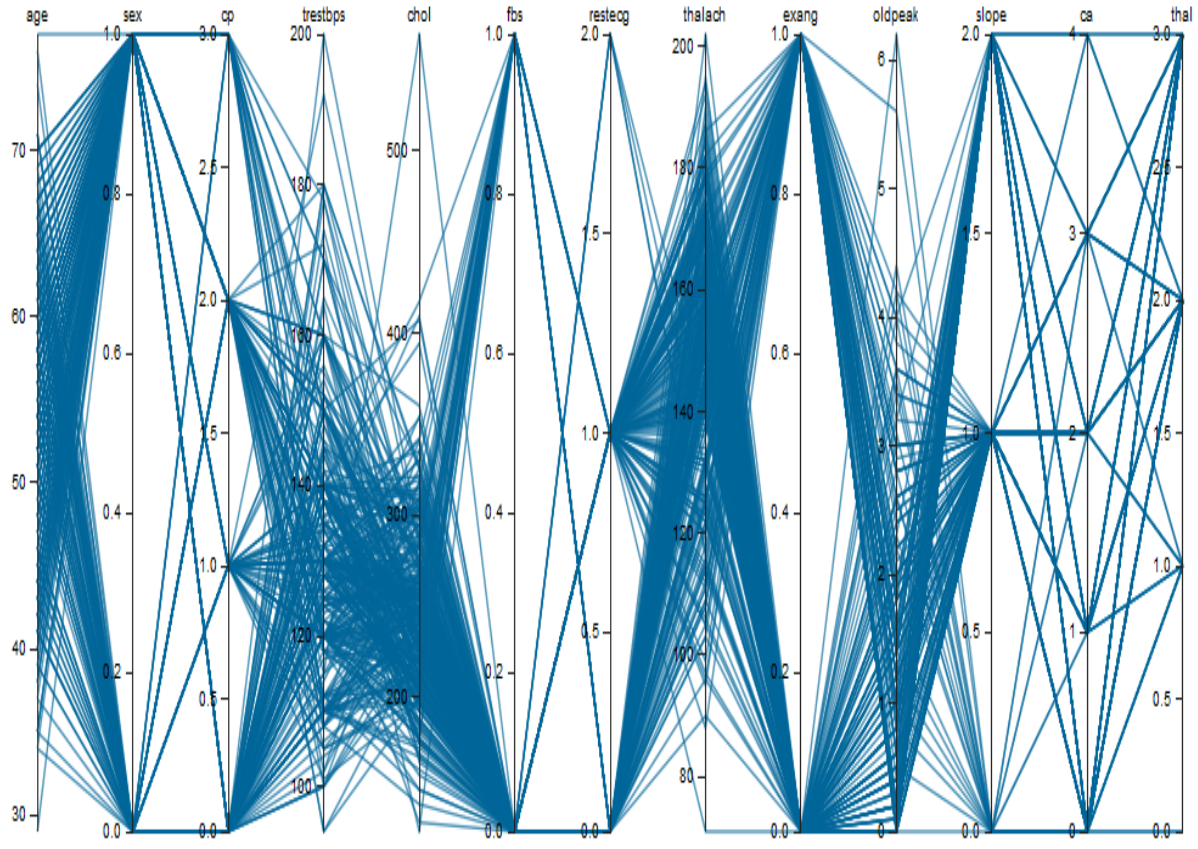
Qualitative Evaluation

The two selected datasets; Heart and Australian Credit Approval are plotted using Parallel Coordinate Plot. The reduced feature sets from Laplacian method and our approach are also plotted and the observations are compared.

Heart Dataset



Parallel Coordinate Plot with all Features in the Heart Dataset

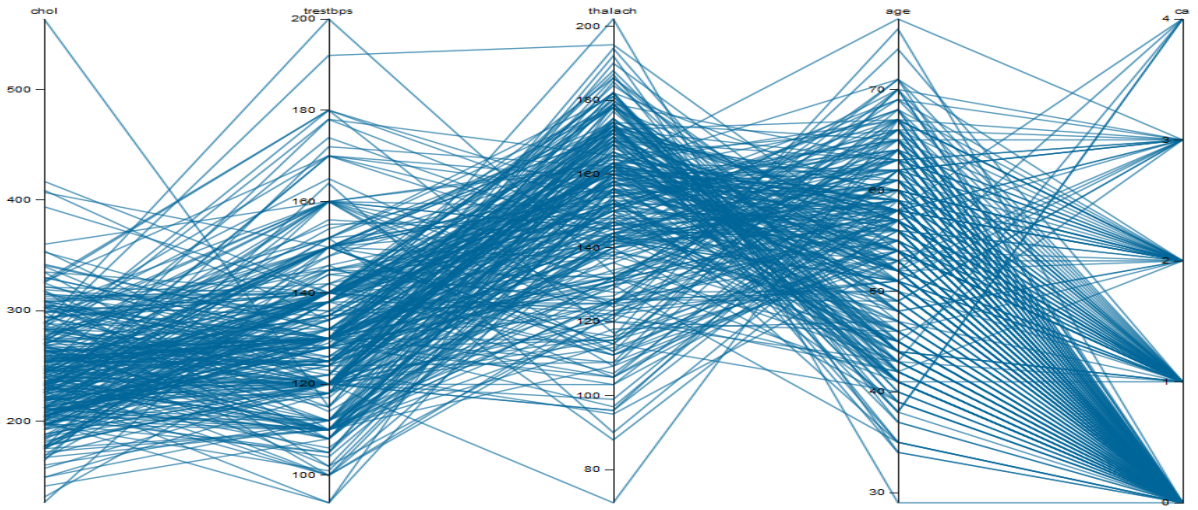


Figure: Parallel Coordinate Plot with Reduced Feature Set from the Laplacian Method in Heart dataset

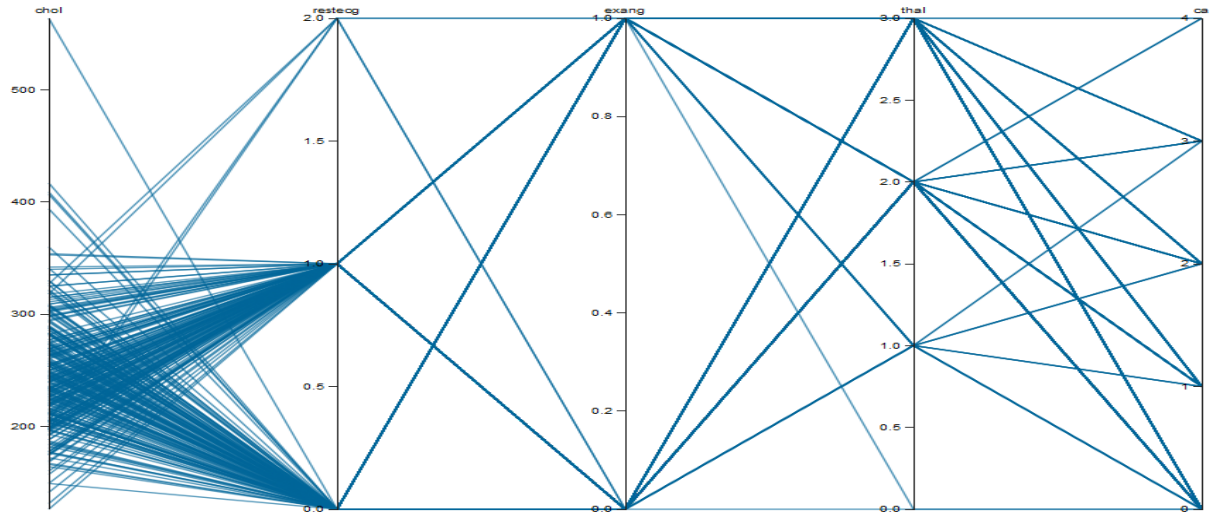


Figure: Parallel Coordinate Plot with Reduced Feature Set from the Proposed Approach in Heart dataset

Australian Credit Approval Dataset

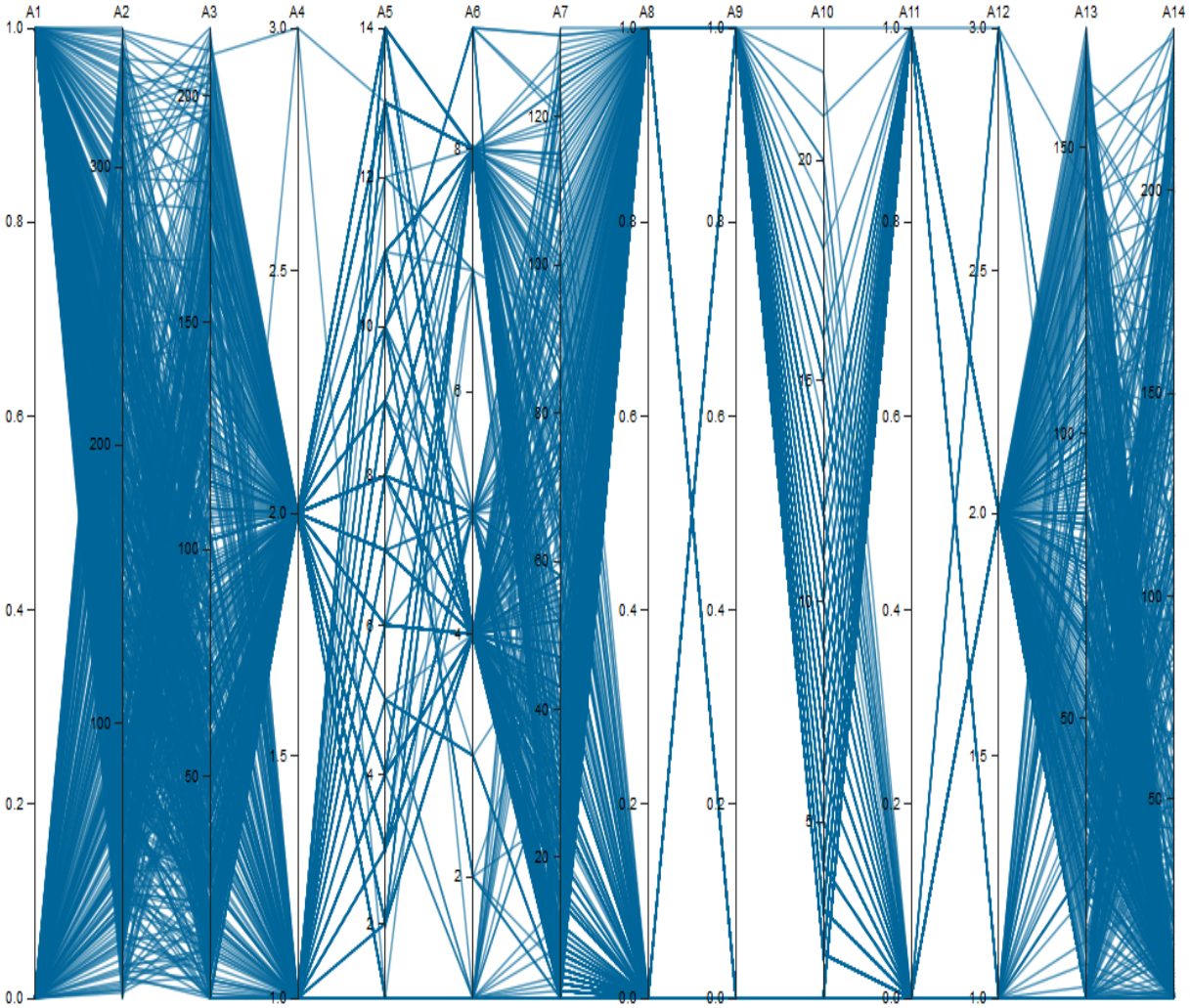


Figure: Parallel Coordinate Plot with all Features in the Aus. Credit Dataset

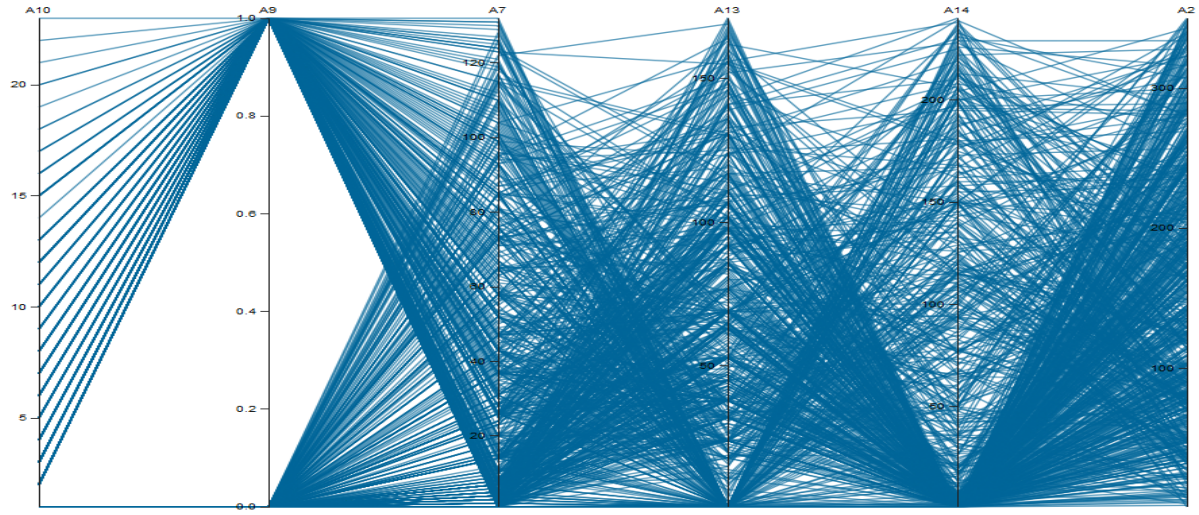


Figure : Parallel Coordinate Plot with Reduced Feature Set from the Laplacian Method in Aus. Credit dataset

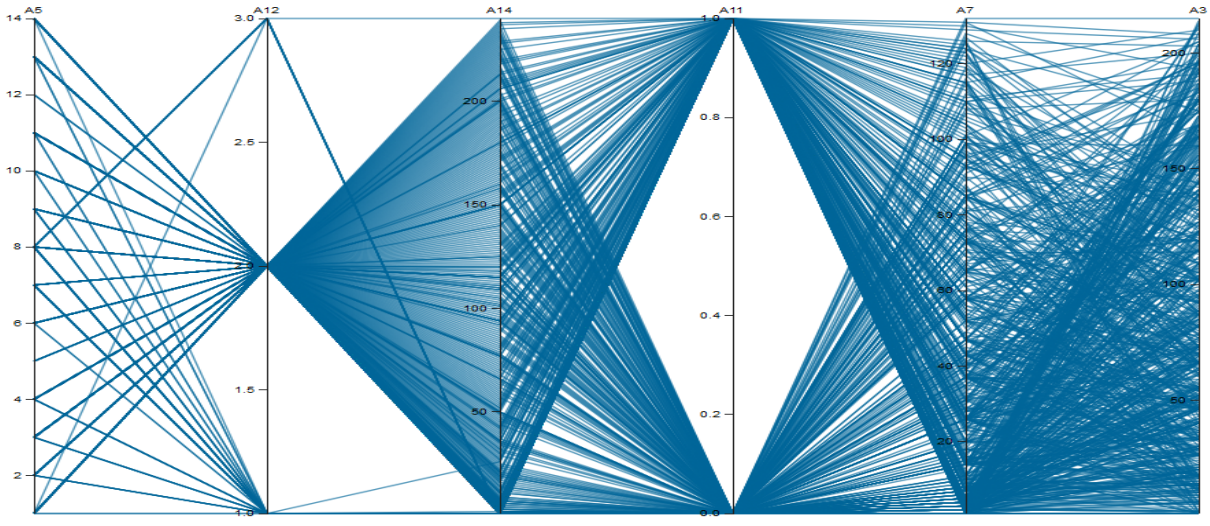


Figure: Parallel Coordinate Plot with Reduced Feature Set from the Proposed Approach in Aus. Credit dataset

Observations

- The Parallel Coordinate Plot with all the features from the Heart dataset is difficult to interpret because it has a lot of dimensions or axes. The clutter is mainly observed more when the feature has several factors. For example, chol, trestbps, thalach has a lot of clutter because all of them have more than 10 factors. The reduced feature set obtained from the Laplacian method is clear compared to the full feature set because it has 5 axes. The Parallel Coordinate Plot obtained from our proposed approach also has 5 axes, but it is much clear compared to the previous plots. The reason is that 4 out of the 5 selected features have less number of factors. The proposed approach works by segregating the dataset into categorical and numerical parts based on the factors. Since 9 out of 13 features in the original dataset have a limited number of factors, they are considered as categorical in our approach. The algorithm proceeds by determining the clusters and the Representative Features from each of the categorical and numerical feature groups. As the majority of the features are treated as categorical, the reduced feature set also has more categorical features making the Parallel Coordinate far less cluttered.

- The Parallel Coordinate Plot from all the features in the Australian Credit Approval dataset has a lot of clutter because of many axes and numerical features. The Laplacian method encodes the categorical features to numerical, and the Parallel Coordinate Plot has 6 features. Our proposed approach starts by segregating the dataset. Since the categorical features constitute almost half of the total features, there are 3 categorical and 3 numerical features in the reduced feature set. The clutter observed is a lot less than the previous plots.