Hans De Sterck

Fast Multilevel Co-Clustering: Unraveling the Multilevel Overlapping Cluster Structure of Social Network Data

Department of Applied Mathematics
University of Waterloo
200 University Ave W
Waterloo
ON N2L 3G1
Canada
hdesterck@uwaterloo.ca
Haifeng Xu
Geoffrey Sanders

There are clear indications that many online social networks contain multilevel overlapping cluster structure, but it is difficult to unravel this structure using existing methods. We propose fast algorithms inspired on algebraic multigrid for finding the multilevel overlapping co-cluster structure of feature matrices that encode social network relations. Starting from the weighted bipartite graph structure of the feature matrix, the algorithms use graph agglomeration procedures to recursively coarsen the bipartite graphs that represent the relations between the co-clusters on increasingly coarser levels. New fast heuristic coarsening routines are described that circumvent the bottleneck of all-to-all similarity computations by exploiting measures of direct connection strength between row and column variables in the feature matrix. For traditional (single-level) co-clustering problems of gene expression data, our algorithms compare favourably to current methods in terms of speed, scalability and accuracy. The scalability of our methods is illustrated on hierarchical problems with input size up to tens of millions of data elements, where our fastest method needs about a second and produces more accurate results than a leading existing (single-level) method that requires more than 1,000 seconds. Our approach successfully uncovers multilevel overlapping cluster structure in a data set that relates LinkedIn users to their skills and expertise.