Stability Progress

Timothy Sudijono

May 3, 2019

Introduction

We would like to better understand which graph distance we are approximating with our pipeline.

Recall the following setup:

- We define a *graph* $G = (V, E)$ with edge weights and vertex weights, defined by $w_v : V \to \mathbb{R}, w_e : E \to \mathbb{R}$. Let the space of such graphs be denoted $\mathcal{G}$, and we can metrize this with several choices.
  - Generally, we can use the graph edit distance. We have seen that certain choices of the cost functions lead to a stability result with bottleneck distance.
  - In any situation, we can also use the *network distance* $d_{\mathcal{N}}$, defined in *Persistent Homology of Asymmetric Networks*.
  - If (M, g) is a (Riemannian) manifold and f a function on it, we can turn a finite collection of points $\{x_i\}_{i=1,\dots,n}$ to a graph by considering $w_v(x_i) = f(x_i)$, $(x_i, x_j) \in E$ for every i, j and $w_e(x_i, x_j) = d(x_i, x_j)$. We can measure distance between these graphs by the d_{TL^p}. Interpreting each collection of points as a point mass probability measure, we can define d_{TL^p} distance between elements $(\mu, f), (\nu, g) \in \mathcal{P}(M) \times \mathbb{L}^p(M)$ via

$$d_{TL^p}((\mu, f), (\nu, g)) = \left(\inf_{\pi \text{ coupling of } \mu, \nu} \int |x - y|^p + |f(x) - g(y)|^p d\pi(x, y) \right)^{1/p}$$

- We have a dynamic network, i.e. a time series of graphs $\{G_t\}_{t \in T}$, with $T \subset \mathbb{R}^+$.
- We convert this to a time series of filtered simplicial complexes, by converting each graph $G_t = (V_t, E_t)$ to a filtered simplicial complex on V_t where the birth time of a simplex $\sigma \subset V_t$ is given by

$$\max_{v_1, v_2 \in \sigma} (w_v(v_1), w_e(v_1, v_2)).$$

- This can be converted to a time series of persistence diagrams.
- Run a sliding window analysis with bottleneck distance.

Stability

We would like to understand what distance we are approximating when we convert a graph into a persistence diagram via the steps above. In essence, we are looking for a stability type result between the graph distance we choose and the bottleneck distance between the persistence diagrams as defined above.

0.1 Graph Edit Distance

Recall the definition of the graph edit distance. Define an *edit* between two graphs as an operation of the following form: edge/node addition, edge/node deletion, edge/node substitution. To each of these edits e_i is assigned an edit cost $\mathrm{Cost}(e_i)$. The graph edit distance between two graphs G_1, G_2 is the infimum of total costs over all edit paths between G_1, G_2. That is, if $e_0, \ldots, e_n$ transform G_1 to G_2, then the total cost of this edit path is $\sum_{i=0}^{n} \mathrm{Cost}(e_i)$. The graph edit distance is the infimum over such costs.

We have established that with the following choices of cost functions, we can establish stability.

- Edge substitution has the cost of the absolute value of the change in edge weight.
- Edge insertion and deletion has the cost of the absolute value of the weight of the edge being edited, perhaps plus a fixed cost.
- Node substitution has the cost of λ times the absolute value of the change in the node weight.
- Node insertion has the same cost of λ times the absolute value of the added node weight, perhaps plus a fixed cost.
- Node deletion has the cost of removing the link of the node, since we had to reduce the weight of the node to 0 and the surrounding edges to 0 to obtain a bound on the network distance. That is, we can take the cost of deletion to be the max of the node weight and the surrounding edge weights:
$$\max\big(|w_v(v)|, \max_{x:(x,v)\in E}(|w_e((x,v))|\big).$$

Proposition 0.1. *With this, we have*

$$d_B(\mathrm{Dgm}_k(G_1), \mathrm{Dgm}_k(G_2)) \leq K d_{GED}(G_1, G_2),$$

for a certain constant K.

0.2 Network Distance

The network distance is defined in [Persistent Homology of Asymmetric Networks]. We define a *network* (X, ω_X) as a set X with weight function $\omega_X : X \times X \to \mathbb{R}^+$.

We define the *network distance* as the following pseudometric:

Definition 0.1 (Network distance)**.**

$$d_{\mathcal{N}}(X,Y) := \frac{1}{2} \min_{R \in \mathscr{R}} \operatorname{dis}(R)$$

where $\mathscr{R}$ is the set of all correspondences between X, Y, which are subsets R of $X \times Y$ such that $\pi_X(R) = X, \pi_Y(R) = Y$, and the *distortion* of such a relation is given by

$$\operatorname{dis}(R) = \max_{(x,y),(x',y') \in R} \left| \omega_X(x,x') - \omega_Y(y,y') \right|.$$

One can define the filtered Rips complex $\left\{ \mathfrak{R}^{\delta}_X \hookrightarrow \mathfrak{R}^{\delta'}_X \right\}_{\delta \leq \delta'}$ where

$$\mathfrak{R}^{\delta}_X := \left\{ \omega \in 2^X : \max_{x,x' \in \sigma} \omega_X(x,x') \leq \delta \right\}.$$

Note that the definition of the Rips diagram above for networks is exactly the construction we use for our pipeline: edge weights correspond to $\omega_X(x,x')$, or ∞ if the edge is not present, and node weights correspond to $\omega_X(x,x)$.

Then denote the H_k-homology persistence diagrams of these complexes by $\operatorname{Dgm}^{\mathfrak{R}}_k(X)$.

The authors prove the following stability result:

Proposition 0.2. *We have*

$$d_B(\operatorname{Dgm}^{\mathfrak{R}}_k(X), \operatorname{Dgm}^{\mathfrak{R}}_k(Y)) \leq 2 d_{\mathcal{N}}(X,Y).$$

So we could use this distance as the one we are approximating with our pipeline. **Question: how easy is it to compute the network distance**?

0.3 d_{TL^p} distance

We do not know if such a stability result can be established. As before, suppose we have a manifold M with a smooth function $f : M \to \mathbb{R}^+$ on it. We consider graphs whose nodes are finite collections of points $\{x_i\}_{i=1,\dots,n} \subset M$ and whose edge weights are given by geodesic distance on M. The intuition is that the graphs serve as discrete approximations to our space M.

Given two graphs G_1, G_2 with nodes $(x_1, \ldots, x_n), (y_1, \ldots, y_m)$ respectively, consider them as probability measures on M by associating G_1 with $\mu_{G_1} := \frac{1}{n}\sum \delta_{x_i}$ and similarly for G_2. Then, given two functions f, g on M and the graphs G_1, G_2, we would like the following inequality to be true:

$$d_B(\mathrm{Dgm}((G_1, f)), \mathrm{Dgm}((G_2, g))) \leq K d_{TL^p}((\mu_{G_1}, f), (\mu_{G_2}, g)).$$

There were a few ideas on how to go about proving this: suppose one had two graphs and functions with d_{TL^p} distance bounded by δ. We seek to prove that their persistence diagrams have bottleneck distance bounded by $K\delta$ for K some constant. Intuitively, the result seems true since the d_{TL^p} bound limits how much the function values f, g can differ, so the function values are close, as well as how much the structure of the graphs can differ (however, edge values aren't really captured by this distance)?