1 Team

Christopher James Brooks -
Gurupragaash Annasamy Mani - u1010032

2 Project Overview

We will use Mapper [**?**] on large datasets to see if we can find any interesting clustering that differs from the usual *k*-means, etc. in ways that depend on the shape of the data.

3 Technical Details

We are going to analyze a large collection of Netflix users' movie ratings [**?**] using the mapper algorithm. To use mapper, we need to represent the data as a point cloud and we need to have a notion of distance between data points that reflects the similarity of users' preferences. Finally, we need a filter function mapping users to R1 or R2. We will be using L-infinity centrality as our filter function.

We will also use more traditional clustering technique(*k*-means) as a concrete way of comparing the two methods.

4 Intellectual merit

We've found that recommendation systems in general give suggestions that are not always unexpected or illuminating, so we hope to find that the mapper based clustering will reveal some heuristic for evaluating the similarities between users' preferences that is non-trivial.

5 Expected Outcome

We expect to find that the clustering from mapper and from *k*-means give different results, and that these results can be used to make better recommendations in the best case.