1158-62-348 **Guangliang Chen*** (guangliang.chen@sjsu.edu), San Jose, CA 95192. *All data are "documents": A scalable spectral clustering framework based on landmark points and cosine similarity.*

We present a unified scalable computing framework for various versions of spectral clustering, such as the Ng-Jordan-Weiss algorithm (NIPS '01), Normalized Cut (Shi and Malik, 2000), and Diffusion Maps (Coifman and Lafon, 2006). We first consider the special setting of cosine similarity for clustering sparse data (e.g., documents under the bag-of-words model) or data with at most a few hundred dimensions (e.g., small images). We show that in such cases, spectral clustering can be implemented solely based on three kinds of efficient matrix operations on the data matrix – elementwise manipulation, matrix-vector multiplication and low-rank SVD. Next, for any given similarity (and for any kind of data), we introduce a landmark-based technique to convert the data (and the selected landmarks) into a "document-term" matrix and then apply the scalable implementation of spectral clustering with cosine similarity to cluster them. Lastly, we demonstrate the performance of our proposed algorithm by comparing it with a few existing methods on several benchmark data sets. (Received March 03, 2020)