Comparing Embedded and Immersed Graphs

Fréchet-like distances for graphs

Carola Wenk
Department of Computer Science
Tulane University

Collaborators:

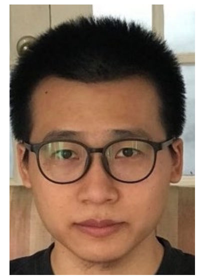
(Highly incomplete list)

Erfan Hosseini
Tulane U.

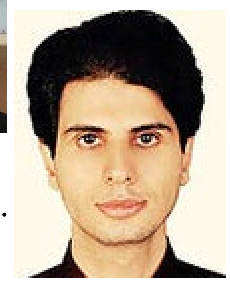
Maike Buchin
U. Bochum

Brittany Fasy
Montana State U.

Erin Chambers
Saint Louis U.



Pan Fan
Florida Sta U.



Majid Mirzanezhad
U. Michigan

Kevin Buchin
TU Dortmund

Ellen Gasparovic
Union College

Dieter Pfoser
George Mason U.

Mahmuda Ahmed
Google

Rodrigo Silveira
U. Poli. de Catalunia

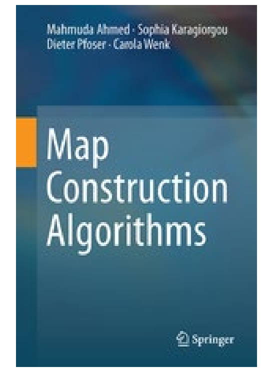
Liz Munch
Michigan State U.

Sophia Karagiorgou
Ubitech

Websites:

cs.tulane.edu/~carola
mapconstruction.org

Book:



Grant support:

NSF CCF-1618469, CCF-1637576, and CCF-2107434

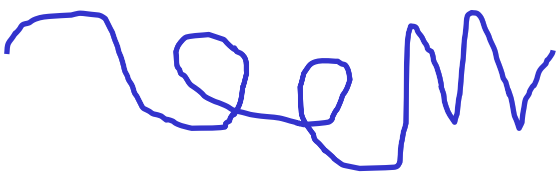
Upcoming paper:

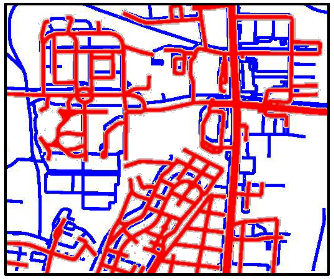
"Distances Between Immersed Graphs: Metric Properties", M. Buchin, E. Chambers, P. Fan, B.T. Fasy, E. Gasparovic, E. Munch, C. Wenk; in revision to La Matematica, 2022

Outline

1. 1D embedded data: Curves and embedded & immersed graphs
2. Hausdorff and Fréchet-like distances:
   - Hausdorff distance
   - Fréchet distance
   - Path-based distance
   - Traversal distance
   - Strong/weak graph distance
   - Contour tree distance
3. Other distances
   - Edit distance for geometric graphs
   - Point sampling distance

1. 1D Embedded Data



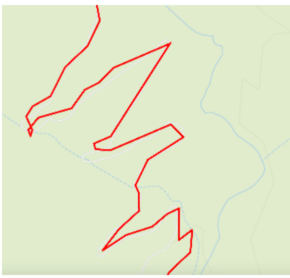


1D Embedded Data

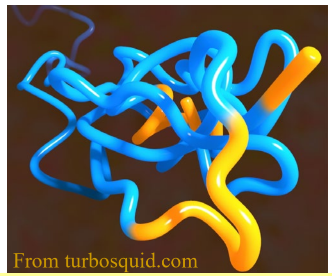
embedded in abient (usually Euclidean) space

C u r v e s

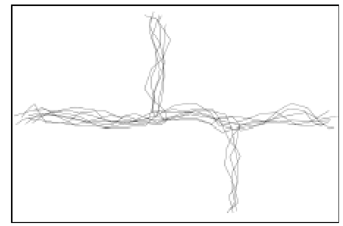
GPS trajectories



Protein chains

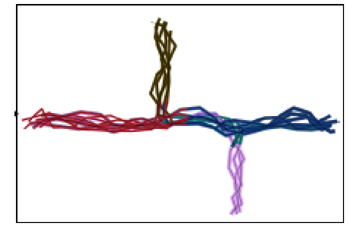


Set of trajectories



⇒

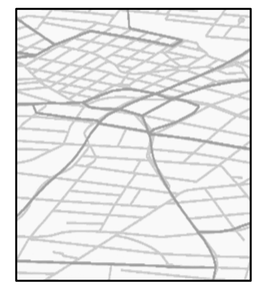
Sub-trajectory clusters



⇒Want to
order to fi

- Want to compare such 1D embedded data ⇒ Geometric shapes
- There are lots of distance measures and algorithms for comparing curves, and some for trees. But not so many for embedded (geometric) graphs.
- Graphs are the most general 1D shapes.

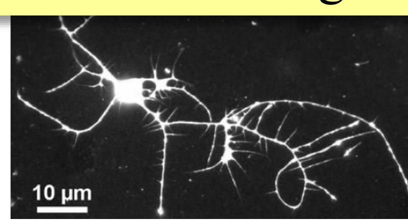
⇒



Constructed roadmap

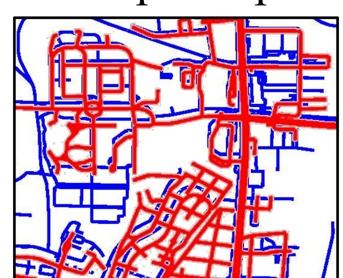
T r e e s

Plant mor



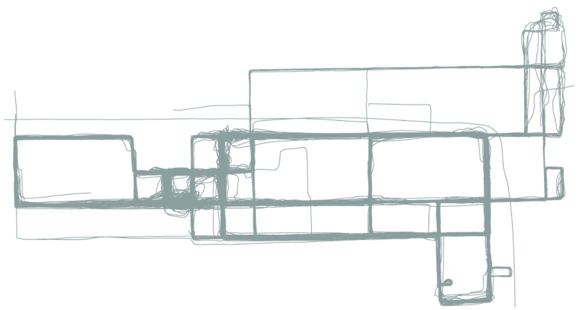
a p h s

Roadmap comparison

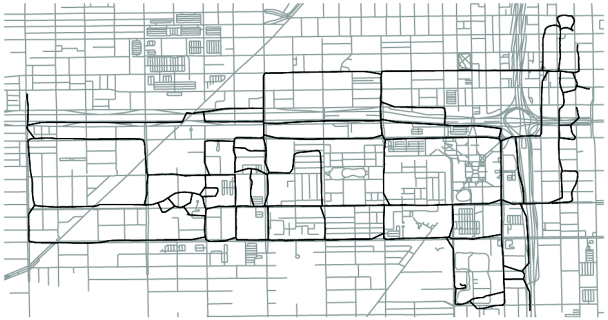


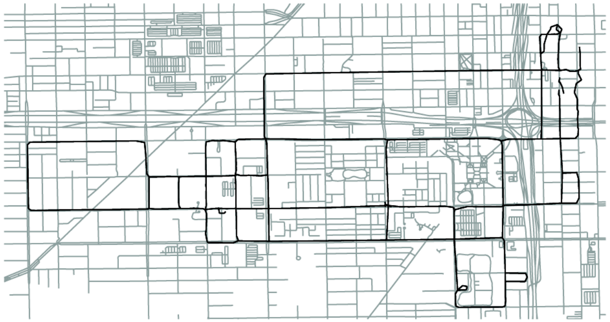
Compare Reconstructed Roadmaps

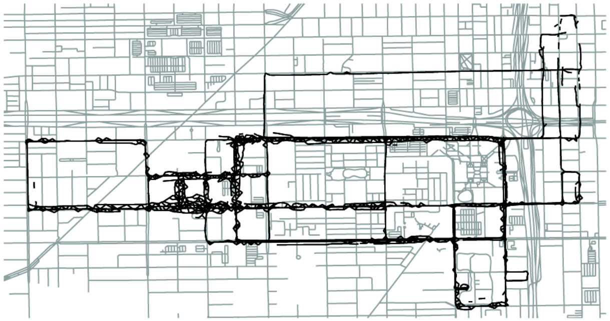
GPS Trajectory Data

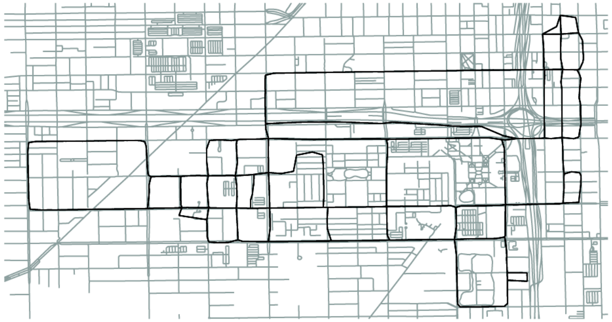


Reconstructed Roadmaps



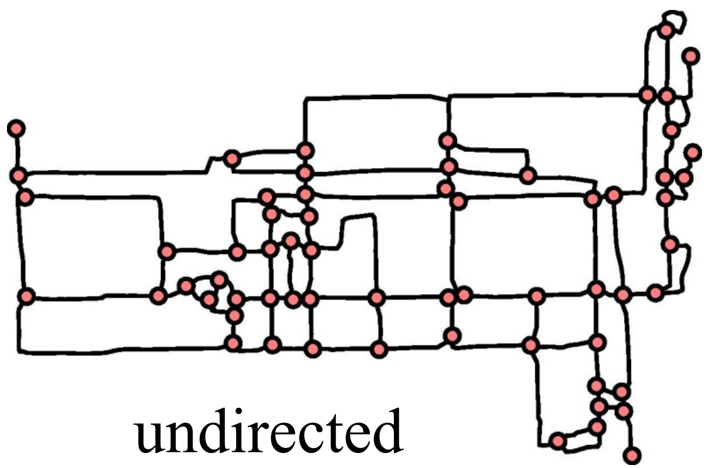


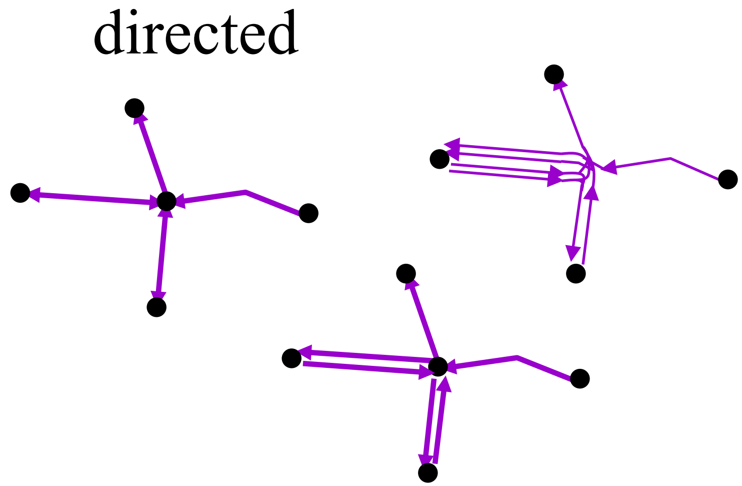




Embedded/Immersed Graphs

- Graph $G = (V, E)$ with a set of vertices V and edges E.
- Road network: Planar embedded





- Can consider G as a topological space (e.g., 1D simplicial complex)
- **Embedded graph**: Have a continuous function $\phi: G \to \mathbb{R}^d$, $d \geq 2$, that is homeomorphic onto its image.
- **Immersed graph**: $\phi: G \to \mathbb{R}^d$ is only **locally** homeomorphic onto its image.

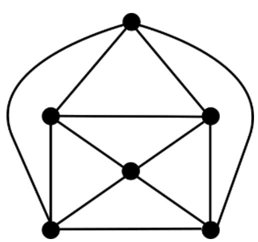
Homeomorphism: A continuous, bijective map whose inverse is continuous.

Embedded/Immersed Graphs

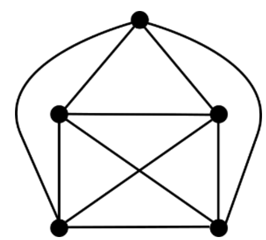
- **Embedded graph**: Have a continuous function $\phi: G \to \mathbb{R}^d$, $d \geq 2$, that is homeomorphic onto its image.
- **Immersed graph**: $\phi: G \to \mathbb{R}^d$ is only **locally** homeomorphic onto its image.

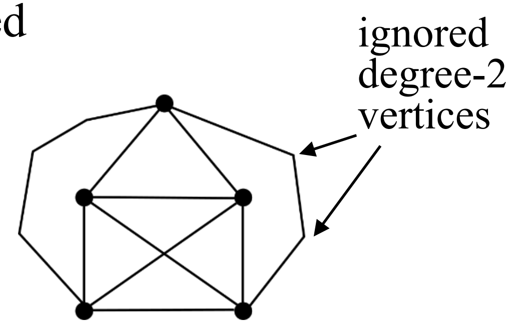
=> Each vertex is mapped to a point and edges are mapped to curves in $\mathbb{R}^d$ in such a way that the graph structure is maintained.

Embedding:
all edge-curves are non-crossing
(every crossing is a vertex)



Immersion:
“Bridges” are allowed



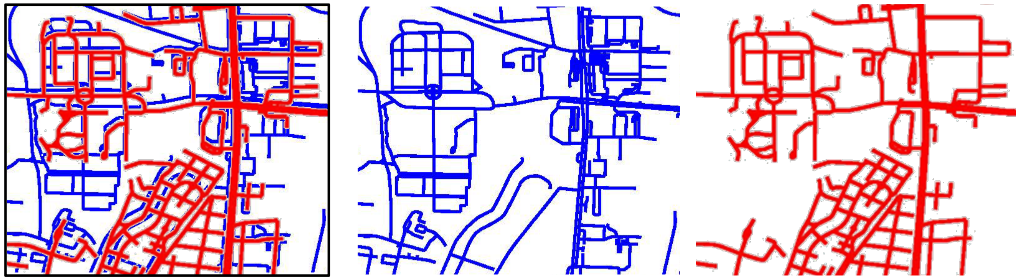


- Assume edge curves are piecewise linear, and may ignore deg-2 vertices

Immersed Graph Comparison

Given two immersed graphs (G, ϕ_G) and (H, ϕ_H), we want to compare them.

- How similar / different are they?
- What does it mean to be similar?
  - Depends on the application.
  - Graph isomorphism?



Here: Assume G and H are embedded in the same space and aligned.

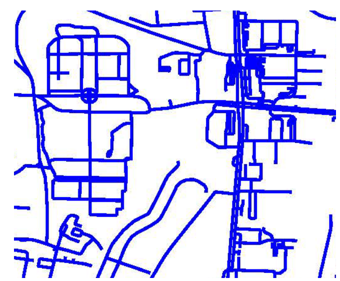
1. Define different distances between G and H, and study their properties (e.g., metric) and computational complexities.
2. Compute correspondences between portions of G and H.

When considering spaces of (immersed or embedded) graphs, we will consider underlying graphs up to homeomorphism.

Graph Isomorphism

- An isomorphism of $G = (V_G, E_G)$ and $\mathrm{H} = (V_H, E_H)$ is a
  - bijective map $f: V_G \to V_H$ for which holds
  - $\{u, v\} \in E_G \Leftrightarrow \{f(u), f(v)\} \in E_H$

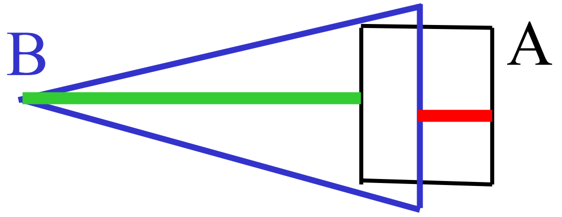
  Can be computed in linear time for planar graphs [HW74]
- Subgraph isomorphism: An isomorphism between G and a subgraph of H
  - NP-complete
  - Can be computed in linear time if G and H are planar and G has constant complexity [E95]

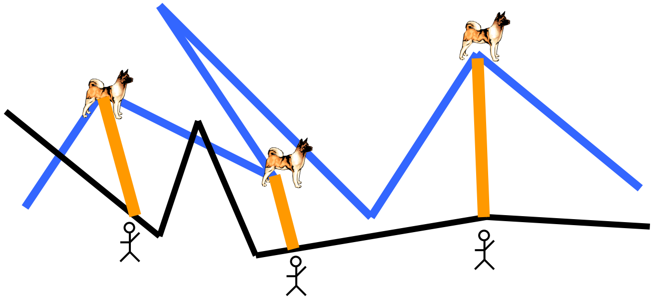



- Isomorphisms are bijective (1-to-1). However, we may want to allow 1-to-many assignments.
- We may also want to allow partial matchings.
- Isomorphisms are combinatorial in nature and don't take the embeddings/immersions into account.

[E95] D. Eppstein, Subgraph isomorphism in planar graphs and related problems, SODA: 632–640, 1995.
[HW74] J. Hopcroft, J. Wong, Linear time algorithm for isomorphism of planar graphs, STOC: 172–184, 1974.

2. Hausdorff and Fréchet-Like Distances

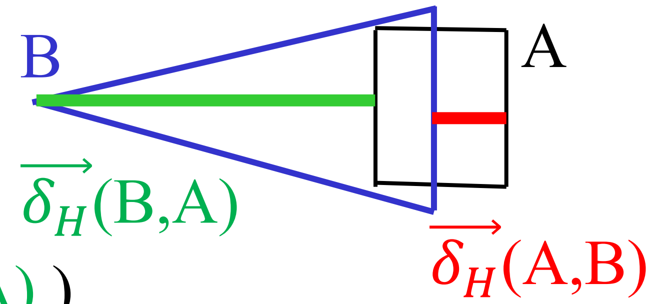




Hausdorff Distance

- Directed Hausdorff distance

$$\overrightarrow{\delta_H}(A,B) = \max_{a\in A} \min_{b\in B} \| a-b \|$$

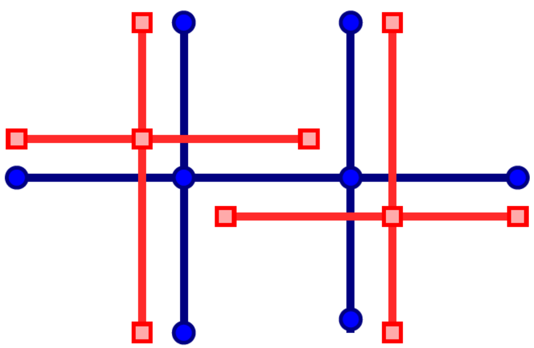


- Undirected Hausdorff-distance

$$\delta_H(A,B) = \max\left(\overrightarrow{\delta_H}(A,B)\,,\ \overrightarrow{\delta_H}(B,A)\right)$$

- Can be computed in polynomial time; O(N log N) in the plane.

- **Con:** When applied to graph comparison, δ_H only compares the geometry but not the topology
- **Pro:** $\overrightarrow{\delta_H}$ allows for partial comparison of one graph
- δ_H is a metric on the set of compact subsets of $\mathbb{R}^d$



Metric Properties

Definition 1 (Key Properties of Dissimilarity Functions). Let $\mathbb{X}$ be a set. Consider a function $d\colon \mathbb{X} \times \mathbb{X} \to \mathbb{R}_{\geq 0}$. We define the following properties:

1. Identity: $d(x,x) = 0$.
2. Symmetry: for all $x, y \in \mathbb{X}$, $d(x,y) = d(y,x)$.
3. Separability: for all $x, y \in \mathbb{X}$, $d(x,y) = 0$ implies $x = y$.
4. Subadditivity (Triangle Inequality): for all $x, y, z \in \mathbb{X}$, $d(x,y) \leq d(x,z) + d(z,y)$.

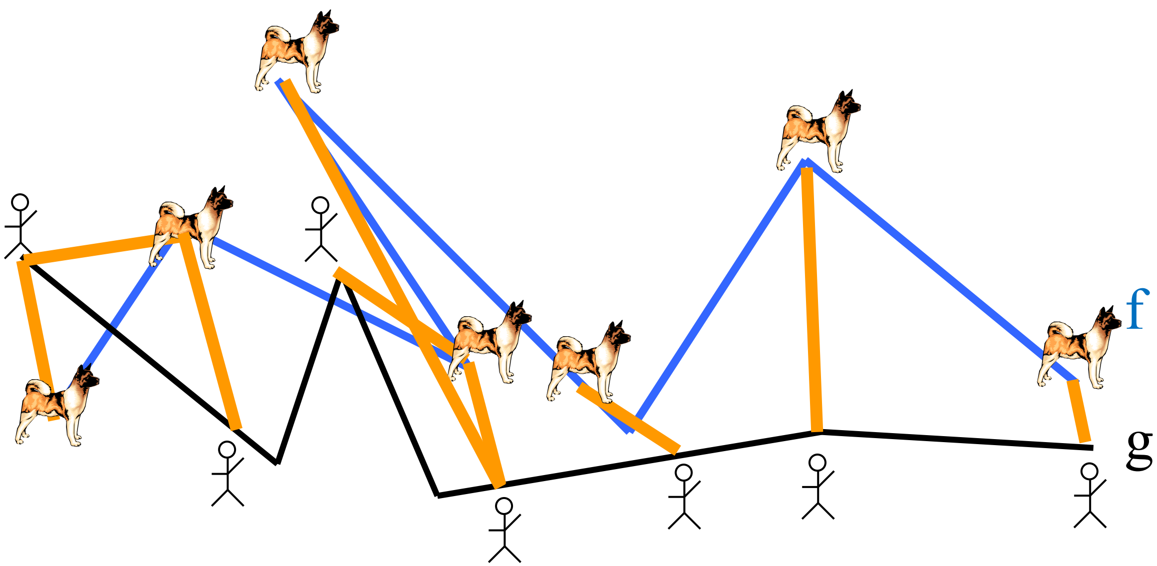
- **Metric:** Fulfills 1.-4.
- **Directed (as a modifier):** Does not fulfill 2.
- **Pseudo-metric:** 1., 2., 4.
- **Semi-metric:** 1., 2., 3.
- **Quasi-metric:** 1., 3., 4. (a directed metric)

$\Rightarrow \overrightarrow{\delta_H}$ is a directed pseudo-metric

Fréchet Distance for Curves

$$\delta_F(f,g) = \inf_{\alpha,\beta:[0,1]\to[0,1]} \; \max_{t\in[0,1]} \|f(\alpha(t)) - g(\beta(t))\|$$

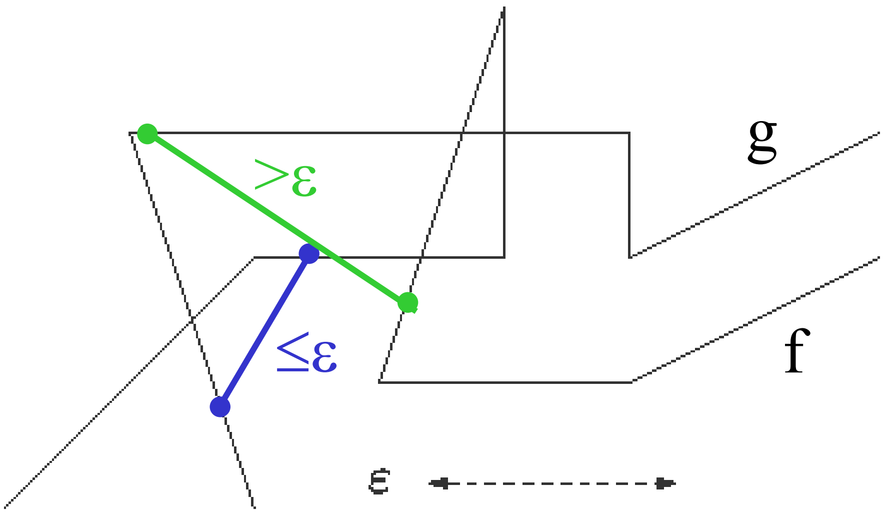
where α and β range over continuous monotone increasing reparameterizations only.

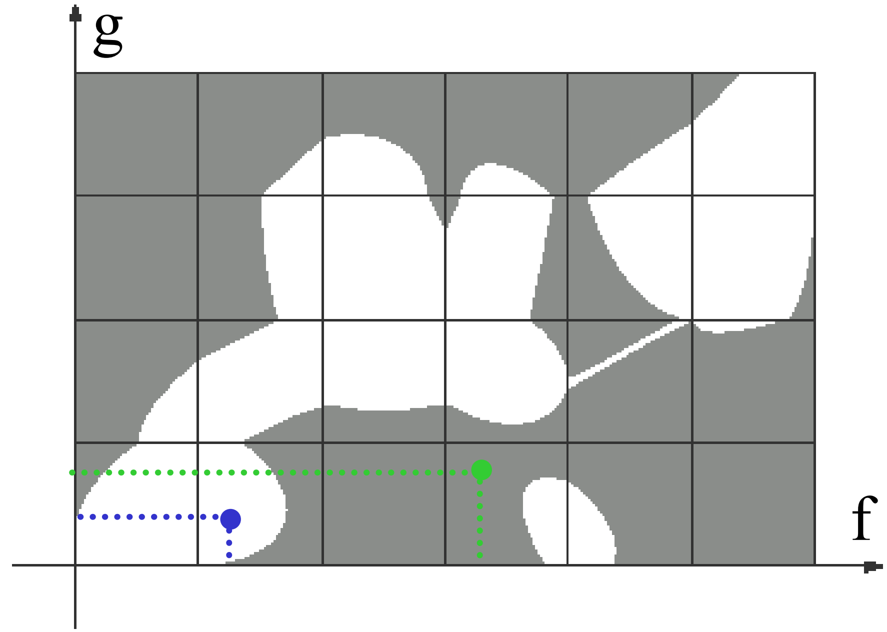


- Man and dog walk on one curve each
- They hold each other at a **leash**
- They are only allowed to go forward
- δ_F is the minimal possible leash length

[F06] M. Fréchet, Sur quelques points de calcul fonctionel, *Rendiconti del Circolo Mathematico di Palermo* 22: 1-74, 1906.

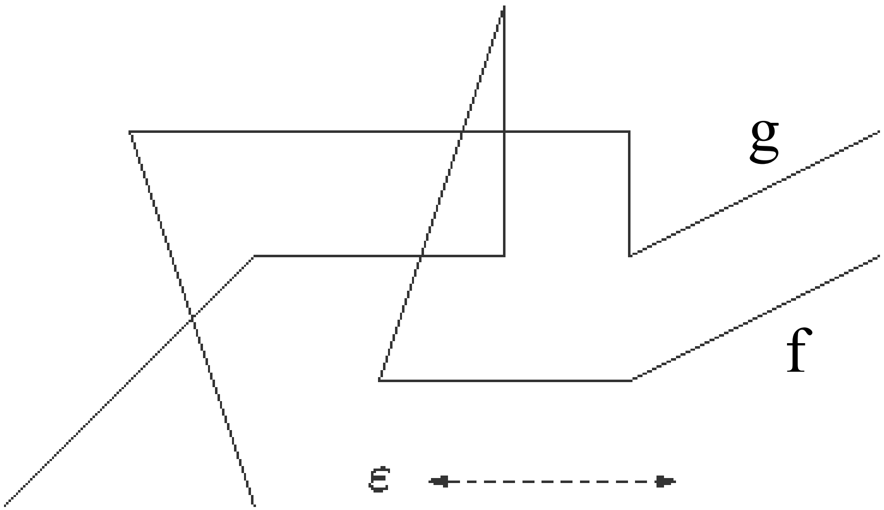
Free Space Diagram

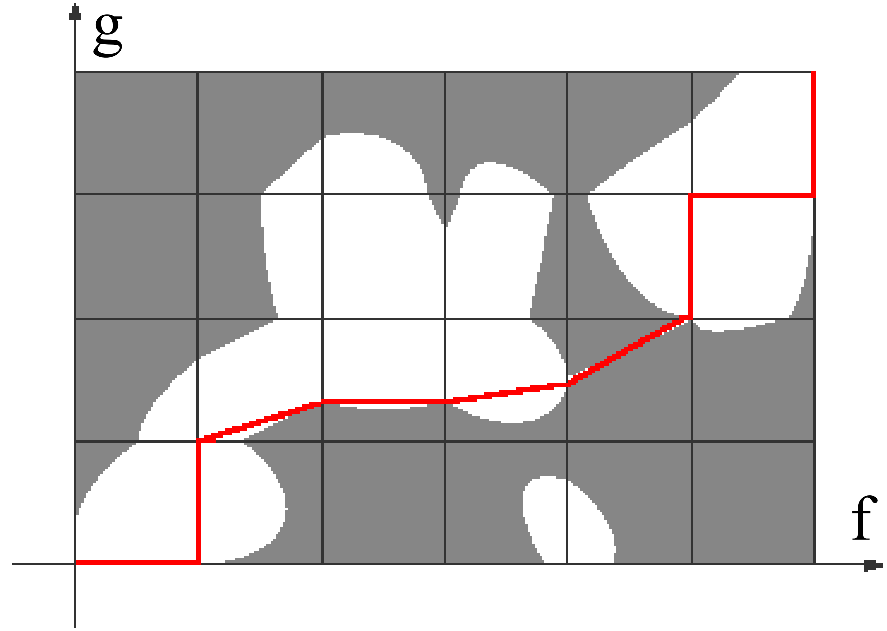




- Let ε > 0 fixed (eventually solve decision problem)
- $F_\varepsilon(f,g) = \{ (s,t) \in [0,1]^2 \mid \| f(s) - g(t) \| \leq \varepsilon \}$ *white points*
  free space of f and g
- The free space in one cell is an ellipse.

Free Space Diagram





- Monotone path encodes reparametrizations of f and g
- $\delta_F(f,g) \leq \varepsilon$ iff there is a monotone path in the free space from (0,0) to (1,1)
- Such a path can be computed using DP in O(*mn*) time

Fréchet Distance, General

Let $A, B \subseteq \mathbb{R}^k$ be two oriented manifolds. And let $f: A \to \mathbb{R}^d$ and g: $B \to \mathbb{R}^d$ be two immersions. Then

$$\delta_F(f, g) = \inf_{\alpha} \max_{t \in A} \|f(t) - g(\alpha(t))\|,$$

where $\alpha: A \to B$ ranges over all orientation-preserving homeomorphisms.

- The Fréchet distance is a metric (up to orientation-preserving homeomorphism)
- Originally defined for oriented manifolds, but can be generalized even further.

[M24] M. Fréchet. Sur la distance de deux surfaces. Ann. Soc. Polonaise Math. 3, 4–19, 1924.

Fréchet Distance, Immersed Graphs

Let (G, ϕ_G) and (H, ϕ_H) be two immersed graphs.

- We can apply the Fréchet distance definition in principle on the maps ϕ_G and ϕ_G.
- Drop the „orientation-preserving“ requirement.
- Equivalent definition:

$$\delta_F(G, H) = \min_{\alpha} \max_{e \in E_G} \delta_F(\phi_G(e), \phi_H(\alpha(e))),$$

where α ranges over all edge mappings corresponding to **isomorphisms** of G and H.

- Is graph-isomorphism hard. Can be computed in poly time for trees and for graphs of bounded tree-width. [BKN20]
- For connected planar graphs, can enumerate orientation-preserving isomorphisms in O(mn log(mn)) time*. [FW21]

[BKN20] M. Buchin, A. Krivosija, A. Neuhaus. Computing the Fréchet distance of trees and graphs of bounded tree width. EuroCG. 2020
[F**W**21] P. Fang, **C. Wenk**. The Fréchet distance for plane graphs. EuroCG 21.

Path-Based Distance

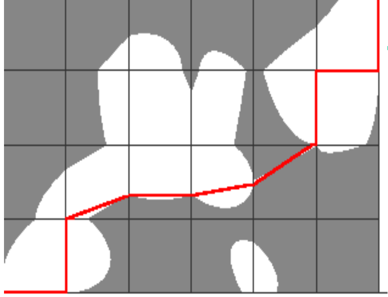
- Directed Hausdorff distance on path-sets:

$$\overrightarrow{d_{path}}(G,H) = \max_{p \in \Pi_G} \min_{q \in \Pi_H} \delta_F(\phi_G(p), \phi_H(q))$$

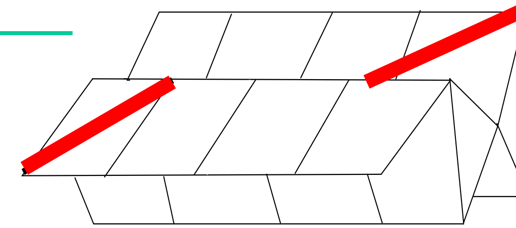
Fréchet distance

- Π_G path-set in G, and Π_H path-set in H
- Asymmetry of distance definition is desirable, if G is a reconstructed map and H a ground-truth map.

[AFH**W**14] M. Ahmed, B. Fasy, K. Hickmann, **C. Wenk**, Path-based distance for street map comparison, arXiv:1309.6131, 2014.



Path-Based Distance

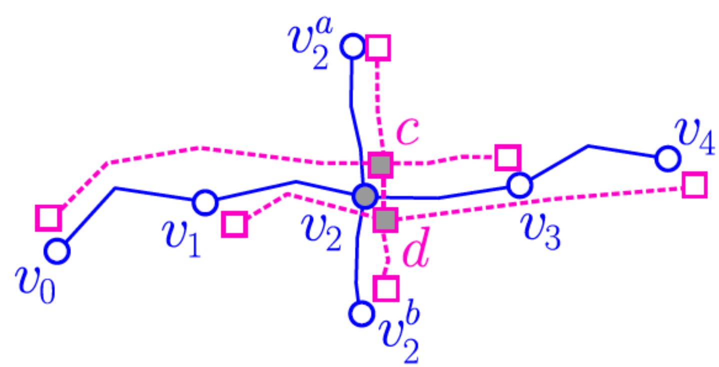


- Ideally, Π_G and Π_H are the set of all paths in G and H

$$\overrightarrow{d_{path}}(G,H) = \max_{p \in \Pi_G} \underbrace{\min_{q \in \Pi_H} \delta_F(\phi_G(p), \phi_H(q))}_{\text{map-matching}}$$

- It is a directed pseudo-metric.
- One can use the set of paths of link-length three to **approximate** the overall distance in polynomial time, if vertices in G are well-separated and have **degree ≠ 3**.

→ Stitch link-length three paths together to form longer paths



[AFHW14] M. Ahmed, B. Fasy, K. Hickmann, **C. Wenk**, Path-based distance for street map comparison, TSAS 1(1): article 3, 28 pages, 2015.

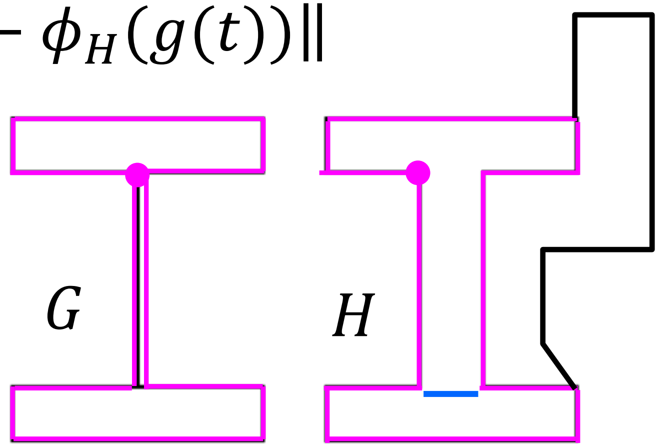
Traversal Distance

Let (G, ϕ_G) and (H, ϕ_H) be two immersed graphs.

- Represent G by traversals $f: [0,1] \to G$ (continuous, surjective) and H by partial traversals $\mathrm{g}: [0,1] \to H$:

$$\overrightarrow{d_T}(G,H) = \inf_{f,g} \max_{t \in [0,1]} \|\phi_G(f(t)) - \phi_H(g(t))\|$$

- Can be computed in O(mn log mn) time using free space diagram.
- Is a directed distance, but fulfills neither separability nor triangle inequality.
- Concides with the weak Fréchet distance when G and H are polygonal curves.



Small traversal distance

[AERW03] **H. Alt**, A. Efrat, **G. Rote**, **C. Wenk**, Matching Planar Maps, Journal of Algorithms 49: 262-283, 2003.

Strong and Weak Graph Distances

Let (G, ϕ_G) and (H, ϕ_H) be two immersed graphs.

- A graph mapping is a **continuous map** $s: G \rightarrow H$.

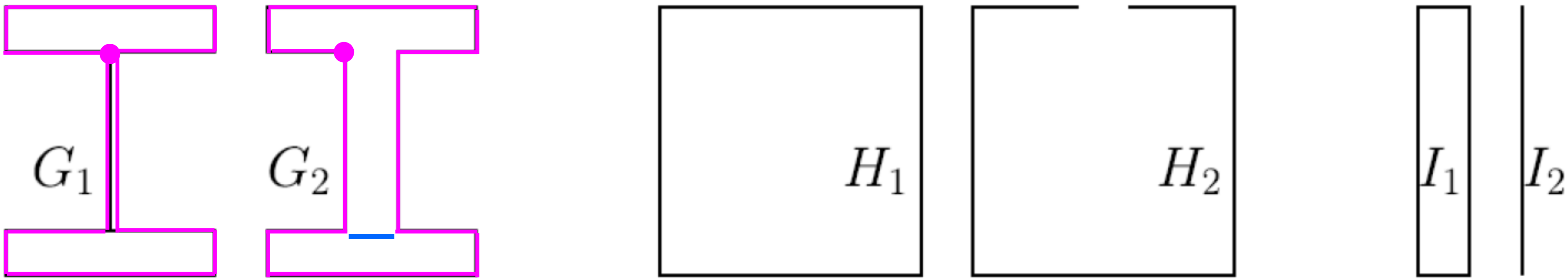
  Can be combinatorially represented as:
  - s sends each $v \in V_G$ to a point $s(v) \in H$
  - s sends each $\mathrm{e} \in E_G$ to a path from $s(u)$ to $s(v)$ in H.
- Then the strong graph distance is

$$\vec{\delta}(G, H) = \inf_{s:G \rightarrow H} \max_{e \in E_G} \delta_F(\phi_G(e), \phi_H(s(e)))$$

- The weak graph distance $\overrightarrow{\delta_w}$ uses δ_{wF} instead of δ_F.
- We have $\overrightarrow{d_T}(G, H) \leq \overrightarrow{\delta_w}(G, H) \leq \vec{\delta}(G, H)$
- NP-hard to decide, but poly time for trees. The weak distance can be computed in poly time for (spike-free) plane graphs.

[ABKS**W**21] H.A. Akitaya, M. Buchin, B. Kilgus, S. Sijben, **C. Wenk**, Distance Measures for Embedded Graphs, CGTA 95, 101743, 2021.

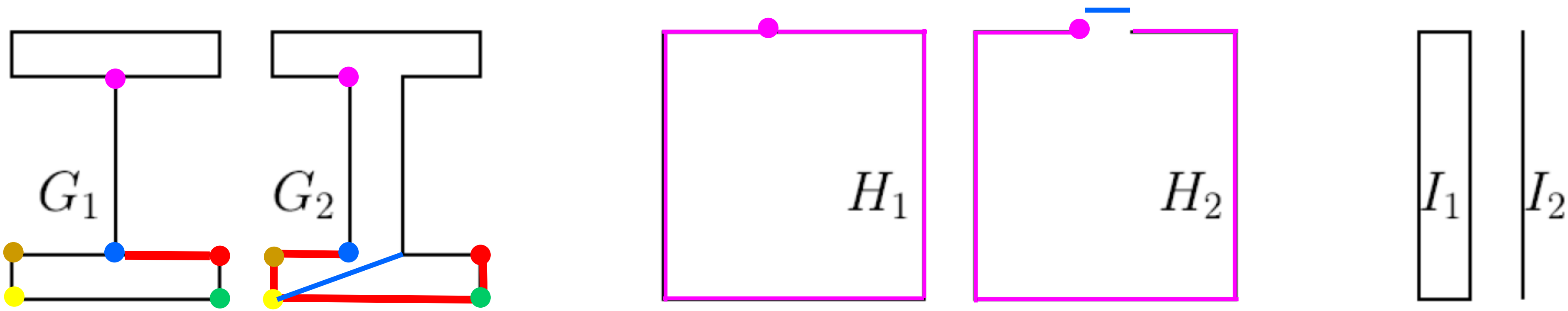
Traversal and Graph Distance



- Small traversal distance

[ABKSW21] H.A. Akitaya, M. Buchin, B. Kilgus, S. Sijben, **C. Wenk**, Distance Measures for Embedded Graphs, CGTA 95, 101743, 2021.

Traversal and Graph Distance

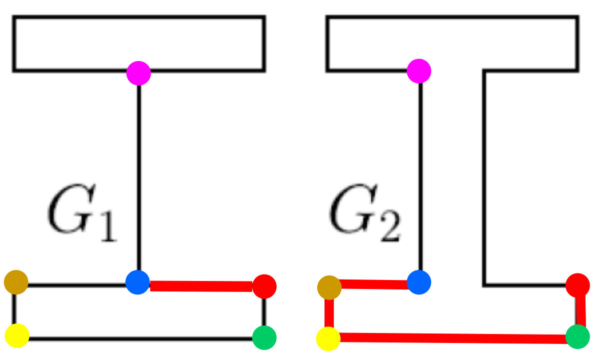


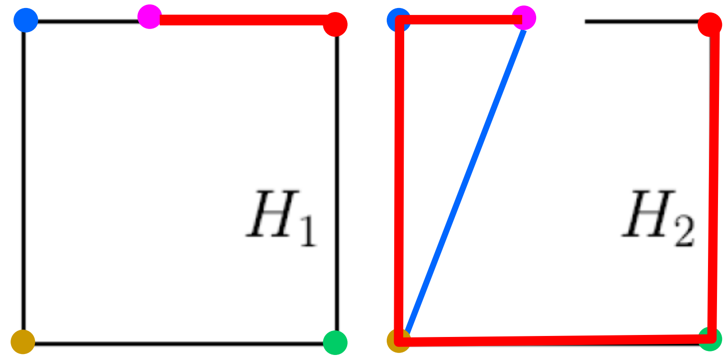
- Small traversal distance
- Large graph distance

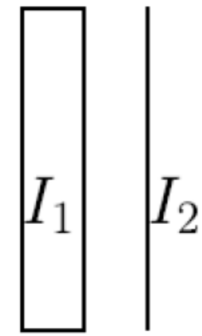
- Small traversal distance

[ABKSW21] H.A. Akitaya, M. Buchin, B. Kilgus, S. Sijben, **C. Wenk**, Distance Measures for Embedded Graphs, CGTA 95, 101743, 2021.

Traversal and Graph Distance







- Small traversal distance
- Large graph distance

- Small traversal distance
- Large graph distance

- Both small

Stay tuned for traversal distance code!




[ABKS**W**21] H.A. Akitaya, M. Buchin, B. Kilgus, S. Sijben, **C. Wenk**, Distance Measures for Embedded Graphs, CGTA 95, 101743, 2021.

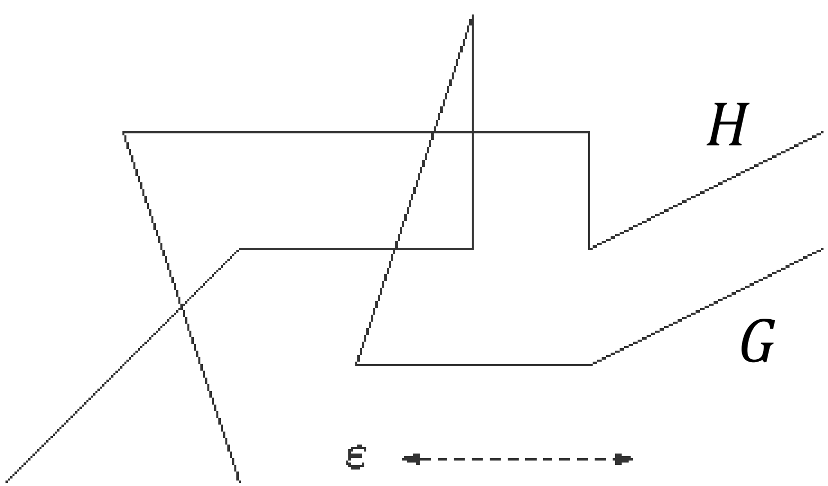
Contour Tree Distance

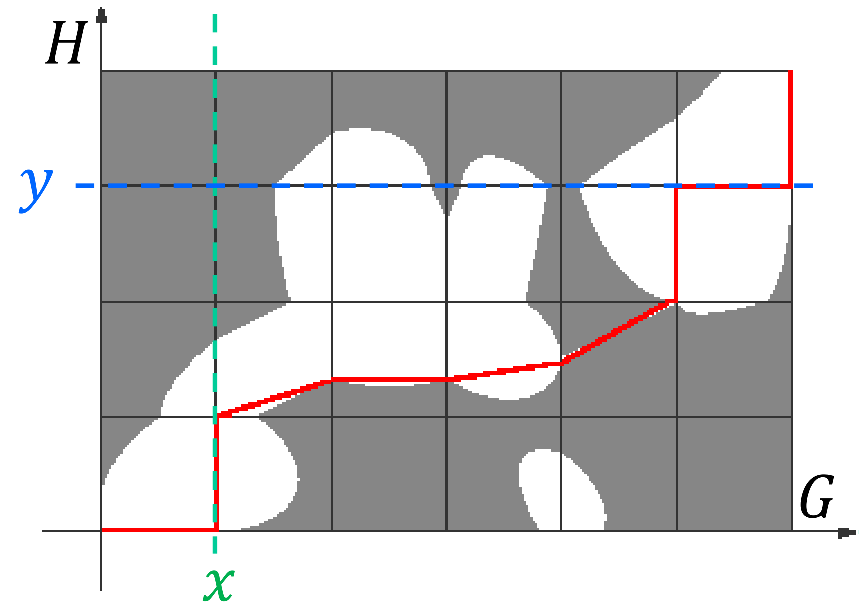
Let (G, ϕ_G) and (H, ϕ_H) be two connected immersed graphs.

$$d_C(G,H) = \inf_{\tau} \sup_{(x,y)\in\tau} \|\phi_G(x) - \phi_H(y)\|,$$

where G ranges over all correspondences τ between G and H such that

1. $\tau \subseteq G \times H$ is connected
2. For each $x \in G$: The set $\tau \cap (\{x\} \times H)$ is non-empty and connected
3. For each $\mathrm{y} \in H$: The set $\tau \cap (G \times \{y\})$ is non-empty and connected





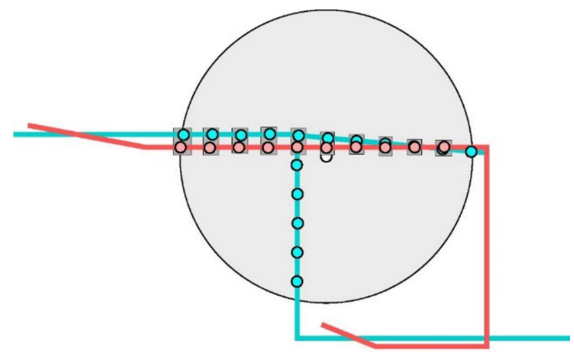
[BOS17] K. Buchin, T. Ophelders, B. Speckmann. Computing the Fréchet distance between real-valued surfaces. SODA: 2443–2455, 2017.

Contour Tree Distance

- The contour tree distance is a metric on connected graphs.
- But it is NP-complete, already for trees.
- This distance seems to correspond to a symmetric version of the (strong or weak) graph distances.

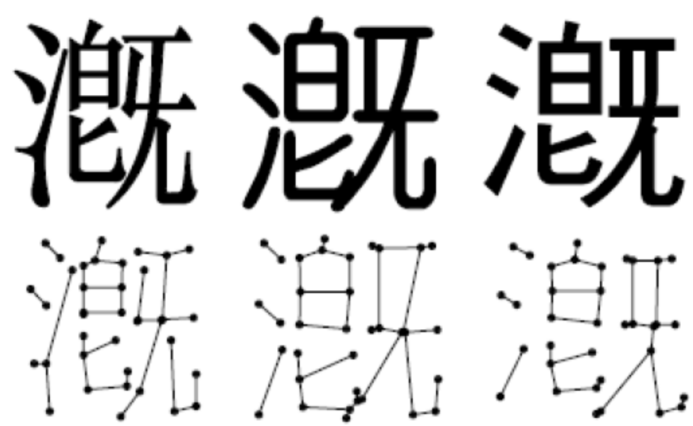
[BOS17] K. Buchin, T. Ophelders, B. Speckmann. Computing the Fréchet distance between real-valued surfaces. SODA: 2443–2455, 2017.
[BK**W**21] M. Buchin, B. Kilgus, **C. Wenk**. Ongoing work.

3. Other Distances



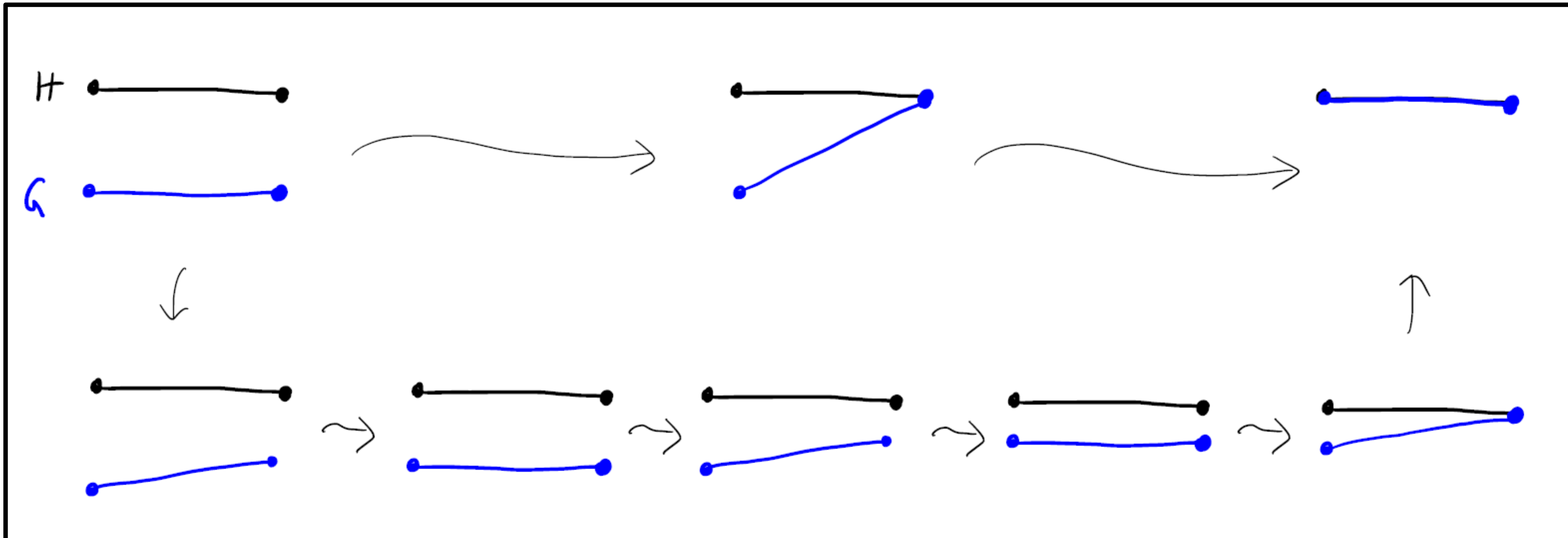
Geometric Edit Distance

- Geometric Edit Distance in $\mathbb{R}^2$:
  - Defined for straight-line embedded graphs.
  - Motivated by Chinese character comparison
  - Perform the following edit operations in this order:
    Edge deletion, vertex deletion,
    vertex translation,
    vertex insertion, edge insertion
  - Costs are proportional to edge length differences and to the distance a vertex has moved.
  - Is a metric. But NP-hard.



[CGKSS09] O. Cheong, J. Gudmundsson, H.-S. Kim, D. Schymura, F. Stehn, Measuring the similarity of geometric graphs, SEA: 101–112, 2009.

Geometric Edit Distance

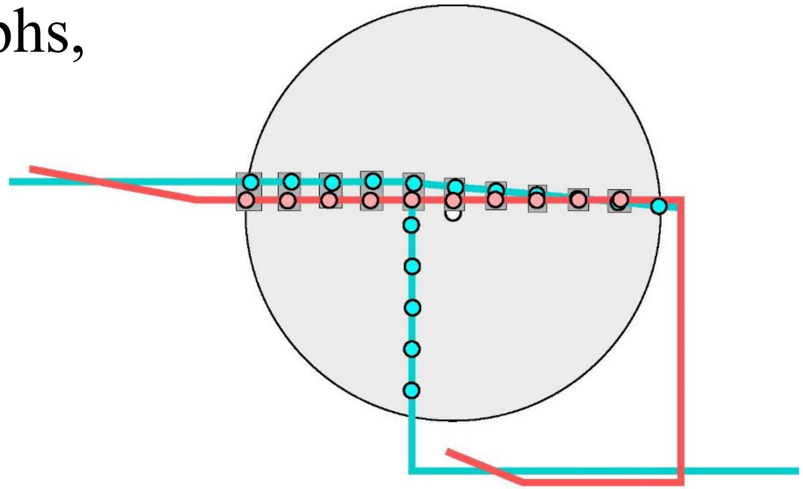


- Costs are proportional to edge length differences and to the distance a vertex has moved.
- Is a metric. But NP-hard.

[CGKSS09] O. Cheong, J. Gudmundsson, H.-S. Kim, D. Schymura, F. Stehn, Measuring the similarity of geometric graphs, SEA: 101–112, 2009.

Graph Sampling Distance

- In a local neighborhood of both graphs, traverse the graphs (from random seeds) and place point samples. (Only graph edges of length ≤ τ.)

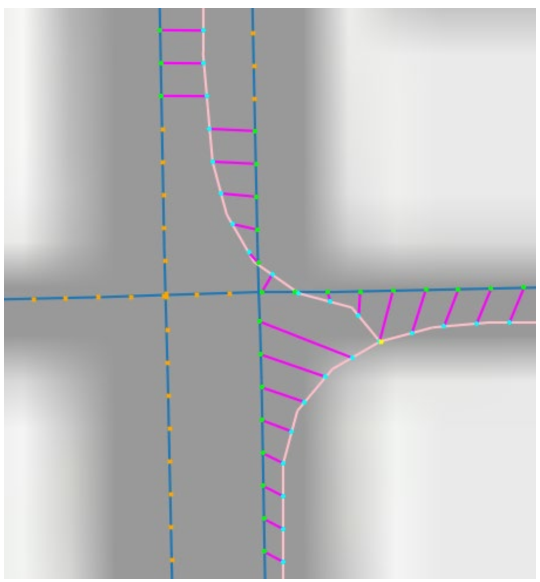


- $\boldsymbol{\tau}$: match_distance threshold
  $\boldsymbol{m}=m(\tau)$: #samples in G
  $\boldsymbol{n}=n(\tau)$: #samples in H
  $\boldsymbol{k}=k(\tau)$ = #matched samples (1-1) within distance τ
- **Precision:** $p = k/n$ **Recall:** $r = k/m$ **F-score:** $2pr/(p+r) = 2k/(n+m)$

[BE12b] J. Biagioni, J. Eriksson, Inferring road maps from global… TRR: J. of the Transportation Research Board 2291, 61-71, 2012.

Graph Sampling Distance

- Lacks theoretical foundation but is practical.
- Does not work well if the reconstructed graph is compared to a more detailed ground-truth graph (e.g., OSM).
- Provides a matching (1-to-1) between a subset of points in G and H



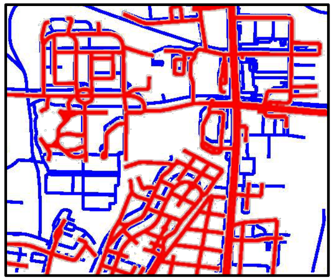
- What is a good matching? [ABBHSW21]
- Graph sampling toolkit
github.com/Erfanh1995/GraphSamplingToolkit
- Continuous definition via length-preserving Fréchet correspondence [BFHW21]

[ABBHS**W**21] J. Aguilar, K.&M Buchin, E. Hosseini S., R.I. Silveira, **C. Wenk**, Graph Sampling for Map Comparison, Spatial Gems, 2021

[BFH**W**21] K. Buchin, B.T. Fasy, E. Hosseini S., **C. Wenk**, Measuring Length-Preserving Fréchet Correspondence for Graphs in R^2, FWCG, 2021

Conclusion & Discussion

1. We've seen a lot of distances for immersed graphs.
   - Are they useful in practice? (Noisy input, runtimes)
   - What are their mathematical properties? (Metric, topological) [CFHMW22]
2. Would like to compute a correspondence / mapping between the two graphs efficiently.
   - An application: Merge multiple road networks



3. What's a good edit distance definition for immersed graphs?
4. Optimize under transformations

[CFHM**W**22] E. Chambers, B.T. Fasy, B. Holmgren, S. Majhi, **C. Wenk**, Path-Connectivity of Fréchet Spaces of Graphs, YRF 2022