# An Iterative Spectral Clustering Algorithm for Directed Networks

James Martin, Tim Rogers, Luca Zanetti

University of Bath
{jlm80,ma3tcr, lz2040}@bath.ac.uk

The goal of clustering is to partition data points into $k$ groups such that points in the same group share similar properties, while points in different groups are dissimilar. In an undirected network, this involves dividing the graph into regions such that nodes in the same cluster are, on average, better connected than nodes from different clusters. When looking at directed networks, however, the notion of similarity depends on the varying directions of edges in addition to the edge densities.

Given a digraph $G = (V, E)$, the goal of our algorithm is to partition $V$ into disjoint subsets $S_1, \ldots, S_k$ such that, for all $i$ and $j$, the directed edges between $S_i$ and $S_j$ typically start and end in the same cluster. Alternatively, this can be seen as embedding $G$ into a smaller digraph of $k$ nodes where each node corresponds to a cluster and node $i$ has a directed edge towards node $j$ if most of the edges between $S_i$ and $S_j$ are directed from $S_i$ to $S_j$. We call this reduced graph the *meta-graph* w.r.t. the clustering $\mathcal{S} = \{S_1, \ldots, S_k\}$.

A popular approach to partition undirected graphs is spectral clustering, von Luxburg [Lux], which uses the eigenvectors of a matrix representation of the graph to embed the nodes into a low dimensional Euclidean space. However, standard representations of directed graphs, such as their adjacency matrix, are not Hermitian and therefore not diagonalisable, making it unclear how to apply spectral techniques. For this reason, Cucuringu et al. [CLSZ] recently introduced the Hermitian complex-valued representation $(A - A^{\mathrm{T}}) \cdot i$, where $A$ is the adjacency matrix of the digraph.

While this approach often produces good experimental results, it has a notable shortcoming: by studying the quadratic forms associated with this Hermitian matrix, we observe that the top eigenvectors of $(A - A^{\mathrm{T}}) \cdot i$ intrinsically embed the input graph $G$ into a meta-graph that always corresponds to a directed 4-cycle.

Our approach, instead, aims to embed the input graph $G$ into the meta-graph that best represents the underlying cluster-structure of $G$. For this reason, for any clustering $\mathcal{S}$, we define a matrix $H_{\mathcal{S}}$ whose quadratic forms tend to penalise edges in $G$ that do not follow the direction dictated by the meta-graph associated with $\mathcal{S}$. We develop an iterative algorithm that works as follows: starting from an arbitrary clustering $\mathcal{S}_0$, for any iteration $t = 1, \ldots, T$, we apply spectral clustering on $H_{\mathcal{S}_{t-1}}$ and recover a new clustering $\mathcal{S}_t$. Finally, we output the *best* clustering $\mathcal{S}_{t^\star}$ according to a relevant metric. The intuition behind our algorithm is that we want to explore in an efficient way the space of all possible meta-graphs and find the one that best fits the input graph.

The performance of our algorithm is evaluated experimentally using synthetic data generated from the directed stochastic block model (DSBM), which is a generalisation of the standard stochastic block model to digraphs. The experiments show our algorithm yields a clear improvement compared to previous algorithms for a wide range of parameters of the DSBM (see Figure 1 for an example).

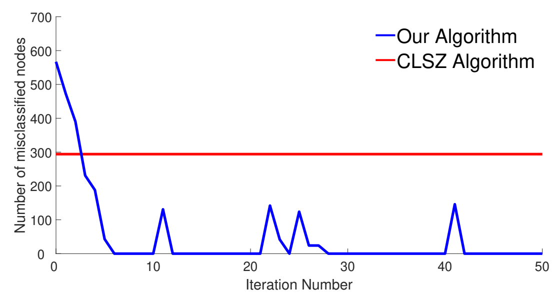


**Figure 1: Comparison of the misclassification error on a graph of 700 nodes generated from a DSBM with 7 clusters between the proposed algorithm (Our Algorithm) over 50 iterations and spectral clustering on $(A - A^{\mathrm{T}}) \cdot i$ (CLSZ Algorithm).**

To showcase the performances of our algorithm on real-world data, we explore the online competitive deckbuilding card game *Hearthstone*. In this two player game, each opponent chooses a hero and selects a deck of cards containing certain cards exclusive to that hero. The objective is to tactically play cards to defeat your opponent. Due to the heroes having exclusive access to certain cards, heroes typically require different strategies that work better with their cards. By using nodes to represent decks and edges to represent the win-rate between decks, our algorithm finds a clustering which groups together decks that achieve similar win-rates against other groups of decks. We observe that the clustering found by the algorithm tends to group together decks that share their corresponding hero. This suggests that, depending on the choice of the hero, one player might have an underlying advantage before the match even begins.

## References

Cucuringu, M., Li, H., Sun, H., & Zanetti, L. (2020). Hermitian matrices for clustering directed graphs: insights and applications. In *AISTATS'20*.

Von Luxburg, U. (2007). A tutorial on spectral clustering. *Statistics and computing*, *17*(4), 395–416.