Introduction and results

Motivation: What is a natural measure of latent distance on a graph? Our answer: The Laplace transform of the hitting time.

Our contributions:
1. A new technique for analyzing random walks on graphs through continuum limits.
2. Extension of von Luxburg-Rudi-Hein 2014: the expected hitting time is a degenerate measure of distance.
3. A principled estimator — Laplace transformed hitting time:
   • Consistent for recovering a latent distance metric
   • Respects underlying density and cluster structure
   • Robust to model misspecification

The spatial graph model

Summary: We define a random network model with edges determined by a latent metric. The spatial graph model depends on the following latent quantities:

- \( p(x) \): latent probability density in compact connected smooth domain \( D \subset \mathbb{R}^d \),
- \( x_i, (x_{i-1}, x_{i+1}) \): coordinate points drawn i.i.d. from \( p(x) \),
- \( R(x_i) \): local scale function (may depend on \( x_i \)),
- \( h : R \rightarrow (0, 1) \): connectivity kernel with \( h(x) = 0 \) for \( x > 1 \), \( h(1) = 0 \), and \( h \) left-continuous at 1.

Definition 1: Spatial graph

The simple random walk

Theorem 1: Continuum limit of the walk (Hashimoto-Sun-Jaakkola 2015)

Figure 1: An heuristic explanation of the scaling limit of Theorem 1.

Figure 2: Estimated distance from orange starting point on a k-nearest neighbor graph constructed on two clusters. A and B show degree of Jennings approximating continuum limit (Theorem 2). D, E, and F show that log-LTHT interpolates between hitting time and shortest path.

Simulating Brownian motion on the latent space

Summary: We give a method to modify the transition probabilities of the random walk on a spatial graph so trajectorily converges to Brownian motion on the latent metric space.

Notations: \( q_i(x, y) \) the probability of transitioning to \( y \) started at \( x \) in \( t \) time.

We make a regularity assumption on transitions of the simple random walk:

For \( t = \Theta(q_{i,j}(x,y)) \) a.s. eventually uniformly equicontinuous. \( \sigma \) Hashimoto-Sun-Jaakkola 2015: under \( \sigma \) can consistently estimate \( p(x) \) and \( h(x) \).

Theorem 3: Simulating Brownian motion on the latent space

Figure 3: Distributions of 40-step random walks on a k-rewired graph with original and Brownian weighting. Points drawn from Gaussian restricted to a disk.

Application: Degeneracy of expected hitting time

Summary: Although expected hitting time between two vertices is a commonly used measure of distance, we show it is degenerate for spatial graphs and recovers only the inverse of the stationary density.

Notation: \( \overline{T}(x) \) the hitting time of \( x \) started at \( x \) to \( y \).

A common measure of distance is:

Expected hitting time as \( \overline{T}(x) \):

Note: Generalizes surprising result of von Luxburg-Rudi-Hein 2014 in the undirected case.

Proof intuition:
1. Compare the hitting time of the simple random walk to its fluid process equivalent.
2. Compare the fluid process with Brownian motion and show by transience that it is unlikely to hit quickly.
3. Conditioned on slow hitting, the random walk moves before it hits, yielding the stationary distribution.

Laplace transformed hitting time (LTHT)

Summary: We propose the Laplace transform of the hitting time as a metric estimator. We have an expectation hitting time reveals the following drawbacks:

1. The expectation is dominated by long paths.
2. Long paths depend on regimes of the graph unrelated to the geodesic.

Resolution: Consider the Laplace transform of the hitting time:

for \( D \) the simple random graph transformed hitting time (LTHT) is defined as:

and can be computed via the matrix inverse for any transition matrix \( W \):

\[ \log \frac{\overline{T}(x)}{\overline{T}(y)} \sim \log \left( 1 - \left( 1 - 1/n \right) W_{x,y} / W_{y,x} \right) \] \]

LTHT is closely related to rooted page rank \( \overline{T} \) and potential distance \( \delta \).

LTHT puts greater weight on shorter paths.

LTHT avoids instability of shortest paths by averaging paths near the geodesic.

PDE Characterization: LTHT is solution to Feynman-Kac boundary value problem

Theorem 6: Properties of LTHT

Proof:
1. By Theorem 1, \( \overline{T}(x,y) = \min \{ n \} \) is the solution to the boundary value problem with boundary condition \( \overline{T}(y) = 0 \).
2. By Theorem 1, \( \overline{T}(x,y) = \min \{ n \} \) is the solution to the boundary value problem with boundary condition \( \overline{T}(y) = 0 \).

LTHT preserves clusters

Summary: LTHT learns a cluster preserving metric in 1.0 without reweighting.

Consider the density-dependent metric:

which separates points in different clusters.

Suppose \( D \) and \( D' \) are LTHT-consistent.

The hitting time \( \overline{T}(x,y) \) of a simple random walk from \( x \) to the out-neighborhood of \( y \) converges to:

\[ \log \overline{T}(x,y) \sim \log \left( 1 + n \right) \] \]

LTHT is consistent

Summary: For scale parameter \( \beta = \exp(-\beta T) \) a modification of LTHT gives consistent metric recovery. We will consider hitting times to a spherical neighborhood of a vertex in \( G \), the graph equivalent of the ball \( B(x,s) \).

Modify LTHT using the following notions:

- A neighborhood \( N_B(x,s) \) is estimated set of vertices within a distance of \( s \).
- \( \overline{T}(x,y) : \) hitting time of the transformed walk on \( G \), from \( x \) to \( N_B(x,s) \).

Theorem 5: Consistency of LTHT

Proof:
1. By Brownian motion \( W_t \) started at \( x \), log-LTHT for \( T_{out}(x,s) \) the hitting time of \( W_t \) to \( N_B(x,s) \) converges to the latent metric.

Empirical consistency:

Evaluation: Link prediction task

Summary: LTHT-based methods outperform benchmarks on two link prediction tasks.

Figure 4: Estimated distance vs. true latent distance for different values of \( \beta \) on reweighted walks (simulated dataset) on a 20-Network graph with 500 vertices and \( k > 100 \).

References