Spring 2023

S&DS 684:

Statistical inference on graphs

Final reading project

Due: May 8th, 2023

1 Rule

As the final project, please choose one or two papers of a given topic and write a coherent essay summarizing the paper(s). In preparing the essay, please note the following:

- Clearly summarize the statistical/mathematical model, the technical assumptions and the main results (e.g., upper bounds, lower bounds, etc).
- Read the proof carefully (for conference papers see their full version), sketch the major steps, and present the key ideas.
- Discuss in your opinion the most innovative/challenging part of the paper and why it is interesting. It is also valuable to present your critiques on the paper such as certain argument which you find unclear or suboptimal.

2 Reading list

Community detection and synchronization

- Reconstruction on trees and SBM: [YP22], [MSS22]
- Information-theoretic bounds [BMNN16] – *Yutong*, [DAM15] – *Max*, [GLM18], [COKPZ18]
- Message passing and non-backtracking matrix: [BLM18], [DM15] – *Yihan*, [AS18]
- Spectral method: [LL20], [MM18]
- Convex relaxation: [HWX16], [CL14], [CLX15], [FM17], [FC22], [GZ22], [CY21] – *Heejune*
- Information-percolation method: [PW20] – *Ruixiao*, [ABA20]

Statistical-computational tradeoffs

- Average-case reduction: [MW15], [HWX15], [GMZ], [BBH18], [BBH19], [LZ22]
- Low-degree polynomial framework: [KWB22] (survey, do not choose), [SW22], [Wei22b], [MWZ23], [Wei22a] – *Peiyuan* (and the earlier [GJW20]), [MW21], [BBH+21] – *Fred*
- Overlap gap property and free energy well: [Gam21] (survey, do not choose), [GJS21], [AWZ22], [GK21]

Random graph matching

- Information-theoretic analysis: [CK17] and [WXY22] – *Kaylee*, [GML21], [DD22], [RS21]
- From tree matching to graph matching: [GMS22], [GM20] – *Haoyu*
- Spectral method and convex relaxation: [FMWX22a, FMWX22b] – *Leon*
- Combinatorial algorithms: [DMWX21], [MRT23], [DDG22], [DL22]
- Planted Matching problem (linear assignment): [MMX21], [DWXY21]
- Geometric models: [KNW22], [WWXY22]

References

[ABA20] Emmanuel Abbe and Enric Boix-Adserà. An information-percolation bound for spin synchronization on general graphs. *The Annals of Applied Probability*, 30(3):1066–1090, 2020.

[AS18] Emmanuel Abbe and Colin Sandon. Proof of the achievability conjectures for the general stochastic block model. *Communications on Pure and Applied Mathematics*, 71(7):1334–1406, 2018.

[AWZ22] Gérard Ben Arous, Alexander S Wein, and Ilias Zadik. Free energy wells and overlap gap property in sparse pca. *Communications on Pure and Applied Mathematics*, 2022.

[BBH18] Matthew Brennan, Guy Bresler, and Wasim Huleihel. Reducibility and computational lower bounds for problems with planted sparse structure. In *Conference On Learning Theory*, pages 48–166. PMLR, 2018.

[BBH19] Matthew Brennan, Guy Bresler, and Wasim Huleihel. Universality of computational lower bounds for submatrix detection. In *Conference on Learning Theory*, pages 417–468. PMLR, 2019.

[BBH+21] Matthew S Brennan, Guy Bresler, Sam Hopkins, Jerry Li, and Tselil Schramm. Statistical query algorithms and low degree tests are almost equivalent. In *Conference on Learning Theory*, pages 774–774. PMLR, 2021.

[BLM18] Charles Bordenave, Marc Lelarge, and Laurent Massoulié. Nonbacktracking spectrum of random graphs: Community detection and nonregular ramanujan graphs. *The Annals of Probability*, 46(1):1–71, 2018.

[BMNN16] Jess Banks, Cristopher Moore, Joe Neeman, and Praneeth Netrapalli. Information-theoretic thresholds for community detection in sparse networks. In *Proceedings of the 29th Conference on Learning Theory, COLT 2016, New York, NY, June 23-26 2016*, pages 383–416, 2016.

[CK17] Daniel Cullina and Negar Kiyavash. Exact alignment recovery for correlated Erdős-Rényi graphs. *arXiv 1711.06783*, 2017.

[CL14] T. Cai and X. Li. Robust and computationally feasible community detection in the presence of arbitrary outlier nodes. *arXiv preprint arXiv:1404.6000*, 2014.

[CLX15] Y. Chen, X. Li, and J. Xu. Convexified modularity maximization for degree-corrected stochastic block models. arXiv 1512.08425, December 2015.

[COKPZ18] Amin Coja-Oghlan, Florent Krzakala, Will Perkins, and Lenka Zdeborová. Information-theoretic thresholds from the cavity method. *Advances in Mathematics*, 333:694–795, 2018.

[CY21] Xiaohui Chen and Yun Yang. Cutoff for exact recovery of gaussian mixture models. *IEEE Transactions on Information Theory*, 67(6):4223–4238, 2021.

[DAM15] Yash Deshpande, Emmanuel Abbe, and Andrea Montanari. Asymptotic mutual information for the two-groups stochastic block model. *arXiv:1507.08685*, 2015.

[DD22] Jian Ding and Hang Du. Detection threshold for correlated erd\h {o} sr\'enyi graphs via densest subgraphs. *arXiv preprint arXiv:2203.14573*, 2022.

[DDG22] Jian Ding, Hang Du, and Shuyang Gong. A polynomial-time approximation scheme for the maximal overlap of two independent erdos-rényi graphs. *arXiv preprint arXiv:2210.07823*, 2022.

[DL22] Jian Ding and Zhangsong Li. A polynomial time iterative algorithm for matching gaussian matrices with non-vanishing correlation. *arXiv preprint arXiv:2212.13677*, 2022.

[DM15] Y. Deshpande and A. Montanari. Finding hidden cliques of size $\sqrt{N/e}$ in nearly linear time. *Foundations of Computational Mathematics*, 15(4):1069–1128, August 2015.

[DMWX21] Jian Ding, Zongming Ma, Yihong Wu, and Jiaming Xu. Efficient random graph matching via degree profiles. *Probability Theory and Related Fields*, 179(1):29–115, 2021.

[DWXY21] Jian Ding, Yihong Wu, Jiaming Xu, and Dana Yang. The planted matching problem: Sharp threshold and infinite-order phase transition. *arXiv preprint arXiv:2103.09383*, 2021.

[FC22] Yingjie Fei and Yudong Chen. Hidden integrality and semirandom robustness of SDP relaxation for sub-Gaussian mixture model. *Mathematics of Operations Research*, 47(3):2464–2493, 2022.

[FM17] Zhou Fan and Andrea Montanari. How well do local algorithms solve semidefinite programs? In *Proceedings of the 49th Annual ACM SIGACT Symposium on Theory of Computing*, pages 604–614, 2017.

[FMWX22a] Zhou Fan, Cheng Mao, Yihong Wu, and Jiaming Xu. Spectral graph matching and regularized quadratic relaxations I algorithm and Gaussian analysis. *Foundations of Computational Mathematics*, pages 1–55, Jun 2022.

[FMWX22b] Zhou Fan, Cheng Mao, Yihong Wu, and Jiaming Xu. Spectral graph matching and regularized quadratic relaxations II: Erdős-Rényi graphs and universality. *Foundations of Computational Mathematics*, pages 1–51, Jun 2022.

[Gam21] David Gamarnik. The overlap gap property: A topological barrier to optimizing over random structures. *Proceedings of the National Academy of Sciences*, 118(41):e2108492118, 2021.

[GJS21] David Gamarnik, Aukosh Jagannath, and Subhabrata Sen. The overlap gap property in principal submatrix recovery. *Probability Theory and Related Fields*, 181:757–814, 2021.

[GJW20] David Gamarnik, Aukosh Jagannath, and Alexander S Wein. Low-degree hardness of random optimization problems. In *2020 IEEE 61st Annual Symposium on Foundations of Computer Science (FOCS)*, pages 131–140. IEEE, 2020.

[GK21] David Gamarnik and Eren C Kızıldağ. Algorithmic obstructions in the random number partitioning problem. *arXiv preprint arXiv:2103.01369*, 2021.

[GLM18] Lennart Gulikers, Marc Lelarge, and Laurent Massoulié. An impossibility result for reconstruction in the degree-corrected stochastic block model. *The Annals of Applied Probability*, 28(5):3002–3027, 2018.

[GM20] Luca Ganassali and Laurent Massoulié. From tree matching to sparse graph alignment. In *Conference on Learning Theory*, pages 1633–1665. PMLR, 2020.

[GML21] Luca Ganassali, Laurent Massoulié, and Marc Lelarge. Impossibility of partial recovery in the graph alignment problem. In *Conference on Learning Theory*, pages 2080–2102. PMLR, 2021.

[GMS22] Luca Ganassali, Laurent Massoulié, and Guilhem Semerjian. Statistical limits of correlation detection in trees. *arXiv preprint arXiv:2209.13723*, 2022.

[GMZ] Chao Gao, Zongming Ma, and Harrison Zhou. Sparse CCA: Adaptive estimation and computational barriers. *The Annals of Statistics*, 45(5):2074–2101.

[GZ22] Chao Gao and Anderson Y Zhang. SDP achieves exact minimax optimality in phase synchronization. *IEEE Transactions on Information Theory*, 68(8):5374–5390, 2022.

[HWX15] B. Hajek, Y. Wu, and J. Xu. Computational lower bounds for community detection on random graphs. In *Proceedings of COLT 2015*, pages 899–928, June 2015.

[HWX16] B. Hajek, Y. Wu, and J. Xu. Achieving exact cluster recovery threshold via semidefinite programming: Extensions. *IEEE Trans. on Information Theory*, 62(10):5918–5937, Oct. 2016.

[KNW22] Dmitriy Kunisky and Jonathan Niles-Weed. Strong recovery of geometric planted matchings. In *Proceedings of the 2022 Annual ACM-SIAM Symposium on Discrete Algorithms (SODA)*, pages 834–876. SIAM, 2022.

[KWB22] Dmitriy Kunisky, Alexander S Wein, and Afonso S Bandeira. Notes on computational hardness of hypothesis testing: Predictions using the low-degree likelihood ratio. In *Mathematical Analysis, its Applications and Computation: ISAAC 2019, Aveiro, Portugal, July 29–August 2*, pages 1–50. Springer, 2022.

[LL20] Yue M Lu and Gen Li. Phase transitions of spectral initialization for high-dimensional non-convex estimation. *Information and Inference: A Journal of the IMA*, 9(3):507–541, 2020.

[LZ22] Yuetian Luo and Anru R Zhang. Tensor clustering with planted structures: Statistical optimality and computational limits. *The Annals of Statistics*, 50(1):584–613, 2022.

[MM18] Marco Mondelli and Andrea Montanari. Fundamental limits of weak recovery with applications to phase retrieval. In *Conference On Learning Theory*, pages 1445–1450. PMLR, 2018.

[MMX21] 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.

[MRT23] Cheng Mao, Mark Rudelson, and Konstantin Tikhomirov. Exact matching of random graphs with constant correlation. *Probability Theory and Related Fields*, pages 1–63, 2023.

[MSS22] Elchanan Mossel, Allan Sly, and Youngtak Sohn. Exact phase transitions for stochastic block models and reconstruction on trees. *arXiv preprint arXiv:2212.03362*, 2022.

[MW15] Z. Ma and Y. Wu. Computational barriers in minimax submatrix detection. *The Annals of Statistics*, 43(3):1089–1116, 2015.

[MW21] Cheng Mao and Alexander S Wein. Optimal spectral recovery of a planted vector in a subspace. *arXiv preprint arXiv:2105.15081*, 2021.

[MWZ23] Cheng Mao, Alexander S Wein, and Shenduo Zhang. Detection-recovery gap for planted dense cycles. *arXiv preprint arXiv:2302.06737*, 2023.

[PW20] Yury Polyanskiy and Yihong Wu. Application of the information-percolation method to reconstruction problems on graphs. *Mathematical Statistics and Learning*, 2(1):1–24, 2020.

[RS21] Miklos Racz and Anirudh Sridhar. Correlated stochastic block models: Exact graph matching with applications to recovering communities. *Advances in Neural Information Processing Systems*, 34:22259–22273, 2021.

[SW22] Tselil Schramm and Alexander S Wein. Computational barriers to estimation from low-degree polynomials. *The Annals of Statistics*, 50(3):1833–1858, 2022.

[Wei22a] Alexander S Wein. Average-case complexity of tensor decomposition for low-degree polynomials. *arXiv preprint arXiv:2211.05274*, 2022.

[Wei22b] Alexander S Wein. Optimal low-degree hardness of maximum independent set. *Mathematical Statistics and Learning*, 4(3):221–251, 2022.

[WWXY22] Haoyu Wang, Yihong Wu, Jiaming Xu, and Israel Yolou. Random graph matching in geometric models: the case of complete graphs. In *Conference on Learning Theory*, pages 3441–3488. PMLR, 2022.

[WXY22] Yihong Wu, Jiaming Xu, and Sophie H. Yu. Settling the sharp reconstruction thresholds of random graph matching. *IEEE Transactions on Information Theory*, 68(8):5391–5417, Apr 2022.

[YP22] Qian Yu and Yury Polyanskiy. Ising model on locally tree-like graphs: Uniqueness of solutions to cavity equations. *arXiv preprint arXiv:2211.15242*, 2022.