FACULTY OF ENGINEERING
AND ARCHITECTURE

Department of Electronics and Information Systems
research groups AIDA - DaMBi

Graph Approximations to Geodesics on Metric Graphs

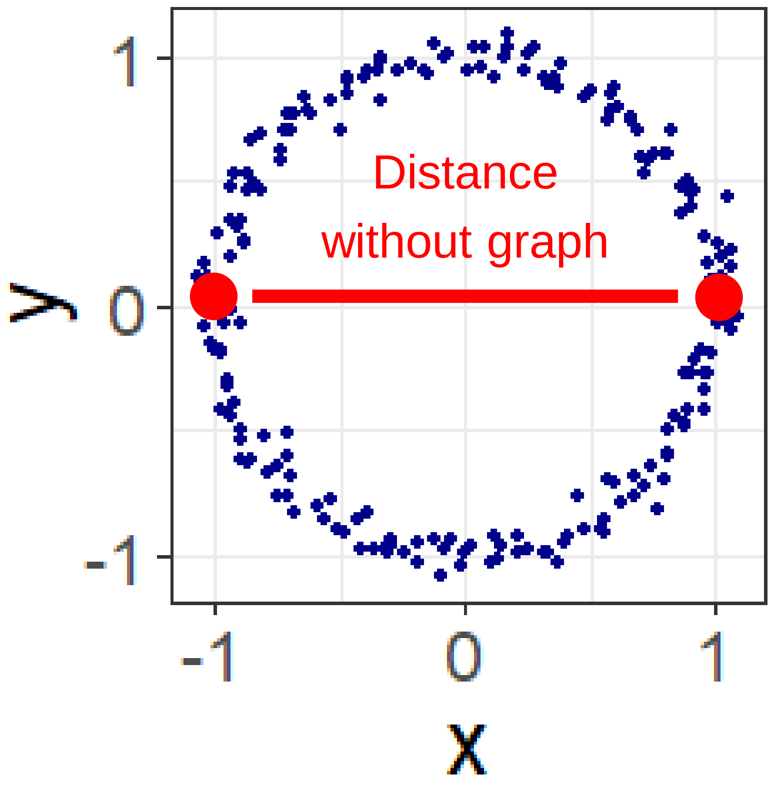
Robin Vandaele, Yvan Saeys, and Tijl De Bie

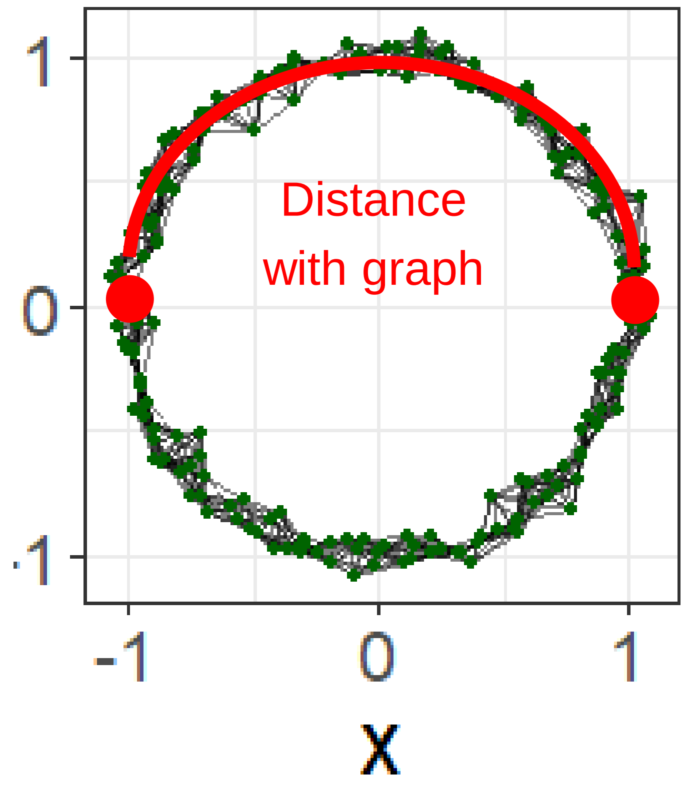
GHENT UNIVERSITY
ARTIFICIAL INTELLIGENCE & DATA ANALYTICS
fwo
erc European Research Council Established by the European Commission
IDLab INTERNET & DATA LAB
VIB science meets life

Proximity graphs are often used to approximate geodesics

– ***k*NN:** connect each point to its *k* closest neighbors

– **Rips/unit disk:** connect every pair of points within distance ε





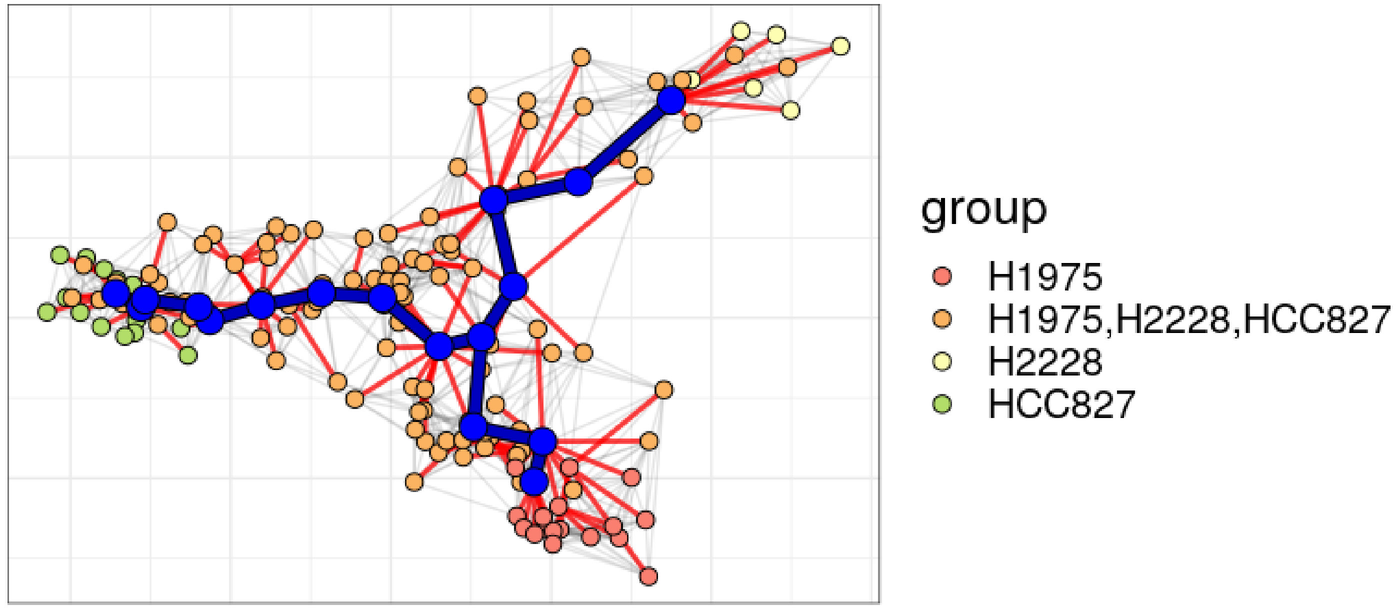
GHENT UNIVERSITY

Proximity graphs are often used to approximate geodesics

- Used for various dimensionality reductions
  - Laplacian Eigenmaps
  - Locally Linear Embedding
  - ISOMAP
  - Maximum Variance Unfolding
  - Local Tangent Space Alignment
  - …

- More recently as intermediate representation for topological inference (after initial dimensionality reduction such as PCA, diffusion maps. …)
  - E.g., Cell Trajectory Inference

Model = Metric Graph



GHENT UNIVERSITY

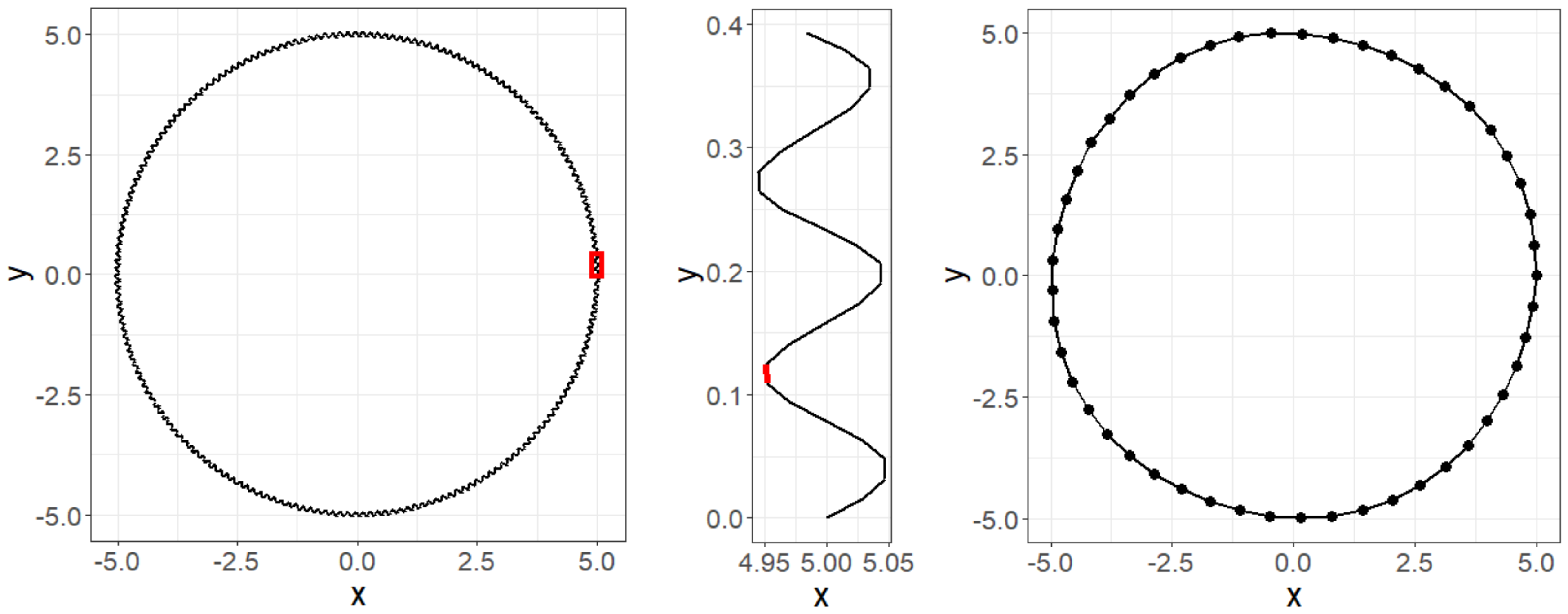
Prior theory

[1] M. Bernstein, V. D. Silva, J. C. Langford, and J. B. Tenenbaum, "Graph approximations to geodesics on embedded manifolds," 2000.

- Mainly serves as a theoretical justification of the ISOMAP dimensionality reduction
- Provides conditions under which shortest path distances in proximity graphs are close to true geodesics
- Restricted to smooth submanifolds of $\mathbb{R}^n$.

Prior theory – the issues

- Restricted to smooth submanifolds of $\mathbb{R}^n$ → no singularities allowed
- Stringent conditions: all local patterns must be approximated well



Proximity graph fails to capture local 'wiggled' pattern
→ Violates conditions in [1]

Captures global cyclic pattern correctly

GHENT UNIVERSITY

Our work – new and more flexible characteristics for metric graphs

Definition 3. *Given a connected metric graph M. For any $\epsilon > 0$ we define the* branch seperation $s_\epsilon(M) \in \mathbb{R}^+ \cup \{\infty\}$ of M at resolution $1/\epsilon$ *as*

$$s_\epsilon(M) := \sup\{s \in \mathbb{R} : \|x - y\| < s \implies d_M(x,y) \leq \epsilon\}.$$

Specifies the scale of interest

(solves flexibility)

Definition 4. *Given a connected metric graph M. For any $0 \leq \epsilon' < \epsilon$ we define the* linearity of M between resolutions $1/\epsilon$ and $1/\epsilon' \in \mathbb{R}^+ \cup \{\infty\}$ *as*

$$\lambda_{\epsilon',\epsilon}(M) := \sup\{\lambda \in \mathbb{R} : \epsilon' \leq d_M(x,y) \leq \epsilon \implies \lambda d_M(x,y) \leq \|x - y\|\}.$$

Constrains from 'how close' we look at the data

(necessary for extension to singularities)

Proof in paper

Theorem 3. *Let M be a connected metric graph in $\mathbb{R}^d$ and let X be a finite set of data points in M. Suppose a graph $G = (X, E)$ is given, defining the following three thresholds:*

1) $\|x - y\| \geq \epsilon'$ *for all* $\{x, y\} \in E$,
2) $\|x - y\| < s_\epsilon(M)$ *for all* $\{x, y\} \in E$, *with* $\epsilon > 0$,
3) *for all* $x, y \in X$ *with* $\|x - y\| \leq \tau$, *we have* $\{x, y\} \in E$.

*If for $0 < 4\delta < \tau$, X satisfies the δ-*sampling condition, *i.e., for every $m \in M$ there is $x \in X$ with $d_M(m, x) \leq \delta$, then for all $x, y \in M$*

$$\lambda_{\epsilon,\epsilon'}(M) d_M(x,y) \leq d_G(x,y) \leq (1 + 4\delta/\tau) d_M(x,y). \quad (1)$$

GHENT UNIVERSITY

References

[2] Robin Vandaele, Yvan Saeys, and Tijl De Bie. “Mining Topological Structure in Graphs through Forest Representations,” Journal of Machine Learning Research, 21(215):1–68, 2020.

– Topological inference of metric graphs through proximity graphs

[3] Robin Vandaele. “Topological Data Analysis of Metric Graphs for Evaluating Cell Trajectory Data Representations,” Master’s thesis, Ghent University, 2020.

– Extension of current theoretical results (to probabilistic results and noise)
– Applications for evaluating cell trajectory data representation quality

[4] Robin Vandaele. “Topological Inference in Graphs and Images,” Doctoral thesis, Ghent University, 2020.

– Further applications of current results for comparing geometric patterns.

GHENT UNIVERSITY

FACULTY OF ENGINEERING
AND ARCHITECTURE

Robin Vandaele

Department of Electronics
and Information systems

E Robin.Vandaele@UGent.be
T +32 9 331 49 56

www.ugent.be

Universiteit Gent
@ugent
@ugent
Ghent University

GHENT
UNIVERSITY