

Approximate Spectral Clustering via Randomized Sketching

Christos Boutsidis, Yahoo! Labs, New York

Spectral clustering is arguably one of the most important algorithms in data mining and machine intelligence; however, its computational complexity makes it a challenge to use it for large scale data analysis. Recently, several approximation algorithms for spectral clustering have been developed in order to alleviate the relevant costs, but theoretical results are lacking. In this paper, we present a novel approximation algorithm for spectral clustering with strong theoretical evidence of its performance. Our algorithm is based on approximating the eigenvectors of the Laplacian matrix using a randomized subspace iteration process, which might be of independent interest.