Improved Road Network Reconstruction using Discrete Morse Theory

Tamal K. Dey, Jiayuan Wang, Yusu Wang*
Department of Computer Science and Engineering, The Ohio State University
Columbus, Ohio 43210

ABSTRACT

With the rapid growth of publicly available GPS traces, robust and efficient automatic road network reconstruction has become a crucial task in GIS data analysis and applications. In [20], an effective and robust road network reconstruction algorithm was developed based on the discrete Morse theory, which has the state-of-the-art performance in automatic road-network reconstruction. Based on a discrete Morse-based graph reconstruction framework, we provide two improvements of the previous algorithm [20]: (1) we further simplify it and obtain a better empirical time performance; and (2) we develop a simple but effective editing strategy that helps adding missing road segments in the output reconstruction.

CCS CONCEPTS

• **Information systems → Geographic information systems**; • **Theory of computation → Theory and algorithms for application domains**;

KEYWORDS

Map generation, GPS traces, Topological method, Morse theory

ACM Reference format:
Tamal K. Dey, Jiayuan Wang, Yusu Wang. 2017. Improved Road Network Reconstruction using Discrete Morse Theory. In *Proceedings of SIGSPATIAL'17, Los Angeles Area, CA, USA, November 7–10, 2017,* 4 pages.
https://doi.org/10.1145/3139958.3140031

1 INTRODUCTION

Robust and efficient automatic road network reconstruction from GPS traces has become a crucial task in GIS data analysis and applications. Indeed, a range of reconstruction algorithms have been proposed and developed in the past few years. In particular, an algorithm based on the discrete Morse theory proposed in [20] turns out to be quite effective for robust road network reconstruction. This Morse-based approach can reconstruct the road network (roads and their connections) in a conceptually simple and clean manner. The framework also provides a meaningful and systematic way to remove noise based on the concept of persistent homology, which is one of the most important developments in the field of topological data analysis in the past two decades [11]. This approach provides the state-of-the-art performance in automatic road-network reconstruction, and is particularly effective for handling noisy and non-uniformly distributed trajectories.

Our contribution. 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. Persistence homology and discrete Morse theory are used for these two purposes respectively in the algorithm of [20]. To understand this algorithm, one needs to understand the persistence based discrete Morse cancellation, specifically in the context of graph reconstruction.

Based on the persistence-guided discrete Morse complex simplification in the literature, especially the work on Morse cancellation by Bauer et al. [3] and the view of region merging from [9], we provide further understanding of the approach of [20]. These insights allow us to eliminate two steps, which simplify such a discrete Morse-based graph reconstruction framework. The insights further let us develop a simple yet effective editing strategy to adding missing road pieces in the output reconstruction.

Related work. In recent years, reconstructing road networks from a collection of trajectories has generated significant interests. A large number of automatic reconstruction algorithms have been proposed [2, 4–7, 10, 16, 18, 20] including the surveys [1] [4].

The idea of using persistence-guided discrete Morse simplification for skeleton recovery is not new, see e.g, several works in [3, 8, 14, 15, 17]. The most relevant work are perhaps [3, 8], where the relation between persistent homology for sub-level set filtration and simplification of discrete Morse functions has been explored. Specifically, in [8], Delgado-Friedrichs et al. clarify the connection between persistence-pairing and the simplification of *discrete Morse chain complex* for 2D and 3D domains, which is closely related, but different from the cancellation in the discrete gradient vector field. In particular, a cancellable pair of critical cells in the discrete Morse chain complex may not be cancellable in the discrete gradient vector field in 3D, although the inverse is always true. For 2D domains, Bauer et al. [3] developed a theory for Morse cancellation which they use for optimal simplification of a function, which uses a persistence guided discrete Morse simplification for combinatorial surfaces. The relation of persistence and discrete Morse functions is also studied and leveraged (together with edge contraction) to build a multi-scale model for discrete Morse functions (gradient vector fields) [15].

*tamaldey,yusu@cse.ohio-state.edu,wang.6195@osu.edu

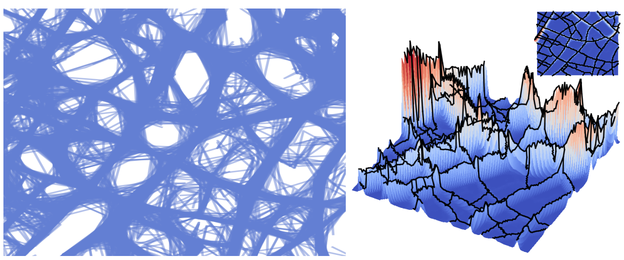
SIGSPATIAL'17, November 7–10, 2017, Los Angeles Area, CA, USA

ACM ISBN 978-1-4503-5490-5/17/11.
https://doi.org/10.1145/3139958.3140031

2 PROBLEM SETUP AND NOTATIONS

2.1 Algorithm of [20]

Given a set of noisy GPS traces, the goal is to have an algorithm that can reconstruct the hidden road network (which can be modeled as a planar geometric graph) automatically. In [20], Wang *et al* propose the following approach for this *automatic road network reconstruction problem*: See Figure 1. First, they convert the input set of GPS traces into a density map $\rho : \Omega \to \mathbb{R}$ defined on the planar domain $\Omega = [0,1] \times [0,1]$. Viewing the graph of this density function as a terrain, Wang *et al* argue that the "mountain ridges" of this terrain tend to correspond to road segments. These are the flow lines following the steepest descending direction and connecting saddles and maximas. They propose to recover the mountain ridges (and thus the road network) via the use of 1-unstable manifolds in Morse theory. To obtain a robust implementation, they use the discrete Morse theory to compute the 1-unstable manifolds, as well as to simplify the output (removing noise).



(a) (b)

Figure 1: Pipeline of algorithm of [20]. (a) Input GPS traces. (b) Terrain corresponding to the graph of the density function computed from input GPS traces. Black lines are the output of algorithm of [20], which captures the 'mountain ridges' of the terrain, corresponding to the reconstructed road-network. The upper right is a top view of the terrain.

2.2 Discrete Morse theory

To realize the idea described above, the algorithm of [20] uses the discrete Morse theory, originally introduced by Forman [13].

A k-simplex $\tau = \{p_0, \ldots, p_k\}$ is the convex hull of $k+1$ affinely independent points. Now suppose we are given a simplicial complex $\mathcal{K}$, which is a collection of simplicies and all their faces so that two simplices may intersect only at a common face. A *discrete (gradient) vector* is a pair of simplices (σ, τ) such that σ is a facet of τ. A *Morse pairing* in $\mathcal{K}$ is a collection of discrete vectors $M(\mathcal{K}) = \{(\sigma, \tau)\}$ where each simplex appears in *at most* one pair; simplices that are not in any pair are called *critical.*

Given a Morse pairing $M(\mathcal{K})$, a *V-path* is a sequence

$$\tau_0, \sigma_1, \tau_1, \ldots, \sigma_\ell, \tau_\ell, \sigma_{\ell+1},$$

where $(\sigma_i, \tau_i) \in M(\mathcal{K})$ 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 $(\sigma_{\ell+1}, \tau_0) \in M(\mathcal{K})$; 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*$(\sigma_i) = 0$. Similarly, it is an edge-triangle gradient path if *dimension*$(\sigma_i) = 1$. We call $M(\mathcal{K})$ a *discrete gradient vector field* or *gradient Morse pairing* if there is no cyclic V-path induced by $M(\mathcal{K})$.

For a critical edge e, its *stable manifold* is 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.

Morse cancellation / simplification. A pair of critical simplices $\langle \sigma, \tau \rangle$ 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 $M(\mathcal{K})$ 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 $i = 1, \ldots, \ell+1$. 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* .

2.3 Persistent pairing

Given a simplicial complex $\mathcal{K}$, let S denote an ordered sequence $\sigma_1, \ldots, \sigma_n$ of all n simplices in $\mathcal{K}$ such that for any simplex $\sigma_i \in \mathcal{K}$, its faces must appear before it in this sequence. Then S induces a *(simplex-wise) filtration*

$$F(\mathcal{K}) : \emptyset = \mathcal{K}_0 \subset \mathcal{K}_1 \subset \mathcal{K}_2 \subset \cdots \subset \mathcal{K}_N = \mathcal{K},$$

where $\mathcal{K}_i = \bigcup_{j \le i} \sigma_j$ is the subcomplex formed by the prefix $\sigma_1, \ldots, \sigma_i$ of S. Given such a filtration, the *birth* and *death* of homological features are tracked and encoded by the persistent homology [12], which can be uniquely represented by a collection of pairs of simplices $P(\mathcal{K}) = \{(\sigma_i, \sigma_j)\}$ called the *persistence-pairing* $P(\mathcal{K})$. Each *persistence pair* $(\sigma_i, \sigma_j) \in P(\mathcal{K})$ indicates that a new k-homological feature with $k =$ *dimension*(σ_i) is created at $\mathcal{K}_i$ and destroyed at $\mathcal{K}_j$. If a simplex σ creates a homological feature which is never destroyed, we introduce a pair $(\sigma, \infty) \in P(\mathcal{K})$.

3 SIMPLIFICATION AND EDITING STRATEGIES

3.1 Simplification of algorithm in [20]

First, we present our simplified version of algorithm of [20] in Algorithm 1. We will describe shortly the main differences (simplifications). The input to the algorithm is a triangulation $\mathcal{K}$ of the domain Ω, the density function $\rho : V \to \mathbb{R}$ defined at vertices V of $\mathcal{K}$, and a parameter δ for simplification (noise-removal). The algorithm outputs a graph G that aims to recover the true road network encoded by the density map ρ. Setting the function value of a simplex σ to be the highest f-value of its vertices, with $f = -\rho$, the *persistence* of a persistence-pair (σ, τ) is defined to be $\mathrm{pers}(\sigma, \tau) = |f(\tau) - f(\sigma)|$. If $\tau = \infty$, then the persistence is ∞.

Our Algorithm SimpMorseRecon() follows the same high-level structure as the algorithm of [20], but has two main simplifications:

(S1) In Line 6, originally, the algorithm of [20] needs to first check whether a persistence pair $(\sigma, \tau) \in P(\mathcal{K})$ is cancellable or not (which requires one to maintain a data structure that enables checking if there is a unique gradient path between σ and τ). The Morse cancellation operation is carried out only when the pair is cancellable. However, it turns out that once the persistence pairs are computed and ordered, each

Algorithm 1: SimpMorseRecon($\mathcal{K}$,ρ, δ)

Data: Triangulation $\mathcal{K}$ of Ω, density function $\rho : \mathcal{K} \to \mathbb{R}$, threshold δ
Result: Graph G
1 **begin**
2 Compute persistence pairings $P(\mathcal{K})$ induced by the sub-level set filtration of K by $-\rho$
3 Order persistence pairs in $P(\mathcal{K})$ in increasing order of their persistences
4 Set initial discrete gradient field M on K to be trivial
5 **for** *each vertex-edge pair* $(\sigma, \tau) \in P(\mathcal{K})$ *with* $\mathrm{pers}(\sigma, \tau) \leq \delta$ **do**
6 Perform Morse cancellation of (σ, τ) and update the discrete gradient vector field M
7 $G = \emptyset$
8 **for** *each remaining critical edge* e *with persistence larger than* δ **do**
9 $G = G \bigcup \{$1-unstable manifold of $e\}$
10 Return G

such persistence pair remains cancellable and thus the check for the same becomes redundant.

The second one is more significant:

(S2) There are two types of pairs in the collection of persistence-pairing $P(\mathcal{K})$ induced by the sub-level set filtration of $\mathcal{K}$ by $-\rho$: vertex-edge pairs and edge-triangle pairs. We claim that it is not necessary to perform edge-triangle pair cancellations, and the output remains the same. Hence in Line 5 of the algorithm, we only consider vertex-edge pairs, while the original algorithm of [20] processes both types of pairs.

The validity of these two simplifications is supported by the lemma below, which can be derived from results in [3]:

Lemma 3.1. *Let $\mathcal{K}$ be the triangulation of an orientable 2-manifold. At the beginning of each* **for**-*loop in lines 5-6 of Algorithm* SimpMorseRecon()*, the simplices σ and τ remain critical in the current discrete gradient vector field and the pair (σ, τ) is cancellable.*

If the **for**-*loop were executed also for edge-triangle pairs along with the vertex-edge pairs, the critical simplices in the discrete gradient vector field M would have been exactly those simplices of $\mathcal{K}$ w.r.t. $f = -\rho$ with persistence larger than δ.*

In a persistence-guided discrete Morse simplification, a priori, it is possible that many persistent pairs from $P(\mathcal{K})$ cannot be canceled, and/or high-persistent critical edges (correspond to important saddles in the smooth case) may no longer remain critical after the modification of the discrete gradient vector field M during the Morse cancellation process . Lemma 3.1 suggests that this does not happen for the special case when the input domain $\mathcal{K}$ is a triangulation of a 2-manifold such as the plane. Simplification (S1) is thus valid as the pair (τ, σ) is always cancellable according Lemma 3.1.

To see why Simplification (S2) is valid, first, by Lemma 3.1, observe that if we would have canceled pairs $(\sigma_1, \tau_1), \ldots, (\sigma_m, \tau_m)$ in order (with m being the largest index such that $\mathrm{pers}(\sigma_m, \tau_m) \leq \delta$), simplices with higher persistence would have remained critical in the discrete gradient vector fields $M = M_m$. Hence the set of critical edges that algorithm SimpMorseRecon() considers at lines 8–9 can be retrieved directly once persistence pairings are computed in line 2 **without** performing any Morse cancellation at all.

However, we still need to argue that that the discrete gradient vector field obtained by cancelling only vertex-edge pairs as in the for loop offers the same 1-unstable manifolds as it would have with cancelling edge-triangle pairs as well. For 1-unstable manifolds, we only need to maintain the vertex-edge type of gradient vectors in M. As the initial gradient vector field is trivial (i.e, all simplices are critical), it is easy to verify that any such vertex-edge gradient vector can only be obtained by inverting an vertex-edge gradient path during the cancellation of a vertex-edge persistence pair. This justifies Simplification (S2).

In general, the vertex-edge type gradient vectors are much easier to maintain than the edge-triangle type (as they form spanning forests in the 1-skeleton of the input complex $\mathcal{K}$). Hence these two observations above significantly simplify the algorithm as well as clarify the essential operations necessary.

We have implemented our simplified algorithm. (The original algorithm of [20] uses the software of [19] to compute discrete Morse cancellation and 1-unstable manifold extraction.) Below we give the size of the data and the comparison of the running times on the data sets used in [20]. In general, we see a factor of 4 time-speedup by simplification.

City	#traces	#points	time[20](s)	Our time(s)
Athens	118	1778400	311	48
Beijing	19287	3754580	390	111
Berlin	27189	326740	35	9

Table 1: Size and running time of three data sets. # points is the number of points in the grid after computing the density. The running time of computing 1-unstable manifolds (in seconds) by the algorithm of [20] and by our simplified algorithm is listed in the 3rd and 4th columns, respectively.

3.2 Editing strategies to add missing branches

Proposition 3.2. *Let $v \in \mathcal{K}$ be any vertex participating in a persistence pair (v, σ) with $\mathrm{pers}(v, \sigma) > \delta$. Then the output graph G of* SimpMorseRecon() *must contain v unless it is empty.*

Proof. One can view the cancellation of vertex-edge Morse pair as follows: at any moment, we maintain a spanning forest where each tree in this forest is represented by the *only* critical vertex in it at this point. At the beginning, all vertices are critical and each tree contains only one vertex. When we cancel a Morse pair (u, e) with $e = (u_1, u_2)$, we are merging two trees one of which is represented by u. After running SimpMorseRecon(), v is still a critical vertex. This follows from Lemma 3.1 as $\mathrm{pers}(v, \sigma) > \delta$. For simplicity, we assume that we have not yet simplified all vertex-edge persistence pairs; the latter case can be handled by a slightly modified argument which we omit due to limited space. The tree T_v containing v must have a critical edge e' in its co-boundary, which will merge T_v to another tree had we continued to perform Morse-cancellation of all vertex-edge persistence pairs. As this edge $e' = (w_1, w_2)$ is critical,

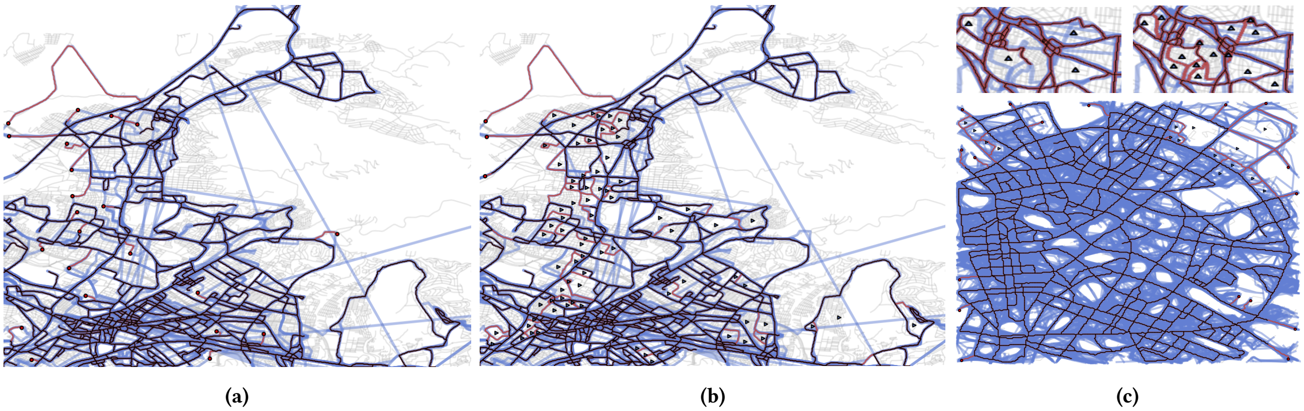


Figure 2: (a) Red points are added, red branches are newly reconstructed for the Athens map (black curves are original reconstruction, blue curves are input GPS traces). (b) We also add blue triangles to capture many missing loops as well. (c) Upper: An example to show that adding extra triangles will capture more loops. Bottom: Berlin with adding both branches and loops.

in Lines 8–9 of SimpMorseRecon(), we will extract its 1-unstable manifold and add to G. Assume w.o.l.g that the endpoint w_1 of e' is from T_v, then the gradient path from e' to v is the unique tree-path in T_v from w_1 to v. Hence v is included in the output graph G. □

We remark that the output graph G is empty **only if** after Line 7, there is no critical edge left – all mountain ridges are considered to be noise, and the algorithm rightfully output an empty set.

This result suggests the following strategy to augment the reconstruction with some missing branches: Suppose we obtain a reconstruction G from an input density field via SimpMorseRecon($\mathcal{K}, \rho, \delta$). The user can inspect the output, and identify places with potential missing road segments. For each such location, the user only needs to click at a vertex $u \in \mathcal{K}$. Let $U = \{u_1, \ldots, u_\ell\}$ be a collection of such locations. See the set of red dots in Figure 2 (a). We can then modify the input density field ρ into ρ' by setting $\rho'(u_i)$, for each $i \in [1, \ell]$, to be sufficiently large, say $\Delta + 2\delta$, where $\Delta = \max_{v \in V(\mathcal{K})} \rho(v)$; and set $\rho'(v) = \rho(v)$ for all other vertices in $\mathcal{K}$. We then re-run SimpMorseRecon($\mathcal{K}, \rho', \delta$) with the modified density field. Vertices in U necessarily have persistence larger than δ in the lower-star filtration induced by $-\rho'$. Hence by Proposition 3.2, the output graph will find "mountain ridges" to connect them.

See Figure 2 (a). We note however, this strategy does not complete missing loops. Hence we also propose an analogous strategy for the dual scenario, where the users can click a few triangles around regions with missing roads that form loops. We then set the new density ρ' value of some vertex of each of such triangle very low (meaning that this triangle eventually will correspond to a maximum in $f = -\rho'$). By a dual argument as the one used for Proposition 3.2 (which we omit due to lack of space), these triangles will remain critical after δ-Morse cancellation and loops around them will be captured. See Figure 2 (b) and (c).

Acknowledgment

This work is supported by NSF grants CCF1526513 and CCF1318595 and by National Institute of Health (NIH) under project R01 EB022899. We thank Suyi Wang for generously helping with the software.

REFERENCES

[1] M. Ahmed, S. Karagiorgou, D. Pfoser, and C. Wenk. 2015. A comparison and evaluation of map construction algorithms using vehicle tracking data. *GeoInformatica* 19, 3 (2015), 601–632.

[2] M. Ahmed and C. Wenk. 2012. Constructing Street Networks from GPS Trajectories. In *Proc. 20th Ann. Euro. Sympos. on Algo.* 60–71.

[3] U. Bauer, C. Lange, and M. Wardetzky. 2012. Optimal Topological Simplification of Discrete Functions on Surfaces. *Discr. Comput. Geom.* 47, 2 (2012), 347–377.

[4] J. Biagioni and J. Eriksson. 2012. Map inference in the face of noise and disparity. In *Proc. 20th ACM SIGSPATIAL*. 79–88.

[5] L. Cao and J. Krumm. 2009. From GPS traces to a routable road map. In *Proc. 17th ACM SIGSPATIAL*. 3–12.

[6] D. Chen, L. J. Guibas, J. Hershberger, and J. Sun. 2010. Road network reconstruction for organizing paths. In *Proc. 21st ACM-SIAM Symp. on Discr. Algo.* 1309–1320.

[7] J. J. Davies, A. R. Beresford, and A. Hopper. 2006. Scalable, Distributed, Real-Time Map Generation. *IEEE Pervasive Computing* 5, 4 (Oct. 2006), 47–54.

[8] O. Delgado-Friedrichs, V. Robins, and A. Sheppard. 2015. Skeletonization and Partitioning of Digital Images Using Discrete Morse Theory. *IEEE Trans. Pattern Anal. Machine Intelligence* 37, 3 (March 2015), 654–666.

[9] T.K. Dey, K. Li, C. Luo, P. Ranjan, I. Safa, and Y. Wang. 2010. Persistent heat signature for pose-oblivious matching of incomplete models. *Computer Graphics Forum* 29, 5 (2010), 1545–1554.

[10] S. Edelkamp and S. Schrödl. 2003. Route planning and map inference with global positioning traces. In *Computer Science in Perspective*. Springer, 128–151.

[11] H. Edelsbrunner and J. Harer. 2010. *Computational Topology - an Introduction*. American Mathematical Soc. I–XII, 1–241 pages.

[12] H. Edelsbrunner, D. Letscher, and A. Zomorodian. 2002. Topological persistence and simplification. *Discr. Comput. Geom.* 28 (2002), 511–533.

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

[14] A. Gyulassy, M. Duchaineau, V. Natarajan, V. Pascucci, E. Bringa, A. Higginbotham, and B. Hamann. 2007. Topologically Clean Distance Fields. *IEEE Trans. Visualization Computer Graphics* 13, 6 (Nov 2007), 1432–1439.

[15] F. Iuricich and L. De Floriani. 2017. Hierarchical Forman Triangulation: a Multiscale Model for Scalar Field Analysis. *Computers & Graphics* (2017).

[16] S. Karagiorgou and D. Pfoser. 2012. On Vehicle Tracking Data-Based Road Network Generation. In *Proc. 20th ACM SIGSPATIAL*. New York, NY, USA, 89–98.

[17] V. Robins, P. J. Wood, and A. P. Sheppard. 2011. Theory and Algorithms for Constructing Discrete Morse Complexes from Grayscale Digital Images. *IEEE Trans. Pattern Anal. Machine Intelligence* 33, 8 (Aug 2011), 1646–1658.

[18] W. Shi, S. Shen, and Y. Liu. 2009. Automatic generation of road network map from massive GPS vehicle trajectories. In *Proc. 12th International IEEE Conference on Intelligent Transportation Systems*. 48–53.

[19] T. Sousbie. 2011. The persistent cosmic web and its filamentary structure - I. Theory and implementation. 414 (June 2011), 350–383. arXiv:astro-ph.CO/1009.4015

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