Graph Reconstruction by Discrete Morse Theory

Tamal K. Dey
The Ohio State University, Columbus, OH, USA
tamaldey@cse.ohio-state.edu

Jiayuan Wang
The Ohio State University, Columbus, OH, USA
wang.6195@buckeyemail.osu.edu

Yusu Wang
The Ohio State University, Columbus, OH, USA
yusu@cse.ohio-state.edu

Abstract

Recovering hidden graph-like structures from potentially noisy data is a fundamental task in modern data analysis. Recently, a persistence-guided discrete Morse-based framework to extract a geometric graph from low-dimensional data has become popular. However, to date, there is very limited theoretical understanding of this framework in terms of graph reconstruction. This paper makes a first step towards closing this gap. Specifically, first, leveraging existing theoretical understanding of persistence-guided discrete Morse cancellation, we provide a simplified version of the existing discrete Morse-based graph reconstruction algorithm. We then introduce a simple and natural noise model and show that the aforementioned framework can correctly reconstruct a graph under this noise model, in the sense that it has the same loop structure as the hidden ground-truth graph, and is also geometrically close. We also provide some experimental results for our simplified graph-reconstruction algorithm.

1 Introduction

Recovering hidden structures from potentially noisy data is a fundamental task in modern data analysis. A particular type of structure often of interest is the geometric graph-like structure. For example, given a collection of GPS trajectories, recovering the hidden road network can be modeled as reconstructing a geometric graph embedded in the plane. Given the simulated density field of dark matters in universe, finding the hidden filamentary structures is essentially a problem of geometric graph reconstruction.

Different approaches have been developed for reconstructing a curve or a metric graph from input data. For example, in computer graphics, much work have been done in extracting
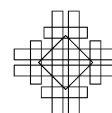
34th International Symposium on Computational Geometry (SoCG 2018).
Editors: Bettina Speckmann and Csaba D. Tóth; Article No. 31; pp. 31:1–31:15

Leibniz International Proceedings in Informatics
LIPIcs Schloss Dagstuhl – Leibniz-Zentrum für Informatik, Dagstuhl Publishing, Germany

1D skeleton of geometric models using the medial axis or Reeb graphs [7, 26, 19, 15, 21, 4]. In computer vision and machine learning, a series of work has been developed based on the concept of *principal curves*, originally proposed by Hastie and Steutzle [17]. Extensions to graphs include the work in [18] for 2D images and in [22] for high dimensional point data.

In general, there is little theoretical guarantees for most approaches developed in practice to extract hidden graphs. One exception is some recent work in computational topology: Aanijaneya et al. [1] proposed the first algorithm to approximate a metric graph from an input metric space with guarantees. The authors of [5, 15] used Reeb-like structures to approximate a hidden (metric) graph with some theoretical guarantees. These work however only handles (Gromov-)Hausdorff-type of noise. When input points are embedded in an ambient space, they requires the input points to lie within a small tubular neighborhood of the hidden graph. Empirically, these methods do not seem to be effective when the input contains ambient noise allowing some faraway points from the hidden graph.

Recently, a discrete Morse-based framework for recovering hidden structures was proposed and studied [6, 16, 23]. This line of work computes and simplifies a discrete analog of (un)stable manifolds of a Morse function by using the (Forman's) discrete Morse theory coupled with persistent homology for 2D or 3D volumetric data. One of the main issues in such simplification is the inherent obstructions that may occur for cancelling critical pairs. The authors of [23] suggest sidestepping this and consider a combinatorial representation of critical pairs for further processing. The authors in [6] identify a restricted set of pairs called "cancellable close pairs" which are guaranteed to admit cancellation. This framework has been applied to, for example, extracting filament structures from simulated dark matter density fields [24] and reconstructing road networks from GPS traces [25].

This persistence-guided discrete Morse-based framework has shown to be very effective in recovering a hidden geometric graph from (non-Hausdorff type) noise and non-homogeneous data. The method draws upon the global topological structure hidden in the input scalar field and thus is particularly effective at identifying junction nodes which has been a challenge for previous approaches that rely mostly on local information. However, to date, theoretical understanding of such a framework remains limited. Simplification of a discrete Morse gradient vector field using persistence has been studied before. For example, the work of [6] clarifies the connection between persistence-pairing and the simplification of *discrete Morse chain complex* (which is closely related, but different from the cancellation in the discrete gradient vector field) for 2D and 3D domains. Bauer et al. [3] obtain several results on persistence guided discrete Morse simplification for combinatorial surfaces. The simplification of vertex-edge persistence pairing used in [3] has also been observed in [2] independently for simplifying Morse functions on surfaces. Leveraging these existing developments, we aim to provide a theoretical understanding of a persistence-guided discrete Morse based approach to reconstruct a hidden geometric graph.

Main contributions and organization of paper.   In Section 3, we start with one version of the existing persistence-guided discrete Morse-based graph reconstruction algorithm (as employed in [24, 25, 8]). We show that this algorithm can be significantly simplified while still yielding the same output. To establish the theoretical guarantee of the reconstruction algorithm, we introduce a simple yet natural noise model in Section 4. Intuitively, this noise model assumes that we are given an input density field $\rho : \mathbb{R}^d \to \mathbb{R}$ where densities are significantly higher within a small neighborhood around a hidden graph than outside it. Under this noise model, we show that the reconstructed graph has the same loop structure as the hidden graph, and is also geometrically close to it; the technical details are in Sections 5 and

6 for the general case and the 2-dimensional case (with additional guarantees), respectively.

While our noise model is simple, our theoretical guarantees are first of a kind developed for a discrete Morse-based approach applied to graph reconstruction. In fact, prior to this, it was not clear whether a discrete Morse based approach can recover a graph even if there is no noise, that is, the density function has positive values *only* on the hidden graph. For our specific noise model, it may be possible to develop thresholding strategies perhaps with theoretical guarantees. However, previous work (e.g, [24, 25]) have shown that discrete Morse approach succeeds in many cases handling non-homogeneous data where thresholding fails. We have implemented the proposed simplified algorithm and tested it on several data sets, which generally gives a speed-up of at least a factor of 2 over a state-of-the-art approach. We present more discussions and experimental results in the full version of this paper [9].

2 Preliminaries

2.1 Morse theory

For simplicity, we consider only a smooth function $f : \mathbb{R}^d \to \mathbb{R}$. See [10, 20] for more general discussions.

For a point $p \in \mathbb{R}^d$, the gradient vector of f at a point p is $\nabla f(p) = -[\frac{\partial f}{\partial x_1} \cdots \frac{\partial f}{\partial x_d}]^T$, which represents the steepest descending direction of f at p, with its magnitude being the rate of change. An integral line of f is a path $\pi : (0, 1) \to \mathbb{R}^d$ such that the tangent vector at each point p of this path equals $\nabla f(p)$, which is intuitively a flow line following the steepest descending direction at any point. A point $p \in \mathbb{R}^d$ is *critical* if its gradient vector vanishes, i.e, $\nabla f(p) = [0 \cdots 0]^T$. A *maximal* integral line necessarily "starts" and "ends" at critical points of f; that is, $\lim_{t \to 0} \pi(t) = p$ with $\nabla f(p) = [0 \cdots 0]^T$, and $\lim_{t \to 1} \pi(t) = q$ with $\nabla f(q) = [0 \cdots 0]^T$. See Figure 1a where we show the graph of a function $f : \mathbb{R}^2 \to \mathbb{R}$, and there is an integral line from p' to the minimum v_1.

For a critical point p, the union of p and all the points from integral lines flowing into p is referred to as the *stable manifold of p*. Similarly, for a critical point q, the union of q and all the points on integral lines starting from q is called the *unstable manifold of q*. The stable manifold of a minimum p intuitively corresponds the basin/valley around p in the terrain of f. The 1-stable manifolds of index $(d-1)$ saddles consist of pieces of curves connecting $(d-1)$-saddles to maxima – These curves intuitively capture "mountain ridges" of the terrain (graph of the function f); see Figure 1a for an example. Symmetrically, the unstable manifold of a maximum q corresponds to the mountain around q. The 1-unstable manifolds consist of a collection of curves connecting 1-saddles to minima, corresponding intuitively to the "valley ridges".

In this paper, we focus on a graph-reconstruction framework using Morse-theory (as in e.g, [16, 6, 24, 25]). Intuitively, the 1-stable manifolds of saddles (mountain ridges) of the density field ρ are used to capture the hidden graphs. To implement such an idea in practice, the *discrete Morse theory* is used for robustness and simplicity contributed by its combinatorial nature; and a simplification scheme guided by the persistence pairings is employed to remove noise. Below, we introduce some necessary background notions in these topics.

2.2 Discrete Morse theory

First we briefly describe some notions from discrete Morse theory (originally introduced by Forman [14]) in the simplicial setting.

(a)   **(b)**   **(c)**   **(d)**
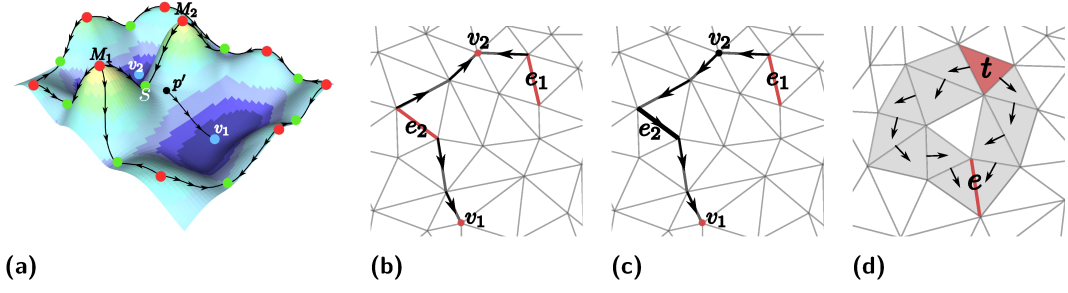
■ **Figure 1** (a) M_1 and M_2 are maxima (red dots), v_1 and v_2 are minima (blue dots), s is a saddle (green dots) with its stable manifolds flowing to it from M_1 and M_2. If we put a drop of water at p' it will flow to v_1. If we put it on the other side of the mountain ridge it will flow to minimum v_2. (b) Before cancellation of pair $\langle v_2, e_2 \rangle$. (c) After cancellation, the path from e_2 to v_2 is inverted, giving rise to a gradient path from e_1 to v_1, making $\langle v_1, e_1 \rangle$ now potentially cancellable. (d) An edge-triangle pair $\langle e, t \rangle$ which is not cancellable as there are two gradient paths between them.

A k-simplex $\tau = \{p_0, \ldots, p_k\}$ is the convex hull of $k+1$ affinely independent points; k is called the *dimension* of τ. A *face σ of τ* is a simplex spanned by a proper subset of vertices of τ; σ is a *facet* of the k-simplex τ, denoted by $\sigma < \tau$, if its dimension is $k-1$.

Suppose we are given a simplicial complex K which is simply a collection of simplices and all their faces so that if two simplices intersect, they do so in a common face. A *discrete (gradient) vector* is a pair of simplices (σ, τ) such that $\sigma < \tau$. A *Morse pairing* in K is a collection of discrete vectors $M(K) = \{(\sigma, \tau)\}$ where each simplex appears in *at most* one pair; simplices that are not in any pair are called *critical*.

Given a Morse pairing $M(K)$, a *V-path* is a sequence $\tau_0, \sigma_1, \tau_1, \ldots, \sigma_\ell, \tau_\ell, \sigma_{\ell+1}$, where $(\sigma_i, \tau_i) \in M(K)$ for every $i = 1, \ldots, \ell$, and each σ_{i+1} is a facet of τ_i for each $i = 0, \ldots, \ell$. If $\ell = 0$, the V-path is *trivial*. This V-path is *cyclic* if $\ell > 0$ and $(\sigma_{\ell+1}, \tau_0) \in M(K)$; otherwise, it is *acyclic* in which case we call this V-path a *gradient path*. We say that a gradient path is a vertex-edge gradient path if $dimension(\sigma_i) = 0$, implying that $dimension(\tau_i) = 1$. Similarly, it is a edge-triangle gradient path if $dimension(\sigma_i) = 1$. A Morse pairing $M(K)$ becomes a *discrete gradient vector field* (or equivalently a *gradient Morse pairing*) if there is no cyclic V-path induced by $M(K)$.

Intuitively, given a discrete gradient vector field $M(K)$, a gradient path $\tau_0, \sigma_1, \ldots, \tau_\ell, \sigma_{\ell+1}$ is the analog of an integral line in the smooth setting. But different from the smooth setting, a maximal gradient path may not start or end at critical simplices. However, those that do (i.e, when τ_0 and σ_{k+1} are critical simplices) are analogous to maximal integral line in the smooth setting which "start" and "end" at critical points, and for convenience one can think of *critical k-simplices* in the discrete Morse setting as *index-k critical points* in the smooth setting. For example, for a function on $\mathbb{R}^2$, critical 0-, 1- and 2-simplices in the discrete Morse setting correspond to minima, saddles and maxima in the smooth setting, respectively.

For a critical edge e, we define its *stable manifold* to be the union of edge-triangle gradient paths that ends at e. Its *unstable manifold* is defined to be the union of vertex-edge gradient paths that begins with e. While earlier we use "mountain ridges" (1-stable manifolds) to motivate the graph reconstruction framework, algorithmically (especially for the Morse cancellations below), vertex-edge gradient paths are simpler to handle. Hence in our algorithm below, we in fact consider the function $g_\rho = -\rho$ (instead of the density field ρ itself) and the algorithm outputs (a subset of) the *1-unstable manifolds* (vertex-edge paths in the discrete setting) as the recovered hidden graph.

Morse cancellation / simplification.   One can simplify a discrete gradient vector field $M(K)$ (i.e, reducing the number of critical simplices) by the following Morse cancellation operation: A pair of critical simplices $\langle \sigma, \tau \rangle$ with $dimension(\tau) = dimension(\sigma) + 1$ is *cancellable*, if there is a *unique* gradient path $\tau = \tau_0, \sigma_1, \ldots, \tau_\ell, \sigma_{\ell+1} = \sigma$ starting at the $k + 1$-simplex τ and ends at the k-simplex σ. The *Morse cancellation operation on $\langle \sigma, \tau \rangle$* then modifies the vector field $M(K)$ by removing all gradient vectors (σ_i, τ_i), for $i = 1, \ldots, \ell$, while adding new gradient vectors (σ_i, τ_{i-1}), for $i = 1, \ldots, \ell + 1$. Intuitively, the gradient path is inverted. Note that $\tau = \tau_0$ and $\sigma = \sigma_{\ell+1}$ are no longer critical after the cancellation as they now participate in discrete gradient vectors. If there is no gradient path, or more than one gradient path between this pair of critical simplices $\langle \sigma, \tau \rangle$, then this pair is *not cancellable* – the uniqueness condition is to ensure that no cyclic V-paths are formed after the cancellation operation. See Figure 1 (b) – (d) for examples.

2.3   Persistence pairing

The Morse cancellation can be applied to any sequence of critical simplices pairs as long as they are cancellable at the time of cancellation. There is no canonical cancellation sequence. To cancel features corresponding to "noise" w.r.t. an input piecewise-linear function $f : |K| \to \mathbb{R}$, a popular strategy is to guide the Morse cancellation by the persistent homology induced by the lower-star filtration [16, 24], which we introduce now.

Filtrations and lower-star filtration.   Given a simplicial complex K, let S be an ordered sequence $\sigma_1, \ldots, \sigma_N$ of all n simplices in K so that for any simplex $\sigma_i \in K$, all of its faces appear before it in S. Then S induces a *(simplex-wise) filtration $F(K)$*: $K_1 \subset K_2 \subset \cdots \subset K_N = K$, where $K_i = \bigcup_{j \leq i} \sigma_j$ is the subcomplex formed by the prefix $\sigma_1, \ldots, \sigma_i$ of S. Passing to homology groups, we have a *persistence module $H_*(K_1) \to \cdots \to H_*(K_N)$*, which has a unique decomposition into the direct sum of a set of indecomposable summands that can be represented by the set of *persistence-pairing $P(K)$ induced by $F(K)$*: Each *persistence pair $(\sigma_i, \sigma_j) \in P(K)$* indicates that a new k-th homological class, $k = dimension(\sigma_i)$, is created at K_i and destroyed at K_j; σ_i is thus called a *positive simplex* as it creates, and σ_j a *negative simplex*. Assuming that there is a *simplex-wise function $\bar{f} : K \to \mathbb{R}$* such that $\bar{f}(\sigma_i) \leq \bar{f}(\sigma_j)$ if $i < j$, then the *persistence* of the pair (σ, τ) is defined as $\mathrm{pers}(\sigma) = \mathrm{pers}(\tau) = \mathrm{pers}(\sigma, \tau) = \bar{f}(\tau) - \bar{f}(\sigma)$. Some simplices σ_ℓ's may be unpaired, meaning that homological features created at K_ℓ are never destroyed. We augment $P(K)$ by adding (σ_ℓ, ∞) for every unpaired simplex σ_ℓ to it, and set $\mathrm{pers}(\sigma_\ell, \infty) = \infty$.

The persistent homology can be defined for any filtration of K. In our setting, there is an input function $f : V(K) \to \mathbb{R}$ defined at the vertices $V(K)$ of K whose linear extension leads to a piecewise-linear (PL) function still denoted by $f : |K| \to \mathbb{R}$. To reflect topological features of f, we use the lower-star filtration of K induced by f: Specifically, for any vertex $v \in V(K)$, its lower-star $\mathrm{LowSt}(v)$ is the set of simplicies containing v where v has the highest f value among their vertices. Now sort vertices of K in non-decreasing order of their f-values: $v_1, \ldots, v_n$. An ordered sequence $S = \langle \sigma_1, \ldots, \sigma_N \rangle$ induces a *lower-star filtration $F_f(K)$ of K w.r.t. f* if S can be partitioned to n consecutive pieces $\langle \sigma_1, \ldots, \sigma_{I_1} \rangle$, $\langle \sigma_{I_1+1}, \ldots, \sigma_{I_2} \rangle$, $\ldots, \langle \sigma_{I_{n-1}+1}, \ldots, \sigma_N \rangle$, such that the i-th piece $\langle \sigma_{I_{i-1}+1}, \ldots, \sigma_{I_i} \rangle$ equals $\mathrm{LowSt}(v_i)$.

Now let $P_f(K)$ be the resulting set of persistence pairs induced by the lower-star filtration $F_f(K)$. Extend the function $f : V(K) \to \mathbb{R}$ to a simplex-wise function $\bar{f} : K \to \mathbb{R}$ where $\bar{f}(\sigma) = \max_{v \in \sigma} f(v)$ (i.e, $\bar{f}(\sigma)$ is the highest f-value of any of its vertices). For each pair (σ, τ), we measure its persistence by $\mathrm{pers}(\sigma, \tau) = \bar{f}(\tau) - \bar{f}(\sigma)$. Every simplex in K contributes to a persistence pair in $P_f(K)$. However, assuming the value of f is distinct on all vertices,

then those persistence pairs with zero-persistence are "trivial" in the sense they correspond to the local pairing of two simplices from the lower-star of the same vertex. A persistence pair (σ, τ) *with positive persistence* corresponds to a pair of (homological) critical points (p, q) for the PL-function $f : |K| \to \mathbb{R}$ [10] induced by the function f on $V(K)$, with $p \in \sigma$ and $q \in \tau$.

3 Reconstruction algorithm

Problem setup. Suppose we have a domain Ω (which will be a cube in $\mathbb{R}^d$ in this paper) and a density function $\rho : \Omega \to \mathbb{R}$ (that "concentrates" around a hidden geometric graph $G \subset \Omega$). In the discrete setting, our input will be a triangulation K of Ω and a density function given as a PL-function $\rho : K \to \mathbb{R}$. Our goal is to compute a graph $\widehat{G}$ approximating the hidden graph G. In Algorithm 1, we first present a *known* discrete Morse-based graph (1-skeleton) reconstruction framework, which is based on the approaches in [16, 6, 24, 25].

Algorithm 1: MorseRecon(K,ρ, δ)

Data: Triangulation K of Ω, density function $\rho : K \to \mathbb{R}$, threshold δ

Result: Reconstructed graph $\widehat{G}$

begin

1 Compute persistence pairings $P(K)$ by the lower-star filtration of K w.r.t $g_\rho = -\rho$

2 $M =$PerSimpVF($P(K), \delta$)

3 $\widehat{G} = $ CollectOutputG(M)

4 **return** $\widehat{G}$

Procedure PerSimpVF*($P(K), \delta$)*

1 Set initial discrete gradient field M on K to be trivial

2 Rank all persistence pairs in $P(K)$ in increasing order of their persistence

3 **for** *each $(\sigma, \tau) \in P(K)$ with* $\mathrm{pers}(\sigma, \tau) \le \delta$ **do**

4 ⎿ If possible, perform discrete-Morse cancellation of (σ, τ) and update the discrete gradient vector field M

5 **return** M

Procedure CollectOutputG*(M)*

1 $\widehat{G} = \emptyset$

2 **for** *each remaining critical edge e with* $\mathrm{pers}(e) > \delta$ **do**

3 ⎿ $\widehat{G} = \widehat{G} \bigcup \{\text{1-unstable manifold of } e\}$

4 **return** $\widehat{G}$

Intuitively, we wish to use "mountain ridges" of the density field to approximate the hidden graph, which are computed as the 1-unstable manifolds of $g_\rho = -\rho$, the negation of the density function. Specifically, after initializing the discrete gradient vector field M to be a trivial one, a persistence-guided Morse cancellation step is performed in Procedure PerSimpVF() to compute a new discrete gradient vector field M_δ so as to capture only important (high persistent) features of g_ρ. In particular, Morse-cancellation is performed for each pair of critical simplices from $P(K)$ (if possible) in increasing order of persistence values (for pairs with equal persistence, we use the nested order as in [3]). Finally, the union of the 1-unstable manifolds of all remaining high-persistence critical edges is taken as the output graph $\widehat{G}$, as outlined in Procedure CollectOutputG().

Algorithm 2: MorseReconSimp(K,ρ, δ)

    Procedure PerSimpTree*($P(K), \delta$) /∗ This procedure replaces original* PerSimpVF*() ∗/*

1 $\Pi :=$ the set of vertex-edge persistence pairs from $P(K)$

2   Set $\Pi_{\leq \delta} \subseteq \Pi$ to be $\Pi_{\leq \delta} = \{(v, e) \in \Pi \mid \mathrm{pers}(v, e) \leq \delta\}$

3 $\mathcal{T} := \bigcup_{(v,\sigma) \in \Pi_{\leq \delta}} \{\sigma = \langle u_1, u_2 \rangle, u_1, u_2\}$

4   **return** $\mathcal{T}$

    Procedure Treebased-OutputG*($\mathcal{T}$) /∗ This procedure replaces* CollectOutputG*() ∗/*

1 $\widehat{G} = \emptyset$

2   **for** *each edge* $e = \langle u, v \rangle$ *with* $\mathrm{pers}(e) > \delta$ **do**

3   Let $\pi(u)$ be the unique path from u to the sink of the tree T_i containing u

4   Define $\pi(v)$ similarly;   Set $\widehat{G} = \widehat{G} \cup \pi(u) \cup \pi(v) \cup \{e\}$

5   **return** $\widehat{G}$

Since we only need 1-unstable manifolds, K is assumed to be a 2-complex. It is pointed out in [8] that in fact, instead of performing Morse-cancellation for all critical pairs involving edges (i.e, vertex-edge pairs and edge-triangle pairs), one only needs to cancel vertex-edge pairs – This is because only vertex-edge gradient vectors will contribute to the 1-unstable manifolds, and also new vertex-edge vectors can only be generated while canceling other vertex-edge pairs. Hence in PerSimpVF(), we can consider *only vertex-edge pairs* $(\sigma, \tau) \in P$ in order. Furthermore, it is not necessary to check whether the cancellation is valid or not – it will always be valid as long as the pairs are processed in increasing orders of persistence [3][1].

However, we can further simplify the algorithm as follows: First, we replace procedure PerSimpVF() by procedure PerSimpTree() as shown in Algorithm 2, which is much simpler both conceptually and implementation speaking. Note that there is *no explicit cancellation operation* any more.

The 1-dimensional simplicial complex $\mathcal{T}$ output by procedure PerSimpTree() may have multiple connected components $\mathcal{T} = \{T_1, \ldots, T_k\}$ – In fact, it is known that each T_i is a tree and $\mathcal{T}$ is a forest (see results from [2, 3] as summarized in Lemma 2 below). For each component T_i, we define its *sink*, denoted by $\mathrm{si}(T_i)$, as the vertex $\mathsf{v}_i \in T_i$ with the lowest function $g_\rho = -\rho$ value. Lemma 2 also states that the sink of T_i would have been the *only critical simplex* among all simplices in T_i, if we had performed the δ-simplification as specified in procedure PerSimpVF(). Next, we replace procedure CollectOutputG() by procedure Treebased-OutputG() shown in Algorithm 2. We use MorseReconSimp() to denote our simplified version of Algorithm 1 (with PerSimpVF() replaced by PerSimpTree(), and CollectOutputG() replaced by Treebased-OutputG(). In summary, algorithm MorseReconSimp(K, ρ, δ) works by first computing all persistence pairs as before. It then collects all vertex-edge persistence pairs (v, e) with $\mathrm{pers}(v, e) \leq \delta$. These edges along with the set of all vertices form a spanning forest $\mathcal{T}$. Then, for every edge $e = \langle u, v \rangle$ with $\mathrm{pers}(e) > \delta$, it outputs the 1-unstable manifold of e, which is simply the union of e and the unique paths from u and v to the sink (root) of the tree containing them respectively. Its time complexity is stated below; note for the previous algorithm MorseRecon(), the cancellation step can take $\tilde{O}(n^2)$ time.

---

[1] We remark that though [3] states that the cancellation is not valid in higher dimension or non-manifold 2-complexes, all cancellations in PerSimpVF() are for vertex-edge pairs in a spanning tree which can be viewed as a 1-complex, and thus are always valid.

▶ **Theorem 1.** *The time complexity of our Algorithm* PerSimpVF*() is $O(PerT(K)+n)$, where $PerT(K)$ is the time to compute persistence pairings for K, and n is the total number of vertices and edges in K.*

We remark that the $O(n)$ term is contributed by the step collecting all 1-unstable manifolds, which takes linear time if one avoids revisiting edges while tracing the paths.

Justification of the modified algorithm MorseReconSimp().   Let M_δ denote the resulting discrete gradient field after canceling all *vertex-edge* persistence pairs in $P(K)$ with persistence at most δ; that is, M_δ is the output of the procedure PerSimpVF() (although we only compute the relevant part of the discrete gradient vector field). Using observations from [2, 3], we show that the output $\mathcal{T}$ of procedure PerSimpTree() includes all information of M_δ. Furthermore, procedure Treebased-OutputG() computes the correct 1-unstable manifolds for all critical edges with persistence larger than δ. Indeed, observe that edges in Morse pairings from M_δ (for any $\delta \geq 0$) form a spanning forest of edges in K. Results of [3] imply that the output $\mathcal{T}$ constructed by our modified procedure corresponds exactly to this spanning forest:

▶ **Lemma 2.** *The following statements hold for the output $\mathcal{T}$ of procedure* PerSimpTree*() w.r.t any $\delta \geq 0$:*
(i) *$\mathcal{T}$ is a spanning forest consisting of potentially multiple trees $\{T_1, \ldots, T_k\}$.*
(ii) *For each tree T_i, its sink v_i is the only critical simplex in M_δ. The collection of v_is corresponds exactly to those vertices whose persistence is bigger than δ.*
(iii) *Any edge with $\mathrm{pers}(e) > \delta$ remains critical in M_δ (and cannot be contained in $\mathcal{T}$).*

Note that, (ii) above implies that for each T_i, any discrete gradient path of M_δ in T_i terminates at its sink v_i. See the right figure for an illustration. Hence for any vertex $v \in T_i$, the path $\pi(v)$ computed in procedure Treebased-OutputG() is the unique discrete gradient path starting at v. This immediately leads to the following result:
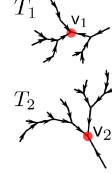
▶ **Corollary 3.** *For each critical edge $e = \langle u, v \rangle$ with $\mathrm{pers}(e) \geq \delta$, $\pi(u) \cup \pi(v) \cup \{e\}$ as computed in procedure* Treebased-OutputG*() is the 1-unstable manifold of e in M_δ. Hence the output of our simplified algorithm* MorseReconSimp*() equals that of the original algorithm* MorseRecon*().*

4   Noise model

We first describe the noise model in the continuous setting where the domain is $\Omega = [0,1]^d$. We then explain the setup in the discrete setting when the input is a triangulation K of Ω.

Given a connected "true graph" G $\subset \Omega$, consider a ω-*neighborhood* $G^\omega \subseteq \Omega$, meaning that (i) G $\subseteq G^\omega$, and (ii) for any $x \in G^\omega$, $d(x, \mathrm{G}) \leq \omega$ (i.e, G^ω is sandwiched between G and its ω-offset). Given G^ω, we use $\mathrm{cl}(\overline{G^\omega})$ to denote the closure of its complement $\mathrm{cl}(\overline{G^\omega}) = \mathrm{cl}(\Omega \setminus G^\omega)$. See the right figure, showing G (red graph) with its ω-neighborhood G^ω (orange).
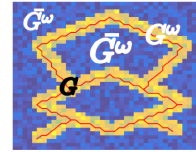
▶ **Definition 4.** A density function $\rho : \Omega \rightarrow \mathbb{R}$ is a (β, ν, ω)-*approximation of a connected graph* G if the following holds:
C-1 There is a ω-neighborhood G^ω of G such that G^ω deformation retracts to G.
C-2 $\rho(x) \in [\beta, \beta + \nu]$ for $x \in G^\omega$; and $\rho(x) \in [0, \nu]$ otherwise. Furthermore, $\beta > 2\nu$.

Intuitively, this noise model requires that the density ρ concentrates around the true graph G in the sense that the density is significantly higher inside G^ω than outside it; and the density fluctuation inside or outside G^ω is small compared to the density value in G^ω (condition C-2). Condition C-1 says that the neighborhood has the same topology of the hidden graph. Such a density field could for example be generated as follows: Imagine that there is an ideal density field $f_G : \Omega \to \mathbb{R}$ where $f_G(x) = \beta$ for $x \in G^\omega$ and 0 otherwise. There is a noisy perturbation $g : \Omega \to \mathbb{R}$ whose size is always bounded by $g(x) \in [0, \nu]$ for any $x \in \Omega$. The observed density field $\rho = f_G + g$ is an (β, ν, ω)-approximation of G.

In the discrete setting when we have a triangulation K of Ω, we define a ω-neighborhood G^ω to be a subcomplex of K, i.e, $G^\omega \subseteq K$, such that (i) G is contained in the underlying space of G^ω and (ii) for any vertex $v \in V(G^\omega)$, $d(v, G) \leq \omega$. The outside-region $\mathrm{cl}(\overline{G^\omega}) \subseteq K$ is simply the smallest subcomplex of K that contains all simplices from $K \setminus G^\omega$ (i.e, all simplices **not** in G^ω and their faces). A PL-function $\rho : K \to \mathbb{R}$ (β, ν, ω)-approximation of G can be extended to this setting by requiring the underlying space of G^ω deformation retracts to G as in (C-1), and having those density conditions in (C-2) only at vertices of K.

We remark that the noise model is still limited – In particular, it does not allow significant non-uniform density distribution. However, this is the first time that theoretical guarantees are provided for a discrete Morse based reconstruction framework, despite that such a framework has been used for different applications before. We also give experiments and discussions in Appendix B of the full version [9] that the algorithm works beyond this noise model empirical, where thresholding type approaches do not work.

5   Theoretical guarantee

In this section, we prove results that are applicable to any dimension. Recall that M_δ is the discrete gradient field after the δ-*Morse cancellation process*, where we perform Morse-cancellation for all *vertex-edge* persistence pairs from $P(K)$. (While our algorithm does not maintain M_δ explicitly, we use it for theoretical analysis.) At this point, all positive edges (i.e, those paired with triangles or unpaired in $P(K)$) remain critical in M_δ. Some negative edges (i.e, those paired with vertices in $P(K)$) are also critical in M_δ – these are exactly the negative edges with persistence bigger than δ. Treebased-OutputG() only takes the 1-unstable manifolds of those critical edges (positive or negative) with persistence bigger than δ; so those positive edges whose persistence is $\leq \delta$ (if there is any) are ignored.

From now on, we use "*under our noise model*" to refer to (1) the input is a (β, ν, ω)-approximated density field w.r.t. G, and (2) $\delta \in [\nu, \beta - \nu)$. Let $\widehat{G}$ be the output of algorithm MorseReconSimp(K, ρ, δ). The proof of the following result is in the full version [9].

▶ **Proposition 5.** *Under our noise model, we have:*
  (i) *There is a single critical vertex left after* PerSimpVF*() which is in* G^ω.
  (ii) *Every critical edge considered by* Treebased-OutputG*() forms a persistence pair with a triangle.*
  (iii) *Every critical edge considered by* Treebased-OutputG*() is in* G^ω.

▶ **Theorem 6.** *Under our noise model, the output graph satisfies* $\hat{G} \subseteq G^\omega$.

Proof. Recall that the output graph $\widehat{G}$ consists of the union of 1-unstable manifolds of all the edges $e_1^*, \ldots, e_g^*$ with persistence larger than δ – By Propositions 5 (ii) and (iii), they are all positive (paired with triangles), and contained inside G^ω.

Take any $i \in [1, g]$ and consider $e_i^* = \langle u, v \rangle$. Without loss of generality, consider the gradient path starting from u: $\pi : u = u_1, e_1, u_2, e_2, \ldots, u_s, e_s, u_{s+1}$. By Lemma 2 and

Proposition 5, u_{s+1} must be a critical vertex (a sink) and is necessarily the global minimum v_0, which is also contained inside G^ω. We now argue that the entire path π (i.e, all simplices in it) is contained inside G^ω. In fact, we argue a stronger statement: First, we say that a gradient vector (v, e) is *crossing* if $v \in G^\omega$ and $e \notin G^\omega$ (i.e, $e \in \mathrm{cl}(\overline{G^\omega})$) – Since v is an endpoint of e, this means that the other endpoint of e must lie in $K \setminus G^\omega$.

▶ **Claim 1.** *During the δ-Morse cancellation, no crossing gradient vector is ever produced.*

Proof. Suppose the lemma is not true: Then let (v, e) be the *first* crossing gradient vector ever produced during the δ-Morse cancellation process. Since we start with a trivial discrete gradient vector field, the creation of (v, e) can only be caused by reversing of some gradient path π' connecting two critical simplices v' and e' while we are performing Morse-cancellation for the persistence pair (v', e'). Obviously, $\mathrm{pers}(v', e') \leq \delta$. On the other hand, due to our (β, ν, ω)-noise model and the choice of δ, it must be that either both $v', e' \in G^\omega$ or both $v', e' \in K \setminus G^\omega$ – as otherwise, the persistence of this pair will be larger than $\beta - \nu > \delta$.

Now consider this gradient path π' connecting v' and e' in the current discrete gradient vector field M'. Since the pair (v, e) becomes a gradient vector after the inversion of this path, it must be that (w, e) currently is a gradient vector where $e = \langle v, w \rangle$. Furthermore, since the path π' begins and ends with simplices either both in G^ω or both outside it, the path π' must contain a gradient vector (v'', e'') going in the opposite direction crossing inside/outside, that is, $v'' \in G^\omega$ and $e'' \notin G^\omega$. In other words, it must contain a crossing gradient vector. This however contradicts to our assumption that (v, e) would be the first crossing gradient vector. Hence the assumption is wrong and no crossing gradient vector can ever be created. ◀

As there is no crossing gradient vector during and after δ-Morse cancellation, it follows that π, which is one piece of the 1-unstable manifold of the critical edge e_i^*, has to be contained inside G^ω. The same argument works for the other piece of 1-unstable manifold of e_i^* (starting from the other endpoint of e_i^*). Since this is for any $i \in [1, g]$, the theorem holds. ◀

The previous theorem shows that $\widehat{G}$ is close to G in geometry. Next we will show that they are also close in topology.

▶ **Proposition 7.** *Under our noise model, $\widehat{G}$ is homotopy equivalent to G.*

Proof. We show that $\widehat{G}$ has the same first Betti number as that of G which implies the claim as any two graphs in $\mathbb{R}^d$ with the same first Betti number are homotopy equivalent.

The underlying space of ω-neighborhood G^ω of G deformation retracts to G by definition. Observe that, by our noise model, G^ω is a sublevel set in the filtration that determines the persistence pairs. This sublevel set being homotopy equivalent to G must contain exactly g positive edges where g is the first Betti number of G. Each of these positive edges pairs with a triangle in $\overline{G^\omega}$. Therefore, $\mathrm{pers}(e, t) > \delta$ for each of the g positive edges in G^ω. By our earlier results, these are exactly the edges that will be considered by procedure Treebased-OutputG(). Our algorithm constructs $\widehat{G}$ by adding these g positive edges to the spanning tree each of which adds a new cycle. Thus, $\widehat{G}$ has first Betti number g. ◀

We have already proved that $\widehat{G}$ is contained in G^ω. This fact along with Proposition 7 can be used to argue that any deformation retraction taking (underlying space) G^ω to G also takes $\widehat{G}$ to a subset $G' \subseteq G$ where G' and G have the same first Betti number. In what follows, we use G^ω to denote also its underlying space.

▶ **Theorem 8.** *Let $F : G^\omega \times [0,1] \to G^\omega$ be any deformation retraction. Then, the restriction $F|_{\widehat{G}} : \widehat{G} \times [0,1] \to G^\omega$ is a homotopy from the embedding $\widehat{G}$ to $G' \subseteq G$ where G' is the minimal subset so that G and G' have the same first Betti number.*

Proof. The fact that $F|_{\widehat{G}}(\cdot, \ell)$ is continuous for any $\ell \in [0,1]$ is obvious from the continuity of F. Only thing that needs to be shown is that $F|_{\widehat{G}}(\widehat{G}, 1) = G'$. Suppose not. Then, $G'' = F|_{\widehat{G}}(\widehat{G}, 1)$ is a proper subset of G which has a first Betti number less than that of G. We observe that the cycle in $\widehat{G}$ created by a positive edge e along with the paths to the root of the spanning tree is also non-trivial in G^ω because this is a cycle created by adding the edge e during persistence filtration and the edge e is not killed in G^ω. Therefore, a cycle basis for $\widehat{G}$ is also a homology basis for G^ω. Since the map $F(\cdot, 1) : G^\omega \to G$ is a homotopy equivalence, it induces an isomorphism in the respective homology groups; in particular, a homology basis in G^ω is mapped to a homology basis in G. Therefore, the image $G'' = F|_{\widehat{G}}(\widehat{G}, 1)$ must have a basis of cardinality g if $\widehat{G}$ has first Betti number g. But, G'' cannot have a cycle basis of cardinality g if it is a proper subset of G' reaching a contradiction. ◀

6 Additional guarantee for 2D

For $\mathbb{R}^2$, we now show that G^ω actually deformation retracts to $\widehat{G}$, which is stronger than saying G and $\widehat{G}$ are homotopy equivalent. We are unable to prove this result for dimensions higher than 2, as our current proof needs that the edge-triangle persistence pairs can always be canceled (even though our algorithm does not depend on edge-triangle cancellations at all). It would be interesting, as a future work, to see whether a different approach can be developed to avoid this obstruction for the special case under our noise model. The main result of this section is as follows.

▶ **Theorem 9.** *Under our noise model, G^ω deformation retracts to G and $\hat{G}$.*

This main result follows from Proposition 10 and Theorem 11 below. To prove them, we will show that there exists a partition $\mathcal{R} := \{R_i\}$ of the set of triangles in K for which Theorem 11 holds. (This theorem is our main tool in establishing the deformation retract.) We first state the results below before giving their proofs. Let $B_i = \partial R_i$ where ∂ is the boundary operator operating on the 2-chain R_i. We also abuse the notations R_i and B_i to denote the geometric space that is the point-wise union of simplices in the respective chains. Let t_i be a triangle in R_i whose choice will be explained later. In the following, let H be the maximal set of edges in $\widehat{G}$ whose deletions do not eliminate a cycle (assume that a vertex is deleted only if all of its edges are deleted). Observe that H necessarily consists of negative edges forming "hairs" attached to the loops of $\widehat{G}$ and hence to $\cup_i B_i$ because of the following proposition.

▶ **Proposition 10.** *Under our noise model, $\widehat{G} = \cup B_i \bigcup H$.*

▶ **Theorem 11.** *Under our noise model, there exists a partition $\{R_i\}$ of triangles in K such that, there is a deformation retraction of $\cup_i (R_i \setminus t_i)$ to $\widehat{G}$ that comprises of two deformation retractions, one from $\cup_i (R_i \setminus t_i)$ to G^ω and another one from G^ω to $\cup_i B_i \bigcup H$ which is $\widehat{G}$.*

Now we describe the construction of a partition $\mathcal{R}$ of the triangles in K to prove Proposition 10 and Theorem 11. For technicality we assume that K is augmented to a triangulation of a sphere by putting a vertex v at infinity and joining it to the boundary of K with edges and triangles all of whom have function value ∞. Let $P(K)$ be the collection of persistence pairs of the form either (σ, τ) or (σ, ∞) generated from the lower-star filtration $F(K)$ as described before. Since K is 2-dimensional, each pair (σ, τ) is either a vertex-edge pair or an

edge-triangle pair. We order persistence pairs in $P(K)$ by their persistence, where ties are broken via the nested order in the filtration $F(K)$, and obtain:

$$P(K) = \{(\sigma_1, \tau_1), \ldots, (\sigma_n, \tau_n), (\alpha_1, \infty), \ldots, (\alpha_s, \infty)\}. \tag{1}$$

Starting with a trivial discrete gradient vector field M_0 where all simplices in K are critical, the algorithm PerSimpVF() performs Morse cancellations for the first $m \leq n$ persistence pairs $(\sigma_1, \tau_1), \ldots, (\sigma_m, \tau_m)$ in order where $\mathrm{pers}(\sigma_m, \tau_m) \leq \delta$ but $(\sigma_{m+1}, \tau_{m+1}) > \delta$, Let M_i denote the gradient vector field after canceling (σ_i, τ_i). Recall that in the implementation of the algorithm we do not need to perform Morse cancellation for any edge-triangle pairs. However in this section, for the theoretical analysis, we will cancel edge-triangle pairs as well. Recall that a positive edge is one that creates a 1-cycle, namely, it is either paired with a triangle or unpaired; while a negative edge is one that destroys a 0-cycle (i.e, paired with a vertex).

Consider the ordered sequence of edge-triangle persistence pairs, $(e_1, t_1), \ldots, (e_n, t_n)$, which is a subsequence of the one in (1). Consider the sequence $t_1, \ldots, t_n$ of triangles in K ordered by the above sequence. Recall the standard persistence algorithm [11]. It implicitly associates a 2-chain with a triangle t when searching for the edge it is about to pair with. This 2-chain is non-empty if t is a destructor, and is empty otherwise. Let D_i denote this 2-chain associated with t_i for $i \in [1, n]$. Initially, the algorithm asserts $D_i = t_i$. At any stage, if D_i is not empty, the persistence algorithm identifies the edge e in the boundary ∂D_i that has been inserted into the filtration $F(K)$ most recently. If e has not been paired with anyone, the algorithm creates the persistence pair (e, t_i). Otherwise, if e has already been paired with a triangle, say $t_{i'}$, then D_i is updated with $D_i = D_i + D_{i'}$ and the search continues. Given an index $j \in [1, n]$, we define a modified set of chains C_i^j inductively as follows. For $j = 1$, $C_i^1 = t_i$. Assume that C_i^{j-1} has been already defined. To define C_i^j, similar to the persistence algorithm, check if the edge e_{j-1} is on the boundary ∂C_i^{j-1}. If so, define $C_i^j := C_i^{j-1} + C_{j-1}^{j-1}$ and $C_i^j := C_i^{j-1}$ otherwise. The following result is proved in [9].

▶ **Proposition 12.** *For $i \in [1, n]$, e_i is in ∂C_i^i. Furthermore, e_i is the most recent edge in ∂C_i^i according to the filtration order $F(K)$.*

Procedure PerSimpVF() also implicitly maintains a 2-chain R_i^* with each triangle t_i. Initially, $R_i^* = t_i$ as in the case of D_i. Then, inductively assume that R_i^* is the 2-chain implicitly associated with t_i when a persistence pair $(e_{i'}, t_{i'})$ is about to be considered by PerSimpVF() and the boundary ∂R_i^* contains $e_{i'}$. By reversing a gradient path between $t_{i'}$ and $e_{i'}$, it implicitly updates the 2-chain R_i^* as $R_i^* := R_i^* + R_{i'}^*$. We observe that R_i^* is identical with $C_i^{i'}$. Proposition 13 below establishes this fact along with some other inductive properties useful to prove Theorem 11. The proof can be found in the full version [9].

▶ **Proposition 13.** *Let (e_j, t_j) be the edge-triangle persistence pair PerSimpVF() is about to consider and let C_i^j be the 2-chains defined as above . Then, the following statements hold:*
(a) *For each triangle t_i, $i = 1, \ldots, n$, in the persistence order, the 2-chain R_i^* satisfies the following conditions:(a.i) $R_i^* = C_i^j$, (a.ii) interpreting R_i^* as a set of triangles, one has that the sets R_i^*, $i = j, \ldots, n$, partition the set of all triangles in K.*
(b) *There is a gradient path from t_i to all edges of the triangles in R_i^*, and (b.i) the path is unique if the edge is in the boundary ∂R_i^* for every $i = j, \ldots, n$; (b.ii) if there is more than one gradient path from t_i to an edge e, then e must be a negative edge.*

We are now ready to setup the regions R_is needed for Theorem 11 and Proposition 10. Suppose the first m edge-triangle pairs have persistence less than or equal to δ, the parameter

supplied to PerSimpVF(). Then, we set $R_i = R_i^*$ as in Proposition 13 for $i \geq m+1, \ldots, n$. The proof for Proposition 10 is in the full version [9].

Finally, similar to the vertex-edge gradient vectors, we say that a gradient vector (e, t) is *crossing* if $e \in G^\omega$ and $t \notin G^\omega$. The following claim can be proved similarly as Claim 1.

▶ **Claim 2.** *During the δ-Morse cancellation of edge-triangle pairs, no crossing gradient vector is ever produced.*

Proof of Theorem 11. Set $\hat{R}_i = R_i \setminus t_i$. Let $\mathcal{T}$ be the spanning tree formed by all negative edges and their vertices. Let L_i be the set of edges in R_i that has more than one gradient path from t_i to them; $L_i \subset \mathcal{T}$ by Proposition 13 (b.ii). First, we want to establish a deformation retraction from $\cup(\hat{R}_i \setminus t_i)$ to G^ω. To do this, for $k = 0, 1, \ldots, s$, we will define $\hat{R}_i^k$ inductively where $\hat{R}_i^{k-1}$ deformation retracts to $\hat{R}_i^k$ and $\hat{R}_i^s \subseteq G^\omega \cup L_i$. Let $\hat{R}_i^0 = \hat{R}_i$. For $k = 1, \ldots, s$, consider a *positive* edge e in $\hat{R}_i^{k-1}$ where (a) e is not in G^ω and (b) there is a unique gradient path in R_i from t_i to e that passes through triangles all of which are in $R_i \setminus \hat{R}_i^{k-1}$. If such an edge e exists, then e is necessarily incident to a single triangle, say t, in $\hat{R}_i^{k-1}$. We collapse the pair (e, t), which is necessarily an edge-triangle gradient vector pair because e is positive. We take $\hat{R}_i^k$ to be $\hat{R}_i^{k-1} \setminus \{e, t\}$. If no such e exists, then either (A) there is no positive edge in $\hat{R}_i^{k-1} \setminus G^\omega$ any more; or (B) for each positive edge $e' \in \hat{R}_i^{k-1} \setminus G^\omega$, (B-1) there is a unique gradient path from t_i to e' but this path passes through some triangle in $\hat{R}_i^{k-1}$; or (B-2) there are two gradient paths from t_i to e'.

If there is no positive edge in $\hat{R}_i^{k-1} \setminus G^\omega$ any more, then $\hat{R}_i^{k-1} \subseteq G^\omega \cup L_i$, as otherwise, there will be at least some triangle from $\hat{R}_i^{k-1} \setminus G^\omega \cup L_i$ with at least one boundary edge of it being positive. The induction then terminates; we set $s = k-1$ and reach our goal.

We now show that case (B-1) is not possible. Suppose it happens, that is, $e' \in \hat{R}_i^{k-1} \setminus L_i$ is an edge not in G^ω for which the unique gradient path from t_i passes through triangles in $\hat{R}_i^{k-1}$. Let e'' be the first edge in this path that is in $\hat{R}_i^{k-1} \setminus L_i$. Then, if $e'' \notin G^\omega$, it qualifies for the conditions (a) and (b) required for e reaching a contradiction. So, assume $e'' \in G^\omega$. But, in that case, we have a gradient path that goes into G^ω and then comes out to reach $e' \notin G^\omega$. There has to be a gradient pair in this path where the edge is in G^ω and the triangle is not in G^ω. This contradicts Claim 2. Thus, case (B-1) is not possible. Now consider (B-2): e' must be negative by Proposition 13 (b.ii). So, it is not possible either.

To summarize, the induction terminates in case (A), at which time we would have that $\hat{R}_i^s \subseteq G^\omega \cup L_i$. Furthermore, this process also establishes a deformation retraction from $\hat{R}_i$ to $\hat{R}_i^s$ realized by successive collapses of edge-triangle pairs. Furthermore, by construction, each collapsed pair (e, t) must be from $\mathrm{cl}(\overline{G^\omega})$, hence $\cup_i \hat{R}_i^s$ contains all simplices in G^ω. Combined with that $\hat{R}_i^s \subseteq G^\omega \cup L_i$, we have that $\cup_i \hat{R}_i^s = G^\omega \cup L$, where $L = \cup_i L_i$ is a subset of the spanning tree $\mathcal{T}$. The edges in L being part of a spanning tree cannot form a cycle and thus can be retracted along the tree to G^ω, which gives rise to a deformation retraction from $\cup_i(R_i \setminus t_i)$ to $G^\omega \cup L$ and then to G^ω, establishing the first part of Theorem 11.

We now show that $(\cup_i \hat{R}_i^s) \bigcap G^\omega = G^\omega$ deformation retracts to $\cup B_i \bigcup H$. Let $\hat{L}_i$ be the edges in $\hat{R}_i^s \cap G^\omega$ with more than one gradient path from t_i to them. These edges are negative by Proposition 13 (b.ii). Replacing $\hat{R}_i$ with $\hat{R}_i^s \cap G^\omega$ and edges in $B_i \cup \hat{L}_i$ playing the role of edges in $G^\omega \cup L_i$ in the above induction, we can obtain that $\hat{R}_i^s \cap G^\omega$ deformation retracts to $B_i \cup \hat{L}_i$. Observe that now instead of Claim 2, we use the fact that no edge-triangle gradient path crosses B_i that consists of only negative and critical edges. To this end, we also observe that $\cup \hat{L}_i = \mathcal{T} \cap G^\omega$ where $\mathcal{T}$ is the spanning tree formed by all negative edges, as we only collapse edge-triangle pairs that are gradient pairs (hence the participating edges are always positive). This implies that $H \subset \cup \hat{L}_i$. Again, edges in $\hat{L}_i$ (being part of a spanning tree) can

be retracted along the spanning tree till one reaches B_i or edges in H. Performing this for each i, we thus obtain a deformation retraction from $(\cup_i \hat{R}_i^s) \bigcap G^\omega = G^\omega$ to $\cup B_i \bigcup \cup \hat{L}_i$ and further to $\cup B_i \bigcup H = \widehat{G}$. This finishes the proof of Theorem 11.   ◄

In the full version of this paper [9], we also provide some experiments demonstrating the efficiency of the simplified algorithm, as well as discussion on thresholding strategies.

References

1   M. Aanjaneya, F. Chazal, D. Chen, M. Glisse, L. Guibas, and D. Morozov. Metric graph reconstruction from noisy data. *International Journal of Computational Geometry & Applications*, 22(04):305–325, 2012.

2   D. Attali, M. Glisse, S. Hornus, F. Lazarus, and D. Morozov. Persistence-sensitive simplification of functions on surfaces in linear time. *Presented at TOPOINVIS*, 9:23–24, 2009.

3   U. Bauer, C. Lange, and M. Wardetzky. Optimal topological simplification of discrete functions on surfaces. *Discr. Comput. Geom.*, 47(2):347–377, 2012.

4   S. Biasotti, D. Giorgi, M. Spagnuolo, and B. Falcidieno. Reeb graphs for shape analysis and applications. *Theoretical Computer Science*, 392(1-3):5–22, 2008.

5   F. Chazal, R. Huang, and J. Sun. Gromov–hausdorff approximation of filamentary structures using reeb-type graphs. *Discr. Comput. Geom.*, 53(3):621–649, 2015.

6   O. Delgado-Friedrichs, V. Robins, and A. Sheppard. Skeletonization and partitioning of digital images using discrete morse theory. *IEEE Trans. Pattern Anal. Machine Intelligence*, 37(3):654–666, March 2015.

7   T. Dey and J. Sun. Defining and computing curve-skeletons with medial geodesic function. In *Sympos. Geom. Proc.*, volume 6, pages 143–152, 2006.

8   T. Dey, J. Wang, and Y. Wang. Improved road network reconstruction using discrete morse theory. In *Proc. 25th ACM SIGSPATIAL*. ACM, 2017.

9   T. Dey, J. Wang, and Y. Wang. Graph reconstruction by discrete morse theory. *arXiv preprint arXiv:1803.05093*, 2018.

10 H. Edelsbrunner and J. Harer. *Computational Topology - an Introduction.* American Mathematical Soc., 2010. URL: `http://www.ams.org/bookstore-getitem/item=MBK-69`.

11 H. Edelsbrunner, D. Letscher, and A. Zomorodian. Topological persistence and simplification. *Discr. Comput. Geom.*, 28:511–533, 2002.

12 The ENZO project. `http://enzo-project.org`.

13 The Center for Integrative Biomedical Computing (CIBC). Micro-CT Dataset Archive. `https://www.sci.utah.edu/cibc-software/cibc-datasets.html`.

14 R. Forman. Morse theory for cell complexes. *Advances in mathematics*, 134(1):90–145, 1998.

15 X. Ge, I. I Safa, M. Belkin, and Y. Wang. Data skeletonization via reeb graphs. In *Advances in Neural Info. Proc. Sys.*, pages 837–845, 2011.

16 A. Gyulassy, M. Duchaineau, V. Natarajan, V. Pascucci, E. Bringa, A. Higginbotham, and B. Hamann. Topologically clean distance fields. *IEEE Trans. Visualization Computer Graphics*, 13(6):1432–1439, Nov 2007.

17 T. Hastie and W. Stuetzle. Principal curves. *Journal of the American Statistical Association*, 84(406):502–516, 1989.

18 B. Kégl and A. Krzyzak. Piecewise linear skeletonization using principal curves. *IEEE Trans. Pattern Anal. Machine Intelligence*, 24(1):59–74, 2002.

19 L. Liu, E. W Chambers, D. Letscher, and T. Ju. Extended grassfire transform on medial axes of 2d shapes. *Computer-Aided Design*, 43(11):1496–1505, 2011.

20 J. Milnor. *Morse Theory.* Princeton Univ. Press, New Jersey, 1963.

21 M. Natali, S. Biasotti, G. Patanè, and B. Falcidieno. Graph-based representations of point clouds. *Graphical Models*, 73(5):151–164, 2011.

22 U. Ozertem and D. Erdogmus. Locally defined principal curves and surfaces. *Journal of Machine learning research*, 12(Apr):1249–1286, 2011.

23 V. Robins, P. J. Wood, and A. P. Sheppard. Theory and algorithms for constructing discrete morse complexes from grayscale digital images. *IEEE Trans. Pattern Anal. Machine Intelligence*, 33(8):1646–1658, Aug 2011.

24 T. Sousbie. The persistent cosmic web and its filamentary structure–i. theory and implementation. *Monthly Notices of the Royal Astronomical Society*, 414(1):350–383, 2011.

25 S. Wang, Y. Wang, and Y. Li. Efficient map reconstruction and augmentation via topological methods. In *Proc. 23rd ACM SIGSPATIAL*, page 25. ACM, 2015.

26 Y. Yan, K. Sykes, E. Chambers, D. Letscher, and T. Ju. Erosion thickness on medial axes of 3d shapes. *ACM Trans. on Graphics*, 35(4):38, 2016.