2022 International Symposium on Nonlinear Theory and Its Applications,
NOLTA2022, Virtual, December 12-15, 2022

Spectral clustering of directed and time-evolving graphs using Koopman operator theory

Stefan Klus[†] and Nataša Djurdjevac Conrad[‡]

[†] School of Mathematical & Computer Sciences, Heriot–Watt University, UK
[‡] Zuse Institute Berlin, Germany
Email: s.klus@hw.ac.uk, natasa.conrad@zib.de

Abstract—Transport networks, electrical grids, and computer networks such as the internet, but also gene regulatory networks, neural networks, and social networks can be represented as directed or undirected graphs by abstracting individual components or entities as nodes and relationships between them as edges. In order to understand such complex networked systems, it is essential to identify community structures or clusters, i.e., sets of nodes that share similar properties. A popular and well-established approach to detect community structures in undirected graphs is spectral clustering. Detecting clusters in directed and time-varying graphs, however, remains a challenging problem. We extend spectral clustering algorithms to directed and time-evolving graphs using transfer operator theory, which is often used to study complex dynamical systems such as molecular dynamics and fluid dynamics problems.

1. Introduction

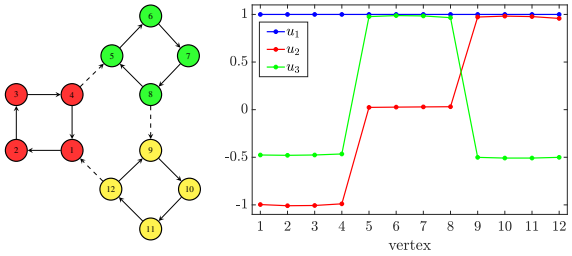
We derive spectral clustering algorithms by applying data-driven methods to estimate transfer operators to random walk data. A detailed description and analysis including numerical results for various benchmark problems can be found in [1].

2. Spectral clustering for directed graphs

Given a graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$, where $\mathcal{V} = \{v_1, \ldots, v_n\}$ is the set of vertices and $\mathcal{E} \subseteq \mathcal{V} \times \mathcal{V}$ the set of edges, and the associated weighted adjacency matrix $A = (a_{ij})_{i,j=1}^n$, the *degree matrix* is defined by $D_\circ = \operatorname{diag}(\circ(v_1), \ldots, \circ(v_n))$, with $\circ(v_i) = \sum_{j=1}^n a_{ij}$, and the *random-walk Laplacian* by $L_{\mathrm{rw}} = I - D_\circ^{-1} A$. The row-stochastic transition probability matrix $P := D_\circ^{-1} A$ can be used to generate random walks on the graph. We define transfer operators on graphs and show that the matrix P can be regarded as the Koopman operator associated with the random walk process. Similarly, the Perron–Frobenius operator on the graph is given by $P^\top$. For the analysis of time-inhomogeneous systems, typically a generalized operator that is related to the forward-backward dynamics, is utilized. For a random-walk process on a graph, the matrix representation of the operator is given by $Q := P D_\nu^{-1} P^\top$, with $D_\nu = \operatorname{diag}(\nu(v_1), \ldots, \nu(v_n))$ and $\nu(v_j) = \sum_{\ell=1}^n p_{\ell j}$. We call the matrix $L_{\mathrm{fb}} := I - Q$ the forward-backward Laplacian.

Spectral clustering algorithm for directed graphs:

1. Generate $L_{\mathrm{fb}} = I - Q$.
2. Compute the k smallest eigenvalues λ_ℓ and associated eigenvectors u_ℓ of L_{fb}.
3. Define $U = [u_1, \ldots, u_k] \in \mathbb{R}^{n \times k}$ and let r_i denote the ith row of U.
4. Cluster the points $\{r_i\}_{i=1}^n$ using, e.g., k-means.



The number of clusters k is typically chosen in such a way that there is a spectral gap between λ_k and λ_{k+1}. In order to illustrate the clustering approach, we construct a graph comprising three unidirectionally coupled clusters. Here, the weight of the solid edges is 1 and the weight of the dashed edges 0.01. The forward-backward Laplacian has three dominant eigenvalues, which indicates that there are three clusters. The corresponding eigenvectors are almost constant within the clusters, with sharp transitions in between. Applying k-means results in the clustering shown above.

3. Conclusion

The proposed algorithm can be easily extended to time-evolving graphs using either random walk data or products of transition probability matrices [1]. The resulting time-evolving clusters can then be regarded as *coherent sets*, which play an important role in the analysis of transport and mixing processes in fluid flows. Random walkers starting within the same cluster will—on average—remain in close proximity over long time scales.

References

[1] S. Klus and N. Djurdjevac Conrad. Koopman-based spectral clustering of directed and time-evolving graphs, 2022. doi:10.48550/ARXIV.2204.02951.

ORCID iDs Stefan Klus: 0000-0002-9672-3806, Nataša Djurdjevac Conrad: 0000-0001-8422-4930