

The Walk-Length Filtration for Persistent Homology on Weighted Directed Graphs

David E. Muñoz ✉🏠

Computational Mathematics, Science and Engineering, Michigan State University, United States

Elizabeth Munch ✉🏠

Computational Mathematics, Science and Engineering, Michigan State University, United States

Brittany Terese Fasy ✉🏠

Montana State University, United States

Firas A. Khasawneh ✉🏠

Mechanical Engineering, Michigan State University, United States

1 Introduction

Topological data analysis, and particularly persistent homology, has been used in many applications where input data arises as graphs [1]. Recently, particularly with the rise in neuroscience data [6], there is a push for more methods for measuring structure that incorporates directionality [2, 3, 4, 5, 7].

We introduce a new method for analyzing directed graphs with persistent homology via a simplicial filtration, called the *walk-length filtration*, that encodes information about when a walk visits all vertices in a particular set to determine when a simplex should be included. Note that if we asked for a path that visits every vertex just once, this would be equivalent to finding Hamiltonian paths, which is an NP-complete problem, hence we focus on walks. While this filtration is not stable under the network distance used in related work, we define a modification that replaces the maximum with a summation for which stability can be obtained. Finally, we show our results of applying walk-length persistence in cycle networks.

2 Definition

A *weighted directed graph* (*digraph*) is a triple $D = (V, E, w)$, where V is a finite set of vertices, $E \subseteq V \times V$ is a set of directed edges and there is a weight on the edges given by $w : E \rightarrow \mathbb{R}_{\geq 0}$. We assume that $w(v, v') = 0$ if and only if $v = v'$. A *walk of length n* in D is any sequence of vertices $\gamma = (v_0, \dots, v_n)$ where $(v_i, v_{i+1}) \in E$ for $0 \leq i \leq n-1$. Here the vertices can be repeated. The *weight of a walk γ* is given by $w(\gamma) := \sum_{i=0}^{n-1} w(v_i, v_{i+1})$. The collection of all complete weighted digraphs $(V, V \times V, w)$ will be denoted by $\mathcal{N}$.

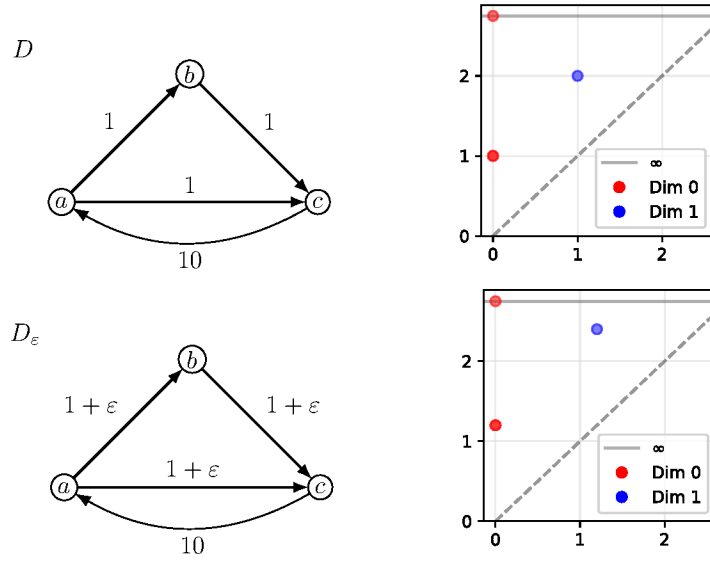
Let $D = (V, E, w)$ be a weighted digraph. For any subset of vertices $\sigma \subseteq V$, define

$$f_D(\sigma) = \inf\{w(\gamma) : \gamma \text{ is a walk in } D \text{ that contains all vertices in } \sigma\}.$$

If there is no such walk, we set $f(\sigma) = \infty$.

For $\delta \in \mathbb{R}$, we can define a simplicial complex $K_\delta = \{\sigma \subseteq V : f(\sigma) \leq \delta\}$. Then, the *walk-length filtration* is the parameterized collection of simplicial complexes $\{K_\delta\}_{\delta \in \mathbb{R}}$ and $\text{Dgm}_k^{\text{WL}}(D)$ is defined as the k -dimensional *walk-length persistence diagram* associated to the corresponding walk-length filtration $\{K_\delta\}_{\delta \in \mathbb{R}}$.

This is an abstract of a presentation given at CG:YRF 2024. It has been made public for the benefit of the community and should be considered a preprint rather than a formally reviewed paper. Thus, this work is expected to appear in a conference with formal proceedings and/or in a journal.



■ **Figure 1** A counterexample to stability under an ℓ_∞ -distance. The one-dimensional persistence points are $(1, 2)$ for D and $(1 + \varepsilon, 2 + 2\varepsilon)$ for D_ε .

Given a strongly connected weighted digraph $D = (V, E, w)$, the *shortest-distance function* $\omega_D : V \times V \rightarrow \mathbb{R}$ is given by $\omega_D(u, v) := \min\{w(\gamma) : \gamma \text{ is a walk in } D \text{ from } u \text{ to } v\}$. The complete weighted digraph (V, ω_D) is called the *shortest-distance digraph associated to* D .

► **Proposition 1.** *Let $D = (V, E, w)$ be a strongly connected weighted digraph, and let $\mathcal{X} = (V, \omega_D)$ be the shortest-distance digraph associated to D . Then, the walk-length filtrations for D and $\mathcal{X}$ are the same.*

3 Stability

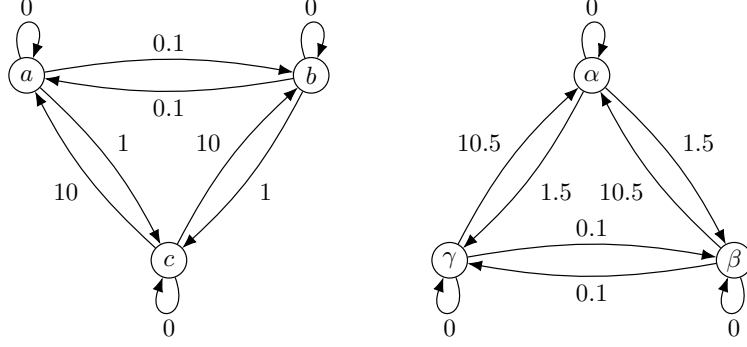
One useful property is that of stability: if the input data is close in some metric, then the resulting representation is closer in its own metric. For this purpose, we work with complete weighted digraphs, where there is an edge for any two distinct vertices; these can also be called *networks*. The standard in the persistence for digraphs literature is to use the network distance d_N [2, 3, 7]. We first point out that our filtration is not stable under this distance. A counterexample is shown in Figure 1. In essence, the issue arises because our filtrating function f is a sum in the walk length but the network distance d_N is a maximum. For this reason, we define a summation version of the network distance.

► **Definition 2.** *Let $\mathcal{X} = (X, \omega_X)$ and $\mathcal{Y} = (Y, \omega_Y)$ be two networks. Define the ℓ_1 -distortion of a correspondence $R \subset X \times Y$ as*

$$\text{dis}^1(R) := \sum_{(x, y), (x', y') \in R} |\omega_X(x, x') - \omega_Y(y, y')|.$$

Then the network ℓ_1 -distance $d_N^1 : \mathcal{N} \times \mathcal{N} \rightarrow \mathbb{R}$ is defined as $d_N^1(\mathcal{X}, \mathcal{Y}) := \frac{1}{2} \min_R \text{dis}^1(R)$.

► **Proposition 3.** *The network ℓ_1 -distance d_N^1 is a pseudometric.*



■ **Figure 2** Networks $\mathcal{X}$ (left) and $\mathcal{Y}$ (right) showing the inequality of Proposition 5 can be strict.

As done in [2], the standard network distance $d_{\mathcal{N}}$ can be reformulated using pairs of maps between networks. Here we define the parallel version in our alternate formulation.

► **Definition 4.** For $\mathcal{X}, \mathcal{Y} \in \mathcal{N}$ and any two maps $\varphi : X \rightarrow Y$ and $\psi : Y \rightarrow X$ on their sets of vertices, the ℓ_1 -distortion and ℓ_1 -codistortion terms are defined respectively as

$$\begin{aligned} \text{dis}^1(\varphi) &:= \sum_{x, x' \in X} |\omega_X(x, x') - \omega_Y(\varphi(x), \varphi(x'))|, \\ \text{codis}^1(\varphi, \psi) &:= \sum_{(x, y) \in X \times Y} |\omega_X(x, \psi(y)) - \omega_Y(\varphi(x), y)|. \end{aligned}$$

Then, define $d_{\mathcal{N}}^{1, \text{map}}(X, Y) := \frac{1}{2} \min_{\varphi, \psi} \{ \max\{\text{dis}^1(\varphi), \text{dis}^1(\psi), \text{codis}^1(\varphi, \psi), \text{codis}^1(\psi, \varphi)\} \}$.

Unlike the ℓ_∞ case, here we have $d_{\mathcal{N}}^{1, \text{map}} \neq d_{\mathcal{N}}^1$. However, for the purpose of stability, it suffices to use the inequality shown in Proposition 5 below.

► **Proposition 5.** Let $\mathcal{X}$ and $\mathcal{Y}$ be two networks. Then, $d_{\mathcal{N}}^{1, \text{map}}(\mathcal{X}, \mathcal{Y}) \leq d_{\mathcal{N}}^1(\mathcal{X}, \mathcal{Y})$.

An example where the inequality in Proposition 5 is strict is shown in Figure 2. We now want to obtain stability for the ℓ_1 -distance:

$$d_B(\text{Dgm}_k^{\text{WL}}(\mathcal{X}), \text{Dgm}_k^{\text{WL}}(\mathcal{Y})) \leq 2 d_{\mathcal{N}}^{1, \text{map}}(\mathcal{X}, \mathcal{Y}) \leq 2 d_{\mathcal{N}}^1(\mathcal{X}, \mathcal{Y}).$$

A proof for this inequality will be included in the full version of the paper.

4 Cycle Networks

Lastly, we turn our attention to a specific type of directed graph: cycle networks. For $n \geq 3$, let D_n be the cycle graph given by a full directed cycle with n vertices where all edges have weight 1. Next, we define $\tilde{D}_n$ as a modification: Swap one of the edges, say (x_1, x_2) , to (x_2, x_1) , keeping its weight 1; to maintain strong connectivity, also add the original edge (x_1, x_2) with weight n . We define the *cycle network* (G_n, ω_{G_n}) and *semicycle network* $(\tilde{G}_n, \omega_{\tilde{G}_n})$ as the respective associated shortest-distance digraphs. We now compare results with Dowker persistence [2]. See Figure 3 for an example with 6 vertices.

For cycle networks, we have $\text{Dgm}_1^{\text{WL}}(G_n) = \text{Dgm}_1^{\mathfrak{D}}(G_n) = \{(1, \lceil n/2 \rceil) \in \mathbb{R}^2\}$. For semicycle networks, we conjecture that

$$\text{Dgm}_1^{\mathfrak{D}}(\tilde{G}_n) = \{(1, \lceil (n-1)/2 \rceil) \in \mathbb{R}^2\} \quad \text{and} \quad \text{Dgm}_1^{\text{WL}}(\tilde{G}_n) = \{(1, n-1) \in \mathbb{R}^2\}.$$

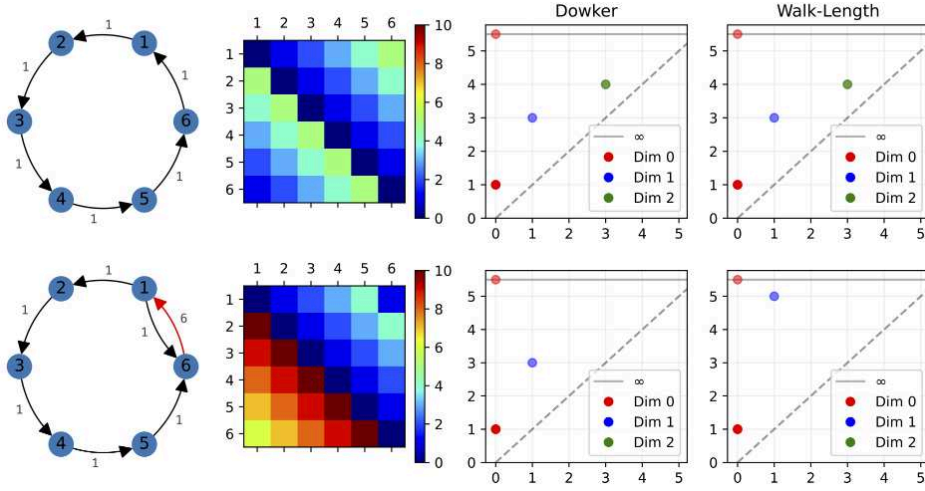


Figure 3 Comparison between Dowker and walk-length filtrations. On the top, a cycle with all weights equal to 1. Then, add a vertex $(1,6)$ with weight 1 and change weight of $(6,1)$ to 6.

The relevance of these semicycle networks $\tilde{G}_n$ lies in that they hint at an interesting property of the walk-length filtration: a higher sensitivity to directionality in cycles, resulting in non-directed cycles dying later than complete directed cycles.

References

- 1 Mehmet E. Aktas, Esra Akbas, and Ahmed El Fatmaoui. Persistence homology of networks: methods and applications. *Applied Network Science*, 4(1), aug 2019. doi:10.1007/s41109-019-0179-3.
- 2 Samir Chowdhury and Facundo Mémoli. A functorial Dowker theorem and persistent homology of asymmetric networks. *Journal of Applied and Computational Topology*, 2(1):115–175, 2018. doi:10.1007/s41468-018-0020-6.
- 3 Samir Chowdhury and Facundo Mémoli. Persistent path homology of directed networks. In *Proceedings of the 2018 Annual ACM-SIAM Symposium on Discrete Algorithms (SODA)*, pages 1152–1169. ACM-SIAM, 2018. URL: <https://epubs.siam.org/doi/abs/10.1137/1.9781611975031.75>, arXiv:<https://epubs.siam.org/doi/pdf/10.1137/1.9781611975031.75>, doi:10.1137/1.9781611975031.75.
- 4 Dejan Govc. Computing homotopy types of directed flag complexes, June 2020. arXiv preprint arXiv:2006.05333. arXiv:2006.05333, doi:10.48550/ARXIV.2006.05333.
- 5 David Méndez and Rubén J. Sánchez-García. A directed persistent homology theory for dissimilarity functions. *Journal of Applied and Computational Topology*, 7(4):771–813, June 2023. doi:10.1007/s41468-023-00124-x.
- 6 Michael W. Reimann, Max Nolte, Martina Scolamiero, Katharine Turner, Rodrigo Perin, Giuseppe Chindemi, Pawel Dlotko, 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. URL: <http://journal.frontiersin.org/article/10.3389/fncom.2017.00048>, doi:10.3389/fncom.2017.00048.
- 7 Katharine Turner. Rips filtrations for quasi-metric spaces and asymmetric functions with stability results. *Algebraic & Geometric Topology*, 19(3):1135–1170, August 2019. arXiv:1608.00365, doi:10.2140/agt.2019.19.1135.