Graph Reconstruction via Discrete Morse Theory

Yusu Wang

Computer Science and Engineering Dept
The Ohio State University

Approximation Theory and Machine Learning Conference 2018
Introduction

- Graphs naturally occur in many applications
  - Hidden space: graph-like structures
  - Simple, non-linear structure behind data

http://www2.iap.fr/users/sousbie/web/html/indexd41d.html
Overall Goal:

Using geometric and topological ideas to develop graph reconstruction algorithms for various settings with theoretical understanding / guarantees
Some Related Work

- Principal curve based approaches
Some Related Work

- Principal curve based approaches

- Reeb graph based
  - [Natali et al., Graphical Models 2011], [Ge et al. W., NIPS 2011], [Chazal et al, DCG 2015]…

This talk: an effective graph reconstruction algorithm to handle ambient noise
This Talk

Overall Goal:
Using geometric and topological ideas to develop graph reconstruction algorithms for various settings with theoretical understanding / guarantees

- Geometric graph reconstruction via discrete Morse + persistence
  - A motivating example from road-network reconstruction
  - Algorithms and theoretical understanding
  - [Wang, Li, W., SIGSPATIAL 2015], [Dey, Wang, W., SIGSPATIAL 2017, SoCG 2018]
A Motivating Application

- Automatic road network reconstruction

Input: GPS trajectories
Goal: Road network
Motivation cont

- Reconstruction from satellite images
A Motivating Application

- Automatic road network reconstruction

- Two main challenges:
  - Noisy trajectories
  - Non-homogeneous distribution

- Previous methods:
  - Local information for decision making, sensitive to noise
  - Often thresholding involved, challenging in handling non-uniform input
  - Junction nodes identification and connectivity challenging

Input: GPS trajectories  
Goal: Road network
Morse-based Reconstruction

- Persistence-guided (discrete) Morse-based reconstruction framework for road network reconstruction
  - uses global structure behind data; robust against noise, small gaps, and non-uniformity in data
  - conceptually clean, easy to implement; also extension to map integration / augmentation
  - [Wang, Li, W., SIGSPATIAL 2015]
  - [Gyulassy, PhD thesis 2008], [Robins et al. 2011], [Delgado-Friedrichs et al 2015], [Sousbie, 2015]

Input:
- Large collection of trajectories

Convert to a density field $\rho: I \rightarrow R$

Discrete Morse-based graph extraction

Persistence-based simplification

Conversion Process:

1. Input: Large collection of trajectories
2. Convert to a density field $\rho: I \rightarrow R$
3. Discrete Morse based graph extraction
4. Persistence-based simplification
Main Idea

- Assume input is a scalar (density) field
  - $f: I \to R$, where high value of $f$ indicates high signal value
- View graph of $f$ as a terrain (mountain range) on $I \times R$
  - $I = [0,1]^2 \subset R^2$ for the case of road network reconstruction
- Road $\approx$ mountain ridge
  - Captured by 1-stable manifold of $f$
Morse Theory: Smooth Case

- Let \( f : \mathbb{R}^d \to \mathbb{R} \) be a Morse function

- Gradient of \( f \) at \( x \):
  \[
  \nabla f(x) = - \left[ \frac{\partial f}{\partial x_1}, \frac{\partial f}{\partial x_2}, \ldots, \frac{\partial f}{\partial x_d} \right]^T
  \]

- Critical points of \( f \): \( \{ x \in \mathbb{R}^d \mid \nabla f(x) = 0 \} \)

- An integral line \( L : (0, 1) \to \mathbb{R}^d \):
  - a maximal path in \( \mathbb{R}^d \) whose tangent vectors agree with gradient of \( f \) at every point of the path
  - origin/destination at critical points
    - \( \text{Dest}(L) = \lim_{p \to 1} L(p) \)
    - \( \text{Ori}(L) = \lim_{p \to 0} L(p) \)

- 1-stable manifolds
  - Integral lines ending at \((d - 1)\)-saddles
1-stable Manifold

1-stable manifold (of index $d - 1$ saddle points) $\approx$ mountain ridges
Discrete Case

- Smooth case
  - 1-stable manifold from Morse theory

- Discrete case
  - Piecewise-linear (PL) approximation?
Discrete Morse Theory

- [Forman 1998, 2002]
- Combinatorial version of Morse theory
- Many results analogous to classical Morse theory
- Works for cell complexes

- Combinatorial, thus numerically stable
- Algorithmically often easy to handle, especially simplification
Discrete Gradient Vector Field

- Given a simplicial complex $K$, a discrete (gradient) vector $(\sigma, \tau)$ s.t. $\sigma < \tau$ (vertex-edge or edge-triangle pair in our case)
- A Morse pairing $M(K)$ of $K$
  - A set of discrete vectors s.t. each simplex appears in at most one vector
- A simplex $\sigma$ is critical, if
  - it does not appear in any pair in $M(K)$
- A V-path in $M(K)$
  - $\tau_0, \sigma_1, \tau_1, \sigma_2, \tau_2, \ldots, \tau_k, \sigma_{k+1}$ s.t. $(\sigma_i, \tau_i) \in M(K)$
  - cyclic: if $k > 0$, and $(\sigma_{k+1}, \tau_0) \in M(K)$
  - acyclic (gradient path) otherwise
Discrete Gradient Vector Field

- Given a simplicial complex $K$, a discrete (gradient) vector $(\sigma, \tau)$ s.t. $\sigma < \tau$ (vertex-edge or edge-triangle pair in our case)
- A Morse pairing $M(K)$ of $K$
  - A set of discrete vectors s.t. each simplex appears in at most one vector
- A simplex $\sigma$ is critical, if
  - it does not appear in any pair in $M(K)$
- A V-path in $M(K)$
  - $\tau_0, \sigma_1, \tau_1, \sigma_2, \tau_2, \ldots, \tau_k, \sigma_{k+1}$ s.t. $(\sigma_i, \tau_i) \in M(K)$
  - cyclic: if $k > 0$, and $(\sigma_{k+1}, \tau_0) \in M(K)$
  - acyclic (gradient path) otherwise
- $M(K)$: discrete gradient vector field
  - if there is no cyclic V-path in $M(K)$
Discrete Gradient Vector Field

- Discrete Morse function $\leftrightarrow$ discrete gradient vector field

- A discrete Gradient Vector field $\approx$ gradient field for Morse functions
  - critical $k$-simplex $\approx$ index-$k$ critical point
  - critical edge $\approx$ saddles for function on $R^2$
  - 1-stable manifolds: edge-triangle V-paths
  - 1-unstable manifolds: vertex-edge V-paths (``valley ridges'')
Simplification via Morse Cancellation

- Morse cancellation operation (to simplify the vector field):
  - A pair of critical simplices $\langle \sigma, \tau \rangle$ can be cancelled
    - if there is a unique gradient path between them
    - By reverting that gradient path

- Reversing $\langle v_2, e_2 \rangle$ cancels the pair
- $\langle s, t \rangle$ is not cancellable
Morse cancellation of critical pairs simplify the discrete gradient vector fields
- which further simplifies 1-(un)stable manifolds

But which critical pairs should we cancel?
- intuitively: should respect input function! Less important ones corresponding to noise

Persistence homology induced by the density function to guide the cancellation of critical pairs
- “persistence” capturing “importance” of critical pairs
- [Edelsbrunner, Letscher, Zomorodian 2002], [Zomorodian, Carlsson 2005], …
Sublevel-set Persistence – Simplified view

- Input: \( f : \mathbb{R} \rightarrow \mathbb{R} \)
Sublevel-set Persistence – Simplified view

- Input: $f : R \rightarrow R$

$$H_*(f^{-1}(a))$$
Sublevel-set Persistence – Simplified view

- Input: $f : R \rightarrow R$
Sublevel-set Persistence – Simplified view

- Input: $f : \mathbb{R} \rightarrow \mathbb{R}$
Sublevel-set Persistence – Simplified view

- **Input:** $f : \mathbb{R} \rightarrow \mathbb{R}$
Sublevel-set Persistence – Simplified view

- **Input:** $f: \mathbb{R} \rightarrow \mathbb{R}$

- **Induced persistence pairings** $P(f)$
  - $\langle x_3, x_4 \rangle, \text{pers} = f(x_4) - f(x_3)$
Sublevel-set Persistence – Simplified view

- **Input:** $f : \mathbb{R} \rightarrow \mathbb{R}$

- **Induced persistence pairings** $P(f)$
  - $\langle x_3,x_4 \rangle, \text{pers} = f(x_4) - f(x_3)$
  - $\langle x_2,x_5 \rangle$
Sublevel-set Persistence – Simplified view

- **Input**: $f : \mathbb{R} \rightarrow \mathbb{R}$

- **Induced persistence pairings**
  
  $P(f)$

  - $\langle x_3, x_4 \rangle, \text{pers} = f(x_4) - f(x_3)$
  - $\langle x_2, x_5 \rangle, \langle x_1, x_6 \rangle, \ldots$
Discrete Case

- A piecewise-linear (PL) function $\rho: |K| \rightarrow R$ defined on a simplicial complex domain $K$

- Persistence algorithm via lower-star filtration
  - [Edelsbrunner, Letscher, Zomorodian 2002],
  - A collection of persistence pairings:
    - $P_{\rho}(K) = \{ (\sigma, \tau) \}$, where $k = \dim(\sigma) = \dim(\tau) - 1$
      - $\sigma$: creator, creating $k$-th homological features
      - $\tau$: destroyer, killing feature created at $\sigma$
    - $\text{per}(\sigma, \tau) = \rho(\tau) - \rho(\sigma)$: life time of this feature

Intuitively, pairs of simplices with positive persistence corresponding to persistence pairing of critical points in the smooth case.
Main Algorithm

- **Input:**
  - Triangulation $K$ of domain $I \subset \mathbb{R}^d$, function $f: K \to \mathbb{R}$, threshold $\delta$

- Initialize discrete gradient vector field $W$ on $K$ to be the **trivial one**

- **Step 1:** *persistence computation*
  - Compute persistence pairings $P(K)$ induced by function $-f$

- **Step 2:** *Morse simplification*
  - Simplify $W$ by performing Morse cancellation for all critical pairs from $P(K)$ with persistence $\leq \delta$, if possible

- **Step 3:** *collect output*
  - For all remaining critical edges with persistence $> \delta$

The algorithm works for any $d$-dimensional domain $I \subset \mathbb{R}^d$ but only 2-skeleton of the triangulation $K$ is needed
Results – Road network reconstruction

Athens

Beijing

Berlin
Effect of Simplification

Berlin, 27189 trajectories

(a) Persistent 0.0001  (b) Persistent 0.001  (c) Persistent 0.01  (d) Persistent 0.1
Thresholding?

Increasing threshold

(a) Persistent 0.0001  (b) Persistent 0.001  (c) Persistent 0.01  (d) Persistent 0.1
Comparison

(a) Karagiorgou (2013)  
(b) Our result
Map Integration

(a) Our reconstruction

(b) Karagiorgou 2013

(c) Our integration
Map Augmentation
Reconstruction from Satellite Images

- CNN + reconstruction framework
Reconstruction from Satellite Images

- CNN + reconstruction framework
Results – Neuron Reconstruction

- Single neuron reconstruction

DIADAM dataset OP 2
Results – Neuron reconstruction

- Mouse brain LM images from an AAV viral tracer-injection
  - from Mitra laboratory at CSHL
Great!

But what can we guarantee?
Provide theoretical justification / understanding for the persistence-guided discrete Morse-based graph reconstruction framework

- Further simplification of the algorithm/editing strategy
- Reconstruction guarantees under a (simple) noise model

- [Dey, Wang, W, ACM SIGSPATIAL 2017], [Dey, Wang, W., SoCG 2018]
Reconstruction Editing

- Simple strategies to allow adding missing parts
  - Enforce minima (vertices): allow adding missing free branches
  - Enforce maxima (triangles): allow adding missing loops

![Adding missing branches](image1)

![Adding missing loops](image2)
Main Algorithm

- **Input:**
  - Triangulation $K$ of domain $I \subset \mathbb{R}^d$, function $f : K \rightarrow \mathbb{R}$, threshold $\delta$

- Initialize discrete gradient vector field $W$ on $K$

- **Step 1:** *persistence computation*
  - Compute persistence pairings $P(K)$ induced by function $-f$

- **Step 2:** *Morse simplification*
  - Simplify $W$ by performing Morse cancellation for all critical pairs from $P(K)$ with persistence $\leq \delta$, if possible

- **Step 3:** *collect output*
  - For all remaining critical edges with persistence $> \delta$
  - collect their 1-unstable manifolds and output
Simplified Algorithms

- **Step 2 (Morse simplification) is replaced by**

  Procedure $\text{PerSimpTree}(P(K), \delta)$

  1. $\Pi :=$ the set of vertex-edge persistence pairs from $P = P(K)$
  2. Set $\Pi_{\leq \delta} \subseteq \Pi$ to be $\Pi_{\leq \delta} = \{ (v, e) \in \Pi \mid \text{pers}(v, e) \leq \delta \}$
  3. $\mathcal{T} := \bigcup_{(v, \sigma) \in \Pi_{\leq \delta}} \{ \sigma = \langle u_1, u_2 \rangle, u_1, u_2 \}$
  4. return $(\mathcal{T})$

- No need to cancel edge-triangle critical pair
- No need to check whether cancellation is valid or not
- No explicit cancellation operation is needed!
  - simply collect all ``negative” edges whose persistence is at most $\delta$

Simplified Step 2:
Linear time to collect a set of edges, and they form a spanning forest that contain all necessary information of discrete gradient field
Simplified Algorithm – cont.

- Step 3 (collecting output) is replaced by:

  ```
  Procedure Treebased-OutputG(T)
  1. for each critical edge e = ⟨u, v⟩ with pers(e) ≥ δ do
  2.   Let π(u) be the unique path from u to the sink of the tree Tᵢ containing u
  3.   Define π(v) similarly; Set \( \hat{G} = \hat{G} ∪ π(u) ∪ π(v) ∪ \{e\} \)
  ```

- No explicit discrete gradient vector field maintained!
- Simplified algorithm even easier and faster
  - [Attali et al 2009], [Bauer et al 2012]

- Theorem
  
  Time complexity of the simplified algorithms is \( O(n + Time(Per)) \) where 
  \( n \) is the total number of vertices and edges in \( K \).
  
  This holds for any dimensions.
Provide theoretical understanding / justification for the persistence-guided discrete Morse-based graph reconstruction framework

- Further simplification of the algorithm/editing strategy
- Reconstruction guarantees under a (simple) noise model

- [Dey, Wang, W., ACM SIGSPATIAL 2017], [Dey, Wang, W., 2018]
Noise Model

- True graph $G \subset \Omega := [0, 1]^d$
- $G^\omega \subset \Omega$: an $\omega$-neighborhood of $G$
  - such that for (i) any $x \in G^\omega$, $d(x, G) \leq \omega$; and (ii) $G^\omega$ deformation retracts to $G$

- A function $\rho: \Omega \rightarrow R$ is $(\beta, \mu, \omega)$-approximation of $G$
  - if there exists an $\omega$-neighborhood $G^\omega$ of $G$ so that
    - $\rho(x) \in [\beta, \beta + \mu]$, for any $x \in G^\omega$
    - $\rho(x) \in [0, \mu]$, otherwise
    - $\beta > 2\mu$

\[\begin{array}{c}
\beta \\
\nu \\
f_G \\
G^\omega \\
\Omega \\
\end{array}\] 

\[\begin{array}{c}
\beta + \nu \\
\nu \\
f_G + g \\
G^\omega \\
\Omega \\
\end{array}\]
Noise Model

- True graph $G \subset \Omega := [0, 1]^d$
- $G^\omega \subset \Omega$: an $\omega$-neighborhood of $G$
  - such that for (i) any $x \in G^\omega$, $d(x, G) \leq \omega$; and (ii) $G^\omega$ deformation retracts to $G$

- A function $\rho: \Omega \to R$ is $(\beta, \mu, \omega)$-approximation of $G$
  - if there exists an $\omega$-neighborhood $G^\omega$ of $G$ so that
    - $\rho(x) \in [\beta, \beta + \mu]$, for any $x \in G^\omega$
    - $\rho(x) \in [0, \mu]$, otherwise
    - $\beta > 2\mu$

- In discrete case,
  - $K$ a triangulation of $\Omega$, $G^\omega \subset K$, $\rho$ defined at vertices of $K$
Main Results

Theorem (Geometry)
For any dimension $d$, under our noise model and for appropriate $\delta$, the output graph $\hat{G}$ satisfies $\hat{G} \subset G^\omega$.

Theorem (Topology)
For any dimension $d$, under our noise model and for appropriate $\delta$, the output graph $\hat{G}$ is homotopy equivalent to $G$.

Theorem (Topology in 2D)
For $d = 2$, under our noise model and for appropriate $\delta$, there is a deformation retraction from $G^\omega$ to $\hat{G}$.
Proof Ideas

- Suppose true graph $G$ has $g$ independent loops
- Lemma A:
  - Under the noise model, after $\delta$-simplification for appropriate $\delta$, exactly 1 critical vertex (global minimum), $g$ critical edges and $g$ critical triangles are left.
Proof Ideas

- Suppose the true graph $G$ has $g$ independent loops

- Lemma A:
  - Under the noise model, after $\delta$-simplification for appropriate $\delta$, exactly 1 critical vertex (global minimum), $g$ critical edges and $g$ critical triangles are left.

- Lemma B:
  - All critical edges are in the region $G^\omega$,
  - and all critical triangles are outside it.
Each critical triangle $t$
  
  - corresponds to a region spanned by triangles reachable from $t$ via discrete gradient paths

Simplification process
  
  - merges such regions
Each critical triangle $t$
- corresponds to a region spanned by triangles reachable from $t$ via discrete gradient paths

Simplification process
- merges such regions

Lemma C:
- In $\mathbb{R}^2$, at the end of simplification, the boundary of the $g$ regions corresponding to the remaining critical triangles form a subset of output graph $\hat{G}$.
- The associated edge-triangle discrete gradient vectors inside each region lead to a deformation retraction from $G^\omega$ to $\hat{G}$. 
Remarks

- Noise model simple
  - Thresholding-based approach may potentially work for this model
  - However, not for real data

Increasing threshold
Remarks

- Noise model simple
  - Thresholding-based approach may potentially work for this model
  - However, not for real data
Concluding Remarks

- Explored the power of a discrete Morse+persistence based framework for graph reconstruction
  - Application to both 2D (road network) and 3D (neuron reconstruction)
- Provided theoretical understanding and justification of its reconstruction ability

- Only a first step!
  - More general noise models
  - High dimensional points data input
THANK YOU!