

Fast and memory-efficient low rank approximation of massive graphs

Inderjit Dhillon (UT Austin)

September 1, 2011

In this talk, I will present a fast and accurate procedure called clustered low rank matrix approximation for massive graphs. The procedure involves a fast clustering of the graph followed by approximation of each cluster separately using existing methods, e.g. the singular value decomposition, or stochastic algorithms. The clusterwise approximations are then extended to approximate the entire graph. This approach has several benefits: (1) important structure of the graph is preserved due to the clustering; (2) accurate low rank approximations are achieved; (3) the procedure is efficient both in terms of computational speed and memory usage. A salient feature of the approximation scheme is that it attempts to preserve the sparsity of the input graph. The recently introduced stochastic algorithms can be incorporated into the clustered low rank approximation framework to get theoretical bounds for the approximation error. Finally, I will present experimental results of our algorithms on link prediction problems in massive social networks, and on an application in biology of linking genes to polygenic diseases.

This is joint work with Berkant Savas at UT Austin.