

Getting lost in space

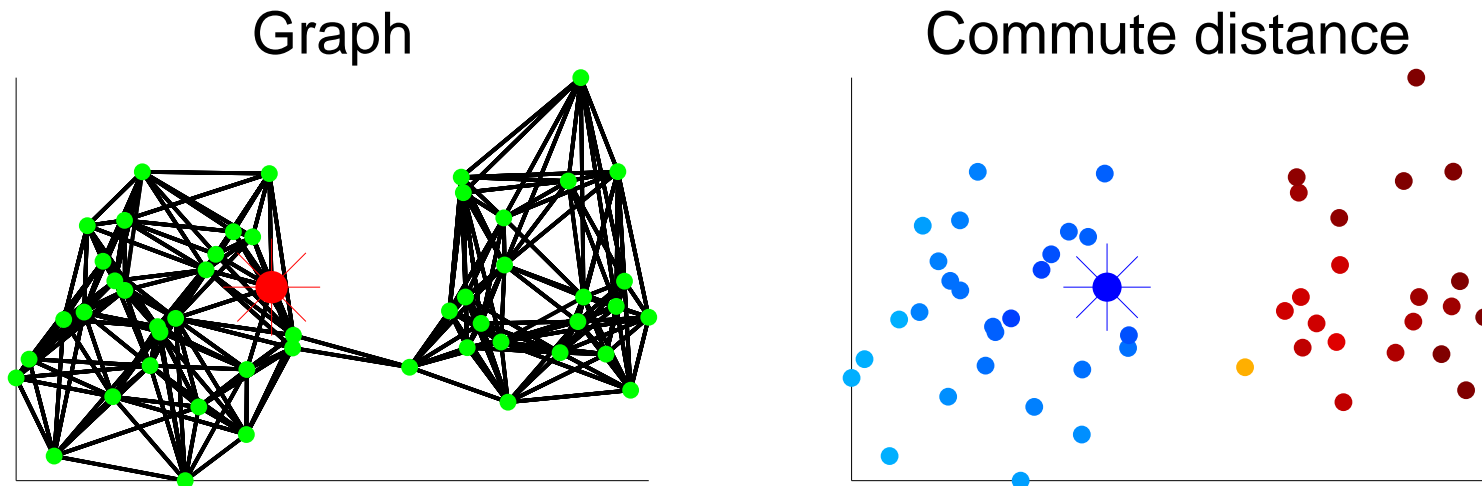
Large sample analysis of the resistance/commute distance

Ulrike von Luxburg, Agnes Radl and Matthias Hein

Speaker: Matthias Hein, Saarland University

Idea of similarity graphs in machine learning:

- connect similar points - build global structure from local structure



Commute distance on graphs:

- expected number of steps for the random walk on the graph to go from one vertex to another vertex and back.

Usage: ranking, clustering, semi-supervised learning, network analysis

Intuition: commute distance captures cluster structure in the data !

Random geometric graphs:

- samples are drawn i.i.d. from a probability measure in \mathbb{R}^d
- k -nearest neighbor or ε -neighborhood graph

Sketch of Main Result:

Commute distance C_{ij} is meaningless for **large** random geometric graphs:

$$C_{ij} \approx \frac{1}{d_i} + \frac{1}{d_j} \quad \text{for } i \neq j,$$

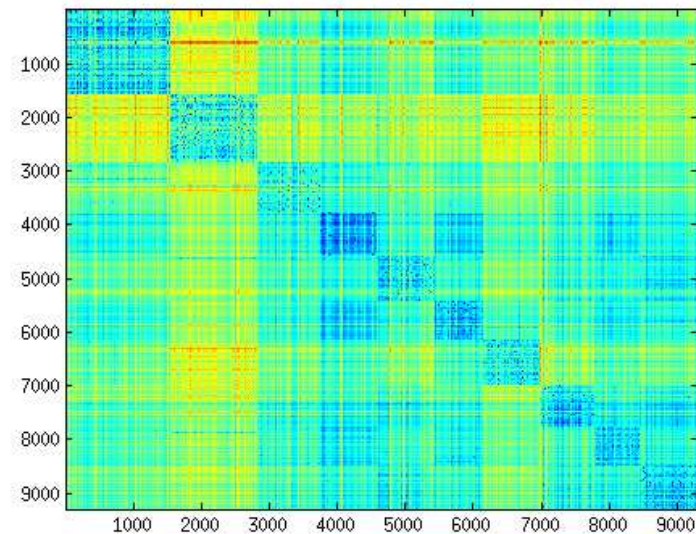
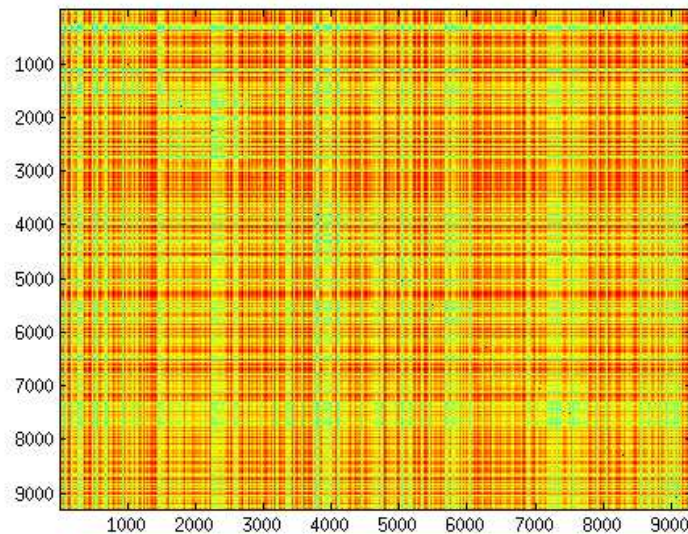
where $d_i = \sum_{s=1}^n w_{is}$ is the degree of vertex i .

Implications:

- depends only on **local** connectivity - **no global** structure incorporated
- all points have the same nearest neighbor: the point with maximal degree

Result generalizes to other (random) graph types

USPS: 10-nearest neighbor graph



Left: Original commute distance, Right: Amplified commute distance

At the poster:

- detailed limit results + discussion of more general graph types
- amplified commute distance and associated kernel
- comparison of different distances for semi-supervised learning on USPS