Robust Spectral Clustering for Noisy Data
Modeling Sparse Corruptions Improves Latent Embeddings

Aleksandar Bojchevski, Yves Matkovic, and Stephan Günnemann

Technical University of Munich, Munich, Germany
Department of Informatics and Institute for Advanced Study
{a.bojchevski,matkovic,guennemann}@in.tum.de

Abstract. Spectral clustering is one of the most prominent clustering approaches, with a multitude of successful applications from computer vision to network analysis. Since spectral clustering relies on a similarity graph only, it is applicable in many domains. However, one of the biggest limitations of spectral clustering is its sensitivity to noisy input data.
In this work, we introduce a principle to robustify spectral clustering. We propose a sparse and latent decomposition of the similarity graph used in spectral clustering. Thus, instead of operating on the original graph, the spectral clustering will be performed on the latent representation. In our model, we jointly learn the spectral embedding as well as the latent decomposition – thus, enhancing the clustering performance overall. We provide algorithmic solutions for our model for all three established versions of spectral clustering using different Laplacian matrices. For our solutions we relate to principles such as Eigenvalue perturbation and the multidimensional Knapsack problem. In each case, the complexity of the overall method is linear in the number of edges. Our experimental analysis confirms the significant potential of our approach for robust spectral clustering, with up to 15 percentage points improvement in accuracy on real-world data compared to standard spectral clustering. Moreover, we propose two novel measures – *local purity* and *global separation* – which enable us to evaluate the intrinsic quality of an embedding without relying on a specific clustering technique.

References

1. Bojchevski, A., Matkovic, Y., Günnemann, S.: Robust Spectral Clustering for Noisy Data. ACM SIGKDD Conference on Knowledge Discovery and Data Mining, 2017