

# Using Mapper in a Recommendation System

Christopher James Brooks  
Gurupragaash Annasamy Mani

March 27, 2017

## 1 Project Overview

We will use Mapper [2] on large datasets to see if we can find any interesting clustering that differs from the usual  $k$ -means in ways that depend on the shape of the data.

## 2 Technical Details

We are going to analyze a large collection of Netflix users' movie ratings [1] 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. We can do this in two ways.

- Sum of difference between their movie ratings.
- Sort the ratings and find how many swaps are needed to make both the lists identical

We also need a filter function mapping users to  $\mathbb{R}^1$  or  $\mathbb{R}^2$ . We will be using L-infinity centrality as our filter function, which will map the users to  $\mathbb{R}^1$ .

Finally, we will also use a more traditional clustering technique( $k$ -means) as a concrete way of comparing the two methods.

## 3 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.

## 4 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.

## References

- [1] Netflix prize dataset. [https://archive.org/details/nf\\_prize\\_dataset.tar](https://archive.org/details/nf_prize_dataset.tar). Accessed: 2017-03-27.
- [2] P. Y. Lum, G. Singh, A. Lehman, T. Ishkanov, M. Vejdemo-Johansson, M. Alagappan, J. Carlsson, and G. Carlsson. Extracting insights from the shape of complex data using topology. *Scientific Reports*, 3:1236 EP –, Feb 2013. Article.