Computational Topology for Data Analysis: Notes from Book by

Tamal Krishna Dey
Department of Computer Science
Purdue University
West Lafayette, Indiana, USA 46907

Yusu Wang
Halıcıoğlu Data Science Institute
University of California, San Diego
La Jolla, California, USA 92093

Topic 12: Discrete Morse theory and persistence

Discrete Morse theory is a combinatorial version of the classical Morse theory. Invented by Forman [10], the theory combines topology with the combinatorial structure of a cell complex. Specifically, much like the fact that critical points of a smooth Morse function on a manifold determines its topological entities such as homology groups and Euler Characteristics, an analogous concept called critical simplices of a discrete Morse function also can determine similar structures for the complex it is defined on. Gradient vectors associated with smooth Morse functions give rise to integral lines and eventually the notion of stable and unstable manifolds [14]. Similarly, a discrete Morse function defines discrete gradient vectors and V-paths analogous to the integral lines. Using these V-paths, one can define the analogues of stable and unstable manifolds of the critical simplices.

It turns out that an acyclic pairing between simplices and their faces so that every simplex participates in at most one pair provides a discrete Morse function and conversely a discrete Morse function defines such a pairing. This pairing termed as a Morse matching is a main building block of the discrete Morse theory. In this chapter, we connect this matching with the pairing obtained through persistence algorithm. Specifically, we present an algorithm for computing a Morse matching and hence a discrete Morse vector field by connecting persistent pairs through V-paths. This requires an operation called critical pair cancellation which may not succeed all the time. However, for 1-complexes and simplicial 2-manifolds (pseudomanifolds), it always succeeds. Section 12.1 and 12.2 are devoted to these results.

In Section 12.4, we apply our persistence based discrete Morse vector field to reconstruct graphs from their noisy samples. Here we show that unstable manifolds of critical edges can recover a graph with guarantees from a density data that captures the hidden graph reasonably well. We provide two applications of using this graph reconstruction algorithm, one for road reconstructions from GPS trajectories and satellite images and another for neuron reconstructions from their images. Section 12.5 describes these applications.

12.1 Discrete Morse function

Following Forman [10] we define a *discrete Morse function* (henceforth called *Morse function* in this chapter) as a function $f : K \to \mathbb{R}$ on a simplicial complex K where for every p-simplex $\sigma^p \in K$ the following two conditions hold[1]:

- $\#\{\sigma^{p-1} \,|\, \sigma^{p-1} \text{ is a facet of } \sigma^p \text{ and } f(\sigma^{p-1}) \geq f(\sigma^p)\} \leq 1$
- $\#\{\sigma^{p+1} \,|\, \sigma^{p+1} \text{ is a coface of } \sigma^p \text{ and } f(\sigma^{p+1}) \leq f(\sigma^p)\} \leq 1$

The first condition says that at most one facet of a simplex σ has higher or equal function value than $f(\sigma)$ and the second condition says that at most one co-face of a simplex σ can have lower or equal function value than $f(\sigma)$. By a result of Chari [5], the two conditions imply that the two sets above are disjoint, that is, if a pair (σ^{p-1}, σ^p) satisfies the first condition, there is no pair (σ^p, σ^{p+1}) satisfying the second condition and vice versa. This means that a discrete Morse function f defines a matching:

[1] Forman formulated discrete Morse function for more general cell complexes

Definition 1 (Matching)**.** A set of ordered pairs $M = \{(\sigma, \tau)\}$ is a *matching* in K if the following conditions hold:

1. for any $(\sigma, \tau) \in M$, σ is a facet of τ (recall a facet is a face of co-dimension 1), and
2. any simplex $\sigma \in K$ can appear in at most one pair in M

Such a matching M defines two disjoint subsets $L \subseteq K, U \subseteq K$ where there is a bijection $\mu : L \to U$ such that $M = \{(\sigma, \mu(\sigma) \mid \sigma \in L\}$.

In Figure 12.1, we indicate a matching by putting an arrow from the lower dimensional simplex to the higher dimensional simplex. Observe that the source of each arrow is a facet of the target of the arrow.

Note however, the matching in K defined by a Morse function has an additional property of acyclicity which we show next. Let us define a relation $\sigma_i \prec \sigma_{i+1}$ if $\sigma_{i+1} = \mu(\sigma_i)$ or σ_{i+1} is a facet of σ_i but $\sigma_i \neq \mu(\sigma_{i+1})$.

Definition 2 (V-path and Morse matching)**.** Given a matching M in K, for $k > 0$, a *V-path* π is a sequence

$$\pi : \sigma_0 \prec \sigma_1 \prec \cdots \sigma_{i-1} \prec \sigma_i \prec \sigma_{i+1} \cdots \prec \sigma_k \tag{12.1}$$

where $\sigma_i \neq \mu(\sigma_{i-1})$ implies $\sigma_{i+1} = \mu(\sigma_i)$. In other words, a V-path is an alternating sequence of facets and cofaces thus alternating in dimensions where every consecutive pair also alternates between matched and unmatched pairs. A V-path is *cyclic* if the first simplex σ_0 is a facet of the last simplex σ_k or $\sigma_0 = \mu(\sigma_k)$ and the matching M is called *cyclic* if there is such a path in it. Otherwise, M is called acyclic. An acyclic matching in K is called a *Morse matching*.

In Figure 12.1(left), the matching indicated by the arrows is not a Morse matching whereas the matching in Figure 12.1(right) is a Morse matching. Observe that in a sequence like (12.1), the function values on facets of the matched pairs strictly decreases. This observation leads to the following fact.

Fact 1. *The matching induced by a Morse function on K is acyclic, thus is a Morse matching.*

We also have the following relation in the opposite direction.

Fact 2. *A Morse matching M in K defines a Morse function on K.*

Proof. First order those simplices which are in some pair of M. A simplex σ^{p-1} is ordered before σ^p if $(\sigma^{p-1}, \sigma^p) \in M$ and it is ordered after σ^p if it is a facet of σ^p but $(\sigma^{p-1}, \sigma^p) \notin M$. Such an ordering is possible because M is acyclic. Then, simply order the rest of the simplices not in any pair of M according to their increasing dimensions. Assign the order numbers as the function values of the simplices, which can easily be verified to satisfy the conditions (1) and (2) of a discrete Morse function on K. □

Since a given Morse matching M in K can be associated with a Morse function f on K, we call the simplices not covered by M the *critical simplices* of f. Let $c_i = c_i(M)$ denote the number of i-dimensional critical simplices. Recall that $\beta_i = \beta_i(K)$ denotes the ith Betti number, the dimension of the homology group $\mathsf{H}_i(K)$. Assume that $c_i, \beta_i = 0$ for $i > p$ where K is p-dimensional. The following result is due to Forman [10]. It is analogous to a theorem for smooth Morse function in the smooth setting.

Proposition 1. *Given a Morse function f on K with its induced Morse matching M, let c_is and β_is defined as above. We have:*

- *(weak Morse inequality)*

  (i) $c_i \geq \beta_i$ *for all* $i \geq 0$.

  (ii) $c_p - c_{p-1} + \cdots \pm c_0 = \beta_p - \beta_{p-1} + \cdots \pm \beta_0$ *where K is p-dimensional.*

- *(strong Morse inequality)*

  $$c_i - c_{i-1} + c_{i-2} - \cdots \pm c_0 \geq \beta_i - \beta_{i-1} + \beta_{i-2} \cdots \pm \beta_0 \text{ for all } i \geq 0.$$

The weak Morse inequality can be derived from the strong Morse inequality (Exercise 7)

12.1.1 Discrete Morse vector field

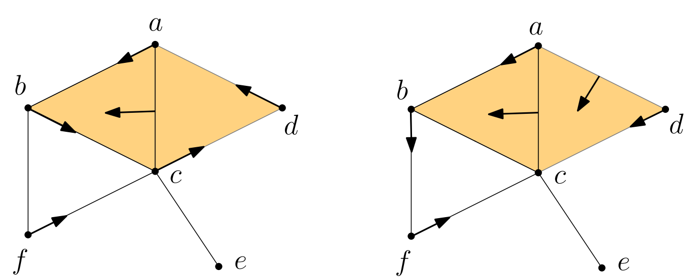


Figure 12.1: Two DMVFs: (left) the matching is not Morse because the sequence $a \prec ab \prec b \prec bc \prec c \prec cd \prec d \prec da$ is cyclic; (right) the matching is Morse, and there is no cyclic sequence.

Morse matchings can be interpreted naturally as a discrete counterpart of a vector field.

Definition 3 (DMVF)**.** A *discrete Morse vector field (DMVF) V* in a simplicial complex K is a partition $V = C \sqcup L \sqcup U$ of K where L is the set of facets paired with a unique coface in U in a Morse matching M giving $\mu(L) = U$ and C is the set of unpaired simplices called *critical simplices*. We also say that V is induced by matching M in this case.

We interpret each pair $(\sigma, \tau = \mu(\sigma))$ as a vector originating at σ and terminating at τ and draw the vector by an arrow with tail in σ and head in τ; see Figures 12.1 and 12.2. The critical simplices are treated as critical points of the vector field justifying their names. The vertex e and edge ce in both left and right pictures in Figure 12.1 are critical whereas the vertex c is critical only in the right picture and the edge bf is only critical in the left picture.

In analogy to the integral lines for smooth vector fields, we define the so called critical V-paths for discrete Morse vector fields.

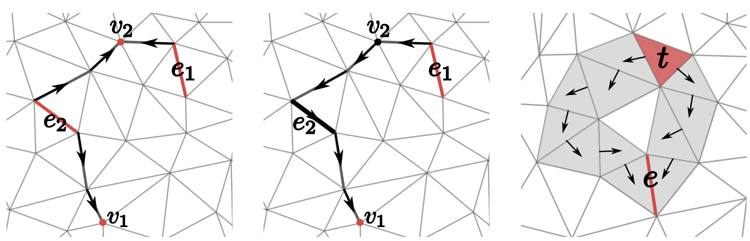
Definition 4 (Critical V-path)**.** Given a DMVF $V = C \sqcup L \sqcup U$ induced by a matching M, a V-path $\pi : \sigma_0 \prec \sigma_1 \prec \cdots \sigma_{i-1} \prec \sigma_i \prec \sigma_{i+1} \cdots \prec \sigma_k$ is critical in M if both σ_0 and σ_k are critical.

Observe that, σ_0 and σ_k in the above definition are necessarily a p- and $(p-1)$-simplex respectively if the V-path alternates between p and $(p-1)$-simplices. The V-path corresponding to a critical V-path cannot be cyclic due to this observation. The critical triangle cda with any of its edges in Figure 12.1(left) forms a V-path wheres the pair $ce \prec e$ forms a critical V-path in Figure 12.1(right).

In a critical V-path π, the pairs $(\sigma_1, \sigma_2), \cdots, (\sigma_{2i-1}, \sigma_{2i}), \cdots (\sigma_{k-2}, \sigma_{k-1})$ are matched. We can cancel the pairs of critical simplices (σ_0, σ_k) by reversing the matched pairs.

Definition 5 (Cancellation)**.** We say a pair of critical simplices (σ_0, σ_k) is *(Morse) cancellable* if there exists a *unique* critical V-path $\pi : \sigma_0 \prec \sigma_1 \prec \cdots \sigma_{i-1} \prec \sigma_i \prec \sigma_{i+1} \cdots \prec \sigma_k$. The pair (σ_0, σ_k) is cancelled if one modifies the matching by shifting the matched pairs by one position, that is, by asserting that the pairs $(\sigma_k, \sigma_{k-1}), \cdots, (\sigma_{2i+1}, \sigma_i), \cdots, (\sigma_1, \sigma_0)$ are matched instead – we refer to this as the *Morse cancellation operation* on (σ_0, σ_k). Observe that a cancellation essentially reverses the vectors in the V-path π and additionally converts critical simplices σ_0 and σ_k to be non-critical; see Figure 12.2.

Observe that a cancellation preserves the property of matching, that is, the new pairs together with the undisturbed pairs indeed form a matching. *Uniqueness* of the critical V-path connecting a pair of critical simplices ensures that the resulting new matching remains Morse. If there are more than one such critical V-path, the new matching may become cyclic – for example, in Figure 12.2(c), the cancellation of one critical V-path between the triangle-edge pair creates a cyclic V-path. The uniqueness of critical V-path is sufficient to ensure that such cyclic matching cannot be produced. In particular, we have:



(a) (b) (c)

Figure 12.2: Critical vertices and edges are marked red; (a) before cancellation of edge-vertex pair (v_2, e_2); (b) after cancellation, the path from e_2 to v_2 is inverted, giving rise to a critical V-path from e_1 to v_1, making (v_1, e_1) now potentially cancellable; (c) the edge-triangle pair (e, t), if cancelled, creates cycle as there are two V-paths between them.

Proposition 2. *Given a Morse matching M, suppose we cancel a pair of critical simplices σ and σ' in a DMVF V via a critical V-path to obtain a new matching M'. Then M' remains a Morse matching if and only if this V-path is the only critical V-path connecting σ and σ' in V (i.e, the pair (σ, σ') is cancellable as in Definition 5).*

Proof. First, assume that there are two V-paths π and π' originating at σ and ending at σ'. Since π and π' are distinct and have common simplices σ at the beginning and σ' at the end, there are simplices τ and τ' where the two paths differ for the first time after τ and join again for the first time at τ'. Reversing one V-path, say π, will create a V-path from τ' to τ. This sub-path along with the V-path from τ to τ' on π' creates a cyclic V-path, thus proving the 'only if' part.

Next, suppose that there is only a single V-path from σ to σ'. After reversing this path, we claim that no cyclic V-path is created. For contradiction, assume that a cyclic V-path is created as the result of reversal of π. Let the maximal sub-path of reveresd π on this cyclic path starts at τ and ends at τ'. We have $\tau \neq \tau'$ because otherwise the original matching needs to be cyclic in the first place. But, then the cyclic V-path has a sub-path from τ' to τ that is not in π. Since the reversed V-path π has a sub-path from τ to τ', the original path has a sub-path from τ' to τ. It means that the DMVF V originally had two V-paths from σ to σ', with one of them being π while the other one containing a sub-path not in π. This forms a contradiction that there are no distinct V-paths from σ to σ'. Hence the assumption that a cyclic V-path is created is wrong, which completes the proof of the 'if' part. □

12.2 Persistence based DMVF

Given a simplicial complex K, one can set up a trivial DMVF where every simplex is critical, that is, $V = K \sqcup \emptyset \sqcup \emptyset$. Then, one may use cancellations to build vector field further by constructing more matchings. The key to the success of this approach is to identify pairs of critical simplices that can be cancelled without creating cyclic paths. One way to do this is by taking advantage of persistence pairs among simplices.

12.2.1 Persistence-guided cancellation

First, we consider the case of simplicial 1-complexes which consist of only vertices and edges. Such a complex admits a DMVF obtained by cancelling the persistence pairs successively. Here we consider pairs with finite persistence only. Recall that some of the creator simplices are never paired with a destructor because the class created by them never dies. They are paired with ∞. Such *essential pairs* are not considered in the following proposition.

Proposition 3. *Let $(v_1, e_1), (v_2, e_2), \cdots, (v_n, e_n)$ be the sequence of all non-essential persistence pairs of vertices and edges sorted in increasing order of the appearance of the edges e_i's in a filtration of a 1-complex K. Let V_0 be the DMVF in K with all simplices being critical. Suppose DMVF V_{i-1} can be obtained by cancelling successively $(v_1, e_1), (v_2, e_2), \cdots (v_{i-1}, e_{i-1})$. Then, (v_i, e_i) can be cancelled in V_{i-1} providing a DMVF V_i for all $i \geq 1$.*

Proof. Inductively assume that (i) V_{i-1} is a DMVF obtained as claimed in the proposition and (ii) any matched edge in V_{i-1} is a paired edge in a persistence pair. We argue that these two hypotheses hold for V_i proving the claim due to the hypothesis (i).

The base case for $i = 1$ is true trivially because V_0 is a DMVF and there is no matched edge. Inductively assume that V_{i-1} satisfies the inductive hypothesis for $i > 1$. Consider the persistence pair (v_i, e_i). First, we observe that a V-path $e_i = e_{i_1} \prec v_{i_1} \prec \ldots \prec e_{i_n} \prec v_{i_n} = v_i$ exists in V_{i-1}. If not, starting from the two endpoints of e_i, we attempt to follow the two V-paths and let $v, v' \neq v_i$ be the first two critical vertices encountered during this construction. Without loss of generality, assume that v' appears before v in the filtration. Then, the 0-dimensional class

$[v + v']$ is born when v is introduced. It is destroyed by e_i. It follows that (v, e_i) is a persistence pair contradicting that actually (v_i, e_i) is a persistence pair. For induction, consider the V-path $e_i = e_{i_1} \prec v_{i_1} \prec \ldots \prec e_{i_n} \prec v_{i_n} = v_i$ in V_{i-1} which is cancelled to create V_i. For V_i not to be a DMVF, due to Proposition 2, we must have another distinct V-path from e_i to v_i in V_{i-1}, $e_i = e_{j_1} \prec v_{j_1} \prec \ldots \prec e_{j_{n'}} \prec v_{i_{n'}} = v_i$. These two non-identical paths form a 1-cycle. Every edge in this cycle except possibly e_i are matched edges in V_{i-1} and hence participates in a persistence pair by the inductive hypothesis. Then, all edges in the 1-cycle participate in some persistence pair because e_i is also such an edge by assumption. But, this is impossible because in any 1-cycle at least one edge has to remain unpaired in persistence. It follows that by cancelling (v_i, e_i), we obtain a DMVF V_i satisfying the inductive hypothesis (i). Also, inductive hypothesis (ii) follows because the new matched pairs in V_i involve edges that were already matched in V_{i-1} and the edge e_i which participates in a persistence pair by assumption. □

The result above holds for vertex-edge pairing in any simplicial complex. Furthermore, using dual graphs, it can be used for edge-triangle pairing in triangulations of 2-manifolds. Given a simplicial 2-complex K whose underlying space is a 2-manifold without boundary, consider the dual graph (1-complex) K^* where each triangle $t \in K$ becomes a vertex $t^* \in K^*$ and two vertices t_1^* and t_2^* are joined with an edge e^* if triangles t_1 and t_2 share an edge e in K.

The following result connects the persistence of a filtration of K and its dual graph K^*.

Proposition 4. *Let $\sigma_1, \sigma_2, \cdots, \sigma_n$ be a subsequence of a simplex-wise filtration $\mathcal{F}$ of K consisting of only edges and triangles. A edge-triangle pair (σ_i, σ_j) is a persistence pair for $\mathcal{F}$ if and only if (σ_j^*, σ_i^*) is a persistence pair for the filtration $\sigma_n^*, \sigma_{n-1}^*, \cdots, \sigma_1^*$ of the dual graph K^*.*

Proof. Recall the following fact. An edge-triangle persistence pair (homology groups) produced by the filtered boundary matrix D_2 for filtration of K are exactly same as the triangle-edge persistence pair (cohomology groups) obtained from the twisted (transposed and reversed) matrix D_2^*. The matrix D_2^* is exactly the filtered boundary matrix of a filtration of K^* that reverses the subsequence of triangle and edges where a triangle t becomes a vertex t^* and an edge e becomes an edge e^*. □

We can compute a DMVF V^* for K^* by cancelling all persistence pairs in non-decreasing order of their persistence. By duality, this also produces a DMVF V for the 2-manifold K. The action of cancelling a vertex-edge pair in K^* can be translated into a cancellation of an edge-triangle pair in K. Combining Propositions 3 and 4, we obtain the following result.

Theorem 5. *Let K be a simplicial 2-complex whose underlying space is a 2-manifold without boundary and $\mathcal{F}$ be a simplex-wise filtration of K. Starting from the trivial DMVF where each simplex is critical, one can obtain a DMVF in K by cancelling the vertex-edge and edge-triangle persistence pairs given by $\mathcal{F}$.*

In general, by duality one can apply the above theorem to cancel all persistence pairs between $(d-1)$-simplices and d-simplices in a filtration of a simplicial d-complex where each $(d-1)$-simplex has at most two d-simplices as cofaces. This includes simplicial d-manifolds with boundary. For this extension, one has to introduce a 'dummy' vertex in the dual graph that connects to all dual vertices of d-simplices incident to a boundary $(d-1)$-simplex. We leave it as an exercise (Exercise 8).

Unfortunately, it does not extend any further. In particular, the result in Theorem 5 does not extend to arbitrary simplicial 2-complexes and hence arbitrary simplicial complexes. The main difficulty arises because such a complex does not admit a dual graph in general. Indeed, there are counterexamples which exhibit that every persistence pair for a filtration of a simplicial 2-complex cannot be cancelled leading to a DMVF. The following Dunce hat example exhibits this obstruction.

Dunce hat. Consider a 2-manifold with boundary which is a cone with apex v and the boundary circle c. Let u be a point on c. Modify the cone by identifying the line segment uv with the circle c. Because of the similarity, the space obtained by this identification is called the Dunce hat. Consider a triangulation K of the Dunce hat. Notice that Dunce hat and hence $|K|$ is not a 2-manifold. The edges discretizing uv in K have three triangles incident to them. We show that there is no DMVF without any critical edge and triangle for K. The complex K is known to have $\beta_i(K) = 0$ for all $i > 0$ and has two or more triangles adjoining every edge in it. For any filtration of K, there cannot be any edge or triangle that remains unpaired because otherwise that would contradict that $\beta_1(K) = 0$ and $\beta_2(K) = 0$. If a DMVF V were possible to be created by cancelling persistence pairs, there would be a finite maximal V-path that cannot be extended any further. Consider such a path π starting at a simplex σ. If σ is a triangle, the edge $\mu^{-1}(\sigma)$ matched with it can be added before it to extend π. If σ is an edge, there is a triangle adjoining σ not in the V-path because at least two triangles adjoin e and the V-path starting at e cannot be cyclic. We can add that triangle to extend π. In both cases, we contradict that π is maximal.

12.2.2 Algorithms

The above results naturally suggest an algorithm for computing a persistence based DMVF for a simplicial 2-manifold K. We compute the persistence pairs on a chosen filtration $\mathcal{F}$ of K and then cancel them successively as Theorem 5 suggests. Both of these tasks can be combined by modifying the well known Kruskal's algorithm for computing minimum spanning tree of a graph.

Consider a graph $G = (U, E)$ which can be either the 1-skeleton of a complex K or the dual graph K^* if K is a simplicial 2-manifold. Assume that $u_1, u_2, \ldots, u_k$ and $e_1, e_2, \ldots, e_\ell$ be an ordered sequence of vertices and edges in G. For minimum spanning tree, the sequence of edges are taken in non-decreasing order of their weights. Here we describe the algorithm by assuming any order. Kruskal's algorithm maintains a spanning forest of the vertex set. It brings one edge e at a time in the given order either to join two trees in the current forest or to discover that the edge makes a cycle and hence does not belong to the spanning forest. If the two endpoints of e belong to two different trees in the forest, then it joins those two trees. Otherwise, e connects two vertices in the same tree creating a cycle. The main computation involves of determining if two vertices of an edge belong to the same tree or not. This can be done by union-find data structure which maintains the set of vertices of a tree in a single set and two sets are united if an edge joins the two respective trees. This is similar to FindSet and Union operations in the algorithm ZeroPerDG described previously. All such find and union operations can be done in $O(k + \ell\alpha(\ell))$ time assuming there are k vertices and ℓ edges in the graph which dominates the overall complexity.

We can incorporate the persistence computation and Morse cancellations simultaneously in the above algorithm with some simple modifications. We process the vertices and edges in their order of the input filtration. Usually, the filtration $\mathcal{F} = \mathcal{F}_f$ is given by a simplex-wise monotone function f as described previously. We compute the persistence $\mathrm{Pers}\,(e)$ of an edge e as $\mathrm{Pers}\,(e) = |f(e) - f(r)|$ if e pairs with the vertex r and ∞ otherwise.

For a vertex u in the filtration $\mathcal{F}_f$, we do not do anything other than creating a new set containing u only. When an edge $e = (u, u')$ comes in, we check if u and u' belong to the same tree by using the union-find data structure. If they do, the edge e is designated as a creator for persistence and as a critical edge in DMVF that is being built on G. Otherwise, we compute $\mathrm{Pers}\,(e)$ after finding the persistence pair for e and at the same time cancel e with its pair in the DMVF as follows. Assume inductively that the current DMVF matches every vertex other than the roots of the trees to one of its adjacent edge as follows. For a leaf vertex v, consider the path $v = v_1, e_1, \ldots, e_{k-1}, v_k = r$ from v to the root r which consist of matched pairs $(v_1, e_1), \ldots, (v_{k-1}, e_{k-1})$ and the critical vertex r. For the edge $e = (u, u')$, let the roots of the two trees containing u and u' be r and r' respectively. Assume without loss of generality that r succeeds r' in the input filtration. Then, e pairs with r in persistence because e joins the two components created by r and r' between which r comes later in the filtration. We cancel the persistence pair (r, e) by shifting the matched pairs on the path from u to r as stated in Definition 5. The root of the joined tree becomes r'. Cancelling (r, e) maintains the invariant that every path from the leaf to the root of the new tree remains a V-path. See Figure 12.3 for an illustration.

Algorithm 1 PERSDMVF$(G, \mathcal{F}_f)$

Let $G = (U, E)$ and $\mathcal{F}$ be the input filtration of its n vertices and edges.
$\mathcal{T} := \{\emptyset\}$; $V := \emptyset \sqcup \emptyset \sqcup \{(U \cup E)\}$; Initialize $\mathcal{U} := U$;
for all $i = 1, \ldots, n$ **do**
  if $\sigma_i \in \mathcal{F}_f$ is a vertex u **then**
    Create a tree T rooted at u; $\mathcal{T} := \mathcal{T} \cup \{T\}$;
  else if $\sigma_i \in F$ is an edge $e = (u, u')$ **then**
    if $t :=$FINDSET$(u)= t' :=$FINDSET(u') **then**
      designate e as creator and critical in V; $\mathrm{Pers}\,(e) := \infty$
    else
      UNION(t,t') updating $\mathcal{U}$
      Let T_u and $T_{u'}$ be trees containing u and u';
      Find V-paths π_u from u to root r and $\pi_{u'}$ from u' to r' in T_u and $T_{u'}$ respectively;
      Let r succeed r' in F; Cancel (e, r) considering the V-path π_u and update DMVF V; $\mathrm{Pers}\,(e) := |f(e) - f(r)|$;
      JOIN$(T_u, T_{u'})$ in $\mathcal{T}$;
    end if
  end if
end for
Output V and persistence pairs with persistence values

The costly step in algorithm PERSDMVF is the cancellation step which takes $O(n)$ time and thus incurs a running time $O(n^2)$ in total. However, we observe that all matchings in the final DMVF are made between a node v and the edge e that connects v to its parent parent(v) in the

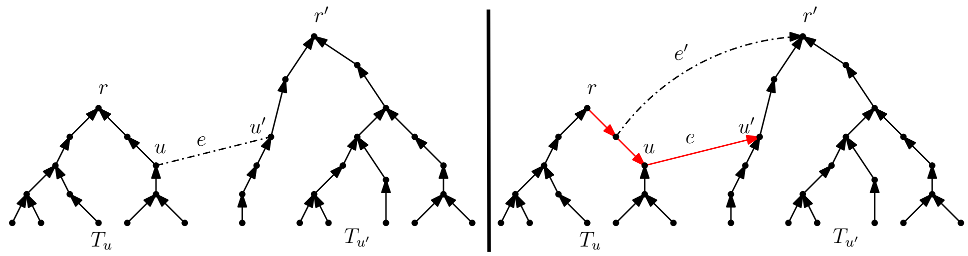


Figure 12.3: Illustration for Algorithm PersDMVF: destroyer edge $e = (u, u')$ is joining two trees T_u and $T_{u'}$ with roots r and r' respectively. The pair (r, e) is cancelled reversing the arrows on edges marked red; edge e' in the right picture is a creator and does not make any change in the forest.

Algorithm 2 SimplePersDMVF(G, $\mathcal{F}_f$)

Input:

A graph G and a filtration $\mathcal{F}_f$ on its n vertices and edges

Output:

A DMVF V and persistence pairs of $\mathcal{F}_f$ which are cancelled for creating V

Let $G = (U, E)$ and $\mathcal{F}_f$ be the input filtration of its n vertices and edges.
$\mathcal{T} := \{\emptyset\}$; $V := \emptyset$; Initialize $\mathcal{U} := U$;
for all $i = 1, \ldots, n$ **do**
    if $\sigma_i \in \mathcal{F}_f$ is a vertex u **then**
        Create a tree T rooted at u; $\mathcal{T} := \mathcal{T} \cup \{T\}$;
    else if $\sigma_i \in F$ is an edge $e = (u, u')$ **then**
        if t :=FindSet(u)= t' :=FindSet(u') **then**
            designate e as creator and critical in V; Pers $(e) := \infty$
        else
            Union(t,t') updating $\mathcal{U}$
            Let T_u and $T_{u'}$ be trees containing u and u' with roots r and r';
            Let r succeed r' in F; Pers $(e) := |f(e) - f(r)|$;
            Join(T_u, $T_{u'}$) in $\mathcal{T}$ with edge e;
        end if
    end if
end for
for each tree $T \in \mathcal{T}$ **do**
    for each node v in T **do**
        $e := (v, \text{parent}(v))$, $V := V \sqcup (v, e)$
    end for
    Put the root of T as a critical vertex in V;
end for
Output V and persistence pairs with persistence values

---

respective rooted tree and the root remains critical. All non-tree edges remain critical. Thus, we can eliminate the cancellation step in PersDMVF and after computing the final forest we can determine all matched pairs by traversing the trees upward from the leaves to the roots while matching a vertex with the edge visited next in this upward traversal. This matching takes $O(n)$ time. Accounting for the union-find operations, all other steps in PersDMVF take $O(n\alpha(n))$ time in total. The simplified algorithm SimplePersDMVF incorporates these changes. We have the following result.

Theorem 6. *Given a simplicial* 1*-complex or a simplicial* 2*-manifold K with n simplices, one can compute*

1. *a persistence based DMVF consistent with a given filtration in K in $O(n\alpha(n))$ time;*
2. *a DMVF consistent with a given PL function on K in $O(n \log n)$ time.*

Proof. We argue for all statements in the theorem when K is a 1-complex. By considering the dual graph K^*, and combining Propositions 3 and Proposition 4, the arguments also hold for K when it is a simplicial 2-manifold. The complexity analysis of the algorithm SimplePersDMVF establishes the first statement. For the second statement, given the function values at the vertices of K, we can compute a simplex-wsie lower star filtration in $O(n \log n)$ time after sorting these function values. A subsequent application of SimplePersDMVF on this lower star filtration provides us the desired DMVF. □

We can modify SimplePersDMVF slightly to take into account a threshold δ for persistence, that is, we can cancel pairs only with persistence up to δ. Interestingly, we do not need to compute all persistence pairs to determine which pairs qualify for the threshold. The new algorithm called PartialPersDMVF takes δ as input and modifies step 14 of Algorithm 1 as:

- **If** Pers $(e) \leq \delta$ **then** Join$(T_u, T_{u'})$ else designate e critical

Claim 1. PartialPersDMVF*($\mathcal{F}$,δ) computes a DMVF obtained by cancelling persistence pairs in non-decreasing order of persistence values which do not exceed the input threshold δ.*

Let V_δ denote the resulting discrete gradient field after canceling all *vertex-edge* persistence pairs with persistence at most δ.

Proposition 7. *The following statements hold for the output $\mathcal{T}$ of procedure* PartialPersDMVF *w.r.t any $\delta \geq 0$:*

(i) For each tree T_i, its root r_i is the only critical simplex in $V_\delta \cap T_i$. The collection of these roots corresponds exactly to those vertices whose persistence is bigger than δ.

(ii) Any edge with Pers $(e) > \delta$ *remains critical in V_δ and cannot be contained in $\mathcal{T}$.*

12.3 Stable and unstable manifolds

In the book, we introduce the concept of smooth Morse functions. These are smooth functions $f : \mathbb{R}^d \to \mathbb{R}$ satisfying certain conditions. We defined critical points of these functions and analyzed topological structures using the neighborhoods of these critical points. Here, we introduce another well known structure associated with Morse functions and then draw a parallel between these smooth continuous structures to their discrete counterparts with the discrete Morse functions.

12.3.1 Morse theory revisited

For a point $p \in \mathbb{R}^d$, recall that 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 ascending direction of f at p, with its magnitude being the rate of change. An *integral path* of f is a maximal path $\pi : (0, 1) \to \mathbb{R}^d$ where the tangent vector at each point p of this path equals $\nabla f(p)$, which is intuitively a flow path following the steepest ascending direction at any point. Recall that a point $p \in \mathbb{R}^d$ is critical if its gradient vector vanishes, i.e, $\nabla f(p) = [0 \cdots 0]^T$. An integral path 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 12.4 where we show the graph of a function $f : \mathbb{R}^2 \to \mathbb{R}$, and there is an integral path from a minimum v to a maximum t_2 and also to a saddle point e_2.

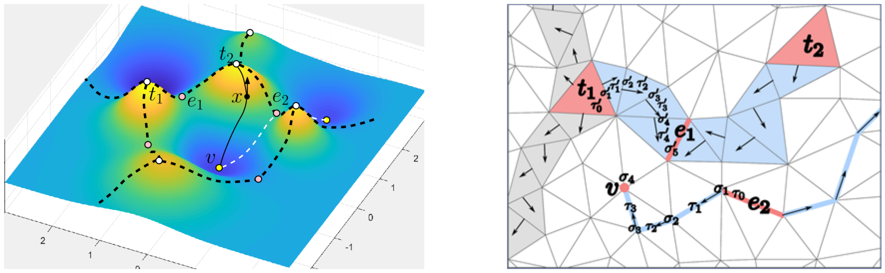
For a critical point p, the union of p and all the points from integral lines flowing into p is referred to as the *stable manifold of p*. Similarly, for a critical point q, the union of q and all the points on integral lines starting from q is called the *unstable manifold of q*. The unstable manifold of a minimum p intuitively corresponds the basin/valley around p in the terrain of f. The 1-unstable manifold of an index $(d - 1)$ saddle consist of flow paths connecting this saddle to maxima– These curves intuitively capture "mountain ridges" of the terrain (graph of the function f); see Figure 12.4 for an example. Symmetrically, the stable manifold of a maximum q corresponds to the mountain around q. The 1-stable manifolds consist of a collection of curves connecting minima to 1-saddles, corresponding intuitively to the "valley ridges".

Now, we focus on a graph-reconstruction approach using Morse-theory. Suppose that a density field $\rho : \Omega \to \mathbb{R}$ on a domain $\Omega \subseteq \mathbb{R}^d$ is given where ρ concentrates around a hidden geometric graph G embedded in $\mathbb{R}^d$. We want to reconstruct G from ρ. Intuitively, we wish to use the 1-unstable manifolds of saddles (mountain ridges) of the density field ρ to capture the hidden graph.

However, to implement this idea, we will use *discrete Morse theory*, which provides robustness and simplicity due to its combinatorial nature. We will also see that the cancellations guided by the persistence pairings could help us removing noise introduced both by discretization and measurement errors. Below, we introduce some concept necessary for transitioning to the discrete versions of (un)stable manifolds.

12.3.2 (Un)Stable manifolds in DMVF

The V-paths in a DMVF are analogues to the integral paths in the smooth setting. A *V-path* $\pi : \sigma_0 \prec \sigma_1 \prec \cdots \sigma_{i-1} \prec \sigma_i \prec \sigma_{i+1} \cdots \prec \sigma_k$ is a vertex-edge gradient path if σ_i alternate between edges and vertices. Similarly, it is a edge-triangle gradient path if they alternate between triangles and edges. Different from the smooth setting, a maximal V-path may not start or end at



(a) (b)

Figure 12.4: (Un)stable manifolds for a smooth Morse function on left and its discrete version (shown partially) on right. t_1 and t_2 are maxima (critical triangles in discrete Morse), v is a minimum, e_1 and e_2 are saddles (critical edges in discrete Morse). The unstable manifold of e_1 flows out of it to t_1 and t_2. On the other hand, its stable manifolds flows out of minima such as v and come to it. These flows work in the opposite direction of ‘gravity’ because if we put a drop of water at x it will flow to v. If we put it on the other side of the mountain ridge it will flow to other minimum. Notice this reversal of flow direction from the smooth case to the discrete case.

critical simplices. However, those that do (i.e, when σ_0 and σ_k are critical simplices) are exactly the critical V-paths. These paths are discrete analogues of maximal integral paths in the smooth setting which “start” and “end” at critical points. 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.

There is one more caveat that one should be aware of. The direction of the integral paths and the V-paths run in the opposite direction by definition: In the smooth setting, function values increase along an integral path, while in the discrete setting, it decreases along a V-path. This means that the stable and unstable manifolds reverses their roles in the two settings. 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. In the graph reconstruction approach presented below, we use “mountain ridges” for the reconstruction. We have seen that these are 1-unstable manifolds of saddles in the smooth setting and hence correspond to 1-stable manifolds in the discrete gradient fields consisting of triangle-edge paths. Notice that these mountain ridges on a triangulation of d-manifold correspond to a V-path alternating between d and $(d-1)$ dimensional simplices. Computationally, however, vertex-edge gradient paths are simpler to handle especially for the Morse cancellations below. Hence in our algorithm below, we negate the density function ρ and consider the function $-\rho$. The algorithm outputs a subset of the *1-unstable manifolds* that are vertex-edge paths in the discrete setting as the recovered hidden graph.

With the above set up, we have an input function $f : V(K) \to \mathbb{R}$ defined at the vertices $V(K)$ of a complex K whose linear extension leads to a PL function still denoted by $f : |K| \to \mathbb{R}$. For computing persistence, we use the lower-star filtration $\mathcal{F}_f$ of f and its simplex-wise version.

12.4 Graph reconstruction

Suppose we have a domain Ω (which will be a cube in $\mathbb{R}^d$) 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 $\rho : |K| \to \mathbb{R}$. The algorithm can be easily modified to take a cell complex as input. Our goal is to compute a graph $\hat{G}$ approximating the hidden graph G.

12.4.1 Algorithm

Intuitively, we wish to use "mountain ridges" of the density field to approximate the hidden graph as Figure 12.6 shows. We compute these ridges as the 1-stable manifolds ("valley ridges") of $f = -\rho$, the negation of the density function. In the discrete setting, these become 1-unstable manifolds consisting of vertex-edge gradient paths in an appropriate DMVF. We compute this DMVF by cancelling vertex-edge persistence pairs whose persistence is at most a threshold δ. The rational behind this choice is that the small undulations in a 1-unstable manifold caused by noise and discretization need to be ignored by cancellation. The procedure PartialPersDMVF described earlier in Section 12.2.2 achieves this goal. Finally, the union of the 1-unstable manifolds of all remaining high-persistence critical edges is taken as the output graph $\hat{G}$, as outlined in Procedure in CollectG.

Algorithm 3 MorseRecon(K, ρ, δ)

Input:
A 2-complex K, a vertex function ρ on K, a threshold δ
Output:
A graph

1: Let $\mathcal{F}$ be a simplex-wise lower star filtration of K w.r.t. $f = -\rho$.
2: Compute persistence Pers(e) for every edge e for the filtration $\mathcal{F}$.
3: Let K^1 be the 1-skeleton of K and $\mathcal{F}^1$ be $\mathcal{F}$ restricted to vertices and edges only
4: Let $\mathcal{T}$ be the forest computed by PartialPersDMVF(K^1,$\mathcal{F}^1$,δ)
5: CollectG(K^1,$\mathcal{T}$, Pers$(\cdot)$, δ)

Since we only need 1-unstable manifolds, K is assumed to be a 2-complex. Notice that one only needs to cancel vertex-edge pairs – this is because only vertex-edge gradient vectors contribute to the 1-unstable manifolds, and also new vertex-edge vectors can only be generated while canceling other vertex-edge pairs.

Let $T_1, T_2, \ldots, T_k$ be the set of trees returned by PartialPersDMVF. The routine CollectG outputs the 1-unstable manifold of every edge $e = (u, v)$ with Pers$(e) > \delta$, which is simply the union of e and the unique paths from u and v to root of the tree containing them respectively.

Notice that we still need to compute the persistence for all edges. If it were only for those edges that pair with vertices, we could have eliminated step 2 in MorseRecon and computed the persistence of these edges in PartialPersDMVF in almost linear time (Theorem 6). However, to compute persistence for edges that pair with triangles, we have to use the standard persistence algorithm whose complexity again depends on the complex K. For example, if K is a simplicial

Algorithm 4 COLLECTG(K^1,$\mathcal{T}$, Pers$(\cdot)$, δ)

Input:
A 1-skeleton K^1, a forest $\mathcal{T} \subseteq K^1$, persistence values for edges in K^1, a threshold δ
Output:
A graph

$\hat{G} := \emptyset$
for every edge $e = (u, v) \in K^1 \setminus \mathcal{T}$ **do**
  if Pers$(e) > \delta$ **then**
    Let $\pi(u)$ and $\pi(v)$ be the two paths from u and v to the roots respectively;
    Set $\hat{G} := \hat{G} \cup \pi(u) \cup \pi(v) \cup \{e\}$
  end if
end for
Return $\hat{G}$

2-manifold, this can run in $O(n \log n)$ time; but this time complexity does not hold for general 2-complex K. To take into account this dependability of the time complexity on the type of K, we simply denote the time for computing persistence with Pert(K) in the following theorem.

Theorem 8. *The time complexity of our Algorithm* MORSERECON *is* $O(\text{Pert}(K))$*, where* Pert(K) is the time to compute persistence pairings for K.

We remark that, for K with n vertices and edges, collecting all 1-unstable manifolds takes $O(n)$ time if one avoids revisiting edges while tracing paths. This $O(n)$ term is subsumed by Pert(K) because there are at least $n/2$ such pairs.

Notice that, Proposition 7(i) implies that for each T_i, any V-path of V_δ starting at a vertex or an edge in T_i terminates at its root r_i. See figure 12.3 for an example. Hence for any vertex $v \in T_i$, the path $\pi(v)$ computed in procedure COLLECTG is the unique V-path starting at v. This immediately leads to the following result:

Corollary 9. *For each critical edge* $e = (u, v)$ *with* Pers$(e) \geq \delta$, $\pi(u) \cup \pi(v) \cup \{e\}$ *as computed in procedure* COLLECTG *is the* 1*-unstable manifold of e in* V_δ.

12.4.2 Noise model

To establish theoretical guarantees for the graph reconstructed by Algorithm MORSERECON, we assume a noise model for the input. We first describe the noise model in the continuous setting where the domain is k-dimensional unit cube $\Omega = [0, 1]^k$. 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 $\mathrm{cl}(\overline{G^\omega}) = \mathrm{cl}(\Omega \backslash G^\omega)$. Figure 12.5 illustrates the noise model in the discrete setting, showing G (red graph) with its ω-neighborhood G^ω (yellow).

Definition 6 ((β, ν, ω)-approximation)**.** A density function $\rho : \Omega \to \mathbb{R}$ is a (β, ν, ω)*-approximation of a connected graph* G if the following holds:

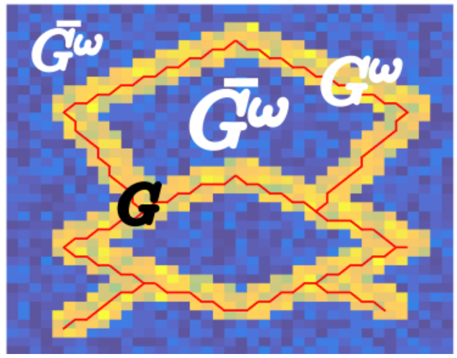


Figure 12.5: Noise model for graph reconstruction.

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; 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 $v \in V(G^\omega)$, $d(v, \mathrm{G}) \leq \omega$. The complex $\mathrm{cl}(\overline{G^\omega}) \subseteq K$ 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 (β, ν, ω)-approximation of G is extended to this setting by a PL-function $\rho : |K| \to \mathbb{R}$ while requiring that the underlying space of G^ω deformation retracts to G as in (C-1), and density conditions in (C-2) are satisfied at vertices of K.

We remark that the noise model is rather limited – In particular, it does not allow significant non-uniform density distribution. However, this is the only case that theoretical guarantees are known at the moment for a discrete Morse based reconstruction framework. In practice, the algorithm has often been applied to non-uniform density distributions.

12.4.3 Theoretical guarantees

In this subsection, we prove results that are applicable to hypercube domains of any dimensions. Recall that V_δ is the discrete gradient field after the cancellation process with threshold δ, where we perform cancellation for *vertex-edge* persistence pairs generated by a simplex-wise filtration induced by the PL-function $f = -\rho$ that negates the density PL-function. At this point, all positive edges, i.e, those not paired with vertices, remain critical in V_δ. Some negative edges, i.e, those paired with vertices also remain critical in V_δ – these are exactly the negative edges with persistence bigger than δ. CollectG only takes the 1-unstable manifolds of those critical edges (positive or negative) with persistence bigger than δ; so those edges whose persistence is at most δ are ignored.

Input assumption. Let ρ be an input density field which is a (β, ν, ω)-approximation of a connected graph G, and $\delta \in [\nu, \beta - \nu)$.

Under the above input assumption, let $\hat{G}$ be the output of algorithm MorseRecon(K, ρ, δ). The proof of the following result can be found in [8].

Proposition 10. *Under the input assumption, we have:*

(i) There is a single critical vertex left after MorseRecon *returns, which is in G^ω.*

(ii) Every critical edge considered by CollecG *forms a persistence pair with a triangle.*

(iii) Every critical edge considered by CollectG *is in G^ω.*

Theorem 11. *Under the input assumption, the output graph satisfies $\hat{G} \subseteq G^\omega$.*

Proof. Recall that the output graph $\hat{G}$ consists of the union of 1-unstable manifolds of all the edges $e_1^*, \ldots, e_g^*$ with persistence larger than δ – by Propositions 10 (ii) and (iii), they are all positive (paired with triangles), and contained inside G^ω. Below we show that other simplicies in their 1-unstable manifolds are also contained in G^ω.

Take any $i \in [1, g]$ and consider $e_i^* = (u, v)$. Without loss of generality, consider the critical V-path $\pi : e_i^* \prec (u = u_1) \prec e_1 \prec u_2 \prec \ldots \prec e_s \prec u_{s+1}$. By definition u_{s+1} is a critical vertex and is necessarily the global minimum v_0 for the density field ρ, which is also contained inside G^ω. We now argue that all simpliecs in the path π lie 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 2. *During the cancellation with threshold δ in the algorithm* MorseRecon*, no crossing gradient vector is ever produced.*

Proof. Suppose the claim is not true. Then, let (v, e) be the *first* crossing gradient vector ever produced during the 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 cancellation for the persistence pair (v', e'). Obviously, $\mathrm{Pers}\,(e') \le \delta$ because otherwise cancellation would not have been performed. 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 the V-path π' connecting e' and v' in the current discrete gradient vector field V'. The path π' begins and ends with simplices that are either both in G^ω or both are outside G^ω and also it has simplices both inside and outside G^ω. It follows that the path π' contains 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 our assumption that (v, e) is 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 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^* which starts from the other endpoint of e_i^*. Since this holds for any $i \in [1, g]$, the theorem follows. □

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

Proposition 12. *Under the input assumption, $\hat{G}$ is homotopy equivalent to G.*

Proof. First we show that $\hat{G}$ is connected. Then, we show that $\hat{G}$ has the same first Betti number as that of G which implies the claim as any two connected graphs in $\mathbb{R}^k$ with the same first Betti number are homotopy equivalent. Suppose that $\hat{G}$ has at least two components. These two components should come from two trees in the forest computed by PartialPersDMVF. The roots, say r and r', of these two trees must reside in G^ω due to Claim 2 and Proposition 10(iii). Furthermore, the supporting complex of G^ω is connected because it contains the connected graph G. It follows that there is a path connecting r and r' within G^ω. All vertices and edges in G^ω appear earlier than other vertices and edges in the filtration that PartialPersDMVF works on. This two facts mean that the first edge which connects the two trees rooted at r and r' resides in G^ω. This edge has a persistence less than δ and should be included in the reconstruction by MorseRecon. It follows that CollectG returns 1-unstable manifolds of edges ending at a common root of the tree containing both r and r'. In other words, $\hat{G}$ cannot have two components as assumed.

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, $\mathrm{Pers}\,(e) > \delta$ 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 CollectG. Our algorithm constructs $\hat{G}$ by adding these g positive edges to the spanning tree each of which adds a new cycle. Thus, $\hat{G}$ has first Betti number g as well, thus proving the proposition. □

We have already proved that $\hat{G}$ is contained in G^ω. This fact along with Proposition 12 can be used to argue that any deformation retraction taking (underlying space) G^ω to G also takes $\hat{G}$ 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 13. *Let $H : G^\omega \times [0, 1] \to G^\omega$ be any deformation retraction so that $H(G^\omega, 1) = G$. Then, the restriction $H|_{\hat{G}} : \hat{G} \times [0, 1] \to G^\omega$ is a homotopy from the embedding $\hat{G}$ to $G' \subseteq G$ where G and G' have the same first Betti number.*

Proof. The fact that $H|_{\hat{G}}(\cdot, \ell)$ is continuous for any $\ell \in [0, 1]$ is obvious from the continuity of H. Only thing that needs to be shown is that $G' := H|_{\hat{G}}(\hat{G}, 1)$ has the same first Betti number as that of

G. We observe that a cycle in $\hat{G}$ 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 cycle created by the edge e is not destroyed in G^{ω}. Therefore, a cycle basis for $\mathsf{H}_1(\hat{G})$ is also a homology basis for $\mathsf{H}_1(G^{\omega})$. Since the map $H(\cdot, 1) : G^{\omega} \to G$ is a homotopy equivalence, it induces an isomorphism in the respective homology groups; in particular, a basis in $\mathsf{H}_1(G^{\omega})$ is mapped bijectively to a basis in $\mathsf{H}_1(G)$. Therefore, the image $G' = H|_{\hat{G}}(\hat{G}, 1)$ must have a basis of cardinality $g = \beta_1(\hat{G}) = \beta_1(G^{\omega}) = \beta_1(G)$ proving that $\beta_1(G') = \beta_1(G)$. □

12.5 Applications

12.5.1 Road network

Robust and efficient automatic road network reconstruction from GPS traces and satelite images is an important task in GIS data analysis and applications. The Morse-based approach can help reconstructing the road network in both cases in a conceptually simple and clean manner. The framework provides a meaningful and robust way to remove noise because it is based on the concept of persistent homology. Intuitively, reconstruction of a road network from a noisy data is tantamount to reconstructing a graph from a noisy function on a 2D domain. One needs to eliminate noise and at the same time preserve the signal. Persistent homology and discrete Morse theory help address both of these aspects. We can simply use the graph reconstruction algorithm detailed in the previous section for this road network recovery.

GPS trajectories. Here the input is a set of GPS traces, and the goal is to reconstruct the underline road network automatically from these traces. The input set of GPS traces can be converted into a density map $\rho : \Omega \to \mathbb{R}$ defined on the planar domain $\Omega = [0, 1] \times [0, 1]$. We then use our graph reconstruction algorithm MorseRecon to recover the “mountain ridges" of the density field; see Figure 12.6.

In Figure 12.7, we show reconstructed road network after improving the discrete-Morse based output graphs with an *editing strategy* [7]. After the automatic reconstruction, the user can observe the missing branches and can recover them by artificially making a vertex near the tip of each such branch a minimum. This forces a 1-unstable manifold from a saddle edge to each of these minima. Similarly, if a distinct loop in the network is missing, the user can artificially make a triangle in the center of the loop a maximum which forces the loop to be detected.

Satellite images. In this case, we combine the Morse based graph reconstruction with a neural network framework to recover the road network from input satellite images. First, we feed the grayscale values of the input satellite image as a density function to MorseRecon. The output graphs from a set of images are used to train a convolutional neural network CNN, which output an image aiming to capture only the foreground (roads) in the satellite images. After training this CNN, we feed the original satellite images to it to obtain a set of hopefully “cleaner" images. These cleaned images are again fed to MorseRecon to output a graph which can again be used to further train the CNN. Repeated use of this reconstruct-and-train step clean the noise considerably. In Figure **??** (f) from Chapter Prelude, we show an example of the output of this strategy.

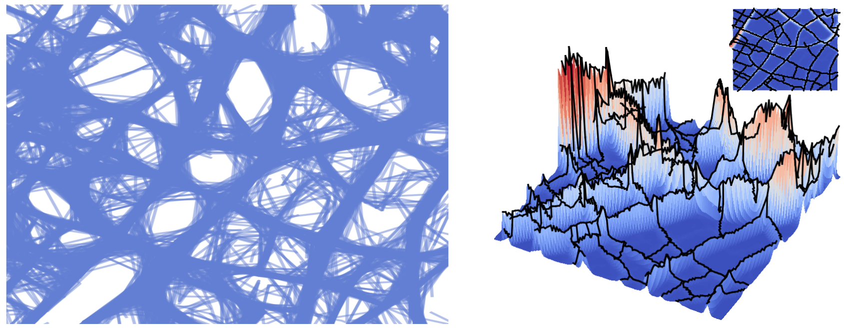


Figure 12.6: Road network reconstruction [22]: (Left) Input GPS traces. (Right) Terrain corresponding to the graph of the density function computed from input GPS traces; Black lines are the output of algorithm MorseRecon, which captures the 'mountain ridges' of the terrain, corresponding to the reconstructed road-network. The upper right is a top view of the terrain.

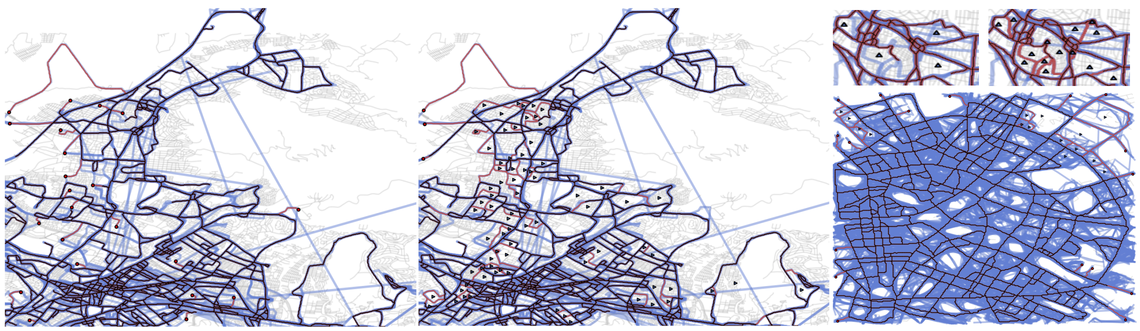


Figure 12.7: Road network reconstruction with editing [7]: (Left) Red points (minima) are added, red branches are newly reconstructed for the Athens map (black curves are original reconstruction, blue curves are input GPS traces). (Middle) We also add blue triangles as maxima to capture many missing loops. (Right) Upper: An example to show that adding extra triangles as maxima will capture more loops. Bottom: Berlin with adding both branches and loops.

Notice that this strategy eliminates the need for curating the satellite images manually for creating training samples.

12.5.2 Neuron network

To understand neuronal circuitry in the brain, a first step is often to reconstruct the neuronal cell morphology and cell connectivity from microscopic neuroanatomical image data. Earlier work often focuses on single neuron reconstruction from high resolution images of specific region in the brain. With the advancement of imaging techniques, whole braining imaging data are becoming more and more common. Robust and efficient methods that can segment and reconstruct neurons and/or connectivities from such images are highly desirable.

The discrete Morse based graph reconstruction algorithms have been applied to both fronts. Neuron cells have tree morphology and can commonly be modeled as a rooted tree, where the root of the tree locates in the soma (cell body) of the neuron. In Figure 12.8, we show the reconstructed

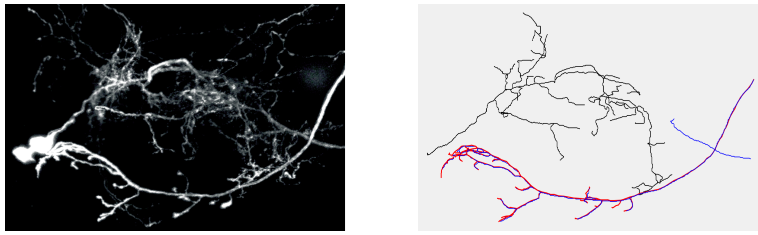


Figure 12.8: Discrete Morse based neuron morphology reconstruction from [21].

neuron morphology by applying the discrete Morse algorithm directly to an Olfactory Projection Fibers data set (specifically, OP-2 data set) from the DIADEM challenge [17]. Specifically, the input is an image stack acquired by 2-channel confocal microscopy method. In the approach proposed in [21], after some preprocessing, the discrete Morse based algorithm is applied to the 3D volumetric data to construct a graph skeleton. A tree-extraction based algorithm is then applied to extract a tree structure from the graph output.

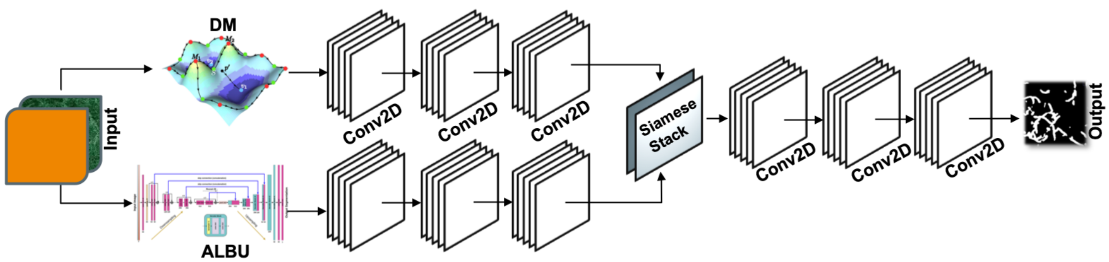


Figure 12.9: The DM++ framework proposed by [1], which combines both the DM output with standard neural-network based output together via a Siamese neural network stack so as to use these two inputs to augment each other and obtain better connected final segmentation. Image courtesy of [1].

The discrete Morse based graph reconstruction algorithm can also be used in a more sophisticated manner to handle more challenging data. Indeed, a new neural network framework is proposed in [1] to combine the reconstructed Morse graph as topological prior with a UNet [18] like neural network architecture for cell process segmentation from various neuroanatomical image data. Intuitively, while UNet has been quite successful in image segmentation, such approaches lack a global view (e.g, connectivity) of the structure behind the segmented signal. Consequentially, the output can contain broken pieces for noisy images, and features such as junction nodes in input signal can be particularly challenging to recover. On the other hand, while DM-based graph reconstruction algorithm is particularly effective in capturing global graph structures, it may produce many false positives. The framework proposed in [1], called DM++ uses output from discrete Morse as a separate channel of input, and co-train it together with the output of a specific UNet-like architecture called ALBU [4] so as to use these two input to complement each other. See Figure 12.9. In particular, UNet otuput helps to remove false positives from discrete Morse output, while the Morse graph output helps to obtain better connectivity.

12.6 Notes and Exercises

Forman [10] developed the discrete analogue of the classical Morse theory in mathematics. This analogy is exemplified by the following fact. Let C_p denote the p-th chain group formed by the p-dimensional critical cells in a discrete Morse vector field. It means that C_p is a free Abelian group with critical p-cells forming a basis assuming $\mathbb{Z}_2$-additions. For a critical cell c_p, define the boundary operator $\partial_p c_p = \Sigma_i (m_p \mod 2) c^i_{p-1}$ where c^i_p is a critical $(p-1)$-cell reachable by m_p number of V-paths from c_p. Extending the boundary operator to the chains we get the boundary homomorphism $\partial_p : \mathsf{C}_p \to \mathsf{C}_{p-1}$. One can verify that $\partial_{p-1} \circ \partial_p = 0$ (Exercise 9) thus leading to a valid *discrete Morse chain complex*. Naturally, we get a homology group H_p from this construction. It turns out that this homology group is isomorphic to the homology group of the complex on which the DMVF is defined.

Several researchers brought the concept to the area of topological data analysis [2, 12, 13, 15]. King et al. [12] presented an algorithm to produce a discrete Morse function on a complex from a given real-valued function on its vertices. Bauer et al. [2] showed that persistent pairs can be cancelled in order of their persistence values for any simplicial 2-manifolds. They also gave an $O(n \log n)$-time algorithm for cancelling pairs that have persistence below a given threshold. The cancellation algorithm and its analysis in this chapter follow this result though with a slightly different presentation. This cancellation does not generalize to simplicial 2-complexes and beyond as we have illustrated. Mischaikow and Nanda [15] proposed Morse cancellation as a tool to simplify an input complex before computing persistence pairs. The combinatorial view of the vector field given by the discrete Morse theory has recently been extended to dynamical systems, see, e.g., [3, 16].

Starting with Lewiner et al. [13], several researchers proposed discrete Morse theory for applications in visualization and image processing. Gyulassy et al. [11], Delgado-Friedricks et al. [6] and Robins et al. [20] used discrete Morse theory in conjunction with persistence based cancellations for processing images and analyzing features for e.g., porous solids. Sousbie [19] proposed using the theory for detecting filamentary structures in data for cosmic webs. These work proposed using cancellations as long as they are permitted acknowledging the fact that all cancellations in a 2- or 3-complex may not be possible. Wang et al. proposed to use discrete Morse complexes to compute unstable 1-manifolds as an output for a road network from GPS data [22]. Using unstable 1-manifolds in a discrete Morse complex defined on a triangulation in $\mathbb{R}^2$ to capture the hidden road network was proposed in this paper. Ultimately, this proposed approach was implemented with a simplified algorithm and a proof of guarantee in [8]. The material in section 12.4 is taken from this paper. The application to road network reconstruction from GPS trajectories and satellite images in section 12.5 appeared in [7] and [9] respectively. The application to neuron imaging data is taken from [1, 21].

Exercises

1. A Hasse diagram of a simplicial complex K is a directed graph that has a vertex v_σ for every simplex σ in K and a directed edge from v_σ to $v_{\sigma'}$ if and only if σ' is a codimension-1 face of σ. Let M be a matching in K. Modify the Hasse diagram by reversing every edge that is directed from v_σ to $v_{\sigma'}$ and (σ', σ) is in the matching M. Show that M induces a DMVF if

and only if the modified Hasse diagram does not have any directed cycle.

2. Let f be a Morse function defined on a simplicial complex K. We say K collapses to K' if there is a simplex σ with a single coface σ' and $K' = K \setminus \{\sigma, \sigma'\}$. Let $K_a \subseteq K$ be the subcomplex where $K_a = \{\sigma \,|\, f(\sigma) \leq a\}$. Show that there is a series of collapses (possibly empty) that brings K_a to K_b for any $b \leq a$ if there is no critical simplex with function value c where $b < c < a$.

3. Call a V-path *extendible* if it can be extended by a simplex at any of the two ends. Show examples:
   (a) A non-extendible V-path that is not critical.
   (b) Show that every non-extendible V-path in a simplicial 2-manifold without boundary must have at least one critical simplex.

4. Show that a discrete Morse function defines a Morse matching.

5. Let K be a simplicial Möbius strip with all its vertices on the boundary. Design a DMVF on K so that there is only one critical edge and only one critical vertex and no critical triangle.

6. Prove that two V-paths that meet must have a common suffix.

7. Show the following:
   (a) The strong Morse inequality implies the weak Morse inequality in Proposition 1.
   (b) A matching which is not Morse may not satisfy Morse inequalities as in Proposition 1 but always satisfies the equality $c_p - c_{p-1} + \cdots \pm c_0 = \beta_p - \beta_{p-1} + \cdots \pm \beta_0$ for a p-dimensional complex K.

8. Let K be a simplicial d-complex that has every $(d-1)$-simplex incident to at most two d-simplices. Extend Theorem 5 to prove that all persistent pairs between $(d-1)$-simplices and d-simplices arising from a filtration of K can be cancelled.

9. Prove $\partial_{p-1} \circ \partial_p = 0$ for the boundary operator defined for chain groups of critical cells as described for disrete Morse chain complex in the notes above.

Bibliography

[1] Samik Banerjee, Lucas Magee, Dingkang Wang, Xu Li, Bing xing Huo, Jaikishan Jayakumar, Katherine Matho, Meng-Kuan Lin, Keerthi Ram, Mohanasankar Sivaprakasam, Josh Huang, Yusu Wang, and Partha Mitra. Semantic segmentation of microscopic neuroanatomical data by combining topological priors with encoder-decoder deep networks. *Nature Machine Intelligence*, 2:585–594, 2020. Also available on biorxiv at 2020.02.18.955237.

[2] Ulrich Bauer, Carsten Lange, and Max Wardetzky. Optimal topological simplification of discrete functions on surfaces. *Discrete & Computational Geometry*, 47(2):347–377, 2012.

[3] Erik Boczko, William D. Kalies, and Konstantin Mischaikow. Polygonal approximation of flows. *Topology and its Applications*, 154:2501–2520, 2007.

[4] Alexander Buslaev, Selim S Seferbekov, Vladimir Iglovikov, and Alexey Shvets. Fully convolutional network for automatic road extraction from satellite imagery. In *CVPR Workshops*, pages 207–210, 2018.

[5] Manoj K. Chari. On discrete morse functions and combinatorial decompositions. *Discrete Math.*, 217(1-3):101–113, 2000.

[6] Olaf Delgado-Friedrichs, Vanessa Robins, and Adrian P. Sheppard. Skeletonization and partitioning of digital images using discrete Morse theory. *IEEE Trans. Pattern Anal. Machine Intelligence*, 37(3):654–666, 2015.

[7] Tamal K. Dey, Jiayuan Wang, and Yusu Wang. Improved road network reconstruction using discrete morse theory. In Erik G. Hoel, Shawn D. Newsam, Siva Ravada, Roberto Tamassia, and Goce Trajcevski, editors, *Proceedings of the 25th ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems, GIS 2017, Redondo Beach, CA, USA, November 7-10, 2017*, pages 58:1–58:4. ACM, 2017.

[8] Tamal K. Dey, Jiayuan Wang, and Yusu Wang. Graph reconstruction by discrete morse theory. In *Proc. Internat. Sympos. Comput. Geom.*, pages 31:1–31:15, 2018.

[9] Tamal K. Dey, Jiayuan Wang, and Yusu Wang. Road network reconstruction from satellite images with machine learning supported by topological methods. In Farnoush Banaei Kashani, Goce Trajcevski, Ralf Hartmut Güting, Lars Kulik, and Shawn D. Newsam, editors, *Proceedings of the 27th ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems, SIGSPATIAL 2019, Chicago, IL, USA, November 5-8, 2019*, pages 520–523. ACM, 2019.

[10] Robin Forman. Morse theory for cell complexes. *Adv. Math.*, 134:90–145, 1998.

[11] Attila Gyulassy, Natallia Kotava, Mark Kim, Charles Hansen, Hans Hagen, and Valerio Pascucci. Direct feature visualization using morse-smale complexes. *IEEE transactions on visualization and computer graphics (TVCG)*, 18(9):1549—1562, September 2012.

[12] Henry King, Kevin P. Knudson, and Neza Mramor. Generating discrete morse functions from point data. *Exp. Math.*, 14(4):435–444, 2005.

[13] Thomas Lewiner, Hélio Lopes, and Geovan Tavares. Applications of forman's discrete morse theory to topology visualization and mesh compression. *IEEE Trans. Vis. Comput. Graph.*, 10(5):499–508, 2004.

[14] John W. Milnor. *Topology from a differentiable viewpoint*. Virginia Univ. Press, 1965.

[15] Konstantin Mischaikow and Vidit Nanda. Morse theory for filtrations and efficient computation of persistent homology. *Discrete & Computational Geometry*, 50(2):330–353, 2013.

[16] Marian Mrozek. Conley-morse-forman theory for combinatorial multivector fields on lefschetz complexes. *Found. Comput. Math.*, 17(6):1585–1633, 2017.

[17] Diadem challenge. http://diademchallenge.org.

[18] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. U-net: Convolutional networks for biomedical image segmentation. In *International Conference on Medical image computing and computer-assisted intervention*, pages 234–241. Springer, 2015.

[19] Thierry Sousbie. The persistent cosmic web and its filamentary structure - i. theory and implementation. *Monthly Notices of the Royal Astronomical Soc.*, 414(1):350–383, 2011.

[20] Peter J. Wood Vanessa Robins and Adrian P. Sheppard. Theory and algorithms for constructing discrete Morse complexes from grayscale digital images. *IEEE Trans. Pattern Anal. Machine Intelligence*, 33(8):1646–1658, 2011.

[21] Suyi Wang, Xu Li, Partha Mitra, and Yusu Wang. Topological skeletonization and tree-summarization of neurons using discrete Morse theory, 2018. available at arXiv:1805.04997.

[22] Suyi Wang, Yusu Wang, and Yanjie Li. Efficient map reconstruction and augmentation via topological methods. In Jie Bao, Christian Sengstock, Mohammed Eunus Ali, Yan Huang, Michael Gertz, Matthias Renz, and Jagan Sankaranarayanan, editors, *Proceedings of the 23rd SIGSPATIAL International Conference on Advances in Geographic Information Systems, Bellevue, WA, USA, November 3-6, 2015*, pages 25:1–25:10. ACM, 2015.