Community detection with the Bethe-Hessian

Ludovic Stephan LUDOVIC.STEPHAN@ENSAI.FR
Univ Rennes, Ensai, CNRS, CREST – UMR 9194 F-35000 Rennes, France

Yizhe Zhu YIZHEZHU@USC.EDU
Department of Mathematics, University of Southern California

Editors: Nika Haghtalab and Ankur Moitra

Abstract

We consider the problem of community detection in very sparse networks, for which a spectral clustering algorithm based on the usual adjacency matrix fails to recover information about the communities. A popular alternative is the use of the non-backtracking matrix, which succeeds down to the conjectured algorithmic limit – also known as the Kesten-Stigum (KS) threshold. However, this matrix is non-Hermitian and its size grows with the average degree of the graph, which limits its applicability in practice.

In contrast, the Bethe-Hessian matrix, introduced by Saade, Krzakala, and Zdeborová (Saade et al., 2014), is a Hermitian matrix whose size does not depend on the density of the graph. Numerical evidence in Saade et al. (2014); Dall'Amico et al. (2019, 2021), as well as heuristic arguments, led to the conjecture that the eigenvectors associated to the negative eigenvalues of this matrix correlate with the underlying community structure of the graph. Previously, rigorous confirmation of this conjecture was scarce, with only Mohanty et al. (2024) achieving partial results on counting the negative eigenvalues of the Bethe-Hessian.

We provide the first rigorous analysis of the Bethe-Hessian spectral method in the SBM under both the bounded expected degree and the growing degree regimes. Specifically, we demonstrate that:

(i) When the expected degree $d \geq 2$, the number of negative outliers of the Bethe-Hessian matrix can consistently estimate the number of blocks above the Kesten-Stigum threshold, thus confirming a conjecture from Saade et al. (2014) for $d \geq 2$.

(ii) For sufficiently large d, the eigenvectors associated to those negative outliers are correlated with the community structure, and can be exploited to achieve weak recovery.

(iii) As $d \to \infty$, the negative eigenvalues of the Bethe Hessian concentrate around deterministic values, and the associated eigenvectors become aligned with those of the expected adjacency matrix. As a result, a spectral method based on the negative outliers of the Bethe Hessian can achieve almost exact recovery (also called weak consistency) in the SBM.[1]

Keywords: Bethe-Hessian matrix, spectral clustering, Kesten-Stigum threshold, stochastic block model

1. Extended abstract. Full version appears as [arXiv:2411.02835v2]

References

Emmanuel Abbe. Community detection and stochastic block models: Recent developments. *Journal of Machine Learning Research*, 18(177):1–86, 2018.

Charles Bordenave, Marc Lelarge, and Laurent Massoulié. Nonbacktracking spectrum of random graphs: Community detection and nonregular ramanujan graphs. *Annals of probability*, 46(1): 1–71, 2018.

Charles Bordenave, Simon Coste, and Raj Rao Nadakuditi. Detection thresholds in very sparse matrix completion. *Foundations of Computational Mathematics*, 23(5):1619–1743, 2023.

Simon Coste and Yizhe Zhu. Eigenvalues of the non-backtracking operator detached from the bulk. *Random Matrices: Theory and Applications*, 10(03):2150028, 2021.

Lorenzo Dall'Amico, Romain Couillet, and Nicolas Tremblay. Revisiting the Bethe-Hessian: Improved community detection in sparse heterogeneous graphs. *Advances in neural information processing systems*, 32, 2019.

Lorenzo Dall'Amico, Romain Couillet, and Nicolas Tremblay. A unified framework for spectral clustering in sparse graphs. *Journal of Machine Learning Research*, 22(217):1–56, 2021.

Neil Hwang, Jiarui Xu, Shirshendu Chatterjee, and Sharmodeep Bhattacharyya. On the estimation of the number of communities for sparse networks. *Journal of the American Statistical Association*, pages 1–22, 2023.

Can M Le and Elizaveta Levina. Estimating the number of communities by spectral methods. *Electronic Journal of Statistics*, 16(1):3315–3342, 2022.

Can M Le, Elizaveta Levina, and Roman Vershynin. Concentration and regularization of random graphs. *Random Structures & Algorithms*, 51(3):538–561, 2017.

Sidhanth Mohanty, Prasad Raghavendra, and David X Wu. Robust recovery for stochastic block models, simplified and generalized. In *Proceedings of the 56th Annual ACM Symposium on Theory of Computing*, pages 367–374, 2024.

Alaa Saade, Florent Krzakala, and Lenka Zdeborová. Spectral clustering of graphs with the Bethe Hessian. *Advances in Neural Information Processing Systems*, 27, 2014.

Ludovic Stephan and Laurent Massoulié. Robustness of spectral methods for community detection. In *Conference on Learning Theory*, pages 2831–2860. PMLR, 2019.

Ludovic Stephan and Laurent Massoulié. Non-backtracking spectra of weighted inhomogeneous random graphs. *Mathematical Statistics and Learning*, 5(3):201–271, 2022.

Ludovic Stephan and Yizhe Zhu. Sparse random hypergraphs: Non-backtracking spectra and community detection. *Information and Inference: A Journal of the IMA*, 13(1):iaae004, 2024a.

Ludovic Stephan and Yizhe Zhu. A non-backtracking method for long matrix and tensor completion. In *The Thirty Seventh Annual Conference on Learning Theory*, pages 4636–4690. PMLR, 2024b.