Interactive Correlation Clustering

Floris Geerts & Reuben Ndindi

Department of Mathematics and Computer Science
University of Antwerp

Correlation clustering is to partition a set of objects into clusters such that the number of false positives and negatives is minimised. In this paper, we combine correlation clustering and user interaction. More specifically, we allow the user to control the quality of the clustering by providing error bounds on the number of false positives and negatives. If no clusterings exist that satisfy these bounds, a set of edges is returned for user inspection such that the deletion or relabelling of these edges guarantees the existence of a clustering consistent with the error bounds. However, a user may reject the deletion or relabelling of certain edges and ask for an alternative set of edges to be provided. If no such set of edges exists, a minimal change to the error bounds should be provided, after which the interactive process continues. The focus of this paper is on the algorithmic challenges involved in returning a minimal set of edges to the user. More specifically, we formalise the Interactive Correlation Clustering problem and show that it is intractable. Therefore, we propose an approximation algorithm based on the well-known region growing technique. We experimentally validate the efficiency and accuracy of the approximation algorithm.