The Dominant Spectral Subspace for Nodal Decomposition of a Network Interconnecting Tight-Knit Communities

Dimitris Floros   Nikos Pitsianis   Xiaobai Sun

We introduce a new connection—established through both theoretical analysis and empirical investigation—between graph spectral analysis and community detection in networks, i.e., graph clustering. The new finding has three key components: (1) A pronounced, dominant gap in the Laplacian spectrum of a graph is indicative of the presence of tight-knit community clusters. (2) The underlying cluster structure emerges when the graph is embedded into a low-dimensional invariant subspace associated with the dominant gap. In this space, the cluster subgraphs are identifiable with the nodal domains and can be robustly recognized, without supervision. (3) The dominant gap and the dominant subspace are spectral characteristics of the graph cluster structure, as the Fiedler value and vector to the graph connectivity. In comparison, a nodal partition by an eigenvector associated with a Laplacian eigenvalue below (or above) the gap merges (or splits) the clusters, which is a known resolution problem, or worse, it blends the clusters into false groups.

According to the graph-Laplacian version of Courant's nodal domain theorem, with the eigenvalues in non-descending order, the j-th eigenvector f_j partitions the graph into at most j nodal domains. A nodal domain containing a vertex v is defined as the largest connected component that can be reached from v without a sign change in f_j (without zero crossing). Below, we briefly review existing and ongoing related work—both theoretical and applied—and highlight several issues that warrant greater attention and deeper investigation.

(a) The Fiedler vector f_2 cuts a graph into exactly two nodal domains. The two divided subgraphs have the weakest interconnection in a specific sense. This foundational theory was initially used for graph partitioning in parallel computation via recursive bisection [5]. It later found applications in automatic image segmentation [7], and, more recently, for data partition at various stages of machine learning systems [9]. However, a naive use of the Fiedler vector can be unreliable for graph clustering, resulting in inadequate resolution or false classification [2].

(b) Some recent studies focus on sharpening the upper and/or lower bounds on the number of

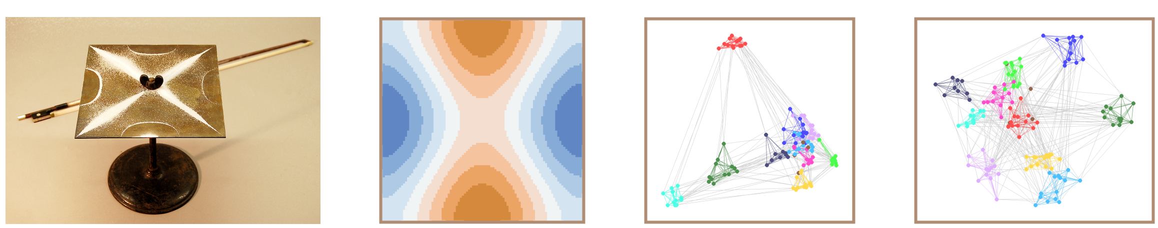

Figure 1: Left to right: (1) The nodal pattern of white sand on a Chladni plate at a resonance frequency (`https://americanhistory.si.edu/science/chladni.htm`). (2) The nodal pattern of a 2D grid graph by eigenvector f_5. (3) The spectral embedding of `NFL-2023`, a real-world network of 132 football teams in 10 conferences (clusters), and 4 independent teams, in the space by the Fiedler vector f_2 and the next eigenvector f_3. (4) The spectral embedding of `NFL-2023` by the eigenvectors f_9 and f_{10} associated with the dominant spectral gap. The conference configuration becomes discernible.

nodal domains for any eigenvector f_j under certain conditions [1, 8]. We address a different, complementary question: which eigenvector or set of eigenvectors produces a nodal partition that best aligns with the community structure?

(c) Vertex-to-vector encoding is an indispensable upstream task in machine learning [4]. Spectral graph embedding for vector encoding typically uses a dimension-reduced spectral space at the lower end of the Laplacian spectrum to preserve near-neighbor connectivity [6]. Unfortunately, this particular subspace selection poses challenges for the downstream task of data clustering, which aims to separate adjacent neighbors into distinct functional or structural units for scientific analysis or group them by common attributes for recommendation systems [9]. The subspace associated with the dominant spectral gap is more suitable for community detection. In this space, intra-cluster neighbors are kept close together (attraction), inter-cluster neighbors are pushed apart (repulsion), as shown in Fig. 1. The spectral index for the gap location increases with the number of intrinsic communities; a lower-indexed eigenvector lacks the capacity for sufficient cluster differentiation.

(d) Theoretical studies of spectral graph analysis have shown success with regular graphs (both random and non-random) [3], where the adjacency matrix, combinatorial Laplacian, and normalized Laplacian share the same eigenvectors and give the same nodal partitions. Yet, perturbation theory is limited in extending such analysis to graphs with varying degree distributions. To address this, we construct and analyze two sets of ideally parameterized networks with homogeneous clusters interconnected through simple and elementary typologies. The graphs are topologically defined in one set and probabilistically characterized in the other set. They are not regular, except in degenerate cases. They serve as structural reference graphs for studying a broader class of graphs with heterogeneous clusters that fall within the scope of perturbation analysis. Together, their spectral properties offer new insights into community structures in real-world networks.

We present both theoretical and empirical results from our investigations on synthetic graphs and real-world networks. We conclude with comments on remaining questions. The bibliography is substantially truncated to meet the space limit.

References

[1] S. Chanillo, A. Logunov, E. Malinnikova, and D. Mangoubi. Local version of Courant's nodal domain theorem. *Journal of Differential Geometry*, 126(1), Jan. 2024.

[2] D. Floros, N. Pitsianis, and X. Sun. The Fiedler connection to the parametrized modularity optimization for community detection, Oct. 2023. arXiv:2310.14359.

[3] S. Ganguly, T. McKenzie, S. Mohanty, and N. Srivastava. Many nodal domains in random regular graphs. *Communications in Mathematical Physics*, 401(2):1291–1309, July 2023.

[4] T. Mikolov, I. Sutskever, K. Chen, G. S. Corrado, and J. Dean. Distributed representations of words and phrases and their compositionality. In *Advances in Neural Information Processing Systems*, volume 26, 2013.

[5] A. Pothen, H. D. Simon, and K.-P. Liou. Partitioning sparse matrices with eigenvectors of graphs. *SIAM Journal on Matrix Analysis and Applications*, 11(3):430–452, 1990.

[6] S. T. Roweis and L. K. Saul. Nonlinear dimensionality reduction by locally linear embedding. *Science*, 290(5500):2323–2326, 2000.

[7] J. Shi and J. Malik. Normalized cuts and image segmentation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 22(8):888–905, 2000.

[8] J. C. Urschel. Nodal decompositions of graphs. *Linear Algebra Appl.*, 539:60–71, Feb. 2018.

[9] S. Zhang, L. Yao, A. Sun, and Y. Tay. Deep learning based recommender system: A survey and new perspectives. *ACM Computing Surveys*, 52(1):1–38, Jan. 2020.