omega_Z(z, z')| : (x, z), (x', z') \in \text{supp}(\mu_{13})\} \\ &= \sup\{|\omega_X(x, x') - \omega_Y(y, y') + \omega_Y(y, y') - \omega_Z(z, z')| : (x, y, z), (x', y', z') \in \text{supp}(\mu)\} \\ &\leq \sup\{|\omega_X(x, x') - \omega_Y(y, y')| + |\omega_Y(y, y') - \omega_Z(z, z')| \\ &\quad : (x, y), (x', y') \in \text{supp}(\mu_{12}), (y, z), (y', z') \in \text{supp}(\mu_{23})\} \\ &\leq \text{dis}_p(\mu_{12}) + \text{dis}_p(\mu_{23}) \\ &\leq 2d_{\mathcal{N},p}(X, Y) + 2d_{\mathcal{N},p}(Y, Z) + 2\varepsilon.\end{aligned}$$

Letting $\varepsilon \rightarrow 0$ now proves the triangle inequality in the case $p = \infty$. This concludes our proof. \square

Proof of Theorem 112. First suppose $p \in [1, \infty)$. By the construction in Section 1.9.3, we pass into interval representations of X and Y . As noted in Section 1.9.3, the choice of parametrization is not necessarily unique, but this does not affect the argument. Let (I, σ_X, λ_I) , (I, σ_Y, λ_I) denote these representations. By Lemma 106, the dis_p functional is continuous on the space of couplings between these two networks. By Lemma 105, this space of couplings is compact. Thus dis_p achieves its infimum.

Let $\mu \in \mathcal{C}(\lambda_I, \lambda_I)$ denote this minimizer of dis_p . By Remark 102, we can also take couplings $\mu_X \in \mathcal{C}(\mu_X, \lambda_I)$ and $\mu_Y \in \mathcal{C}(\lambda_I, \mu_Y)$ which have zero distortion. By Lemma 107, we can glue together μ_X , μ , and μ_Y to obtain a coupling $\nu \in \mathcal{C}(\mu_X, \mu_Y)$. By the proof of the triangle inequality in Theorem 111, we have:

$$\text{dis}_p(\nu) \leq \text{dis}_p(\mu_X) + \text{dis}_p(\mu) + \text{dis}_p(\mu_Y) = \text{dis}_p(\mu) = 2d_{\mathcal{N},p}((I, \sigma_X, \lambda_I), (I, \sigma_Y, \lambda_I)).$$

Also by the triangle inequality, we have $d_{\mathcal{N},p}((I, \sigma_X, \lambda_I), (I, \sigma_Y, \lambda_I)) \leq d_{\mathcal{N},p}(X, Y)$. It follows that $\nu \in \mathcal{C}(\mu_X, \mu_Y)$ is optimal.

The case $p = \infty$ is analogous, because lower semicontinuity (Lemma 106) combined with compactness (Lemma 105) is sufficient to guarantee that dis_∞ achieves its infimum on $\mathcal{C}(\lambda_I, \lambda_I)$. \square

Proof of Theorem 113. Fix $p \in [1, \infty)$. For the backwards direction, suppose there exist Z and measurable maps $f : Z \rightarrow X$ and $g : Z \rightarrow Y$ such that the appropriate conditions are satisfied. We first claim that $d_{\mathcal{N},p}((X, \omega_X, \mu_X), (Z, f^*\omega_X, \mu_Z)) = 0$.

To see the claim, define $\mu \in \text{Prob}(X \times Z)$ by $\mu := (f, \text{id})_*\mu_Z$. Then,

$$\begin{aligned} & \int_{X \times Z} \int_{X \times Z} |\omega_X(x, x') - f^*\omega_X(z, z')|^p d\mu(x, z) d\mu(x', z') \\ & \int_{X \times Z} \int_{X \times Z} |\omega_X(x, x') - \omega_X(f(z), f(z'))|^p d\mu(x, z) d\mu(x', z') \\ & \int_Z \int_Z |\omega_X(f(z), f(z')) - \omega_X(f(z), f(z'))|^p d\mu_Z(z) d\mu_Z(z') = 0. \end{aligned}$$

This verifies the claim. Similarly we have $d_{\mathcal{N},p}((Y, \omega_Y, \mu_X), (Z, g^*\omega_Y, \mu_Z)) = 0$. Using the diagonal coupling along with the assumption, we have $d_{\mathcal{N},p}((Z, f^*\omega_X, \mu_Z), (Z, g^*\omega_Y, \mu_Z)) = 0$. By triangle inequality, we then have $d_{\mathcal{N},p}(X, Y) = 0$.

For the forwards direction, let $\mu \in \mathcal{C}(\mu_X, \mu_Y)$ be an optimal coupling with $\text{dis}_p(\mu) = 0$ (Theorem 112). Define $Z := X \times Y$, $\mu_Z := \mu$. Then the projection maps $\pi_X : Z \rightarrow X$ and $\pi_Y : Z \rightarrow Y$ are measurable. We also have $(\pi_X)_*\mu = \mu_X$ and $(\pi_Y)_*\mu = \mu_Y$. Since $\text{dis}_p(\mu) = 0$, we also have $\|(\pi_X)^*\omega_X - (\pi_Y)^*\omega_Y\|_\infty = \|\omega_X - \omega_Y\|_\infty = 0$.

The $p = \infty$ case is proved analogously. This concludes the proof. \square

Proof of Theorem 115. Let $(X, \omega_X, \mu_X), (Y, \omega_Y, \mu_Y), (Z, \omega_Z, \mu_Z) \in \mathcal{N}_m$. The proofs that $d_{\mathcal{N},\alpha}^{\mathcal{GP}}(X, Y) \geq 0$, $d_{\mathcal{N},\alpha}^{\mathcal{GP}}(X, X) = 0$, and that $d_{\mathcal{N},\alpha}^{\mathcal{GP}}(X, Y) = d_{\mathcal{N},\alpha}^{\mathcal{GP}}(Y, X)$ are analogous to those used in Theorem 111. Hence we only check the triangle inequality. Let $\varepsilon_{XY} > 2d_{\mathcal{N},\alpha}^{\mathcal{GP}}(X, Y)$, $\varepsilon_{YZ} > 2d_{\mathcal{N},\alpha}^{\mathcal{GP}}(Y, Z)$, and let μ_{XY}, μ_{YZ} be couplings such that

$$\begin{aligned}\mu_{XY}^{\otimes 2}(\{(x, y, x', y') : |\omega_X(x, x') - \omega_Y(y, y')| \geq \varepsilon_{XY}\}) &\leq \alpha \varepsilon_{XY}, \\ \mu_{YZ}^{\otimes 2}(\{(y, z, y', z') : |\omega_Y(y, y') - \omega_Z(z, z')| \geq \varepsilon_{YZ}\}) &\leq \alpha \varepsilon_{YZ}.\end{aligned}$$

For convenience, define:

$$\begin{aligned}A &:= \{((x, y, z), (x', y', z')) \in (X \times Y \times Z)^2 : |\omega_X(x, x') - \omega_Y(y, y')| \geq \varepsilon_{XY}\} \\ B &:= \{((x, y, z), (x', y', z')) \in (X \times Y \times Z)^2 : |\omega_Y(y, y') - \omega_Z(z, z')| \geq \varepsilon_{YZ}\} \\ C &:= \{((x, y, z), (x', y', z')) \in (X \times Y \times Z)^2 : |\omega_X(x, x') - \omega_Z(z, z')| \geq \varepsilon_{XY} + \varepsilon_{YZ}\}.\end{aligned}$$

Next let μ denote the probability measure obtained from gluing μ_{XY} and μ_{YZ} (cf. Lemma 107). This has marginals μ_{XY}, μ_{YZ} , and a marginal $\mu_{XZ} \in \mathcal{C}(\mu_X, \mu_Z)$. We need to show:

$$\mu_{XZ}^{\otimes 2}((\pi_X, \pi_Z)(C)) \leq \alpha(\varepsilon_{XY} + \varepsilon_{YZ}).$$

To show this, it suffices to show $C \subseteq A \cup B$, because then we have $\mu^{\otimes 2}(C) \leq \mu^{\otimes 2}(A) + \mu^{\otimes 2}(B)$ and consequently

$$\begin{aligned}\mu_{XZ}^{\otimes 2}((\pi_X, \pi_Z)(C)) = \mu^{\otimes 2}(C) &\leq \mu^{\otimes 2}(A) + \mu^{\otimes 2}(B) = \mu_{XY}^{\otimes 2}((\pi_X, \pi_Y)(A)) + \mu_{YZ}^{\otimes 2}((\pi_Y, \pi_Z)(B)) \\ &\leq \alpha(\varepsilon_{XY} + \varepsilon_{YZ}).\end{aligned}$$

Let $((x, y, z), (x', y', z')) \in (X \times Y \times Z)^2 \setminus (A \cup B)$. Then we have

$$|\omega_X(x, x') - \omega_Y(y, y')| < \varepsilon_{XY} \text{ and } |\omega_Y(y, y') - \omega_Z(z, z')| < \varepsilon_{YZ}.$$

By the triangle inequality, we then have:

$$|\omega_X(x, x') - \omega_Z(z, z')| \leq |\omega_X(x, x') - \omega_Y(y, y')| + |\omega_Y(y, y') - \omega_Z(z, z')| < \varepsilon_{XY} + \varepsilon_{YZ}.$$

Thus $((x, y, z), (x', y', z')) \in (X \times Y \times Z)^2 \setminus C$. This shows $C \subseteq A \cup B$.

The preceding work shows that $2d_{\mathcal{N},\alpha}^{\mathcal{GP}}(X, Z) \leq \varepsilon_{XY} + \varepsilon_{YZ}$. Since $\varepsilon_{XY} > 2d_{\mathcal{N},\alpha}^{\mathcal{GP}}(X, Y)$ and $\varepsilon_{YZ} > 2d_{\mathcal{N},\alpha}^{\mathcal{GP}}(Y, Z)$ were arbitrary, it follows that $d_{\mathcal{N},\alpha}^{\mathcal{GP}}(X, Z) \leq d_{\mathcal{N},\alpha}^{\mathcal{GP}}(X, Y) + d_{\mathcal{N},\alpha}^{\mathcal{GP}}(Y, Z)$. \square

Proof of Lemma 116. Write $\mathcal{C} := \mathcal{C}(X, Y)$. When $\alpha = 0$ in the $d_{\mathcal{N},\alpha}^{\mathcal{GP}}$ formulation, we have:

$$\begin{aligned}2d_{\mathcal{N},0}^{\mathcal{GP}}(X, Y) &= \inf_{\mu \in \mathcal{C}} \inf \{\varepsilon > 0 : \mu^{\otimes 2}(\{x, y, x', y' \in (X \times Y)^2 : |\omega_X(x, x') - \omega_Y(y, y')| \geq \varepsilon\}) = 0\} \\ &= \inf_{\mu \in \mathcal{C}} \sup \{\varepsilon > 0 : \mu^{\otimes 2}(\{x, y, x', y' \in (X \times Y)^2 : |\omega_X(x, x') - \omega_Y(y, y')| < \varepsilon\}) = 1\} \\ &= \inf_{\mu \in \mathcal{C}} \sup \{|\omega_X(x, x') - \omega_Y(y, y')| : (x, y), (x', y') \in \text{supp}(\mu)\} \\ &= 2d_{\mathcal{N},\infty}(X, Y).\end{aligned}$$

Proof of Theorem 118. Let $t_0 \in \mathbb{R}$, and let $(X, \omega_X, \mu_X), (Y, \omega_Y, \mu_Y) \in \mathcal{N}_m$. Via Lemma 116, write $\varepsilon := d_{\mathcal{N},0}^{\mathcal{GP}}(X, Y) = d_{\mathcal{N},\infty}(X, Y)$. Using Theorem 112, let μ be an optimal coupling between μ_X and μ_Y for which $d_{\mathcal{N},\infty}(X, Y)$ is achieved.

For each $t \in \mathbb{R}$, write $A(X, t) := \{(x, x') \in X \times X : \omega_X(x, x') \leq t\} = \{\omega_X \leq t\}$. Similarly write $A(Y, t) := \{\omega_Y \leq t\}$ for each $t \in \mathbb{R}$.

Let $B := \{(x, y, x', y') \in (X \times Y)^2 : |\omega_X(x, x') - \omega_Y(y, y')| \geq \varepsilon\}$. Also let G denote the complement of B , i.e. $G := \{(x, y, x', y') \in (X \times Y)^2 : |\omega_X(x, x') - \omega_Y(y, y')| < \varepsilon\}$. In particular, notice that for any $(x, y, x', y') \in G$, we have $\omega_X(x, x') < \varepsilon + \omega_Y(y, y')$.

By the definition of ε , we have $\mu^{\otimes 2}(B) = 0$, and hence $\mu^{\otimes 2}(G) = 1$.

In what follows, we will focus on the case $p \in [1, \infty)$ and write out the integrals explicitly. An analogous proof holds for $p = \infty$. We have:

$$\begin{aligned} \text{sub}_{p,t_0}^w(X) &= \left(\int_{A(X,t_0)} |\omega_X(x, x')|^p d\mu_X^{\otimes 2}(x, x') \right)^{1/p} \\ &= \left(\int_{A(X,t_0) \times Y^2} |\omega_X(x, x')|^p d\mu^{\otimes 2}(x, y, x', y') \right)^{1/p} \\ &= \left(\int_{G \cap (A(X,t_0) \times Y^2)} |\omega_X(x, x')|^p d\mu^{\otimes 2}(x, y, x', y') \right)^{1/p} \\ &= \left(\int_{(X \times Y)^2} \mathbb{1}_{G \cap (A(X,t_0) \times Y^2)}^p |\omega_X(x, x') - \omega_Y(y, y') + \omega_Y(y, y')|^p d\mu^{\otimes 2}(x, y, x', y') \right)^{1/p}. \end{aligned}$$

For convenience, write $H := G \cap (A(X, t_0) \times Y^2)$. Using Minkowski's inequality, we have:

$$\begin{aligned} &\leq \left(\int_{(X \times Y)^2} \mathbb{1}_H^p |\omega_X(x, x') - \omega_Y(y, y')|^p d\mu^{\otimes 2}(x, y, x', y') \right)^{1/p} \dots \\ &\quad + \left(\int_{(X \times Y)^2} \mathbb{1}_H^p |\omega_Y(y, y')|^p d\mu^{\otimes 2}(x, y, x', y') \right)^{1/p}. \end{aligned}$$

For any $(x, y, x', y') \in H$, we have $|\omega_X(x, x') - \omega_Y(y, y')| < \varepsilon$. Also we have $\omega_Y(y, y') < \varepsilon + \omega_X(x, x') \leq \varepsilon + t_0$. From the latter, we know $G \cap (A(X, t_0) \times Y^2) \subseteq X^2 \times A(Y, t_0 + \varepsilon)$. So we continue the previous expression as below:

$$\begin{aligned} &\leq \left(\int_{(X \times Y)^2} \mathbb{1}_H^p |\varepsilon|^p d\mu^{\otimes 2} \right)^{1/p} + \left(\int_{X^2 \times A(Y, t_0 + \varepsilon)} |\omega_Y(y, y')|^p d\mu^{\otimes 2}(x, y, x', y') \right)^{1/p} \\ &\leq \varepsilon + \left(\int_{A(Y, t_0 + \varepsilon)} |\omega_Y(y, y')|^p d\mu_Y^{\otimes 2}(y, y') \right)^{1/p} \\ &= \text{sub}_{p,t_0+\varepsilon}^w(Y) + \varepsilon. \end{aligned} \tag{2.4}$$

Analogously, we have

$$\text{sub}_{p,t_0+\varepsilon}^w(Y) \leq \text{sub}_{p,t_0+2\varepsilon}^w(X) + \varepsilon.$$

This yields interleaving for $p \in [1, \infty)$. For $p = \infty$, we use the same arguments about G and B to obtain:

$$\begin{aligned}\text{sub}_{p,t_0}^w(X) &= \sup\{|\omega_X(x, x')| : x, x' \in \text{supp}(\mu_X), \omega_X(x, x') \leq t_0\} \\ &\leq \sup\{|\omega_Y(y, y')| + \varepsilon : y, y' \in \text{supp}(\mu_Y), \omega_Y(y, y') \leq t_0 + \varepsilon\} \\ &\leq \text{sub}_{p,t_0+\varepsilon}^w(Y) + \varepsilon.\end{aligned}$$

Thus we have interleaving for all $p \in [1, \infty]$.

The case for the \sup_p^w invariant is similar, except in step (2.4) above. In this case, we note that for any $(x, y, x', y') \in H$, we have $\omega_Y(y, y') > \omega_X(x, x') - \varepsilon \geq t_0 - \varepsilon$. Thus we have $H = G \cap (A(X, t_0) \times Y^2) \subseteq X^2 \times A(Y, t_0 - \varepsilon)$, and so:

$$\left(\int_{(X \times Y)^2} \mathbb{1}_H^p |\omega_Y(y, y')|^p d\mu^{\otimes 2}(x, y, x', y') \right)^{1/p} \leq \left(\int_{X^2 \times A(Y, t_0 - \varepsilon)} |\omega_Y(y, y')|^p d\mu^{\otimes 2}(x, y, x', y') \right)^{1/p}.$$

It follows that we have:

$$\text{sup}_{p,t_0}^w(X) \leq \text{sup}_{p,t_0-\varepsilon}^w(X) + \varepsilon \leq \text{sup}_{p,t_0-2\varepsilon}^w(Y) + 2\varepsilon.$$

The $p = \infty$ is proved analogously. \square

Proof of Theorem 121. Let $(X, \omega_X, \mu_X), (Y, \omega_Y, \mu_Y) \in \mathcal{N}_m$. For each $s \in X$ and $t \in Y$, define $\varphi_{XY}^{st} : (X \times Y)^2 \rightarrow \mathbb{R}$ by writing $\varphi_{XY}^{st}(x, y, x', y') := \omega_X(s, x')$, and define $\psi_{XY}^{st} : (X \times Y)^2 \rightarrow \mathbb{R}$ by writing $\psi_{XY}^{st}(x, y, x', y') := \omega_Y(t, y')$. For convenience, we write $\mathcal{C} := \mathcal{C}(\mu_X, \mu_Y)$.

First let $p \in [1, \infty)$. Let $\eta > d_{\mathcal{N}, p}(X, Y)$, and let $\mu \in \mathcal{C}(\mu_X, \mu_Y)$ be a coupling such that $\text{dis}_p(\mu) < 2\eta$. Then by applying Minkowski's inequality, we obtain

$$|\|\varphi_{XY}^{st}\|_{L^p(\mu \otimes \mu)} - \|\psi_{XY}^{st}\|_{L^p(\mu \otimes \mu)}| \leq \|\varphi_{XY}^{st} - \psi_{XY}^{st}\|_{L^p(\mu \otimes \mu)}. \quad (2.5)$$

In particular, because $x \mapsto x^p$ is increasing on \mathbb{R}_+ , we also have

$$|\|\varphi_{XY}^{st}\|_{L^p(\mu \otimes \mu)} - \|\psi_{XY}^{st}\|_{L^p(\mu \otimes \mu)}|^p \leq \|\varphi_{XY}^{st} - \psi_{XY}^{st}\|_{L^p(\mu \otimes \mu)}^p. \quad (2.6)$$

Next we observe:

$$\begin{aligned}\|\varphi_{XY}^{st}\|_{L^p(\mu \otimes \mu)} &= \left(\int_{X \times Y} \int_{X \times Y} |\varphi_{XY}^{st}(x, y, x', y')|^p d\mu(x, y) d\mu(x', y') \right)^{1/p} \\ &= \left(\int_{X \times Y} \int_{X \times Y} |\omega_X(s, x')|^p d\mu(x, y) d\mu(x', y') \right)^{1/p} \\ &= \left(\int_X |\omega_X(s, x')|^p d\mu_X(x') \right)^{1/p} = \text{ecc}_{p,X}^{\text{out}}(s).\end{aligned}$$

Similarly, $\|\psi_{XY}^{st}\|_{L^p(\mu \otimes \mu)} = \text{ecc}_{p,Y}^{\text{out}}(t)$.

For the right side of Inequality (2.6), we have:

$$\begin{aligned}
\|\varphi_{XY}^{st} - \psi_{XY}^{st}\|_{L^p(\mu \otimes \mu)}^p &= \int_{X \times Y} \int_{X \times Y} |\varphi_{XY}^{st}(x, y, x', y') - \psi_{XY}^{st}(x, y, x', y')|^p d\mu(x, y) d\mu(x', y') \\
&= \int_{X \times Y} \int_{X \times Y} |\omega_X(s, x') - \omega_Y(t, y')|^p d\mu(x, y) d\mu(x', y') \\
&= \int_{X \times Y} |\omega_X(s, x') - \omega_Y(t, y')|^p d\mu(x', y').
\end{aligned}$$

Putting all these observations together with Inequality (2.6), we have:

$$|\text{ecc}_{p,X}^{\text{out}}(s) - \text{ecc}_{p,Y}^{\text{out}}(t)|^p \leq \int_{X \times Y} |\omega_X(s, x') - \omega_Y(t, y')|^p d\mu(x', y').$$

The left hand side above is independent of the coupling μ , so we can infimize over $\mathcal{C}(\mu_X, \mu_Y)$:

$$|\text{ecc}_{p,X}^{\text{out}}(s) - \text{ecc}_{p,Y}^{\text{out}}(t)|^p \leq \inf_{\nu \in \mathcal{C}} \int_{X \times Y} |\omega_X(s, x') - \omega_Y(t, y')|^p d\nu(x', y') = (\text{ecc}_{p,X,Y}^{\text{out}}(s, t))^p.$$

Also observe:

$$\begin{aligned}
&\left(\int_{X \times Y} \|\varphi_{XY}^{st} - \psi_{XY}^{st}\|_{L^p(\mu \otimes \mu)}^p d\mu(s, t) \right)^{1/p} \\
&= \left(\int_{X \times Y} \int_{X \times Y} |\omega_X(s, x') - \omega_Y(t, y')|^p d\mu(x', y') d\mu(s, t) \right)^{1/p} \\
&= \text{dis}_p(\mu) < 2\eta.
\end{aligned}$$

Thus we obtain:

$$\left(\int_{X \times Y} |\text{ecc}_{p,X}^{\text{out}}(s) - \text{ecc}_{p,Y}^{\text{out}}(t)|^p d\mu(s, t) \right)^{1/p} \leq \left(\int_{X \times Y} (\text{ecc}_{p,X,Y}^{\text{out}}(s, t))^p d\mu(s, t) \right)^{1/p} < 2\eta.$$

Since $\eta > 2d_{\mathcal{N},p}(X, Y)$ was arbitrary, it follows that

$$\begin{aligned}
&\left(\inf_{\mu \in \mathcal{C}} \int_{X \times Y} |\text{ecc}_{p,X}^{\text{out}}(s) - \text{ecc}_{p,Y}^{\text{out}}(t)|^p d\mu(s, t) \right)^{1/p} \leq \left(\inf_{\mu \in \mathcal{C}} \int_{X \times Y} (\text{ecc}_{p,X,Y}^{\text{out}}(s, t))^p d\mu(s, t) \right)^{1/p} \\
&\quad \leq 2d_{\mathcal{N},p}(X, Y).
\end{aligned} \tag{2.7}$$

This proves the $p \in [1, \infty)$ case. The $p = \infty$ case follows by applying Minkowski's inequality to obtain Inequality (2.5), and working analogously from there. Finally, we remark that the same proof holds for the $\text{ecc}_{p,X}^{\text{in}}$ and $\text{ecc}_{p,X,Y}^{\text{in}}$ functions. \square

The following lemma is a particular statement of the change of variables theorem that we use later.

Lemma 151 (Change of variables). *Let $(X, \mathcal{F}_X, \mu_X)$ and $(Y, \mathcal{F}_Y, \mu_Y)$ be two probability spaces. Let $f : X \rightarrow \mathbb{R}$ and $g : Y \rightarrow \mathbb{R}$ be two measurable functions. Write $f_*\mu_X$ and $g_*\mu_Y$ to denote the pushforward distributions on \mathbb{R} . Let $T : X \times Y \rightarrow \mathbb{R}^2$ be the map $(x, y) \mapsto (f(x), g(y))$ and let $h : \mathbb{R}^2 \rightarrow \mathbb{R}_+$ be measurable. Next let $\mu \in \mathcal{C}(\mu_X, \mu_Y)$. Then $T_*\mu \in \mathcal{C}(f_*\mu_X, g_*\mu_Y)$, and the following inequality holds:*

$$\left(\inf_{\nu \in \mathcal{C}(f_*\mu_X, g_*\mu_Y)} \int_{\mathbb{R}^2} h(a, b) d(T_*\mu)(a, b) \right)^{1/p} \leq \left(\int_{X \times Y} h(T(x, y)) d\mu(x, y) \right)^{1/p}.$$

This is essentially the same as [86, Lemma 6.1] but stated for general probability spaces instead of metric measure spaces. The form of the statement in [86, Lemma 6.1] is slightly different, but it can be obtained from the statement presented above by using [120, Remark 2.19].

Proof of Lemma 151. First we check that $T_*\mu \in \mathcal{C}(f_*\mu_X, g_*\mu_Y)$. Let $A \in \text{Borel}(\mathbb{R})$. Then,

$$\begin{aligned} T_*\mu(A \times \mathbb{R}) &= \mu(\{(x, y) \in X \times Y : T(x, y) \in A \times \mathbb{R}\}) = \mu(\{(x, y) \in X \times Y : f(x) \in A\}) \\ &= f_*\mu_X(A). \end{aligned}$$

Similarly we check $T_*\mu(\mathbb{R} \times A) = g_*\mu_Y(A)$.

Next we check the inequality. By the change of variables formula, we have:

$$\left(\int_{\mathbb{R}^2} h(a, b) d(T_*\mu)(a, b) \right)^{1/p} = \left(\int_{X \times Y} h(T(x, y)) d\mu(x, y) \right)^{1/p}.$$

We have already verified that $T_*\mu \in \mathcal{C}(f_*\mu_X, g_*\mu_Y)$. The inequality is obtained by infimizing the left hand side over all possible couplings $\nu \in \mathcal{C}(f_*\mu_X, g_*\mu_Y)$. This does not affect the right hand side, which is independent of such couplings. \square

Proof of Theorem 127. Consider the probability spaces $X \times X$ and $Y \times Y$, equipped with the product measures $\mu_X \otimes \mu_X$ and $\mu_Y \otimes \mu_Y$. For convenience, we define the shorthand notation $\nu_X := (\omega_X)_*(\mu_X \otimes \mu_X)$ and $\nu_Y := (\omega_Y)_*(\mu_Y \otimes \mu_Y)$. Let $T : (X \times X) \times (Y \times Y) \rightarrow \mathbb{R}^2$ be the map $(x, x', y, y') \mapsto (\omega_X(x, x'), \omega_Y(y, y'))$. Also let $h : \mathbb{R}^2 \rightarrow \mathbb{R}$ be the map $(a, b) \mapsto |a - b|^p$.

Let $\eta > d_{\mathcal{N}, p}(X, Y)$, and let $\mu \in \mathcal{C}(\mu_X, \mu_Y)$ be a coupling such that $\text{dis}_p(\mu) < 2\eta$. Also let τ be a measure on $X \times X \times Y \times Y$ defined by writing $\tau(A, A', B, B') := \mu(A, B)\mu(A', B')$ for $A, A' \in \text{Borel}(X)$ and $B, B' \in \text{Borel}(Y)$. Then $\tau \in \mathcal{C}(\mu_X^{\otimes 2}, \mu_Y^{\otimes 2})$.

By Lemma 151, we know that $T_*\tau \in \mathcal{C}(\nu_X, \nu_Y)$. By the change of variables formula and Fubini's theorem,

$$\begin{aligned} \left(\int_{\mathbb{R}^2} |a - b|^p d(T_*\tau)(a, b) \right)^{1/p} &= \left(\int_{X^2 \times Y^2} |\omega_X(x, x') - \omega_Y(y, y')|^p d\tau(x, x', y, y') \right)^{1/p} \\ &= \left(\int_{X^2 \times Y^2} |\omega_X(x, x') - \omega_Y(y, y')|^p d(\mu(x, y)\mu(x', y')) \right)^{1/p} \\ &= \left(\int_{X \times Y} \int_{X \times Y} |\omega_X(x, x') - \omega_Y(y, y')|^p d\mu(x, y) d\mu(x', y') \right)^{1/p} \\ &< 2\eta. \end{aligned}$$

We infimize over $\mathcal{C}(\mu_X^{\otimes 2}, \mu_Y^{\otimes 2})$, use the fact that $\eta > d_{\mathcal{N},p}(X, Y)$ was arbitrary, and apply Lemma 151 to obtain:

$$\begin{aligned} 2d_{\mathcal{N},p}(X, Y) &\geq \left(\inf_{\mu \in \mathcal{C}(\mu_X^{\otimes 2}, \mu_Y^{\otimes 2})} \int_{X^2 \times Y^2} |\omega_X(x, x') - \omega_Y(y, y')|^p d\mu(x, x', y, y') \right)^{1/p} \\ &\geq \inf_{\gamma \in \mathcal{C}(\nu_X, \nu_Y)} \left(\int_{\mathbb{R}^2} |a - b|^p d\gamma(a, b) \right)^{1/p}. \end{aligned}$$

This yields Inequalities (1.7)-(1.8).

Next we consider the distributions induced by the ecc^{out} function. For convenience, write $e_X := (\text{ecc}_{p,X}^{\text{out}})_*\mu_X$ and $e_Y := (\text{ecc}_{p,Y}^{\text{out}})_*\mu_Y$. Now let $T : X \times Y \rightarrow \mathbb{R}$ be the map $(x, y) \mapsto (\text{ecc}_{p,X}^{\text{out}}(x), \text{ecc}_{p,Y}^{\text{out}}(y))$, and let $h : \mathbb{R}^2 \rightarrow \mathbb{R}$ be the map $(a, b) \mapsto |a - b|^p$. By the change of variables formula and Theorem 121, we know

$$\begin{aligned} \inf_{\mu \in \mathcal{C}(\mu_X, \mu_Y)} \left(\int_{\mathbb{R}^2} |a - b|^p d(T_*\mu)(a, b) \right)^{1/p} &= \inf_{\mu \in \mathcal{C}(\mu_X, \mu_Y)} \left(\int_{X \times Y} |\text{ecc}_{p,X}^{\text{out}}(x) - \text{ecc}_{p,X}^{\text{out}}(y)|^p d\mu(x, y) \right)^{1/p} \\ &\leq 2d_{\mathcal{N},p}(X, Y). \end{aligned}$$

By Lemma 151, we know that $T_*\mu \in \mathcal{C}(e_X, e_Y)$ and also the following:

$$\begin{aligned} \inf_{\gamma \in \mathcal{C}(e_X, e_Y)} \left(\int_{\mathbb{R}^2} |a - b|^p d\gamma(a, b) \right)^{1/p} &\leq \inf_{\mu \in \mathcal{C}(\mu_X, \mu_Y)} \left(\int_{X \times Y} |\text{ecc}_{p,X}^{\text{out}}(x) - \text{ecc}_{p,X}^{\text{out}}(y)|^p d\mu(x, y) \right)^{1/p} \\ &\leq 2d_{\mathcal{N},p}(X, Y). \end{aligned}$$

This proves Inequalities (1.9)-(1.10). Inequalities (1.11)-(1.12) are proved analogously.

Finally we consider the distributions obtained as pushforwards of the joint eccentricity function, i.e. Inequalities (1.13)-(1.16). For each $x \in X$ and $y \in Y$ let $T^{xy} : X \times Y \rightarrow \mathbb{R}$ be the map $(x', y') \mapsto (\omega_X(x, x'), \omega_Y(y, y'))$, and let $h : \mathbb{R}^2 \rightarrow \mathbb{R}$ be the map $(a, b) \mapsto |a - b|^p$. Let $\gamma \in \mathcal{C}(\mu_X, \mu_Y)$. By the change of variables formula, we have

$$\begin{aligned} \int_{\mathbb{R}^2} |a - b|^p d(T_*^{xy}\gamma)(a, b) &= \int_{X \times Y} |\omega_X(x, x') - \omega_Y(y, y')|^p d\gamma(x', y'), \text{ and so} \\ \int_{X \times Y} \int_{\mathbb{R}^2} |a - b|^p d(T_*^{xy}\gamma)(a, b) d\mu(x, y) &= \int_{X \times Y} \int_{X \times Y} |\omega_X(x, x') - \omega_Y(y, y')|^p d\gamma(x', y') d\mu(x, y). \end{aligned}$$

By Lemma 151, $T_*^{xy}\mu \in \mathcal{C}(\lambda_X(x), \lambda_Y(y))$. Applying Theorem 121 and Lemma 151, we have:

$$\begin{aligned} 2d_{\mathcal{N},p}(X, Y) &\geq \inf_{\mu \in \mathcal{C}(\mu_X, \mu_Y)} \left(\int_{X \times Y} \inf_{\gamma \in \mathcal{C}(\mu_X, \mu_Y)} \int_{X \times Y} |\omega_X(x, x') - \omega_Y(y, y')|^p d\gamma(x', y') d\mu(x, y) \right)^{1/p} \\ &\geq \inf_{\mu \in \mathcal{C}(\mu_X, \mu_Y)} \left(\int_{X \times Y} \inf_{\gamma \in \mathcal{C}(\lambda_X(x), \lambda_Y(y))} \int_{\mathbb{R}^2} |a - b|^p d\gamma(a, b) d\mu(x, y) \right)^{1/p}. \end{aligned}$$

This verifies Inequalities (1.13)-(1.14). Inequalities (1.15)-(1.16) are proved analogously. \square

Chapter 3: Persistent Homology on Networks

In this chapter, we give proofs and auxiliary results related to persistent homology.

3.1 Background on persistence and interleaving

3.1.1 Homology, persistence, and tameness

We assume that the reader is familiar with terms and concepts related to simplicial homology, and refer to [88] for details. Here we describe our choices of notation. Whenever we have a simplicial complex over a set X and a k -simplex $\{x_0, x_1, \dots, x_k\}$, $k \in \mathbb{Z}_+$, we will assume that the simplex is *oriented* by the ordering $x_0 < x_1 < \dots < x_k$. We will write $[x_0, x_1, \dots, x_k]$ to denote the equivalence class of the even permutations of this chosen ordering, and $-[x_0, x_1, \dots, x_k]$ to denote the equivalence class of the odd permutations of this ordering. Given a simplicial complex Σ , we will denote its geometric realization by $|\Sigma|$. The *weak topology* on $|\Sigma|$ is defined by requiring that a subset $A \subseteq |\Sigma|$ is closed if and only if $A \cap |\sigma|$ is closed in $|\sigma|$ for each $\sigma \in \Sigma$. A simplicial map $f : \Sigma \rightarrow \Xi$ between two simplicial complexes induces a map $|f| : |\Sigma| \rightarrow |\Xi|$ between the geometric realizations, defined as $|f|(\sum_{v \in \Sigma} a_v v) := \sum_{v \in \Sigma} a_v f(v)$. These induced maps satisfy the usual composition identity: given simplicial maps $f : \Sigma \rightarrow \Xi$ and $g : \Xi \rightarrow \Upsilon$, we have $|g \circ f| = |g| \circ |f|$. To see this, observe the following:

$$|g \circ f|(\sum_{v \in \Sigma} a_v v) = \sum_{v \in \Sigma} a_v g(f(v)) = |g|(\sum_{v \in \Sigma} a_v f(v)) = |g| \circ |f|(\sum_{v \in \Sigma} a_v v). \quad (3.1)$$

A *filtration* of a simplicial complex Σ (also called a *filtered simplicial complex*) is defined to be a nested sequence $\{\Sigma^\delta \subseteq \Sigma^{\delta'}\}_{\delta \leq \delta' \in \mathbb{R}}$ of simplicial complexes satisfying the condition that there exist $\delta_I, \delta_F \in \mathbb{R}$ such that $\Sigma^\delta = \emptyset$ for all $\delta \leq \delta_I$, and $\Sigma^\delta = \Sigma$ for all $\delta \geq \delta_F$.

Fix a field \mathbb{K} . Given a finite simplicial complex Σ and a dimension $k \in \mathbb{Z}_+$, we will denote a k -chain in Σ as $\sum_i a_i \sigma_i$, where each $a_i \in \mathbb{K}$ and each $\sigma_i \in \Sigma$ is a k -simplex. We write $C_k(\Sigma)$ or just C_k to denote the \mathbb{K} -vector space of all k -chains. We will write ∂_k to denote the associated *boundary map* $\partial_k : C_k \rightarrow C_{k-1}$:

$$\partial_k[x_0, \dots, x_k] := \sum_i (-1)^i [x_0, \dots, \hat{x}_i, \dots, x_k],$$

where \hat{x}_i denotes omission of x_i from the sequence.

We will write $\mathcal{C} = (C_k, \partial_k)_{k \in \mathbb{Z}_+}$ to denote a *chain complex*, i.e. a sequence of vector spaces with boundary maps such that $\partial_{k-1} \circ \partial_k = 0$. Given a chain complex \mathcal{C} and any $k \in \mathbb{Z}_+$, the k -th homology of the chain complex \mathcal{C} is denoted $H_k(\mathcal{C}) := \ker(\partial_k) / \text{im}(\partial_{k+1})$. The k -th *Betti number* of \mathcal{C} is denoted $\beta_k(\mathcal{C}) := \dim(H_k(\mathcal{C}))$.

Given a simplicial map f between simplicial complexes, we write f_* to denote the induced chain map between the corresponding chain complexes [88, §1.12], and $(f_k)_\#$ to denote the linear map on k th homology vector spaces induced for each $k \in \mathbb{Z}_+$.

The operations of passing from simplicial complexes and simplicial maps to chain complexes and induced chain maps, and then to homology vector spaces with induced linear maps, will be referred to as *passing to homology*. Recall the following useful fact, often referred to as *functoriality of homology* [88, Theorem 12.2]: given a composition $g \circ f$ of simplicial maps, we have

$$(g_k \circ f_k)_\# = (g_k)_\# \circ (f_k)_\# \quad \text{for each } k \in \mathbb{Z}_+. \quad (3.2)$$

A *persistence vector space* is defined to be a family of vector spaces $\{U^\delta \xrightarrow{\mu_{\delta,\delta'}} U^{\delta'}\}_{\delta \leq \delta' \in \mathbb{R}}$ such that: (1) $\mu_{\delta,\delta}$ is the identity for each $\delta \in \mathbb{R}$, and (2) $\mu_{\delta,\delta''} = \mu_{\delta',\delta''} \circ \mu_{\delta,\delta'}$ for each $\delta \leq \delta' \leq \delta'' \in \mathbb{R}$. The persistence vector spaces that we consider in this work also satisfy the following conditions: (1) $\dim(U^\delta) < \infty$ at each $\delta \in \mathbb{R}$, (2) there exist $\delta_I, \delta_F \in \mathbb{R}$ such that all maps $\mu_{\delta,\delta'}$ are isomorphisms for $\delta, \delta' \geq \delta_F$ and for $\delta, \delta' \leq \delta_I$, and (3) there are only finitely many values of $\delta \in \mathbb{R}$ such that $U^{\delta-\varepsilon} \not\cong U^\delta$ for each $\varepsilon > 0$. Here δ is referred to as a *resolution parameter*, and such a persistence vector space is described as being \mathbb{R} -indexed. The collection of all such persistence vector spaces is denoted $\mathbf{PVec}(\mathbb{R})$. Observe that by fixing $k \in \mathbb{Z}_+$ and passing to the k th homology vector space at each step Σ^δ of a filtered simplicial complex $(\Sigma^\delta)_{\delta \in \mathbb{R}}$, the functoriality of homology gives us the k th persistence vector space associated to $(\Sigma^\delta)_{\delta \in \mathbb{R}}$, denoted

$$\mathcal{H}_k(\Sigma) := \{H_k(\mathcal{C}^\delta) \xrightarrow{(\iota_{\delta,\delta'})_\#} H_k(\mathcal{C}^{\delta'})\}_{\delta \leq \delta' \in \mathbb{R}}.$$

The elements of $\mathbf{PVec}(\mathbb{R})$ contain only a finite number of vector spaces, up to isomorphism. By the classification results in [21, §5.2], it is possible to associate a full invariant, called a *persistence barcode* or *persistence diagram*, to each element of $\mathbf{PVec}(\mathbb{R})$. This barcode is a multiset of *persistence intervals*, and is represented as a set of lines over a single axis. The barcode of a persistence vector space \mathcal{V} is denoted $\mathbf{Pers}(\mathcal{V})$. The intervals in $\mathbf{Pers}(\mathcal{V})$ can be represented as the *persistence diagram* of \mathcal{V} , which is as a multiset of points lying on or above the diagonal in $\overline{\mathbb{R}}^2$, counted with multiplicity. More specifically,

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

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

Persistence diagrams can be compared using the *bottleneck distance*, which we denote by d_B . Details about this distance, as well as the other material related to persistent homology, can be found in [26]. Numerous other formulations of the material presented above can be found in [49, 123, 19, 47, 50, 9, 48].

Remark 152. Whenever we describe a persistence diagram as being *trivial*, we mean that it does not have any off-diagonal points.

3.1.2 Interleaving distance and stability of persistence vector spaces.

In what follows, we will consider \mathbb{R} -indexed persistence vector spaces $\mathbf{PVec}(\mathbb{R})$.

Given $\varepsilon \geq 0$, two \mathbb{R} -indexed persistence vector spaces $\mathcal{V} = \{V^\delta \xrightarrow{\nu_{\delta,\delta'}} V^{\delta'}\}_{\delta \leq \delta'}$ and $\mathcal{U} = \{U^\delta \xrightarrow{\mu_{\delta,\delta'}} U^{\delta'}\}_{\delta \leq \delta'}$ are said to be ε -interleaved [24, 9] if there exist two families of linear maps

$$\begin{aligned} &\{\varphi_{\delta,\delta+\varepsilon} : V^\delta \rightarrow U^{\delta+\varepsilon}\}_{\delta \in \mathbb{R}}, \\ &\{\psi_{\delta,\delta+\varepsilon} : U^\delta \rightarrow V^{\delta+\varepsilon}\}_{\delta \in \mathbb{R}} \end{aligned}$$

such that the following diagrams commute for all $\delta \leq \delta' \in \mathbb{R}$:

$$\begin{array}{ccc} \begin{array}{ccc} V^\delta & \xrightarrow{\nu_{\delta,\delta'}} & V^{\delta'} \\ & \searrow \varphi_\delta & \swarrow \varphi_{\delta'} \\ & U^{\delta+\varepsilon} & \xrightarrow{\mu_{\delta+\varepsilon,\delta'+\varepsilon}} U^{\delta'+\varepsilon} \end{array} & \quad & \begin{array}{ccc} V^{\delta+\varepsilon} & \xrightarrow{\nu_{\delta+\varepsilon,\delta'+\varepsilon}} & V^{\delta'+\varepsilon} \\ \psi_\delta \nearrow & & \swarrow \psi_{\delta'} \\ U^\delta & \xrightarrow{\mu_{\delta,\delta'}} & U^{\delta'} \end{array} \\ \\ \begin{array}{ccc} V^\delta & \xrightarrow{\nu_{\delta,\delta+2\varepsilon}} & V^{\delta+2\varepsilon} \\ & \searrow \varphi_\delta & \swarrow \psi_{\delta+\varepsilon} \\ & U^{\delta+\varepsilon} & \end{array} & \quad & \begin{array}{ccc} V^{\delta+\varepsilon} & & \\ \psi_\delta \nearrow & & \swarrow \varphi_{\delta+\varepsilon} \\ U^\delta & \xrightarrow{\mu_{\delta,\delta+2\varepsilon}} & U^{\delta+2\varepsilon} \end{array} \end{array}$$

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

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

This definition induces an extended pseudometric on the collection of persistence vector spaces [24, 9, 26].

Before stating the following lemma, recall that two simplicial maps $f, g : \Sigma \rightarrow \Xi$ are *contiguous* if for any simplex $\sigma \in \Sigma$, $f(\sigma) \cup g(\sigma)$ is a simplex of Ξ . Contiguous maps satisfy the following useful properties:

Proposition 153 (Properties of contiguous maps). *Let $f, g : \Sigma \rightarrow \Xi$ be two contiguous simplicial maps. Then,*

1. $|f|, |g| : |\Sigma| \rightarrow |\Xi|$ are homotopic [112, §3.5], and
2. The chain maps induced by f and g are chain homotopic, and as a result, the induced maps $f_{\#}$ and $g_{\#}$ for homology are equal [88, Theorem 12.5].

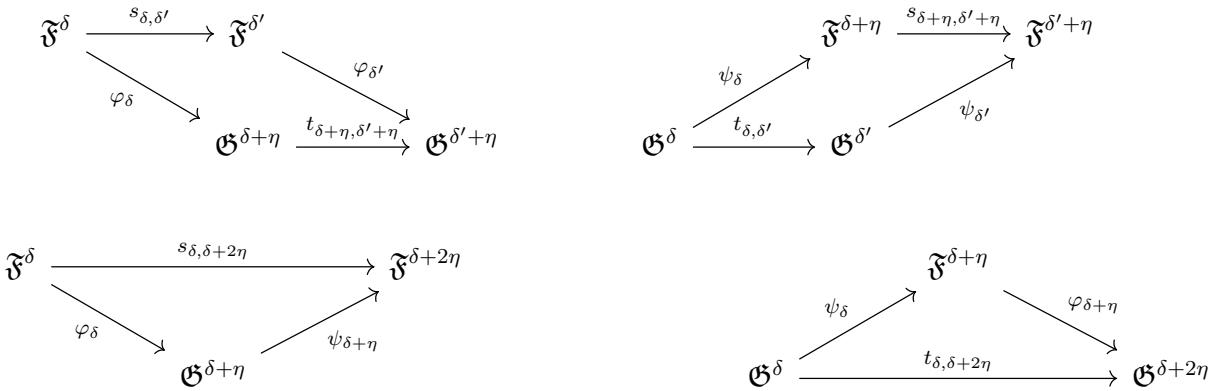
Lemma 154 (Stability Lemma, [45]). *Let $\mathfrak{F}, \mathfrak{G}$ be two filtered simplicial complexes written as*

$$\{\mathfrak{F}^{\delta} \xrightarrow{s_{\delta, \delta'}} \mathfrak{F}^{\delta'}\}_{\delta' \geq \delta \in \mathbb{R}} \text{ and } \{\mathfrak{G}^{\delta} \xrightarrow{t_{\delta, \delta'}} \mathfrak{G}^{\delta'}\}_{\delta' \geq \delta \in \mathbb{R}},$$

where $s_{\delta, \delta'}$ and $t_{\delta, \delta'}$ denote the natural inclusion maps. Suppose $\eta \geq 0$ is such that there exist families of simplicial maps $\{\varphi_{\delta} : \mathfrak{F}^{\delta} \rightarrow \mathfrak{G}^{\delta+\eta}\}_{\delta \in \mathbb{R}}$ and $\{\psi_{\delta} : \mathfrak{G}^{\delta} \rightarrow \mathfrak{F}^{\delta+\eta}\}_{\delta \in \mathbb{R}}$ such that the following are satisfied for any $\delta \leq \delta'$:

1. $t_{\delta+\eta, \delta'+\eta} \circ \varphi_{\delta}$ and $\varphi_{\delta'} \circ s_{\delta, \delta'}$ are contiguous
2. $s_{\delta+\eta, \delta'+\eta} \circ \psi_{\delta}$ and $\psi_{\delta'} \circ t_{\delta, \delta'}$ are contiguous
3. $\psi_{\delta+\eta} \circ \varphi_{\delta}$ and $s_{\delta, \delta+2\eta}$ are contiguous
4. $\varphi_{\delta+\eta} \circ \psi_{\delta}$ and $t_{\delta, \delta+2\eta}$ are contiguous.

All the diagrams are as below:



For each $k \in \mathbb{Z}_+$, let $\mathbf{PVec}_k(\mathfrak{F})$, $\mathbf{PVec}_k(\mathfrak{G})$ denote the k -dimensional persistent vector spaces associated to \mathfrak{F} and \mathfrak{G} . Then for each $k \in \mathbb{Z}_+$,

$$d_1(\mathbf{PVec}_k(\mathfrak{F}), \mathbf{PVec}_k(\mathfrak{G})) \leq \eta.$$

3.2 Simplicial constructions

3.2.1 Stability of Vietoris-Rips and Dowker constructions

Proof of Proposition 29. Both cases are similar, so we just prove the result for PVec^{si} . Let $\eta > 2d_N(X, Y)$. Then by Proposition 7, there exist maps $\varphi : X \rightarrow Y, \psi : Y \rightarrow X$ such that

$$\max(\text{dis}(\varphi), \text{dis}(\psi), C_{X,Y}(\varphi, \psi), C_{Y,X}(\psi, \varphi)) < \eta.$$

First we check that φ, ψ induce simplicial maps $\varphi_\delta : \mathfrak{D}_{\delta, X}^{\text{si}} \rightarrow \mathfrak{D}_{\delta+\eta, Y}^{\text{si}}$ and $\psi_\delta : \mathfrak{D}_{\delta, Y}^{\text{si}} \rightarrow \mathfrak{D}_{\delta+\eta, X}^{\text{si}}$ for each $\delta \in \mathbb{R}$.

Let $\delta' \geq \delta \in \mathbb{R}$. Let $\sigma = [x_0, \dots, x_n] \in \mathfrak{D}_{\delta, X}^{\text{si}}$. Then there exists $x' \in X$ such that $\omega_X(x_i, x') \leq \delta$ for each $0 \leq i \leq n$. Fix such an x' . Since $\text{dis}(\varphi) < \eta$, we have the following for each i :

$$|\omega_X(x_i, x') - \omega_Y(\varphi(x_i), \varphi(x'))| < \eta.$$

So $\omega_Y(\varphi(x_i), \varphi(x')) < \omega_X(x_i, x') + \eta \leq \delta + \eta$ for each $0 \leq i \leq n$. Thus $\varphi_\delta(\sigma) := \{\varphi(x_0), \dots, \varphi(x_n)\}$ is a simplex in $\mathfrak{D}_{\delta+\eta, Y}^{\text{si}}$. Thus the map on simplices φ_δ induced by φ is simplicial for each $\delta \in \mathbb{R}$.

Similarly we check that the map ψ_δ on simplices induced by ψ is simplicial. Now to prove the result, it will suffice to check the contiguity conditions in the statement of Lemma 154. Consider the following diagram:

$$\begin{array}{ccccc} \mathfrak{D}_{\delta, X}^{\text{si}} & \xrightarrow{s_{\delta, \delta'}} & \mathfrak{D}_{\delta', X}^{\text{si}} & & \\ & \searrow \varphi_\delta & \swarrow \varphi_{\delta'} & & \\ & & \mathfrak{D}_{\delta+\eta, Y}^{\text{si}} & \xrightarrow{t_{\delta+\eta, \delta'+\eta}} & \mathfrak{D}_{\delta'+\eta, Y}^{\text{si}} \end{array}$$

Here $s_{\delta, \delta'}$ and $t_{\delta+\eta, \delta'+\eta}$ are the inclusion maps. We claim that $t_{\delta+\eta, \delta'+\eta} \circ \varphi_\delta$ and $\varphi_{\delta'} \circ s_{\delta, \delta'}$ are contiguous simplicial maps. To see this, let $\sigma \in \mathfrak{D}_{\delta, X}^{\text{si}}$. Since $s_{\delta, \delta'}$ is just the inclusion, it follows that $t_{\delta+\eta, \delta'+\eta}(\varphi_\delta(\sigma)) \cup \varphi_{\delta'}(s_{\delta, \delta'}(\sigma)) = \varphi_\delta(\sigma)$, which is a simplex in $\mathfrak{D}_{\delta+\eta, Y}^{\text{si}}$ because φ_δ is simplicial, and hence a simplex in $\mathfrak{D}_{\delta'+\eta, Y}^{\text{si}}$ because the inclusion $t_{\delta+\eta, \delta'+\eta}$ is simplicial. Thus $t_{\delta+\eta, \delta'+\eta} \circ \varphi_\delta$ and $\varphi_{\delta'} \circ s_{\delta, \delta'}$ are contiguous, and their induced linear maps for homology are equal. By a similar argument, we verify that $s_{\delta+\eta, \delta'+\eta} \circ \psi_\delta$ and $\psi_{\delta'} \circ t_{\delta, \delta'}$ are contiguous simplicial maps as well.

Next we check that the maps $\psi_{\delta+\eta} \circ \varphi_\delta$ and $s_{\delta, \delta+2\eta}$ in the figure below are contiguous.

$$\begin{array}{ccc} \mathfrak{D}_{\delta, X}^{\text{si}} & \xrightarrow{s_{\delta, \delta+2\eta}} & \mathfrak{D}_{\delta+2\eta, X}^{\text{si}} \\ & \searrow \varphi_\delta & \nearrow \psi_{\delta+\eta} \\ & \mathfrak{D}_{\delta+\eta, Y}^{\text{si}} & \end{array}$$

Let $x_i \in \sigma$. Note that for our fixed $\sigma = [x_0, \dots, x_n] \in \mathfrak{D}_{\delta, X}^{\text{si}}$ and x' , we have:

$$\begin{aligned} |\omega_X(x_i, x') - \omega_X(\psi(\varphi(x_i)), \psi(\varphi(x')))| &\leq |\omega_X(x_i, x') - \omega_Y(\varphi(x_i), \varphi(x'))| \\ &\quad + |\omega_Y(\varphi(x_i), \varphi(x')) - \omega_X(\psi(\varphi(x_i)), \psi(\varphi(x')))| \\ &< 2\eta. \end{aligned}$$

Thus we obtain $\omega_X(\psi(\varphi(x_i)), \psi(\varphi(x'))) < \omega_X(x_i, x') + 2\eta \leq \delta + 2\eta$.

Since this holds for any $x_i \in \sigma$, it follows that $\psi_{\delta+\eta}(\varphi_\delta(\sigma)) \in \mathfrak{D}_{\delta+2\eta, X}^{\text{si}}$. We further claim that

$$\tau := \sigma \cup \psi_{\delta+\eta}(\varphi_\delta(\sigma)) = \{x_0, x_1, \dots, x_n, \psi(\varphi(x_0)), \dots, \psi(\varphi(x_n))\}$$

is a simplex in $\mathfrak{D}_{\delta+2\eta, X}^{\text{si}}$. Let $0 \leq i \leq n$. It suffices to show that $\omega_X(x_i, \psi(\varphi(x'))) \leq \delta + 2\eta$.

Notice that from the reformulation of $d_{\mathcal{N}}$ (Proposition 7), we have

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

Let $y = \varphi(x')$. Then $|\omega_X(x_i, \psi(y)) - \omega_Y(\varphi(x_i), y)| < \eta$. In particular,

$$\omega_X(x_i, \psi(\varphi(x'))) < \omega_Y(\varphi(x_i), \varphi(x')) + \eta \leq \omega_X(x_i, x') + 2\eta \leq \delta + 2\eta.$$

Since $0 \leq i \leq n$ were arbitrary, it follows that $\tau \in \mathfrak{D}_{\delta+2\eta, X}^{\text{si}}$. Thus $\psi_{\delta+\eta} \circ \varphi_\delta$ and $s_{\delta, \delta+2\eta}$ are contiguous. Similarly, we use the $\text{dis}(\psi)$ and $C_{Y, X}(\psi, \varphi)$ terms to verify that $t_{\delta, \delta+2\eta}$ and $\varphi_{\delta+\eta} \circ \psi_\delta$ are contiguous.

Since $\eta > 2d_{\mathcal{N}}(X, Y)$ was arbitrary, the result now follows by an application of Lemma 154. \square

3.2.2 The Functorial Dowker Theorem and equivalence of diagrams

Let $(X, \omega_X) \in \mathcal{CN}$, and let $\delta \in \mathbb{R}$ be such that $R_{\delta, X}$ is nonempty. By applying Dowker's theorem (Theorem 33) to the setting $Y = X$, we have $H_k(\mathfrak{D}_{\delta, X}^{\text{si}}) \cong H_k(\mathfrak{D}_{\delta, X}^{\text{so}})$, for any $k \in \mathbb{Z}_+$. We still have this equality in the case where $R_{\delta, X}$ is empty, because then $\mathfrak{D}_{\delta, X}^{\text{si}}$ and $\mathfrak{D}_{\delta, X}^{\text{so}}$ are both empty. Thus we obtain:

Corollary 155. *Let $(X, \omega_X) \in \mathcal{CN}$, $\delta \in \mathbb{R}$, and $k \in \mathbb{Z}_+$. Then,*

$$H_k(\mathfrak{D}_{\delta, X}^{\text{si}}) \cong H_k(\mathfrak{D}_{\delta, X}^{\text{so}}).$$

In the persistent setting, Theorem 33 and Corollary 155 suggest the following question:

Given a network (X, ω_X) and a fixed dimension $k \in \mathbb{Z}_+$, are the persistence diagrams of the Dowker sink and source filtrations of (X, ω_X) necessarily equal?

In what follows, we provide a positive answer to the question above. Our strategy is to use the Functorial Dowker Theorem (Theorem 35), for which we will provide a complete proof below. The Functorial Dowker Theorem implies equality between sink and source persistence diagrams.

Corollary 156 (Dowker duality). *Let $(X, \omega_X) \in \mathcal{FN}$, and let $k \in \mathbb{Z}_+$. Then,*

$$\mathrm{Dgm}_k^{\mathrm{si}}(X) = \mathrm{Dgm}_k^{\mathrm{so}}(X).$$

Thus we may call either of the diagrams above the k -dimensional Dowker diagram of X , denoted $\mathrm{Dgm}_k^{\mathfrak{D}}(X)$.

Before proving the corollary, we state an \mathbb{R} -indexed variant of the Persistence Equivalence Theorem [47]. This particular version follows from the *isometry theorem* [9], and we refer the reader to [26, Chapter 5] for an expanded presentation of this material.

Theorem 157 (Persistence Equivalence Theorem). *Consider two persistent vector spaces $\mathcal{U} = \{U^\delta \xrightarrow{\mu_{\delta, \delta'}} U^{\delta'}\}_{\delta \leq \delta' \in \mathbb{R}}$ and $\mathcal{V} = \{V^\delta \xrightarrow{\nu_{\delta, \delta'}} V^{\delta'}\}_{\delta \leq \delta' \in \mathbb{R}}$ with connecting maps $f_\delta : U^\delta \rightarrow V^\delta$.*

$$\begin{array}{ccccccc} \cdots & \longrightarrow & U^\delta & \longrightarrow & U^{\delta'} & \longrightarrow & U^{\delta''} \longrightarrow \cdots \\ & & \downarrow f_\delta & & \downarrow f_{\delta'} & & \downarrow f_{\delta''} \\ \cdots & \longrightarrow & V^\delta & \longrightarrow & V^{\delta'} & \longrightarrow & V^{\delta''} \longrightarrow \cdots \end{array}$$

If the f_δ are all isomorphisms and each square in the diagram above commutes, then:

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

Proof of Corollary 156. Let $\delta \leq \delta' \in \mathbb{R}$, and consider the relations $R_{\delta, X} \subseteq R_{\delta', X} \subseteq X \times X$. Suppose first that $R_{\delta, X}$ and $R_{\delta', X}$ are both nonempty. By applying Theorem 35, we obtain homotopy equivalences between the source and sink complexes that commute with the canonical inclusions up to homotopy. Passing to the k -th homology level, we obtain persistence vector spaces that satisfy the commutativity properties of Theorem 157. The result follows from Theorem 157.

In the case where $R_{\delta, X}$ and $R_{\delta', X}$ are both empty, there is nothing to show because all the associated complexes are empty. Suppose $R_{\delta, X}$ is empty, and $R_{\delta', X}$ is nonempty. Then $\mathfrak{D}_{\delta, X}^{\mathrm{si}}$ and $\mathfrak{D}_{\delta, X}^{\mathrm{so}}$ are empty, so their inclusions into $\mathfrak{D}_{\delta', X}^{\mathrm{si}}$ and $\mathfrak{D}_{\delta', X}^{\mathrm{so}}$ induce zero maps upon passing to homology. Thus the commutativity of Theorem 157 is satisfied, and the result follows by Theorem 157. \square

The proof of the Functorial Dowker Theorem It remains to prove Theorem 35. Because the proof involves numerous maps, we will adopt the notational convention of adding a subscript to a function to denote its codomain—e.g. we will write f_B to denote a function with codomain B .

First we recall the construction of a combinatorial barycentric subdivision (see [46, §2], [81, §4.7], [7, Appendix A]).

Definition 47 (Barycentric subdivisions). For any simplicial complex Σ , one may construct a new simplicial complex $\Sigma^{(1)}$, called the *first barycentric subdivision*, as follows:

$$\Sigma^{(1)} := \{[\sigma_1, \sigma_2, \dots, \sigma_p] : \sigma_1 \subseteq \sigma_2 \subseteq \dots \subseteq \sigma_p, \text{ each } \sigma_i \in \Sigma\}.$$

Note that the vertices of $\Sigma^{(1)}$ are the simplices of Σ , and the simplices of $\Sigma^{(1)}$ are nested sequences of simplices of Σ . Furthermore, note that given any two simplicial complexes Σ, Ξ and a simplicial map $f : \Sigma \rightarrow \Xi$, there is a natural simplicial map $f\mathbb{1} : \Sigma\mathbb{1} \rightarrow \Xi\mathbb{1}$ defined as:

$$f\mathbb{1}([\sigma_1, \dots, \sigma_p]) := [f(\sigma_1), \dots, f(\sigma_p)], \quad \sigma_1 \subseteq \sigma_2 \subseteq \dots, \sigma_p, \text{ each } \sigma_i \in \Sigma.$$

To see that this is simplicial, note that $f(\sigma_i) \subseteq f(\sigma_j)$ whenever $\sigma_i \subseteq \sigma_j$. As a special case, observe that any inclusion map $\iota : \Sigma \hookrightarrow \Xi$ induces an inclusion map $\iota\mathbb{1} : \Sigma\mathbb{1} \hookrightarrow \Xi\mathbb{1}$.

Given a simplex $\sigma = [x_0, \dots, x_k]$ in a simplicial complex Σ , one defines the *barycenter* to be the point $\mathcal{B}(\sigma) := \sum_{i=0}^k \frac{1}{k+1} x_i \in |\Sigma|$. Then the spaces $|\Sigma\mathbb{1}|$ and $|\Sigma|$ can be identified via a homeomorphism $\mathcal{E}_{|\Sigma|} : |\Sigma\mathbb{1}| \rightarrow |\Sigma|$ defined on vertices by $\mathcal{E}_{|\Sigma|}(\sigma) := \mathcal{B}(\sigma)$ and extended linearly.

Details on the preceding list of definitions can be found in [88, §2.14-15, 2.19], [112, §3.3-4], and also [7, Appendix A]. The next proposition follows from the discussions in these references, and is a simple restatement of [7, Proposition A.1.5]. We provide a proof in the appendix for completeness.

Proposition 158 (Simplicial approximation to \mathcal{E}_\bullet). *Let Σ be a simplicial complex, and let $\Phi : \Sigma\mathbb{1} \rightarrow \Sigma$ be a simplicial map such that $\Phi(\sigma) \in \sigma$ for each $\sigma \in \Sigma$. Then $|\Phi| \simeq \mathcal{E}_{|\Sigma|}$.*

We now introduce some auxiliary constructions dating back to [46] that use the setup stated in Theorem 35. For any nonempty relation $R \subseteq X \times Y$, one may define [46, §2] an associated map $\Phi_{E_R} : E_R\mathbb{1} \rightarrow E_R$ as follows: first define Φ_{E_R} on vertices of $E_R\mathbb{1}$ by $\Phi_{E_R}(\sigma) = s_\sigma$, where s_σ is the least vertex of σ with respect to the total order. Next, for any simplex $[\sigma_1, \dots, \sigma_k]$ of $E_R\mathbb{1}$, where $\sigma_1 \subseteq \dots \subseteq \sigma_k$, we have $\Phi_{E_R}(\sigma_i) = s_{\sigma_i} \in \sigma_k$ for all $1 \leq i \leq k$. Thus $[\Phi_{E_R}(\sigma_1), \dots, \Phi_{E_R}(\sigma_k)] = [s_{\sigma_1}, s_{\sigma_2}, \dots, s_{\sigma_k}]$ is a face of σ_k , hence a simplex of Σ . This defines Φ_{E_R} as a simplicial map $E_R\mathbb{1} \rightarrow E_R$. This argument also shows that Φ_{E_R} is order-reversing: if $\sigma \subseteq \sigma'$, then $\Phi_{E_R}(\sigma) \geq \Phi_{E_R}(\sigma')$.

Remark 159. Applying Proposition 158 to the setup above, one sees that $|\Phi_{E_R}| \simeq \mathcal{E}_{|E_R|}$. After passing to a second barycentric subdivision $E_R^{(2)}$ (obtained by taking a barycentric subdivision of $E_R\mathbb{1}$) and obtaining a map $\Phi_{E_R\mathbb{1}} : E_R^{(2)} \rightarrow E_R\mathbb{1}$, one also has $|\Phi_{E_R\mathbb{1}}| \simeq \mathcal{E}_{|E_R\mathbb{1}|}$.

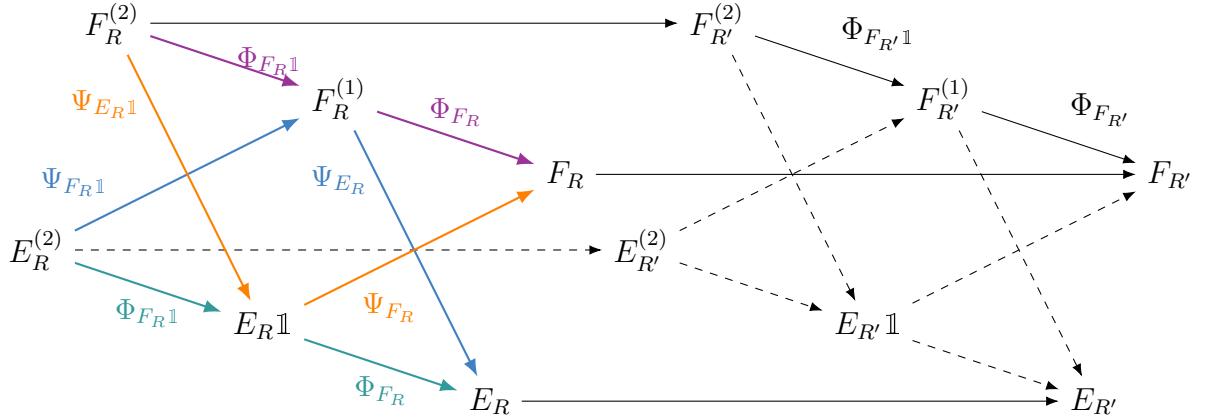
One also defines [46, §3] a simplicial map $\Psi_{F_R} : E_R \mathbb{1} \rightarrow F_R$ as follows. Given a vertex $\sigma = [x_0, \dots, x_k] \in E_R \mathbb{1}$, one defines $\Psi_{F_R}(\sigma) = y_\sigma$, for some $y_\sigma \in Y$ such that $(x_i, y_\sigma) \in R$ for each i . To see why this vertex map is simplicial, let $\sigma \mathbb{1} = [\sigma_0, \dots, \sigma_k]$ be a simplex in $E_R \mathbb{1}$. Let $x \in \sigma_0$. Then, because $\sigma_0 \subseteq \sigma_1 \subseteq \dots \subseteq \sigma_k$, we automatically have that $(x, \Psi_{F_R}(\sigma_i)) \in R$, for each $i = 0, \dots, k$. Thus $\Psi_{F_R}(\sigma \mathbb{1})$ is a simplex in F_R . This definition involves a choice of y_σ when writing $\Psi_{F_R}(\sigma) = y_\sigma$, but all the maps resulting from such choices are contiguous [46, §3].

The preceding map induces a simplicial map $\Psi_{F_R \mathbb{1}} : E_R^{(2)} \rightarrow F_R \mathbb{1}$ as follows. Given a vertex $\tau \mathbb{1} = [\tau_0, \dots, \tau_k] \in E_R^{(2)}$, i.e. a simplex in $E_R \mathbb{1}$, we define $\Psi_{F_R \mathbb{1}}(\tau \mathbb{1}) := [\Psi_{F_R}(\tau_0), \dots, \Psi_{F_R}(\tau_k)]$. Since Ψ_{F_R} is simplicial, this is a simplex in F_R , i.e. a vertex in $F_R \mathbb{1}$. Thus we have a vertex map $\Psi_{F_R \mathbb{1}} : E_R^{(2)} \rightarrow F_R \mathbb{1}$. To check that this map is simplicial, let $\tau^{(2)} = [\tau \mathbb{1}_0, \dots, \tau \mathbb{1}_p]$ be a simplex in $E_R^{(2)}$. Then $\tau \mathbb{1}_0 \subseteq \tau \mathbb{1}_1 \subseteq \dots \subseteq \tau \mathbb{1}_p$, and because Ψ_{F_R} is simplicial, we automatically have

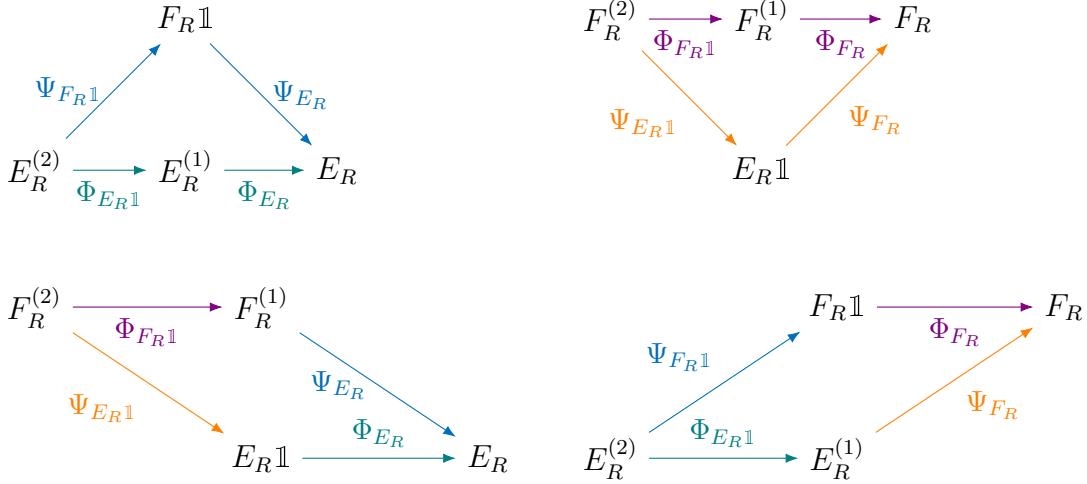
$$\Psi_{F_R}(\tau \mathbb{1}_0) \subseteq \Psi_{F_R}(\tau \mathbb{1}_1) \subseteq \dots \subseteq \Psi_{F_R}(\tau \mathbb{1}_p).$$

Thus $\Psi_{F_R \mathbb{1}}(\tau^{(2)})$ is a simplex in $F_R \mathbb{1}$.

Proof of Theorem 35. We write $F_R^{(2)}$ to denote the barycentric subdivision of $F_R \mathbb{1}$, and obtain simplicial maps $\Phi_{F_R \mathbb{1}} : F_R^{(2)} \rightarrow F_R \mathbb{1}$, $\Phi_{F_R} : F_R^{(1)} \rightarrow F_R$, $\Psi_{E_R \mathbb{1}} : F_R^{(2)} \rightarrow E_R \mathbb{1}$, and $\Psi_{F_R} : E_R^{(1)} \rightarrow F_R$ as above. Consider the following diagram:



We proceed by claiming contiguity of the following.



Claim 13. More specifically:

1. $\Phi_{E_R} \circ \Phi_{E_R\mathbb{1}}$ and $\Psi_{E_R} \circ \Psi_{F_R\mathbb{1}}$ are contiguous.
2. $\Phi_{F_R} \circ \Phi_{F_R\mathbb{1}}$ and $\Psi_{F_R} \circ \Psi_{E_R\mathbb{1}}$ are contiguous.
3. $\Psi_{E_R} \circ \Phi_{F_R\mathbb{1}}$ and $\Phi_{E_R} \circ \Psi_{E_R\mathbb{1}}$ are contiguous.
4. $\Psi_{F_R} \circ \Phi_{E_R\mathbb{1}}$ and $\Phi_{F_R} \circ \Psi_{F_R\mathbb{1}}$ are contiguous.

Items (1) and (3) appear in the proof of Dowker's theorem [46, Lemmas 5, 6], and it is easy to see that a symmetric argument shows Items (2) and (4). For completeness, we will verify these items in this paper, but defer this verification to the end of the proof.

By passing to the geometric realization and applying Proposition 153 and Remark 159, we obtain the following from Item (3) of Claim 13:

$$\begin{aligned} |\Psi_{E_R}| \circ |\Phi_{F_R\mathbb{1}}| &\simeq |\Phi_{E_R}| \circ |\Psi_{E_R\mathbb{1}}|, \\ |\Psi_{E_R}| \circ \mathcal{E}_{|F_R\mathbb{1}|} &\simeq \mathcal{E}_{|E_R|} \circ |\Psi_{E_R\mathbb{1}}|, \quad (\text{Remark 159}) \\ \mathcal{E}_{|E_R|}^{-1} \circ |\Psi_{E_R}| \circ \mathcal{E}_{|F_R\mathbb{1}|} &\simeq |\Psi_{E_R\mathbb{1}}|. \quad (\mathcal{E} \text{ is a homeomorphism, hence invertible}) \end{aligned}$$

Replacing this term in the expression for Item (2) of Claim 13, we obtain:

$$\begin{aligned} |\Psi_{F_R}| \circ |\Psi_{E_R\mathbb{1}}| &\simeq |\Phi_{F_R}| \circ |\Phi_{F_R\mathbb{1}}| \simeq \mathcal{E}_{|F_R|} \circ \mathcal{E}_{|F_R\mathbb{1}|}, \\ |\Psi_{F_R}| \circ \mathcal{E}_{|E_R|}^{-1} \circ |\Psi_{E_R}| \circ \mathcal{E}_{|F_R\mathbb{1}|} &\simeq \mathcal{E}_{|F_R|} \circ \mathcal{E}_{|F_R\mathbb{1}|}, \\ |\Psi_{F_R}| \circ \mathcal{E}_{|E_R|}^{-1} \circ |\Psi_{E_R}| \circ \mathcal{E}_{|F_R|}^{-1} &\simeq \text{id}_{|F_R|}. \end{aligned}$$

Similarly, we obtain the following from Item (4) of Claim 13:

$$|\Psi_{F_R}| \circ |\Phi_{E_R\mathbb{1}}| \simeq |\Phi_{F_R}| \circ |\Psi_{F_R\mathbb{1}}|, \text{ so } \mathcal{E}_{|F_R|}^{-1} \circ |\Psi_{F_R}| \circ \mathcal{E}_{|E_R\mathbb{1}|} \simeq |\Psi_{F_R\mathbb{1}}|.$$

Replacing this term in the expression for Item (1) of Claim 13, we obtain:

$$\begin{aligned} |\Psi_{E_R}| \circ |\Psi_{F_R \mathbb{1}}| &\simeq |\Phi_{E_R}| \circ |\Phi_{E_R \mathbb{1}}| \simeq \mathcal{E}_{|E_R|} \circ \mathcal{E}_{|E_R \mathbb{1}|}, \\ |\Psi_{E_R}| \circ \mathcal{E}_{|F_R|}^{-1} \circ |\Psi_{F_R}| \circ \mathcal{E}_{|E_R \mathbb{1}}| &\simeq \mathcal{E}_{|E_R|} \circ \mathcal{E}_{|E_R \mathbb{1}|} \\ |\Psi_{E_R}| \circ \mathcal{E}_{|F_R|}^{-1} \circ |\Psi_{F_R}| \circ \mathcal{E}_{|E_R|}^{-1} &\simeq \text{id}_{|E_R|} \end{aligned}$$

Define $\Gamma_{|E_R|} : |F_R| \rightarrow |E_R|$ by $\Gamma_{|E_R|} := |\Psi_{E_R}| \circ \mathcal{E}_{|F_R|}^{-1}$. Then $\Gamma_{|E_R|}$ is a homotopy equivalence, with inverse given by $|\Psi_{F_R}| \circ \mathcal{E}_{|E_R|}^{-1}$. This shows that $|F_R| \simeq |E_R|$, for any nonempty relation $R \subseteq X \times Y$.

Next we need to show that $\Gamma_{|E_R|}$ commutes with the canonical inclusion. Consider the following diagram, where the ι_\bullet maps denote the respective canonical inclusions (cf. Definition 47):

$$\begin{array}{ccccc} F_R \mathbb{1} & \xrightarrow{\iota_{F \mathbb{1}}} & F_{R'} \mathbb{1} & & \\ \downarrow \Phi_{F_R} & & \downarrow \Phi_{F_{R'}} & & \\ F_R & \xrightarrow{\iota_F} & F_{R'} & & \\ \downarrow \Psi_{E_R} & & \downarrow \Psi_{E_{R'}} & & \\ E_R & \xrightarrow{\iota_E} & E_{R'} & & \end{array}$$

Claim 14. $\iota_E \circ \Psi_{E_R}$ and $\Psi_{E_{R'}} \circ \iota_{F \mathbb{1}}$ are contiguous.

Claim 15. $\iota_F \circ \Phi_{F_R}$ and $\Phi_{F_{R'}} \circ \iota_{F \mathbb{1}}$ are contiguous.

Suppose Claim 15 is true. Then, upon passing to geometric realizations, we have:

$$\begin{aligned} |\iota_F| \circ \mathcal{E}_{|F_R|} &\simeq |\iota_F| \circ |\Phi_{F_R}| \simeq |\Phi_{F_{R'}}| \circ |\iota_{F \mathbb{1}}| \simeq \mathcal{E}_{|F_{R'}|} \circ |\iota_{F \mathbb{1}}|, \\ \mathcal{E}_{|F_{R'}|}^{-1} \circ |\iota_F| \circ \mathcal{E}_{|F_R|} &\simeq |\iota_{F \mathbb{1}}|. \end{aligned}$$

Suppose also that Claim 14 is true. Then we have:

$$\begin{aligned} |\Psi_{E_{R'}}| \circ |\iota_{F \mathbb{1}}| &\simeq |\iota_E| \circ |\Psi_{E_R}|, \\ |\Psi_{E_{R'}}| \circ \mathcal{E}_{|F_{R'}|}^{-1} \circ |\iota_F| \circ \mathcal{E}_{|F_R|} &\simeq |\iota_E| \circ |\Psi_{E_R}|, \\ |\Psi_{E_{R'}}| \circ \mathcal{E}_{|F_{R'}|}^{-1} \circ |\iota_F| &\simeq |\iota_E| \circ |\Psi_{E_R}| \circ \mathcal{E}_{|F_R|}^{-1}, \text{ i.e.} \\ \Gamma_{|E_{R'}|} \circ |\iota_F| &\simeq |\iota_E| \circ \Gamma_{|E_R|}. \end{aligned}$$

This proves the theorem. It only remains to prove the various claims.

Proof of Claim 13. In proving Claim 13, we supply the proofs of Items (2) and (4). These arguments are adapted from [46, Lemmas 1, 5, and 6], where the proofs of Items (1) and (3) appeared.

For Item (2), let $\tau^{(2)} = [\tau_0 \mathbb{1}, \dots, \tau_k \mathbb{1}]$ be a simplex in $F_R^{(2)}$, where $\tau_0 \mathbb{1} \subseteq \dots \subseteq \tau_k \mathbb{1}$ is a chain of simplices in $F_R \mathbb{1}$. By the order-reversing property of the map $\Phi_{F_R \mathbb{1}}$, we have that $\Phi_{F_R \mathbb{1}}(\tau_0 \mathbb{1}) \supseteq \Phi_{F_R \mathbb{1}}(\tau_i \mathbb{1})$ for each $i = 0, \dots, k$. Define $x := \Psi_{E_R}(\Phi_{F_R \mathbb{1}}(\tau_0 \mathbb{1}))$. Then $(x, y) \in R$ for each $y \in \Phi_{F_R \mathbb{1}}(\tau_0 \mathbb{1})$. But we also have $(x, \Phi_{F_R}(\Phi_{F_R \mathbb{1}}(\tau_i \mathbb{1}))) \in R$ for each $i = 0, \dots, k$, because $\Phi_{F_R}(\Phi_{F_R \mathbb{1}}(\tau_i \mathbb{1})) \in \Phi_{F_R \mathbb{1}}(\tau_i \mathbb{1}) \subseteq \Phi_{F_R}(\tau_i \mathbb{1})$ for each $i = 0, \dots, k$.

Next let $0 \leq i \leq k$. For each $\tau \in \tau \mathbb{1}_i$, we have $\Psi_{E_R}(\tau) \in \Psi_{E_R \mathbb{1}}(\tau \mathbb{1}_i)$ (by the definition of $\Psi_{E_R \mathbb{1}}$). Because $\Phi_{F_R \mathbb{1}}(\tau_0 \mathbb{1}) \in \tau_0 \mathbb{1} \subseteq \tau_i \mathbb{1}$, we then have $x = \Psi_{E_R}(\Phi_{F_R \mathbb{1}}(\tau_0 \mathbb{1})) \in \Psi_{E_R \mathbb{1}}(\tau \mathbb{1}_i)$, which is a vertex of $E_R \mathbb{1}$ or alternatively a simplex of E_R . But then, by definition of Ψ_{F_R} , we have that $(x, \Psi_{F_R}(\Psi_{E_R \mathbb{1}}(\tau_i \mathbb{1}))) \in R$. This holds for each $0 \leq i \leq k$. Since $\tau^{(2)}$ was arbitrary, this shows that $\Phi_{F_R} \circ \Phi_{F_R \mathbb{1}}$ and $\Psi_{F_R} \circ \Psi_{E_R \mathbb{1}}$ are contiguous.

For Item (4), let $\sigma^{(2)} = [\sigma \mathbb{1}_0, \dots, \sigma \mathbb{1}_k]$ be a simplex in $E_R^{(2)}$. Let $0 \leq i \leq k$. Then $\sigma \mathbb{1}_0 \subseteq \dots \subseteq \sigma \mathbb{1}_k$, and $\Phi_{E_R \mathbb{1}}(\sigma \mathbb{1}_i) \in \sigma \mathbb{1}_i \subseteq \sigma \mathbb{1}_k$. So $\Psi_{F_R}(\Phi_{E_R \mathbb{1}}(\sigma \mathbb{1}_i)) \in \Psi_{F_R \mathbb{1}}(\sigma \mathbb{1}_k)$. On the other hand, we have $\Psi_{F_R \mathbb{1}}(\sigma \mathbb{1}_i) \subseteq \Psi_{F_R \mathbb{1}}(\sigma \mathbb{1}_k)$. Then $\Phi_{F_R}(\Psi_{F_R \mathbb{1}}(\sigma \mathbb{1}_i)) \in \Psi_{F_R \mathbb{1}}(\sigma \mathbb{1}_i) \subseteq \Psi_{F_R \mathbb{1}}(\sigma \mathbb{1}_k)$. Since i was arbitrary, this shows that $\Psi_{F_R} \circ \Phi_{E_R \mathbb{1}}$ and $\Phi_{F_R} \circ \Psi_{F_R \mathbb{1}}$ both map the vertices of $\sigma^{(2)}$ to the simplex $\Psi_{F_R \mathbb{1}}(\sigma \mathbb{1}_k)$, hence are contiguous. This concludes the proof of the claim. ■

Proof of Claim 14. Let $\tau \mathbb{1} = [\tau_0, \tau_1, \dots, \tau_k] \in F_R \mathbb{1}$, where $\tau_0 \subseteq \tau_1 \subseteq \dots \subseteq \tau_k$ is a chain of simplices in F_R . Then $\iota_{F \mathbb{1}}(\tau \mathbb{1}) = \tau \mathbb{1}$, and $\Psi_{E_{R'}}(\tau \mathbb{1}) = [x_{\tau_0}, \dots, x_{\tau_k}]$, for some choice of x_{τ_i} terms. Also we have $\iota_E \circ \Psi_{E_R}(\tau \mathbb{1}) = [x'_{\tau_0}, \dots, x'_{\tau_k}]$ for some other choice of x'_{τ_i} terms. For contiguity, we need to show that

$$[x_{\tau_0}, \dots, x_{\tau_k}, x'_{\tau_0}, \dots, x'_{\tau_k}] \in E_{R'}.$$

But this is easy to see: letting $y \in \tau_0$, we have $\{(x_{\tau_0}, y), \dots, (x_{\tau_k}, y), (x'_{\tau_0}, y), \dots, (x'_{\tau_k}, y)\} \subseteq R$. Since $\tau \mathbb{1}$ was arbitrary, it follows that we have contiguity. ■

Proof of Claim 15. Let $\tau \mathbb{1} = [\tau_0, \tau_1, \dots, \tau_k] \in F_R \mathbb{1}$, where $\tau_0 \subseteq \tau_1 \subseteq \dots \subseteq \tau_k$ is a chain of simplices in F_R . Then $\Phi_{F_R}(\tau_i) \in \tau_k$ for each $0 \leq i \leq k$. Thus $\iota_F \circ \Phi_{F_R}(\tau \mathbb{1})$ is a face of τ_k . Similarly, $\Phi_{F_{R'}} \circ \iota_{F \mathbb{1}}(\tau \mathbb{1})$ is also a face of τ_k . Since $\tau \mathbb{1}$ was an arbitrary simplex of $F_R \mathbb{1}$, it follows that $\iota_F \circ \Phi_{F_R}$ and $\Phi_{F_{R'}} \circ \iota_{F \mathbb{1}}$ are contiguous. ■

□

3.2.3 The equivalence between the finite FDT and the simplicial FNTs

In this section, we present our answer to Question 1. We present the proof of Theorem 39 over the course of the next few subsections.

Remark 160. By virtue of Theorem 39, we will write *simplicial FNT* to mean either of the FNT I or FNT II.

Theorem 36 implies Theorem 37

Proof of Theorem 37. Let V, V' denote the vertex sets of Σ, Σ' , respectively. We define the relations $R \subseteq V \times I$ and $R' \subseteq V' \times I'$ as follows: $(v, i) \in R \iff v \in \Sigma_i$ and $(v', i') \in R' \iff v' \in \Sigma'_i$. Then $R \subseteq R'$, the set I' is finite by assumption, and so we are in the setting of the finite FDT (Theorem 36) (perhaps invoking the Axiom of Choice to obtain the total order on V'). It suffices to show that $E_R = \Sigma$, $E_{R'} = \Sigma'$, $F_R = \mathcal{N}(\mathcal{A}_\Sigma)$, and $F_{R'} = \mathcal{N}(\mathcal{A}_{\Sigma'})$, where $E_R, E_{R'}, F_R, F_{R'}$ are as defined in Theorem 35.

First we claim the $E_R = \Sigma$. By the definitions of R and E_R , we have $E_R = \{\sigma \subseteq V : \exists i \in I, (v, i) \in R \forall v \in \sigma\} = \{\sigma \subseteq V : \exists i \in I, v \in \Sigma_i \forall v \in \sigma\}$. Let $\sigma \in E_R$, and let $i \in I$ be such that $v \in \Sigma_i$ for all $v \in \sigma$. Then $\sigma \subseteq V(\Sigma_i)$, and since $\Sigma_i = \text{pow}(V(\Sigma_i))$ by the assumption about covers of simplices, we have $\sigma \in \Sigma_i \subseteq \Sigma$. Thus $E_R \subseteq \Sigma$. Conversely, let $\sigma \in \Sigma$. Then $\sigma \in \Sigma_i$ for some i . Thus for all $v \in \sigma$, we have $(v, i) \in R$. It follows that $\sigma \in E_R$. This shows $E_R = \Sigma$. The proof that $E_{R'} = \Sigma'$ is analogous.

Next we claim that $F_R = \mathcal{N}(\mathcal{A}_\Sigma)$. By the definition of F_R , we have $F_R = \{\tau \subseteq I : \exists v \in V, (v, i) \in R \forall i \in \tau\}$. Let $\tau \in F_R$, and let $v \in V$ be such that $(v, i) \in R$ for each $i \in \tau$. Then $\cap_{i \in \tau} \Sigma_i \neq \emptyset$, and so $\tau \in \mathcal{N}(\mathcal{A}_\Sigma)$. Conversely, let $\tau \in \mathcal{N}(\mathcal{A}_\Sigma)$. Then $\cap_{i \in \tau} \Sigma_i \neq \emptyset$, so there exists $v \in V$ such that $v \in \Sigma_i$ for each $i \in \tau$. Thus $\sigma \in F_R$. This shows $F_R = \mathcal{N}(\mathcal{A}_\Sigma)$. The case for R' is analogous.

An application of Theorem 36 now completes the proof. \square

Theorem 38 implies Theorem 36

Proof. Let X and Y be two sets, and suppose X is finite. Let $R \subseteq R' \subseteq X \times Y$ be two relations. Consider the simplicial complexes $E_R, F_R, E_{R'}, F_{R'}$ as defined in Theorem 35. Let $V_R := V(E_R)$. For each $x \in V_R$, define $A_x := \{\tau \in F_R : (x, y) \in R \text{ for all } y \in \tau\}$. Then A_x is a subcomplex of F_R . Furthermore, $\cup_{x \in V_R} A_x = F_R$. To see this, let $\tau \in F_R$. Then there exists $x \in X$ such that $(x, y) \in R$ for all $y \in \tau$, and so $\tau \in A_x$.

Let $\mathcal{A} := \{A_x : x \in V_R\}$. We have seen that \mathcal{A} is a cover of subcomplexes for F_R . It is finite because the indexing set V_R is a subset of X , which is finite by assumption. Next we claim that $\mathcal{N}(\mathcal{A}) = E_R$. Let $\sigma \in E_R$. Then there exists $y \in Y$ such that $(x, y) \in R$ for all $x \in \sigma$. Thus $\cap_{x \in \sigma} A_x \neq \emptyset$, and so $\sigma \in \mathcal{N}(\mathcal{A})$. Conversely, let $\sigma \in \mathcal{N}(\mathcal{A})$. Then $\cap_{x \in \sigma} A_x \neq \emptyset$, and so there exists $y \in Y$ such that $(x, y) \in R$ for all $x \in \sigma$. Thus $\sigma \in E_R$.

Next we check that nonempty finite intersections of elements in \mathcal{A} are contractible. Let $\sigma \in \mathcal{N}(\mathcal{A}) = E_R$. Let $V_\sigma := \cap_{x \in \sigma} V(A_x) \subseteq V(F_R)$. We claim that $\cap_{x \in \sigma} A_x = \text{pow}(V_\sigma)$, i.e. that the intersection is a full simplex in F_R , hence contractible. The inclusion $\cap_{x \in \sigma} A_x \subseteq \text{pow}(V_\sigma)$ is clear, so we show the reverse inclusion. Let $\tau \in \text{pow}(V_\sigma)$, and let $y \in \tau$. Then $y \in \cap_{x \in \sigma} A_x$, so $(x, y) \in R$ for each $x \in \sigma$. This holds for each $y \in \tau$, so it follows that $\tau \in \cap_{x \in \sigma} A_x$. Thus $\cap_{x \in \sigma} A_x = \text{pow}(V_\sigma)$. We remark that this also shows that \mathcal{A} is a cover of simplices for F_R .

Now for each $x \in V(E_{R'})$, define $A'_x := \{\tau \in F_{R'} : (x, y) \in R' \text{ for all } y \in \tau\}$. Also define $\mathcal{A}' := \{A'_x : x \in V(E_{R'})\}$. The same argument shows that \mathcal{A}' is a finite cover of subcomplexes (in particular, a cover of simplices) for $F_{R'}$ with all finite intersections either empty or contractible, and that $E_{R'} = \mathcal{N}(\mathcal{A}')$. An application of Theorem 38 now shows that $|E_R| \simeq |F_R|$ and $|E_{R'}| \simeq |F_{R'}|$, via maps that commute up to homotopy with the inclusions $|E_R| \hookrightarrow |E_{R'}|$ and $|F_R| \hookrightarrow |F_{R'}|$. \square

Theorem 37 implies Theorem 38

We lead with some remarks about the ideas involved in this proof. Theorem 38 is a functorial statement in the sense that it is about an arbitrary inclusion $\Sigma \subseteq \Sigma'$. Restricting the statement to just Σ would lead to a non-functorial statement. A proof of this non-functorial statement, via a non-functorial analogue of Theorem 37, can be obtained using techniques presented in [12] (see also [78, Theorem 15.24]). We strengthen these techniques to our functorial setting and thus obtain a proof of Theorem 38 via Theorem 37.

We first present a lemma related to barycentric subdivisions and several lemmas about gluings and homotopy equivalences. These will be used in proving Theorem 38.

Definition 48 (Induced subcomplex). Let Σ be a simplicial complex, and let Δ be a subcomplex. Then Δ is an *induced subcomplex* if $\Delta = \Sigma \cap \text{pow}(V(\Delta))$.

Lemma 161. *Let Σ be a simplicial complex, and let Δ be a subcomplex. Then $\Delta\mathbb{1}$ is an induced subcomplex of $\Sigma\mathbb{1}$, i.e. $\Delta\mathbb{1} = \Sigma\mathbb{1} \cap \text{pow}(V(\Delta\mathbb{1}))$.*

Proof. Let σ be a simplex of $\Delta\mathbb{1}$. Then σ belongs to $\Sigma\mathbb{1}$, and also to the full simplex $\text{pow}(V(\Delta\mathbb{1}))$. Thus $\Delta\mathbb{1} \subseteq \Sigma\mathbb{1} \cap \text{pow}(V(\Delta\mathbb{1}))$. Conversely, let $\sigma \in \Sigma\mathbb{1} \cap \text{pow}(V(\Delta\mathbb{1}))$. Since $\sigma \in \Sigma\mathbb{1}$, we can write $\sigma = [\tau_0, \dots, \tau_k]$, where $\tau_0 \subseteq \dots \subseteq \tau_k$. Since $\sigma \in \text{pow}(V(\Delta\mathbb{1}))$ and the vertices of $\Delta\mathbb{1}$ are simplices of Δ , we also know that each τ_i is a simplex of Δ . Thus $\sigma \in \Delta\mathbb{1}$. The equality follows. \square

Lemma 162 (Carrier Lemma, [12] §4). *Let X be a topological space, and let Σ be a simplicial complex. Also let $f, g : X \rightarrow |\Sigma|$ be any two continuous maps such that $f(x), g(x)$ belong to the same simplex of $|\Sigma|$, for any $x \in X$. Then $f \simeq g$.*

Lemma 163 (Gluing Lemma, see Lemmas 4.2, 4.7, 4.9, [12]). *Let Σ be a simplicial complex, and let $U \subseteq V(\Sigma)$. Suppose $|\Sigma \cap \text{pow}(U)|$ is contractible. Then there exists a homotopy equivalence $\varphi : |\Sigma \cup \text{pow}(U)| \rightarrow |\Sigma|$.*

The Gluing and Carrier Lemmas presented above are classical. We provide full details for the Gluing lemma inside the proof of the following functorial generalization of Lemma 163.

Lemma 164 (Functorial Gluing Lemma). *Let $\Sigma \subseteq \Sigma'$ be two simplicial complexes. Also let $U \subseteq V(\Sigma)$ and $U' \subseteq V(\Sigma')$ be such that $U \subseteq U'$. Suppose $|\Sigma \cap \text{pow}(U)|$ and $|\Sigma' \cap \text{pow}(U')|$ are contractible. Then,*

1. There exists a homotopy equivalence $\varphi : |\Sigma \cup \text{pow}(U)| \rightarrow |\Sigma|$ such that $\varphi(x)$ and $\text{id}_{|\Sigma \cup \text{pow}(U)|}(x)$ belong to the same simplex of $|\Sigma \cup \text{pow}(U)|$ for each $x \in |\Sigma \cup \text{pow}(U)|$. Furthermore, the homotopy inverse is given by the inclusion $\iota : |\Sigma| \hookrightarrow |\Sigma \cup \text{pow}(U)|$.
2. Given a homotopy equivalence $\varphi : |\Sigma \cup \text{pow}(U)| \rightarrow |\Sigma|$ as above, there exists a homotopy equivalence $\varphi' : |\Sigma' \cup \text{pow}(U')| \rightarrow |\Sigma'|$ such that $\varphi'|_{|\Sigma \cup \text{pow}(U)|} = \varphi$, and $\varphi'(x)$ and $\text{id}_{|\Sigma' \cup \text{pow}(U')|}(x)$ belong to the same simplex of $|\Sigma' \cup \text{pow}(U')|$ for each $x \in |\Sigma' \cup \text{pow}(U')|$. Furthermore, the homotopy inverse is given by the inclusion $\iota' : |\Sigma'| \hookrightarrow |\Sigma' \cup \text{pow}(U')|$.

Proof of Lemma 164. The proof uses this fact: any continuous map of an n -sphere \mathbb{S}^n into a contractible space Y can be continuously extended to a mapping of the $(n+1)$ -disk \mathbb{D}^{n+1} into Y , where \mathbb{D}^{n+1} has \mathbb{S}^n as its boundary [112, p. 27]. First we define φ . On $|\Sigma|$, define φ to be the identity. Next let σ be a minimal simplex in $|\text{pow}(U) \setminus \Sigma|$. By minimality, the boundary of σ (denoted $\text{Bd}(\sigma)$) belongs to $|\Sigma \cap \text{pow}(U)|$, and $|\Sigma|$ in particular. Thus φ is defined on $\text{Bd}(\sigma)$, which is an n -sphere for some $n \geq 0$. Furthermore, φ maps $\text{Bd}(\sigma)$ into the contractible space $|\Sigma \cap \text{pow}(U)|$. Then we use the aforementioned fact to extend φ continuously to all of σ so that φ maps σ into $|\Sigma \cap \text{pow}(U)|$. Furthermore, both $\text{id}_{|\Sigma \cup \text{pow}(U)|}(\sigma) = \sigma$ and $\varphi(\sigma)$ belong to the simplex $|\text{pow}(U)|$. By iterating this procedure, we obtain a retraction $\varphi : |\Sigma \cup \text{pow}(U)| \rightarrow |\Sigma|$ such that $\varphi(x)$ and x belong to the same simplex in $|\Sigma \cup \text{pow}(U)|$, for each $x \in |\Sigma \cup \text{pow}(U)|$. Thus φ is homotopic to $\text{id}_{|\Sigma \cup \text{pow}(U)|}$ by Lemma 162. Thus we have a homotopy equivalence:

$$\text{id}_{|\Sigma|} = \varphi \circ \iota, \quad \iota \circ \varphi \simeq \text{id}_{|\Sigma \cup \text{pow}(U)|}. \quad (\text{here } \iota := \iota_{|\Sigma| \hookrightarrow |\Sigma \cup \text{pow}(U)|})$$

For the second part of the proof, suppose that a homotopy equivalence $\varphi : |\Sigma \cup \text{pow}(U)| \rightarrow |\Sigma|$ as above is provided. We need to extend φ to obtain φ' . Define φ' to be equal to φ on $|\Sigma \cup \text{pow}(U)|$, and equal to the identity on $G := |\Sigma'| \setminus |\Sigma \cup \text{pow}(U)|$. Let σ be a minimal simplex in $|\text{pow}(U')| \setminus G$. Then by minimality, $\text{Bd}(\sigma)$ belongs to $|\Sigma' \cap \text{pow}(U')|$. As before, we have φ' mapping $\text{Bd}(\sigma)$ into the contractible space $|\Sigma' \cap \text{pow}(U')|$, and we extend φ' continuously to a map of σ into $|\Sigma' \cap \text{pow}(U')|$. Once again, $\text{id}_{|\Sigma' \cup \text{pow}(U')|}(x)$ and $\varphi'(x)$ belong to the same simplex $|\text{pow}(U')|$, for all $x \in \sigma$. Iterating this procedure gives a continuous map $\varphi' : |\Sigma' \cup \text{pow}(U')| \rightarrow |\Sigma'|$. This map is not necessarily a retraction, because there may be a simplex $\sigma \in |\Sigma \cup \text{pow}(U)| \cap |\Sigma'|$ on which φ' is not the identity. However, it still holds that φ' is continuous, and that $x, \varphi'(x)$ get mapped to the same simplex for each $x \in |\Sigma' \cup \text{pow}(U')|$. Thus Lemma 162 still applies to show that φ' is homotopic to $\text{id}_{|\Sigma' \cup \text{pow}(U')|}$.

We write ι' to denote the inclusion $\iota' : |\Sigma'| \hookrightarrow |\Sigma' \cup \text{pow}(U')|$. By the preceding work, we have $\iota' \circ \varphi' \simeq \text{id}_{|\Sigma' \cup \text{pow}(U')|}$. Next let $x \in |\Sigma'|$. Then either $x \in |\Sigma'| \cap |\Sigma \cup \text{pow}(U)|$, or $x \in G$. In the first case, we know that $\varphi'(x) = \varphi(x)$ and $\text{id}_{|\Sigma'|}(x) = \text{id}_{|\Sigma \cup \text{pow}(U)|}(x)$ belong to the same simplex of $|\Sigma \cup \text{pow}(U)|$ by the assumption on φ . In the second case, we know that $\varphi'(x) = x = \text{id}_{|\Sigma'|}(x)$. Thus for any $x \in |\Sigma'|$, we know that $\varphi'(x)$ and $\text{id}_{|\Sigma'|}(x)$ belong

to the same simplex in $|\Sigma' \cup \text{pow}(U')|$. By Lemma 162, we then have $\varphi'|_{|\Sigma'|} \simeq \text{id}_{|\Sigma'|}$. Thus $\varphi' \circ \iota' \simeq \text{id}_{|\Sigma'|}$. This shows that φ' is the necessary homotopy equivalence. \square

Now we present the proof of Theorem 38.

Notation. Let I be an ordered set. For any subset $J \subseteq I$, we write (J) to denote the sequence (j_1, j_2, j_3, \dots) , where the ordering is inherited from the ordering on I .

Proof of Theorem 38. The first step is to functorially deform \mathcal{A}_Σ and $\mathcal{A}_{\Sigma'}$ into covers of simplices while still preserving all associated homotopy types. Then we will be able to apply Theorem 37. We can assume by Lemma 161 that each subcomplex Σ_i is induced, and likewise for each Σ'_i . We start by fixing an enumeration $I' = \{l_1, l_2, \dots\}$. Thus I' becomes an ordered set.

Passing to covers of simplices. We now define some inductive constructions. In what follows, we will define complexes denoted $\Sigma^\bullet, \Sigma'^\bullet$ obtained by “filling in” Σ and Σ' while preserving homotopy equivalence, as well as covers of these larger complexes denoted $\Sigma_{*,\bullet}, \Sigma'_{*,\bullet}$. First define:

$$\begin{aligned}\Sigma^{(l_1)} &:= \begin{cases} \Sigma \cup \text{pow}(V(\Sigma_{l_1})) & : \text{if } l_1 \in I \\ \Sigma & : \text{otherwise.} \end{cases} \\ \Sigma'^{(l_1)} &:= \Sigma' \cup \text{pow}(V(\Sigma'_{l_1})).\end{aligned}$$

Next, for all $i \in I$, define

$$\Sigma_{i,(l_1)} := \begin{cases} \Sigma_i \cup \text{pow}(V(\Sigma_i) \cap V(\Sigma_{l_1})) & : \text{if } l_1 \in I \\ \Sigma_i & : \text{otherwise.} \end{cases}$$

And for all $i \in I'$, define

$$\Sigma'_{i,(l_1)} := \Sigma'_i \cup \text{pow}(V(\Sigma'_i) \cap V(\Sigma'_{l_1})).$$

Now by induction, suppose $\Sigma^{(l_1, \dots, l_n)}$ and $\Sigma_{i,(l_1, \dots, l_n)}$ are defined for all $i \in I$. Also suppose $\Sigma'^{(l_1, \dots, l_n)}$ and $\Sigma'_{i,(l_1, \dots, l_n)}$ are defined for all $i \in I'$. Then we define:

$$\begin{aligned}\Sigma^{(l_1, \dots, l_n, l_{n+1})} &:= \begin{cases} \Sigma^{(l_1, \dots, l_n)} \cup \text{pow}(V(\Sigma_{l_{n+1}, (l_1, \dots, l_n)})) & : \text{if } l_{n+1} \in I \\ \Sigma^{(l_1, \dots, l_n)} & : \text{otherwise.} \end{cases} \\ \Sigma'^{(l_1, \dots, l_n, l_{n+1})} &:= \Sigma'^{(l_1, \dots, l_n)} \cup \text{pow}(V(\Sigma'_{l_{n+1}, (l_1, \dots, l_n)})).\end{aligned}$$

For all $i \in I$, we have

$$\Sigma_{i,(l_1, l_2, \dots, l_{n+1})} := \begin{cases} \Sigma_{i,(l_1, l_2, \dots, l_n)} \cup \text{pow}(V(\Sigma_{i,(l_1, l_2, \dots, l_n)}) \cap V(\Sigma_{l_{n+1}, (l_1, l_2, \dots, l_n)})) & : \text{if } l_{n+1} \in I \\ \Sigma_{i,(l_1, l_2, \dots, l_n)} & : \text{otherwise.} \end{cases}$$

And for all $i \in I'$, we have

$$\Sigma'_{i,(l_1,l_2,\dots,l_{n+1})} := \Sigma'_{i,(l_1,l_2,\dots,l_n)} \cup \text{pow}(V(\Sigma'_{i,(l_1,l_2,\dots,l_n)}) \cap V(\Sigma'_{l_{n+1},(l_1,l_2,\dots,l_n)})).$$

Finally, for any $n \leq \text{card}(I')$, we define $\mathcal{A}_{\Sigma,(l_1,\dots,l_n)} := \{\Sigma_{i,(l_1,\dots,l_n)} : i \in I\}$ and $\mathcal{A}_{\Sigma',(l_1,\dots,l_n)} := \{\Sigma'_{i,(l_1,\dots,l_n)} : i \in I'\}$. We will show that these are covers of $\Sigma^{(l_1,l_2,\dots,l_n)}$ and $\Sigma'^{(l_1,l_2,\dots,l_n)}$, respectively.

The next step is to prove by induction that for any $n \leq \text{card}(I')$, we have $|\Sigma| \simeq |\Sigma^{(l_1,l_2,\dots,l_n)}|$ and $|\Sigma'| \simeq |\Sigma'^{(l_1,l_2,\dots,l_n)}|$, that $\mathcal{N}(\mathcal{A}_\Sigma) = \mathcal{N}(\mathcal{A}_{\Sigma,(l_1,l_2,\dots,l_n)})$ and $\mathcal{N}(\mathcal{A}_{\Sigma'}) = \mathcal{N}(\mathcal{A}_{\Sigma',(l_1,l_2,\dots,l_n)})$, and that nonempty finite intersections of the new covers $\mathcal{A}_{\Sigma,(l_1,l_2,\dots,l_n)}$, $\mathcal{A}_{\Sigma',(l_1,l_2,\dots,l_n)}$ remain contractible. For the base case $n = 0$, we have $\Sigma = \Sigma^0$, $\Sigma' = \Sigma'^0$. Thus the base case is true by assumption. We present the inductive step next.

Claim 16. For this claim, let \bullet denote l_1, \dots, l_n , where $0 < n < \text{card}(I')$. Define $l := l_{n+1}$. Suppose the following is true:

1. The collections $\mathcal{A}_{\Sigma,(\bullet)}$ and $\mathcal{A}_{\Sigma',(\bullet)}$ are covers of $\Sigma^{(\bullet)}$ and $\Sigma'^{(\bullet)}$.
2. The nerves of the coverings are unchanged: $\mathcal{N}(\mathcal{A}_\Sigma) = \mathcal{N}(\mathcal{A}_{\Sigma,(\bullet)})$ and $\mathcal{N}(\mathcal{A}_{\Sigma'}) = \mathcal{N}(\mathcal{A}_{\Sigma',(\bullet)})$.
3. Each of the subcomplexes $\Sigma_{i,(\bullet)}$, $i \in I$, and $\Sigma'_{j,(\bullet)}$, $j \in I'$ is induced in $\Sigma^{(\bullet)}$ and $\Sigma'^{(\bullet)}$, respectively.
4. Let $\sigma \subseteq I$. If $\cap_{i \in \sigma} \Sigma_{i,(\bullet)}$ is nonempty, then it is contractible. Similarly, let $\tau \subseteq I'$. If $\cap_{i \in \tau} \Sigma'_{i,(\bullet)}$ is nonempty, then it is contractible.
5. We have homotopy equivalences $|\Sigma| \simeq |\Sigma^{(\bullet)}|$ and $|\Sigma'| \simeq |\Sigma'^{(\bullet)}|$ via maps that commute with the canonical inclusions.

Then the preceding statements are true for $\Sigma^{(\bullet,l)}$, $\Sigma'^{(\bullet,l)}$, $\mathcal{A}_{\Sigma,(\bullet,l)}$, and $\mathcal{A}_{\Sigma',(\bullet,l)}$ as well.

Proof. For the first claim, we have $\Sigma^{(\bullet,l)} = \Sigma^{(\bullet)} \cup \text{pow}(V(\Sigma_{l,(\bullet)})) \subseteq \cup_{i \in I} \Sigma_{i,(\bullet,l)}$. For the inclusion, we used the inductive assumption that $\Sigma^{(\bullet)} = \cup_{i \in I} \Sigma_{i,(\bullet)}$. Similarly, $\Sigma'^{(\bullet,l)} \subseteq \cup_{i \in I'} \Sigma'_{i,(\bullet,l)}$.

For the second claim, let $i \in I$. Then $V(\Sigma_{i,(l_1)}) = V(\Sigma_i)$, and in particular, we have $V(\Sigma_{i,(\bullet,l)}) = V(\Sigma_{i,(\bullet)}) = V(\Sigma_i)$. Next observe that for any $\sigma \subseteq I$, the intersection

$$\cap_{i \in \sigma} \Sigma_i \neq \emptyset \iff \cap_{i \in \sigma} V(\Sigma_i) \neq \emptyset \iff \cap_{i \in \sigma} V(\Sigma_{i,(\bullet,l)}) \neq \emptyset \iff \cap_{i \in \sigma} \Sigma_{i,(\bullet,l)} \neq \emptyset.$$

Thus $\mathcal{N}(\mathcal{A}_\Sigma) = \mathcal{N}(\mathcal{A}_{\Sigma,(\bullet)}) = \mathcal{N}(\mathcal{A}_{\Sigma,(\bullet,l)})$, and similarly $\mathcal{N}(\mathcal{A}_{\Sigma'}) = \mathcal{N}(\mathcal{A}_{\Sigma',(\bullet)}) = \mathcal{N}(\mathcal{A}_{\Sigma',(\bullet,l)})$.

For the third claim, again let $i \in I$. If $l \notin I$, then $\Sigma_{i,(\bullet,l)} = \Sigma_{i,(\bullet)}$, so we are done by the inductive assumption. Suppose $l \in I$. Since $\Sigma_{i,(\bullet)}$ is induced by the inductive assumption,

we have:

$$\begin{aligned}
\Sigma_{i,(\bullet,l)} &= \Sigma_{i,(\bullet)} \cup (\text{pow}(V(\Sigma_{i,(\bullet)})) \cap V(\Sigma_{l,(\bullet)}))) \\
&= (\Sigma^{(\bullet)} \cap \text{pow}(V(\Sigma_{i,(\bullet)}))) \cup (\text{pow}(V(\Sigma_{i,(\bullet)})) \cap \text{pow}(V(\Sigma_{l,(\bullet)}))) \\
&= (\Sigma^{(\bullet)} \cup \text{pow}(V(\Sigma_{l,(\bullet)}))) \cap \text{pow}(V(\Sigma_{i,(\bullet)})) \\
&= \Sigma^{(\bullet,l)} \cap \text{pow}(V(\Sigma_{i,(\bullet)})) = \Sigma^{(l)} \cap \text{pow}(V(\Sigma_{i,(\bullet,l)})).
\end{aligned}$$

Thus $\Sigma_{i,(\bullet,l)}$ is induced. The same argument holds for the I' case.

For the fourth claim, let $\sigma \subseteq I$, and suppose $\cap_{i \in \sigma} \Sigma_{i,(\bullet,l)}$ is nonempty. By the previous claim, each $\Sigma_{i,(\bullet,l)}$ is induced. Thus we write:

$$\begin{aligned}
\cap_{i \in \sigma} \Sigma_{i,(\bullet,l)} &= \Sigma^{(\bullet,l)} \cap \text{pow}(\cap_{i \in \sigma} V(\Sigma_{i,(\bullet,l)})) \\
&= \Sigma^{(\bullet,l)} \cap \text{pow}(\cap_{i \in \sigma} V(\Sigma_{i,(\bullet)})) \\
&= (\Sigma^{(\bullet)} \cup \text{pow}(V(\Sigma_{l,(\bullet)}))) \cap \text{pow}(\cap_{i \in \sigma} V(\Sigma_{i,(\bullet)})) \\
&= (\cap_{i \in \sigma} (\Sigma^{(\bullet)} \cap \text{pow}(V(\Sigma_{i,(\bullet)})))) \cup \text{pow}(\cap_{i \in \sigma} V(\Sigma_{i,(\bullet)}) \cap V(\Sigma_{l,(\bullet)})) \\
&= (\cap_{i \in \sigma} \Sigma_{i,(\bullet)}) \cup \text{pow}(\cap_{i \in \sigma} V(\Sigma_{i,(\bullet)}) \cap V(\Sigma_{l,(\bullet)})).
\end{aligned}$$

For convenience, define $A := (\cap_{i \in \sigma} \Sigma_{i,(\bullet)})$ and $B := \text{pow}(\cap_{i \in \sigma} V(\Sigma_{i,(\bullet)}) \cap V(\Sigma_{l,(\bullet)}))$. Then $|A|$ is contractible by inductive assumption, and $|B|$ is a full simplex, hence contractible. Also, $A \cap B$ has the form

$$\begin{aligned}
&(\cap_{i \in \sigma} (\Sigma^{(\bullet)} \cap \text{pow}(V(\Sigma_{i,(\bullet)})))) \cap \text{pow}(\cap_{i \in \sigma} V(\Sigma_{i,(\bullet)}) \cap V(\Sigma_{l,(\bullet)})) \\
&= \Sigma^{(\bullet)} \cap \text{pow}(\cap_{i \in \sigma} V(\Sigma_{i,(\bullet)}) \cap V(\Sigma_{l,(\bullet)})) \\
&= \cap_{i \in \sigma} \Sigma_{i,(\bullet)} \cap \Sigma_{l,(\bullet)},
\end{aligned}$$

and the latter is contractible by inductive assumption. Thus by Lemma 163, we have $|A \cup B|$ contractible. This proves the claim for the case $\sigma \subseteq I$. The case $\tau \subseteq I'$ is similar.

Now we proceed to the final claim. Since $\Sigma_{l,(\bullet)}$ is induced, we have $\Sigma_{l,(\bullet)} = \Sigma^{(\bullet)} \cap \text{pow}(V(\Sigma_{l,(\bullet)}))$. By the contractibility assumption, we know that $|\Sigma_{l,(\bullet)}|$ is contractible. Also we know that $|\Sigma'_{l,(\bullet)}| = |\Sigma'^{(\bullet)} \cap \text{pow}(V(\Sigma'_{l,(\bullet)}))|$ is contractible. By assumption we also have $V(\Sigma_{l,(\bullet)}) \subseteq V(\Sigma'_{l,(\bullet)})$. Thus by Lemma 164, we obtain homotopy equivalences $\Phi_l : |\Sigma^{(\bullet,l)}| \rightarrow |\Sigma^{(\bullet)}|$ and $\Phi'_l : |\Sigma'^{(\bullet,l)}| \rightarrow |\Sigma'^{(\bullet)}|$ such that Φ'_l extends Φ_l . Furthermore, the homotopy inverses of Φ_l and Φ'_l are just the inclusions $|\Sigma^{(\bullet)}| \hookrightarrow |\Sigma^{(\bullet,l)}|$ and $|\Sigma'^{(\bullet)}| \hookrightarrow |\Sigma'^{(\bullet,l)}|$.

Now let $\iota : |\Sigma^{(\bullet)}| \rightarrow |\Sigma'^{(\bullet)}|$ and $\iota_l : |\Sigma^{(\bullet,l)}| \rightarrow |\Sigma'^{(\bullet,l)}|$ denote the canonical inclusions. We wish to show the equality $\Phi'_l \circ \iota_l = \iota \circ \Phi_l$. Let $x \in |\Sigma^{(\bullet,l)}|$. Because Φ'_l extends Φ_l (this is why we needed the *functorial gluing lemma*), we have

$$\Phi'_l(\iota_l(x)) = \Phi'_l(x) = \Phi_l(x) = \iota(\Phi_l(x)).$$

Since $x \in |\Sigma^{(\bullet,l)}|$ was arbitrary, the equality follows immediately. By the inductive assumption, we already have homotopy equivalences $|\Sigma^{(\bullet)}| \rightarrow |\Sigma|$ and $|\Sigma'^{(\bullet)}| \rightarrow |\Sigma'|$ that

commute with the canonical inclusions. Composing these maps with Φ_l and Φ'_l completes the proof of the claim. \blacksquare

By the preceding work, we replace the subcomplexes Σ_l, Σ'_l by full simplices of the form $\Sigma_{l,(\bullet,l)}, \Sigma'_{l,(\bullet,l)}$. In this process, the nerves remain unchanged and the complexes Σ, Σ' are replaced by homotopy equivalent complexes $\Sigma^{(\bullet,l)}, \Sigma'^{(\bullet,l)}$. Furthermore, this process is functorial—the homotopy equivalences commute with the canonical inclusions $\Sigma \hookrightarrow \Sigma^{(\bullet,l)}$ and $\Sigma' \hookrightarrow \Sigma'^{(\bullet,l)}$.

Repeating the inductive process in Claim 16 for all the finitely many $l \in I$ yields a simplicial complex $\Sigma^{(I)}$ along with a cover of simplices $\mathcal{A}_{\Sigma,(I)}$. We also perform the same procedure for all $l \in I' \setminus I$ (this does not affect $\Sigma^{(I)}$) to obtain a simplicial complex $\Sigma'^{(I')}$ along with a cover of simplices $\mathcal{A}_{\Sigma',(I')}$. Furthermore, $\Sigma^{(I)}$ and $\Sigma'^{(I')}$ are related to Σ and Σ' by a finite sequence of homotopy equivalences that commute with the canonical inclusions. Also, we have $\mathcal{N}(\mathcal{A}_\Sigma) = \mathcal{N}(\mathcal{A}_{\Sigma,(I)})$ and $\mathcal{N}(\mathcal{A}_{\Sigma'}) = \mathcal{N}(\mathcal{A}_{\Sigma',(I')})$. Thus we obtain the following picture:

$$\begin{array}{ccccccc}
|\Sigma| & \xleftarrow{\simeq} & |\Sigma^{(l_1)}| & \xleftarrow{\simeq} & \cdots & \xleftarrow{\simeq} & |\Sigma^{(I)}| \xrightarrow{\simeq} |\mathcal{N}(\mathcal{A}_{\Sigma,(I)})| = |\mathcal{N}(\mathcal{A}_\Sigma)| \\
\downarrow \iota & & \downarrow \iota_{(l_1)} & & & & \downarrow \iota_{(I)} \quad \downarrow \iota_{\mathcal{N},(I)} \quad \downarrow \iota_{\mathcal{N}} \\
|\Sigma'| & \xleftarrow{\simeq} & |\Sigma'^{(l_1)}| & \xleftarrow{\simeq} & \cdots & \xleftarrow{\simeq} & |\Sigma'^{(I')}| \xrightarrow{\simeq} |\mathcal{N}(\mathcal{A}_{\Sigma',(I')})| = |\mathcal{N}(\mathcal{A}_{\Sigma'})|
\end{array}$$

By applying Theorem 37 to the block consisting of $|\Sigma^{(I)}|, |\Sigma'^{(I')}|, |\mathcal{N}(\mathcal{A}_{\Sigma,I})|$ and $|\mathcal{N}(\mathcal{A}_{\Sigma',I'})|$, we obtain a square that commutes up to homotopy. Then by composing the homotopy equivalences constructed above, we obtain a square consisting of $|\Sigma|, |\Sigma'|, |\mathcal{N}(\mathcal{A}_\Sigma)|$, and $|\mathcal{N}(\mathcal{A}_{\Sigma'})|$ that commutes up to homotopy. Thus we obtain homotopy equivalences $|\Sigma| \simeq |\mathcal{N}(\mathcal{A}_\Sigma)|$ and $|\Sigma'| \simeq |\mathcal{N}(\mathcal{A}_{\Sigma'})|$ via maps that commute up to homotopy with the canonical inclusions. \square

3.2.4 Dowker persistence diagrams of cycle networks

The contents of this section rely on results in [3] and [4]. We introduce some minimalist versions of definitions from the referenced papers to use in this section. The reader should refer to these papers for the original definitions.

Given a metric space (M, d_M) and $m \in M$, we will write $\overline{B(m, \varepsilon)}$ to denote a closed ε -ball centered at m , for any $\varepsilon > 0$. For a subset $X \subseteq M$ and some $\varepsilon > 0$, the Čech complex of X at resolution ε is defined to be the following simplicial complex:

$$\check{C}(X, \varepsilon) := \left\{ \sigma \subseteq X : \cap_{x \in \sigma} \overline{B(x, \varepsilon)} \neq \emptyset \right\}.$$

In the setting of metric spaces, the Čech complex coincides with the Dowker source and sink complexes. We will be interested in the special case where the underlying metric space

is the circle. We write S^1 to denote the circle with unit circumference. Next, for any $n \in \mathbb{N}$, we write $\mathbb{X}_n := \{0, \frac{1}{n}, \frac{2}{n}, \dots, \frac{n-1}{n}\}$ to denote the collection of n equally spaced points on S^1 with the restriction of the arc length metric on S^1 . Also let G_n denote the n -node cycle network with vertex set \mathbb{X}_n (in contrast with \mathbb{X}_n , here G_n is equipped with the asymmetric weights defined in §1.3.2). The connection between \mathbb{X}_n and Dowker complexes of the cycle networks G_n is highlighted by the following observation:

Proposition 165. *Let $n \in \mathbb{N}$. Then for any $\delta \in [0, 1]$, we have $\check{C}(\mathbb{X}_n, \frac{\delta}{2}) = \mathfrak{D}_{n\delta, G_n}^{\text{si}}$.*

The scaling factor arises because G_n has diameter $\sim n$, whereas $\mathbb{X}_n \subseteq S^1$ has diameter $\sim 1/2$. This proposition provides a pedagogical step which helps us transport results from the setting of [3] and [4] to that of the current paper.

Proof. For $\delta = 0$, both the Čech and Dowker complexes consist of the n vertices, and are equal. Similarly for $\delta = 1$, both $\check{C}(\mathbb{X}_n, 1)$ and $\mathfrak{D}_{n\delta, G_n}^{\text{si}}$ are equal to the $(n - 1)$ -simplex.

Now suppose $\delta \in (0, 1)$. Let $\sigma \in \mathfrak{D}_{n\delta, G_n}^{\text{si}}$. Then σ is of the form $[\frac{k}{n}, \frac{k+1}{n}, \dots, \frac{\lfloor k+n\delta \rfloor}{n}]$ for some integer $0 \leq k \leq n - 1$, where the $n\delta$ -sink is $\frac{\lfloor k+n\delta \rfloor}{n}$ and all the numerators are taken modulo n . We claim that $\sigma \in \check{C}(\mathbb{X}_n, \frac{\delta}{2})$. To see this, observe that $d_{S^1}(\frac{k}{n}, \frac{\lfloor k+n\delta \rfloor}{n}) \leq \delta$, and so $\overline{B(\frac{k}{n}, \frac{\delta}{2})} \cap \overline{B(\frac{\lfloor k+n\delta \rfloor}{n}, \frac{\delta}{2})} \neq \emptyset$. Then we have $\sigma \in \bigcap_{i=0}^{n\delta} \overline{B\left(\frac{\lfloor k+i \rfloor}{n}, \frac{\delta}{2}\right)}$, and so $\sigma \in \check{C}(\mathbb{X}_n, \frac{\delta}{2})$.

Now let $\sigma \in \check{C}(\mathbb{X}_n, \frac{\delta}{2})$. Then σ is of the form $[\frac{k}{n}, \frac{k+1}{n}, \dots, \frac{k+j}{n}]$ for some integer $0 \leq k \leq n - 1$, where j is an integer such that $\frac{j}{n} \leq \delta$. In this case, we have $\sigma = \mathbb{X}_n \cap_{i=0}^j \overline{B\left(\frac{k+i}{n}, \delta\right)}$. Then in G_n , after applying the scaling factor n , we have $\sigma \in \mathfrak{D}_{n\delta, G_n}^{\text{si}}$, with $\frac{k+j}{n}$ as an $n\delta$ -sink in G_n . This shows equality of the two simplicial complexes. \square

Theorem 166 (Theorem 3.5, [4]). *Fix $n \in \mathbb{N}$, and let $0 \leq k \leq n - 2$ be an integer. Then,*

$$\check{C}(\mathbb{X}_n, \frac{k}{2n}) \simeq \begin{cases} \bigvee^{n-k-1} S^{2l} & \text{if } \frac{k}{n} = \frac{l}{l+1}, \\ S^{2l+1} & \text{or if } \frac{l}{l+1} < \frac{k}{n} < \frac{l+1}{l+2}, \end{cases}$$

for some $l \in \mathbb{Z}_+$. Here \bigvee denotes the wedge sum, and \simeq denotes homotopy equivalence.

Theorem 44 (Even dimension). *Fix $n \in \mathbb{N}$, $n \geq 3$. If $l \in \mathbb{N}$ is such that n is divisible by $(l + 1)$, and $k := \frac{nl}{l+1}$ is such that $0 \leq k \leq n - 2$, then $\text{Dgm}_{2l}^{\mathfrak{D}}(G_n)$ consists of precisely the point $(\frac{nl}{l+1}, \frac{nl}{l+1} + 1)$ with multiplicity $\frac{n}{l+1} - 1$. If l or k do not satisfy the conditions above, then $\text{Dgm}_{2l}^{\mathfrak{D}}(G_n)$ is trivial.*

Proof of Theorem 44. Let $l \in \mathbb{N}$ be such that $(l + 1)$ divides n and $0 \leq k \leq n - 2$. Then $\mathfrak{D}_{k, G_n}^{\text{si}} = \check{C}(\mathbb{X}_n, \frac{k}{2n})$ has the homotopy type of a wedge sum of $(n - k - 1)$ copies of S^{2l} , by Theorem 166. Here the equality follows from Proposition 165. Notice that $n - k - 1 = \frac{n}{l+1} - 1$. Furthermore, by another application of Theorem 166, it is always possible to choose $\varepsilon > 0$ small enough so that $\mathfrak{D}_{k-\varepsilon, G_n}^{\text{si}} = \check{C}(\mathbb{X}_n, \frac{k-\varepsilon}{2n})$ and $\mathfrak{D}_{k+\varepsilon, G_n}^{\text{si}} = \check{C}(\mathbb{X}_n, \frac{k+\varepsilon}{2n})$ have the homotopy types of odd-dimensional spheres. Thus the inclusions

$\mathfrak{D}_{k-\varepsilon, G_n}^{\text{si}} \subseteq \mathfrak{D}_{k, G_n}^{\text{si}} \subseteq \mathfrak{D}_{k+\varepsilon, G_n}^{\text{si}}$ induce zero maps upon passing to homology. It follows that $\text{Dgm}_{2l}^{\mathfrak{D}}(G_n)$ consists of the point $(\frac{nl}{l+1}, \frac{nl}{l+1} + 1)$ with multiplicity $\frac{n}{l+1} - 1$.

If $l \in \mathbb{N}$ does not satisfy the condition described above, then there does not exist an integer $1 \leq j \leq n-2$ such that $j/n = l/(l+1)$. So for each $1 \leq j \leq n-2$, $\mathfrak{D}_{j, G_n}^{\text{si}} = \check{C}(\mathbb{X}_n, \frac{j}{2n})$ has the homotopy type of an odd-dimensional sphere by Theorem 166, and thus does not contribute to $\text{Dgm}_{2l}^{\mathfrak{D}}(G_n)$. If l satisfies the condition but $k \geq n-1$, then $\check{C}(\mathbb{X}_n, \frac{k}{2n})$ is just the $(n-1)$ -simplex, hence contractible. \square

Theorem 44 gives a characterization of the even dimensional Dowker persistence diagrams of cycle networks. The most interesting case occurs when considering the 2-dimensional diagrams: we see that cycle networks of an even number of nodes have an interesting barcode, even if the bars are all short-lived. For dimensions 4, 6, 8, and beyond, there are fewer and fewer cycle networks with nontrivial barcodes (in the sense that only cycle networks with number of nodes equal to a multiple of 4, 6, 8, and so on have nontrivial barcodes). For a complete picture, it is necessary to look at odd-dimensional persistence diagrams. This is made possible by the next set of constructions.

We have already recalled the definition of a Rips complex of a metric space. To facilitate the assessment of the connection to [3], we temporarily adopt the notation $\mathbf{VR}(X, \varepsilon)$ to denote the Vietoris-Rips complex of a metric space (X, d_X) at resolution $\varepsilon > 0$, i.e. the simplicial complex $\{\sigma \subseteq X : \text{diam}(\sigma) \leq \varepsilon\}$.

Theorem 167 (Theorem 9.3, Proposition 9.5, [3]). *Let $0 < r < \frac{1}{2}$. Then there exists a map $T_r : \text{pow}(S^1) \rightarrow \text{pow}(S^1)$ and a map $\pi_r : S^1 \rightarrow S^1$ such that there is an induced homotopy equivalence*

$$\mathbf{VR}(T_r(X), \frac{2r}{1+2r}) \xrightarrow{\sim} \check{C}(X, r).$$

Next suppose $X \subseteq S^1$ and let $0 < r \leq r' < \frac{1}{2}$. Then there exists a map $\eta : S^1 \rightarrow S^1$ such that the following diagram commutes:

$$\begin{array}{ccc} \mathbf{VR}(T_r(X), \frac{2r}{1+2r}) & \xrightarrow{\eta} & \mathbf{VR}(T_{r'}(X), \frac{2r'}{1+2r'}) \\ \pi_r \downarrow \simeq & & \pi_{r'} \downarrow \simeq \\ \check{C}(X, r) & \xleftarrow{\subseteq} & \check{C}(X, r') \end{array}$$

Theorem 168. *Consider the setup of Theorem 167. If $\check{C}(X, r)$ and $\check{C}(X, r')$ are homotopy equivalent, then the inclusion map between them is a homotopy equivalence.*

Before providing the proof, we show how it implies Theorem 45.

Theorem 45 (Odd dimension). *Fix $n \in \mathbb{N}$, $n \geq 3$. Then for $l \in \mathbb{N}$, define $M_l := \left\{m \in \mathbb{N} : \frac{nl}{l+1} < m < \frac{n(l+1)}{l+2}\right\}$. If M_l is empty, then $\text{Dgm}_{2l+1}^{\mathfrak{D}}(G_n)$ is trivial. Otherwise, we have:*

$$\text{Dgm}_{2l+1}^{\mathfrak{D}}(G_n) = \left\{ \left(a_l, \left\lceil \frac{n(l+1)}{l+2} \right\rceil \right) \right\},$$

where $a_l := \min \{m \in M_l\}$. We use set notation (instead of multisets) to mean that the multiplicity is 1.

Proof of Theorem 45. By Proposition 165 and Theorem 166, we know that $\mathfrak{D}_{k,G_n}^{\text{si}} = \check{C}(\mathbb{X}_n, \frac{k}{2n}) \simeq S^1$ for integers $0 < k < \frac{n}{2}$. Let $b \in \mathbb{N}$ be the greatest integer less than $n/2$. Then by Theorem 168, we know that each inclusion map in the following chain is a homotopy equivalence:

$$\mathfrak{D}_{1,G_n}^{\text{si}} \subseteq \dots \subseteq \mathfrak{D}_{b,G_n}^{\text{si}} = \mathfrak{D}_{\lceil n/2 \rceil^-, G_n}^{\text{si}}.$$

It follows that $\text{Dgm}_1^{\mathfrak{D}}(G_n) = \{(1, \lceil \frac{n}{2} \rceil)\}$. The notation in the last equality means that $\mathfrak{D}_{b,G_n}^{\text{si}} = \mathfrak{D}_{\delta,G_n}^{\text{si}}$ for all $\delta \in [b, b+1)$, where $b+1 = \lceil n/2 \rceil$.

In the more general case, let $l \in \mathbb{N}$ and let M_l be as in the statement of the result. Suppose first that M_l is empty. Then by Proposition 165 and Theorem 166, we know that $\mathfrak{D}_{k,G_n}^{\text{si}}$ has the homotopy type of a wedge of even-dimensional spheres or an odd-dimensional sphere of dimension strictly different from $(2l+1)$, for any choice of integer k . Thus $\text{Dgm}_{2l+1}^{\mathfrak{D}}(G_n)$ is trivial.

Next suppose M_l is nonempty. By another application of Proposition 165 and Theorem 166, we know that $\mathfrak{D}_{k,G_n}^{\text{si}} = \check{C}(\mathbb{X}_n, \frac{k}{2n}) \simeq S^{2l+1}$ for integers $\frac{nl}{l+1} < k < \frac{n(l+1)}{l+2}$. Write $a_l := \min \{m \in M_l\}$ and $b_l := \max \{m \in M_l\}$. Then by Theorem 168, we know that each inclusion map in the following chain is a homotopy equivalence:

$$\mathfrak{D}_{a_l,G_n}^{\text{si}} \subseteq \dots \subseteq \mathfrak{D}_{b_l,G_n}^{\text{si}} = \mathfrak{D}_{\lceil n(l+1)/(l+2) \rceil^-, G_n}^{\text{si}}.$$

It follows that $\text{Dgm}_{2l+1}^{\mathfrak{D}}(G_n) = \left\{ \left(a_l, \lceil \frac{n(l+1)}{l+2} \rceil \right) \right\}$. □

It remains to provide a proof of Theorem 168. For this, we need some additional machinery.

Cyclic maps and winding fractions We introduce some more terms from [3], but for efficiency, we try to minimize the scope of the definitions to only what is needed for our purpose. Recall that we write S^1 to denote the circle with unit circumference. Thus we naturally identify any $x \in S^1$ with a point in $[0, 1)$. We fix a choice of $0 \in S^1$, and for any $x, x' \in S^1$, the length of a clockwise arc from x to x' is denoted by $\overrightarrow{d}_{S^1}(x, x')$. Then, for any finite subset $X \subseteq S^1$ and any $r \in (0, 1/2)$, the *directed Vietoris-Rips graph* $\overrightarrow{\text{VR}}(X, r)$ is defined to be the graph with vertex set X and edge set $\{(x, x') : 0 < \overrightarrow{d}_{S^1}(x, x') < r\}$. Next, let \overrightarrow{G} be a Vietoris-Rips graph such that the vertices are enumerated as x_0, x_1, \dots, x_{n-1} , according to the *clockwise* order in which they appear. A *cyclic map* between \overrightarrow{G} and a Vietoris-Rips graph \overrightarrow{H} is a map of vertices f such that for each edge $(x, x') \in \overrightarrow{G}$, we have either $f(x) = f(x')$, or $(f(x), f(x')) \in \overrightarrow{H}$, and $\sum_{i=0}^{n-1} \overrightarrow{d}_{S^1}(f(x_i), f(x_{i+1})) = 1$. Here $x_n := x_0$.

Next, the *winding fraction* of a Vietoris-Rips graph \overrightarrow{G} with vertex set $V(\overrightarrow{G})$ is defined to be the infimum of numbers $\frac{k}{n}$ such that there is an order-preserving map $V(\overrightarrow{G}) \rightarrow \mathbb{Z}/n\mathbb{Z}$ such that each edge is mapped to a pair of numbers at most k apart. A key property of the

winding fraction, denoted wf , is that if there is a cyclic map between Vietoris-Rips graphs $\vec{G} \rightarrow \vec{H}$, then $\text{wf}(\vec{G}) \leq \text{wf}(\vec{H})$.

Theorem 169 (Corollary 4.5, Proposition 4.9, [3]). *Let $X \subseteq S^1$ be a finite set and let $0 < r < \frac{1}{2}$. Then,*

$$\mathbf{VR}(X, r) \simeq \begin{cases} S^{2l+1} & : \frac{l}{2l+1} < \text{wf}(\overrightarrow{\text{VR}}(X, r)) < \frac{l+1}{2l+3} \text{ for some } l \in \mathbb{Z}_+, \\ \bigvee^j S^{2l} & : \text{wf}(\overrightarrow{\text{VR}}(X, r)) = \frac{l}{2l+1}, \text{ for some } j \in \mathbb{N}. \end{cases}$$

Next let $X' \subseteq S^1$ be another finite set, and let $r \leq r' < \frac{1}{2}$. Suppose $f : \overrightarrow{\text{VR}}(X, r) \rightarrow \overrightarrow{\text{VR}}(X', r')$ is a cyclic map between Vietoris-Rips graphs and $\frac{l}{2l+1} < \text{wf}(\overrightarrow{\text{VR}}(X, r)) \leq \text{wf}(\overrightarrow{\text{VR}}(X', r')) < \frac{l+1}{2l+3}$. Then f induces a homotopy equivalence between $\mathbf{VR}(X, r)$ and $\mathbf{VR}(X', r')$.

We now have the ingredients for a proof of Theorem 168.

Proof of Theorem 168. Since the maps π_r and $\pi_{r'}$ induce homotopy equivalences, it follows that

$$\mathbf{VR}(T_r(X), \frac{2r}{1+2r}) \simeq \mathbf{VR}(T_{r'}(X), \frac{2r'}{1+2r'}).$$

By the characterization result in Theorem 169, we know that there exists $l \in \mathbb{Z}_+$ such that

$$\frac{l}{2l+1} < \text{wf}(\overrightarrow{\text{VR}}(T_r(X), \frac{2r}{1+2r})) \leq \text{wf}(\overrightarrow{\text{VR}}(T_{r'}(X), \frac{2r'}{1+2r'})) < \frac{l+1}{2l+3}.$$

The map η in Theorem 167 appears in [3, Proposition 9.5] through an explicit construction. Moreover, it is shown that η induces a cyclic map

$$\text{wf}(\overrightarrow{\text{VR}}(T_r(X), \frac{2r}{1+2r})) \rightarrow \text{wf}(\overrightarrow{\text{VR}}(T_{r'}(X), \frac{2r'}{1+2r'})).$$

Thus by Theorem 169, η induces a homotopy equivalence between $\mathbf{VR}(T_r(X), \frac{2r}{1+2r})$ and $\mathbf{VR}(T_{r'}(X), \frac{2r'}{1+2r'})$. Finally, the commutativity of the diagram in Theorem 167 shows that the inclusion $\check{C}(X, r) \subseteq \check{C}(X, r')$ induces a homotopy equivalence. \square

Remark 170. The analogue of Theorem 168 for Čech complexes appears as Proposition 4.9 of [3] for Vietoris–Rips complexes. We prove Theorem 168 by connecting Čech and Vietoris–Rips complexes using Proposition 9.5 of [3]. However, as remarked in §9 of [3], one could prove Theorem 168 directly using a parallel theory of winding fractions for Čech complexes.

3.3 Persistent path homology

3.3.1 Digraph maps and functoriality

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

notation means:

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

To extend path homology constructions to a persistent framework, we need to verify the *functoriality* of path homology. As a first step, one must understand how digraph maps transform into maps between vector spaces. Some of the material below can be found in [63]; we contribute a statement and verification of the functoriality of path homology (Proposition 172) that is central to the PPH framework (Definition 20).

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

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

Note also that for any $p \in \mathbb{Z}_+$ and any generator $[x_0, \dots, x_p] \in \Lambda_p(X)$, we have:

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

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

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

$$(\tilde{f}_*)_p : \Lambda_p(X)/\mathcal{I}_p(X) \rightarrow \Lambda_p(Y)/\mathcal{I}_p(Y)$$

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

$$(f_*)_p([x_0, \dots, x_p]) := \begin{cases} [f(x_0), \dots, f(x_p)] & : f(x_i), f(x_{i+1}) \text{ distinct for } 0 \leq i \leq p-1 \\ 0 & : \text{otherwise.} \end{cases}$$

We refer to f_* as *the chain map induced by the set map $f : X \rightarrow Y$* .

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

Proposition 171 (Theorem 2.10, [63]). *Let $G_X = (X, E_X), G_Y = (Y, E_Y)$ be two digraphs, and let $f : G_X \rightarrow G_Y$ be a digraph map. Let $f_* : \mathcal{R}_\bullet(X) \rightarrow \mathcal{R}_\bullet(Y)$ denote the chain map induced by the underlying set map $f : X \rightarrow Y$. Let $(\Omega_p(G_X), \partial_p^{G_X})_{p \in \mathbb{Z}_+}$, $(\Omega_p(G_Y), \partial_p^{G_Y})_{p \in \mathbb{Z}_+}$ denote the chain complexes of the ∂ -invariant paths associated to each of these digraphs. Then $(f_*)_p(\Omega_p(G_X)) \subseteq \Omega_p(G_Y)$ for each $p \in \mathbb{Z}_+$, and the restriction of f_* to $\Omega_\bullet(G_X)$ is a chain map.*

Henceforth, given two digraphs G, G' and a digraph map $f : G \rightarrow G'$, we refer to the chain map f_* given by Proposition 171 as the *chain map induced by the digraph map f*. Because f_* is a chain map, we then obtain an induced linear map $(f_\#)_p : H_p(G) \rightarrow H_p(G')$ for each $p \in \mathbb{Z}_+$.

The preceding concepts are necessary for developing the theory of path homology. We use this set up to state and prove the following result, which is used in defining PPH (Definition 20) and also for proving stability (Theorem 56).

Proposition 172 (Functionality of path homology). *Let G, G', G'' be three digraphs.*

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

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

$$(\text{id}_{G*})_p([x_0, \dots, x_p]) = [\text{id}_G(x_0), \dots, \text{id}_G(x_p)] = [x_0, \dots, x_p].$$

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

$$\begin{aligned} ((g \circ f)_*)_p([x_0, \dots, x_p]) &= [g(f(x_0)), \dots, g(f(x_p))] && g(f(x_i)), g(f(x_{i+1})) \text{ distinct} \\ &= (g)_p([f(x_0), \dots, f(x_p)]) && \text{because } f(x_i), f(x_{i+1}) \text{ distinct} \\ &= (g)_p((f_*)_p([x_0, \dots, x_p])). \end{aligned}$$

Next suppose that for some $0 \leq i < p$, we have $g(f(x_i)) = g(f(x_{i+1}))$. Then we obtain:

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

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

Remark 173. We thank Paul Ignacio for pointing out an error in a version of the preceding proof that appeared in [40].

3.3.2 Homotopy of digraphs

The constructions of path homology are accompanied by a theory of homotopy developed in [63]. An illustrated example is provided in Figure 3.1.

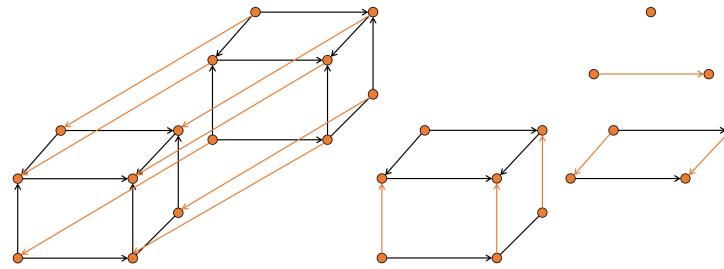


Figure 3.1: Directed d -cubes that are all homotopy equivalent.

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

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

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

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

This condition is equivalent to requiring:

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

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

An example of digraph homotopy equivalence is illustrated in Figure 3.1. Informally, the homotopy equivalence is given by “crushing” the orange arrows according to the directions they mark. This operation crushes the 4-tesseract to the 3-cube, to the 2-square, to the line, and finally to the point.

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

Theorem 174 (Theorem 3.3, [63]). *Let G, G' be two digraphs.*

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

$$(f\#)_p : H_p(G) \rightarrow H_p(G') \quad (g\#)_p : H_p(G) \rightarrow H_p(G').$$

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

3.3.3 The Persistent Path Homology of a Network

Proof of Theorem 56. Let $\eta > 2d_{\mathcal{N}}(\mathcal{X}, \mathcal{Y})$. By virtue of Proposition 7, we obtain maps $\varphi : X \rightarrow Y$ and $\psi : Y \rightarrow X$ such that $\text{dis}(\varphi) < \eta$, $\text{dis}(\psi) < \eta$, $C_{X,Y}(\varphi, \psi) < \eta$, and $C_{Y,X}(\psi, \varphi) < \eta$.

Claim 17. For each $\delta \in \mathbb{R}$, the map φ induces a digraph map $\varphi_\delta : \mathfrak{G}_{\mathcal{X}}^\delta \rightarrow \mathfrak{G}_{\mathcal{Y}}^{\delta+\eta}$ given by $x \mapsto \varphi(x)$, and the map ψ induces a digraph map $\psi_\delta : \mathfrak{G}_{\mathcal{Y}}^\delta \rightarrow \mathfrak{G}_{\mathcal{X}}^{\delta+\eta}$ given by $y \mapsto \psi(y)$.

Proof. Let $\delta \in \mathbb{R}$, and let $(x, x') \in E_X^\delta$. Then $A_X(x, x') \leq \delta$. Because $\text{dis}(\varphi) < \eta$, we have $A_Y(\varphi(x), \varphi(x')) < \delta + \eta$. Thus $(\varphi(x), \varphi(x')) \in E_Y^{\delta+\eta}$, and so φ_δ is a digraph map. Similarly, ψ_δ is a digraph map. Since $\delta \in \mathbb{R}$ was arbitrary, the claim now follows. ■

Claim 18. Let $\delta \leq \delta' \in \mathbb{R}$, and let $s_{\delta, \delta'}, t_{\delta+\eta, \delta'+\eta}$ denote the digraph inclusion maps $\mathfrak{G}_{\mathcal{X}}^\delta \hookrightarrow \mathfrak{G}_{\mathcal{X}}^{\delta'}$ and $\mathfrak{G}_{\mathcal{Y}}^{\delta+\eta} \hookrightarrow \mathfrak{G}_{\mathcal{Y}}^{\delta'+\eta}$, respectively. Then $\varphi_{\delta'} \circ s_{\delta, \delta'}$ and $t_{\delta+\eta, \delta'+\eta} \circ \varphi_\delta$ are one-step homotopic.

Proof. Let $x \in X$. We wish to show $\varphi_{\delta'}(s_{\delta, \delta'}(x)) \xrightarrow{\sim} t_{\delta+\eta, \delta'+\eta}(\varphi_\delta(x))$. But notice that:

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

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

Claim 19. Let $\delta \in \mathbb{R}$, and let $s_{\delta, \delta+2\eta}$ denote the digraph inclusion map $\mathfrak{G}_{\mathcal{X}}^\delta \hookrightarrow \mathfrak{G}_{\mathcal{X}}^{\delta+2\eta}$. Then $s_{\delta, \delta+2\eta}$ and $\psi_{\delta+\eta} \circ \varphi_\delta$ are one-step homotopic.

Proof. Recall that $C_{X,Y}(\varphi, \psi) < \eta$, which means that for any $x \in X, y \in Y$, we have:

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

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

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

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

By combining the preceding claims and Theorem 174, we obtain the following, for each $p \in \mathbb{Z}_+$:

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

By invoking functoriality of path homology (Proposition 172), we obtain:

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

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

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

Thus $\mathbf{PVec}_p^{\Xi}(\mathcal{X})$ and $\mathbf{PVec}_p^{\Xi}(\mathcal{Y})$ are η -interleaved, for each $p \in \mathbb{Z}_+$. The result now follows by an application of Lemma 154. \square

3.3.4 PPH and Dowker persistence

Definition 49 (Type I and II Dowker simplices). Let $(X, A_X) \in \mathcal{FN}$, fix $\delta \in \mathbb{R}$, and let σ be a simplex in $\mathfrak{D}_{\delta,X}^{\text{si}}$. Then we define σ to be a *Type I* simplex if some $x \in \sigma$ is a δ -sink for σ . Otherwise, σ is a *Type II* simplex. Notice that if σ is a Type II simplex, then there exists $x \notin \sigma$ such that x is a δ -sink for σ .

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

Lemma 175 (Proposition 2.9, [63]). *Let \mathfrak{G} be a finite digraph. Then any $v \in \Omega_2(\mathfrak{G})$ is a linear combination of the following three types of ∂ -invariant 2-paths:*

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

Lemma 176 (Parity lemma). *Fix a simplicial complex K and a field $\mathbb{Z}/p\mathbb{Z}$ for some prime p . Let $w := \sum_{i \in I} b_i \tau_i$ be a 2-chain in $C_2(K)$ where I is a finite index set, each $b_i \in \mathbb{Z}/p\mathbb{Z}$, and each τ_i is a 2-simplex in K . Let σ be a 1-simplex contained in some τ_i such that σ does not appear in $\partial_2^\Delta(w)$. Define $J_\sigma := \{j \in I : \sigma \text{ a face of } \tau_j\}$. Then there exists $n(\sigma) \in \mathbb{N}$ such that:*

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

where $\partial_2^\Delta(\tau_j^+ + \tau_j^-)$ is independent of σ for each $1 \leq j \leq n(\sigma)$.

Proof of Lemma 176. Since we are working over $\mathbb{Z}/p\mathbb{Z}$, we adopt the convention that $b_i \in \{0, 1, \dots, p-1\}$ for each $i \in I$. Then for each $j \in J_\sigma$, we know that $\partial_2^\Delta(\tau_j)$ contributes either $+\sigma$ or $-\sigma$ with multiplicity b_j . Write $w = \sum_{i \in I \setminus J_\sigma} b_i \tau_i + \sum_{j \in J_\sigma} b_j \tau_j$.

Since σ is not a summand of $\partial_2^\Delta(w)$, it follows that $\sum_{j \in J_\sigma} b_j = 0$. Define:

$$J_\sigma^+ := \{j \in J_\sigma : \tau_j \text{ contributes } +\sigma\}, \quad J_\sigma^- := \{j \in J_\sigma : \tau_j \text{ contributes } -\sigma\}.$$

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

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

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

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

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

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

Theorem 61. Let $\mathcal{X} = (X, A_X) \in \mathcal{CN}$ be a square-free network, and fix $\mathbb{K} = \mathbb{Z}/p\mathbb{Z}$ for some prime p . Then $\mathrm{Dgm}_1^\Xi(\mathcal{X}) = \mathrm{Dgm}_1^\mathfrak{D}(\mathcal{X})$.

Proof of Theorem 61. Let $\delta \in \mathbb{R}$. First we wish to find an isomorphism $\varphi_\delta : H_1^\Xi(\mathfrak{G}_X^\delta) \rightarrow H_1^\Delta(\mathfrak{D}_{\delta, X}^{\mathrm{si}})$. We begin with the basis B for $\Omega_1(\mathfrak{G}_X^\delta)$. We claim that B is just the collection of allowed 1-paths in \mathfrak{G}_X^δ . To see this, let ab be an allowed 1-path. Then $\partial_1(ab) = b - a$, which is allowed because the vertices a and b are automatically allowed. Thus $ab \in \Omega_1(\mathfrak{G}_X^\delta)$, and so B generates $\Omega_1(\mathfrak{G}_X^\delta)$.

Whenever ab is an allowed 1-path, we have a directed edge (a, b) in \mathfrak{G}_X^δ , and so $A_X(a, b) \leq \delta$ by the definition of \mathfrak{G}_X^δ . Thus the simplex $[a, b]$ belongs to $\mathfrak{D}_{\delta, X}^{\mathrm{si}}$, with b

as a δ -sink. Hence $[a, b]$ is a 1-chain in $C_1(\mathfrak{D}_{\delta, X}^{\text{si}})$. Define a map $\tilde{\varphi}_\delta : \Omega_1(\mathfrak{G}_X^\delta) \rightarrow C_1(\mathfrak{D}_{\delta, X}^{\text{si}})$ by setting $\tilde{\varphi}_\delta(ab) = [a, b]$ and extending linearly. The image of $\tilde{\varphi}_\delta$ restricted to B is linearly independent because any linear dependence relation would contradict the independence of B . Furthermore, $\tilde{\varphi}_\delta$ induces a map $\tilde{\varphi}'_\delta : \ker(\partial_1^\Xi) \rightarrow \ker(\partial_1^\Delta)$. We need to check that this descends to a map $\varphi_d : \ker(\partial_1^\Xi)/\text{im}(\partial_2^\Xi) \rightarrow \ker(\partial_1^\Delta)/\text{im}(\partial_2^\Delta)$ on quotients. To see this, we need to verify that $\tilde{\varphi}'_\delta(\text{im}(\partial_2^\Xi)) \subseteq \text{im}(\partial_2^\Delta)$.

By Lemma 175, we have a complete characterization of $\Omega_2(\mathfrak{G}_X^\delta)$. Thus we know that any element of $\text{im}(\partial_2^\Xi)$ is of the form $ba + ab, bc - ac + ab$, or $bc + ab - dc - ad$. In the first case, we have $\tilde{\varphi}'_\delta(ba + ab) = [b, a] + [a, b] = [b, a] - [b, a] = 0 \in \text{im}(\partial_2^\Delta)$. The next case corresponds to the situation where we have $abc \in \Omega_2(\mathfrak{G}_X^\delta)$ with edges $(a, b), (b, c), (a, c)$ in \mathfrak{G}_X^δ . In this case, $[a, b, c]$ is a 2-simplex in $\mathfrak{D}_{\delta, X}^{\text{si}}$, with c as a δ -sink. Thus $[b, c] - [a, c] + [a, b] = \tilde{\varphi}'_\delta(bc - ac + ab)$ belongs to $\text{im}(\partial_2^\Delta)$.

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

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

Claim 20. $w := \sum_{j=0}^m b_j \tau_j$ is homologous to a 2-cycle $\sum_{k=0}^n b'_k \tau'_k$ in $C_2(\mathfrak{D}_{\delta, X}^{\text{si}})$, where each τ'_k is of the form $[a, b, c]$ and abc is a triangle in \mathfrak{G}_X^δ .

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

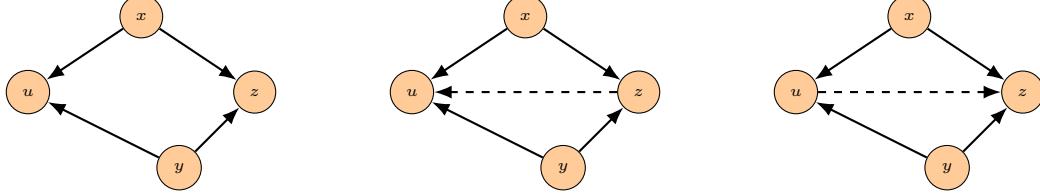
Proof of Claim 20. Let us now prove the claim. Suppose τ_j is a Type II simplex, for some $0 \leq j \leq m$. Write $\tau_j = [u, x, y]$. Then there exists $z \in X$ such that z is a δ -sink for τ_j . But then $[u, x, y, z] \in \mathfrak{D}_{\delta, X}^{\text{si}}$, and $\partial_3^\Delta([u, x, y, z]) = [x, y, z] - [u, y, z] + [u, x, z] - [u, x, y]$. Since $\partial_2^\Delta \circ \partial_3^\Delta = 0$, it follows that $[u, x, y]$ is homologous to $[x, y, z] - [u, y, z] + [u, x, z]$, each of which is a Type I simplex. Using this argument, we first replace all Type II simplices in w by Type I simplices.

Next let τ be a Type I simplex in the rewritten expression for w . By taking a permutation and appending a (-1) coefficient if needed, we can write $\tau = [x, y, z]$, where z is the δ -sink for τ . Thus $(x, z), (y, z)$ are edges in \mathfrak{G}_X^δ . If (x, y) or (y, x) is also an edge, then xyz is a triangle, and we are done. Suppose that neither is an edge, i.e. neither of xy, yx is in $\Omega_1(\mathfrak{G}_X^\delta)$. Then, since xy is not a summand of v , we know that $[x, y]$ is not a summand of $\varphi_\delta(v)$. Thus we are in the setting of Lemma 176, because $\partial_2^\Delta(w) = \varphi_\delta(v)$. Define $J := \{0 \leq j \leq m : [x, y] \text{ a face of } \tau_j\}$. By applying Lemma 176, we can rewrite w :

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

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

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



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

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

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

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

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

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

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

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

To complete the proof, let $\delta \leq \delta' \in \mathbb{R}$. Consider the inclusion maps $\iota_{\mathfrak{G}} : \mathfrak{G}_X^\delta \hookrightarrow \mathfrak{G}_X^{\delta'}$ and $\iota_{\mathfrak{D}} : \mathfrak{D}_{\delta, X}^{\text{si}} \hookrightarrow \mathfrak{D}_{\delta', X}^{\text{si}}$, and let $(\iota_{\mathfrak{G}})_\#$, $(\iota_{\mathfrak{D}})_\#$ denote the induced maps at the respective

homology levels. Let $v = \sum_{i=0}^n a_i x_i y_i$ be a 1-cycle in $H_1^\Xi(\mathfrak{G}_X^\delta)$. Then we have:

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

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

Theorem 63. *Let G_n be a cycle network for some integer $n \geq 3$. Fix a field $\mathbb{K} = \mathbb{Z}/p\mathbb{Z}$ for some prime p . Then $\text{Dgm}_1^\Xi(G_n) = \{(1, \lceil n/2 \rceil)\}$.*

Proof of Theorem 63. From [37], we know that $\text{Dgm}_1^\mathfrak{D}(G_n) = \{(1, \lceil n/2 \rceil)\}$. Thus by Theorem 61, it suffices to show that G_n is square-free. Suppose $n \geq 4$, and let a, b, c, d be four nodes that appear in G_n in clockwise order. First let $\delta \in \mathbb{R}$ be such that $(a, b), (b, c), (a, d), (d, c)$ are edges in $\mathfrak{G}_{G_n}^\delta$. Then $\omega_{G_n}(d, c) \leq \delta$, and because of the clockwise orientation $d \preceq a \preceq c$, we automatically $\omega_{G_n}(a, c) \leq \delta$. Hence (a, c) is an edge in $\mathfrak{G}_{G_n}^\delta$, and so the subgraph induced by a, b, c, d is not a long square.

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

Theorem 59. *Let $\mathcal{X} = (X, A_X) \in \mathcal{CN}$ be a symmetric network, and fix $\mathbb{K} = \mathbb{Z}/p\mathbb{Z}$ for some prime p . Then $\text{Dgm}_1^\Xi(\mathcal{X}) = \text{Dgm}_1^\mathfrak{D}(\mathcal{X})$.*

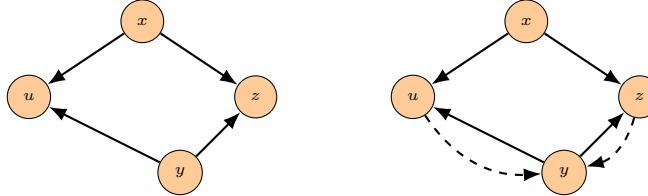
Proof of Theorem 59. The proof is similar to that of Theorem 61; instead of repeating all details, we will show how the argument changes when the square-free assumption is replaced by the symmetry assumption. Let $\delta \in \mathbb{R}$, and consider the map $\tilde{\varphi}'_\delta : \ker(\partial_1^\Xi) \rightarrow \ker(\partial_1^\Delta)$ defined as in Theorem 61. As before, we need to check that this descends to a map $\varphi_\delta : \ker(\partial_1^\Xi)/\text{im}(\partial_2^\Xi) \rightarrow \ker(\partial_1^\Delta)/\text{im}(\partial_2^\Delta)$ on quotients. For this we need to verify that $\tilde{\varphi}'_\delta(\text{im}(\partial_2^\Xi)) \subseteq \text{im}(\partial_2^\Delta)$.

By Lemma 175, we know that any element of $\text{im}(\partial_2^\Xi)$ is of the form $ba+ab, bc-ac+ab$, or $bc+ab-dc-ad$. For the first two cases, we can repeat the argument used in Theorem 61. The final case corresponds to the situation where we have a long square in \mathfrak{G}_X^δ consisting of edges $(a, b), (b, c), (a, d)$, and (d, c) . This gives the 2-chain $abc - adc$. Now by the symmetry condition, we also have edges (c, d) and (c, b) . Thus $[a, b, c]$ is a 2-simplex in $\mathfrak{D}_{\delta, X}^{\text{si}}$, with b as a δ -sink, and $[a, d, c]$ is a 2-simplex with d as a δ -sink. Hence $[a, b, c] - [a, d, c]$ is a 2-chain in $C_2(\mathfrak{D}_{\delta, X}^{\text{si}})$. Thus $\tilde{\varphi}'_\delta(bc+ab-dc-ad) = [b, c] + [a, b] - [d, c] - [a, d]$ belongs to $\text{im}(\partial_2^\Delta)$. Thus we obtain a well-defined map $\varphi_\delta : H_1^\Xi(\mathfrak{G}_X^\delta) \rightarrow H_1^\Delta(\mathfrak{D}_{\delta, X}^{\text{si}})$.

Next we need to check that φ_δ is injective. As in Theorem 61, let $v \in \ker(\varphi_\delta)$. Then $\varphi_\delta(v) = \varphi_\delta(\sum_{i=0}^k a_i \sigma_i) = \partial_2^\Delta(\sum_{j=0}^m b_j \tau_j)$, where the a_i, b_j terms belong to the field \mathbb{K} , each σ_i is a 1-path in \mathfrak{G}_X^δ , and each τ_j is a 2-simplex in $\mathfrak{D}_{\delta, X}^{\text{si}}$. We proceed by proving an

analogue of Claim 20 in the symmetric setting. Write $w := \sum_{j=0}^m b_j \tau_j$. We need to show that w is homologous to a 2-cycle $\sum_{k=0}^n b'_k \tau'_k$ in $C_2(\mathfrak{D}_{\delta,X}^{\text{si}})$, where each τ'_k is of the form $[a, b, c]$ and abc is either a triangle or part of a square in \mathfrak{G}_X^δ .

As in the proof of Claim 20, we first replace all Type II simplices in w by Type I simplices. Next let $\tau = [x, y, z]$ be a Type I simplex in w , and suppose z is the δ -sink for τ , but neither of $(x, y), (y, x)$ is an edge. As in the proof of Theorem 61, we apply Lemma 176 to separate the summands of w containing $[x, y]$ as a face into pairs of the form $(\tau^+ + \tau^-)$. Writing $\tau^+ = [x, z, y]$ and $\tau^- = [x, y, u]$, we obtain the following arrangement:



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

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

We omit the remainder of the argument, because it is a repeat of the corresponding part of the proof of Theorem 61. In summary, it turns out that φ_δ is surjective, hence an isomorphism, and furthermore that it commutes with the linear maps induced by the canonical inclusions. This concludes the proof. \square

3.4 Diagrams of compact networks

We now prove the well-definedness of persistence diagrams arising from compact networks, as well as convergence properties.

Theorem 86. *Let $(X, \omega_X) \in \mathcal{CN}$, $k \in \mathbb{Z}_+$. Then the persistence vector spaces associated to the Vietoris-Rips, Dowker, and PPH constructions are all q -tame.*

Proof of Theorem 86. All the cases are similar, so we just prove the case of $\text{PVec}_k^\mathfrak{D}(X)$. For convenience, write $\text{PVec}_k^\mathfrak{D}(X) = \{V^\delta \xrightarrow{\nu_{\delta, \delta'}} V^{\delta'}\}_{\delta \leq \delta' \in \mathbb{R}}$. Let $\delta < \delta'$. We need to show $\nu_{\delta, \delta'}$ has finite rank. Write $\varepsilon := (\delta' - \delta)/2$. Let \mathcal{U} be an $\varepsilon/4$ -system on X (this requires Theorem 64). Then by Theorem 66 we pick a finite subset $X' \subseteq X$ such that $d_N(X, X') < \varepsilon/2$. By stability results, we have that $\text{PVec}_k^\mathfrak{D}(X')$ and $\text{PVec}_k^\mathfrak{D}(X)$ are ε -interleaved. For convenience, write $\text{PVec}_k^\mathfrak{D}(X') = \{U^\delta \xrightarrow{\mu_{\delta, \delta'}} U^{\delta'}\}_{\delta \leq \delta' \in \mathbb{R}}$. Then the map $\nu_{\delta, \delta'} : V^\delta \rightarrow V^{\delta'}$ factorizes through $U^{\delta+\varepsilon}$ via interleaving maps $V^\delta \rightarrow U^{\delta+\varepsilon} \rightarrow V^{\delta+2\varepsilon} = V^{\delta'}$. Since $U^{\delta+\varepsilon}$ is finite dimensional, it follows that $\nu_{\delta, \delta'}$ has finite rank. This concludes the proof. \square

Corollary 177 (Stability). *Let $(X, \omega_X), (Y, \omega_Y) \in \mathcal{CN}$, $k \in \mathbb{Z}_+$. Then,*

$$d_B(\text{Dgm}_k^\bullet(X), \text{Dgm}_k^\bullet(Y)) \leq 2d_N(X, Y),$$

where Dgm^\bullet denotes each of the Vietoris-Rips, Dowker, or path persistence diagrams.

Proof. By Theorem 86, both the Rips and Dowker persistent vector spaces of X and Y are q -tame. Thus they have well-defined persistence diagrams (Theorem 85), and we have equality of d_I and d_B . \square

Theorem 87 (Convergence). *Let (X, ω_X) be a measure network equipped with a Borel probability measure μ_X . For each $i \in \mathbb{N}$, let $x_i : \Omega \rightarrow X$ be an independent random variable defined on some probability space $(\Omega, \mathcal{F}, \mathbb{P})$ with distribution μ_X . For each $n \in \mathbb{N}$, let $\mathbb{X}_n = \{x_1, x_2, \dots, x_n\}$. Let $\varepsilon > 0$. Then we have:*

$$\mathbb{P}\left(\{\omega \in \Omega : d_B(\text{Dgm}^\bullet(\text{supp}(\mu_X)), \text{Dgm}^\bullet(\mathbb{X}_n(\omega))) \geq \varepsilon\}\right) \leq \frac{(1 - \mathfrak{M}_{\varepsilon/4}(\text{supp}(\mu_X)))^n}{\mathfrak{M}_{\varepsilon/4}(\text{supp}(\mu_X))},$$

where $\mathbb{X}_n(\omega)$ is the subnetwork induced by $\{x_1(\omega), \dots, x_n(\omega)\}$ and Dgm^\bullet is either of the Vietoris-Rips, Dowker, or PPH diagrams. In particular, either of these three persistent vector spaces of the subnetwork \mathbb{X}_n converges almost surely to that of $\text{supp}(\mu_X)$ in bottleneck distance.

Proof of Theorem 87. We can consider $\text{supp}(\mu_X)$ as a network with full support by endowing it with the restriction of ω_X to $\text{supp}(\mu_X) \times \text{supp}(\mu_X)$, so for convenience, we assume $X = \text{supp}(\mu_X)$. Let $\omega \in \Omega$ be such that $d_N(X, \mathbb{X}_n(\omega)) < \varepsilon/2$. Then by Corollary 177, we have that $d_B(\text{Dgm}^\bullet(X), \text{Dgm}^\bullet(\mathbb{X}_n)) < \varepsilon$. By applying Theorem 68, we then have:

$$\begin{aligned} \mathbb{P}\left(\{\omega \in \Omega : d_B(\text{Dgm}^\bullet(X), \text{Dgm}^\bullet(\mathbb{X}_n(\omega))) \geq \varepsilon\}\right) &\leq \mathbb{P}\left(\{\omega \in \Omega : d_N(X, \mathbb{X}_n(\omega)) \geq \varepsilon/2\}\right) \\ &\leq \frac{(1 - \mathfrak{M}_{\varepsilon/4}(\text{supp}(\mu_X)))^n}{\mathfrak{M}_{\varepsilon/4}(\text{supp}(\mu_X))}. \end{aligned}$$

We conclude the proof with an application of the Borel-Cantelli lemma, as in the proof of Theorem 68. \square

Chapter 4: Algorithms, computation, and experiments

4.1 The complexity of computing $d_{\mathcal{N}}$

By Remark 10 and Proposition 12 we know that in the setting of finite networks, it is possible to obtain an upper bound on $d_{\mathcal{N}}$, in the case $\text{card}(X) = \text{card}(Y)$, by using $\widehat{d}_{\mathcal{N}}$. Solving for $\widehat{d}_{\mathcal{N}}(X, Y)$ reduces to minimizing the function $\max_{x, x' \in X} f(\varphi)$ over all bijections φ from X to Y . Here $f(\varphi) := \max_{x, x'} |\omega_X(x, x') - \omega_Y(\varphi(x), \varphi(x'))|$. However, this is an instance of an NP-hard problem known as the *quadratic bottleneck assignment problem* [95]. The structure of the optimization problem induced by $d_{\mathcal{N}}$ is very similar to that of $\widehat{d}_{\mathcal{N}}$, so it seems plausible that computing $d_{\mathcal{N}}$ would be NP-hard as well. This intuition is confirmed in Theorem 178. We remark that similar results were obtained for the Gromov-Hausdorff distance by F. Schmiedl in his PhD thesis [105].

Theorem 178. *Computing $d_{\mathcal{N}}$ is NP-hard.*

Proof. To obtain a contradiction, assume that $d_{\mathcal{N}}$ is not NP-hard. Let $X, Y \in \mathcal{FN}^{\text{dis}}$ such that $\text{card}(X) = \text{card}(Y)$. We write $\mathcal{R}(X, Y) = R_B \sqcup R_N$, where R_B consists of correspondences for which the projections π_X, π_Y are injective, thus inducing a bijection between X and Y , and $R_N = \mathcal{R}(X, Y) \setminus R_B$. Note that for any $R \in R_N$, there exist x, x', y such that $(x, y), (x', y) \in R$, or there exist x, y, y' such that $(x, y), (x, y') \in R$. Define $\Psi : \mathbb{R} \rightarrow \mathbb{R}$ by:

$$\begin{aligned} \Psi(\zeta) &= \begin{cases} \zeta + C & : \zeta \neq 0 \\ 0 & : \zeta = 0 \end{cases}, \text{ where} \\ C &= \max_{R \in \mathcal{R}(X, Y)} \text{dis}(R) + 1. \end{aligned}$$

For convenience, we will write $\Psi(X), \Psi(Y)$ to mean $(X, \Psi \circ \omega_X)$ and $(Y, \Psi \circ \omega_Y)$ respectively. We will also write:

$$\text{dis}_{\Psi}(R) := \max_{(x, y), (x', y') \in R} |\Psi(\omega_X(x, x')) - \Psi(\omega_Y(y, y'))|.$$

Consider the problem of computing $d_{\mathcal{N}}(\Psi(X), \Psi(Y))$. First observe that for any $R \in R_B$, we have $\text{dis}(R) = \text{dis}_{\Psi}(R)$. To see this, let $R \in R_B$. Let $(x, y), (x', y') \in R$, and note that $x \neq x', y \neq y'$. Then:

$$|\Psi(\omega_X(x, x')) - \Psi(\omega_Y(y, y'))| = |\omega_X(x, x') + C - \omega_Y(y, y') - C| = |\omega_X(x, x') - \omega_Y(y, y')|.$$

Since $(x, y), (x', y')$ were arbitrary, it follows that $\text{dis}(R) = \text{dis}_{\Psi}(R)$. This holds for all $R \in R_B$.

On the other hand, let $R \in R_N$. By a previous observation, we assume that there exist x, x', y such that $(x, y), (x', y) \in R$. For such a pair, we have:

$$|\Psi(\omega_X(x, x')) - \Psi(\omega_Y(y, y))| = |\omega_X(x, x') + C - 0| \geq \max_{S \in \mathcal{R}(X, Y)} \text{dis}(S) + 1.$$

It follows that $\text{dis}_{\Psi}(R) > \text{dis}_{\Psi}(S)$, for any $S \in R_B$. Hence:

$$\begin{aligned} d_{\mathcal{N}}(\Psi(X), \Psi(Y)) &= \frac{1}{2} \min_{R \in \mathcal{R}(X, Y)} \text{dis}_{\Psi}(R) \\ &= \frac{1}{2} \min_{R \in R_B} \text{dis}_{\Psi}(R) \\ &= \frac{1}{2} \min_{R \in R_B} \text{dis}(R) \\ &= \frac{1}{2} \min_{\varphi} \text{dis}(\varphi), \text{ where } \varphi \text{ ranges over bijections } X \rightarrow Y \\ &= \widehat{d}_{\mathcal{N}}(X, Y). \end{aligned}$$

It is known (see Remark 179 below) that computing $\widehat{d}_{\mathcal{N}}$ is NP-hard. But the preceding calculation shows that $\widehat{d}_{\mathcal{N}}$ can be computed through $d_{\mathcal{N}}$, which, by assumption, is not NP-hard. This is a contradiction. Hence $d_{\mathcal{N}}$ is NP-hard. \square

Remark 179. We can be more precise about why computing $\widehat{d}_{\mathcal{N}}$ is a case of the QBAP. Let $X = \{x_1, \dots, x_n\}$ and let $Y = \{y_1, \dots, y_n\}$. Let Π denote the set of all $n \times n$ permutation matrices. Note that any $\pi \in \Pi$ can be written as $\pi = ((\pi_{ij}))_{i,j=1}^n$, where each $\pi_{ij} \in \{0, 1\}$. Then $\sum_j \pi_{ij} = 1$ for any i , and $\sum_i \pi_{ij} = 1$ for any j . Computing $\widehat{d}_{\mathcal{N}}$ now becomes:

$$\widehat{d}_{\mathcal{N}}(X, Y) = \frac{1}{2} \min_{\pi \in \Pi} \max_{1 \leq i, k, j, l \leq n} \Gamma_{ijkl} \pi_{ij} \pi_{kl}, \text{ where } \Gamma_{ijkl} = |\omega_X(x_i, x_k) - \omega_Y(y_j, y_l)|.$$

This is just the QBAP, which is known to be NP-hard [18].

4.2 Computing lower bounds for $d_{\mathcal{N}}$

In this section we first discuss some algorithmic details on how to compute the lower bounds for $d_{\mathcal{N}}$ involving local spectra and then present some computational examples. All networks in this section are assumed to be finite. Our software and datasets are available on <https://github.com/fmemoli/PersNet> as part of the PersNet software package.

4.2.1 An algorithm for computing minimum matchings

Lower bounds for $d_{\mathcal{N}}$ involving the comparison of local spectra of two networks such as those in Proposition 147 require computing the minimum of a functional $J(R) := \max_{(x,y) \in R} C(x, y)$ where $C : X \times Y \rightarrow \mathbb{R}_+$ is a given *cost* function and R ranges in $\mathcal{R}(X, Y)$. This is an instance of a bottleneck linear assignment problem (or LBAP) [18]. We remark that the current instance differs from the standard formulation in that one is now optimizing over correspondences and not over permutations. Hence the standard algorithms need to be modified.

Assume $n = \text{card}(X)$ and $m = \text{card}(Y)$. In this section we adopt matrix notation and regard R as a matrix $((r_{i,j})) \in \{0, 1\}^{n \times m}$. The condition $R \in \mathcal{R}(X, Y)$ then requires that $\sum_i r_{i,j} \geq 1$ for all j and $\sum_j r_{i,j} \geq 1$ for all i . We denote by $C = ((c_{i,j})) \in \mathbb{R}_+^{n \times m}$ the matrix representation of the cost function C described above. With the goal of identifying a suitable algorithm, the key observation is that the optimal value $\min_{R \in \mathcal{R}} J(R)$ must coincide with a value realized in the matrix C .

An algorithm with complexity $O(n^2 \times m^2)$ is the one in Algorithm 1 (we give it in Matlab pseudo-code). The algorithm belongs to the family of *thresholding algorithms* for solving matching problems over permutations, see [18]. Notice that R is a binary matrix and that procedure **TestCorrespondence** has complexity $O(n \times m)$. In the worst case, the matrix C has $n \times m$ distinct entries, and the **while** loop will need to exhaustively test them all, hence the claimed complexity of $O(n^2 \times m^2)$. Even though a more efficient version (with complexity $O((n \times m) \log(n \times m))$) can be obtained by using a bisection strategy on the range of possible values contained in the matrix C (in a manner similar to what is described for the case of permutations in [18]), here for clarity we limit our presentation to the version detailed above.

4.2.2 Computational example: randomly generated networks

As a first application of our ideas we generated a database of weighted directed networks with different numbers of “communities” and different total cardinalities using the software provided by [54]. Using this software, we generated 35 random networks as follows: 5 networks with 5 communities and 200 nodes each (class c5-n200), 5 networks with 5 communities and 100 nodes each (class c5-n100), 5 networks with 4 communities and 128 nodes each (class c4-n128), 5 networks with 2 communities and 20 nodes each (class c2-n20), 5 networks with 1 community and 50 nodes each (class c1-n50), and 10 networks with 1 community and 128 nodes each (class c1-n128). In order to make the comparison more realistic, as a preprocessing step we divided all the weights in each network by the diameter of the network. In this manner, discriminating between networks requires differentiating their structure and not just the scale of the weights. Note that the (random) weights produced by the software [54] are all non-negative.

Algorithm 1 MinMax matching

```

1: procedure MINMAXMATCH(C)
2:   v = sort(unique(C(:)));
3:   k = 1;
4:   while ~done do
5:     c = v(k);
6:     R = (C <= c);
7:     done = TESTCORRESPONDENCE(R);
8:     k = k + 1;
9:   end while
10:  return c
11: end procedure
12: procedure TESTCORRESPONDENCE(R)
13:   done = prod(sum(R))*prod(sum(R')) > 0;
14:   return done
15: end procedure

```

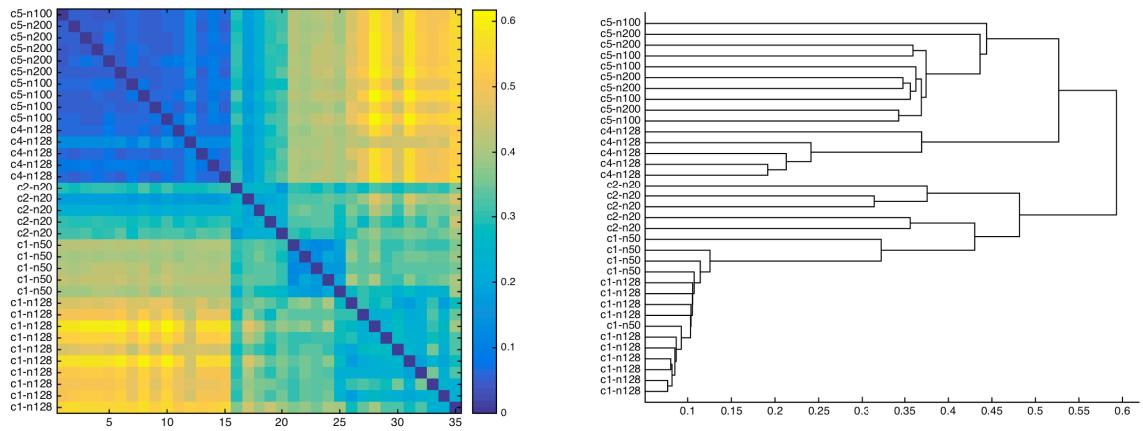


Figure 4.1: **Left:** Lower bound matrix arising from matching local spectra on the database of community networks. **Right:** Corresponding single linkage dendrogram. The labels indicate the number of communities and the total number of nodes. Results correspond to using local spectra as described in Proposition 180.

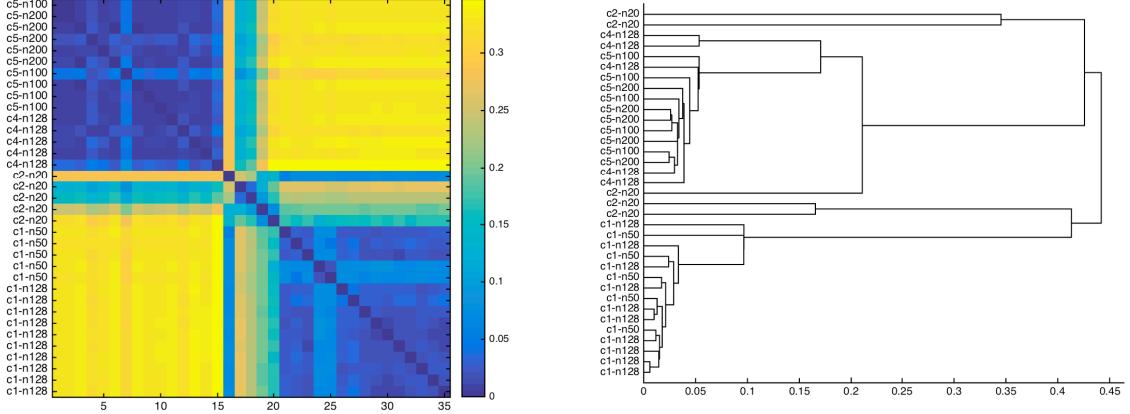


Figure 4.2: **Left:** Lower bound matrix arising from matching global spectra on the database of community networks. **Right:** Corresponding single linkage dendrogram. The labels indicate the number of communities and the total number of nodes. Results correspond to using global spectra as signatures.

Using a Matlab implementation of Algorithm 1 we computed a 35×35 matrix of values corresponding to a lower bound based simultaneously on both the in and out local spectra. This strengthening of Proposition 147 is stated below.

Proposition 180. *For all $X, Y \in \mathcal{FN}$,*

$$d_{\mathcal{N}}(X, Y) \geq \frac{1}{2} \min_{R \in \mathcal{R}} \max_{(x, y) \in R} C(x, y), \text{ where} \\ C(x, y) = \max(d_{\mathcal{H}}^{\mathbb{R}}(\text{spec}_X^{\text{in}}(x), \text{spec}_Y^{\text{in}}(y)), d_{\mathcal{H}}^{\mathbb{R}}(\text{spec}_X^{\text{out}}(x), \text{spec}_Y^{\text{out}}(y))).$$

This bound follows from Proposition 147 by the discussion at the beginning of §2.6.1.

The results are shown in the form of the lower bound matrix and its single linkage dendrogram in Figure 4.1. Notice that the labels in the dendrogram permit ascertaining the quality of the classification provided by the local spectra bound. With only very few exceptions, networks with similar structure (same number of communities) were clustered together regardless of their cardinality. Notice furthermore how networks with 4 and 5 communities merge together before merging with networks with 1 and 2 communities, and vice versa. For comparison, we provide details about the performance of the global spectra lower bound on the same database in Figure 4.2. The results are clearly inferior to those produced by the local version, as predicted by the inequality in Proposition 147.

4.2.3 Computational example: simulated hippocampal networks

To compare with persistent homology methods, we repeated the experiment described in §1.10.2 on a smaller scale using local spectra lower bounds.

In this experiment, there were two environments: (1) a square of side length L , and (2) a square of side length L , with a disk of radius $0.33L$ removed from the center. In what follows, we refer to the environments of the second type as *1-hole environments*, and those of the first type as *0-hole environments*. For each environment, a random-walk trajectory of 5000 steps was generated, where the animal could move above, below, left, or right with equal probability. If one or more of these moves took the animal outside the environment (a disallowed move), then the probabilities were redistributed uniformly among the allowed moves. The length of each step in the trajectory was $0.1L$.

In the first set of 20 trials for each environment, 200 place fields of radius $0.1L$ were scattered uniformly at random. In the next two sets, the place field radii were changed to $0.2L$ and $0.05L$. This produced a total of 60 trials for each environment. For each trial, the corresponding network (X, ω_X) was constructed as follows: X consisted of 200 place cells, and for each $1 \leq i, j \leq 200$, the weight $\omega_X(x_i, x_j)$ was given by:

$$\omega_X(x_i, x_j) = 1 - \frac{\# \text{ times cell } x_j \text{ spiked in a window of five time units after cell } x_i \text{ spiked}}{\# \text{ times cell } x_j \text{ spiked}}.$$

The results of applying the local spectra lower bound are shown in Figure 4.3. The labels `env-0`, `env-1` correspond to 0 and 1-hole environments, respectively.

As a final remark, we note that at least for this experiment, it appears that superior results were obtained using Dowker persistent homology, as shown in §1.10.2.

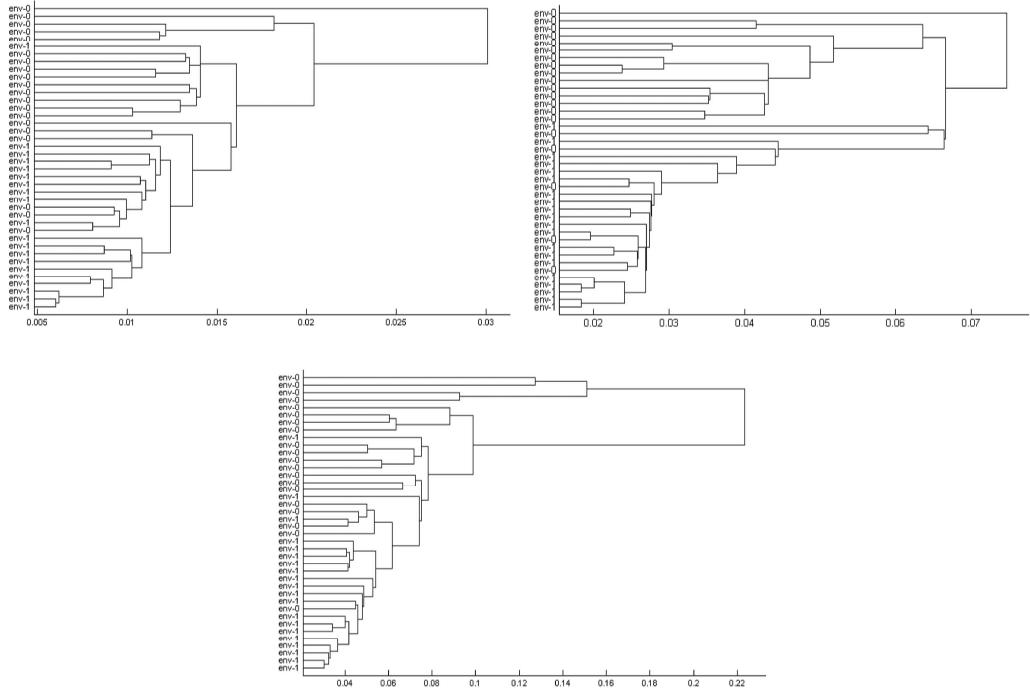


Figure 4.3: Single linkage dendrogram based on local spectrum lower bound of Proposition 180 corresponding to hippocampal networks with place field radii $0.2L$, $0.1L$, and $0.05L$ (clockwise from top left).

4.3 Lower bounds for $d_{\mathcal{N},p}$

In this section, we discuss the computation of a lower bound obtained by taking the maximum of the outer and inner joint eccentricity bounds in Theorem 127, specifically the bounds given by Inequalities (1.14) and (1.16). In what follows, we write *TLB bound* to mean the maximum of these two bounds. Each TLB is obtained by repeatedly invoking Equation (1.17) and finally solving a linear program. To speed up computation of the linear program, we use *entropic regularization*, as used in [42] and further developed in [10, 111, 100, 119]. As a model on which to run our computations, we develop the notion of a *stochastic block model (SBM) network*, which is based on the popular generative model for random graphs [2].

We first explain the notion of entropic regularization and associated difficulties with numerical stability.

4.3.1 Numerical stability of entropic regularization

Let μ_X, μ_Y be probability measures on sets X, Y with $|X| = m, |Y| = n$. For a general $m \times n$ cost matrix M , one may consider the entropically regularized optimal transport problem below, where $\lambda \geq 0$ is a regularization parameter and H denotes entropy:

$$\inf_{p \in \mathcal{C}(\mu_X, \mu_Y)} \sum_{i,j} M_{ij} p_{ij} - \frac{1}{\lambda} H(p), \quad H(m) = - \sum_{i,j} p_{ij} \log p_{ij}.$$

As shown in [42], the solution to this problem has the form $\text{diag}(a) * K * \text{diag}(b)$, where $K := e^{-\lambda M}$ is a kernel matrix and a, b are nonnegative scaling vectors in $\mathbb{R}^m, \mathbb{R}^n$, respectively. Here $*$ denotes matrix multiplication, and exponentiation is performed elementwise. An approximation to this solution can be obtained by iteratively scaling K to have row and column sums equal to μ_X and μ_Y , respectively. More specifically, the updates are simply:

$$a = \text{ones}(m, 1); \% \text{initialize}, \quad b \leftarrow \frac{\mu_Y}{K' * a}, \quad a \leftarrow \frac{\mu_X}{K * b}.$$

As pointed out in [106, 31, 30], using a large value of λ (corresponding to a small regularization) leads to numerical instability, where values of K can go below machine precision and entries of the scaling updates a, b can also blow up. For example, Matlab will interpret e^{-1000} as 0, which is a problem with even a moderate choice of $\lambda = 200$ and $M_{ij} = 50$. Theoretically, it is necessary to have K be a positive matrix for the Sinkhorn algorithm to converge to the correct output [109, 110].

In [106, 31, 30], the authors proposed a variety of clever methods for stabilizing the structure of the algorithm. One such idea is to incorporate some amount of log-domain computations to stabilize the iterations, i.e. during the iterations, large values of a and b are occasionally “absorbed” into K . This leads to some cancellation with the small values of K so that the resulting matrix \tilde{K} is stabilized. The iterations then continue with the stabilized kernel until another absorption step is required, which stabilizes \tilde{K} even further. Even with this strategy, some entries of K might be zero at initialization. Another strategy described in the preceding works is to start with a conservative value of λ , obtain some scaling updates that are used to stabilize K , and then gradually increase λ to the desired value while further stabilizing the kernel.

As discussed in [30], many entries of the stabilized kernel obtained as above could be below machine precision, but the entries corresponding to those on which the optimal plan is supported are likely to be above the machine limit. Indeed, this sparsity may even be leveraged for additional computational tricks.

The techniques for stabilizing the entropy regularized OT problem are not the focus of our work, but because these considerations naturally arose in our computational experiments, we describe some strategies we undertook that are complementary to the techniques available in the current literature. In order to provide a perspective complementary to that presented in [30], we impose the requirement that *all* entries of the kernel matrix remain above machine precision.

Initializing in the log domain. A simple adaptation of the “log domain absorption” step referred to above yields a “log initialization” method that works well in most cases for initializing K to have values above machine precision. To explain this method, we first present an algorithm (Algorithm 2) for the log domain absorption method. We follow the presentation provided in [30], making notational changes as necessary.

Algorithm 2 Sinkhorn with partial log domain steps

```

procedure SINKHORNLOG( $M, \lambda, mX, mY$ )       $\triangleright M$  an  $m \times n$  cost matrix,  $mX, mY$ 
prob. measures
   $a \leftarrow 1_m, b \leftarrow 1_n$                        $\triangleright$  scaling updates
   $u \leftarrow 0_m, v \leftarrow 0_n$                      $\triangleright$  log domain storage of large  $a, b$ 
   $K_{ij} \leftarrow \exp(\lambda(-M_{ij} + u_i + v_j))$      $\triangleright$  initialize kernel
  while stopping criterion not met do
     $b \leftarrow mB ./ (K'a)$ 
     $a \leftarrow mA ./ (Kb)$ 
    if  $\max(\max(a), \max(b)) > \text{threshold}$  then
       $u \leftarrow u + (1/\lambda) \log(a)$                    $\triangleright$  store  $a, b$  in  $u, v$ 
       $v \leftarrow v + (1/\lambda) \log(b)$ 
       $K_{ij} \leftarrow \exp(\lambda(-M_{ij} + u_i + v_j))$      $\triangleright$  absorb  $a, b$  into  $K$ 
       $a \leftarrow 1_m, b \leftarrow 1_n$                        $\triangleright$  after absorption, reset  $a, b$ 
    end if
  end while
  return  $\text{diag}(a)K \text{diag}(b)$ 
end procedure

```

Notice that in Algorithm 2, K might already have values below machine precision at initialization. To circumvent this, we can add a preprocessing step that yields a stable initialization of K . This is outlined in Algorithm 3. An important point to note about Algorithm 3 is that the user needs to choose a function `decideParam(α, β)` which returns a number γ between α and β , where α and β are as stated in the algorithm. This number γ should be such that $\exp(-\lambda\beta + \lambda\gamma)$ is above machine precision, but $\exp(-\lambda\alpha + \lambda\gamma)$ is not too large. The crux of Algorithm 3 is that by choosing large initial scaling vectors a, b and immediately absorbing them into the log domain, the extreme values of M are canceled out before exponentiation.

A geometric trick in the $p = 2$ case. The preceding initialization method has its limitations: depending on how far $\min(M), \max(M)$ are spread apart, the log initialization step might not be able to yield an initial kernel K that has all entries above machine precision and below the machine limit. This suggests that it would be beneficial to normalize the cost matrix M to control the spread of $\min(M), \max(M)$. However, it is crucial to remember

Algorithm 3 Log domain initialization of K

```

procedure LOGINITIALIZE( $M, \lambda, mX, mY$ )       $\triangleright M$  an  $m \times n$  cost matrix,  $mX, mY$ 
prob. measures
   $\alpha \leftarrow \min(M), \beta \leftarrow \max(M)$            $\triangleright$  scan  $M$  for max and min values
   $\gamma \leftarrow \text{decideParam}(\alpha, \beta)$             $\triangleright \text{decideParam}$  is an independent function
   $a \leftarrow \exp(\lambda\gamma)1_m, b \leftarrow \exp(\lambda\gamma)1_n$ 
   $u \leftarrow (1/\lambda) \log(a), v \leftarrow (1/\lambda) \log(b)$ 
   $K_{ij} \leftarrow \exp(\lambda(-M_{ij} + u_i + v_j))$         $\triangleright K$  is stably initialized
   $a \leftarrow 1_m, b \leftarrow 1_n$ 
  perform rest of SINKHORNLOG(a)s usual
end procedure

```

that OT problems arise in our setting when computing the TLB between two networks, so any normalization would have to be theoretically justified.

It turns out that in the case $p = 2$, the particular geometry of $(\mathcal{N}_m, d_{\mathcal{N},p})$ allows for an elegant normalization scheme. In this particular case, it is possible to use a certain ‘‘cosine rule formula’’ [117] to compute the TLB between rescaled versions of the original networks X and Y , and then rescale the solution to get the TLB between X and Y . We describe this method in detail below. In what follows, we always have $p = 2$ for $d_{\mathcal{N},p}$ unless specified otherwise.

The caveat to this normalization scheme, pointed out to us by Justin Solomon, is that scaling the network weights down in turn requires the λ values to be scaled up, which once again leads to numerical instability. In practice, we have used the following approach. When running a computation over a database of networks which have widely varying weights (such that any fixed choice of λ causes some of the initial kernels to have values above/below machine precision), we rescale all the networks simultaneously using the normalization described below. Then we proceed with the Sinkhorn algorithm, employing log domain absorption steps as needed (and some of the computations will indeed need more of these absorption steps).

Let $(X, \omega_X, \mu_X), (Y, \omega_Y, \mu_Y) \in \mathcal{N}_m$. Recall from Example 109 that $d_{\mathcal{N},2}(X, N_1(0)) = \frac{1}{2} \text{size}_2(X)$. Define $s := \frac{1}{2} \text{size}_2(X, \omega_X, \mu_X)$, $t := \frac{1}{2} \text{size}_2(Y, \omega_Y, \mu_Y)$. Notice also that for an optimal coupling $\mu \in \mathcal{C}(\mu_X, \mu_Y)$, we have:

$$\begin{aligned} d_{\mathcal{N},2}(X, Y)^2 &= \frac{1}{4} \int \int |\omega_X(x, x')|^2 + |\omega_Y(y, y')|^2 - 2\omega_X(x, x')\omega_Y(y, y') d\mu(x, y) d\mu(x', y') \\ &= s^2 + t^2 - \frac{1}{2} \int \int \omega_X(x, x')\omega_Y(y, y') d\mu(x, y) d\mu(x', y'), \end{aligned}$$

where the first equality holds because $|a - b|^2 = |a|^2 + |b|^2 - 2ab$ for all $a, b \in \mathbb{R}$, and the last equality holds because $\omega_X(x, x'), \omega_Y(y, y')$ do not depend on μ_Y, μ_X , respectively.

Sturm [117, Lemma 4.2] observed the following “cosine rule” structure. Define

$$\omega'_X := \frac{\omega_X}{2s}, \quad \omega'_Y := \frac{\omega_Y}{2t}. \quad (4.1)$$

Then $\text{size}_2(X, \omega'_X) = \frac{1}{2s} \text{size}_2(X, \omega_X) = 1 = \frac{1}{2t} \text{size}_2(Y, \omega_Y) = \text{size}_2(Y, \omega'_Y)$. A geometric fact about this construction is that $(X, \omega_X, \mu_X), (Y, \omega_Y, \mu_Y)$ lie on geodesic rays connecting $\mathcal{X} := (X, \omega'_X, \mu_X)$ and $\mathcal{Y} := (Y, \omega'_Y, \mu_Y)$ respectively to $N_1(0)$. Actually, once (X, ω_X, μ_X) and (Y, ω_Y, μ_Y) are chosen, the geodesic rays are automatically defined to be given by their scalar multiples. Then we independently define \mathcal{X} and \mathcal{Y} to be representatives of the weak isomorphism class of networks at $d_{\mathcal{N},2}$ distance $1/2$ from $N_1(0)$ that lie on these geodesics. We illustrate a related situation in Figure 4.4, and refer the reader to [117] for further details. Implicitly using this geometric fact, we fix \mathcal{X}, \mathcal{Y} as above and treat $(X, \omega_X, \mu_X), (Y, \omega_Y, \mu_Y)$ as $2s$ and $2t$ -scalings of \mathcal{X} and \mathcal{Y} , respectively (i.e. such that Equation 4.1 is satisfied). Then we have:

$$\begin{aligned} & 4d_{\mathcal{N},2}((X, \omega_X, \mu_Y), (Y, \omega_X, \mu_Y))^2 - 4s^2 - 4t^2 \\ &= \int \int 4s^2 |\omega'_X(x, x')|^2 + 4t^2 |\omega'_Y(y, y')|^2 - 4s^2 - 4t^2 \\ &\quad - 8st \omega'_X(x, x') \omega'_Y(y, y') d\mu(x, y) d\mu(x', y') \\ &= 4s^2 \text{size}_2(X, \omega'_X)^2 + 4t^2 \text{size}_2(Y, \omega'_Y)^2 - 4s^2 - 4t^2 \\ &\quad - 2st \int \int |\omega'_X(x, x')| |\omega'_Y(y, y')| d\mu(x, y) d\mu(x', y') \\ &= -8st \int \int |\omega'_X(x, x')| |\omega'_Y(y, y')| d\mu(x, y) d\mu(x', y'), \end{aligned}$$

where the last equality holds because $\text{size}_2(X, \omega'_X) = 1 = \text{size}_2(Y, \omega'_Y)$.

Since $(X, \omega_X, \mu_X), (Y, \omega_Y, \mu_Y)$ were $2s, 2t$ -scalings of \mathcal{X} and \mathcal{Y} for arbitrary $s, t > 0$, this shows in particular that the quantity

$$(1/2st) (d_{\mathcal{N},2}((X, \omega_X, \mu_Y), (Y, \omega_Y, \mu_Y))^2 - s^2 - t^2) \quad (\text{cosine rule})$$

depends only on the reference networks \mathcal{X} and \mathcal{Y} , and is independent of s and t .

We leverage the preceding observations to produce a scaling method as follows. Suppose $(X, \omega_X, \mu_X), (Y, \omega_Y, \mu_Y)$ are given. Define $s := \frac{1}{2} \text{size}_2(X, \omega_X, \mu_X)$ and $t := \frac{1}{2} \text{size}_2(Y, \omega_Y, \mu_Y)$. For $\alpha, \beta \geq 1$, define

$$\omega_{\mathcal{X}} := \frac{\omega_X}{\alpha \|\omega_X\|_{\infty}}, \quad \omega_{\mathcal{Y}} := \frac{\omega_Y}{\beta \|\omega_Y\|_{\infty}}.$$

Also define $\mathcal{X} := (X, \omega_{\mathcal{X}}, \mu_X), \mathcal{Y} := (Y, \omega_{\mathcal{Y}}, \mu_Y)$. Then we have

$$\sigma := \frac{1}{2} \text{size}_2(\mathcal{X}) = \frac{s}{\alpha \|\omega_X\|_{\infty}}, \quad \tau := \frac{1}{2} \text{size}_2(\mathcal{Y}) = \frac{t}{\beta \|\omega_Y\|_{\infty}}.$$

Each weight in the support of \mathcal{X}, \mathcal{Y} is in the range $[-1, 1]$. Forming a TLB cost matrix from these rescaled networks and performing Sinkhorn iterations with the rescaled weights

is more stable because the entries in the corresponding kernel matrix K are more likely to be above machine precision. Indeed, for larger values of λ , one can scale down ω_X, ω_Y sufficiently via α and β to ensure that K is well-behaved. The cosine rule can then be used to recover $d_{\mathcal{N},2}(X, Y)$ in terms of $d_{\mathcal{N},2}(\mathcal{X}, \mathcal{Y})$.

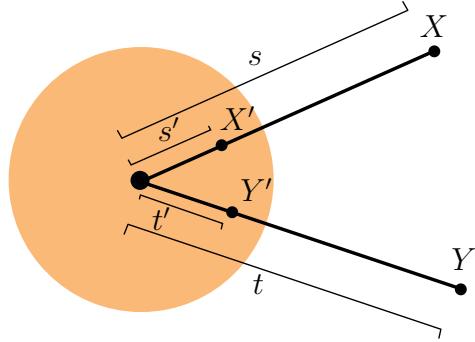


Figure 4.4: Sinkhorn computations for $d_{\mathcal{N},2}(\mathcal{X}, \mathcal{Y})$ are carried out in the “stable region” for K , and the end result is rescaled to recover $d_{\mathcal{N},2}(X, Y)$.

This geometric idea is illustrated in Figure 4.4. The spaces $(X, \omega_X), (X, \omega'_X), \mathcal{X}$ all live on a geodesic ray emanating from $N_1(0)$, and likewise for Y . See [117] for more details about the geodesic structure of gauged measure spaces; the analogous results hold for $(\mathcal{N}_m, d_{\mathcal{N},2})$.

From the preceding observation, we have:

$$\begin{aligned} \frac{1}{2st} (d_{\mathcal{N},2}(X, Y)^2 - s^2 - t^2) &= \frac{1}{2\sigma\tau} (d_{\mathcal{N},2}(\mathcal{X}, \mathcal{Y})^2 - \sigma^2 - \tau^2) \\ d_{\mathcal{N},2}(X, Y)^2 - s^2 - t^2 &= \alpha\beta\|\omega_X\|_\infty\|\omega_Y\|_\infty \left(d_{\mathcal{N},2}(\mathcal{X}, \mathcal{Y})^2 - \frac{s^2}{\alpha^2\|\omega_X\|_\infty^2} - \frac{t^2}{\beta^2\|\omega_Y\|_\infty^2} \right). \end{aligned}$$

The final step is summarized in the following

Lemma 181. *Let $X, Y, \mathcal{X}, \mathcal{Y}, \alpha, \beta, s, t$ be as above. Then,*

$$d_{\mathcal{N},2}(X, Y)^2 = \alpha\beta\|\omega_X\|_\infty\|\omega_Y\|_\infty d_{\mathcal{N},2}(\mathcal{X}, \mathcal{Y})^2 - \frac{s^2\beta\|\omega_Y\|_\infty}{\alpha\|\omega_X\|_\infty} - \frac{t^2\alpha\|\omega_X\|_\infty}{\beta\|\omega_Y\|_\infty} + s^2 + t^2.$$

Remark 182. From the perspective of computations, the preceding lemma should be interpreted as follows. The quantities $s, t, \|\omega_X\|_\infty, \|\omega_Y\|_\infty$ are all easy to compute. One can either attempt to obtain a local minimizer for the $d_{\mathcal{N},2}(\mathcal{X}, \mathcal{Y})$ -functional (e.g. following [111]), or to obtain a TLB-type lower bound for $d_{\mathcal{N},2}(\mathcal{X}, \mathcal{Y})$, as in the current work. In either case, the output can be rescaled by the formula in the lemma to approximate (or lower-bound) $d_{\mathcal{N},2}(X, Y)^2$.

Remark 183. In our experiments, it was best to fix $\alpha, \beta = 1$. This cosine rule method works best when the networks in question have large sizes; if the networks are provided in normalized form with weights in a small range (e.g. in $[0,1]$), then it is better to transfer some of the computations to the log domain using Algorithms 2 and 3.

Zeros on the diagonal for $\text{TLB}(A, A)$ and adaptive λ -search. When comparing a network to itself, we expect to get a TLB output of zero. The corresponding optimal coupling should be the diagonal coupling. However, some care needs to be taken to achieve this during computation, because entropic regularization produces “smoothed” couplings by design, and the diagonal coupling is highly nonsmooth. We now briefly explain a simple heuristic we used to achieve these nonsmooth couplings.

Suppose we are comparing a matrix A to itself via the TLB. The corresponding cost matrix M should have zeros on the diagonal, which translates to a kernel matrix K with 1s on the diagonal. When M has all 0s on the diagonal and values strictly above 0 everywhere else, and the two marginals are equal, then the optimal coupling should be the diagonal coupling: the optimal transport plan is to not transport anything at all. To achieve this via Sinkhorn iterations, we noticed that the off-diagonal entries in each column of K needed to be several orders of magnitude below 1. For our desired precision, three orders of magnitude were sufficient. Since orders of separation in K are easily related to differences of values in M , we performed the following procedure: for each column j of M , we computed the minimal difference between values in the column, and computed λ_j so that after computing $K = e^{-\lambda_j M}$, the entries in column j of K would be separated by at least three orders of magnitude. Thus we obtained a pool of λ values. From this list, we used a binary search to pick the largest λ that would not cause entries of K to go below machine precision.

While this approach naturally suggests using log domain initialization (as in Algorithm 3) to choose the largest possible λ , we did not use any log domain computations so that we could independently observe the behavior of this simple heuristic. However, to ensure that at least one λ in the pool would work without causing entries of K to go below machine precision, we preprocessed the networks via the cosine rule normalization strategy used above and inserted a moderate value of $\lambda = 200$ into the pool of λ values.

We used this adaptive λ -search heuristic when computing the TLB for any pair of networks, not just the TLB of a network with itself (which was the motivation for this heuristic). For an illustration of the operation of this method, see the TLB dissimilarity matrix corresponding to Experiment 1.10.3 in Figure 1.25.

4.4 Complexity of PPH and algorithmic aspects

The origin of a general persistent homology algorithm for simplicial complexes can be traced back to [49] for $\mathbb{Z}/2\mathbb{Z}$ coefficients, and to [123] for arbitrary field coefficients. Here it was observed that the persistence algorithm has the same running time as Gaussian elimination over fields, i.e. $O(m^3)$ in the worst case, where m is the number of simplices.

Algorithm 4 TLB computation with cosine rule normalization

```
procedure GETCOSINETLB( $X, Y, mX, mY$ )       $\triangleright X$  is  $n \times n$ ,  $Y$  is  $m \times m$ ,  $mX, mY$ 
prob. measures
     $A \leftarrow X ./ \max(\text{abs}(X))$ ,  $B \leftarrow Y ./ \max(\text{abs}(Y))$   $\triangleright ./$  denotes elementwise division
     $\rho \leftarrow \text{GETTLB}(A, B, mX, mY)$ 
     $s \leftarrow (1/2)\text{GETSIZE}(A, mX)$ ,  $t \leftarrow (1/2)\text{GETSIZE}(B, mY)$ 
    perform scaling of  $\rho$  using  $s, t$  as in Lemma 181, save as  $\pi$ 
    return  $\pi$ 
end procedure
procedure GETSIZE( $A, mA$ )                       $\triangleright$  Get the 2-size of a network
     $\sigma \leftarrow \sqrt{mA'(A \wedge 2)mA}$ 
    return  $\sigma$ 
end procedure
procedure GETTLB( $A, B, mA, mB$ )                 $\triangleright$  get TLB over  $\mathbb{R}$  with  $p = 2$ 
    for  $1 \leq i \leq n$  and  $1 \leq j \leq m$  do
         $vAout \leftarrow A(i, :)$ ,  $vBout \leftarrow B(j, :)$            $\triangleright$  get both  $\text{ecc}^{\text{out}}$  and  $\text{ecc}^{\text{in}}$ 
         $vAin \leftarrow A(:, i)$ ,  $vBin \leftarrow B(:, j)$ 
         $Cout(i, j) \leftarrow \text{COMPAREDISTRIBUTIONS}(vAout, vBout, mA, mB)$ 
         $Cin(i, j) \leftarrow \text{COMPAREDISTRIBUTIONS}(vAin, vBin, mA, mB)$ 
    end for
    perform Sinkhorn iterations for OT problem with  $Cout, Cin$  as cost matrices
    store results in  $tlbIn, tlbOut$ 
    return  $\max(tlbIn, tlbOut)$   $\triangleright$  both are valid lower bounds for  $d_{\mathcal{N},2}$ , so take the max
end procedure
procedure COMPAREDISTRIBUTIONS( $vA, vB, mA, mB$ )
     $\gamma \leftarrow$  use Equation (1.17) to get 2-Wasserstein distance between
    pushforward distributions over  $\mathbb{R}$  induced by  $vA$  and  $vB$ 
    return  $\gamma$ 
end procedure
```

The PPH setting is more complicated, due to two reasons: (1) because of directionality, the number of p -paths on a vertex set is much larger than the number of p -simplices, for any $p \in \mathbb{N}$, and (2) one must first obtain bases for the ∂ -invariant p -paths $\{\Omega_p : p \geq 2\}$. The first item is unavoidable, and even desirable—we capture the asymmetry in the data, thus retaining more information. For the second item, note that Ω_0 and Ω_1 are just the allowed 0 and 1-paths, so their bases can just be read off from the network weight function. After obtaining compatible bases for the filtered chain complex $\{\Omega_\bullet^i \rightarrow \Omega_\bullet^{i+1}\}_{i \in \mathbb{N}}$, however, one can use the general persistent homology algorithm [49, 123, 29]. By *compatible* bases, we mean a set of bases $\{B_p^i \subseteq \Omega_p^i : 0 \leq p \leq D + 1, i \in \mathbb{N}\}$ such that $B_p^i \subseteq B_p^{i+1}$ for each i , and relative to which the transformation matrices M_p of ∂_p are known. Here D is the dimension up to which we compute persistence.

We now present a procedure for obtaining compatible bases for the ∂ -invariant paths. Fix a network (X, A_X) . We write \mathcal{R}_p to denote $\mathcal{R}_p(X, \mathbb{K})$, for each $p \in \mathbb{Z}_+$. Given a digraph filtration on X , we obtain a filtered vector space $\{\mathcal{A}_\bullet^i \rightarrow \mathcal{A}_\bullet^{i+1}\}_{i=1}^N$ and a filtered chain complex $\{\Omega_\bullet^i \rightarrow \Omega_\bullet^{i+1}\}_{i=1}^N$ for some $N \in \mathbb{N}$. For any p -path v , define its *allow time* as $\text{at}(v) := \min\{k \geq 0 : v \in \mathcal{A}_p^k\}$. Similarly define its *entry time* as $\text{et}(v) := \min\{k \geq 0 : v \in \Omega_p^k\}$. The allow time and entry time coincide when $p = 0, 1$, but are not necessarily equal in general. In Figure 1.13, for example, we have $\text{at}(x_4x_1x_2) = 1 < 2 = \text{et}(x_4x_1x_2)$.

Now fix $p \geq 2$, and consider the map $\partial_p : \mathcal{R}_p \rightarrow \mathcal{R}_{p-1}$. Let M_p denote the matrix representation of ∂_p , relative to an arbitrary choice of bases B_p and B_{p-1} for \mathcal{R}_p and \mathcal{R}_{p-1} . For convenience, we write the bases as $B_p = \{v_i^p : 1 \leq i \leq \dim(\mathcal{R}_p)\}$ and $B_{p-1} = \{v_i^{p-1} : 1 \leq i \leq \dim(\mathcal{R}_{p-1})\}$, respectively. Each basis element has an allow time that can be computed efficiently, and the allow times belong the set $\{1, 2, \dots, N\}$. By performing row and column swaps as needed, we can arrange M_p so that the basis vectors for the domain are in increasing allow time, and the basis vectors for the codomain are in decreasing allow time. This is illustrated in Figure 4.5.

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

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

In words, $hr(i)$ lists the codomain basis vectors that have allow time i . Next we transform M_p into a *column echelon form* $M_{p,G}$, using left-to-right Gaussian elimination. In this form, all nonzero columns are to the left of any zero column, and the leading coefficient (the topmost nonzero element) of any column is strictly above the leading coefficient of the column on its right. The leading coefficients are usually called *pivots*. An illustration of $M_{p,G}$ is provided in Figure 4.5. To obtain this column echelon form, the following elementary column operations are used:

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

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

Remark 184. In the Gaussian elimination step above, we only eliminate entries by adding paths that have already been allowed in the filtration. This means that for any operation of the form $v_j^p \leftarrow v_j^p - kv_i^p$, we must have $\text{at}(v_i^p) \leq \text{at}(v_j^p)$. Thus $\text{at}(v_j^p - kv_i^p) = \text{at}(v_j^p)$. It follows that the allow times of the domain basis vectors do not change as we pass from M_p to $M_{p,G}$, i.e. M_p and $M_{p,G}$ have the same number of domain basis vectors corresponding to any particular allow time.

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

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

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

Now we arrive at the contradiction. Since $\tilde{v} \notin B_{p,G}^i$, we must have either $\text{at}(\tilde{v}) > i$, or $\text{at}(\partial_p(\tilde{v})) > i$. Suppose first that $\text{at}(\tilde{v}) > i$. Since $\tilde{u} \in \Omega_p^i$, we must have $\text{et}(\tilde{u}) \leq i$, and so $\text{at}(\tilde{u}) \leq i$. By the way in which we sorted $M_p^{\tilde{u}}$, we know that u is obtained by adding terms from $B_{p,G}^i$ to \tilde{u} . Each term in $B_{p,G}^i$ has allow time $\leq i$, so $\text{at}(u) \leq i$ by Remark 184. But then B_p^u has one more basis vector with allow time $\leq i$ than B_p , i.e. one fewer basis vector with allow time $> i$. This is a contradiction, because taking linear combinations of linearly independent vectors to arrive at B_p^u can only increase the allow time. Next suppose that $\text{at}(\partial_p(\tilde{v})) > i$. Then, because $M_{p,G}$ is already reduced, the column of \tilde{v} has a pivot at a height that does *not* belong to $hr(i)$. Now consider $\partial_p(u)$. Suppose first that $\partial_p(u) = 0$. Then the column of u clearly does not have a pivot, and it does not affect the pivots of the columns to its right in M_p^u . Thus M_p^u has one fewer pivot than $M_{p,G}$, which is a contradiction because both matrices have the same reduced column echelon form and hence the same pivot positions. Finally, suppose $\partial_p(u) \neq 0$. Since u is obtained from \tilde{u} by reduction, we also have $\text{at}(\partial_p(u)) \leq \text{at}(\partial_p(\tilde{u})) \leq i$. Thus M_p^u has one more pivot at height range i than $M_{p,G}$, which is again a contradiction. Thus $B_{p,G}^i$ spans Ω_p^i . Since $1 \leq i \leq N$ was arbitrary, the result follows.

Data structure Our work shows that left-to-right column reduction is sufficient to obtain compatible bases for the filtered chain complex $\{\Omega_\bullet^i \rightarrow \Omega_\bullet^{i+1}\}_{i=1}^N$. As shown in [123], this is precisely the operation needed in computing persistence intervals, so we can compute PPH with little more work. It is known that there are simple ways to optimize the left-to-right persistence computation [29, 8], but in this paper we follow the classical treatment. Following [49, 123], our data structure is a linear array T labeled by the elementary regular p -paths, $0 \leq p \leq D + 1$, where D is the dimension up to which homology is computed. For completeness, we show below how to modify the algorithms in [123] to obtain PPH.

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

Compare this with the problem of producing simplicial complexes from networks, and then computing simplicial persistent homology. For a network on n nodes, assume that the simplicial filtration is such that every D -simplex on n points eventually enters the filtration (see [37] for such filtrations). The number of D -simplices over n points is $\binom{n}{D+1}$, which is of the same order as n^{D+1} . Thus computing simplicial persistent homology in dimension $D - 1$ via such a filtration (using the general algorithm of [123]) still has complexity $O(n^{3+3D})$.

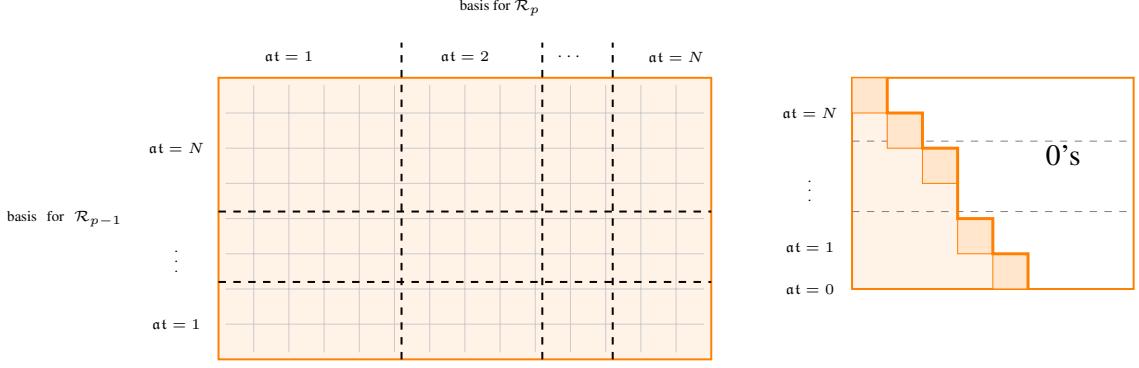


Figure 4.5: **Left:** The rows and columns of M_p are initially arranged so that the domain and codomain vectors are in increasing and decreasing allow time, respectively. If there are no domain (codomain) vectors having a particular allow time, then the corresponding vertical (horizontal) strip is omitted. **Right:** After converting to column echelon form, the domain vectors of $M_{p,G}$ need not be in the original ordering. But the codomain vectors are still arranged in decreasing allow time.

4.4.1 The modified algorithm

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

Below we present a modified version of the algorithm in [123] that computes PPH. We make one last remark, based on an observation in [123]: because of the relation $\partial_p \circ \partial_{p+1}$, a pivot column of the reduced boundary matrix $M_{p,G}$ corresponds to a zero row in $M_{p+1,G}$. Thus whenever we compute $\partial_p(v)$ in our algorithm, we can immediately remove the summands that correspond to pivot terms in $M_{p-1,G}$. This is done in Algorithm 6.

Algorithm 5 Computing persistent path homology

```
1: procedure COMPUTEPPH( $\mathcal{X}, D + 1$ )  $\triangleright$  Compute PPH of network  
    $\mathcal{X}$  up to dimension  $D$   
2:   for  $p = 0, \dots, D$  do  
3:      $\text{Pers}_p = \emptyset$ ;  $\triangleright$  Store intervals here  
4:     for  $j = 1, \dots, \dim(\mathcal{R}_{p+1})$  do  
5:        $[u, i, et] = \text{BASISCHANGE}(v_j^{p+1}, p + 1);$   
6:       if  $u = 0$  then Mark  $T_{p+1}[j]$ ;  
7:       else  
8:          $T_p[i] \leftarrow (u, et);$   
9:         Add  $(\text{et}(v_i^p), et)$  to  $\text{Pers}_p$ ;  
10:        end if  
11:      end for  
12:      for  $j = 1, \dots, \dim(\mathcal{R}_p)$  do  
13:        if  $T_p[j]$  marked and empty then  
14:          Add  $(\text{et}(v_j^p), \infty)$  to  $\text{Pers}_p$ ;  
15:        end if  
16:      end for  
17:    end for  
18:    return  $\text{Pers}_0, \dots, \text{Pers}_D$ ;  
19: end procedure
```

Algorithm 6 Left-to-right column reduction

```
1: procedure BASISCHANGE( $v, \dim$ )  $\triangleright$  Find pivot or zero columns  
2:    $p \leftarrow \dim; u = \partial_p(v)$ ; Remove unmarked (pivot) terms from  $u$ ;  
3:   while  $u \neq 0$  do  
4:      $\sigma \leftarrow \text{argmax}\{\text{index}(\tau) : \tau \text{ is a summand of } u\};$   
5:      $i \leftarrow \text{index}(\sigma);$   
6:      $et \leftarrow \max(\text{at}(v), \text{at}(\sigma));$   
7:     if  $T_{p-1}[i]$  is unoccupied then break;  
8:     end if  
9:     Let  $a, b$  be coefficients of  $\sigma$  in  $T_{p-1}[i]$  and  $u$ , respectively;  
10:     $u \leftarrow u - (a/b)T_{p-1}[i];$   $\triangleright$  Column reduction step  
11:  end while  
12:  return  $u, i, et$ ;  
13: end procedure
```

4.5 More experiments using Dowker persistence

4.5.1 U.S. economy input-output accounts

Each year, the U.S. Department of Commerce Bureau of Economic Analysis (www.bea.gov) releases the U.S. input-output accounts, which show the commodity inputs used by each industry to produce its output, and the total commodity output of each industry. Economists use this data to answer two questions: (1) what is the total output of the US economy, and (2) what is the process by which this output is produced and distributed [72].

One of the core data types in these accounts is a “make” table, which shows the production of commodities by industries. The industries are labeled according to the North American Industry Classification System (NAICS), and the commodities (i.e. goods or services) produced by each industry are also labeled according to the NAICS. This make table can be viewed as a network (E, m) consisting of a set of NAICS labels E , and a function $m : E \times E \rightarrow \mathbb{Z}_+$. Note that the same labels are used to denote industries and commodities. After fixing an enumeration (e_1, \dots, e_n) of E , the entry $m(e_i, e_j)$ corresponds to the dollar value (in millions) of commodity type e_j produced by industry e_i . For example, if e corresponds to the economic sector “Farms,” then $m(e, e)$ is the dollar value (in millions) of farming commodities produced by the farming industry.

In our next example, we analyze make table data from the U.S. Department of Commerce for the year 2011. Specifically, we begin with a set E of 71 economic sectors, and view it as a network by the process described above. We remark that a complementary data set, the “use” table data for 2011, has been analyzed thoroughly via hierarchical clustering methods in [20].

Our analysis is motivated by question (2) described above, i.e. what is the process by which commodities are produced and distributed across economic sectors. Note that one can simply read off values from the make table that show *direct* investment between commodities and industries, i.e. which commodities are being produced by which industry. A more interesting question is to find patterns of *indirect* investment, e.g. chains (e_i, e_j, e_k) where industry e_i produces commodities of type e_j , and industry e_j in turn produces commodities of type e_k . Note that if e_i does not produce any commodities of type e_k , then this indirect influence of e_i on e_k is not immediately apparent from the make table. One can manually infer this kind of indirect influence by tracing values across a make table, but this process can become cumbersome when there are large numbers of economic sectors, and when one wants to find chains of greater length.

To automate this process so that finding flows of investment can be used for exploratory data analysis, we take the viewpoint of using persistent homology. Beginning with the 71×71 make table matrix, we first obtained a matrix $\overline{\omega}_E$ defined by:

$$\overline{\omega}_E(e_i, e_j) := \begin{cases} m(e_i, e_j) & : i \neq j \\ 0 & : i = j. \end{cases}, \text{ for each } 1 \leq i, j \leq 71.$$

Because our goal was to analyze the interdependence of industries and the flow of commodities across industrial sectors, we removed the diagonal as above to discount the commodities produced by each industry in its own type. Next we defined a network (E, ω_E) , where ω_E was given by:

$$\omega_E(e_i, e_j) = f\left(\frac{\bar{\omega}_E(e_i, e_j)}{\sum_{e \in E} \bar{\omega}_E(e, e_j)}\right) \text{ for each } 1 \leq i, j \leq 71.$$

Here $f(x) = 1 - x$ is a function used to convert the original similarity network into a dissimilarity network. The greater the dissimilarity, the weaker the investment, and vice versa. So if $\omega_E(e, e') = 0.85$, then sector e is said to make an investment of 15% on sector e' , meaning that 15% of the commodities of type e' produced externally (i.e. by industries other than e') are produced by industry e . After this preprocessing step, we computed the 0 and 1-dimensional Dowker persistence diagrams of the resulting network. The corresponding barcodes are presented in Figure 4.6, and our interpretation is given below.

Dependent sectors

We open our discussion with the 0-dimensional Dowker persistence barcode presented in Figure 4.6. Recall that Javaplex produces representative 0-cycles for each persistence interval in this 0-dimensional barcode. Typically, these representative 0-cycles are given as the boundary of a 1-simplex, so that we know which pair of sectors merges together into a 1-simplex and converts the 0-cycle into a 0-boundary. We interpret the representative sectors of the shortest 0-dimensional persistence intervals as pairs of sectors where one is strongly dependent on the other. To justify this interpretation, we observe that the right endpoint of a 0-dimensional persistence interval corresponds to a resolution δ at which two industries e, e' find a common δ -sink. Typically this sink is one of e or e' , although this sink is allowed to be a third industry e'' . We suggest the following interpretation for being a common δ -sink: of all the commodities of type e'' produced by industries other than e'' , over $(1 - \delta) * 100\%$ is produced by each of the industries e, e' . Note that for $\delta < 0.50$, this interpretation suggests that e'' is actually e' (or e), and that over 50% of the commodities of type e' produced by external industries (i.e. by industries other than e') are actually produced by e (resp. e').

In Table 4.1 we list some sample representative 0-cycles produced by Javaplex. Note that these cycles are representatives for bars that we can actually see on the 0-dimensional barcode in Figure 4.6. We do not focus any more on finding dependent sectors and direct investment relations, and point the reader to [20] where this topic has been covered in great detail under the lens of hierarchical clustering (albeit with a slightly different dataset, the “use” table instead of the make table). In the following subsection, we study some of the persistent 1-dimensional intervals shown in Figure 4.6, specifically the two longest bars that we have colored in red.

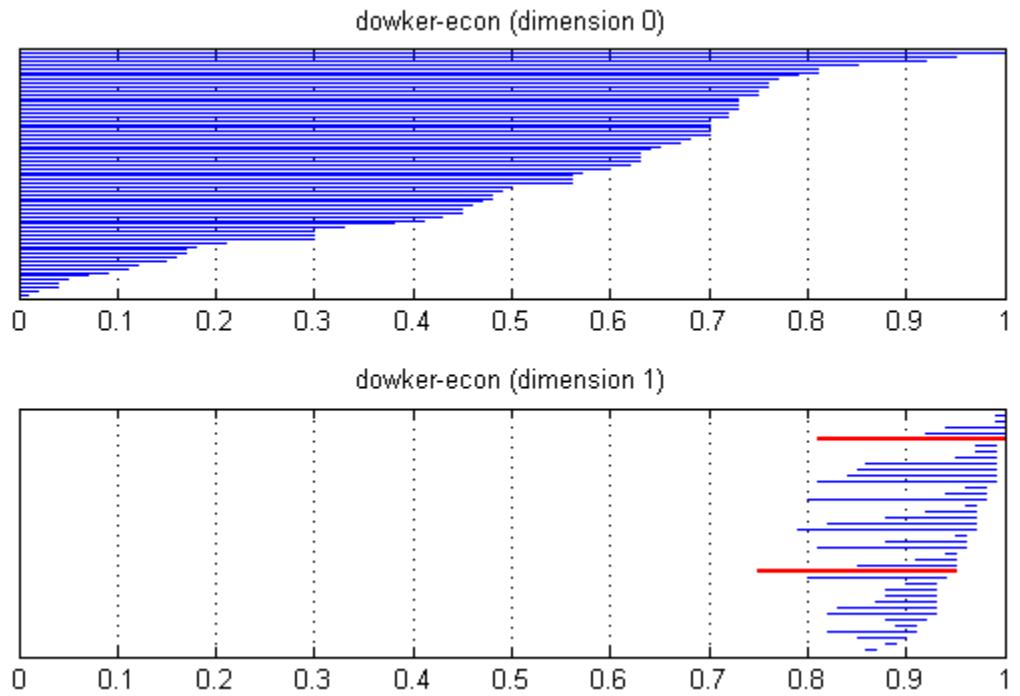


Figure 4.6: 0 and 1-dimensional Dowker persistence barcodes for US economic sector data, obtained by the process described in §4.5.1. The long 1-dimensional persistence intervals that are colored in red are examined in §4.5.1 and Figures 4.9, 4.12.

Sample 0-dimensional persistence intervals from economy data			
Interval $[0, \delta]$	Representative 0-cycle	Labels	δ -sink
$[0.0, 0.1)$	$[47] + [70]$	47=Funds, trusts, and other financial vehicles 70=State and local general government	47
$[0.0, 0.4)$	$[45] + [44]$	45=Securities, commodity contracts, and investments 44=Federal Reserve banks, credit intermediation, and related activities	45
$[0.0, 0.07)$	$[24] + [3]$	3=Oil and gas extraction 24=Petroleum and coal products	24

Table 4.1: The first two columns contain sample 0-dimensional persistence intervals, as produced by Javaplex. We have added the labels in column 3, and the common δ -sinks in column 4.

Patterns of investment

Examining the representative cycles of the persistent 1-dimensional intervals in Figure 4.6 allows us to discover patterns of investment that would not otherwise be apparent from the raw data. Javaplex produces representative nodes for each nontrivial persistence interval, so we were able to directly obtain the industrial sectors involved in each cycle. Note that for a persistence interval $[\delta_0, \delta_1]$, Javaplex produces a representative cycle that emerges at resolution δ_0 . As more 1-simplices enter the Dowker filtration at greater resolutions, the homology equivalence class of this cycle may coincide with that of a shorter cycle, until finally it becomes the trivial class at δ_1 . We have illustrated some of the representative cycles produced by Javaplex in Figures 4.9 and 4.12. To facilitate our analysis, we have also added arrows in the figures according to the following rule: for each representative cycle at resolution δ , there is an arrow $e_i \rightarrow e_j$ if and only if $\omega_E(e_i, e_j) \leq \delta$, i.e. if and only if e_j is a sink for the simplex $[e_i, e_j]$ in $\mathfrak{D}_{\delta, E}^{\text{si}}$.

Consider the 1-dimensional persistence interval $[0.75, 0.95]$, colored in red in Figure 4.6. The industries involved in a representative cycle for this interval at $\delta = 0.75$ are: Wood products (WO), Primary metals (PM), Fabricated metal products (FM), Petroleum and coal products (PC), Chemical products (CH), and Plastics and rubber products (PL). The entire cycle is illustrated in Figure 4.9. Starting at the bottom right, note that PC has an arrow going towards CH, suggesting the dependence of the chemical industry on petroleum and coal products. This makes sense because petroleum and coal products are the major organic components used by chemical plants. Chemical products are a necessary ingredient for the synthesis of plastics, which could explain the arrow (CH→PL). Plastic products are commonly used to produce wood-plastic composites, which are low-cost alternatives to products made entirely out of wood. This can explain the arrow PL→WO. Next consider the arrows FM→WO and FM→PM. As a possible interpretation of these arrows, note that fabricated metal frames and other components are frequently used in wood products, and fabricated metal structures are used in the extraction of primary metals from ores. Also note that the metal extraction industry is one of the largest consumers of energy. Since energy is mostly produced from petroleum and coal products, this is a possible reason for the arrow PC→PM.

We now consider the 1-dimensional persistence interval $[0.81, 1]$ colored in red in Figure 4.6. The sectors involved in a representative cycle for this interval at $\delta = 0.81$ are: Petroleum and coal products (PC), Oil and gas (OG), Waste management (WM), State and local general government (SLGG), Apparel and leather and allied products (AP), Textile mills (TE), Plastics and rubber products (PL), and Chemical products (CH). The pattern of investment in this cycle is illustrated in Figure 4.12, at resolutions $\delta = 0.81$ and $\delta = 0.99$. We have already provided interpretations for the arrows OG→PC→CH→PL above. Consider the arrow PL→TE. This likely reflects the widespread use of polyester and polyester blends in production of fabrics. These fabrics are then cut and sewn to manufacture clothing, hence the arrow TE→AP. Also consider the arrow WM→OG: this suggests the role of waste management services in the oil and gas industry, which makes sense because the

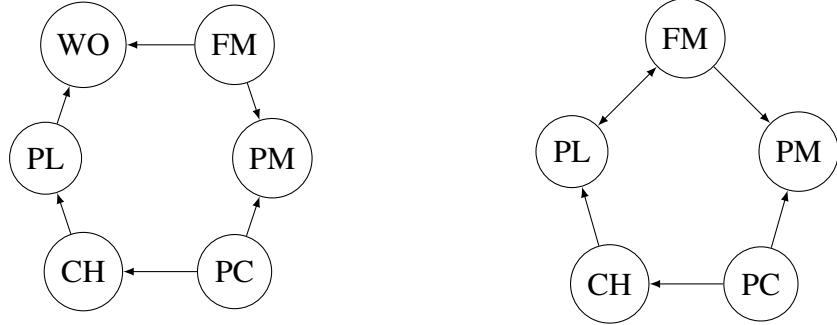


Figure 4.7: Investment patterns at $\delta = 0.75$ Figure 4.8: Investment patterns at $\delta = 0.94$

Figure 4.9: Here we illustrate the representative nodes for one of the 1-dimensional persistence intervals in Figure 4.6. This 1-cycle $[PC,CH] + [CH,PL] + [PL,WO] - [WO,FM] + [FM,PM] - [PM,PC]$ persists on the interval $[0.75, 0.95]$. At $\delta = 0.94$, we observe that this 1-cycle has joined the homology equivalence class of the shorter 1-cycle illustrated on the right. Unidirectional arrows represent an asymmetric flow of investment. A full description of the meaning of each arrow is provided in §4.5.1.

waste management industry has a significant role in the treatment and disposal of hazardous materials produced in the oil and gas industry. Finally, note that the arrows $SLGG \rightarrow WM$ and $SLGG \rightarrow AP$ likely suggest the dependence of the waste management and apparel industries on state and local government support.

We note that there are numerous other 1-dimensional persistence intervals in Figure 4.6 that could be worth exploring, especially for economists who regularly analyze the make tables in input-output accounts and are better prepared to interpret this type of data. The results obtained in our analysis above suggest that viewing these tables as asymmetric networks and then computing their persistence diagrams is a reasonable method for uncovering their hidden attributes.

4.5.2 U.S. migration

The U.S. Census Bureau (www.census.gov) publishes data on the annual migration of individuals between the 50 states of the U.S., the District of Columbia, and Puerto Rico. An individual is said to have migrated from State A to State B in year Y if their residence in year Y is in State B, and in the previous year, their residence was in State A. In this section, we define a network structure that encapsulates migration flows within the U.S., and study this network via its Dowker persistence diagrams. We use migration flow data from 2011

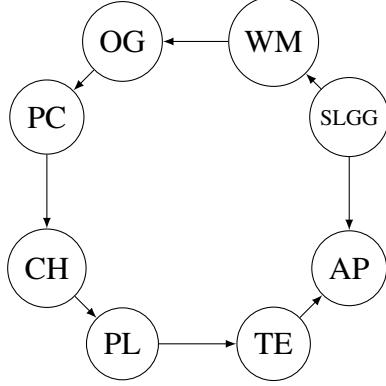


Figure 4.10: Investment patterns at $\delta = 0.81$

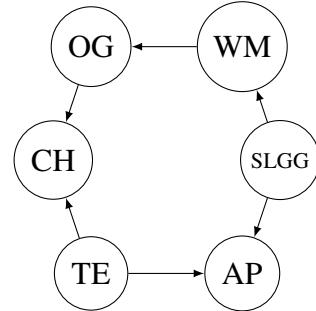


Figure 4.11: Investment patterns at $\delta = 0.99$

Figure 4.12: Representative nodes for another 1-dimensional persistence interval in Figure 4.6. A full description of this cycle is provided in §4.5.1.

in our work. We begin with a set $S = \{s_1, \dots, s_{52}\}$ of these 52 regions and a function $m : S \times S \rightarrow \mathbb{Z}_+$, where $m(s_i, s_j)$ represents the number of migrants moving from s_i to s_j . We define a network (S, ω_S) , with ω_S given by:

$$\omega_S(s_i, s_j) = f \left(\frac{m(s_i, s_j)}{\sum_{s_i \in S, i \neq j} m(s_i, s_j)} \right) \text{ if } s_i \neq s_j, \quad \omega_S(s_i, s_i) = 0, \quad s_i, s_j \in S,$$

where $f(x) = 1 - x$. The purpose of f is to convert similarity data into dissimilarity data. A large value of $\omega_S(s_i, s_j)$ means that few people move from s_i to s_j . The diagonal is removed to ensure that we focus on migration patterns, not on the base population of each state.

The Dowker complex of migration

Because we will eventually compute the Dowker persistence diagram of the network (S, ω_S) , we first suggest interpretations of some aspects of the Dowker complex that we construct as an intermediate step. Let $\delta \geq 0$, and write $\mathfrak{D}_\delta^{\text{si}} := \mathfrak{D}_{\delta, S}^{\text{si}}$, $C_2^\delta := C_2(\mathfrak{D}_\delta^{\text{si}})$. We proceed by defining, for any state $s \in S$, the *migrant influx* of s :

$$\text{influx}(s) := \sum_{s' \in S, s' \neq s} m(s', s).$$

Interpretation of a δ -sink Next we probe the meaning of a δ -sink in the migration context. For simplicity, we first discuss δ -sinks of 1-simplices. Let $[s, s']$ be a 1-simplex for $s, s' \in S$, and let $s'' \in S$ be a δ -sink for $[s, s']$. Then, by unwrapping the definitions above, we see that s'' receives *at least* $(1 - \delta)(\text{influx}(s''))$ migrants from each of s, s' . This suggests the following physical interpretation of the 1-simplex $[s, s']$: in 2010, there were at least $(1 - \delta)(\text{influx}(s''))$ residents in each of s and s' who had a common goal of moving to s'' in 2011. There could be a variety of reasons for interregional migration—people might be moving for employment purposes, for better climate, and so on—but the important point here is that we have a quantitative estimate of residents of s and s' with similar relocation preferences. On the other hand, letting $r, r' \in S$ be states such that $[r, r'] \notin \mathfrak{D}_\delta^{\text{si}}$, the lack of a common δ -sink suggests that residents of r and r' might have significantly different migration preferences. Following this line of thought, we hypothesize the following:

Residents of states that span a 1-simplex in $\mathfrak{D}_\delta^{\text{si}}$ are more similar to each other (in terms of migrational preferences) than residents of states that do not span a 1-simplex.

More generally, when n states form an n -simplex in $\mathfrak{D}_\delta^{\text{si}}$, we say that they exhibit *coherence of preference* at resolution δ . The idea is that the residents of these n states have a mutual preference for a particular attractor state, which acts as a δ -sink. Conversely, a collection of n states that do not form an n -simplex are said to exhibit *incoherence of preference* at resolution δ —residents of these states do not agree strongly enough on a common destination for migration.

Interpretation of a connected component Now we try to understand the physical interpretation of a *connected component* in $\mathfrak{D}_\delta^{\text{si}}$. Recall that two states $s, s' \in S$ belong to a connected component in $\mathfrak{D}_\delta^{\text{si}}$ if and only if there exist $s_1 = s, \dots, s_n = s' \in S$ such that:

$$[s_1, s_2], [s_2, s_3], [s_3, s_4], \dots, [s_{n-1}, s_n] \in \mathfrak{D}_\delta^{\text{si}}. \quad (4.2)$$

Let $s_1, \dots, s_n \in S$ be such that Condition 4.2 above is satisfied. Note that this implies that there exists a δ -sink r_i for each $[s_i, s_{i+1}]$, for $1 \leq i \leq n - 1$. Let r_1, \dots, r_{n-1} be δ -sinks for $[s_1, s_2], \dots, [s_{n-1}, s_n]$. We can further verify, using the fact that ω_S vanishes on the diagonal, that the sinks r_1, \dots, r_{n-1} themselves belong to this connected component:

$$[s_1, r_1], [r_1, s_2], [s_2, r_2], [r_2, s_3], \dots, [r_{n-1}, s_n] \in \mathfrak{D}_\delta^{\text{si}}.$$

The moral of the preceding observations can be summarized as follows:

The vertex set of any connected component of $\mathfrak{D}_\delta^{\text{si}}$ contains a special subset of “attractor” or “sink” states at resolution δ .

So in 2010, for any $i \in \{1, n - 1\}$, there were at least $(1 - \delta)(\text{influx}(r_i))$ people in s_i and in s_{i-1} who had a common goal of moving to r_i in 2011. Moreover, for any $i, j \in \{1, \dots, n\}, i \neq j$, there were at least $\min_{1 \leq i \leq n-1} (1 - \delta)(\text{influx}(r_i))$ people in each

of s_i and s_j in 2010 who migrated elsewhere in 2011. From a different perspective, we are able to distinguish all the states in a connected component that are significantly attractive to migrants (the sinks/receivers), and we have quantitative estimates on the migrant flow within this connected component into its sink/receiver states.

Consider the special case where each state in a connected component of n states, written as s_1, s_2, \dots, s_n , loses $(1 - \delta)(\text{influx}(r))$ residents to a single state $r \in S$. By the preceding observations, r belongs to this connected component, and we can write $r = s_1$ (relabeling as needed). Then we observe that the n states $\{r, s_2, \dots, s_n\}$ form an n -simplex, with r as a common sink. In this case, we have $\omega_S(s_i, r) \leq \delta$ for each $2 \leq i \leq n$. Also note that if we write

$$\begin{aligned} v_n^\delta &:= [r, s_2] + [s_2, s_3] + [s_3, s_4] + \dots + [s_{n-1}, s_n] + [s_n, r] \in C_1^\delta \\ \gamma_n^\delta &:= [r, s_2, s_3] + [r, s_3, s_4] + \dots + [r, s_{n-1}, s_n] \in C_2^\delta, \end{aligned}$$

then we can verify that $\partial_1^\delta(v_n) = 0$, and $\partial_2^\delta(\gamma_n) = v_n$. In other words, we obtain a 1-cycle that is automatically the boundary of a 2-chain, i.e. is trivial upon passing to homology.

In general, a connected component in $\mathfrak{D}_\delta^{\text{si}}$ might contain chains of states that form loops, i.e. states s_1, s_2, \dots, s_n such that:

$$[s_1, s_2], [s_2, s_3], [s_3, s_4], \dots, [s_{n-1}, s_n], [s_n, s_1] \in \mathfrak{D}_\delta^{\text{si}}. \quad (4.3)$$

Note that Condition 4.3 is of course more stringent than Condition 4.2. By writing such a loop in the form of v_n^δ above, we can verify that it forms a 1-cycle. Thus a connected component containing a loop will be detected in a 1-dimensional Dowker persistence diagram, unless the resolution at which the 1-cycle appears coincides with that at which it becomes a 1-boundary.

Interpretation of 1-cycles The preceding discussion shows that it is necessary to determine not just 1-cycles, but also the 1-boundaries that they eventually form. Any 1-boundary arises as the image of ∂_2^δ applied to a linear combination of 2-simplices in $\mathfrak{D}_\delta^{\text{si}}$. Note that in this context, each 2-simplex is a triple of states $[s_i, s_j, s_k]$ with a common sink r to which each of s_i, s_j, s_k has lost $(1 - \delta)(\text{influx}(r))$ residents between 2010 and 2011. Alternatively, at least $(1 - \delta)(\text{influx}(r))$ residents from each of s_i, s_j, s_k had a common preference of moving to r between 2010 and 2011. Next let $\{[s_1, s'_1, s''_1], [s_2, s'_2, s''_2], \dots, [s_n, s'_n, s''_n]\}$ be a collection of 2-simplices in $\mathfrak{D}_\delta^{\text{si}}$, with sinks $\{r_1, \dots, r_n\}$. One way to consolidate the information they contain is to simply write them as a sum:

$$\tau_n^\delta := [s_1, s'_1, s''_1] + [s_2, s'_2, s''_2] + \dots + [s_n, s'_n, s''_n] \in C_2^\delta.$$

Notice that applying the boundary map to τ_n yields:

$$z_n^\delta := \partial_2^\delta(\tau_n) = \sum_{i=1}^n ([s'_i, s''_i] - [s_i, s''_i] + [s_i, s'_i]).$$

At this point we have a list of triples of states, and for each triple we have a quantitative estimate on the number of residents who have a preferred state for migration in common. Now we consider the following relaxation of this situation: for a fixed $i \in \{1, \dots, n\}$ and some $\delta_0 < \delta$, it might be the case that r_i is no longer a mutual δ_0 -sink for $[s_i, s'_i, s''_i]$, or even that there is *no* δ_0 -sink for $[s_i, s'_i, s''_i]$. However, there might still be δ_0 -sinks u, u', u'' for $[s_i, s'_i], [s'_i, s''_i], [s''_i, s_i]$, respectively. In such a case, we see that $\tau_n^{\delta_0} \notin C_2^{\delta_0}$, but $z_n^{\delta_0} \in C_1^{\delta_0}$. Thus $0 \neq \langle z_n \rangle_{\delta_0} \in H_1(\mathcal{D}_{\delta_0}^{\text{si}})$. Assuming that $\delta > \delta_0$ is the minimum resolution at which $\langle z_n \rangle_\delta = 0$, we then have a general description of the way in which persistent 1-cycles might arise.

A very special case of the preceding example occurs when we are able to choose a δ -sink r_i for each $[s_i, s'_i, s''_i], i \in \{1, \dots, n\}$, such that $r_1 = r_2 = \dots = r_n$. In this case, we say that z_n^δ becomes a 1-boundary due to a single mutual sink r_1 . This situation is illustrated in Figure 4.13. Also note the interpretation of this special case: assuming that z_n^δ is a 1-boundary, we know that each of the states in the collection $\cup_{i=1}^n \{s_i, s'_i, s''_i\}$ loses $(1 - \delta)(\text{influx}(r_1))$ residents to r_1 between 2010 and 2011. This signals that r_1 is an especially strong attractor state.

We remark that none of the 1-cycles in the U.S. migration data set that we analyzed exhibited the property of becoming a boundary due to a single mutual sink. However, we did find several examples of this special phenomenon in the global migration dataset studied in §4.5.3. One of these special sinks turns out to be Djibouti, which is a gateway from the Horn of Africa to the Middle East, and is both a destination and a port of transit for migrants moving between Asia and Africa.

Interpretation of barcodes in the context of migration data. Having suggested interpretations of simplices, cycles, and boundaries, we now turn to the question of interpreting a persistence barcode in the context of migration. Note that when computing persistence barcodes, Javaplex can return a representative cycle for each bar, with the caveat that we do not have any control over which representative is returned. From the 1-dimensional Dowker persistence barcode of a migration dataset, we can use the right endpoint of a bar to obtain a 1-boundary, i.e. a list of triples of states along with quantitative estimates on how many residents from each triple had a preferred migration destination in common. In the special case where the 1-boundary forms due to a single mutual sink, we will have a further quantitative estimate on how many residents from each state in the 1-boundary migrated to the mutual sink. The left endpoint of a bar in the 1-dimensional Dowker persistence barcode corresponds to a special connected component with the structure of a 1-cycle. Notice that all the connected components are listed in the 0-dimensional Dowker persistence diagram. See §4.5.3 for some additional comments.

Interpretation of error between lower bounds and true migration In each of Tables 4.2 and 4.3 (and Tables 4.4, 4.5 in §4.5.3), we have provided lower bounds on migration flow between certain states, following the discussion above. More precisely, we do the following:

0-cycles Given a persistence interval $[0, \delta), \delta \in \mathbb{R}$, and a representative 0-cycle, we find the 1-simplex that converts the 0-cycle into a 0-boundary at resolution δ . We then find a δ -sink for this 1-simplex, and estimate a lower bound on the migrant flow into this δ -sink.

1-cycles Given a persistence interval $[\delta_0, \delta_1), \delta_0, \delta_1 \in \mathbb{R}$, and a representative 1-cycle, we find a δ_0 -sink for each 1-simplex in the 1-cycle, and estimate a lower bound on the migrant flow into this δ_0 -sink from its associated 1-simplex.

We also provide the true migration flows beside our lower bound estimates. However, in each of our analyses, we incur a certain error between our lower bound and the actual migration value. We now provide some interpretations for this error.

For the case of 0-cycles, note that all the networks we analyze are normalized to have edge weights in the interval $[0, 1]$. For efficiency, in order to produce a Dowker filtration, we compute $\mathfrak{D}_\delta^{\text{si}}$ for δ -values in the set

$$\text{delta} := \{0.01, 0.02, 0.03, \dots, 1.00\}.$$

So whenever we have $\omega_S(s_i, s_j) \notin \text{delta}$ for some states $s_i, s_j \in S$, the 1-simplex $[s_i, s_j]$ is not detected until we compute $\mathfrak{D}_{\delta'}^{\text{si}}$, where δ' is the smallest element in delta greater than $\omega_S(s_i, s_j)$. If s_j is a δ -sink in this case, then our predicted lower bound on the migration flow $s_i \rightarrow s_j$ will differ by up to $(0.01)(\text{influx}(s_j))$ from the true value. The situation described here best explains the error values in Table 4.4.

For the case of 1-cycles, we will study a simple motivating example. Suppose we have the following 1-simplices:

$$[s_1, s_2], [s_2, s_3], \dots, [s_{n-1}, s_n], [s_n, s_1], \text{ for } s_1, \dots, s_n \in S.$$

For each $i \in \{1, \dots, n\}$, let $\delta_i \in \mathbb{R}$ denote the resolution at which simplex $[s_i, s_{i+1} \pmod n]$ emerges. For simplicity, suppose we have $\delta_1 \leq \delta_2 \leq \delta_3 \leq \dots \leq \delta_n$, and also that s_2 is a δ_n -sink for $[s_1, s_2]$. For our lower bound, we estimate that the migrant flow $s_1 \rightarrow s_2$ is at least $(1 - \delta_n)(\text{influx}(s_2))$. A better lower bound would be $(1 - \delta_1)(\text{influx}(s_2))$, but the only δ -value that Javaplex gives us access to is δ_n . Because δ_1 could be much smaller than δ_n , it might be the case that our lower bound is much smaller than the true migration.

The preceding discussion suggests the following inference: if a 1-simplex $[s_i, s_{i+1}]$ exhibits a large error between the true migration into a δ_n -sink and the predicted lower bound, then $[s_i, s_{i+1}]$ likely emerged at a resolution proportionately smaller than δ_n . Thus we can interpret the states s_i, s_{i+1} as exhibiting relatively strong coherence of preference. Conversely, 1-simplices that exhibit a smaller error likely emerged at a resolution closer to δ_n —the states forming such 1-simplices exhibited incoherence of preference for a greater range of resolutions. Note that even though we made some simplifying assumptions in our choice of a 1-cycle, a similar analysis can be done for any 1-cycle.

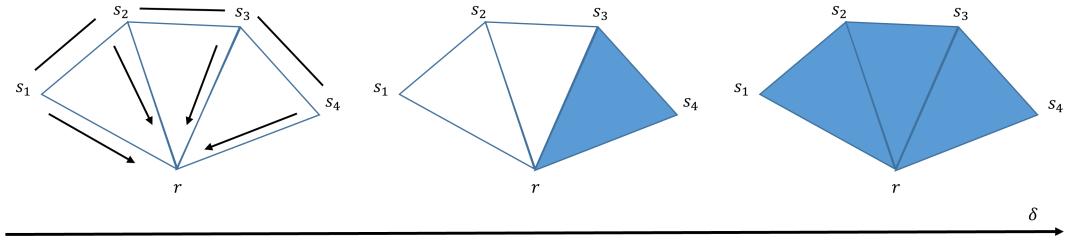


Figure 4.13: An example of a 1-cycle becoming a 1-boundary due to a single mutual sink r , as described in the interpretation of 1-cycles in §4.5.2. The figure on the left shows a connected component of $\mathfrak{D}_{\delta, S}^{\text{si}}$, consisting of $[s_1, s_2]$, $[s_2, s_3]$, $[s_3, s_4]$. The arrows are meant to suggest that r will eventually become a δ -sink for each of these 1-simplices, for some large enough δ . The progression of these simplices for increasing values of δ are shown from left to right. In the leftmost figure, r is not a δ -sink for any of the three 1-simplices. Note that r has become a δ -sink for $[s_3, s_4]$ in the middle figure. Finally, in the rightmost figure, r has become a δ -sink for each of the three 1-simplices.

Analysis of U.S. migration

We computed the 0 and 1-dimensional Dowker persistence barcodes of the network (S, ω_S) obtained in §4.5.2. The result is shown in Figure 4.14. As particular cases, we study the two 1-dimensional bars that are highlighted in red in Figure 4.14. We obtain representative cycles for each of these bars, and superimpose them on a map of the U.S. in Figure 4.15. In the remainder of this section, we will discuss the cycles presented in Figure 4.15.

The (OH-KY-GA-FL) cycle As a representative 1-cycle for the 1-dimensional interval $[0.90, 0.94]$, Javaplex returns the 1-cycle $([\text{FL}, \text{OH}] + [\text{FL}, \text{GA}] + [\text{KY}, \text{OH}] + [\text{GA}, \text{FL}])$. This Ohio-Kentucky-Georgia-Florida cycle first emerges in $C_1^{0.90}$, and becomes a 1-boundary in $C_1^{0.94}$. In Table 4.2, we provide the 0.90-sinks for each of the 1-simplices, the 0.94-sinks for some 2-simplices that convert this cycle into a boundary, our estimates on the number of migrants to each sink, and also the true numbers of migrants.

The (WA-OR-CA-AZ-UT-ID) cycle As a representative 1-cycle for the 1-dimensional interval $[0.87, 0.92]$, Javaplex returns the 1-cycle $([\text{CA}, \text{OR}] + [\text{OR}, \text{WA}] + [\text{AZ}, \text{UT}] + [\text{ID}, \text{UT}] + [\text{ID}, \text{WA}] + [\text{AZ}, \text{CA}])$. This Washington-Oregon-California-Arizona-Utah-Idaho cycle first emerges in $C_1^{0.87}$, and becomes a 1-boundary in $C_1^{0.92}$. In Table 4.3, we provide the 0.87-sinks for each of the 1-simplices, the 0.92-sinks that convert this 1-cycle into a boundary, our estimates on the number of migrants to each sink, and the true migration numbers.

Analysis of OH-KY-GA-FL cycle			
1-simplex	0.90-sinks	Estimated lower bound on migration	True migration
[FL,OH]	WV	$(1 - 0.90)(\text{influx(WV)}) = 4597$	$m(\text{FL, WV}) = 4964$ $m(\text{OH, WV}) = 7548$
[FL,GA]	AL	$(1 - 0.90)(\text{influx(AL)}) = 10684$	$m(\text{FL, AL}) = 12635$ $m(\text{GA, AL}) = 18799$
	GA	$(1 - 0.90)(\text{influx(GA)}) = 24913$	$m(\text{FL, GA}) = 38658$
[KY,OH]	KY	$(1 - 0.90)(\text{influx(KY)}) = 9925$	$m(\text{OH, KY}) = 12744$
[GA,KY]	TN	$(1 - 0.90)(\text{influx(TN)}) = 15446$	$m(\text{GA, TN}) = 16898$ $m(\text{KY, TN}) = 16852$
2-simplex	0.94-sinks	Estimated lower bound on migration	True migration
[FL,OH,KY]	IN	$(1 - 0.94)(\text{influx(IN)}) = 8640$	$m(\text{FL, IN}) = 11472$ $m(\text{OH, IN}) = 11588$ $m(\text{KY, IN}) = 11071$
	OH	$(1 - 0.94)(\text{influx(OH)}) = 12363$	$m(\text{FL, OH}) = 18191$ $m(\text{KY, OH}) = 19617$
[FL,GA,KY]	TN	$(1 - 0.94)(\text{influx(TN)}) = 9268$	$m(\text{FL, TN}) = 10451$ $m(\text{GA, TN}) = 16898$ $m(\text{KY, TN}) = 16852$

Table 4.2: Quantitative estimates on migrant flow, following the interpretation presented in §4.5.2. In each row, we list a simplex of the form $[s_i, s_j]$ (resp. $[s_i, s_j, s_l]$ for 2-simplices) and any possible δ -sinks s_k . We hypothesize that s_k receives at least $(1 - \delta)(\text{influx}(s_k))$ migrants from each of s_i, s_j (resp. s_i, s_j, s_l)—these lower bounds are presented in the third column. The fourth column contains the true migration numbers. Notice that the [FL,GA] simplex appears to show the greatest error between the lower bound and the true migration. Following the interpretation suggested earlier in §4.5.2, this indicates that Florida and Georgia appear to have strong coherence of preference, relative to the other pairs of states spanning 1-simplices in this table.

Analysis of WA-OR-CA-AZ-UT-ID cycle			
1-simplex	0.87-sinks	Estimated lower bound on migration	True migration
[CA,OR]	OR	$(1 - 0.87)(\text{influx(OR)}) = 14273$	$m(\text{CA, OR}) = 18165$
[OR,WA]	OR	$(1 - 0.87)(\text{influx(OR)}) = 14273$	$m(\text{WA, OR}) = 29168$
[AZ,UT]	UT	$(1 - 0.87)(\text{influx(UT)}) = 9517$	$m(\text{AZ, UT}) = 10577$
[ID,UT]	ID	$(1 - 0.87)(\text{influx(ID)}) = 7519$	$m(\text{UT, ID}) = 7538$
[ID,WA]	ID	$(1 - 0.87)(\text{influx(ID)}) = 7519$	$m(\text{WA, ID}) = 10895$
[AZ,CA]	AZ	$(1 - 0.87)(\text{influx(AZ)}) = 27566$	$m(\text{CA, AZ}) = 35650$
2-simplex	0.92-sinks	Estimated lower bound on migration	True migration
[OR,WA,UT]	ID	$(1 - 0.92)(\text{influx(ID)}) = 4627$	$m(\text{OR, ID}) = 6236$ $m(\text{WA, ID}) = 10895$ $m(\text{UT, ID}) = 7538$
[AZ,CA,ID]	UT	$(1 - 0.92)(\text{influx(UT)}) = 5856$	$m(\text{AZ, UT}) = 10577$ $m(\text{CA, UT}) = 8944$ $m(\text{ID, UT}) = 6059$

Table 4.3: Quantitative estimates on migrant flow, following the interpretation presented in §4.5.2. The entries in this table follow the same rules as those of Table 4.2. Notice that the [OR,WA] and [AZ,CA] simplices show the greatest error between the lower bound and the true migration. Following the interpretation in §4.5.2, this suggests that these two pairs of states exhibit stronger coherence of preference than the other pairs of states forming 1-simplices in this table.

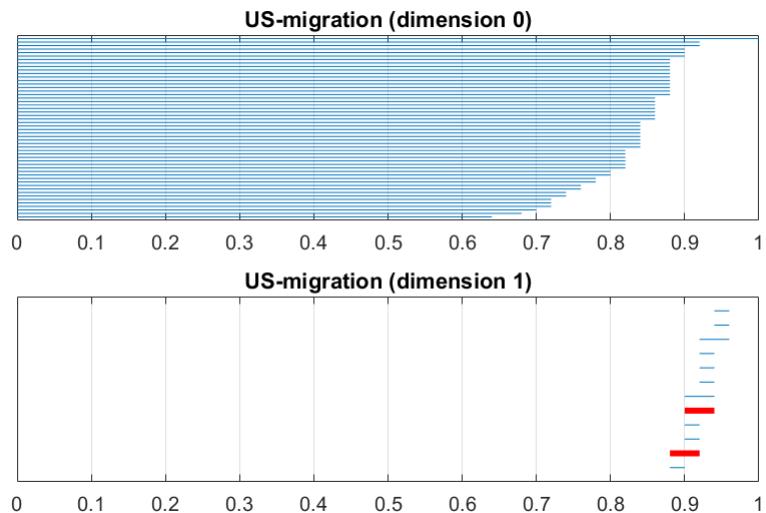


Figure 4.14: 0 and 1-dimensional Dowker persistence barcodes for U.S. migration data



Figure 4.15: U.S. map with representative cycles of the persistence intervals that were highlighted in Figure 4.14. The cycle on the left appears at $\delta_1 = 0.87$, and the cycle on the right appears at $\delta_2 = 0.90$. The red lines indicate the 1-simplices that participate in each cycle. Each red line is decorated with an arrowhead $s_i \rightarrow s_j$ if and only if s_j is a sink for the simplex $[s_i, s_j]$. The blue arrows point towards all possible alternative δ -sinks, and are interpreted as follows: Tennessee is a 0.90-sink for the Kentucky-Georgia simplex, West Virginia is a 0.90-sink for the Ohio-Florida simplex, and Alabama is a 0.90-sink for the Georgia-Florida simplex.

To probe the sociological aspects of a 1-cycle, we recall our hypothesis that residents of states that are not connected by a 1-simplex are less similar to each other than residents of states that are connected as such. The West Coast cycle given above seems to follow this hypothesis: It seems reasonable to think that residents of California would be quite different from residents of Idaho or Utah, and possibly quite similar to those of Oregon. Similarly, one would expect a large group of people from Ohio and Kentucky to be quite similar, especially with Cincinnati being adjacent to the state border with Kentucky. The Ohio-Florida simplex might be harder to justify, but given the very small population of their mutual sink West Virginia, it might be the case that the similarity between Ohio and Florida is being overrepresented.

Our analysis shows that by using Dowker persistence diagrams for exploratory analysis of migration data, we can obtain meaningful lower bounds on the number of residents from different states who share a common migration destination. An interesting extension of this experiment would be to study the persistence barcodes of migration networks over a longer range of years than the 2010-2011 range that we have used here: ideally, we would be able to detect changing trends in migration from changes in the lower bounds that we obtain.

4.5.3 Global migration

For our next example, we study data from the World Bank Global Bilateral Migration Database [94] on global bilateral migration between 1990 and 2000. This dataset is available on <http://databank.worldbank.org/>. As in the case of the U.S. migration database, we begin with a set $C = \{c_1, c_2, \dots, c_{231}\}$ of 231 global regions and a function $m : C \times C \rightarrow \mathbb{Z}_+$, where $m(c_i, c_j)$ represents the number of migrants moving from region c_i to region c_j . We define a network (C, ω_C) , with ω_C given by:

$$\omega_C(c_i, c_j) = f \left(\frac{m(c_i, c_j)}{\sum_{c_i \in C, i \neq j} m(c_i, c_j)} \right) \text{ if } c_i \neq c_j, \quad \omega_C(c_i, c_i) = 0, \quad c_i, c_j \in C,$$

where $f(x) = 1 - x$. The 0 and 1-dimensional Dowker persistence barcodes that we obtain from this network are provided in Figure 4.16. Some of the 0 and 1-dimensional persistence intervals are tabulated in Tables 4.4 and 4.5.

We interpret connected components, simplices, cycles and boundaries for the Dowker sink complexes constructed from the global migration data just as we did for the U.S. migration data in §4.5.2.

We draw the reader's attention to two interesting features in the persistence barcodes of the global migration dataset. First, the 0-dimensional barcode contains many short bars (e.g. bars of length less than 0.2). In contrast, the shortest bars in the 0-dimensional barcode for the U.S. migration data had length greater than 0.65. In our interpretation, which we explain more carefully below, this observation suggests that migration patterns in the U.S. are relatively uniform, whereas global migration patterns can be more skewed. Second,

because there are many more 1-dimensional persistence intervals, it is easier to find a 1-cycle that becomes a boundary due to a single mutual sink, i.e. due to an especially strong “attractor” region.

For the first observation, consider a 0-dimensional persistence interval $[0, \delta]$, where δ is assumed to be small. Formally, this interval represents the persistence of a 0-cycle that emerges at resolution 0, and becomes a 0-boundary at resolution δ . One can further verify the following: this interval represents the resolutions for which a 0-simplex $[c_i], c_i \in C$ remains disconnected from other 0-simplices, and δ is a resolution at which c_i forms a 1-simplex with some $c_j \in C$. Recall from §4.5.2 that this means the following: either there exists a region $c_k \notin \{c_i, c_j\}$ which receives at least $(1 - \delta)(\text{influx}(c_k))$ migrants from each of c_i and c_j , or $c_k = c_i$ (or c_j) and c_k receives over $(1 - \delta)(\text{influx}(c_k))$ migrants from c_j (resp. c_i). The first case cannot happen when $\delta < 0.5$, because this would mean that c_k receives strictly over 50% of its migrant influx from each of c_i and c_j . Thus when $\delta < 0.5$, we know that $c_k = c_i$ (or $c_k = c_j$), and c_k receives over 50% of its migrant influx from c_j (resp. c_i). For very small δ , we then know that *most* of the migrants into c_k arrive from c_j (resp. c_i).

For convenience, let us assume that $\delta < 0.2$, that $c_k = c_i$, and that c_k receives over 80% of its migrant influx from c_j . This might occur for a variety of reasons, some of which are: (1) there might be war or political strife in c_j and c_k might be letting in refugees, (2) c_k might have achieved independence or administrative autonomy and some residents from c_j might be flocking to c_k because they perceive it to be their homeland, and (3) c_j might be overwhelmingly populous in comparison to other neighboring regions of c_k , so that the contribution of c_j to the migrant influx of c_k dominates that of other regions.

Notice that neither of the first two reasons listed above are valid in the case of U.S. migration. The third reason is valid in the case of a few states, but nevertheless, the shortest 0-dimensional persistence interval in the U.S. migration dataset has length greater than 0.65. In other words, the minimal resolution at which a 1-simplex forms in the U.S. migration data is 0.65. This in turn means that there is no state in the U.S. which receives over 35% of its migrant influx from any single other state. Based on this reasoning, we interpret the migration pattern of the U.S. as “diffuse” or “uniform”, and that of the world as a whole as “skewed” or “biased”. This makes intuitive sense, because despite the heterogeneity of the U.S. and differences in state laws and demographics, any resident can easily migrate to any other state of their choice while maintaining similar legal rights, salary, and living standards.

In Table 4.4, we list some short 0-dimensional persistence intervals for the global migration dataset. For each interval $[0, \delta]$, we also include the 1-simplex that emerges at δ , the δ -sink associated to this 1-simplex, and our lower bound on the migrant influx into this sink. Note that the error between the true migration numbers and our predicted lower bounds is explained in §4.5.2. Also notice that many of the migration patterns provided in Table 4.4 seem to fit with the suggestions we made earlier: (1) political turmoil in the West Bank and Gaza (especially following the Gulf War) prompted many Palestinians to

enter Syria, (2) Greenland and Macao are both autonomous regions of Denmark and China, respectively, and (3) India's population far outstrips that of its neighbors, and its migrant flow plays a dominating role in the migrant influx of its neighboring states.

For the second observation, recall from our discussion in §4.5.2 that whenever we have a 1-cycle involving regions c_1, \dots, c_n that becomes a 1-boundary at resolution $\delta \geq 0$ due to a single mutual sink c_{n+1} , we know that c_{n+1} receives at least $(1 - \delta)(\text{influx}(c_{n+1}))$ migrants from each of c_1, \dots, c_n . As such, c_{n+1} can be perceived to be an especially strong attractor region. In Table 4.5 we list some 1-cycles persisting on an interval $[\delta_0, \delta_1]$, their mutual δ_1 -sinks, our lower bound on migration flow, and the true migration numbers. The reader is again encouraged to check that the true migration agrees with the lower bounds that we predicted. We remark that the first row of this table contains a notable example of a strong attractor region: Djibouti. Djibouti is geographically located at a crossroads of Asia and Africa, and is a major commercial hub due to its access to the Red Sea and the Indian Ocean. As such, one would expect it to be a destination for many migrants in the Horn of Africa, as well as a transit point for migrants moving between Africa and the Middle East.

The Oceania cycle listed in the fourth row of Table 4.5 can likely be discarded; the very small migrant influx of Samoa indicates that its attractiveness as a sink state is being overrepresented. The second row lists China as a strong attractor, which is reasonable given its economic growth between 1990 and 2000, and as a consequence, its attractiveness to foreign workers from neighboring countries. The third row lists Vietnam as a strong sink, and one reason could be that in the 1990s, many refugees who had been displaced due to the Vietnam War were returning to their homeland.

We also illustrate the emergence at δ_0 for some of these cycles in Figure 4.17.

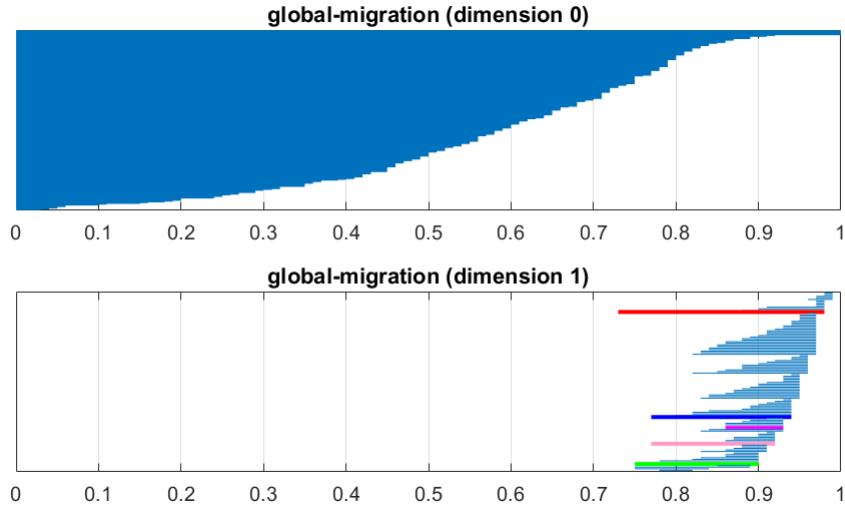


Figure 4.16: Dowker persistence barcodes for global migration dataset.

Short bars in 0-dimensional global migration barcode				
Interval $[0, \delta)$	1-simplex	δ -sink(s)	Lower bound on migration	True migration
[0.0, 0.03)	[India, Sri Lanka]	Sri Lanka	(0.97)(influx(LKA)) = 383034	$m(\text{IND}, \text{LKA}) = 384789$
[0.0, 0.04)	[India, Bangladesh]	Bangladesh	(0.96)(influx(BGD)) = 927270	$m(\text{IND}, \text{BGD}) = 936151$
[0.0, 0.05)	[India, Nepal]	Nepal	(0.95)(influx(NPL)) = 927270	$m(\text{IND}, \text{BGD}) = 936151$
[0.0, 0.05)	[India, Pakistan]	Pakistan	(0.95)(influx(PAK)) = 2508882	$m(\text{IND}, \text{PAK}) = 2512906$
[0.0, 0.06)	[India, Bhutan]	Bhutan	(0.94)(influx(BTN)) = 30206	$m(\text{IND}, \text{BTN}) = 30431$
[0.0, 0.06)	[Denmark, Greenland]	Greenland	(0.94)(influx(GRL)) = 6792	$m(\text{DNK}, \text{GRL}) = 6808$
[0.0, 0.10)	[Greece, Albania]	Albania	(0.90)(influx(ALB)) = 67008	$m(\text{GRC}, \text{ALB}) = 67508$
[0.0, 0.11)	[Timor-Leste, Indonesia]	Timor-Leste	(0.89)(influx(TLS)) = 8246	$m(\text{IDN}, \text{TLS}) = 8334$
[0.0, 0.15)	[West Bank and Gaza, Syria]	Syria	(0.85)(influx(SYR)) = 455515	$m(\text{PSE}, \text{SYR}) = 458611$
[0.0, 0.16)	[Macao, China]	Macao	(0.84)(influx(MAC)) = 201840	$m(\text{CHN}, \text{MAC}) = 203877$

Table 4.4: Short 0-dimensional Dowker persistence intervals capture regions which receive most of their incoming migrants from a single source. Each interval $[0, \delta)$ corresponds to a 0-simplex which becomes subsumed into a 1-simplex at resolution δ . We list these 1-simplices in the second column, and their δ sinks in the third column. The definition of a δ -sink enables us to produce a lower bound on the migration into each sink, which we provide in the fourth column. We also list the true migration numbers in the fifth column, and the reader can consult §4.5.2 for our explanation of the error between the true migration and the lower bounds on migration.

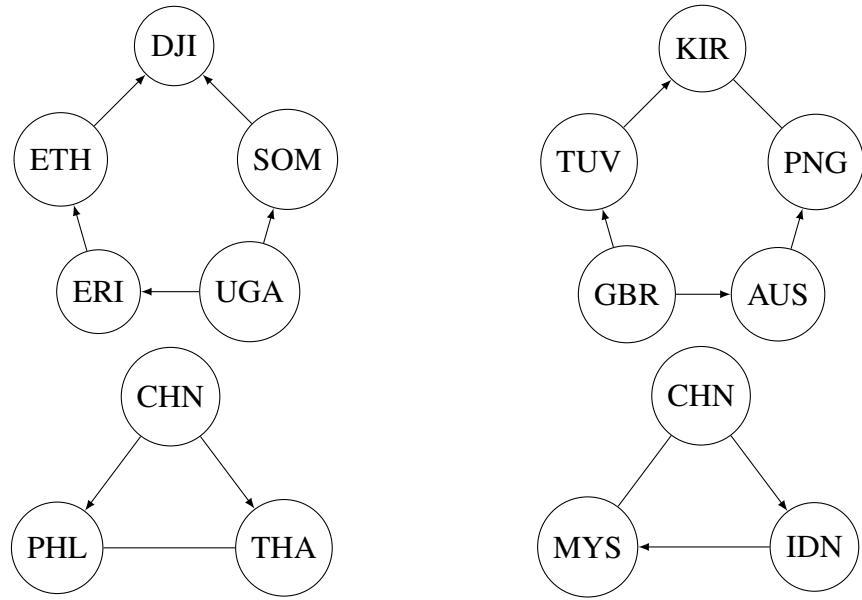


Figure 4.17: **Top:** Two cycles corresponding to the left endpoints of the (Djibouti-Somalia-Uganda-Eritrea-Ethiopia) and (Kiribati-Papua New Guinea-Australia-United Kingdom-Tuvalu) persistence intervals listed in Table 4.5. The δ values are 0.73, 0.77, respectively. **Bottom:** Two cycles corresponding to the left endpoints of the (China-Thailand-Philippines) and (China-Indonesia-Malaysia) persistence intervals listed in Table 4.5. The δ values are 0.77, 0.75, respectively. **Meaning of arrows:** In each cycle, an arrow $s_i \rightarrow s_j$ means that $\omega_S(s_i, s_j) \leq \delta$, i.e. that s_j is a sink for the simplex $[s_i, s_j]$. We can verify separately that for $\delta = 0.77$, the Kiribati-Papua New Guinea simplex has the Solomon Islands as a δ -sink, and that the Philippines-Thailand simplex has Taiwan as a δ -sink. Similarly, the China-Malaysia simplex has Singapore as a δ -sink, for $\delta = 0.75$.

1-cycles with single mutual sink in global migration data				
Interval $[\delta_0, \delta_1)$	Regions involved	Mutual δ_1 -sink(s)	Lower bound on migration	True migration
[0.73,0.98]	Djibouti—Ethiopia—Eritrea—Uganda—Somalia	Djibouti	(0.02)(influx(DJI)) = 1738	$m(\text{ERI}, \text{DJI}) = 3259$ $m(\text{ETH}, \text{DJI}) = 25437$ $m(\text{SOM}, \text{DJI}) = 41968$ $m(\text{UGA}, \text{DJI}) = 1811$
[0.77,0.94]	China—Thailand—Philippines	China	(0.06)(influx(CHN)) = 12858	$m(\text{THA}, \text{CHN}) = 14829$ $m(\text{PHL}, \text{CHN}) = 17828$
[0.75,0.89)	China—Indonesia—Malaysia	Vietnam	(0.11)(influx(VNM)) = 4465	$m(\text{CHN}, \text{VNM}) = 8940$ $m(\text{IDN}, \text{VNM}) = 10529$ $m(\text{MYS}, \text{VNM}) = 4813$
[0.86,0.93)	American Samoa—New Zealand—Samoa—Australia	Samoa	(0.07)(influx(WSM)) = 397	$m(\text{ASM}, \text{WSM}) = 1920$ $m(\text{NZL}, \text{WSM}) = 1803$ $m(\text{AUS}, \text{WSM}) = 404$
[0.77,0.92)	Kiribati—Papua New Guinea—Australia—United Kingdom—Tuvalu	Not applicable	Not Applicable	Not Applicable

Table 4.5: Representative 1-cycles for several intervals in the 1-dimensional persistence barcode for the global migration dataset. Each of the first four cycles has the special property that it becomes a boundary due to a single sink at the right endpoint of its associated persistence interval. This permits us to obtain a lower bound on the migration into this sink from each of the regions in the cycle. The last row contains a cycle without this special property. The font colors of the persistence intervals correspond to the colors of the highlighted 1-dimensional bars in Figure 4.16.

Bibliography

- [1] Dowker’s theorem. <https://ncatlab.org/nlab/show/Dowker%27s+theorem>. Accessed: 2017-04-24.
- [2] Emmanuel Abbe. Community detection and stochastic block models: recent developments. *arXiv preprint arXiv:1703.10146*, 2017.
- [3] Michał Adamaszek and Henry Adams. The Vietoris–Rips complexes of a circle. *Pacific Journal of Mathematics*, 290(1):1–40, 2017.
- [4] Michal Adamaszek, Henry Adams, Florian Frick, Chris Peterson, and Corrine Previte-Johnson. Nerve complexes of circular arcs. *Discrete & Computational Geometry*, 56(2):251–273, 2016.
- [5] Luigi Ambrosio, Nicola Gigli, and Giuseppe Savaré. *Gradient flows: in metric spaces and in the space of probability measures*. Springer Science & Business Media, 2008.
- [6] David Bao, S-S Chern, and Zhongmin Shen. *An introduction to Riemann-Finsler geometry*, volume 200. Springer Science & Business Media, 2012.
- [7] Jonathan A Barmak. *Algebraic topology of finite topological spaces and applications*, volume 2032. Springer, 2011.
- [8] Ulrich Bauer, Michael Kerber, and Jan Reininghaus. Clear and compress: Computing persistent homology in chunks. In *Topological Methods in Data Analysis and Visualization III*, pages 103–117. Springer, 2014.
- [9] Ulrich Bauer and Michael Lesnick. Induced matchings of barcodes and the algebraic stability of persistence. In *Proceedings of the thirtieth annual symposium on Computational geometry*, page 355. ACM, 2014.
- [10] Jean-David Benamou, Guillaume Carlier, Marco Cuturi, Luca Nenna, and Gabriel Peyré. Iterative Bregman projections for regularized transportation problems. *SIAM Journal on Scientific Computing*, 37(2):A1111–A1138, 2015.

- [11] Anders Björner. Topological methods. *Handbook of combinatorics*, 2:1819–1872, 1995.
- [12] Anders Björner, Bernhard Korte, and László Lovász. Homotopy properties of greedoids. *Advances in Applied Mathematics*, 6(4):447–494, 1985.
- [13] Vladimir I Bogachev. *Measure theory*, volume 2. Springer Science & Business Media, 2007.
- [14] Vladimir I Bogachev. *Measure theory*, volume 1. Springer Science & Business Media, 2007.
- [15] Mireille Boutin and Gregor Kemper. Lossless representation of graphs using distributions. *arXiv preprint arXiv:0710.1870*, 2007.
- [16] Martin R Bridson and André Haefliger. *Metric spaces of non-positive curvature*, volume 319. Springer Science & Business Media, 2011.
- [17] Dmitri Burago, Yuri Burago, and Sergei Ivanov. *A Course in Metric Geometry*, volume 33 of *AMS Graduate Studies in Math*. American Mathematical Society, 2001.
- [18] Rainer E Burkard, Mauro Dell’Amico, and Silvano Martello. *Assignment Problems*. SIAM, 2009.
- [19] Gunnar Carlsson and Vin De Silva. Zigzag persistence. *Foundations of computational mathematics*, 10(4):367–405, 2010.
- [20] Gunnar Carlsson, Facundo Mémoli, Alejandro Ribeiro, and Santiago Segarra. Axiomatic construction of hierarchical clustering in asymmetric networks. In *Acoustics, Speech and Signal Processing (ICASSP), 2013 IEEE International Conference on*, pages 5219–5223. IEEE, 2013.
- [21] Gunnar Carlsson, Afra Zomorodian, Anne Collins, and Leonidas J Guibas. Persistence barcodes for shapes. *International Journal of Shape Modeling*, 11(02):149–187, 2005.
- [22] Gunnar E. Carlsson, Facundo Mémoli, Alejandro Ribeiro, and Santiago Segarra. Hierarchical quasi-clustering methods for asymmetric networks. In *Proceedings of the 31th International Conference on Machine Learning, ICML 2014*, 2014.
- [23] CJ Carstens and KJ Horadam. Persistent homology of collaboration networks. *Mathematical Problems in Engineering*, 2013, 2013.

- [24] Frédéric Chazal, David Cohen-Steiner, Marc Glisse, Leonidas J Guibas, and Steve Y Oudot. Proximity of persistence modules and their diagrams. In *Proceedings of the twenty-fifth annual symposium on Computational geometry*, pages 237–246. ACM, 2009.
- [25] Frédéric Chazal, David Cohen-Steiner, Leonidas J Guibas, Facundo Mémoli, and Steve Y Oudot. Gromov-hausdorff stable signatures for shapes using persistence. In *Computer Graphics Forum*, volume 28, pages 1393–1403. Wiley Online Library, 2009.
- [26] Frédéric Chazal, Vin De Silva, Marc Glisse, and Steve Oudot. *The structure and stability of persistence modules*. Springer International Publishing, 2016.
- [27] Frédéric Chazal, Vin De Silva, and Steve Oudot. Persistence stability for geometric complexes. *Geometriae Dedicata*, 173(1):193–214, 2014.
- [28] Frédéric Chazal and Steve Y Oudot. Towards persistence-based reconstruction in Euclidean spaces. In *Proceedings of the twenty-fourth annual Symposium on Computational Geometry*, pages 232–241. ACM, 2008.
- [29] Chao Chen and Michael Kerber. Persistent homology computation with a twist. In *Proceedings 27th European Workshop on Computational Geometry*, volume 11, 2011.
- [30] Lénaïc Chizat. *Transport optimal de mesures positives: modèles, méthodes numériques, applications*. PhD thesis, Université Paris-Dauphine, 2017.
- [31] Lénaïc Chizat, Gabriel Peyré, Bernhard Schmitzer, and François-Xavier Vialard. Scaling algorithms for unbalanced optimal transport problems. *Math. Comp.*, 87(314):2563–2609, 2018.
- [32] Samir Chowdhury, Bowen Dai, and Facundo Mémoli. Topology of stimulus space via directed network persistent homology. *Cosyne Abstracts 2017*.
- [33] Samir Chowdhury, Bowen Dai, and Facundo Mémoli. The importance of forgetting: Limiting memory improves recovery of topological characteristics from neural data. *PloS one*, 13(9):e0202561, 2018.
- [34] Samir Chowdhury and Facundo Mémoli. Convergence of hierarchical clustering and persistent homology methods on directed networks. *arXiv preprint arXiv:1711.04211*, 2017.
- [35] Samir Chowdhury and Facundo Mémoli. Distances and isomorphism between networks and the stability of network invariants. *arXiv preprint arXiv:1708.04727*, 2017.

- [36] Samir Chowdhury and Facundo Mémoli. Explicit geodesics in Gromov-Hausdorff space. *Electronic Research Announcements in Mathematical Sciences*, 2018.
- [37] Samir Chowdhury and Facundo Mémoli. A functorial Dowker theorem and persistent homology of asymmetric networks. *Journal of Applied and Computational Topology*, 2(1-2):115–175, 2018.
- [38] Samir Chowdhury and Facundo Mémoli. The Gromov-Wasserstein distance between networks and stable network invariants. *arXiv preprint arXiv:1808.04337*, 2018.
- [39] Samir Chowdhury and Facundo Mémoli. The metric space of networks. *arXiv preprint arXiv:1804.02820*, 2018.
- [40] Samir Chowdhury and Facundo Mémoli. Persistent path homology of directed networks. In *Proceedings of the Twenty-Ninth Annual ACM-SIAM Symposium on Discrete Algorithms*, pages 1152–1169. SIAM, 2018.
- [41] Carina Curto and Vladimir Itskov. Cell groups reveal structure of stimulus space. *PLoS Computational Biology*, 4(10), 2008.
- [42] Marco Cuturi. Sinkhorn distances: Lightspeed computation of optimal transport. In *Advances in neural information processing systems*, pages 2292–2300, 2013.
- [43] Yuri Dabaghian, Facundo Mémoli, L Frank, and Gunnar Carlsson. A topological paradigm for hippocampal spatial map formation using persistent homology. *PLoS Comput Biol*, 8(8), 2012.
- [44] Vin De Silva and Gunnar Carlsson. Topological estimation using witness complexes. *Proc. Sympos. Point-Based Graphics*, pages 157–166, 2004.
- [45] Tamal K Dey, Facundo Mémoli, and Yusu Wang. Multiscale mapper: Topological summarization via codomain covers. In *Proceedings of the twenty-seventh annual ACM-SIAM Symposium on Discrete Algorithms*, pages 997–1013. SIAM, 2016.
- [46] Clifford H Dowker. Homology groups of relations. *Annals of mathematics*, pages 84–95, 1952.
- [47] Herbert Edelsbrunner and John Harer. *Computational topology: an introduction*. American Mathematical Soc., 2010.
- [48] Herbert Edelsbrunner, Grzegorz Jabłoński, and Marian Mrozek. The persistent homology of a self-map. *Foundations of Computational Mathematics*, 15(5):1213–1244, 2015.

- [49] Herbert Edelsbrunner, David Letscher, and Afra Zomorodian. Topological persistence and simplification. *Discrete and Computational Geometry*, 28(4):511–533, 2002.
- [50] Herbert Edelsbrunner and Dmitriy Morozov. Persistent homology: theory and practice. 2014.
- [51] Gerald A Edgar. Classics on fractals. 1993.
- [52] Alon Efrat, Alon Itai, and Matthew J Katz. Geometry helps in bottleneck matching and related problems. *Algorithmica*, 31(1):1–28, 2001.
- [53] Alexander Engström. Complexes of directed trees and independence complexes. *Discrete Mathematics*, 309(10):3299–3309, 2009.
- [54] Santo Fortunato. Benchmark graphs to test community detection algorithms. <https://sites.google.com/site/santofortunato/inthepress2>.
- [55] M Maurice Fréchet. Sur quelques points du calcul fonctionnel. *Rendiconti del Circolo Matematico di Palermo (1884-1940)*, 22(1):1–72, 1906.
- [56] Patrizio Frosini. Measuring shapes by size functions. In *Intelligent Robots and Computer Vision X: Algorithms and Techniques*, pages 122–133. International Society for Optics and Photonics, 1992.
- [57] Fred Galvin and Samuel Shore. Completeness in semimetric spaces. *Pacific Journal of Mathematics*, 113(1):67–75, 1984.
- [58] Fred Galvin and Samuel Shore. Distance functions and topologies. *The American Mathematical Monthly*, 98(7):620–623, 1991.
- [59] Robert Ghrist. *Elementary applied topology*. Createspace, 2014.
- [60] Chad Giusti, Eva Pastalkova, Carina Curto, and Vladimir Itskov. Clique topology reveals intrinsic geometric structure in neural correlations. *Proceedings of the National Academy of Sciences*, 112(44):13455–13460, 2015.
- [61] Thibaut Le Gouic and Jean-Michel Loubes. Existence and consistency of wasserstein barycenters. *arXiv preprint arXiv:1506.04153*, 2015.
- [62] Alexander Grigor’yan, Yong Lin, Yuri Muranov, and Shing-Tung Yau. Homologies of path complexes and digraphs. *arXiv preprint arXiv:1207.2834*, 2012.
- [63] Alexander Grigor’yan, Yong Lin, Yuri Muranov, and Shing-Tung Yau. Homotopy theory for digraphs. *Pure and Applied Mathematics Quarterly*, 10(4), 2014.

- [64] Alexander Grigor'yan, Yuri Muranov, and Shing-Tung Yau. Homologies of digraphs and the Künneth formula. 2015.
- [65] Misha Gromov. *Metric structures for Riemannian and non-Riemannian spaces*, volume 152 of *Progress in Mathematics*. Birkhäuser Boston Inc., Boston, MA, 1999.
- [66] World Bank Group. Global bilateral migration database. <https://datacatalog.worldbank.org/dataset/global-bilateral-migration-database>. Accessed: October 3, 2018.
- [67] Gary Gruenhage. Generalized metric spaces. *Handbook of set-theoretic topology*, pages 423–501, 1984.
- [68] Allen Hatcher. Algebraic topology. 2002. *Cambridge UP, Cambridge*, 606(9), 2002.
- [69] Jean-Claude Hausmann. On the vietoris-rips complexes and a cohomology theory for metric spaces. *Ann. Math. Studies*, 138:175–188, 1995.
- [70] Reigo Hendrikson. Using Gromov-Wasserstein distance to explore sets of networks. Master's thesis, University of Tartu, 2016.
- [71] Danijela Horak, Slobodan Maletić, and Milan Rajković. Persistent homology of complex networks. *Journal of Statistical Mechanics: Theory and Experiment*, 2009(03):P03034, 2009.
- [72] Karen J Horowitz and Mark A Planting. Concepts and methods of the input-output accounts. 2006.
- [73] Alexandr Ivanov, Nadezhda Nikolaeva, and Alexey Tuzhilin. The Gromov-Hausdorff metric on the space of compact metric spaces is strictly intrinsic. *arXiv preprint arXiv:1504.03830*, 2015.
- [74] Roy A Johnson. Atomic and nonatomic measures. *Proceedings of the American Mathematical Society*, 25(3):650–655, 1970.
- [75] Nigel J Kalton and Mikhail I Ostrovskii. Distances between Banach spaces. In *Forum Mathematicum*, volume 11, pages 17–48. Walter de Gruyter, 1999.
- [76] Arshi Khalid, Byung Sun Kim, Moo K Chung, Jong Chul Ye, and Daejong Jeon. Tracing the evolution of multi-scale functional networks in a mouse model of depression using persistent brain network homology. *NeuroImage*, 101:351–363, 2014.
- [77] Jon M Kleinberg. Authoritative sources in a hyperlinked environment. *Journal of the ACM (JACM)*, 46(5):604–632, 1999.

- [78] Dmitry Kozlov. *Combinatorial algebraic topology*, volume 21. Springer Science & Business Media, 2007.
- [79] Janko Latschev. Vietoris-rips complexes of metric spaces near a closed riemannian manifold. *Archiv der Mathematik*, 77(6):522–528, 2001.
- [80] Hyekyoung Lee, Moo K Chung, Hyejin Kang, Boong-Nyun Kim, and Dong Soo Lee. Computing the shape of brain networks using graph filtration and gromov-hausdorff metric. In *Medical Image Computing and Computer-Assisted Intervention–MICCAI 2011*, pages 302–309. Springer, 2011.
- [81] Solomon Lefschetz. *Algebraic topology*, volume 27. American Mathematical Soc., 1942.
- [82] Laurentiu Leustean, Adriana Nicolae, and Alexandru Zaharescu. Barycenters in uniformly convex geodesic spaces. *arXiv preprint arXiv:1609.02589*, 2016.
- [83] Daniel Lütgehetmann. Flagser. Software available at <https://github.com/luetge/flagser/>, 2018.
- [84] Paolo Masulli and Alessandro EP Villa. The topology of the directed clique complex as a network invariant. *SpringerPlus*, 5(1):1–12, 2016.
- [85] Facundo Mémoli. On the use of Gromov-Hausdorff distances for shape comparison. 2007.
- [86] Facundo Mémoli. Gromov-Wasserstein distances and the metric approach to object matching. *Foundations of Computational Mathematics*, pages 1–71, 2011. 10.1007/s10208-011-9093-5.
- [87] Facundo Mémoli. Some properties of Gromov–Hausdorff distances. *Discrete & Computational Geometry*, pages 1–25, 2012. 10.1007/s00454-012-9406-8.
- [88] James R Munkres. *Elements of algebraic topology*, volume 7. Addison-Wesley Reading, 1984.
- [89] James R Munkres. *Topology*. Prentice Hall, 2000.
- [90] Mark Newman. *Networks*. Oxford university press, 2018.
- [91] VW Niemytzki. On the “third axiom of metric space”. *Transactions of the American Mathematical Society*, 29(3):507–513, 1927.
- [92] Shin-ichi Ohta. Barycenters in Alexandrov spaces of curvature bounded below. 2012.

- [93] John O’Keefe and Jonathan Dostrovsky. The hippocampus as a spatial map. preliminary evidence from unit activity in the freely-moving rat. *Brain research*, 34(1):171–175, 1971.
- [94] Caglar Özden, Christopher R Parsons, Maurice Schiff, and Terrie L Walmsley. Where on earth is everybody? The evolution of global bilateral migration 1960–2000. *The World Bank Economic Review*, 25(1):12–56, 2011.
- [95] Panos M Pardalos and Henry Wolkowicz, editors. *Quadratic assignment and related problems*. DIMACS Series in Discrete Mathematics and Theoretical Computer Science, 16. American Mathematical Society, Providence, RI, 1994.
- [96] Xavier Pennec. Intrinsic statistics on Riemannian manifolds: Basic tools for geometric measurements. *Journal of Mathematical Imaging and Vision*, 25(1):127, 2006.
- [97] Vladimir Pestov. *Dynamics of infinite-dimensional groups: the Ramsey-Dvoretzky-Milman phenomenon*, volume 40. American Mathematical Soc., 2006.
- [98] Peter Petersen. *Riemannian geometry*, volume 171. Springer Science & Business Media, 2006.
- [99] Giovanni Petri, Martina Scolamiero, Irene Donato, and Francesco Vaccarino. Topological strata of weighted complex networks. *PloS one*, 8(6):e66506, 2013.
- [100] Gabriel Peyré, Marco Cuturi, and Justin Solomon. Gromov-Wasserstein averaging of kernel and distance matrices. In *International Conference on Machine Learning*, pages 2664–2672, 2016.
- [101] Arthur Dunn Pitcher and Edward Wilson Chittenden. On the foundations of the calcul fonctionnel of Fréchet. *Transactions of the American Mathematical Society*, 19(1):66–78, 1918.
- [102] Michael W Reimann, Max Nolte, Martina Scolamiero, Katharine Turner, Rodrigo Perin, Giuseppe Chindemi, Paweł Dłotko, Ran Levi, Kathryn Hess, and Henry Markram. Cliques of neurons bound into cavities provide a missing link between structure and function. *Frontiers in computational neuroscience*, 11:48, 2017.
- [103] Vanessa Robins. Towards computing homology from finite approximations. In *Topology proceedings*, volume 24, pages 503–532, 1999.
- [104] Sorin V Sabau, Kazuhiro Shibuya, and Hideo Shimada. Metric structures associated to finsler metrics. *arXiv preprint arXiv:1305.5880*, 2013.

- [105] Felix Schmiedl. *Shape matching and mesh segmentation: mathematical analysis, algorithms and an application in automated manufacturing*. PhD thesis, München, Technische Universität München, Diss., 2015, 2015.
- [106] Bernhard Schmitzer. Stabilized sparse scaling algorithms for entropy regularized transport problems. *arXiv preprint arXiv:1610.06519*, 2016.
- [107] Bernhard Schmitzer and Christoph Schnörr. Modelling convex shape priors and matching based on the Gromov-Wasserstein distance. *Journal of mathematical imaging and vision*, 46(1):143–159, 2013.
- [108] Yi-Bing Shen and Wei Zhao. Gromov pre-compactness theorems for nonreversible finsler manifolds. *Differential Geometry and its Applications*, 28(5):565–581, 2010.
- [109] Richard Sinkhorn. A relationship between arbitrary positive matrices and doubly stochastic matrices. *The annals of mathematical statistics*, 35(2):876–879, 1964.
- [110] Richard Sinkhorn. Diagonal equivalence to matrices with prescribed row and column sums. *The American Mathematical Monthly*, 74(4):402–405, 1967.
- [111] Justin Solomon, Gabriel Peyré, Vladimir G Kim, and Suvrit Sra. Entropic metric alignment for correspondence problems. *ACM Transactions on Graphics (TOG)*, 35(4):72, 2016.
- [112] Edwin H Spanier. *Algebraic topology*, volume 55. Springer Science & Business Media, 1994.
- [113] Sashi Mohan Srivastava. *A course on Borel sets*, volume 180. Springer Science & Business Media, 2008.
- [114] Lynn Arthur Steen, J Arthur Seebach, and Lynn A Steen. *Counterexamples in topology*, volume 18. Springer, 1978.
- [115] Aleksandar Stojmirović and Yi-Kuo Yu. Geometric aspects of biological sequence comparison. *Journal of Computational Biology*, 16(4):579–610, 2009.
- [116] Karl-Theodor Sturm. On the geometry of metric measure spaces. *Acta mathematica*, 196(1):65–131, 2006.
- [117] Karl-Theodor Sturm. The space of spaces: curvature bounds and gradient flows on the space of metric measure spaces. *arXiv preprint arXiv:1208.0434*, 2012.
- [118] Katharine Turner. Generalizations of the Rips filtration for quasi-metric spaces with persistent homology stability results. *arXiv preprint arXiv:1608.00365*, 2016.

- [119] Titouan Vayer, Laetitia Chapel, Rémi Flamary, Romain Tavenard, and Nicolas Courty. Optimal transport for structured data. *arXiv preprint arXiv:1805.09114*, 2018.
- [120] Cédric Villani. *Topics in optimal transportation*. Number 58. American Mathematical Soc., 2003.
- [121] Cédric Villani. *Optimal transport: old and new*, volume 338. Springer Science & Business Media, 2008.
- [122] Paweł Waszkiewicz. The local triangle axiom in topology and domain theory. *Applied General Topology*, 4(1):47–70, 2013.
- [123] Afra Zomorodian and Gunnar Carlsson. Computing persistent homology. *Discrete & Computational Geometry*, 33(2):249–274, 2005.