Discovery Thread: Project 2

In this project you will apply three techniques for dimension reduction and model explanation.

On the weighted undirected graph dataset either assigned to your project, or chosen from an available online database perform the following three tasks:

I. Random graph hypothesis:

1. Turn the weighted graph into an ordered sequence of edges (starting with the largest weight edge and continuing with the next largest weight, and so on).
2. Compute: a) the sequeence of 3-cliques; b) the sequence of 4-cliques; c) the sequence of the second smallest eigenvalue of the normalized symmetric unweighted graph Laplacian.
3. Compare the previous metrics to the predicted sequences under the random graph hypothesis.

II. Geometric graph embedding:

First turn the weigthed graph data into sets of distances using exponential and power laws. Then for each set of distances repeat:

1. Solve the Semidefinite Program (SDP) for the quadratic form G;
2. Perform the SVD factorization of the matrix G and estimate the appropriate dimension and the embedding coordinates;
3. Compare the results for each weight model (exponential and power laws), and several embedding dimensions.

III. Laplacian eigenmaps:

First, for the normalized symmetric weighted Laplacian compute the bottom $D+1$ eigenvectors.

For each $d = 2, 3, ..., D$ repeat

1. Construct the geometric graph in d dimensions; Order edges according to the pairwise distances
2. Compute the sequence of 3-cliques and 4-cliques
3. Compare to the sequence computed on the original weighted graph (Part I.2)

Visualize for $d = 2, 3$ and conclude the dimension d_* that matched closest the 3-clique and 4-clique metrics.