

SPECTRAL METHODS FOR CLUSTERING SIGNED AND DIRECTED NETWORKS AND
HETEROGENEOUS GROUP SYNCHRONIZATION

Mihai Cucuringu

University of Oxford, United Kingdom
mihai.cucuringu@stats.ox.ac.uk

Graph clustering problems typically arise in settings where there exists a discrepancy in the edge density within different parts of the graph. In this work, we consider several problem instances where the underlying cluster structure arises as a consequence of a signal present on the edges or on the nodes of the graph, and is not driven by edge density.

We first consider the problem of clustering in two important families of networks: signed and directed, both relatively less well explored compared to their unsigned and undirected counterparts. Both problems share an important common feature: they can be solved by exploiting the spectrum of certain graph Laplacian matrices or derivations thereof. In signed networks, the edge weights between the nodes may take either positive or negative values, encoding a measure of similarity or dissimilarity. We consider a generalized eigenvalue problem involving graph Laplacians, and provide performance guarantees under the setting of a Signed Stochastic Block Model, along with regularized versions to handle very sparse graphs (below the connectivity threshold), a regime where standard spectral methods are known to underperform. We also propose a spectral clustering algorithm for directed graphs based on a complex-valued representation of the adjacency matrix, which is able to capture the underlying cluster structures, for which the information encoded in the direction of the edges is crucial. We evaluate the proposed algorithm in terms of a cut flow imbalance-based objective function, which, for a pair of given clusters, it captures the propensity of the edges to flow in a given direction. We analyze its theoretical performance on a Directed Stochastic Block Model for digraphs in which the cluster-structure is given not only by variations in edge densities, but also by the direction of the edges.

Finally, we discuss an extension of the classical angular synchronization problem that aims to recover unknown angles $\theta_1, \dots, \theta_n$ from a collection of noisy pairwise measurements of the form $(\theta_i - \theta_j) \bmod 2\pi$, for each $\{i, j\} \in E$. We consider a generalization to the heterogeneous setting where there exist k unknown groups of angles, and the measurement graph has an unknown edge-disjoint decomposition $G = G_1 \cup G_2 \dots \cup G_k$, where the G_i 's denote the subgraphs of noisy edge measurements corresponding to each group. We propose a probabilistic generative model for this problem, along with a spectral algorithm for which we provide a detailed theoretical analysis in terms of robustness against sampling sparsity and noise.