Graph Approximations to Geodesics on Metric Graphs

Robin Vandaele, Yvan Saeys, and Tijl De Bie
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 $\varepsilon$
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

\[ \text{Model} = \text{Metric Graph} \]
Prior theory


- 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 $\mathbb{R}^n \rightarrow$ no singularities allowed
- Stringent conditions: all local patterns must be approximated well

Proximity graph fails to capture local ‘wiggled’ pattern
$\rightarrow$ Violates conditions in [1]
Captures global cyclic pattern correctly
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 separation $s_\epsilon(M) \in \mathbb{R}^+ \cup \{\infty\}$ of $M$ at resolution $1/\epsilon$ as

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

**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

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

Specifies the scale of interest (solves flexibility)

Constrains from ‘how close’ we look at the data (necessary for extension to singularities)

**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 $\delta$-sampling condition, i.e., for every $m \in M$ there is $x \in X$ with $d_M(m, x) \leq \delta$, then for all $x, y \in M$

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

Proof in paper
References


- Topological inference of metric graphs through proximity graphs


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


- Further applications of current results for comparing geometric patterns.
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