

Recovery thresholds for hidden weighted sparse graphs

Zhe Hou
Nanjing University

HOUZHE@SMAIL.NJU.EDU.CN and **Jingcheng Liu**

LIU@NJU.EDU.CN

Editors: Steve Hanneke and Tor Lattimore

Abstract

Recovering structural information from noisy high-dimensional data is a fundamental task in statistical inference. We investigate the information-theoretic recovery thresholds for a graph hidden in a randomly weighted complete graph. Specifically, an unknown graph $H^* \in \mathcal{H}_n$ is chosen uniformly at random, and hidden in a complete graph of n vertices through edge weights: for an edge $e \in H^*$, its weight is distributed independently according to \mathcal{P}_n ; otherwise the edge weight is distributed independently according to \mathcal{Q}_n . The goal is to recover the hidden edge set from these noisy observations. Our primary focus is on *almost exact recovery*, where all but a vanishing fraction of the edges of H^* must be identified. By choosing $\mathcal{P} = \text{Bern}(1)$, $\mathcal{Q} = \text{Bern}(q)$, this model captures the well-studied planted Erdős-Rényi recovery setting, and other weighted formulations such as Gaussian and Exponential distributions.

Assuming a local Lipschitzness of the Rényi divergence between distributions \mathcal{P}_n and \mathcal{Q}_n , and a mild density condition for the graphs \mathcal{H}_n , we give a unified characterization of the information-theoretic limit for recovering almost all of H (also known as almost exact recovery).

Our characterization uses the KL divergence $D_{\text{KL}}(\mathcal{P}_n \parallel \mathcal{Q}_n)$ as the signal-to-noise metric. We show that there is a critical threshold D_c governed by the logarithm of the first moment threshold of \mathcal{H}_n in the Erdős-Rényi random graph model $G(n, p)$. For any constant η , if $D_{\text{KL}} \geq (1 + \eta)D_c$, the maximum likelihood estimator (MLE) achieves almost exact recovery; conversely if $D_{\text{KL}} \leq (1 - \eta)D_c$, almost exact recovery is information-theoretically impossible. Our MLE analysis is based on an extension of [Ding et al. \(2020\)](#); While our lower bound involves deriving a distance-based Fano-type inequality based on Sibson’s α -mutual information, and it also extends to the task of partial recovery, in which only a constant λ -fraction of the hidden graph is required to be recovered. As a result, if $D_{\text{KL}} \leq (\lambda - \eta)D_c$, we also proved that the recovery of any λ -fraction must fail with high probability.

Another interesting phenomenon exhibited by some of these models is the “All-or-Nothing” (AoN), where one can either recover almost all of the hidden graph or essentially nothing at all. For several natural distributional families, including certain Gaussian, Bernoulli and Exponential regimes, we establish an All-or-Nothing (AoN) threshold phenomenon at the exponential scale. For Gaussians, we obtained an AoN by lifting the sharp almost exact recovery threshold through the I-MMSE relation; For the others, we combine a second-moment argument inspired by planted subgraph recovery [Mossel et al. \(2023\)](#) with the second-order Rényi divergence. We also provide distributional examples where AoN is not universal at this level of generality, and our partial recovery lower bound is already the best possible. ¹

1. Extended abstract. Full version appears as [\[arXiv:2606.14335, v1\]](#).

References

- Louigi Addario-Berry, Nicolas Broutin, Luc Devroye, and Gábor Lugosi. On combinatorial testing problems. *The Annals of Statistics*, 2010.
- Alexei Ashikhmin, Gerhard Kramer, and Stephan ten Brink. Code rate and the area under extrinsic information transfer curves. In *Proceedings IEEE International Symposium on Information Theory*, page 115. IEEE, 2002.
- 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.
- Boaz Barak, Samuel Hopkins, Jonathan Kelner, Pravesh K Kothari, Ankur Moitra, and Aaron Potechin. A nearly tight sum-of-squares lower bound for the planted clique problem. *SIAM Journal on Computing*, 48(2):687–735, 2019.
- JM Blair, CA Edwards, and J Howard Johnson. Rational chebyshev approximations for the inverse of the error function. *Mathematics of Computation*, 30(136):827–830, 1976.
- Andrej Bogdanov, Chris Jones, Alon Rosen, and Ilias Zadik. Low-degree security of the planted random subgraph problem. In *Theory of Cryptography Conference*, pages 255–275. Springer, 2024.
- Béla Bollobás and Alan M Frieze. On matchings and hamiltonian cycles in random graphs. *Annals of Discrete Mathematics*, 28:23–46, 1985.
- Bela Bollobas, Trevor I. Fenner, and Alan M. Frieze. An algorithm for finding hamilton paths and cycles in random graphs. *Combinatorica*, 7:327–341, 1987.
- Zongchen Chen, Elchanan Mossel, and Ilias Zadik. Almost-linear planted cliques elude the metropolis process. In *Proceedings of the 2023 Annual ACM-SIAM Symposium on Discrete Algorithms (SODA)*, pages 4504–4539. SIAM, 2023.
- 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.
- Yash Deshpande and Andrea Montanari. Sparse PCA via covariance thresholding. *Advances in Neural Information Processing Systems*, 27, 2014.
- Jian Ding, Yihong Wu, Jiaming Xu, and Dana Yang. Consistent recovery threshold of hidden nearest neighbor graphs. In *Conference on Learning Theory*, pages 1540–1553. PMLR, 2020.
- Jian Ding, Yihong Wu, Jiaming Xu, and Dana Yang. The planted matching problem: Sharp threshold and infinite-order phase transition. *Probability Theory and Related Fields*, 187(1-2):1–71, 2023.
- Yunzi Ding, Dmitriy Kunisky, Alexander S Wein, and Afonso S Bandeira. Subexponential-time algorithms for sparse pca. *Foundations of Computational Mathematics*, 24(3):865–914, 2024.
- David L Donoho and Iain M Johnstone. Minimax risk over lp-balls for lp-error. *Probability theory and related fields*, 99:277–303, 1994.

- Tommaso d’Orsi, Pravesh K Kothari, Gleb Novikov, and David Steurer. Sparse pca: algorithms, adversarial perturbations and certificates. In *2020 IEEE 61st Annual Symposium on Foundations of Computer Science (FOCS)*, pages 553–564. IEEE, 2020.
- John C Duchi and Martin J Wainwright. Distance-based and continuum fano inequalities with applications to statistical estimation. *arXiv preprint arXiv:1311.2669*, 2013.
- Amedeo Roberto Esposito, Michael Gastpar, and Ibrahim Issa. Sibson alpha-mutual information and its variational representations. *IEEE Transactions on Information Theory*, pages 1–1, 2025. doi: 10.1109/TIT.2025.3587340.
- Dongning Guo, Shlomo Shamai, and Sergio Verdú. Mutual information and minimum mean-square error in Gaussian channels. *IEEE transactions on information theory*, 51(4):1261–1282, 2005.
- Shuichi Hirahara and Nobutaka Shimizu. Planted clique conjectures are equivalent. In *Proceedings of the 56th Annual ACM Symposium on Theory of Computing*, pages 358–366, 2024.
- Samuel Hopkins. *Statistical inference and the sum of squares method*. PhD thesis, Cornell University, 2018.
- Mark Jerrum. Large cliques elude the metropolis process. *Random Structures & Algorithms*, 3(4): 347–359, 1992.
- Jeff Kahn and Gil Kalai. Thresholds and expectation thresholds. *Combinatorics, Probability and Computing*, 16(3):495–502, 2007.
- Daniel Z. Lee, Francisco Pernice, Amit Rajaraman, and Ilias Zadik. The fundamental limits of recovering planted subgraphs. In Nika Haghtalab and Ankur Moitra, editors, *Proceedings of Thirty Eighth Conference on Learning Theory*, volume 291 of *Proceedings of Machine Learning Research*, pages 3578–3579. PMLR, 30 Jun–04 Jul 2025.
- Erez Lieberman-Aiden, Nynke L Van Berkum, Louise Williams, Maxim Imakaev, Tobias Ragozcy, Agnes Telling, Ido Amit, Bryan R Lajoie, Peter J Sabo, Michael O Dorschner, et al. Comprehensive mapping of long-range interactions reveals folding principles of the human genome. *science*, 326 (5950):289–293, 2009.
- Brendan D McKay and Nicholas C Wormald. Asymptotic enumeration by degree sequence of graphs with degrees $o(n^{1/2})$. *Combinatorica*, 11(4):369–382, 1991.
- Cyril Méasson, Andrea Montanari, and Rüdiger Urbanke. Maxwell construction: The hidden bridge between iterative and maximum a posteriori decoding. *IEEE Transactions on Information Theory*, 54(12):5277–5307, 2008.
- Mehrdad Moharrami, Cristopher Moore, and Jiaming Xu. The planted matching problem: Phase transitions and exact results. *The Annals of Applied Probability*, 31(6):2663–2720, 2021.
- Mehrdad Moharrami, Cristopher Moore, and Jiaming Xu. The planted spanning tree problems: Exact overlap characterization via local weak convergence extended abstract. In Nika Haghtalab and Ankur Moitra, editors, *Proceedings of Thirty Eighth Conference on Learning Theory*, volume 291 of *Proceedings of Machine Learning Research*, pages 4166–4167. PMLR, 30 Jun–04 Jul 2025.

- Andrea Montanari, Daniel Reichman, and Ofer Zeitouni. On the limitation of spectral methods: From the gaussian hidden clique problem to rank-one perturbations of gaussian tensors. *Advances in Neural Information Processing Systems*, 28, 2015.
- Elchanan Mossel, Jonathan Niles-Weed, Nike Sun, and Ilias Zadik. On the second kahn–kalai conjecture. *arXiv preprint arXiv:2209.03326*, 2022.
- Elchanan Mossel, Jonathan Niles-Weed, Youngtak Sohn, Nike Sun, and Ilias Zadik. Sharp thresholds in inference of planted subgraphs. In *The Thirty Sixth Annual Conference on Learning Theory*, pages 5573–5577. PMLR, 2023.
- Mark EJ Newman and Duncan J Watts. Scaling and percolation in the small-world network model. *Physical review E*, 60(6):7332, 1999.
- Jonathan Niles-Weed and Ilias Zadik. The all-or-nothing phenomenon in sparse tensor pca. *Advances in Neural Information Processing Systems*, 33:17674–17684, 2020.
- 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.
- Jinyoung Park and Huy Pham. A proof of the kahn–kalai conjecture. *Journal of the American Mathematical Society*, 37(1):235–243, 2024.
- Yury Polyanskiy and Yihong Wu. *Information theory: From coding to learning*. Cambridge university press, 2025.
- Yury Polyanskiy, H Vincent Poor, and Sergio Verdú. Channel coding rate in the finite blocklength regime. *IEEE Transactions on Information Theory*, 56(5):2307–2359, 2010.
- Garvesh Raskutti, Martin J Wainwright, and Bin Yu. Minimax rates of estimation for high-dimensional linear regression over ℓ_q -balls. *IEEE transactions on information theory*, 57(10):6976–6994, 2011.
- Guilhem Semerjian, Gabriele Sicuro, and Lenka Zdeborová. Recovery thresholds in the sparse planted matching problem. *Physical Review E*, 102(2):022304, 2020.
- Gabriele Sicuro and Lenka Zdeborová. The planted k-factor problem. *Journal of Physics A: Mathematical and Theoretical*, 54(17):175002, 2021.
- Tim Van Erven and Peter Harremoos. Rényi divergence and kullback-leibler divergence. *IEEE Transactions on Information Theory*, 60(7):3797–3820, 2014.
- Ramji Venkataramanan and Oliver Johnson. A strong converse bound for multiple hypothesis testing, with applications to high-dimensional estimation. *Electronic Journal of Statistics*, 2018.
- Duncan J Watts and Steven H Strogatz. Collective dynamics of ‘small-world’ networks. *nature*, 393(6684):440–442, 1998.
- Yuhong Yang and Andrew Barron. Information-theoretic determination of minimax rates of convergence. *Annals of Statistics*, pages 1564–1599, 1999.

Bin Yu, Assouad, fano, and le cam. In *Festschrift for Lucien Le Cam: research papers in probability and statistics*, pages 423–435. Springer, 1997.

Xifan Yu, Ilias Zadik, and Peiyuan Zhang. Counting stars is constant-degree optimal for detecting any planted subgraph. *arXiv preprint arXiv:2403.17766*, 2024.