Density-Based Clustering for Adaptive Density Variation

Since used dataset size is too big (around 3 million), clustering instances into small groups will decrease complexity and also improve efficiency of link prediction. This is because similar inputs will be in same cluster so that they are trained with themselves. The article Density-Based Clustering for Adaptive Density Variation is selected to give an idea how to cope with noise and outliers.

It lets density based clusters form with varying densities. To achieve that it uses object's neighbors information and calculate adaptive search range of each object. Therefore, not only defining object information but also density distribution of each cluster can be defined too. In addition, article provides a new metric that is to set better boundaries for clusters.