

IMPROVED CLUSTERING ALGORITHMS FOR THE BIPARTITE STOCHASTIC BLOCK MODEL

SUZANNE SIGALLA

Abstract: Consider a Bipartite Stochastic Block Model (BSBM) on vertex sets V_1 and V_2 . We investigate sufficient conditions of exact and almost full recovery for clustering over V_1 using polynomial-time algorithms, in the regime where the cardinalities satisfy $|V_1| \ll |V_2|$. We improve upon the known conditions of almost full recovery for spectral clustering algorithms in BSBM. Furthermore, we propose a new computationally simple and fast procedure achieving exact recovery under milder conditions than the state of the art. This procedure is a variant of Lloyd's iterations initialized with a well-chosen spectral algorithm leading to what we expect to be optimal conditions for exact recovery in this model. The latter fact is further supported by showing that a supervised oracle procedure requires similar conditions to achieve exact recovery. The key elements of the proof techniques are different from classical community detection tools on random graphs. Numerical studies confirm our theory, and show that the suggested algorithm is both very fast and achieves similar performance as the supervised oracle. Finally, using the connection between planted satisfiability problems and the BSBM, we improve upon the sufficient number of clauses to completely recover the planted assignment.