

Correlation Clustering under perturbations

Whereas clustering objective functions are NP-hard to optimize, meaningful instances in which we are interested have additional structure which allows for guaranteed polynomial time algorithms (see <http://arxiv.org/abs/1510.05336> for a high level overview) For instance [1] assumes that under multiplicative perturbations of the pairwise distances between points, the optimal clustering is modified only for a small fraction of points. But this work and many others focus on problems where the number of clusters k is fixed beforehand. Therefore it might be interesting to relax this assumption. One way is to replace pairwise distances by similarity/dissimilarity information. This gives rise to Correlation Clustering (CC), which was shown to be equivalent to weighted multicut problem [2]. Furthermore the stability notion of [1] was recently extended to multiway cut [3]. The goal would be to study and formalize this connection between CC and stability assumption and provide an implementation to verify the agreement between theoretical claims and practical performances.

1. Maria-Florina Balcan and Yingyu Liang. Clustering under perturbation resilience. In ICALP (1), pages 63–74, 2012. <http://www.cs.cmu.edu/~ninamf/papers/ICALP2012-clusteringunderperturbationresilience.pdf>
2. Demaine, E. D., Emanuel, D., Fiat, A., & Immorlica, N. (2006). Correlation clustering in general weighted graphs. Theoretical Computer Science, 361(2-3), 172–187. http://erikdemaine.org/papers/Clustering_TCS/paper.pdf
3. Konstantin Makarychev, Yury Makarychev, and Aravindan Vijayaraghavan. 2014. Bilu-linial stable instances of max cut and minimum multiway cut. In Proceedings of the Twenty-Fifth Annual ACM-SIAM Symposium on Discrete Algorithms (SODA '14). SIAM 890-906 http://users.eecs.northwestern.edu/~aravindv/stability_maxcut.pdf