5339 - Algorithm design under a geometric lens, Spring 2014, CSE, OSU
Project suggestions
Instructor: Anastasios Sidiropoulos

Project 1. From graphs to metrics. Find some data set that can be represented as a graph. For example, the internet graph $G = (V, E)$, a social network graph, etc. Define a metric space $M = (V, d)$, i.e. a distance function $d : V \times V \to \mathbb{R}_{\geq 0}$, that satisfies the metric conditions (symmetry, and triangle inequality). Argue that the distance d captures the "dissimilarity" between the elements in V. For example, for a social network graph G, the set V is the set of users. Then, for $u, v \in V$, the distance $d(u, v)$ should be large if the two users u, v are very dissimilar.

You can also compare several different functions d, and argue about the benefits, of each one. Argue which distance function is more appropriate for each application.

Project 2. Implementations. Chose your favorite embedding / algorithm from class, implement it, and test it on some real, or synthetic data. This can also be done in combination with Project 1.

One idea would be to implement different embeddings, and compare the resulting distortion in practice. E.g., compare the distortion of Bourgain's embedding, with the distortion of embedding into random trees.

If you chose to implement a nearest-neighbor, or clustering algorithm, you can evaluate how well these algorithms work in practice.

Project 3. The complexity of real-world data. Chose a metric space $M = (X, d)$ (this might be synthetic, or it might arise from some application). Estimate the doubling dimension of M. You can also try to estimate the parameter β of Lipschitz partitions.

Your metric can be interpreted as Euclidean if X consists of a collection of real-values vectors, and you chose the distance d to be the Euclidean norm of their difference. Then, you can implement the random projection method we discussed in class, and estimate the minimum dimension required to get distortion $1 + \varepsilon$.

Project 4. Study other related papers. You can pick a paper related to the theme of the class, and present it in class. These are some suggested papers/subjects (but feel free to chose another):

- E. Grant, C. Hegde and P. Indyk, *Nearly Optimal Linear Embeddings into Very Low Dimensions*, IEEE GlobalSIP Symposium on Sensing and Statistical Inference, 2013.
- Sofiane Abbar, Sihem Amer-Yahia, Piotr Indyk, and Sepideh Mahabadi) *Efficient Computation of Diverse News*, WWW, 2013.
- k-d Trees.
- Multidimensional scaling.

Project 5. Theory / algorithms research. There are several open problems related to the subjects we have seen in class. You can pick a problem that seems interesting, and review some of the related research papers. You can also think about some of the open questions, and write your approach for solving them. You do not have to completely solve a problem. The intention is for you to get exposed in some interesting open questions.