Minimum Routing Cost Spanning Tree for node active learning

Since the complexity of learning algorithms in graphs grows with the number of edges, an important pre processing step is to sparsify the graph while preserving some of its key topological properties. For instance, one can look for a low stretch tree, that is a spanning tree such that the average tree distance between two nodes (the stretch) is minimized. This has been used in active node classification [1], where short paths reduce the probability of mistake. Now consider a semi supervised setting, where some node labels are given to the learner beforehand. How to design a tree that minimizes the overall distance between labeled nodes and the others? One could adapt general low stretch techniques [2] or look at similar solution from the network community [3]. A tempting heuristic is to start from an acceptable spanning tree and greedily exchange edges to improve an objective function [4]. The goal would be to implement one (or more) such approach and analyze their performances.

1. Cesa-Bianchi, N., Gentile, C., Vitale, F., & Zappella, G. (2012). A linear time active learning algorithm for link classification. In Advances in Neural Information Processing Systems 25 (NIPS 2012). http://papers.nips.cc/paper/4598-a-linear-time-active-learning-algorithm-for-link-classification

2. Abraham, I., & Neiman, O. (2012). Using petal-decompositions to build a low stretch spanning tree. In Proceedings of the 44th symposium on Theory of Computing - STOC '12 (p. 395). New York, New York, USA: ACM Press. http://www.cs.bgu.ac.il/~neimano/spanning-full1.pdf

3. McMahan, H. B., & Proskurowski, A. (2004). Multi-source spanning trees: algorithms for minimizing source eccentricities. Discrete Applied Mathematics, 137(2), 213–222. http://www.sciencedirect.com/science/article/pii/S0166218X03002622

4. Singh, A. (2008). A New Heuristic for the Minimum Routing Cost Spanning Tree Problem. In International Conference on Information Technology (pp. 9–13). IEEE. http://dx.doi.org/10.1109/ICIT.2008.16