Sharp thresholds in inference of planted subgraphs

Elchanan Mossel ELMOS@MIT.EDU
Department of Mathematics and IDSS, MIT

Jonathan Niles-Weed JNW@CIMS.NYU.EDU
Center for Data Science and Courant Institute of Mathematical Sciences, NYU

Youngtak Sohn YOUNGTAK@MIT.EDU
Department of Mathematics, MIT

Nike Sun NSUN@MIT.EDU
Department of Mathematics, MIT

Ilias Zadik IZADIK@MIT.EDU
Department of Mathematics, MIT

Editors: Gergely Neu and Lorenzo Rosasco

Abstract

We connect the study of phase transitions in high-dimensional statistical inference to the study of threshold phenomena in random graphs.

A major question in the study of the Erdős–Rényi random graph $G(n,p)$ is to understand the probability, as a function of p, that $G(n,p)$ contains a given subgraph $H = H_n$. This was studied for many specific examples of H, starting with classical work of Erdős and Rényi (1960). More recent work studies this question for general H, both in building a general theory of sharp versus coarse transitions (Friedgut and Bourgain 1999; Hatami, 2012) and in results on the location of the transition (Kahn and Kalai, 2007; Talagrand, 2010; Frankston, Kahn, Narayanan, Park, 2019; Park and Pham, 2022).

In inference problems, one often studies the optimal accuracy of inference as a function of the amount of noise. In a variety of sparse recovery problems, an "all-or-nothing (AoN) phenomenon" has been observed: Informally, as the amount of noise is gradually increased, at some critical threshold the inference problem undergoes a sharp jump from near-perfect recovery to near-zero accuracy (Gamarnik and Zadik, 2017; Reeves, Xu, Zadik, 2021). We can regard AoN as the natural inference analogue of the sharp threshold phenomenon in random graphs. In contrast with the general theory developed for sharp thresholds of random graph properties, the AoN phenomenon has only been studied so far in specific inference settings, and a general theory behind its appearance remains elusive.

In this paper we study the general problem of inferring a graph $H = H_n$ planted in an Erdős–Rényi random graph, thus naturally connecting the two lines of research mentioned above. We show that questions of AoN are closely connected to first moment thresholds, and to a generalization of the so-called Kahn–Kalai expectation threshold that scans over subgraphs of H of edge density at least q. In a variety of settings we characterize AoN, by showing that AoN occurs *if and only if* this "generalized expectation threshold" is roughly constant in q. Our proofs combine techniques from random graph theory and Bayesian inference.[1]

Keywords: planted subgraph model; all-or-nothing transition; Kahn–Kalai expectation threshold

1. Extended abstract. Full version appears as [arXiv:2302.14830, v1]

Acknowledgements

We acknowledge the support of Simons-NSF grant DMS-2031883 (E.M., Y.S., N.S., and I.Z.), the Vannevar Bush Faculty Fellowship ONR-N00014-20-1-2826 (E.M., Y.S., and I.Z.), the Simons Investigator Award 622132 (E.M.), the Sloan Research Fellowship (J.N.W.), and NSF CAREER grant DMS-1940092 (N.S.).

References

Dimitris Achlioptas and Amin Coja-Oghlan. Algorithmic barriers from phase transitions. In *2008 49th Annual IEEE Symposium on Foundations of Computer Science*, pages 793–802. IEEE, 2008.

Ryan Alweiss, Shachar Lovett, Kewen Wu, and Jiapeng Zhang. Improved bounds for the sunflower lemma. *Annals of Mathematics*, 194(3):795 – 815, 2021. doi: 10.4007/annals.2021.194.3.5. URL https://doi.org/10.4007/annals.2021.194.3.5.

Ery Arias-Castro and Nicolas Verzelen. Community detection in dense random networks. *The Annals of Statistics*, 42(3):940 – 969, 2014. doi: 10.1214/14-AOS1208.

Vivek Bagaria, Jian Ding, David Tse, Yihong Wu, and Jiaming Xu. Hidden hamiltonian cycle recovery via linear programming. *Operations research*, 68(1):53–70, 2020.

Jean Barbier, Florent Krzakala, Nicolas Macris, Léo Miolane, and Lenka Zdeborová. Optimal errors and phase transitions in high-dimensional generalized linear models. *Proceedings of the National Academy of Sciences*, 116(12):5451–5460, 2019.

Tolson Bell, Suchakree Chueluecha, and Lutz Warnke. Note on sunflowers. *Discrete Math.*, 344 (7):112367, 2021.

Béla Bollobás. Threshold functions for small subgraphs. *Math. Proc. Cambridge Philos. Soc.*, 90(2):197–206, 1981. ISSN 0305-0041. doi: 10.1017/S0305004100058655. URL https://doi-org.libproxy.mit.edu/10.1017/S0305004100058655.

Michael Chertkov, Lukas Kroc, F Krzakala, M Vergassola, and L Zdeborová. Inference in particle tracking experiments by passing messages between images. *Proceedings of the National Academy of Sciences*, 107(17):7663–7668, 2010.

Amin Coja-Oghlan, Oliver Gebhard, Max Hahn-Klimroth, Alexander S Wein, and Ilias Zadik. Statistical and computational phase transitions in group testing. In *Conference on Learning Theory*, pages 4764–4781. PMLR, 2022.

Jian Ding, Yihong Wu, Jiaming Xu, and Dana Yang. The planted matching problem: Sharp threshold and infinite-order phase transition, 2021.

P. Erdős and A. Rényi. On the evolution of random graphs. *Magyar Tud. Akad. Mat. Kutató Int. Közl.*, 5:17–61, 1960. ISSN 0541-9514.

Keith Frankston, Jeff Kahn, Bhargav Narayanan, and Jinyoung Park. Thresholds versus fractional expectation-thresholds. *Ann. of Math. (2)*, 194(2):475–495, 2021. ISSN 0003-486X. doi: 10.4007/annals.2021.194.2.2. URL https://doi-org.libproxy.mit.edu/10.4007/annals.2021.194.2.2.

Ehud Friedgut. Boolean functions with low average sensitivity depend on few coordinates. *Combinatorica*, 18(1):27–35, 1998. ISSN 0209-9683. doi: 10.1007/PL00009809. URL https://doi-org.libproxy.mit.edu/10.1007/PL00009809.

Ehud Friedgut. Sharp thresholds of graph properties, and the k-sat problem. *J. Amer. Math. Soc.*, 12 (4):1017–1054, 1999. ISSN 0894-0347. doi: 10.1090/S0894-0347-99-00305-7. URL https://doi-org.libproxy.mit.edu/10.1090/S0894-0347-99-00305-7. With an appendix by Jean Bourgain.

David Gamarnik and Ilias Zadik. Sparse high-dimensional linear regression. estimating squared error and a phase transition. *The Annals of Statistics*, 50(2):880–903, 2022.

Dongning Guo, Yihong Wu, Shlomo S Shitz, and Sergio Verdú. Estimation in gaussian noise: Properties of the minimum mean-square error. *IEEE Transactions on Information Theory*, 57(4): 2371–2385, 2011.

Hamed Hatami. A structure theorem for Boolean functions with small total influences. *Ann. of Math. (2)*, 176(1):509–533, 2012. ISSN 0003-486X. doi: 10.4007/annals.2012.176.1.9. URL https://doi-org.libproxy.mit.edu/10.4007/annals.2012.176.1.9.

Wasim Huleihel. Inferring hidden structures in random graphs. *IEEE Transactions on Signal and Information Processing over Networks*, 8:855–867, 2022.

Mark Jerrum. Large cliques elude the metropolis process. *Random Structures & Algorithms*, 3(4):347–359, 1992. doi: https://doi.org/10.1002/rsa.3240030402. URL https://onlinelibrary.wiley.com/doi/abs/10.1002/rsa.3240030402.

J Kahn, G Kalai, and N Linial. The influence of variables on Boolean functions. In *Proc. 29th FOCS*, pages 68–80, 1988.

Jeff Kahn and Gil Kalai. Thresholds and expectation thresholds. *Combin. Probab. Comput.*, 16 (3):495–502, 2007. ISSN 0963-5483. doi: 10.1017/S0963548307008474. URL https://doi-org.libproxy.mit.edu/10.1017/S0963548307008474.

Gil Kalai and Shmuel Safra. Perspectives from mathematics, computer science, and economics. *Computational complexity and statistical physics*, page 25, 2006.

A. D. Koršunov. Solution of a problem of P. Erdős and A. Rényi on Hamiltonian cycles in undirected graphs. *Dokl. Akad. Nauk SSSR*, 228(3):529–532, 1976. ISSN 0002-3264.

Florent Krzakała, Andrea Montanari, Federico Ricci-Tersenghi, Guilhem Semerjian, and Lenka Zdeborová. Gibbs states and the set of solutions of random constraint satisfaction problems. *Proceedings of the National Academy of Sciences*, 104(25):10318–10323, 2007.

Shrinivas Kudekar, Santhosh Kumar, Marco Mondelli, Henry D Pfister, Eren Şaşoğlu, and Rüdiger Urbanke. Reed-muller codes achieve capacity on erasure channels. In *Proceedings of the forty-eighth annual ACM symposium on Theory of Computing*, pages 658–669, 2016.

Clément Luneau, Nicolas Macris, and Jean Barbier. Information theoretic limits of learning a sparse rule. *Journal of Statistical Mechanics: Theory and Experiment*, 2022(4):044001, 2022.

Nicolas Macris, Cynthia Rush, et al. All-or-nothing statistical and computational phase transitions in sparse spiked matrix estimation. *Advances in Neural Information Processing Systems*, 33: 14915–14926, 2020.

Laurent Massoulié, Ludovic Stephan, and Don Towsley. Planting trees in graphs, and finding them back. In *Conference on Learning Theory*, pages 2341–2371. PMLR, 2019.

Mehrdad Moharrami, Cristopher Moore, and Jiaming Xu. The planted matching problem: phase transitions and exact results. *Ann. Appl. Probab.*, 31(6):2663–2720, 2021. ISSN 1050-5164. doi: 10.1214/20-aap1660. URL https://doi.org/10.1214/20-aap1660.

Elchanan Mossel, Jonathan Niles-Weed, Nike Sun, and Ilias Zadik. On the second kahn–kalai conjecture. *arXiv preprint arXiv:2209.03326*, 2022.

Jonathan Niles-Weed and Ilias Zadik. The all-or-nothing phenomenon in sparse tensor pca. In H. Larochelle, M. Ranzato, R. Hadsell, M. F. Balcan, and H. Lin, editors, *Advances in Neural Information Processing Systems*, volume 33, pages 17674–17684. Curran Associates, Inc., 2020. URL https://proceedings.neurips.cc/paper/2020/file/cd0b43eac0392accf3624b7372dec36e-Paper.pdf.

Jonathan Niles-Weed and Ilias Zadik. It was “all” for “nothing”: sharp phase transitions for noiseless discrete channels. In *Conference on Learning Theory*, pages 3546–3547. PMLR, 2021.

Hidetoshi Nishimori. *Statistical physics of spin glasses and information processing: an introduction*. Oxford University Press, 2001.

Jinyoung Park and Huy Tuan Pham. A proof of the Kahn–Kalai conjecture. arXiv:2203.17207, 2022.

L. Pósa. Hamiltonian circuits in random graphs. *Discrete Math.*, 14(4):359–364, 1976. ISSN 0012-365X. doi: 10.1016/0012-365X(76)90068-6. URL https://doi.org/10.1016/0012-365X(76)90068-6.

Anup Rao. Coding for sunflowers. *Discrete Anal.*, 2:1–8, 2020. doi: 10.19086/da. URL https://doi.org/10.19086/da.

Galen Reeves and Henry D Pfister. The replica-symmetric prediction for random linear estimation with gaussian matrices is exact. *IEEE Transactions on Information Theory*, 65(4):2252–2283, 2019.

Galen Reeves, Jiaming Xu, and Ilias Zadik. The all-or-nothing phenomenon in sparse linear regression. *Mathematical Statistics and Learning*, 3(3):259–313, 2021.

A Rucinski. Small subgraphs of random graphs—a survey. In *Random graphs*, volume 87, pages 283–303, 1987.

Andrzej Ruciński and Andrew Vince. Strongly balanced graphs and random graphs. *J. Graph Theory*, 10(2):251–264, 1986. ISSN 0364-9024. doi: 10.1002/jgt.3190100214. URL https://doi.org/10.1002/jgt.3190100214.

Guilhem Semerjian, Gabriele Sicuro, and Lenka Zdeborová. Recovery thresholds in the sparse planted matching problem. *Phys. Rev. E*, 102(2):022304, 18, 2020. ISSN 2470-0045. doi: 10.1103/physreve.102.022304. URL https://doi.org/10.1103/physreve.102.022304.

Manuel Stoeckl. Lecture notes on recent improvements for the sunflower lemma. Online post, https://mstoeckl.com/notes/research/sunflower_notes.html, 2022.

Michel Talagrand. Are many small sets explicitly small? In *Proc. 42nd STOC*, pages 13–35. ACM, New York, 2010.

Terence Tao. The sunflower lemma via Shannon entropy. Online post, https://terrytao.wordpress.com/2020/07/20/the-sunflower-lemma-via-shannon-entropy, 2020.

Lan V Truong, Matthew Aldridge, and Jonathan Scarlett. On the all-or-nothing behavior of bernoulli group testing. *IEEE Journal on Selected Areas in Information Theory*, 1(3):669–680, 2020.

Yihong Wu, Jiaming Xu, and H Yu Sophie. Settling the sharp reconstruction thresholds of random graph matching. *IEEE Transactions on Information Theory*, 68(8):5391–5417, 2022.