NONBACKTRACKING SPECTRAL CLUSTERING OF NONUNIFORM HYPERGRAPHS

Philip Chodrow, Nicole Eikmeier, and Jamie Haddock

Spectral methods offer a tractable, global framework for clustering via eigenvector computations. Graph data—in which entities interact in pairs—are easily represented by matrices and therefore amenable to spectral clustering. Hypergraph data—in which entities interact in sets of arbitrary sizes—poses challenges for matrix representations and therefore for spectral methods. These challenges are especially pronounced in *nonuniform* hypergraphs, which contain edges of multiple sizes simultaneously. Many data sets of interest are nonuniform, and the development of performant spectral methods for such data sets is therefore of considerable interest.

We propose spectral clustering techniques for nonuniform hypergraphs based on the hypergraph nonbacktracking operator of Storm [2006]. We prove an Ihara-Bass theorem for this operator, extending results for the uniform case by Angelini et al. [2015]. We show through analytic calculations and computational experiments, however, that the eigenstructure of this operator makes it of limited use for clustering nonuniform hypergraphs generated from stochastic blockmodels. This is especially true when edges of different sizes carry differing information about the latent cluster structure.

We therefore propose an alternating algorithm—Nonbacktracking Belief-Propagation Spectral Clustering, or NBPSC—using a spectral subroutine derived by linearizing belief-propagation. We offer proofs supporting NBPSC that both formalize and extend previous arguments by Angelini et al. [2015] and Krzakala et al. [2013]. Analyzing NBPSC, we prove several results about the eigenstructure of the linearized belief-propagation operator, including both an in-expectation description of the leading eigenpairs and a reduction formula that allows these eigenpairs to be computed from a simpler matrix. We pose conjectures about fundamental detectability limitations of both spectral methods and more general methods in recovery problems for hypergraph stochastic blockmodels. These conjectures extend recently proven detectability thresholds for graph stochastic blockmodels [Mossel et al., 2018, Massoulié, 2014]. Experiments on synthetic data support our conjectures (see Fig. 1). We then study several empirical hypergraph data sets, including school contact networks and co-tagging networks of mathematical concepts on StackExchange. We find favorable performance of NBPSC when compared to graph-based nonbacktracking spectral methods with clique-expansions. Our results underscore the distinctive challenges and benefits of explicit, dedicated hypergraph methods for clustering problems.

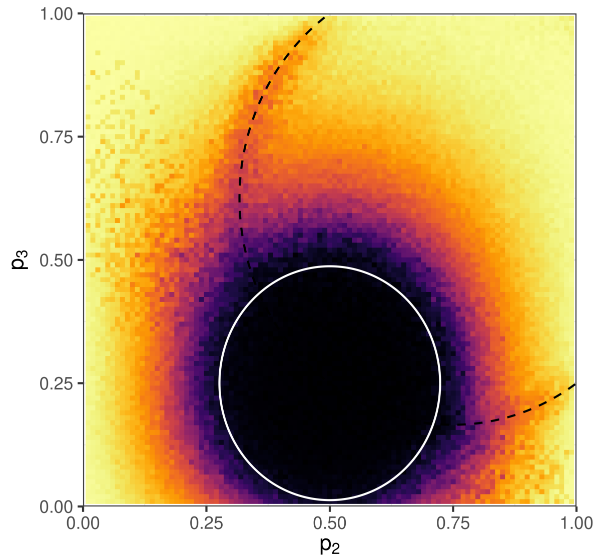

Figure 1: Adjusted Rand Index (ARI) of output clustering from NBPSC in a series of experiments on synthetic data generated from a blockmodel containing hyperedges of size 2 and 3. The white ellipse gives our conjectured detectability threshold.

References

M. C. Angelini, F. Caltagirone, F. Krzakala, and L. Zdeborová. Spectral detection on sparse hypergraphs. In *2015 53rd Annual Allerton Conference on Communication, Control, and Computing (Allerton)*, pages 66–73. IEEE, 2015.

P. Chodrow, N. Eikmeier, and J. Haddock. Nonbacktracking spectral clustering of nonuniform hypergraphs. *arXiv:2204.13586*, 2022.

F. Krzakala, C. Moore, E. Mossel, J. Neeman, A. Sly, L. Zdeborová, and P. Zhang. Spectral redemption in clustering sparse networks. *Proceedings of the National Academy of Sciences*, 110(52):20935–20940, 2013.

L. Massoulié. Community detection thresholds and the weak Ramanujan property. In *Proceedings of the Forty-Sixth Annual ACM Symposium on Theory of Computing*, pages 694–703, 2014.

E. Mossel, J. Neeman, and A. Sly. A proof of the block model threshold conjecture. *Combinatorica. An International Journal on Combinatorics and the Theory of Computing*, 38(3):665–708, 2018.

C. K. Storm. The Zeta Function of a Hypergraph. *The Electronic Journal of Combinatorics*, 13(1):R84, Oct. 2006. ISSN 1077-8926. doi: 10.37236/1110.