

STATISTICAL SCIENCE

Volume 36, Number 1

February 2021

Special Section on Network Data

Introduction to the Special Section.....	<i>Yihong Wu and Harrison Zhou</i>	1
Convex Relaxation Methods for Community Detection	<i>Xiaodong Li, Yudong Chen and Jiaming Xu</i>	2
Minimax Rates in Network Analysis: Graphon Estimation, Community Detection and Hypothesis Testing.....	<i>Chao Gao and Zongming Ma</i>	16
Additive and Multiplicative Effects Network Models	<i>Peter Hoff</i>	34
A Statistical Framework for Modern Network Science....	<i>Harry Crane and Walter Dempsey</i>	51
On Estimation and Inference in Latent Structure Random Graphs	<i>Avanti Athreya, Minh Tang, Youngser Park and Carey E. Priebe</i>	68
Network Modeling in Biology: Statistical Methods for Gene and Brain Networks	<i>Y. X. Rachel Wang, Lexin Li, Jingyi Jessica Li and Haiyan Huang</i>	89

Articles

Bipartite Causal Inference with Interference...	<i>Corwin M. Zigler and Georgia Papadogeorgou</i>	109
A General Framework for Vecchia Approximations of Gaussian Processes	<i>Matthias Katzfuss and Joseph Guinness</i>	124
On General Notions of Depth for Regression	<i>Yijun Zuo</i>	142
A Conversation with Tze Leung Lai.....	<i>Ying Lu, Dylan S. Small and Zhiliang Ying</i>	158

Statistical Science [ISSN 0883-4237 (print); ISSN 2168-8745 (online)], Volume 36, Number 1, February 2021. Published quarterly by the Institute of Mathematical Statistics, 9760 Smith Road, Waite Hill, Ohio 44094, USA. Periodicals postage paid at Cleveland, Ohio and at additional mailing offices.

POSTMASTER: Send address changes to *Statistical Science*, Institute of Mathematical Statistics, Dues and Subscriptions Office, 9650 Rockville Pike—Suite L2310, Bethesda, MD 20814-3998, USA.

Copyright © 2021 by the Institute of Mathematical Statistics
Printed in the United States of America

Statistical Science

Volume 36, Number 1 (1–167) February 2021

Volume 36

Number 1

February 2021

Special Section on Network Data

Introduction to the Special Section

Yihong Wu and Harrison Zhou

Convex Relaxation Methods for Community Detection

Xiaodong Li, Yudong Chen and Jiaming Xu

Minimax Rates in Network Analysis: Graphon Estimation, Community Detection and Hypothesis Testing

Chao Gao and Zongming Ma

Additive and Multiplicative Effects Network Models

Peter Hoff

A Statistical Framework for Modern Network Science

Harry Crane and Walter Dempsey

On Estimation and Inference in Latent Structure Random Graphs

Avanti Athreya, Minh Tang, Youngser Park and Carey E. Priebe

Network Modeling in Biology: Statistical Methods for Gene and Brain Networks

Y. X. Rachel Wang, Lexin Li, Jingyi Jessica Li and Haiyan Huang

Articles

Bipartite Causal Inference with Interference

Corwin M. Zigler and Georgia Papadogeorgou

A General Framework for Vecchia Approximations of Gaussian Processes

Matthias Katzfuss and Joseph Guinness

On General Notions of Depth for Regression

Yijun Zuo

A Conversation with Tze Leung Lai

Ying Lu, Dylan S. Small and Zhiliang Ying

EDITOR

Sonia Petrone
Università Bocconi

DEPUTY EDITOR

Peter Green
University of Bristol and University of Technology Sydney

ASSOCIATE EDITORS

Shankar Bhamidi
University of North Carolina
Jay Breidt
Colorado State University
Peter Bühlmann
ETH Zürich
Philip Dawid
University of Cambridge
Robin Evans
University of Oxford
Stefano Favaro
Università di Torino
Alan Gelfand
Duke University
Peter Green
*University of Bristol and
University of Technology
Sydney*
Chris Holmes
University of Oxford
Susan Holmes
Stanford University

Tailen Hsing
University of Michigan
Michael I. Jordan
*University of California,
Berkeley*
Ian McKeague
Columbia University
Peter Müller
University of Texas
Luc Pronzato
Université Nice
Nancy Reid
University of Toronto
Jason Roy
Rutgers University
Chiara Sabatti
Stanford University
Bodhisattva Sen
Columbia University

Glenn Shafer
*Rutgers Business
School–Newark and
New Brunswick
Royal Holloway College,
University of London*
Michael Stein
University of Chicago
Jon Wellner
University of Washington
Yihong Wu
Yale University
Bin Yu
*University of California,
Berkeley*
Giacomo Zanella
Università Bocconi
Cun-Hui Zhang
Rutgers University
Harrison Zhou
Yale University

MANAGING EDITOR

Robert Keener
University of Michigan

PRODUCTION EDITOR

Patrick Kelly

EDITORIAL COORDINATOR

Kristina Mattson

PAST EXECUTIVE EDITORS

Morris H. DeGroot, 1986–1988	Morris Eaton, 2001
Carl N. Morris, 1989–1991	George Casella, 2002–2004
Robert E. Kass, 1992–1994	Edward I. George, 2005–2007
Paul Switzer, 1995–1997	David Madigan, 2008–2010
Leon J. Gleser, 1998–2000	Jon A. Wellner, 2011–2013
Richard Tweedie, 2001	Peter Green, 2014–2016
	Cun-Hui Zhang, 2017–2019

Introduction to the Special Section

Yihong Wu and Harrison Zhou

Convex Relaxation Methods for Community Detection

Xiaodong Li, Yudong Chen and Jiaming Xu

Abstract. This paper surveys recent theoretical advances in convex optimization approaches for community detection. We introduce some important theoretical techniques and results for establishing the consistency of convex community detection under various statistical models. In particular, we discuss the basic techniques based on the primal and dual analysis. We also present results that demonstrate several distinctive advantages of convex community detection, including robustness against outlier nodes, consistency under weak assortativity, and adaptivity to heterogeneous degrees.

This survey is not intended to be a complete overview of the vast literature on this fast-growing topic. Instead, we aim to provide a big picture of the remarkable recent development in this area and to make the survey accessible to a broad audience. We hope that this expository article can serve as an introductory guide for readers who are interested in using, designing, and analyzing convex relaxation methods in network analysis.

Key words and phrases: Community detection, semidefinite program, strong consistency, weak consistency, assortativity, degree correction, robustness.

REFERENCES

- ABBE, E. (2017). Community detection and stochastic block models: Recent developments. *J. Mach. Learn. Res.* **18** Paper No. 177, 86. [MR3827065](#)
- ABBE, E., BANDEIRA, A. S. and HALL, G. (2016). Exact recovery in the stochastic block model. *IEEE Trans. Inform. Theory* **62** 471–487. [MR3447993](#) <https://doi.org/10.1109/TIT.2015.2490670>
- AGARWAL, N., BANDEIRA, A. S., KOILIARIS, K. and KOLLA, A. (2015). Multisection in the Stochastic Block Model using Semidefinite Programming. Available at [arXiv:1507.02323](#).
- AMES, B. P. W. (2014). Guaranteed clustering and biclustering via semidefinite programming. *Math. Program.* **147** 429–465. [MR3258531](#) <https://doi.org/10.1007/s10107-013-0729-x>
- AMES, B. P. W. and VAVASIS, S. A. (2014). Convex optimization for the planted k -disjoint-clique problem. *Math. Program.* **143** 299–337. [MR3152071](#) <https://doi.org/10.1007/s10107-013-0733-1>
- AMINI, A. A. and LEVINA, E. (2018). On semidefinite relaxations for the block model. *Ann. Statist.* **46** 149–179. [MR3766949](#) <https://doi.org/10.1214/17-AOS1545>
- BANDEIRA, A. S. (2018). Random Laplacian matrices and convex relaxations. *Found. Comput. Math.* **18** 345–379. [MR3777782](#) <https://doi.org/10.1007/s10208-016-9341-9>
- BANDEIRA, A. S., BOUMAL, N. and VORONINSKI, V. (2016). On the low-rank approach for semidefinite programs arising in synchronization and community detection. In *29th Annual Conference on Learning Theory* 361–382.
- BANDEIRA, A. S. and VAN HANDEL, R. (2016). Sharp nonasymptotic bounds on the norm of random matrices with independent entries. *Ann. Probab.* **44** 2479–2506. [MR3531673](#) <https://doi.org/10.1214/15-AOP1025>
- BOYD, S. and VANDENBERGHE, L. (2004). *Convex Optimization*. Cambridge Univ. Press, Cambridge. [MR2061575](#) <https://doi.org/10.1017/CBO9780511804441>
- CAI, T. T. and LI, X. (2015). Robust and computationally feasible community detection in the presence of arbitrary outlier nodes. *Ann. Statist.* **43** 1027–1059. [MR3346696](#) <https://doi.org/10.1214/14-AOS1290>
- CHAUDHURI, K., CHUNG, F. and TSIATAS, A. (2012). Spectral clustering of graphs with general degrees in the extended planted partition model. **23** *J. Mach. Learn. Res.* 35.1–35.23.
- CHEN, Y., LI, X. and XU, J. (2018). Convexified modularity maximization for degree-corrected stochastic block models. *Ann. Statist.* **46** 1573–1602. [MR3819110](#) <https://doi.org/10.1214/17-AOS1595>
- CHEN, Y., SANGHAVI, S. and XU, H. (2014). Improved graph clustering. *IEEE Trans. Inform. Theory* **60** 6440–6455. [MR3265033](#) <https://doi.org/10.1109/TIT.2014.2346205>
- CHEN, Y. and XU, J. (2016). Statistical-computational tradeoffs in planted problems and submatrix localization with a growing num-

Xiaodong Li is Assistant Professor, Department of Statistics, University of California, Davis, 4232 Math Sciences Bldg, 399 Crocker Lane, Davis, California 95616, USA (e-mail: xdgli@ucdavis.edu). Yudong Chen is Assistant Professor, School of Operations Research and Information Engineering, Cornell University, Ithaca, New York 14853, USA (e-mail: yudong.chen@cornell.edu). Jiaming Xu is Assistant Professor, Fuqua School of Business, Duke University, Durham, North Carolina 27708, USA (e-mail: jx77@duke.edu).

- ber of clusters and submatrices. *J. Mach. Learn. Res.* **17** 882–938. [MR3491121](#)
- CHEN, Y., JALALI, A., SANGHAVI, S. and XU, H. (2014). Clustering partially observed graphs via convex optimization. *J. Mach. Learn. Res.* **15** 2213–2238. [MR3231602](#)
- DECELLE, A., KRZAKALA, F., MOORE, C. and ZDEBOROVÁ, L. (2011). Asymptotic analysis of the stochastic block model for modular networks and its algorithmic applications. *Phys. Rev. E* **84** 066106.
- FAN, Z. and MONTANARI, A. (2017). How well do local algorithms solve semidefinite programs? In *STOC'17—Proceedings of the 49th Annual ACM SIGACT Symposium on Theory of Computing* 604–614. ACM, New York. [MR3678214](#)
- FEI, Y. and CHEN, Y. (2018). Hidden integrality of SDP relaxation for sub-Gaussian mixture models. In *Conference on Learning Theory (COLT)*.
- FEI, Y. and CHEN, Y. (2019a). Exponential error rates of SDP for block models: Beyond Grothendieck's inequality. *IEEE Trans. Inform. Theory* **65** 551–571. [MR3901009](#) <https://doi.org/10.1109/TIT.2018.2839677>
- FEI, Y. and CHEN, Y. (2019b). Achieving the Bayes error rate in stochastic block model by SDP, robustly. In *Conference on Learning Theory (COLT)*.
- FEIGE, U. and OFEK, E. (2005). Spectral techniques applied to sparse random graphs. *Random Structures Algorithms* **27** 251–275. [MR2155709](#) <https://doi.org/10.1002/rsa.20089>
- FORTUNATO, S. and BARTHELEMY, M. (2007). Resolution limit in community detection. *Proc. Natl. Acad. Sci. USA* **104** 36–41.
- GOEMANS, M. X. and WILLIAMSON, D. P. (1995). Improved approximation algorithms for maximum cut and satisfiability problems using semidefinite programming. *J. Assoc. Comput. Mach.* **42** 1115–1145. [MR1412228](#) <https://doi.org/10.1145/227683.227684>
- GROTHENDIECK, A. (1996). Résumé de la théorie métrique des produits tensoriels topologiques. *Resenhas* **2** 401–480. [MR1466414](#)
- GUÉDON, O. and VERSHYNIN, R. (2016). Community detection in sparse networks via Grothendieck's inequality. *Probab. Theory Related Fields* **165** 1025–1049. [MR3520025](#) <https://doi.org/10.1007/s00440-015-0659-z>
- HAJEK, B., WU, Y. and XU, J. (2016a). Achieving exact cluster recovery threshold via semidefinite programming. *IEEE Trans. Inform. Theory* **62** 2788–2797. [MR3493879](#) <https://doi.org/10.1109/TIT.2016.2546280>
- HAJEK, B., WU, Y. and XU, J. (2016b). Achieving exact cluster recovery threshold via semidefinite programming: Extensions. *IEEE Trans. Inform. Theory* **62** 5918–5937. [MR3552431](#) <https://doi.org/10.1109/TIT.2016.2594812>
- HAJEK, B., WU, Y. and XU, J. (2016c). Semidefinite programs for exact recovery of a hidden community. In *Proceedings of Conference on Learning Theory (COLT)*. Available at [arXiv:1602.06410](#).
- HOLLAND, P. W., LASKEY, K. B. and LEINHARDT, S. (1983). Stochastic blockmodels: First steps. *Soc. Netw.* **5** 109–137. [MR0718088](#) [https://doi.org/10.1016/0378-8733\(83\)90021-7](https://doi.org/10.1016/0378-8733(83)90021-7)
- IGUCHI, T., MIXON, D. G., PETERSON, J. and VILLAR, S. (2015). On the tightness of an SDP relaxation of k-means. Preprint. Available at [arXiv:1505.04778](#).
- IGUCHI, T., MIXON, D. G., PETERSON, J. and VILLAR, S. (2017). Probably certifiably correct k -means clustering. *Math. Program.* **165** 605–642. [MR3707375](#) <https://doi.org/10.1007/s10107-016-1097-0>
- JAVANMARD, A., MONTANARI, A. and RICCI-TERSENGHI, F. (2015). Phase transitions in semidefinite relaxations. Available at [arXiv:1511.08769](#).
- KRZAKALA, F., MOORE, C., MOSSEL, E., NEEMAN, J., SLY, A., ZDEBOROVÁ, L. and ZHANG, P. (2013). Spectral redemption in clustering sparse networks. *Proc. Natl. Acad. Sci. USA* **110** 20935–20940. [MR3174850](#) <https://doi.org/10.1073/pnas.1312486110>
- LANCICHINETTI, A. and FORTUNATO, S. (2011). Limits of modularity maximization in community detection. *Phys. Rev. E* **84**.
- LI, X., LI, Y., LING, S., STROHMER, T. and WEI, K. (2017). When do birds of a feather flock together? K-means, proximity, and conic programming. Preprint. Available at [arXiv:1710.06008](#).
- LIM, S. H., CHEN, Y. and XU, H. (2017). Clustering from general pairwise observations with applications to time-varying graphs. *J. Mach. Learn. Res.* **18** Paper No. 49, 47. [MR3670494](#)
- LINDENSTRAUSS, J. and PEŁCZYŃSKI, A. (1968). Absolutely summing operators in L_p -spaces and their applications. *Studia Math.* **29** 275–326. [MR0231188](#) <https://doi.org/10.4064/sm-29-3-275-326>
- MAKARYCHEV, K., MAKARYCHEV, Y. and VIJAYARAGHAVAN, A. (2016). Learning communities in the presence of errors. In *Conference on Learning Theory* 1258–1291.
- MASSOULIÉ, L. (2014). Community detection thresholds and the weak Ramanujan property. In *STOC'14—Proceedings of the 2014 ACM Symposium on Theory of Computing* 694–703. ACM, New York. [MR3238997](#)
- MATHIEU, C. and SCHUDY, W. (2010). Correlation clustering with noisy input. In *Proceedings of the Twenty-First Annual ACM-SIAM Symposium on Discrete Algorithms* 712–728. SIAM, Philadelphia, PA. [MR2768627](#)
- MEI, S., MISIAKIEWICZ, T., MONTANARI, A. and OLIVEIRA, R. I. (2017). Solving SDPs for synchronization and MaxCut problems via the Grothendieck inequality. In *Proceedings of Machine Learning Research* **65** 1–40.
- MOITRA, A., PERRY, W. and WEIN, A. S. (2016). How robust are reconstruction thresholds for community detection? In *STOC'16—Proceedings of the 48th Annual ACM SIGACT Symposium on Theory of Computing* 828–841. ACM, New York. [MR3536617](#)
- MONTANARI, A. (2016). A Grothendieck-type inequality for local maxima. Preprint. Available at [arXiv:1603.04064](#).
- MONTANARI, A. and SEN, S. (2015). Semidefinite programs on sparse random graphs. Available at [arXiv:1504.05910](#).
- MOORE, C. (2017). The computer science and physics of community detection: Landscapes, phase transitions, and hardness. *Bull. Eur. Assoc. Theor. Comput. Sci. EATCS* **121** 26–61. [MR3699594](#)
- MOSSEL, E., NEEMAN, J. and SLY, A. (2015). Reconstruction and estimation in the planted partition model. *Probab. Theory Related Fields* **162** 431–461. [MR3383334](#) <https://doi.org/10.1007/s00440-014-0576-6>
- MOSSEL, E., NEEMAN, J. and SLY, A. (2016). Consistency thresholds for the planted bisection model. *Electron. J. Probab.* **21** Paper No. 21, 24. [MR3485363](#) <https://doi.org/10.1214/16-EJP4185>
- MOSSEL, E., NEEMAN, J. and SLY, A. (2018). A proof of the block model threshold conjecture. *Combinatorica* **38** 665–708. [MR3876880](#) <https://doi.org/10.1007/s00493-016-3238-8>
- NEWMAN, M. E. J. (2006). Modularity and community structure in networks. *Proc. Natl. Acad. Sci. USA* **103** 8577–8582. [https://doi.org/10.1073/pnas.0601602103](#)
- NOCEDAL, J. and WRIGHT, S. J. (2006). *Numerical Optimization*, 2nd ed. *Springer Series in Operations Research and Financial Engineering*. Springer, New York. [MR2244940](#)
- OYMAK, S. and HASSIBI, B. (2011). Finding dense clusters via low rank + sparse decomposition. Available at [arXiv:1104.5186](#).
- PENG, J. and WEI, Y. (2007). Approximating K -means-type clustering via semidefinite programming. *SIAM J. Optim.* **18** 186–205. [MR2299680](#) <https://doi.org/10.1137/050641983>
- PERRY, W. and WEIN, A. S. (2015). A semidefinite program for unbalanced multisection in the stochastic block model. Available at [arXiv:1507.05605](#).

- REICHARDT, J. and BORNHOLDT, S. (2006). Statistical mechanics of community detection. *Phys. Rev. E* (3) **74** 016110, 14. [MR2276596](#) <https://doi.org/10.1103/PhysRevE.74.016110>
- ROHE, K., CHATTERJEE, S. and YU, B. (2011). Spectral clustering and the high-dimensional stochastic blockmodel. *Ann. Statist.* **39** 1878–1915. [MR2893856](#) <https://doi.org/10.1214/11-AOS887>
- VU, V. (2018). A simple SVD algorithm for finding hidden partitions. *Combin. Probab. Comput.* **27** 124–140. [MR3734334](#) <https://doi.org/10.1017/S0963548317000463>
- WU, Y. and XU, J. (2018). Statistical problems with planted structures: Information-theoretical and computational limits. Preprint. Available at [arXiv:1806.00118](#).
- YAN, B., SARKAR, P. and CHENG, X. (2018). Provable estimation of the number of blocks in block models. In *Proceedings of the 21st International Conference on Artificial Intelligence and Statistics (AISTATS) 2018, Lanzarote, Spain*.
- ZHANG, A. Y. and ZHOU, H. H. (2015). Minimax rates of community detection in stochastic block models. Available at [arXiv:1507.05313](#).

Minimax Rates in Network Analysis: Graphon Estimation, Community Detection and Hypothesis Testing

Chao Gao and Zongming Ma

Abstract. This paper surveys some recent developments in fundamental limits and optimal algorithms for network analysis. We focus on minimax optimal rates in three fundamental problems of network analysis: graphon estimation, community detection and hypothesis testing. For each problem, we review state-of-the-art results in the literature followed by general principles behind the optimal procedures that lead to minimax estimation and testing. This allows us to connect problems in network analysis to other statistical inference problems from a general perspective.

Key words and phrases: Clustering, Minimax rates, polynomial-time algorithms, random graphs, social networks.

REFERENCES

- ABBE, E. (2017). Community detection and stochastic block models: Recent developments. *J. Mach. Learn. Res.* **18** Paper No. 177, 86. [MR3827065](#)
- ABBE, E., BANDEIRA, A. S. and HALL, G. (2016). Exact recovery in the stochastic block model. *IEEE Trans. Inf. Theory* **62** 471–487. [MR3447993](#) <https://doi.org/10.1109/TIT.2015.2490670>
- ABBE, E. and SANDON, C. (2015). Community detection in general stochastic block models: Fundamental limits and efficient algorithms for recovery. In *2015 IEEE 56th Annual Symposium on Foundations of Computer Science—FOCS 2015* 670–688. IEEE Computer Soc., Los Alamitos, CA. [MR3473334](#) <https://doi.org/10.1109/FOCS.2015.47>
- AIROLDI, E. M., COSTA, T. B. and CHAN, S. H. (2013). Stochastic blockmodel approximation of a graphon: Theory and consistent estimation. In *Advances in Neural Information Processing Systems* 692–700.
- ALDOUS, D. J. (1981). Representations for partially exchangeable arrays of random variables. *J. Multivariate Anal.* **11** 581–598. [MR0637937](#) [https://doi.org/10.1016/0047-259X\(81\)90099-3](https://doi.org/10.1016/0047-259X(81)90099-3)
- AMINI, A. A. and LEVINA, E. (2018). On semidefinite relaxations for the block model. *Ann. Statist.* **46** 149–179. [MR3766949](#) <https://doi.org/10.1214/17-AOS1545>
- AMINI, A. A., CHEN, A., BICKEL, P. J. and LEVINA, E. (2013). Pseudo-likelihood methods for community detection in large sparse networks. *Ann. Statist.* **41** 2097–2122. [MR3127859](#) <https://doi.org/10.1214/13-AOS1138>
- BANERJEE, D. (2018). Contiguity and non-reconstruction results for planted partition models: The dense case. *Electron. J. Probab.* **23** Paper No. 18, 28. [MR3771755](#) <https://doi.org/10.1214/17-EJP128>
- BANERJEE, D. and MA, Z. (2017). Optimal hypothesis testing for stochastic block models with growing degrees. Preprint. Available at [arXiv:1705.05305](#).
- BANERJEE, D. and MA, Z. (2018). Non-backtracking matrices and optimal hypothesis testing for stochastic block models with growing degrees.
- BERTHET, Q. and RIGOLLET, P. (2013). Complexity theoretic lower bounds for sparse principal component detection. In *Conference on Learning Theory* 1046–1066.
- BICKEL, P. J. and CHEN, A. (2009). A nonparametric view of network models and Newman–Girvan and other modularities. *Proc. Natl. Acad. Sci. USA* **106** 21068–21073.
- BICKEL, P. J. and SARKAR, P. (2016). Hypothesis testing for automated community detection in networks. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **78** 253–273. [MR3453655](#) <https://doi.org/10.1111/rssb.12117>
- BORGS, C., CHAYES, J. and SMITH, A. (2015). Private graphon estimation for sparse graphs. In *Advances in Neural Information Processing Systems* 1369–1377.
- BORGS, C., CHAYES, J. T., COHN, H. and ZHAO, Y. (2014). An L^p theory of sparse graph convergence I: Limits, sparse random graph models, and power law distributions. Preprint. Available at [arXiv:1401.2906](#).
- BORGS, C., CHAYES, J. T., COHN, H. and GANGULY, S. (2015). Consistent nonparametric estimation for heavy-tailed sparse graphs. Preprint. Available at [arXiv:1508.06675](#).
- BORGS, C., CHAYES, J. T., COHN, H. and HOLDEN, N. (2017). Sparse exchangeable graphs and their limits via graphon processes. *J. Mach. Learn. Res.* **18** Paper No. 210, 71. [MR3827098](#)
- BORGS, C., CHAYES, J. T., COHN, H. and ZHAO, Y. (2018). An L^p theory of sparse graph convergence II: LD convergence, quotients and right convergence. *Ann. Probab.* **46** 337–396. [MR3758733](#) <https://doi.org/10.1214/17-AOP1187>
- BUTUCEA, C., NDAOUD, M., STEPANOVA, N. A. and TSYBAKOV, A. B. (2018). Variable selection with Hamming loss.

Chao Gao is an Assistant Professor, Department of Statistics, University of Chicago, 5747 S Ellis Ave, Jones 314, Chicago, Illinois 60637, USA (e-mail: chaogao@uchicago.edu). Zongming Ma is an Associate Professor, Department of Statistics, University of Pennsylvania, 3730 Walnut Street, Philadelphia, Pennsylvania 19104, USA (e-mail: zongming@wharton.upenn.edu).

- Ann. Statist.* **46** 1837–1875. MR3845003 <https://doi.org/10.1214/17-AOS1572>
- CAI, T. T. and LI, X. (2015). Robust and computationally feasible community detection in the presence of arbitrary outlier nodes. *Ann. Statist.* **43** 1027–1059. MR3346696 <https://doi.org/10.1214/14-AOS1290>
- CAI, T. T., MA, Z. and WU, Y. (2013). Sparse PCA: Optimal rates and adaptive estimation. *Ann. Statist.* **41** 3074–3110. MR3161458 <https://doi.org/10.1214/13-AOS1178>
- CARON, F. and FOX, E. B. (2017). Sparse graphs using exchangeable random measures. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **79** 1295–1366. MR3731666 <https://doi.org/10.1111/rssb.12233>
- CARPENTIER, A. and VERZELEN, N. (2019). Adaptive estimation of the sparsity in the Gaussian vector model. *Ann. Statist.* **47** 93–126. MR3909928 <https://doi.org/10.1214/17-AOS1680>
- CHAN, S. and AIROLDI, E. (2014). A consistent histogram estimator for exchangeable graph models. In *International Conference on Machine Learning* 208–216.
- CHATTERJEE, S. (2015). Matrix estimation by universal singular value thresholding. *Ann. Statist.* **43** 177–214. MR3285604 <https://doi.org/10.1214/14-AOS1272>
- CHAUDHURI, K., CHUNG, F. and TSIRTAS, A. (2012). Spectral clustering of graphs with general degrees in the extended planted partition model. In *Conference on Learning Theory* 35–1.
- CHEN, P. and GAO, C. (2018). Minimax estimation of permutations.
- CHEN, K. and LEI, J. (2018). Network cross-validation for determining the number of communities in network data. *J. Amer. Statist. Assoc.* **113** 241–251. MR3803461 <https://doi.org/10.1080/01621459.2016.1246365>
- CHEN, Y., LI, X. and XU, J. (2018). Convexified modularity maximization for degree-corrected stochastic block models. *Ann. Statist.* **46** 1573–1602. MR3819110 <https://doi.org/10.1214/17-AOS1595>
- CHENG, Y. and CHURCH, G. M. (2000). Biclustering of expression data. In *Proceedings of the Eighth International Conference on Intelligent Systems for Molecular Biology* 93–103. AAAI Press, Menlo Park, CA.
- CHIN, P., RAO, A. and VU, V. (2015). Stochastic block model and community detection in sparse graphs: A spectral algorithm with optimal rate of recovery. In *Conference on Learning Theory* 391–423.
- CHUNG, F. and LU, L. (2002). The average distances in random graphs with given expected degrees. *Proc. Natl. Acad. Sci. USA* **99** 15879–15882. MR1944974 <https://doi.org/10.1073/pnas.252631999>
- COJA-OGHLAN, A. (2010). Graph partitioning via adaptive spectral techniques. *Combin. Probab. Comput.* **19** 227–284. MR2593622 <https://doi.org/10.1017/S0963548309990514>
- COVER, T. M. and THOMAS, J. A. (2006). *Elements of Information Theory*, 2nd ed. Wiley Interscience, Hoboken, NJ. MR2239987
- CRANE, H. (2018). *Probabilistic Foundations of Statistical Network Analysis. Monographs on Statistics and Applied Probability* **157**. CRC Press, Boca Raton, FL. MR3791467
- CRANE, H. and DEMPSEY, W. (2018). Edge exchangeable models for interaction networks. *J. Amer. Statist. Assoc.* **113** 1311–1326. MR3862359 <https://doi.org/10.1080/01621459.2017.1341413>
- DASGUPTA, A., HOPCROFT, J. E. and MCSHERRY, F. (2004). Spectral analysis of random graphs with skewed degree distributions. In *45th Annual IEEE Symposium on Foundations of Computer Science*, 2004. *Proceedings* 602–610. IEEE, Los Alamitos, CA.
- DAWID, A. P. and SKENE, A. M. (1979). Maximum likelihood estimation of observer error-rates using the em algorithm. *Appl. Stat.* **28** 20–28. <https://doi.org/10.2307/2346806>
- DIACONIS, P. and GRAHAM, R. L. (1977). Spearman’s footrule as a measure of disarray. *J. Roy. Statist. Soc. Ser. B* **39** 262–268. MR0652736
- DIACONIS, P. and JANSON, S. (2008). Graph limits and exchangeable random graphs. *Rend. Mat. Appl.* (7) **28** 33–61. MR2463439
- DONOHU, D. L. and JOHNSTONE, I. M. (1994). Minimax risk over l_p -balls for l_q -error. *Probab. Theory Related Fields* **99** 277–303. MR1278886 <https://doi.org/10.1007/BF01199026>
- FAN, Z. and MONTANARI, A. (2017). How well do local algorithms solve semidefinite programs? In *STOC’17—Proceedings of the 49th Annual ACM SIGACT Symposium on Theory of Computing* 604–614. ACM, New York. MR3678214 <https://doi.org/10.1145/3055399.3055451>
- FEI, Y. and CHEN, Y. (2019). Exponential error rates of SDP for block models: Beyond Grothendieck’s inequality. *IEEE Trans. Inf. Theory* **65** 551–571. MR3901009 <https://doi.org/10.1109/TIT.2018.2839677>
- GAO, C. (2017). Phase transitions in approximate ranking. Preprint. Available at [arXiv:1711.11189](https://arxiv.org/abs/1711.11189).
- GAO, C. and LAFFERTY, J. (2017). Testing for global network structure using small subgraph statistics. Preprint. Available at [arXiv:1710.00862](https://arxiv.org/abs/1710.00862).
- GAO, C., LU, Y. and ZHOU, H. H. (2015). Rate-optimal graphon estimation. *Ann. Statist.* **43** 2624–2652. MR3405606 <https://doi.org/10.1214/15-AOS1354>
- GAO, C., LU, Y. and ZHOU, D. (2016). Exact exponent in optimal rates for crowdsourcing. In *International Conference on Machine Learning* 603–611.
- GAO, C., MA, Z. and ZHOU, H. H. (2017). Sparse CCA: Adaptive estimation and computational barriers. *Ann. Statist.* **45** 2074–2101. MR3718162 <https://doi.org/10.1214/16-AOS1519>
- GAO, C., VAN DER VAART, A. W. and ZHOU, H. H. (2020). A general framework for Bayes structured linear models. *Ann. Statist.* **48** 2848–2878. MR4152123 <https://doi.org/10.1214/19-AOS1909>
- GAO, C., MA, Z., REN, Z. and ZHOU, H. H. (2015). Minimax estimation in sparse canonical correlation analysis. *Ann. Statist.* **43** 2168–2197. MR3396982 <https://doi.org/10.1214/15-AOS1332>
- GAO, C., LU, Y., MA, Z. and ZHOU, H. H. (2016). Optimal estimation and completion of matrices with biclustering structures. *J. Mach. Learn. Res.* **17** Paper No. 161, 29. MR3569248
- GAO, C., MA, Z., ZHANG, A. Y. and ZHOU, H. H. (2017). Achieving optimal misclassification proportion in stochastic block models. *J. Mach. Learn. Res.* **18** Paper No. 60, 45. MR3687603
- GAO, C., MA, Z., ZHANG, A. Y. and ZHOU, H. H. (2018). Community detection in degree-corrected block models. *Ann. Statist.* **46** 2153–2185. MR3845014 <https://doi.org/10.1214/17-AOS1615>
- GIRVAN, M. and NEWMAN, M. E. J. (2002). Community structure in social and biological networks. *Proc. Natl. Acad. Sci. USA* **99** 7821–7826. MR1908073 <https://doi.org/10.1073/pnas.122653799>
- GOLDENBERG, A., ZHENG, A. X., FIENBERG, S. E. and AIROLDI, E. M. (2010). A survey of statistical network models. *Found. Trends Mach. Learn.* **2** 129–233.
- GULIKERS, L., LELARGE, M. and MASSOULIÉ, L. (2017). A spectral method for community detection in moderately sparse degree-corrected stochastic block models. *Adv. in Appl. Probab.* **49** 686–721. MR3694314 <https://doi.org/10.1017/apr.2017.18>
- GULIKERS, L., LELARGE, M. and MASSOULIÉ, L. (2018). An impossibility result for reconstruction in the degree-corrected stochastic block model. *Ann. Appl. Probab.* **28** 3002–3027. MR3847979 <https://doi.org/10.1214/18-AAP1381>
- HAJEK, B., WU, Y. and XU, J. (2016a). Achieving exact cluster recovery threshold via semidefinite programming. *IEEE Trans. Inf. Theory* **62** 2788–2797. MR3493879 <https://doi.org/10.1109/TIT.2016.2546280>
- HAJEK, B., WU, Y. and XU, J. (2016b). Achieving exact cluster recovery threshold via semidefinite programming: Extensions. *IEEE Trans. Inf. Theory* **62** 5918–5937. MR3552431 <https://doi.org/10.1109/TIT.2016.2594812>

- HARTIGAN, J. A. (1972). Direct clustering of a data matrix. *J. Amer. Statist. Assoc.* **67** 123–129.
- HERLAU, T., SCHMIDT, M. N. and MØRUP, M. (2016). Completely random measures for modelling block-structured sparse networks. In *Advances in Neural Information Processing Systems* 4260–4268.
- HOFF, P. D., RAFTERY, A. E. and HANDCOCK, M. S. (2002). Latent space approaches to social network analysis. *J. Amer. Statist. Assoc.* **97** 1090–1098. [MR1951262](#) <https://doi.org/10.1198/016214502388618906>
- HOLLAND, P. W., LASKEY, K. B. and LEINHARDT, S. (1983). Stochastic blockmodels: First steps. *Soc. Netw.* **5** 109–137. [MR0718088](#) [https://doi.org/10.1016/0378-8733\(83\)90021-7](https://doi.org/10.1016/0378-8733(83)90021-7)
- HOOVER, D. N. (1979). Relations on probability spaces and arrays of random variables 2. Institute for Advanced Study, Princeton, NJ. Preprint.
- JANSON, S. (1995). Random regular graphs: Asymptotic distributions and contiguity. *Combin. Probab. Comput.* **4** 369–405. [MR1377557](#) <https://doi.org/10.1017/S0963548300001735>
- JIN, J. (2015). Fast community detection by SCORE. *Ann. Statist.* **43** 57–89. [MR3285600](#) <https://doi.org/10.1214/14-AOS1265>
- JIN, J., KE, Z. T. and LUO, S. (2018). Network global testing by counting graphlets. In *International Conference on Machine Learning* 2338–2346.
- JOSEPH, A. and YU, B. (2016). Impact of regularization on spectral clustering. *Ann. Statist.* **44** 1765–1791. [MR3519940](#) <https://doi.org/10.1214/16-AOS1447>
- KALLENBERG, O. (1989). On the representation theorem for exchangeable arrays. *J. Multivariate Anal.* **30** 137–154. [MR1003713](#) [https://doi.org/10.1016/0047-259X\(89\)90092-4](https://doi.org/10.1016/0047-259X(89)90092-4)
- KARRER, B. and NEWMAN, M. E. J. (2011). Stochastic blockmodels and community structure in networks. *Phys. Rev. E* (3) **83** 016107, 10. [MR2788206](#) <https://doi.org/10.1103/PhysRevE.83.016107>
- KLOPP, O., TSYBAKOV, A. B. and VERZELEN, N. (2017). Oracle inequalities for network models and sparse graphon estimation. *Ann. Statist.* **45** 316–354. [MR3611494](#) <https://doi.org/10.1214/16-AOS1454>
- KLOPP, O., LU, Y., TSYBAKOV, A. B. and ZHOU, H. H. (2019). Structured matrix estimation and completion. *Bernoulli* **25** 3883–3911. [MR4010976](#) <https://doi.org/10.3150/19-bej1114>
- LE, C. M., LEVINA, E. and VERSHYNIN, R. (2015). Sparse random graphs: regularization and concentration of the laplacian. *Random Structures Algorithms* Preprint. Available at [arXiv:1502.03049](#).
- LE, C. M., LEVINA, E. and VERSHYNIN, R. (2017). Concentration and regularization of random graphs. *Random Structures Algorithms* **51** 538–561. [MR3689343](#) <https://doi.org/10.1002/rsa.20713>
- LEI, J. and RINALDO, A. (2015). Consistency of spectral clustering in stochastic block models. *Ann. Statist.* **43** 215–237. [MR3285605](#) <https://doi.org/10.1214/14-AOS1274>
- LLOYD, J., ORBANZ, P., GHAHRAMANI, Z. and ROY, D. M. (2012). Random function priors for exchangeable arrays with applications to graphs and relational data. In *Advances in Neural Information Processing Systems* 998–1006.
- LOVÁSZ, L. (2012). *Large Networks and Graph Limits*. American Mathematical Society Colloquium Publications **60**. Amer. Math. Soc., Providence, RI. [MR3012035](#) <https://doi.org/10.1090/coll/060>
- MA, Z., MA, Z. and YUAN, H. (2020). Universal latent space model fitting for large networks with edge covariates. *J. Mach. Learn. Res.* **21** 1–67.
- MA, Z. and WU, Y. (2015a). Computational barriers in minimax submatrix detection. *Ann. Statist.* **43** 1089–1116. [MR3346698](#) <https://doi.org/10.1214/14-AOS1300>
- MA, Z. and WU, Y. (2015b). Volume ratio, sparsity, and minimaxity under unitarily invariant norms. *IEEE Trans. Inf. Theory* **61** 6939–6956. [MR3430731](#) <https://doi.org/10.1109/TIT.2015.2487541>
- MADEIRA, S. C. and OLIVEIRA, A. L. (2004). Biclustering algorithms for biological data analysis: A survey. *IEEE/ACM Trans. Comput. Biol. Bioinform.* **1** 24–45.
- MASSOULIÉ, L. (2014). Community detection thresholds and the weak Ramanujan property. In *STOC’14—Proceedings of the 2014 ACM Symposium on Theory of Computing* 694–703. ACM, New York. [MR3238997](#)
- MCCULLAGH, P. (2002). What is a statistical model? *Ann. Statist.* **30** 1225–1310. [MR1936320](#) <https://doi.org/10.1214/aos/1035844977>
- MCSherry, F. (2001). Spectral partitioning of random graphs. In *42nd IEEE Symposium on Foundations of Computer Science (Las Vegas, NV, 2001)* 529–537. IEEE Computer Soc., Los Alamitos, CA. [MR1948742](#)
- MONTANARI, A. and SEN, S. (2016). Semidefinite programs on sparse random graphs and their application to community detection. In *STOC’16—Proceedings of the 48th Annual ACM SIGACT Symposium on Theory of Computing* 814–827. ACM, New York. [MR3536616](#) <https://doi.org/10.1145/2897518.2897548>
- MOSSEL, E., NEEMAN, J. and SLY, A. (2014). Consistency thresholds for binary symmetric block models. Preprint. Available at [arXiv:1407.1591](#).
- MOSSEL, E., NEEMAN, J. and SLY, A. (2015). Reconstruction and estimation in the planted partition model. *Probab. Theory Related Fields* **162** 431–461. [MR3383334](#) <https://doi.org/10.1007/s00440-014-0576-6>
- MOSSEL, E., NEEMAN, J. and SLY, A. (2018). A proof of the block model threshold conjecture. *Combinatorica* **38** 665–708. [MR3876880](#) <https://doi.org/10.1007/s00493-016-3238-8>
- NEWMAN, M. E. J. (2010). *Networks: An Introduction*. Oxford Univ. Press, Oxford. [MR2676073](#) <https://doi.org/10.1093/acprof:oso/9780199206650.001.0001>
- NEWMAN, M. E., WATTS, D. J. and STROGATZ, S. H. (2002). Random graph models of social networks. *Proc. Natl. Acad. Sci. USA* **99** 2566–2572.
- NOWICKI, K. and SNIJDERS, T. A. B. (2001). Estimation and prediction for stochastic blockstructures. *J. Amer. Statist. Assoc.* **96** 1077–1087. [MR1947255](#) <https://doi.org/10.1198/016214501753208735>
- OLHEDE, S. C. and WOLFE, P. J. (2014). Network histograms and universality of blockmodel approximation. *Proc. Natl. Acad. Sci. USA* **111** 14722–14727.
- QIN, T. and ROHE, K. (2013). Regularized spectral clustering under the degree-corrected stochastic blockmodel. In *Advances in Neural Information Processing Systems* 3120–3128.
- RASKUTTI, G., WAINWRIGHT, M. J. and YU, B. (2011). Minimax rates of estimation for high-dimensional linear regression over ℓ_q -balls. *IEEE Trans. Inf. Theory* **57** 6976–6994. [MR2882274](#) <https://doi.org/10.1109/TIT.2011.2165799>
- ROHE, K., CHATTERJEE, S. and YU, B. (2011). Spectral clustering and the high-dimensional stochastic blockmodel. *Ann. Statist.* **39** 1878–1915. [MR2893856](#) <https://doi.org/10.1214/11-AOS887>
- SARKAR, P. and BICKEL, P. J. (2015). Role of normalization in spectral clustering for stochastic blockmodels. *Ann. Statist.* **43** 962–990. [MR3346694](#) <https://doi.org/10.1214/14-AOS1285>
- SHI, J. and MALIK, J. (2000). Normalized cuts and image segmentation. *IEEE Trans. Pattern Anal. Mach. Intell.* **22** 888–905.
- TODESCHINI, A., MISCOURIDOU, X. and CARON, F. (2020). Exchangeable random measures for sparse and modular graphs with overlapping communities. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **82** 487–520. [MR4084173](#) <https://doi.org/10.1111/rssb.12363>
- VAN DER HOFSTAD, R. (2017). *Random Graphs and Complex Networks*. Vol. 1. Cambridge Series in Statistical and Probabilistic Mathematics. Cambridge Univ. Press, Cambridge. [MR3617364](#) <https://doi.org/10.1017/9781316779422>
- VEITCH, V. (2015). The class of random graphs arising from exchangeable random measures. Preprint. Available at [arXiv:1512.03099](#).

- VON LUXBURG, U. (2007). A tutorial on spectral clustering. *Stat. Comput.* **17** 395–416. [MR2409803](#) <https://doi.org/10.1007/s11222-007-9033-z>
- VON LUXBURG, U., BELKIN, M. and BOUSQUET, O. (2008). Consistency of spectral clustering. *Ann. Statist.* **36** 555–586. [MR2396807](#) <https://doi.org/10.1214/0090536070000000640>
- WASSERMAN, S. and FAUST, K. (1994). *Social Network Analysis: Methods and Applications* **8**. Cambridge University Press, Cambridge.
- WOLFE, P. J. and OLHEDE, S. C. (2013). Nonparametric graphon estimation. Preprint. Available at [arXiv:1309.5936](#).
- XU, J. (2018). Rates of convergence of spectral methods for graphon estimation. In *International Conference on Machine Learning* 5433–5442. PMLR.
- YE, F. and ZHANG, C.-H. (2010). Rate minimaxity of the Lasso and Dantzig selector for the ℓ_q loss in ℓ_r balls. *J. Mach. Learn. Res.* **11** 3519–3540. [MR2756192](#)
- YUN, S.-Y. and PROUTIERE, A. (2014). Accurate community detection in the stochastic block model via spectral algorithms. Preprint. Available at [arXiv:1412.7335](#).
- YUN, S.-Y. and PROUTIERE, A. (2016). Optimal cluster recovery in the labeled stochastic block model. In *Advances in Neural Information Processing Systems* 965–973.
- ZHANG, A. Y. and ZHOU, H. H. (2016). Minimax rates of community detection in stochastic block models. *Ann. Statist.* **44** 2252–2280. [MR3546450](#) <https://doi.org/10.1214/15-AOS1428>
- ZHANG, A. Y. and ZHOU, H. H. (2020). Theoretical and computational guarantees of mean field variational inference for community detection. *Ann. Statist.* **48** 2575–2598. [MR4152113](#) <https://doi.org/10.1214/19-AOS1898>
- ZHAO, Y., LEVINA, E. and ZHU, J. (2012). Consistency of community detection in networks under degree-corrected stochastic block models. *Ann. Statist.* **40** 2266–2292. [MR3059083](#) <https://doi.org/10.1214/12-AOS1036>
- ZHOU, Z. and AMINI, A. A. (2019). Analysis of spectral clustering algorithms for community detection: The general bipartite setting. *J. Mach. Learn. Res.* **20** Paper No. 47, 47. [MR3948087](#)
- ZHOU, Z. and LI, P. (2018). Non-asymptotic chernoff lower bound and its application to community detection in stochastic block model. Preprint. Available at [arXiv:1812.11269](#).
- AMAZON MECHANICAL TURK (2010). Available at <https://www.mturk.com/mturk/welcome>.

Additive and Multiplicative Effects Network Models

Peter Hoff

Abstract. Network datasets typically exhibit certain types of statistical patterns, such as within-dyad correlation, degree heterogeneity, and triadic patterns such as transitivity and clustering. The first two of these can be well represented with a social relations model, a type of additive effects model originally developed for continuous dyadic data. Higher-order patterns can be represented with multiplicative effects models, which are related to matrix decompositions that are commonly used for matrix-variate data analysis. Additionally, these multiplicative effects models generalize other popular latent feature network models, such as the stochastic blockmodel and the latent space model. In this article, we review a general regression framework for the analysis of network data that combines these two types of effects, and accommodates a variety of network data types, including continuous, binary and ordinal network relations.

Key words and phrases: Bayesian, factor model, generalized linear model, latent variable, matrix decomposition, mixed effects model.

REFERENCES

- AIROLDI, E. M., BLEI, D. M., FIENBERG, S. E. and XING, E. P. (2008). Mixed membership stochastic blockmodels. *J. Mach. Learn. Res.* **9** 1981–2014.
- ALBERT, J. H. and CHIB, S. (1993). Bayesian analysis of binary and polychotomous response data. *J. Amer. Statist. Assoc.* **88** 669–679. [MR1224394](#)
- AMINI, A. A., CHEN, A., BICKEL, P. J. and LEVINA, E. (2013). Pseudo-likelihood methods for community detection in large sparse networks. *Ann. Statist.* **41** 2097–2122. [MR3127859](#) <https://doi.org/10.1214/13-AOS1138>
- ASTA, D. M. and SHALIZI, C. R. (2015). Geometric network comparisons. In *Proceedings of the Thirty-First Conference on Uncertainty in Artificial Intelligence, UAI'15* 102–110. AUAI Press, Arlington, VA.
- BAIER, S. L. and BERGSTRAND, J. H. (2009). Bonus vetus ols: A simple method for approximating international trade-cost effects using the gravity equation. *Journal of International Economics* **77** 77–85.
- BERGSTRAND, J. H. (1985). The gravity equation in international trade: Some microeconomic foundations and empirical evidence. *Rev. Econ. Stat.* **67** 474–481.
- BICKEL, P. J., CHEN, A. and LEVINA, E. (2011). The method of moments and degree distributions for network models. *Ann. Statist.* **39** 2280–2301. [MR2906868](#) <https://doi.org/10.1214/11-AOS904>
- BICKEL, P. J. and RITOV, Y. (1997). Local asymptotic normality of ranks and covariates in transformation models. In *Festschrift for Lucien Le Cam* 43–54. Springer, New York. [MR1462938](#)
- BOND, C. F. and LASHLEY, B. R. (1996). Round-robin analysis of social interaction: Exact and estimated standard errors. *Psychometrika* **61** 303–311.
- BRADU, D. and GABRIEL, K. R. (1974). Simultaneous statistical inference on interactions in two-way analysis of variance. *J. Amer. Statist. Assoc.* **69** 428–436. [MR0359192](#)
- CELISSE, A., DAUDIN, J.-J. and PIERRE, L. (2012). Consistency of maximum-likelihood and variational estimators in the stochastic block model. *Electron. J. Stat.* **6** 1847–1899. [MR2988467](#) <https://doi.org/10.1214/12-EJS729>
- DORFF, C., GALLOP, M. and MINHAS, S. (2019). Networks of violence: Predicting conflict in Nigeria. *The Journal of Politics*. forthcoming.
- DURANTE, D., DUNSON, D. B. and VOGELSTEIN, J. T. (2017). Non-parametric Bayes modeling of populations of networks. *J. Amer. Statist. Assoc.* **112** 1516–1530. [MR3750873](#) <https://doi.org/10.1080/01621459.2016.1219260>
- ECKART, C. and YOUNG, G. (1936). The approximation of one matrix by another of lower rank. *Psychometrika* **1** 211–218.
- GABRIEL, K. R. (1998). Generalised bilinear regression. *Biometrika* **85** 689–700. [MR1665834](#) <https://doi.org/10.1093/biomet/85.3.689>
- GADE, E. K., HAFEZ, M. M. and GABBAY, M. (2019). Fratricide in rebel movements: A network analysis of syrian militant infighting. *Journal of Peace Research* **56** 321–335.
- GENOVART, M., GIMENEZ, O., BERTOLERO, A., CHOQUET, R., ORO, D. and PRADEL, R. (2019). Changes in social cohesion in a long-lived species under a perturbation regime. *bioRxiv*.
- GILL, P. S. and SWARTZ, T. B. (2001). Statistical analyses for round robin interaction data. *Canad. J. Statist.* **29** 321–331. [MR1840712](#) <https://doi.org/10.2307/3316080>
- GOLLOB, H. F. (1968). A statistical model which combines features of factor analytic and analysis of variance techniques. *Psychometrika* **33** 73–115. [MR0221658](#) <https://doi.org/10.1007/BF02289676>

- HANDCOCK, M. S. (2003). Assessing degeneracy in statistical models of social networks Technical Report 39, Center for Statistics and the Social Sciences, Univ. Washington, Seattle, WA.
- HARRIS, K., HALPERN, C., WHITSEL, E., HUSSEY, J., TABOR, J., ENTZEL, P. and UDRY, J. (2009). The national longitudinal study of adolescent health: Research design.
- HOFF, P. D. (2005). Bilinear mixed-effects models for dyadic data. *J. Amer. Statist. Assoc.* **100** 286–295. MR2156838 <https://doi.org/10.1198/016214504000001015>
- HOFF, P. (2008). Modeling homophily and stochastic equivalence in symmetric relational data. In *Advances in Neural Information Processing Systems* (J. Platt, D. Koller, Y. Singer and S. Roweis, eds.) **20** 657–664. MIT Press, Cambridge, MA.
- HOFF, P. D. (2009a). *A First Course in Bayesian Statistical Methods*. Springer Texts in Statistics. Springer, New York. MR2648134 <https://doi.org/10.1007/978-0-387-92407-6>
- HOFF, P. D. (2009b). Multiplicative latent factor models for description and prediction of social networks. *Comput. Math. Organ. Theory* **15** 261–272.
- HOFF, P. D. (2011). Hierarchical multilinear models for multi-way data. *Comput. Statist. Data Anal.* **55** 530–543. MR2736574 <https://doi.org/10.1016/j.csda.2010.05.020>
- HOFF, P. (2015). Dyadic data analysis with amen Technical Report 638, Dept. Statistics, Univ. Washington, Seattle, WA.
- HOFF, P. D. (2016). Equivariant and scale-free Tucker decomposition models. *Bayesian Anal.* **11** 627–648. MR3498040 <https://doi.org/10.1214/14-BA934>
- HOFF, P. D., RAFTERY, A. E. and HANDCOCK, M. S. (2002). Latent space approaches to social network analysis. *J. Amer. Statist. Assoc.* **97** 1090–1098. MR1951262 <https://doi.org/10.1198/016214502388618906>
- HOFF, P., FOSDICK, B., VOLFOVSKY, A. and HE, Y. (2012). amen: Additive and multiplicative effects modeling of networks and relational data.
- HOFF, P., FOSDICK, B., VOLFOVSKY, A. and STOVEL, K. (2013). Likelihoods for fixed rank nomination networks. *Netw. Sci.* **1** 253–277.
- HOLLAND, P. W. and LEINHARDT, S. (1981). An exponential family of probability distributions for directed graphs. *J. Amer. Statist. Assoc.* **76** 33–65. MR0608176
- HUNTER, D. R., GOODREAU, S. M. and HANDCOCK, M. S. (2008). Goodness of fit of social network models. *J. Amer. Statist. Assoc.* **103** 248–258. MR2394635 <https://doi.org/10.1198/016214507000000446>
- HUNTER, D. R. and HANDCOCK, M. S. (2006). Inference in curved exponential family models for networks. *J. Comput. Graph. Statist.* **15** 565–583. MR2291264 <https://doi.org/10.1198/106186006X133069>
- ISARD, W. (1954). Location theory and trade theory: Short-run analysis. *Q. J. Econ.* **68** 305–320.
- KARRER, B. and NEWMAN, M. E. J. (2011). Stochastic blockmodels and community structure in networks. *Phys. Rev. E* (3) **83** 016107, 10. MR2788206 <https://doi.org/10.1103/PhysRevE.83.016107>
- KOSTER, J. (2018). Family ties: The multilevel effects of households and kinship on the networks of individuals. *R. Soc. Open Sci.* **5** 172159. <https://doi.org/10.1098/rsos.172159>
- KRIVITSKY, P. N., HANDCOCK, M. S., RAFTERY, A. E. and HOFF, P. D. (2009). Representing degree distributions, clustering, and homophily in social networks with latent cluster random effects models. *Soc. Netw.* **31** 204–213.
- LI, H. and LOKEN, E. (2002). A unified theory of statistical analysis and inference for variance component models for dyadic data. *Statist. Sinica* **12** 519–535. MR1902723
- LINDSAY, B. G. (1988). Composite likelihood methods. In *Statistical Inference from Stochastic Processes* (Ithaca, NY, 1987). *Contemp. Math.* **80** 221–239. Amer. Math. Soc., Providence, RI. MR0999014 <https://doi.org/10.1090/conm/080/999014>
- LUSHER, D., KOSKINEN, J. and ROBINS, G. (2013). *Exponential Random Graph Models for Social Networks: Theory, Methods, and Applications*. Cambridge University Press, Cambridge.
- NOWICKI, K. and SNIJDERS, T. A. B. (2001). Estimation and prediction for stochastic blockstructures. *J. Amer. Statist. Assoc.* **96** 1077–1087. MR1947255 <https://doi.org/10.1198/016214501753208735>
- PERRY, P. O. (2017). Fast moment-based estimation for hierarchical models. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **79** 267–291. MR3597973 <https://doi.org/10.1111/rssb.12165>
- PINHEIRO, J. C. and BATES, D. M. (2000). *Mixed-Effects Models in S and S-PLUS*. Springer, Berlin.
- PRICE, D. (1976). A general theory of bibliometric and other cumulative advantage processes. *Journal of the Association for Information Science and Technology* **27** 292–306.
- RAFTERY, A. E., NIU, X., HOFF, P. D. and YEUNG, K. Y. (2012). Fast inference for the latent space network model using a case-control approximate likelihood. *J. Comput. Graph. Statist.* **21** 901–919. MR3005803 <https://doi.org/10.1080/10618600.2012.679240>
- RASTELLI, R., FRIEL, N. and RAFTERY, A. E. (2016). Properties of latent variable network models. *Netw. Sci.* **4** 407–432.
- ROHE, K., CHATTERJEE, S. and YU, B. (2011). Spectral clustering and the high-dimensional stochastic blockmodel. *Ann. Statist.* **39** 1878–1915. MR2893856 <https://doi.org/10.1214/11-AOS887>
- SALTER-TOWNSHEND, M. and MURPHY, T. B. (2013). Variational Bayesian inference for the latent position cluster model for network data. *Comput. Statist. Data Anal.* **57** 661–671. MR2981116 <https://doi.org/10.1016/j.csda.2012.08.004>
- SCHWEINBERGER, M. and HANDCOCK, M. S. (2015). Local dependence in random graph models: Characterization, properties and statistical inference. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **77** 647–676. MR3351449 <https://doi.org/10.1111/rssb.12081>
- SCHWEINBERGER, M. and SNIJDERS, T. A. B. (2003). Settings in social networks: A measurement model. *Sociol. Method.* **33** 307–341.
- SCHWEINBERGER, M., KRIVITSKY, P. N., BUTTS, C. T. and STEWART, J. (2017). Exponential-family models of random graphs: Inference in finite-, super-, and infinite population scenarios. Preprint.
- SNIJDERS, T. A., PATTISON, P. E., ROBINS, G. L. and HANDCOCK, M. S. (2006). New specifications for exponential random graph models. *Sociol. Method.* **36** 99–153.
- VAN DUIN, M. A. J., SNIJDERS, T. A. B. and ZIJLSTRA, B. J. H. (2004). p_2 : A random effects model with covariates for directed graphs. *Stat. Neerl.* **58** 234–254. MR2064846 <https://doi.org/10.1046/j.0039-0402.2003.00258.x>
- WALKER, S. C., BOLKER, B. M., MÄCHLER, M. and BATES, D. (2015). Fitting linear mixed-effects models using lme4. *J. Stat. Softw.* **67** 1–48.
- WANG, L., ZHANG, Z. and DUNSON, D. (2019). Common and individual structure of brain networks. *Ann. Appl. Stat.* **13** 85–112. MR3937422 <https://doi.org/10.1214/18-AOAS1193>
- WARNER, R., KENNY, D. A. and STOTO, M. (1979). A new round robin analysis of variance for social interaction data. *J. Pers. Soc. Psychol.* **37** 1742–1757.
- WASSERMAN, S. and FAUST, K. (1994). *Social Network Analysis: Methods and Applications*. Cambridge Univ. Press, Cambridge.
- WASSERMAN, S. and PATTISON, P. (1996). Logit models and logistic regressions for social networks. I. An introduction to Markov graphs and p . *Psychometrika* **61** 401–425. MR1424909 <https://doi.org/10.1007/BF02294547>
- WONG, G. Y. (1982). Round robin analysis of variance via maximum likelihood. *J. Amer. Statist. Assoc.* **77** 714–724. MR0686400

YOUNG, S. J. and SCHEINERMAN, E. R. (2007). Random dot product graph models for social networks. In *Algorithms and Models for the Web-Graph. Lecture Notes in Computer Science* **4863** 138–149. Springer, Berlin. [MR2504912](#) https://doi.org/10.1007/978-3-540-77004-6_11

ZHAO, Y., LEVINA, E. and ZHU, J. (2012). Consistency of community detection in networks under degree-corrected stochastic block models. *Ann. Statist.* **40** 2266–2292. [MR3059083](#) <https://doi.org/10.1214/12-AOS1036>

A Statistical Framework for Modern Network Science

Harry Crane and Walter Dempsey

Abstract. We discuss how sampling design, units, the observation mechanism and other basic statistical notions figure into modern network data analysis. These considerations pose several new challenges that cannot be adequately addressed by merely extending or generalizing classical methods. Such challenges stem from fundamental differences between the domains in which network data emerge and those for which classical tools were developed. By revisiting these basic statistical considerations, we suggest a framework in which to develop theory and methods for network analysis in a way that accounts for both conceptual and practical challenges of network science. We then discuss how some well-known model classes fit within this framework.

Key words and phrases: Network data, sparse network, scale-free network, edge exchangeable network, relational exchangeability, relative exchangeability, data generating process, network sampling.

REFERENCES

- [1] ACHLIOPTAS, D., CLAUSET, A., KEMPE, D. and MOORE, C. (2005). On the bias of traceroute sampling or, power-law degree distributions in regular graphs. In *STOC'05: Proceedings of the 37th Annual ACM Symposium on Theory of Computing* 694–703. ACM, New York. [MR2181674](#) <https://doi.org/10.1145/1060590.1060693>
- [2] ALDOUS, D. J. (1981). Representations for partially exchangeable arrays of random variables. *J. Multivariate Anal.* **11** 581–598. [MR0637937](#) [https://doi.org/10.1016/0047-259X\(81\)90099-3](https://doi.org/10.1016/0047-259X(81)90099-3)
- [3] ALDOUS, D. J. (1985). Exchangeability and related topics. In *École D'été de Probabilités de Saint-Flour, XIII—1983. Lecture Notes in Math.* **1117** 1–198. Springer, Berlin. [MR0883646](#) <https://doi.org/10.1007/BFb0099421>
- [4] BICKEL, P. and CHEN, A. (2009). A nonparametric view of network models and Newman–Girvan and other modularities. *Proc. Natl. Acad. Sci. USA* **106** 21068–21073.
- [5] BORGS, C., CHAYES, J. T., COHN, H. and HOLDEN, N. (2017). Sparse exchangeable graphs and their limits via graphon processes. *J. Mach. Learn. Res.* **18** Paper No. 210, 71. [MR3827098](#)
- [6] BOX, G. E. P. (1976). Science and statistics. *J. Amer. Statist. Assoc.* **71** 791–799. [MR0431440](#)
- [7] BURKE, C. J. and ROSENBLATT, M. (1958). A Markovian function of a Markov chain. *Ann. Math. Stat.* **29** 1112–1122. [MR0101557](#) <https://doi.org/10.1214/aoms/1177706444>
- [8] CARON, F. and FOX, E. B. (2017). Sparse graphs using exchangeable random measures. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **79** 1295–1366. [MR3731666](#) <https://doi.org/10.1111/rssb.12233>
- [9] CHEN, K. and LEI, J. (2018). Network cross-validation for determining the number of communities in network data. *J. Amer. Statist. Assoc.* **113** 241–251. [MR3803461](#) <https://doi.org/10.1080/01621459.2016.1246365>
- [10] COX, D. R. and HINKLEY, D. V. (1974). *Theoretical Statistics*. CRC Press, London. [MR0370837](#)
- [11] CRANE, H. (2015). Time-varying network models. *Bernoulli* **21** 1670–1696. [MR3352057](#) <https://doi.org/10.3150/14-BEJ617>
- [12] CRANE, H. (2016). The ubiquitous Ewens sampling formula. *Statist. Sci.* **31** 1–19. [MR3458585](#) <https://doi.org/10.1214/15-STS529>
- [13] CRANE, H. (2017). Exchangeable graph-valued Feller processes. *Probab. Theory Related Fields* **168** 849–899. [MR3663633](#) <https://doi.org/10.1007/s00440-016-0726-0>
- [14] CRANE, H. (2018). Combinatorial Lévy processes. *Ann. Appl. Probab.* **28** 285–339. [MR3770878](#) <https://doi.org/10.1214/17-AAP1306>
- [15] CRANE, H. (2018). *Probabilistic Foundations of Statistical Network Analysis. Monographs on Statistics and Applied Probability* **157**. CRC Press, Boca Raton, FL. [MR3791467](#)
- [16] CRANE, H. and DEMPSEY, W. (2018). Edge exchangeable models for interaction networks. *J. Amer. Statist. Assoc.* **113** 1311–1326. [MR3862359](#) <https://doi.org/10.1080/01621459.2017.1341413>
- [17] CRANE, H. and DEMPSEY, W. (2019). Relational exchangeability. *J. Appl. Probab.* **56** 192–208. [MR3981153](#) <https://doi.org/10.1017/jpr.2019.13>
- [18] CRANE, H. and TOWNSNER, H. (2018). Relatively exchangeable structures. *J. Symbolic Logic* **83** 416–442. [MR3835071](#) <https://doi.org/10.1017/jsl.2017.61>

Harry Crane is Associate Professor, Department of Statistics & Biostatistics, Rutgers University, 110 Frelinghuysen Road, Piscataway, New Jersey 08854, USA (e-mail: hcrane@stat.rutgers.edu). Walter Dempsey is Assistant Professor, Department of Statistics, Harvard University One Oxford Street, Cambridge, Massachusetts 02138, USA (e-mail: wdempsey@uchicago.edu).

- [19] DE FINETTI, B. (1937). La prévision: Ses lois logiques, ses sources subjectives. *Ann. Inst. Henri Poincaré* **7** 1–68. [MR1508036](#)
- [20] DEMPSEY, W., OSELIO, B. and HERO, A. (2019). Hierarchical network models for structured exchangeable interaction processes. Available at [arXiv:1901.09982](#).
- [21] DRTON, M. and SULLIVANT, S. (2007). Algebraic statistical models. *Statist. Sinica* **17** 1273–1297. [MR2398596](#)
- [22] FENG, S. (2010). *The Poisson–Dirichlet Distribution and Related Topics: Models and Asymptotic Behaviors. Probability and Its Applications (New York)*. Springer, Heidelberg. [MR2663265](#) <https://doi.org/10.1007/978-3-642-11194-5>
- [23] FRANK, O. and STRAUSS, D. (1986). Markov graphs. *J. Amer. Statist. Assoc.* **81** 832–842. [MR0860518](#)
- [24] GELMAN, A., MENG, X.-L. and STERN, H. (1996). Posterior predictive assessment of model fitness via realized discrepancies. *Statist. Sinica* **6** 733–807. [MR1422404](#)
- [25] HANDCOCK, M. S. and GILE, K. J. (2010). Modeling social networks from sampled data. *Ann. Appl. Stat.* **4** 5–25. [MR2758082](#) <https://doi.org/10.1214/08-AOAS221>
- [26] HELLAND, I. S. (2006). Extended statistical modeling under symmetry; the link toward quantum mechanics. *Ann. Statist.* **34** 42–77. [MR2275234](#) <https://doi.org/10.1214/009053605000000868>
- [27] HOFF, P. D., RAFTERY, A. E. and HANDCOCK, M. S. (2002). Latent space approaches to social network analysis. *J. Amer. Statist. Assoc.* **97** 1090–1098. [MR1951262](#) <https://doi.org/10.1198/016214502388618906>
- [28] HOLLAND, P. W., LASKEY, K. B. and LEINHARDT, S. (1983). Stochastic blockmodels: First steps. *Soc. Netw.* **5** 109–137. [MR0718088](#) [https://doi.org/10.1016/0378-8733\(83\)90021-7](https://doi.org/10.1016/0378-8733(83)90021-7)
- [29] HOLLAND, P. W. and LEINHARDT, S. (1981). An exponential family of probability distributions for directed graphs. *J. Amer. Statist. Assoc.* **76** 33–65. [MR0608176](#)
- [30] HOOVER, D. (1979). Relations on probability spaces and arrays of random variables. Institute for Advanced Studies. Preprint.
- [31] KHABBAZIAN, M., HANLON, B., RUSSEK, Z. and ROHE, K. (2017). Novel sampling design for respondent-driven sampling. *Electron. J. Stat.* **11** 4769–4812. [MR3729659](#) <https://doi.org/10.1214/17-EJS1358>
- [32] KLUSOWSKI, J. and WU, Y. (2018). Counting motifs with graph sampling. *COLT* **75** 1966–2011.
- [33] KLUSOWSKI, J. M. and WU, Y. (2020). Estimating the number of connected components in a graph via subgraph sampling. *Bernoulli* **26** 1635–1664. [MR4091087](#) <https://doi.org/10.3150/19-BEJ1147>
- [34] KRIVITSKY, P. N. and HANDCOCK, M. S. (2014). A separable model for dynamic networks. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **76** 29–46. [MR3153932](#) <https://doi.org/10.1111/rssb.12014>
- [35] LEHMANN, E. L. (1983). *Theory of Point Estimation. Wiley Series in Probability and Mathematical Statistics: Probability and Mathematical Statistics*. Wiley, New York. [MR0702834](#) <https://doi.org/10.1007/978-1-4757-2769-2>
- [36] LI, T., LEVINA, E. and ZHU, J. (2020). Network cross-validation by edge sampling. *Biometrika* **107** 257–276. [MR4108931](#) <https://doi.org/10.1093/biomet/asaa006>
- [37] MCCULLAGH, P. (2002). What is a statistical model? *Ann. Statist.* **30** 1225–1310. [MR1936320](#) <https://doi.org/10.1214/aos/1035844977>
- [38] MCCULLAGH, P. (2002). What is a statistical model? *Ann. Statist.* **30** 1225–1310. [MR1936320](#) <https://doi.org/10.1214/aos/1035844977>
- [39] MORENO, J. and JENNINGS, H. (1938). Statistics of social configurations. *Sociometry* **1** 342–374.
- [40] NG, A. Y., JORDAN, M. I. and WEISS, Y. (2001). On spectral clustering: Analysis and an algorithm. In *Proceedings of the 14th International Conference on Neural Information Processing Systems: Natural and Synthetic, NIPS'01* 849–856.
- [41] ORBANZ, P. (2017). Subsampling large graphs and invariance in networks. Available at [arXiv:1710.04217](#).
- [42] ORBANZ, P. and ROY, D. (2014). Bayesian models of graphs, arrays and other exchangeable random structures. *IEEE Trans. Pattern Anal. Mach. Intell.* **37** 437–461.
- [43] PITMAN, J. (2006). *Combinatorial Stochastic Processes. Lecture Notes in Math.* **1875**. Springer, Berlin. [MR2245368](#)
- [44] SHALIZI, C. R. and RINALDO, A. (2013). Consistency under sampling of exponential random graph models. *Ann. Statist.* **41** 508–535. [MR3099112](#) <https://doi.org/10.1214/12-AOS1044>
- [45] VEITCH, V. and ROY, D. (2015). The class of random graphs arising from exchangeable random measures. Available at [arXiv:1512.03099](#).
- [46] WASSERMAN, S. and PATTISON, P. (1996). Logit models and logistic regressions for social networks. I. An introduction to Markov graphs and p . *Psychometrika* **61** 401–425. [MR1424909](#) <https://doi.org/10.1007/BF02294547>
- [47] YU, B. and KUMBIER, K. (2019). Three principles of data science: Predictability, computability, and stability (pcs). Submitted. Available at [arXiv:1901.08152](#).

On Estimation and Inference in Latent Structure Random Graphs

Avanti Athreya, Minh Tang, Youngser Park and Carey E. Priebe

Abstract. We define a *latent structure* random graph as a random dot product graph (RDPG) in which the latent position distribution incorporates both probabilistic and geometric constraints, delineated by a family of *underlying distributions* on some fixed Euclidean space, and a *structural support submanifold* from which are drawn the latent positions for the graph. For a one-dimensional latent structure model with known structural support, we extend existing results on the consistency of spectral estimates in RDPGs to demonstrate that the parameters of the underlying distribution can be efficiently estimated. We describe how to estimate or learn the structural support in cases where it is unknown, with a focus on graphs with latent positions along the Hardy–Weinberg curve. Finally, we use the latent structural model formulation to address a hitherto-open question in neuroscience on the bilateral homology of the *Drosophila* left and right hemisphere connectome.

Key words and phrases: Latent structure random graphs, manifold learning, spectral graph inference, efficiency.

REFERENCES

- AIROLDI, E. M., BLEI, D. M., FIENBERG, S. E. and XING, E. P. (2008). Mixed membership stochastic blockmodels. *J. Mach. Learn. Res.* **9** 1981–2014.
- ATHREYA, A., PRIEBE, C. E., TANG, M., LYZINSKI, V., MARCHETTE, D. J. and SUSSMAN, D. L. (2016). A limit theorem for scaled eigenvectors of random dot product graphs. *Sankhya A* **78** 1–18. [MR3494576 https://doi.org/10.1007/s13171-015-0071-x](https://doi.org/10.1007/s13171-015-0071-x)
- ATHREYA, A., FISHKIND, D. E., TANG, M., PRIEBE, C. E., PARK, Y., VOGELSTEIN, J. T., LEVIN, K., LYZINSKI, V., QIN, Y. et al. (2018). Statistical inference on random dot product graphs: A survey. *J. Mach. Learn. Res.* **18** Paper No. 226, 92. [MR3827114](https://doi.org/10.1214/15-AOS834)
- BICKEL, P. J. and DOKSUM, K. A. (2016). *Mathematical Statistics—Basic Ideas and Selected Topics. Vol. 2*, 2nd ed. *Texts in Statistical Science Series*. CRC Press, Boca Raton, FL. [MR3287337](https://doi.org/10.1002/rsa.20168)
- BOLLOBÁS, B., JANSON, S. and RIORDAN, O. (2007). The phase transition in inhomogeneous random graphs. *Random Structures Algorithms* **31** 3–122. [MR2337396 https://doi.org/10.1002/rsa.20168](https://doi.org/10.1002/rsa.20168)
- BREIMAN, L. (2001). Statistical modeling: The two cultures. *Statist. Sci.* **16** 199–231. [MR1874152 https://doi.org/10.1214/ss/1009213726](https://doi.org/10.1214/ss/1009213726)
- CAPE, J., TANG, M. and PRIEBE, C. E. (2019). The two-to-infinity norm and singular subspace geometry with applications to high-dimensional statistics. *Ann. Statist.* **47** 2405–2439. [MR3988761 https://doi.org/10.1214/18-AOS1752](https://doi.org/10.1214/18-AOS1752)
- COPPERSMITH, G. A. (2014). Vertex nomination. *Wiley Interdiscip. Rev.: Comput. Stat.* **6** 144–153.
- DIACONIS, P. and JANSON, S. (2008). Graph limits and exchangeable random graphs. *Rend. Mat. Appl.* (7) **28** 33–61. [MR2463439](https://doi.org/10.1007/s13171-015-0071-x)
- EICHLER, K., LI, F., KUMAR, A. L., PARK, Y., ANDRADE, I., SCHNEIDER-MIZELL, C., SAUMWEBER, T., HUSER, A., BONNERY, D. et al. (2017). The complete wiring diagram of a high-order learning and memory center, the insect mushroom body. *Nature* **548** 175–182.
- FISHKIND, D. E., LYZINSKI, V., PAO, H., CHEN, L. and PRIEBE, C. E. (2015). Vertex nomination schemes for membership prediction. *Ann. Appl. Stat.* **9** 1510–1532. [MR3418733 https://doi.org/10.1214/15-AOS834](https://doi.org/10.1214/15-AOS834)
- GALLIER, J. (2000). *Curves and Surfaces in Geometric Modeling: Theory and Algorithms. The Morgan Kaufmann Series in Computer Graphics and Geometric Modeling*. Morgan Kaufmann, San Francisco, CA. [MR1823812](https://doi.org/10.1002/rsa.20168)
- HOFF, P. D., RAFTERY, A. E. and HANDCOCK, M. S. (2002). Latent space approaches to social network analysis. *J. Amer. Statist. Assoc.* **97** 1090–1098. [MR1951262 https://doi.org/10.1198/016214502388618906](https://doi.org/10.1198/016214502388618906)
- HOLLAND, P. W., LASKEY, K. B. and LEINHARDT, S. (1983). Stochastic blockmodels: First steps. *Soc. Netw.* **5** 109–137. [MR0718088 https://doi.org/10.1016/0378-8733\(83\)90021-7](https://doi.org/10.1016/0378-8733(83)90021-7)
- KARRER, B. and NEWMAN, M. E. J. (2011). Stochastic blockmodels and community structure in networks. *Phys. Rev. E* (3) **83** 016107,

Avanti Athreya is Assistant Research Professor, Department of Applied Mathematics and Statistics, Johns Hopkins University, 3400 N Charles Street, Baltimore, Maryland 21218, USA (e-mail: dathrey1@jhu.edu). Minh Tang is Assistant Professor, Department of Statistics, North Carolina State University, 2311 Stinson Drive, Raleigh, North Carolina 27607, USA (e-mail: mtang8@ncsu.edu). Youngser Park is Research Scientist, Center for Imaging Science, Johns Hopkins University, 3400 N Charles Street, Baltimore, Maryland 21218, USA (e-mail: youngser@jhu.edu). Carey E. Priebe is Professor, Department of Applied Mathematics, Johns Hopkins University, 3400 N Charles Street, Baltimore, Maryland 21218, USA (e-mail: cep@jhu.edu).

10. MR2788206 <https://doi.org/10.1103/PhysRevE.83.016107>
- LEE, J. M. (2013). *Introduction to Smooth Manifolds*, 2nd ed. *Graduate Texts in Mathematics* **218**. Springer, New York. MR2954043
- LYZINSKI, V., SUSSMAN, D. L., TANG, M., ATHREYA, A. and PRIEBE, C. E. (2014). Perfect clustering for stochastic blockmodel graphs via adjacency spectral embedding. *Electron. J. Stat.* **8** 2905–2922. MR3299126 <https://doi.org/10.1214/14-EJS978>
- LYZINSKI, V., TANG, M., ATHREYA, A., PARK, Y. and PRIEBE, C. E. (2017). Community detection and classification in hierarchical stochastic blockmodels. *IEEE Trans. Netw. Sci. Eng.* **4** 13–26. MR3625952 <https://doi.org/10.1109/TNSE.2016.2634322>
- PRAUTZSCH, H., BOEHM, W. and PALUSZNY, M. (2002). *Bézier and B-Spline Techniques. Mathematics and Visualization*. Springer, Berlin. MR1993341 <https://doi.org/10.1007/978-3-662-04919-8>
- PRIEBE, C. E., PARK, Y., TANG, M., ATHREYA, A., LYZINSKI, V., VOGELSTEIN, J. T., QIN, Y., COCANOUGH, B., EICHLER, K. et al. (2017). Semiparametric spectral modeling of the drosophila connectome. Preprint. Available at [arXiv:1705.03297](https://arxiv.org/abs/1705.03297).
- SMITH, A. L., ASTA, D. M. and CALDER, C. A. (2019). The geometry of continuous latent space models for network data. *Statist. Sci.* **34** 428–453. MR4017522 <https://doi.org/10.1214/19-STS702>
- SUSSMAN, D. L., TANG, M., FISHKIND, D. E. and PRIEBE, C. E. (2012). A consistent adjacency spectral embedding for stochastic blockmodel graphs. *J. Amer. Statist. Assoc.* **107** 1119–1128. MR3010899 <https://doi.org/10.1080/01621459.2012.699795>
- TANG, M., CAPE, J. and PRIEBE, C. E. (2017). Asymptotically efficient estimators for stochastic blockmodels: The naive MLE, the rank-constrained MLE, and the spectral. Preprint. Available at [arXiv:1710.10936](https://arxiv.org/abs/1710.10936).
- TANG, M., ATHREYA, A., SUSSMAN, D. L., LYZINSKI, V. and PRIEBE, C. E. (2017). A nonparametric two-sample hypothesis testing problem for random graphs. *Bernoulli* **23** 1599–1630. MR3624872 <https://doi.org/10.3150/15-BEJ789>
- TENENBAUM, J. B., DE SILVA, V. and LANGFORD, J. C. (2000). A global geometric framework for nonlinear dimensionality reduction. *Science* **290** 2319–2323.
- VAN DER VAART, A. W. and WELLNER, J. A. (1996). *Weak Convergence and Empirical Processes: With Applications to Statistics*. *Springer Series in Statistics*. Springer, New York. MR1385671 <https://doi.org/10.1007/978-1-4757-2545-2>
- YOUNG, S. J. and SCHEINERMAN, E. R. (2007). Random dot product graph models for social networks. In *Algorithms and Models for the Web-Graph. Lecture Notes in Computer Science* **4863** 138–149. Springer, Berlin. MR2504912 https://doi.org/10.1007/978-3-540-77004-6_11

Network Modeling in Biology: Statistical Methods for Gene and Brain Networks

Y. X. Rachel Wang, Lexin Li, Jingyi Jessica Li and Haiyan Huang

Abstract. The rise of network data in many different domains has offered researchers new insights into the problem of modeling complex systems and propelled the development of numerous innovative statistical methodologies and computational tools. In this paper, we primarily focus on two types of biological networks, gene networks and brain networks, where statistical network modeling has found both fruitful and challenging applications. Unlike other network examples such as social networks where network edges can be directly observed, both gene and brain networks require careful estimation of edges using measured data as a first step. We provide a discussion on existing statistical and computational methods for edge estimation and subsequent statistical inference problems in these two types of biological networks.

Key words and phrases: Gene regulatory networks, brain connectivity networks, network reconstruction, network inference.

REFERENCES

- [1] AACH, J. and CHURCH, G. M. (2001). Aligning gene expression time series with time warping algorithms. *Bioinformatics* **17** 495–508. <https://doi.org/10.1093/bioinformatics/17.6.495>
- [2] AGHDAM, R., GANJALI, M., ZHANG, X. and ESLAHCHI, C. (2015). CN: A consensus algorithm for inferring gene regulatory networks using the SORDER algorithm and conditional mutual information test. *Mol. BioSyst.* **11** 942–949.
- [3] AHN, M., SHEN, H., LIN, W. and ZHU, H. (2015). A sparse reduced rank framework for group analysis of functional neuroimaging data. *Statist. Sinica* **25** 295–312. [MR3328816](https://doi.org/10.1007/s11464-015-0416-6)
- [4] AIBAR, S., GONZÁLEZ-BLAS, C. B., MOERMAN, T., IMRICHCHOVA, H., HULSELMANS, G., RAMBOW, F., MARINE, J.-C., GEURTS, P., AERTS, J. et al. (2017). SCENIC: Single-cell regulatory network inference and clustering. *Nat. Methods* **14** 1083–1086.
- [5] AIROLDI, E. M., BLEI, D. M., FIENBERG, S. E. and XING, E. P. (2008). Mixed membership stochastic blockmodels. *J. Mach. Learn. Res.* **9** 1981–2014.
- [6] AITTOKALLIO, T. and SCHWIKOWSKI, B. (2006). Graph-based methods for analysing networks in cell biology. *Brief. Bioinform.* **7** 243–255.
- [7] ALLEN, E. A., DAMARAJU, E., PLIS, S. M., ERHARDT, E. B., EICHELE, T. and CALHOUN, V. D. (2014). Tracking whole-brain connectivity dynamics in the resting state. *Cereb. Cortex* **24** 663–676.
- [8] AMAR, D., SAFER, H. and SHAMIR, R. (2013). Dissection of regulatory networks that are altered in disease via differential co-expression. *PLoS Comput. Biol.* **9** Art. ID e1002955. <https://doi.org/10.1371/journal.pcbi.1002955>
- [9] ANDERS, S. and HUBER, W. (2010). Differential expression analysis for sequence count data. *Genome Biol.* **11** Art. ID R106. <https://doi.org/10.1186/gb-2010-11-10-r106>
- [10] BACHER, R., CHU, L.-F., LENG, N., GASCH, A. P., THOMSON, J. A., STEWART, R. M., NEWTON, M. and KENDZIORSKI, C. (2016). SCnorm: A quantile-regression based approach for robust normalization of single-cell RNA-seq data. *bioRxiv* 090167. <https://doi.org/10.1101/090167>
- [11] BAR-JOSEPH, Z., GERBER, G. K., LEE, T. I., RINALDI, N. J., YOO, J. Y., ROBERT, F., GORDON, D. B., FRAENKEL, E., JAAKKOLA, T. S. et al. (2003). Computational discovery of gene modules and regulatory networks. *Nat. Biotechnol.* **21** 1337–1342.
- [12] BAR-JOSEPH, Z., GITTER, A. and SIMON, I. (2012). Studying and modelling dynamic biological processes using time-series gene expression data. *Nat. Rev. Genet.* **13** 552–564. <https://doi.org/10.1038/nrg3244>
- [13] BARABÁSI, A.-L., GULBAHCE, N. and LOSCALZO, J. (2011). Network medicine: A network-based approach to human disease. *Nat. Rev. Genet.* **12** 56–68. <https://doi.org/10.1038/nrg2918>
- [14] BARABÁSI, A.-L. and OLTVAI, Z. N. (2004). Network biology: Understanding the cell's functional organization. *Nat. Rev. Genet.* **5** 101–113. <https://doi.org/10.1038/nrg1272>
- [15] BASSO, K., MARGOLIN, A. A., STOLOVITZKY, G., KLEIN, U., DALLA-FAVERA, R. and CALIFANO, A. (2005).

Y. X. Rachel Wang is a Senior Lecturer in the School of Mathematics and Statistics, University of Sydney, Sydney, New South Wales, Australia (e-mail: rachel.wang@sydney.edu.au). Lexin Li is a Professor in the Department of Biostatistics and Epidemiology, and School of Public Health, University of California, Berkeley, California, USA (e-mail: lexinli@berkeley.edu). Jingyi Jessica Li is an Associate Professor, Department of Statistics, University of California, Los Angeles, California, USA (e-mail: jli@stat.ucla.edu). Haiyan Huang is a Professor in the Department of Statistics, University of California, Berkeley, California, USA (e-mail: hhuang@stat.berkeley.edu).

- Reverse engineering of regulatory networks in human B cells. *Nat. Genet.* **37** 382–390.
- [16] BEN-DOR, A., SHAMIR, R. and YAKHINI, Z. (1999). Clustering gene expression patterns. *J. Comput. Biol.* **6** 281–297.
- [17] BERG, J. and LÄSSIG, M. (2004). Local graph alignment and motif search in biological networks. *Proc. Natl. Acad. Sci. USA* **101** 14689–14694.
- [18] BUETTNER, F., NATARAJAN, K. N., CASALE, F. P., PROSERPIO, V., SCIALDONE, A., THEIS, F. J., TEICHMANN, S. A., MARIONI, J. C. and STEGLE, O. (2015). Computational analysis of cell-to-cell heterogeneity in single-cell RNA-sequencing data reveals hidden subpopulations of cells. *Nat. Biotechnol.* **33** 155–160. <https://doi.org/10.1038/nbt.3102>
- [19] BULLMORE, E. and SPORNS, O. (2009). Complex brain networks: Graph theoretical analysis of structural and functional systems. *Nat. Rev., Neurosci.* **10** 186–198. <https://doi.org/10.1038/nrn2575>
- [20] BUTTE, A. J. and KOHANE, I. S. (1999). Mutual information relevance networks: Functional genomic clustering using pairwise entropy measurements. In *Biocomputing 2000* 418–429. World Scientific, Singapore.
- [21] CABREROS, I., ABBE, E. and TSIRIGOS, A. (2016). Detecting community structures in hi-c genomic data. In *2016 Annual Conference on Information Science and Systems (CISS)* 584–589. IEEE, Los Alamitos, CA.
- [22] CAI, T., LIU, W. and LUO, X. (2011). A constrained ℓ_1 minimization approach to sparse precision matrix estimation. *J. Amer. Statist. Assoc.* **106** 594–607. [MR2847973 https://doi.org/10.1198/jasa.2011.tm10155](https://doi.org/10.1198/jasa.2011.tm10155)
- [23] CALHOUN, V. D., MILLER, R., PEARLSON, G. and ADALI, T. (2014). The chronnectome: Time-varying connectivity networks as the next frontier in fMRI data discovery. *Neuron* **84** 262–274.
- [24] CAO, X., SANDSTEDE, B. and LUO, X. (2019). A functional data method for causal dynamic network modeling of task-related fMRI. *Front. Neurosci.* **13** Art. ID 127. <https://doi.org/10.3389/fnins.2019.00127>
- [25] CARTER, S. L., BRECHBÜHLER, C. M., GRIFFIN, M. and BOND, A. T. (2004). Gene co-expression network topology provides a framework for molecular characterization of cellular state. *Bioinformatics* **20** 2242–2250.
- [26] CASSIDY, B., RAE, C. and SOLO, V. (2014). Brain activity: Connectivity, sparsity, and mutual information. *IEEE Trans. Med. Imag.* **34** 846–860.
- [27] CHAHROUR, M., JUNG, S. Y., SHAW, C., ZHOU, X., WONG, S. T., QIN, J. and ZOGHBİ, H. Y. (2008). MeCP2, a key contributor to neurological disease, activates and represses transcription. *Science* **320** 1224–1229.
- [28] CHAN, T. E., STUMPF, M. P. and BAPTIE, A. C. (2017). Gene regulatory network inference from single-cell data using multivariate information measures. *Cell Syst.* **5** 251–267.
- [29] CHAVALI, S., BARRENAS, F., KANDURI, K. and BENSON, M. (2010). Network properties of human disease genes with pleiotropic effects. *BMC Syst. Biol.* **4** Art. ID 78. <https://doi.org/10.1186/1752-0509-4-78>
- [30] CHEN, G., NING, B. and SHI, T. (2019). Single-cell RNA-seq technologies and related computational data analysis. *Front. Genet.* **10** Art. ID 317. <https://doi.org/10.3389/fgene.2019.00317>
- [31] CHEN, J., BARDES, E. E., ARONOW, B. J. and JEGGA, A. G. (2009). ToppGene Suite for gene list enrichment analysis and candidate gene prioritization. *Nucleic Acids Res.* **37** W305–W311. <https://doi.org/10.1093/nar/gkp427>
- [32] CHEN, K.-C., WANG, T.-Y., TSENG, H.-H., HUANG, C.-Y. F. and KAO, C.-Y. (2005). A stochastic differential equation model for quantifying transcriptional regulatory network in *Saccharomyces cerevisiae*. *Bioinformatics* **21** 2883–2890.
- [33] CHEN, S., KANG, J., XING, Y. and WANG, G. (2015). A parsimonious statistical method to detect groupwise differentially expressed functional connectivity networks. *Hum. Brain Mapp.* **36** 5196–5206.
- [34] CHICKERING, D. M., HECKERMAN, D. and MEEK, C. (2004). Large-sample learning of Bayesian networks is NP-hard. *J. Mach. Learn. Res.* **5** 1287–1330. [MR2248018](https://doi.org/10.1162/jmlr.2004.5.1287)
- [35] CHUN, H., ZHANG, X. and ZHAO, H. (2015). Gene regulation network inference with joint sparse Gaussian graphical models. *J. Comput. Graph. Statist.* **24** 954–974. [MR3432924 https://doi.org/10.1080/10618600.2014.956876](https://doi.org/10.1080/10618600.2014.956876)
- [36] CIRIELLO, G. and GUERRA, C. (2008). A review on models and algorithms for motif discovery in protein–protein interaction networks. *Brief. Funct. Genomics Proteomics* **7** 147–156.
- [37] CRIBBEN, I., HARALDSDOTTIR, R., ATLAS, L. Y., WAGER, T. D. and LINDQUIST, M. A. (2012). Dynamic connectivity regression: Determining state-related changes in brain connectivity. *NeuroImage* **61** 907–920.
- [38] DA ROCHA, E. L., ROWE, R. G., LUNDIN, V., MALLESHAIAH, M., JHA, D. K., RAMBO, C. R., LI, H., NORTH, T. E., COLLINS, J. J. et al. (2018). Reconstruction of complex single-cell trajectories using CellRouter. *Nat. Commun.* **9** Art. ID 892.
- [39] DAUB, C. O., STEUER, R., SELBIG, J. and KLOSKA, S. (2004). Estimating mutual information using B-spline functions—An improved similarity measure for analysing gene expression data. *BMC Bioinform.* **5** Art. ID 118. <https://doi.org/10.1186/1471-2105-5-118>
- [40] DAUDIN, J.-J., PICARD, F. and ROBIN, S. (2008). A mixture model for random graphs. *Stat. Comput.* **18** 173–183. [MR2390817 https://doi.org/10.1007/s11222-007-9046-7](https://doi.org/10.1007/s11222-007-9046-7)
- [41] DE LA FUENTE, A., BING, N., HOESCHELE, I. and MENDES, P. (2004). Discovery of meaningful associations in genomic data using partial correlation coefficients. *Bioinformatics* **20** 3565–3574.
- [42] DEMBÉLÉ, D. and KASTNER, P. (2003). Fuzzy C-means method for clustering microarray data. *Bioinformatics* **19** 973–980.
- [43] DESIKAN, R. S., SÉGONNE, F., FISCHL, B., QUINN, B. T., DICKERSON, B. C., BLACKER, D., BUCKNER, R. L., DALE, A. M., MAGUIRE, R. P. et al. (2006). An automated labeling system for subdividing the human cerebral cortex on MRI scans into gyral based regions of interest. *NeuroImage* **31** 968–980.
- [44] DI GESÚ, V., GIANCARLO, R., BOSCO, G. L., RAIMONDI, A. and SCATURRO, D. (2005). GenClust: A genetic algorithm for clustering gene expression data. *BMC Bioinform.* **6** Art. ID 289.
- [45] DIXON, J. R., SELVARAJ, S., YUE, F., KIM, A., LI, Y., SHEN, Y., HU, M., LIU, J. S. and REN, B. (2012). Topological domains in mammalian genomes identified by analysis of chromatin interactions. *Nature* **485** 376–380.
- [46] DIXON, S. J., COSTANZO, M., BARYSHNIKOVA, A., ANDREWS, B. and BOONE, C. (2009). Systematic mapping of genetic interaction networks. *Annu. Rev. Genet.* **43** 601–625.
- [47] DOSENBAACH, N. U., NARDOS, B., COHEN, A. L., FAIR, D. A., POWER, J. D., CHURCH, J. A., NELSON, S. M., WIG, G. S., VOGEL, A. C. et al. (2010). Prediction of individual brain maturity using fMRI. *Science* **329** 1358–1361.
- [48] DUÒ, A., ROBINSON, M. D. and SONESON, C. (2018). A systematic performance evaluation of clustering methods for single-cell RNA-seq data. *F1000Res.* **7** Art. ID 1141. <https://doi.org/10.12688/f1000research.15666.2>

- [49] DURANTE, D. and DUNSON, D. B. (2018). Bayesian inference and testing of group differences in brain networks. *Bayesian Anal.* **13** 29–58. [MR3737942 https://doi.org/10.1214/16-BA1030](https://doi.org/10.1214/16-BA1030)
- [50] EBERWINE, J., SUL, J.-Y., BARTFAI, T. and KIM, J. (2014). The promise of single-cell sequencing. *Nat. Methods* **11** 25–27.
- [51] EISEN, M. B., SPELLMAN, P. T., BROWN, P. O. and BOTSTEIN, D. (1998). Cluster analysis and display of genome-wide expression patterns. *Proc. Natl. Acad. Sci. USA* **95** 14863–14868.
- [52] ENRIGHT, A. J., VAN DONGEN, S. and OUZOUNIS, C. A. (2002). An efficient algorithm for large-scale detection of protein families. *Nucleic Acids Res.* **30** 1575–1584.
- [53] ESTRADA, E. (2006). Virtual identification of essential proteins within the protein interaction network of yeast. *Proteomics* **6** 35–40. <https://doi.org/10.1002/pmic.200500209>
- [54] FENG, J., BARBANO, P. E. and MISHRA, B. (2004). Time-frequency feature detection for time-course microarray data. In *Proceedings of the 2004 ACM Symposium on Applied Computing* 128–132. ACM, New York.
- [55] FIECAS, M. and OMBAO, H. (2011). The generalized shrinkage estimator for the analysis of functional connectivity of brain signals. *Ann. Appl. Stat.* **5** 1102–1125. [MR2840188 https://doi.org/10.1214/10-AOAS396](https://doi.org/10.1214/10-AOAS396)
- [56] FIERIS, M. W. E. J., MINNOYE, L., AIBAR, S., GONZÁLEZ-BLAS, C. B., ATAK, Z. K. and AERTS, S. (2018). Mapping gene regulatory networks from single-cell omics data. *Brief. Funct. Genomics* **17** 246–254. <https://doi.org/10.1093/bfpg/ely046>
- [57] FINAK, G., MCDAVID, A., YAJIMA, M., DENG, J., GER-SUK, V., SHALEK, A. K., SLICHTER, C. K., MILLER, H. W., MCEL RATH, M. J. et al. (2015). MAST: A flexible statistical framework for assessing transcriptional changes and characterizing heterogeneity in single-cell RNA sequencing data. *Genome Biol.* **16** Art. ID 278.
- [58] FORNITO, A., ZALESKY, A. and BREAKSPEAR, M. (2013). Graph analysis of the human connectome: Promise, progress, and pitfalls. *NeuroImage* **80** 426–444.
- [59] FOX, M. D. and GREICIUS, M. (2010). Clinical applications of resting state functional connectivity. *Front. Syst. Neurosci.* **4** Art. ID 19.
- [60] FRASER, H. B., HIRSH, A. E., STEINMETZ, L. M., SCHARFE, C. and FELDMAN, M. W. (2002). Evolutionary rate in the protein interaction network. *Science* **296** 750–752.
- [61] FRIEDMAN, J., HASTIE, T. and TIBSHIRANI, R. (2008). Sparse inverse covariance estimation with the graphical lasso. *Biostatistics* **9** 432–441.
- [62] FRISTON, K. J. (1994). Functional and effective connectivity in neuroimaging: A synthesis. *Hum. Brain Mapp.* **2** 56–78.
- [63] FRISTON, K. J. (2011). Functional and effective connectivity: A review. *Brain Connect.* **1** 13–36. <https://doi.org/10.1089/brain.2011.0008>
- [64] FRISTON, K. J., HARRISON, L. and PENNY, W. (2003). Dynamic causal modelling. *NeuroImage* **19** 1273–1302.
- [65] FU, L. and MEDICO, E. (2007). FLAME, a novel fuzzy clustering method for the analysis of DNA microarray data. *BMC Bioinform.* **8** Art. ID 3. <https://doi.org/10.1186/1471-2105-8-3>
- [66] GAMBARDILLA, G., MORETTI, M. N., DE Cegli, R., CARDONE, L., PERON, A. and DI BERNARDO, D. (2013). Differential network analysis for the identification of condition-specific pathway activity and regulation. *Bioinformatics* **29** 1776–1785.
- [67] GAO, L., SUN, P.-G. and SONG, J. (2009). Clustering algorithms for detecting functional modules in protein interaction networks. *J. Bioinform. Comput. Biol.* **7** 217–242.
- [68] GILL, R., DATTA, S. and DATTA, S. (2010). A statistical framework for differential network analysis from microarray data. *BMC Bioinform.* **11** Art. ID 95. <https://doi.org/10.1186/1471-2105-11-95>
- [69] GOLTSEV, Y. and PAPATSENKO, D. (2009). Time warping of evolutionary distant temporal gene expression data based on noise suppression. *BMC Bioinform.* **10** Art. ID 353. <https://doi.org/10.1186/1471-2105-10-353>
- [70] GREENE, C. S., KRISHNAN, A., WONG, A. K., RICCIOTTI, E., ZELAYA, R. A., HIMMELSTEIN, D. S., ZHANG, R., HARTMANN, B. M., ZASLAVSKY, E. et al. (2015). Understanding multicellular function and disease with human tissue-specific networks. *Nat. Genet.* **47** 569–576.
- [71] GREICIUS, M. D., KRASNOW, B., REISS, A. L. and MENON, V. (2003). Functional connectivity in the resting brain: A network analysis of the default mode hypothesis. *Proc. Natl. Acad. Sci. USA* **100** 253–258.
- [72] GUALA, D. and SONNHAMMER, E. L. L. (2017). A large-scale benchmark of gene prioritization methods. *Sci. Rep.* **7** Art. ID 46598. <https://doi.org/10.1038/srep46598>
- [73] GUIMERA, R. and AMARAL, L. A. N. (2005). Functional cartography of complex metabolic networks. *Nature* **433** 895–900.
- [74] GUO, M., WANG, H., POTTER, S. S., WHITSETT, J. A. and XU, Y. (2015). SINCERA: A pipeline for single-cell RNA-seq profiling analysis. *PLoS Comput. Biol.* **11** Art. ID e1004575.
- [75] HA, M. J., BALADANDAYUTHAPANI, V. and DO, K.-A. (2015). DINGO: Differential network analysis in genomics. *Bioinformatics* **31** 3413–3420.
- [76] HAQUE, A., ENGEL, J., TEICHMANN, S. A. and LÖNNBERG, T. (2017). A practical guide to single-cell RNA-sequencing for biomedical research and clinical applications. *Gen. Med.* **9** Art. ID 75. <https://doi.org/10.1186/s13073-017-0467-4>
- [77] HAUSER, A. and BÜHLMANN, P. (2015). Jointly interventional and observational data: Estimation of interventional Markov equivalence classes of directed acyclic graphs. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **77** 291–318. [MR3299409 https://doi.org/10.1111/rssb.12071](https://doi.org/10.1111/rssb.12071)
- [78] HEDDEN, T., VAN DIJK, K. R., BECKER, J. A., MEHTA, A., SPERLING, R. A., JOHNSON, K. A. and BUCKNER, R. L. (2009). Disruption of functional connectivity in clinically normal older adults harboring amyloid burden. *J. Neurosci.* **29** 12686–12694.
- [79] HOLLAND, P. W., LASKEY, K. B. and LEINHARDT, S. (1983). Stochastic blockmodels: First steps. *Soc. Netw.* **5** 109–137. [MR0718088 https://doi.org/10.1016/0378-8733\(83\)90021-7](https://doi.org/10.1016/0378-8733(83)90021-7)
- [80] HUANG, M., WANG, J., TORRE, E., DUECK, H., SHAFFER, S., BONASIO, R., MURRAY, J. I., RAJ, A., LI, M. et al. (2018). SAVER: Gene expression recovery for single-cell RNA sequencing. *Nat. Methods* **15** 539–542.
- [81] HUANG, S., JIN, J. and YAO, Z. (2016). Partial correlation screening for estimating large precision matrices, with applications to classification. *Ann. Statist.* **44** 2018–2057. [MR3546442 https://doi.org/10.1214/15-AOS1392](https://doi.org/10.1214/15-AOS1392)
- [82] HUSMEIER, D. (2003). Sensitivity and specificity of inferring genetic regulatory interactions from microarray experiments with dynamic Bayesian networks. *Bioinformatics* **19** 2271–2282.
- [83] IDEKER, T. and KROGAN, N. J. (2012). Differential network biology. *Mol. Syst. Biol.* **8** Art. ID 565.
- [84] IDEKER, T. and SHARAN, R. (2008). Protein networks in disease. *Genome Res.* **18** 644–652.

- [85] ILICIC, T., KIM, J. K., KOLODZIEJCZYK, A. A., BAGGER, F. O., MCCARTHY, D. J., MARIONI, J. C. and TEICHMANN, S. A. (2016). Classification of low quality cells from single-cell RNA-seq data. *Genome Biol.* **17** Art. ID 29.
- [86] IRRTHUM, A., WEHENKEL, L., GEURTS, P. et al. (2010). Inferring regulatory networks from expression data using tree-based methods. *PLoS ONE* **5** Art. ID e12776.
- [87] ISLAM, S., KJÄLLQUIST, U., MOLINER, A., ZAJAC, P., FAN, J.-B., LÖNNERBERG, P. and LINNARSSON, S. (2011). Characterization of the single-cell transcriptional landscape by highly multiplex RNA-seq. *Genome Res.* **21** 1160–1167.
- [88] ISLAM, S., ZEISEL, A., JOOST, S., MANNO, G. L., ZAJAC, P., KASPER, M., LÖNNERBERG, P. and LINNARSSON, S. (2014). Quantitative single-cell RNA-seq with unique molecular identifiers. *Nat. Methods* **11** 163–166. <https://doi.org/10.1038/nmeth.2772>
- [89] JEONG, H., MASON, S. P., BARABÁSI, A.-L. and OLTVAI, Z. N. (2001). Lethality and centrality in protein networks. *Nature* **411** 41–42.
- [90] JI, Z. and JI, H. (2016). TSCAN: Pseudo-time reconstruction and evaluation in single-cell RNA-seq analysis. *Nucleic Acids Res.* **44** Art. ID e117. <https://doi.org/10.1093/nar/gkw430>
- [91] JIANG, H., SOHN, L. L., HUANG, H. and CHEN, L. (2018). Single cell clustering based on cell-pair differentiability correlation and variance analysis. *Bioinformatics* **34** 3684–3694.
- [92] JIANG, R., TU, Z., CHEN, T. and SUN, F. (2006). Network motif identification in stochastic networks. *Proc. Natl. Acad. Sci. USA* **103** 9404–9409.
- [93] JONSSON, P. F. and BATES, P. A. (2006). Global topological features of cancer proteins in the human interactome. *Bioinformatics* **22** 2291–2297.
- [94] JOY, M. P., BROCK, A., INGBER, D. E. and HUANG, S. (2005). High-betweenness proteins in the yeast protein interaction network. *BioMed Res. Int.* **2005** 96–103.
- [95] KANG, H., OMBAO, H., LINKLETTER, C., LONG, N. and BADRE, D. (2012). Spatio-spectral mixed-effects model for functional magnetic resonance imaging data. *J. Amer. Statist. Assoc.* **107** 568–577. <https://doi.org/10.1080/01621459.2012.664503>
- [96] KANG, H. M., SUBRAMANIAM, M., TARG, S., NGUYEN, M., MALISKOVA, L., MCCARTHY, E., WAN, E., WONG, S., BYRNES, L. et al. (2018). Multiplexed droplet single-cell RNA-sequencing using natural genetic variation. *Nat. Biotechnol.* **36** 89–94.
- [97] KANG, J., BOWMAN, F. D., MAYBERG, H. and LIU, H. (2016). A depression network of functionally connected regions discovered via multi-attribute canonical correlation graphs. *NeuroImage* **141** 431–441.
- [98] KARRER, B. and NEWMAN, M. E. J. (2011). Stochastic block-models and community structure in networks. *Phys. Rev. E* (3) **83** Art. ID 016107. <https://doi.org/10.1103/PhysRevE.83.016107>
- [99] KASHTAN, N., ITZKOVITZ, S., MILO, R. and ALON, U. (2004). Efficient sampling algorithm for estimating subgraph concentrations and detecting network motifs. *Bioinformatics* **20** 1746–1758.
- [100] KATO, M., TSUNODA, T. and TAKAGI, T. (2001). Lag analysis of genetic networks in the cell cycle of budding yeast. *Genome Inform.* **12** 266–267.
- [101] KHARCHENKO, P. V., SILBERSTEIN, L. and SCADDEN, D. T. (2014). Bayesian approach to single-cell differential expression analysis. *Nat. Methods* **11** 740–742. <https://doi.org/10.1038/nmeth.2967>
- [102] KIM, J., WOZNIAK, J. R., MUELLER, B. A., SHEN, X. and PAN, W. (2014). Comparison of statistical tests for group differences in brain functional networks. *NeuroImage* **101** 681–694.
- [103] KIM, K., JIANG, K., TENG, S. L., FELDMAN, L. J. and HUANG, H. (2012). Using biologically interrelated experiments to identify pathway genes in *Arabidopsis*. *Bioinformatics* **28** 815–822.
- [104] KIM, S., IMOTO, S. and MIYANO, S. (2004). Dynamic Bayesian network and nonparametric regression for nonlinear modeling of gene networks from time series gene expression data. *Biosystems* **75** 57–65. <https://doi.org/10.1016/j.biosystems.2004.03.004>
- [105] KISELEV, V. Y., ANDREWS, T. S. and HEMBERG, M. (2019). Challenges in unsupervised clustering of single-cell RNA-seq data. *Nat. Rev. Genet.* **20** 273–282. <https://doi.org/10.1038/s41576-018-0088-9>
- [106] KISELEV, V. Y., KIRSCHNER, K., SCHAUB, M. T., ANDREWS, T., YIU, A., CHANDRA, T., NATARAJAN, K. N., REIK, W., BARAHONA, M. et al. (2017). SC3: Consensus clustering of single-cell RNA-seq data. *Nat. Methods* **14** 483–486.
- [107] KLEIN, A. M., MAZUTIS, L., AKARTUNA, I., TALLAPRAGADA, N., VERES, A., LI, V., PESHKIN, L., WEITZ, D. A. and KIRSCHNER, M. W. (2015). Droplet barcoding for single-cell transcriptomics applied to embryonic stem cells. *Cell* **161** 1187–1201.
- [108] KOLODZIEJCZYK, A. A., KIM, J. K., SVENSSON, V., MARIONI, J. C. and TEICHMANN, S. A. (2015). The technology and biology of single-cell RNA sequencing. *Mol. Cell* **58** 610–620.
- [109] KUMARI, S., NIE, J., CHEN, H.-S., MA, H., STEWART, R., LI, X., LU, M.-Z., TAYLOR, W. M. and WEI, H. (2012). Evaluation of gene association methods for coexpression network construction and biological knowledge discovery. *PLoS ONE* **7** Art. ID e50411.
- [110] KUNDU, S. and KANG, J. (2016). Semiparametric Bayes conditional graphical models for imaging genetics applications. *Stat* **5** 322–337. <https://doi.org/10.1002/sta4.119>
- [111] KWON, A. T., HOOS, H. H. and NG, R. (2003). Inference of transcriptional regulation relationships from gene expression data. In *Proceedings of the 2003 ACM Symposium on Applied Computing* 135–140. ACM, New York.
- [112] LANGFELDER, P. and HORVATH, S. (2008). WGCNA: An R package for weighted correlation network analysis. *BMC Bioinform.* **9** Art. ID 559.
- [113] LAZAR, N. (2008). *The Statistical Analysis of Functional MRI Data*. Springer, New York.
- [114] LE DILY, F., BAÙ, D., POHL, A., VICENT, G. P., SERRA, F., SORONELLAS, D., CASTELLANO, G., WRIGHT, R. H., BALLARE, C. et al. (2014). Distinct structural transitions of chromatin topological domains correlate with coordinated hormone-induced gene regulation. *Genes Dev.* **28** 2151–2162.
- [115] LEE, H., CHUNG, M. K., KANG, H., KIM, B.-N. and LEE, D. S. (2011). Discriminative persistent homology of brain networks. In *2011 IEEE International Symposium on Biomedical Imaging: From Nano to Macro* 841–844. IEEE, Los Alamitos, CA.
- [116] LEHNER, B., CROMBIE, C., TISCHLER, J., FORTUNATO, A. and FRASER, A. G. (2006). Systematic mapping of genetic interactions in *Caenorhabditis elegans* identifies common modifiers of diverse signaling pathways. *Nat. Genet.* **38** 896–903.
- [117] LEMMENS, K., DE BIE, T., DHOLLANDER, T., DE KEERSMAECKER, S. C., THIJS, I. M., SCHOofs, G., DE WEERDT, A., DE MOOR, B., VANDERLEYDEN, J. et al. (2009). DISTILLER: A data integration framework to reveal condition dependency of complex regulons in *Escherichia coli*. *Genome Biol.* **10** Art. ID R27.

- [118] LI, B. and SOLEA, E. (2018). A nonparametric graphical model for functional data with application to brain networks based on fMRI. *J. Amer. Statist. Assoc.* **113** 1637–1655. [MR3902235](https://doi.org/10.1080/01621459.2017.1356726) <https://doi.org/10.1080/01621459.2017.1356726>
- [119] LI, H. and GUI, J. (2005). Gradient directed regularization for sparse Gaussian concentration graphs, with applications to inference of genetic networks. *Biostatistics* **7** 302–317.
- [120] LI, K.-C. (2002). Genome-wide coexpression dynamics: Theory and application. *Proc. Natl. Acad. Sci. USA* **99** 16875–16880.
- [121] LI, K.-C., PALOTIE, A., YUAN, S., BRONNIKOV, D., CHEN, D., WEI, X., CHOI, O.-W., SAARELA, J. and PELTONEN, L. (2007). Finding disease candidate genes by liquid association. *Genome Biol.* **8** Art. ID R205.
- [122] LI, Q., ŞENTÜRK, D., SUGAR, C. A., JESTE, S., DISTEFANO, C., FROHLICH, J. and TELESKA, D. (2019). Inferring brain signals synchronicity from a sample of EEG readings. *J. Amer. Statist. Assoc.* **114** 991–1001. [MR4011753](https://doi.org/10.1080/01621459.2018.1518233) <https://doi.org/10.1080/01621459.2018.1518233>
- [123] LI, W. V. and LI, J. J. (2018). An accurate and robust imputation method scImpute for single-cell RNA-seq data. *Nat. Commun.* **9** Art. ID 997.
- [124] LIN, A., WANG, R. T., AHN, S., PARK, C. C. and SMITH, D. J. (2010). A genome-wide map of human genetic interactions inferred from radiation hybrid genotypes. *Genome Res.* **20** 1122–1132.
- [125] LIN, P., TROUP, M. and HO, J. W. K. (2017). CIDR: Ultra-fast and accurate clustering through imputation for single-cell RNA-seq data. *Genome Biol.* **18** Art. ID 59. <https://doi.org/10.1186/s13059-017-1188-0>
- [126] LIN, Z., WANG, T., YANG, C. and ZHAO, H. (2017). On joint estimation of Gaussian graphical models for spatial and temporal data. *Biometrics* **73** 769–779. [MR3713111](https://doi.org/10.1111/biom.12650) <https://doi.org/10.1111/biom.12650>
- [127] LINDQUIST, M. A. (2008). The statistical analysis of fMRI data. *Statist. Sci.* **23** 439–464. [MR2530545](https://doi.org/10.1214/09-STS282) <https://doi.org/10.1214/09-STS282>
- [128] LIU, F., ZHANG, S.-W., GUO, W.-F., WEI, Z.-G. and CHEN, L. (2016). Inference of gene regulatory network based on local Bayesian networks. *PLoS Comput. Biol.* **12** Art. ID e1005024.
- [129] LIU, K., THEUSCH, E., ZHOU, Y., ASHUACH, T., DOSE, A., BICKEL, P. J., MEDINA, M. W. and HUANG, H. (2019). GeneFishing: A method to reconstruct context-specific portraits of biological processes and its application to cholesterol metabolism. *Proc. Natl. Acad. Sci. USA* **116** 18943–18950.
- [130] LUN, A. T. L., BACH, K. and MARIONI, J. C. (2016). Pooling across cells to normalize single-cell RNA sequencing data with many zero counts. *Genome Biol.* **17** Art. ID 75. <https://doi.org/10.1186/s13059-016-0947-7>
- [131] LUO, X. and WEI, Y. (2018). Nonparametric Bayesian learning of heterogeneous dynamic transcription factor networks. *Ann. Appl. Stat.* **12** 1749–1772. [MR3852696](https://doi.org/10.1214/17-AOAS1129) <https://doi.org/10.1214/17-AOAS1129>
- [132] MA, C., XIN, M., FELDMANN, K. A. and WANG, X. (2014). Machine learning-based differential network analysis: A study of stress-responsive transcriptomes in *Arabidopsis*. *Plant Cell* **26** 520–537.
- [133] MAATHUIS, M. H., KALISCH, M. and BÜHLMANN, P. (2009). Estimating high-dimensional intervention effects from observational data. *Ann. Statist.* **37** 3133–3164. [MR2549555](https://doi.org/10.1214/09-AOS685) <https://doi.org/10.1214/09-AOS685>
- [134] MACOSKO, E. Z., BASU, A., SATIJA, R., NEMESH, J., SHEKHAR, K., GOLDMAN, M., TIROSH, I., BIALAS, A. R., KAMITAKI, N. et al. (2015). Highly parallel genome-wide expression profiling of individual cells using nanoliter droplets. *Cell* **161** 1202–1214.
- [135] MAGWENE, P. M. and KIM, J. (2004). Estimating genomic co-expression networks using first-order conditional independence. *Genome Biol.* **5** Art. ID R100.
- [136] MARBACH, D., COSTELLO, J. C., KÜFFNER, R., VEGA, N. M., PRILL, R. J., CAMACHO, D. M., ALLISON, K. R., ADERHOLD, A., BONNEAU, R. et al. (2012). Wisdom of crowds for robust gene network inference. *Nat. Methods* **9** 796–804.
- [137] MARBACH, D., PRILL, R. J., SCHAFFTER, T., MATTIUSI, C., FLOREANO, D. and STOLOVITZKY, G. (2010). Revealing strengths and weaknesses of methods for gene network inference. *Proc. Natl. Acad. Sci. USA* **107** 6286–6291.
- [138] MARBACH, D., ROY, S., AY, F., MEYER, P. E., CANDEIAS, R., KAHVECI, T., BRISTOW, C. A. and KELIS, M. (2012). Predictive regulatory models in *Drosophila melanogaster* by integrative inference of transcriptional networks. *Genome Res.* **22** 1334–1349.
- [139] MARGOLIN, A. A., NEMENMAN, I., BASSO, K., WIGGINS, C., STOLOVITZKY, G., DALLA FAVERA, R. and CALIFANO, A. (2006). ARACNE: An algorithm for the reconstruction of gene regulatory networks in a mammalian cellular context. *BMC Bioinform.* **7** Art. ID S7.
- [140] MATSUMOTO, H., KIRYU, H., FURUSAWA, C., KO, M. S. H., KO, S. B. H., GOUDA, N., HAYASHI, T. and NIKAIDO, I. (2017). SCODE: An efficient regulatory network inference algorithm from single-cell RNA-Seq during differentiation. *Bioinformatics* **33** 2314–2321. <https://doi.org/10.1093/bioinformatics/btx194>
- [141] MCLINTOSH, A. R. and GONZALEZ-LIMA, F. (1994). Structural equation modeling and its application to network analysis in functional brain imaging. *Hum. Brain Mapp.* **2** 2–22.
- [142] MEINSHAUSEN, N. and BÜHLMANN, P. (2006). High-dimensional graphs and variable selection with the lasso. *Ann. Statist.* **34** 1436–1462. [MR2278363](https://doi.org/10.1214/009053606000000281) <https://doi.org/10.1214/009053606000000281>
- [143] MEINSHAUSEN, N., HAUSER, A., MOOIJ, J. M., PETERS, J., VERSTEEG, P. and BÜHLMANN, P. (2016). Methods for causal inference from gene perturbation experiments and validation. *Proc. Natl. Acad. Sci. USA* **113** 7361–7368.
- [144] MILO, R., SHEN-ORR, S., ITZKOVITZ, S., KASHTAN, N., CHKLOVSKII, D. and ALON, U. (2002). Network motifs: Simple building blocks of complex networks. *Science* **298** 824–827.
- [145] MOIGNARD, V., WOODHOUSE, S., HAGHVERDI, L., LILLY, A. J., TANAKA, Y., WILKINSON, A. C., BUETTNER, F., MACAULAY, I. C., JAWAID, W. et al. (2015). Decoding the regulatory network of early blood development from single-cell gene expression measurements. *Nat. Biotechnol.* **33** 269–276.
- [146] MOREAU, Y. and TRANCHEVENT, L.-C. (2012). Computational tools for prioritizing candidate genes: Boosting disease gene discovery. *Nat. Rev. Genet.* **13** 523–536. <https://doi.org/10.1038/nrg3253>
- [147] MOSCHOPOULOS, C. N., PAVLOPOULOS, G. A., SCHNEIDER, R., LIKOTHANASSIS, S. D. and KOSSIDA, S. (2009). GIBA: A clustering tool for detecting protein complexes. *BMC Bioinform.* **10** Art. ID S11.
- [148] MURO, S., TAKEMASA, I., OBA, S., MATOBA, R., UENO, N., MARUYAMA, C., YAMASHITA, R., SEKIMOTO, M., YAMAMOTO, H. et al. (2003). Identification of expressed genes linked to malignancy of human colorectal carcinoma by parametric clustering of quantitative expression data. *Genome Biol.* **4** Art. ID R21.

- [149] NAIR, A., CHETTY, M. and WANGIKAR, P. P. (2015). Improving gene regulatory network inference using network topology information. *Mol. BioSyst.* **11** 2449–2463.
- [150] NARAYAN, M. (2015). Inferential methods to find differences in populations of graphical models with applications to functional connectomics. Ph.D. thesis, Rice Univ. [MR3781788](#)
- [151] NEWMAN, M. E. J. (2010). *Networks: An Introduction*. Oxford Univ. Press, Oxford. [MR2676073](#) <https://doi.org/10.1093/acprof:oso/9780199206650.001.0001>
- [152] NORTON, H. K., EMERSON, D. J., HUANG, H., KIM, J., TITUS, K. R., GU, S., BASSETT, D. S. and PHILLIPS-CREMINS, J. E. (2018). Detecting hierarchical genome folding with network modularity. *Nat. Methods* **15** 119–122.
- [153] NUNEZ, M. D., NUNEZ, P. L., SRINIVASAN, R., OMBAO, H., LINQUIST, M., THOMPSON, W. and ASTON, J. (2016). Electroencephalography (EEG): Neurophysics, experimental methods, and signal processing. In *Handbook of Neuroimaging Data Analysis. Chapman & Hall/CRC Handbooks of Modern Statistical Methods* 175–197. CRC Press, Boca Raton, FL.
- [154] OKUDA, S., YAMADA, T., HAMAJIMA, M., ITOH, M., KATAYAMA, T., BORK, P., GOTO, S. and KANEHISA, M. (2008). KEGG Atlas mapping for global analysis of metabolic pathways. *Nucleic Acids Res.* **36** W423–W426. <https://doi.org/10.1093/nar/gkn282>
- [155] OMBAO, H., SCHRODER, A. L., EUAN, C., TING, C. M. and SAMDIN, B. (2016). Advanced topics for modeling electroencephalograms. In *Handbook of Neuroimaging Data Analysis* (H. Ombao, M. Linquist, W. Thompson and J. Aston, eds.) 567–626. Chapman & Hall/CRC, Boca Raton, FL.
- [156] OMRANIAN, N., ELOUNDOU-MBEI, J. M. O., MUELLER-ROEBER, B. and NIKOLOSKI, Z. (2016). Gene regulatory network inference using fused LASSO on multiple data sets. *Sci. Rep.* **6** Art. ID 20533. <https://doi.org/10.1038/srep20533>
- [157] OZGÜR, A., VU, T., ERKAN, G. and RADEV, D. R. (2008). Identifying gene–disease associations using centrality on a literature mined gene–interaction network. *Bioinformatics* **24** i277–i285. <https://doi.org/10.1093/bioinformatics/btn182>
- [158] PALADUGU, S. R., ZHAO, S., RAY, A. and RAVAL, A. (2008). Mining protein networks for synthetic genetic interactions. *BMC Bioinform.* **9** Art. ID 426. <https://doi.org/10.1186/1471-2105-9-426>
- [159] PAVLOPOULOS, G. A., SECRIER, M., MOSCHOPOULOS, C. N., SOLDATOS, T. G., KOSSIDA, S., AERTS, J., SCHNEIDER, R. and BAGOS, P. G. (2011). Using graph theory to analyze biological networks. *BioData Min.* **4** Art. ID 10. <https://doi.org/10.1186/1756-0381-4-10>
- [160] PEI, Y., GAO, Q., LI, J. and ZHAO, X. (2014). Identifying local co-regulation relationships in gene expression data. *J. Theoret. Biol.* **360** 200–207.
- [161] PENG, J., ZHOU, N. and ZHU, J. (2009). Partial correlation estimation by joint sparse regression models. *J. Amer. Statist. Assoc.* **104** 735–746. [MR2541591](#) <https://doi.org/10.1198/jasa.2009.0126>
- [162] PETERS, J., BÜHLMANN, P. and MEINSHAUSEN, N. (2016). Causal inference by using invariant prediction: Identification and confidence intervals. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **78** 947–1012. [MR3557186](#) <https://doi.org/10.1111/rssb.12167>
- [163] PETUKHOV, V., GUO, J., BARYAWNO, N., SEVERE, N., SCADDEN, D. T., SAMSONOVA, M. G. and KHARCHENKO, P. V. (2018). dropEst: Pipeline for accurate estimation of molecular counts in droplet-based single-cell RNA-seq experiments. *Genome Biol.* **19** Art. ID 78.
- [164] PICELLI, S. (2017). Single-cell RNA-sequencing: The future of genome biology is now. *RNA Biol.* **14** 637–650. <https://doi.org/10.1080/15476286.2016.1201618>
- [165] PICELLI, S., BJÖRKLUND, Å. K., FARIDANI, O. R., SAGASSER, S., WINBERG, G. and SANDBERG, R. (2013). Smart-seq2 for sensitive full-length transcriptome profiling in single cells. *Nat. Methods* **10** 1096–1098.
- [166] PINA, C., TELES, J., FUGAZZA, C., MAY, G., WANG, D., GUO, Y., SONEJI, S., BROWN, J., EDÉN, P. et al. (2015). Single-cell network analysis identifies DDIT3 as a nodal lineage regulator in hematopoiesis. *Cell Rep.* **11** 1503–1510.
- [167] POWER, J. D., COHEN, A. L., NELSON, S. M., WIG, G. S., BARNES, K. A., CHURCH, J. A., VOGEL, A. C., LAUMANN, T. O., MIEZIN, F. M. et al. (2011). Functional network organization of the human brain. *Neuron* **72** 665–678.
- [168] QIAO, X., GUO, S. and JAMES, G. M. (2019). Functional graphical models. *J. Amer. Statist. Assoc.* **114** 211–222. [MR3941249](#) <https://doi.org/10.1080/01621459.2017.1390466>
- [169] QIU, A., LEE, A., TAN, M. and CHUNG, M. K. (2015). Manifold learning on brain functional networks in aging. *Med. Image Anal.* **20** 52–60.
- [170] QIU, H., HAN, F., LIU, H. and CAFFO, B. (2016). Joint estimation of multiple graphical models from high dimensional time series. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **78** 487–504. [MR3454206](#) <https://doi.org/10.1111/rssb.12123>
- [171] QIU, X., HILL, A., PACKER, J., LIN, D., MA, Y.-A. and TRAPNELL, C. (2017). Single-cell mRNA quantification and differential analysis with Census. *Nat. Methods* **14** 309–315.
- [172] RAMONI, M. F., SEBASTIANI, P. and KOHANE, I. S. (2002). Cluster analysis of gene expression dynamics. *Proc. Natl. Acad. Sci. USA* **99** 9121–9126. [MR1909705](#) <https://doi.org/10.1073/pnas.132656399>
- [173] RAMSKÖLD, D., LUO, S., WANG, Y.-C., LI, R., DENG, Q., FARIDANI, O. R., DANIELS, G. A., KHREBTUKOVA, I., LORING, J. F. et al. (2012). Full-length mRNA-Seq from single-cell levels of RNA and individual circulating tumor cells. *Nat. Biotechnol.* **30** 777–782.
- [174] RAU, A., JAFFRÉZIC, F. and NUEL, G. (2013). Joint estimation of causal effects from observational and intervention gene expression data. *BMC Syst. Biol.* **7** Art. ID 111. <https://doi.org/10.1186/1752-0509-7-111>
- [175] RESHEF, D. N., RESHEF, Y. A., FINUCANE, H. K., GROSSMAN, S. R., MCVEAN, G., TURNBAUGH, P. J., LANDER, E. S., MITZENMACHER, M. and SABETI, P. C. (2011). Detecting novel associations in large data sets. *Science* **334** 1518–1524.
- [176] ROBINSON, M. D. and OSHLACK, A. (2010). A scaling normalization method for differential expression analysis of RNA-seq data. *Genome Biol.* **11** Art. ID R25. <https://doi.org/10.1186/gb-2010-11-3-r25>
- [177] ROUSSEEUW, P. J. (1987). Silhouettes: A graphical aid to the interpretation and validation of cluster analysis. *J. Comput. Appl. Math.* **20** 53–65.
- [178] ROY, S., BHATTACHARYYA, D. K. and KALITA, J. K. (2014). Reconstruction of gene co-expression network from microarray data using local expression patterns. *BMC Bioinform.* **15** Art. ID S10. <https://doi.org/10.1186/1471-2105-15-S7-S10>
- [179] RUDIE, J. D., BROWN, J. A., BECK-PANCER, D., HERNANDEZ, L. M., DENNIS, E. L., THOMPSON, P. M., BOOKHEIMER, S. Y. and DAPRETTO, M. J. N. C. (2013). Altered functional and structural brain network organization in autism. *NeuroImage Clin.* **2** 79–94.
- [180] RYALI, S., CHEN, T., SUPEKAR, K. and MENON, V. (2012). Estimation of functional connectivity in fMRI data using stability selection-based sparse partial correlation with elastic net penalty. *NeuroImage* **59** 3852–3861.

- [181] SALIBA, A.-E., WESTERMANN, A. J., GORSKI, S. A. and VOGEL, J. (2014). Single-cell RNA-seq: Advances and future challenges. *Nucleic Acids Res.* **42** 8845–8860. <https://doi.org/10.1093/nar/gku555>
- [182] SALVADOR, R., SUCKLING, J., SCHWARZBAUER, C. and BULLMORE, E. (2005). Undirected graphs of frequency-dependent functional connectivity in whole brain networks. *Philos. Trans. R. Soc. B* **360** 937–946.
- [183] SANTOS-ZAVALA, A., SALGADO, H., GAMA-CASTRO, S., SÁNCHEZ-PÉREZ, M., GÓMEZ-ROMERO, L., LEDEZMA-TEJEIDA, D., GARCÍA-SOTELO, J. S., ALQUICIRA-HERNÁNDEZ, K., MUÑIZ-RASCADO, L. J. et al. (2018). RegulonDB v 10.5: Tackling challenges to unify classic and high throughput knowledge of gene regulation in *E. coli* K-12. *Nucleic Acids Res.* **47** D212–D220.
- [184] SATIJA, R., FARRELL, J. A., GENNERT, D., SCHIER, A. F. and REGEV, A. (2015). Spatial reconstruction of single-cell gene expression data. *Nat. Biotechnol.* **33** 495–502. <https://doi.org/10.1038/nbt.3192>
- [185] SATULURI, V., PARTHASARATHY, S. and UCAR, D. (2010). Markov clustering of protein interaction networks with improved balance and scalability. In *Proceedings of the First ACM International Conference on Bioinformatics and Computational Biology* 247–256. ACM, New York.
- [186] SCHIFFMAN, C., LIN, C., SHI, F., CHEN, L., SOHN, L. and HUANG, H. (2017). SIDEseq: A cell similarity measure defined by shared identified differentially expressed genes for single-cell RNA sequencing data. *Stat. Biosci.* **9** 200–216.
- [187] SCHULDINER, M., COLLINS, S. R., THOMPSON, N. J., DENIC, V., BHAMIDIPATI, A., PUNNA, T., IHMELS, J., ANDREWS, B., BOONE, C. et al. (2005). Exploration of the function and organization of the yeast early secretory pathway through an epistatic miniarray profile. *Cell* **123** 507–519.
- [188] SETTY, M., TADMOR, M. D., REICH-ZELIGER, S., ANGEL, O., SALAME, T. M., KATHAIL, P., CHOI, K., BENDALL, S., FRIEDMAN, N. et al. (2016). Wishbone identifies bifurcating developmental trajectories from single-cell data. *Nat. Biotechnol.* **34** 637–645.
- [189] SEXTON, T., YAFFE, E., KENIGSBERG, E., BANTIGNIES, F., LEBLANC, B., HOICHMAN, M., PARRINELLO, H., TANAY, A. and CAVALLI, G. (2012). Three-dimensional folding and functional organization principles of the *Drosophila* genome. *Cell* **148** 458–472.
- [190] SHAPIRO, E., BIEZUNER, T. and LINNARSSON, S. (2013). Single-cell sequencing-based technologies will revolutionize whole-organism science. *Nat. Rev. Genet.* **14** 618–630. <https://doi.org/10.1038/nrg3542>
- [191] SHARAN, R., MARON-KATZ, A. and SHAMIR, R. (2003). CLICK and EXPANDER: A system for clustering and visualizing gene expression data. *Bioinformatics* **19** 1787–1799.
- [192] SHI, F. and HUANG, H. (2017). Identifying cell subpopulations and their genetic drivers from single-cell RNA-Seq data using a biclustering approach. *J. Comput. Biol.* **24** 663–674. <https://doi.org/10.1089/cmb.2017.0049>
- [193] SHIN, J., BERG, D. A., ZHU, Y., SHIN, J. Y., SONG, J., BONAGUIDI, M. A., ENIKOLOPOV, G., NAUEN, D. W., CHRISTIAN, K. M. et al. (2015). Single-cell RNA-seq with waterfall reveals molecular cascades underlying adult neurogenesis. *Cell Stem Cell* **17** 360–372.
- [194] SMET, R. D. and MARCHAL, K. (2010). Advantages and limitations of current network inference methods. *Nat. Rev. Microbiol.* **8** 717–729. <https://doi.org/10.1038/nrmicro2419>
- [195] SMITH, T., HEGER, A. and SUDBERY, I. (2017). UMI-tools: Modeling sequencing errors in Unique Molecular Identifiers to improve quantification accuracy. *Genome Res.* **27** 491–499. <https://doi.org/10.1101/gr.209601.116>
- [196] SOLO, V., POLINE, J.-B., LINDQUIST, M. A., SIMPSON, S. L., BOWMAN, F. D., CHUNG, M. K. and CASSIDY, B. (2018). Connectivity in fMRI: Blind spots and breakthroughs. *IEEE Trans. Med. Imag.* **37** 1537–1550. <https://doi.org/10.1109/TMI.2018.2831261>
- [197] SPELLMAN, P. T., SHERLOCK, G., ZHANG, M. Q., IYER, V. R., ANDERS, K., EISEN, M. B., BROWN, P. O., BOSTEIN, D. and FUTCHER, B. (1998). Comprehensive identification of cell cycle-regulated genes of the yeast *Saccharomyces cerevisiae* by microarray hybridization. *Mol. Biol. Cell* **9** 3273–3297.
- [198] SRIVASTAVA, A., MALIK, L., SMITH, T., SUDBERY, I. and PATRO, R. (2019). Alevin efficiently estimates accurate gene abundances from dscRNA-seq data. *Genome Biol.* **20** Art. ID 65. <https://doi.org/10.1186/s13059-019-1670-y>
- [199] STEGLE, O., TEICHMANN, S. A. and MARIONI, J. C. (2015). Computational and analytical challenges in single-cell transcriptomics. *Nat. Rev. Genet.* **16** 133–145. <https://doi.org/10.1038/nrg3833>
- [200] STOECKIUS, M., ZHENG, S., HOUCK-LOOMIS, B., HAO, S., YEUNG, B. Z., MAUCK, W. M., SMIBERT, P. and SATIJA, R. (2018). Cell Hashing with barcoded antibodies enables multiplexing and doublet detection for single cell genomics. *Genome Biol.* **19** Art. ID 224.
- [201] STREET, K., RISSO, D., FLETCHER, R. B., DAS, D., NGAI, J., YOSEF, N., PURDOM, E. and DUDOIT, S. (2018). Slingshot: Cell lineage and pseudotime inference for single-cell transcriptomics. *BMC Genomics* **19** Art. ID 477.
- [202] STUART, J. M., SEGAL, E., KOLLER, D. and KIM, S. K. (2003). A gene-coexpression network for global discovery of conserved genetic modules. *Science* **302** 249–255.
- [203] SUN, W. W. and LI, L. (2017). STORE: Sparse tensor response regression and neuroimaging analysis. *J. Mach. Learn. Res.* **18** Art. ID 135. <https://doi.org/10.1162/jmlr.2017.18.1.135>
- [204] SUN, W. W. and LI, L. (2019). Dynamic tensor clustering. *J. Amer. Statist. Assoc.* **114** 1894–1907. <https://doi.org/10.1080/01621459.2018.1527701>
- [205] TAMAYO, P., SLONIM, D., MESIROV, J., ZHU, Q., KITAREEWAN, S., DMITROVSKY, E., LANDER, E. S. and GOLUB, T. R. (1999). Interpreting patterns of gene expression with self-organizing maps: Methods and application to hematopoietic differentiation. *Proc. Natl. Acad. Sci. USA* **96** 2907–2912.
- [206] TANG, F., BARBACIORU, C., WANG, Y., NORDMAN, E., LEE, C., XU, N., WANG, X., BODEAU, J., TUCH, B. B. et al. (2009). mRNA-Seq whole-transcriptome analysis of a single cell. *Nat. Methods* **6** 377–382.
- [207] TAVAZOIE, S., HUGHES, J. D., CAMPBELL, M. J., CHO, R. J. and CHURCH, G. M. (1999). Systematic determination of genetic network architecture. *Nat. Genet.* **22** 281–285.
- [208] TESCHENDORFF, A. E., WANG, Y., BARBOSA-MORAIS, N. L., BRENTON, J. D. and CALDAS, C. (2005). A variational Bayesian mixture modelling framework for cluster analysis of gene-expression data. *Bioinformatics* **21** 3025–3033.
- [209] TIAN, T. and BURRAGE, K. (2006). Stochastic models for regulatory networks of the genetic toggle switch. *Proc. Natl. Acad. Sci. USA* **103** 8372–8377.
- [210] TIBSHIRANI, R., WALTHER, G. and HASTIE, T. (2001). Estimating the number of clusters in a data set via the gap statistic. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **63** 411–423. <https://doi.org/10.1111/1467-9868.00293>

- [211] TOMASI, D. and VOLKOW, N. D. (2012). Abnormal functional connectivity in children with attention-deficit/hyperactivity disorder. *Biol. Psychiatry* **71** 443–450.
- [212] TONG, A. H. Y., LESAGE, G., BADER, G. D., DING, H., XU, H., XIN, X., YOUNG, J., BERRIZ, G. F., BROST, R. L. et al. (2004). Global mapping of the yeast genetic interaction network. *Science* **303** 808–813.
- [213] TRANCHEVENT, L.-C., ARDESHIRDAVANI, A., ELSHAL, S., ALCAIDE, D., AERTS, J., AUBOEUF, D. and MOREAU, Y. (2016). Candidate gene prioritization with Endeavour. *Nucleic Acids Res.* **44** W117–W121. <https://doi.org/10.1093/nar/gkw365>
- [214] TRANCHEVENT, L.-C., BARRIOT, R., YU, S. and VOOREN, S. V. (2006). Gene prioritization through genomic data fusion. *Nat. Biotechnol.* **24** 537–544.
- [215] TRAPNELL, C., CACCHIARELLI, D., GRIMSBY, J., POKHAREL, P., LI, S., MORSE, M., LENNON, N. J., LIVAK, K. J., MIKKELSEN, T. S. et al. (2014). The dynamics and regulators of cell fate decisions are revealed by pseudotemporal ordering of single cells. *Nat. Biotechnol.* **32** 381–386.
- [216] TZOURIO-MAZOYER, N., LANDEAU, B., PAPATHANASSIOU, D., CRIVELLO, F., ETARD, O., DELCROIX, N. et al. (2002). Automated anatomical labeling of activations in SPM using a macroscopic anatomical parcellation of the MNI MRI single-subject brain. *NeuroImage* **15** 273–289.
- [217] VALLEJOS, C. A., RISSO, D., SCIALDONE, A., DUDOIT, S. and MARIONI, J. C. (2017). Normalizing single-cell RNA sequencing data: Challenges and opportunities. *Nat. Methods* **14** 565–571. <https://doi.org/10.1038/nmeth.4292>
- [218] VAN DIJK, D., SHARMA, R., NAINYS, J., YIM, K., KATHAIL, P., CARR, A. J., BURDZIAK, C., MOON, K. R., CHAFFER, C. L. et al. (2018). Recovering gene interactions from single-cell data using data diffusion. *Cell* **174** 716–729.
- [219] VAN DONGEN, S. M. (2000). Graph clustering by flow simulation. Ph.D. thesis.
- [220] VAROQUAUX, G. and CRADDOCK, R. C. (2013). Learning and comparing functional connectomes across subjects. *NeuroImage* **80** 405–415.
- [221] VINH, N. X., CHETTY, M., COPPEL, R. and WANGIKAR, P. P. (2011). GlobalMIT: Learning globally optimal dynamic Bayesian network with the mutual information test criterion. *Bioinformatics* **27** 2765–2766.
- [222] VLASBLOM, J. and WODAK, S. J. (2009). Markov clustering versus affinity propagation for the partitioning of protein interaction graphs. *BMC Bioinform.* **10** Art. ID 99. <https://doi.org/10.1186/1471-2105-10-99>
- [223] WACHI, S., YONEDA, K. and WU, R. (2005). Interactome-transcriptome analysis reveals the high centrality of genes differentially expressed in lung cancer tissues. *Bioinformatics* **21** 4205–4208.
- [224] WANG, Y., HU, L. and OMBAO, H. (2016). Statistical analysis of electroencephalograms. In *Handbook of Neuroimaging Data Analysis* (H. Ombao, M. Linquist, W. Thompson and J. Aston, eds.) 523–565. Chapman & Hall/CRC, Boca Raton, FL.
- [225] WANG, Y., JOSHI, T., ZHANG, X.-S., XU, D. and CHEN, L. (2006). Inferring gene regulatory networks from multiple microarray datasets. *Bioinformatics* **22** 2413–2420.
- [226] WANG, Y., KANG, J., KEMMER, P. B. and GUO, Y. (2016). An efficient and reliable statistical method for estimating functional connectivity in large scale brain networks using partial correlation. *Front. Neurosci.* **10** Art. ID 123.
- [227] WANG, Y., ZHANG, X.-S. and XIA, Y. (2009). Predicting eukaryotic transcriptional cooperativity by Bayesian network integration of genome-wide data. *Nucleic Acids Res.* **37** 5943–5958.
- [228] WANG, Y. R. and HUANG, H. (2014). Review on statistical methods for gene network reconstruction using expression data. *J. Theoret. Biol.* **362** 53–61.
- [229] WANG, Y. R., LIU, K., THEUSCH, E., ROTTER, J. I., MEDINA, M. W., WATERMAN, M. S. and HUANG, H. (2017). Generalized correlation measure using count statistics for gene expression data with ordered samples. *Bioinformatics* **34** 617–624.
- [230] WANG, Y. R., WATERMAN, M. S. and HUANG, H. (2014). Gene coexpression measures in large heterogeneous samples using count statistics. *Proc. Natl. Acad. Sci. USA* **111** 16371–16376.
- [231] WANG, Y. X. R., JIANG, K., FELDMAN, L. J., BICKEL, P. J. and HUANG, H. (2015). Inferring gene–gene interactions and functional modules using sparse canonical correlation analysis. *Ann. Appl. Stat.* **9** 300–323. MR3341117 <https://doi.org/10.1214/14-AOAS792>
- [232] WANG, Y. X. R., SARKAR, P., URSU, O., KUNDAJE, A. and BICKEL, P. J. (2019). Network modelling of topological domains using Hi-C data. *Ann. Appl. Stat.* **13** 1511–1536. MR4019148 <https://doi.org/10.1214/19-AOAS1244>
- [233] WERNICKE, S. (2006). Efficient detection of network motifs. *IEEE/ACM Trans. Comput. Biol. Bioinform.* **3** 347–359.
- [234] WILLE, A. and BÜHLMANN, P. (2006). Low-order conditional independence graphs for inferring genetic networks. *Stat. Appl. Genet. Mol. Biol.* **5** Art. ID 1. MR2221304 <https://doi.org/10.2202/1544-6115.1170>
- [235] WILLE, A., ZIMMERMANN, P., VRANOVÁ, E., FÜRHOLZ, A., LAULE, O., BLEULER, S., HENNIG, L., PRELIĆ, A., VON ROHR, P. et al. (2004). Sparse graphical Gaussian modeling of the isoprenoid gene network in *Arabidopsis thaliana*. *Genome Biol.* **5** Art. ID R92.
- [236] WOLFE, C. J., KOHANE, I. S. and BUTTE, A. J. (2005). Systematic survey reveals general applicability of “guilt-by-association” within gene coexpression networks. *BMC Bioinform.* **6** Art. ID 227. <https://doi.org/10.1186/1471-2105-6-227>
- [237] WONG, A. K., KRISHNAN, A. and TROYANSKAYA, O. G. (2018). GIANT 2.0: Genome-scale integrated analysis of gene networks in tissues. *Nucleic Acids Res.* **46** W65–W70. <https://doi.org/10.1093/nar/gky408>
- [238] WONG, E., BAUR, B., QUADER, S. and HUANG, C.-H. (2011). Biological network motif detection: Principles and practice. *Brief. Bioinform.* **13** 202–215.
- [239] WONG, R. K. W., LEE, T. C. M., PAUL, D., PENG, J. and ALZHEIMER’S DISEASE NEUROIMAGING INITIATIVE (2016). Fiber direction estimation, smoothing and tracking in diffusion MRI. *Ann. Appl. Stat.* **10** 1137–1156. MR3553216 <https://doi.org/10.1214/15-AOAS880>
- [240] WU, Z., ZHANG, Y., STITZEL, M. L. and WU, H. (2018). Two-phase differential expression analysis for single cell RNA-seq. *Bioinformatics* **34** 3340–3348.
- [241] XIA, Y. and LI, L. (2017). Hypothesis testing of matrix graph model with application to brain connectivity analysis. *Bio-metrics* **73** 780–791. MR3713112 <https://doi.org/10.1111/biom.12633>
- [242] XIA, Y. and LI, L. (2019). Matrix graph hypothesis testing and application in brain connectivity alternation detection. *Statist. Sinica* **29** 303–328. MR3889369
- [243] XIONG, Q., ANCONA, N., HAUSER, E. R., MUKHERJEE, S. and FUREY, T. S. (2012). Integrating genetic and gene expression evidence into genome-wide association analysis of gene sets. *Genome Res.* **22** 386–397.
- [244] XU, C. and SU, Z. (2015). Identification of cell types from single-cell transcriptomes using a novel clustering method. *Bioinformatics* **31** 1974–1980.

- [245] XU, J. and LI, Y. (2006). Discovering disease-genes by topological features in human protein-protein interaction network. *Bioinformatics* **22** 2800–2805.
- [246] XU, Y. and LINDQUIST, M. A. (2015). Dynamic connectivity detection: An algorithm for determining functional connectivity change points in fMRI data. *Front. Neurosci.* **9** Art. ID 285. <https://doi.org/10.3389/fnins.2015.00285>
- [247] YAN, K.-K., LOU, S. and GERSTEIN, M. (2017). MrTADFinder: A network modularity based approach to identify topologically associating domains in multiple resolutions. *PLoS Comput. Biol.* **13** Art. ID e1005647. <https://doi.org/10.1371/journal.pcbi.1005647>
- [248] YEUNG, K. Y., FRALEY, C., MURUA, A., RAFTERY, A. E. and RUZZO, W. L. (2001). Model-based clustering and data transformations for gene expression data. *Bioinformatics* **17** 977–987.
- [249] YU, H., KIM, P. M., SPRECHER, E., TRIFONOV, V. and GERSTEIN, M. (2007). The importance of bottlenecks in protein networks: Correlation with gene essentiality and expression dynamics. *PLoS Comput. Biol.* **3** Art. ID e59. [MR2373048 https://doi.org/10.1371/journal.pcbi.0030059](https://doi.org/10.1371/journal.pcbi.0030059)
- [250] YU, J., SMITH, V. A., WANG, P. P., HARTEMINK, A. J. and JARVIS, E. D. (2002). Using Bayesian network inference algorithms to recover molecular genetic regulatory networks. In *International Conference on Systems Biology* **2002**.
- [251] YUAN, M. and LIN, Y. (2007). Model selection and estimation in the Gaussian graphical model. *Biometrika* **94** 19–35. [MR2367824 https://doi.org/10.1093/biomet/asm018](https://doi.org/10.1093/biomet/asm018)
- [252] YUAN, Y., CHEN, Y.-P. P., NI, S., XU, A. G., TANG, L., VINGRON, M., SOMEL, M. and KHAITOVICH, P. (2011). Development and application of a modified dynamic time warping algorithm (DTW-S) to analyses of primate brain expression time series. *BMC Bioinform.* **12** Art. ID 347.
- [253] ZHANG, B., LI, H., RIGGINS, R. B., ZHAN, M., XUAN, J., ZHANG, Z., HOFFMAN, E. P., CLARKE, R. and WANG, Y. (2008). Differential dependency network analysis to identify condition-specific topological changes in biological networks. *Bioinformatics* **25** 526–532.
- [254] ZHANG, L. and ZHANG, S. (2018). Comparison of computational methods for imputing single-cell RNA-sequencing data. *IEEE/ACM Trans. Comput. Biol. Bioinform.* **17** 376–389.
- [255] ZHANG, T., WU, J., LI, F., CAFFO, B. and BOATMAN-REICH, D. (2015). A dynamic directional model for effective brain connectivity using electrocorticographic (ECoG) time series. *J. Amer. Statist. Assoc.* **110** 93–106. [MR3338489 https://doi.org/10.1080/01621459.2014.988213](https://doi.org/10.1080/01621459.2014.988213)
- [256] ZHANG, Z., DESCOTEAUX, M., ZHANG, J., GIRARD, G., CHAMBERLAND, M., DUNSON, D., SRIVASTAVA, A. and ZHU, H. (2018). Mapping population-based structural connectomes. *NeuroImage* **172** 130–145.
- [257] ZHAO, F., MCCARRICK-WALMSLEY, R., ÅKERBLAD, P., SIGVARDSSON, M. and KADESCH, T. (2003). Inhibition of p300/CBP by early B-cell factor. *Mol. Cell. Biol.* **23** 3837–3846.
- [258] ZHENG, G. X., TERRY, J. M., BELGRADER, P., RYVKIN, P., BENT, Z. W., WILSON, R., ZIRALDO, S. B., WHEELER, T. D., MCDERMOTT, G. P. et al. (2017). Massively parallel digital transcriptional profiling of single cells. *Nat. Commun.* **8** Art. ID 14049.
- [259] ZHOU, S., RÜTIMANN, P., XU, M. and BÜHLMANN, P. (2011). High-dimensional covariance estimation based on Gaussian graphical models. *J. Mach. Learn. Res.* **12** 2975–3026. [MR2854354 https://doi.org/10.1162/jmlr.2011.12.1](https://doi.org/10.1162/jmlr.2011.12.1)
- [260] ZHU, H., CHEN, Y., IBRAHIM, J. G., LI, Y., HALL, C. and LIN, W. (2009). Intrinsic regression models for positive-definite matrices with applications to diffusion tensor imaging. *J. Amer. Statist. Assoc.* **104** 1203–1212. [MR2750245 https://doi.org/10.1198/jasa.2009.tm08096](https://doi.org/10.1198/jasa.2009.tm08096)
- [261] ZHU, J., CHEN, Y., LEONARDSON, A. S., WANG, K., LAMB, J. R., EMILSSON, V. and SCHADT, E. E. (2010). Characterizing dynamic changes in the human blood transcriptional network. *PLoS Comput. Biol.* **6** Art. ID e1000671.
- [262] ZHU, Y. and LI, L. (2018). Multiple matrix Gaussian graphs estimation. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **80** 927–950. [MR3874304 https://doi.org/10.1111/rssb.12278](https://doi.org/10.1111/rssb.12278)
- [263] ZIEGENHAIN, C., VIETH, B., PAREKH, S., REINIUS, B., GUILLAUMET-ADKINS, A., SMETS, M., LEONHARDT, H., HEYN, H., HELLMANN, I. et al. (2017). Comparative analysis of single-cell RNA sequencing methods. *Mol. Cell* **65** 631–643.
- [264] ZOTENKO, E., MESTRE, J., O’LEARY, D. P. and PRZYTYCKA, T. M. (2008). Why do hubs in the yeast protein interaction network tend to be essential: Reexamining the connection between the network topology and essentiality. *PLoS Comput. Biol.* **4** Art. ID e1000140. [MR2443413 https://doi.org/10.1371/journal.pcbi.1000140](https://doi.org/10.1371/journal.pcbi.1000140)
- [265] ZOU, M. and CONZEN, S. D. (2004). A new dynamic Bayesian network (DBN) approach for identifying gene regulatory networks from time course microarray data. *Bioinformatics* **21** 71–79.

Bipartite Causal Inference with Interference

Corwin M. Zigler and Georgia Papadogeorgou

Abstract. Statistical methods to evaluate the effectiveness of interventions are increasingly challenged by the inherent interconnectedness of units. Specifically, a recent flurry of methods research has addressed the problem of *interference* between observations, which arises when one observational unit's outcome depends not only on its treatment but also the treatment assigned to other units. We introduce the setting of *bipartite causal inference with interference*, which arises when (1) treatments are defined on observational units that are distinct from those at which outcomes are measured and (2) there is *interference* between units in the sense that outcomes for some units depend on the treatments assigned to many other units. The focus of this work is to formulate definitions and several possible causal estimands for this setting, highlighting similarities and differences with more commonly considered settings of causal inference with interference. Toward an empirical illustration, an inverse probability of treatment weighted estimator is adapted from existing literature to estimate a subset of simplified, but interesting, estimands. The estimators are deployed to evaluate how interventions to reduce air pollution from 473 power plants in the U.S. causally affect cardiovascular hospitalization among Medicare beneficiaries residing at 18,807 zip code locations.

Key words and phrases: Air pollution, causal inference, interference, network dependence, power plants.

REFERENCES

- [1] ARONOW, P. M. (2012). A general method for detecting interference between units in randomized experiments. *Sociol. Methods Res.* **41** 3–16. MR3190698 <https://doi.org/10.1177/0049124112437535>
- [2] ARONOW, P. M. and SAMII, C. (2017). Estimating average causal effects under general interference, with application to a social network experiment. *Ann. Appl. Stat.* **11** 1912–1947. MR3743283 <https://doi.org/10.1214/16-AOAS1005>
- [3] BELL, M. L., MCDERMOTT, A., ZEGER, S. L., SAMET, J. M. and DOMINICI, F. (2004). Ozone and short-term mortality in 95 US urban communities, 1987–2000. *JAMA J. Am. Med. Assoc.* **292** 2372–2378.
- [4] BOWERS, J., FREDRICKSON, M. M. and PANAGOPOULOS, C. (2013). Reasoning about interference between units: A general framework. *Polit. Anal.* **21** 97–124.
- [5] FORASTIERE, L., AIROLDI, E. M. and MEALLI, F. (2016). Identification and estimation of treatment and interference effects in observational studies on networks. Preprint. Available at [arXiv:1609.06245](https://arxiv.org/abs/1609.06245) [stat].
- [6] GRAHAM, D. J., MCCOY, E. J. and STEPHENS, D. A. (2013). Quantifying the effect of area deprivation on child pedestrian casualties by using longitudinal mixed models to adjust for confounding, interference and spatial dependence. *J. Roy. Statist. Soc. Ser. A* **176** 931–950. MR3120956 <https://doi.org/10.1111/j.1467-985X.2012.01071.x>
- [7] HALLORAN, M. E. and STRUCHINER, C. J. (1991). Study designs for dependent happenings. *Epidemiology* **2** 331–338.
- [8] HALLORAN, M. E. and STRUCHINER, C. J. (1995). Causal inference in infectious diseases. *Epidemiology* **6** 142.
- [9] HONG, G. and RAUDENBUSH, S. W. (2006). Evaluating kindergarten retention policy: A case study of causal inference for multilevel observational data. *J. Amer. Statist. Assoc.* **101** 901–910. MR2324091 <https://doi.org/10.1198/016214506000000447>
- [10] HUDGENS, M. G. and HALLORAN, M. E. (2008). Toward causal inference with interference. *J. Amer. Statist. Assoc.* **103** 832–842. MR2435472 <https://doi.org/10.1198/016214508000000292>
- [11] KARWA, V. and AIROLDI, E. (2018). A systematic investigation of classical causal inference strategies under misspecification due to network interference. Preprint. Available at [arXiv:1810.08259v1](https://arxiv.org/abs/1810.08259v1) [stat.ME].
- [12] KIM, C., DANIELS, M. J., HOGAN, J. W., CHOIRAT, C. and ZIGLER, C. M. (2019). Bayesian methods for multiple mediators: Relating principal stratification and causal mediation in the analysis of power plant emission controls. *Ann. Appl. Stat.* **13** 1927–1956. MR4019162 <https://doi.org/10.1214/19-AOAS1260>

Corwin M. Zigler is Associate Professor of Statistics and Data Sciences, Dell Medical School, University of Texas at Austin, 2317 Speedway D9800, Austin, Texas 78712-1823, USA (e-mail: cory.zigler@austin.utexas.edu). Georgia Papadogeorgou is Postdoctoral Associate, Department of Statistical Science, Duke University, 206 Old Chem Bldg, Durham, North Carolina 27708, USA (e-mail: gp118@duke.edu).

- [13] LIU, L. and HUDGENS, M. G. (2014). Large sample randomization inference of causal effects in the presence of interference. *J. Amer. Statist. Assoc.* **109** 288–301. MR3180564 <https://doi.org/10.1080/01621459.2013.844698>
- [14] LIU, L., HUDGENS, M. G. and BECKER-DREPS, S. (2016). On inverse probability-weighted estimators in the presence of interference. *Biometrika* **103** 829–842. MR3620442 <https://doi.org/10.1093/biomet/asw047>
- [15] LUO, X., SMALL, D. S., LI, C.-S. R. and ROSENBAUM, P. R. (2012). Inference with interference between units in an fMRI experiment of motor inhibition. *J. Amer. Statist. Assoc.* **107** 530–541. MR2980065 <https://doi.org/10.1080/01621459.2012.655954>
- [16] OGBURN, E. L. and VANDERWEELE, T. J. (2017). Vaccines, contagion, and social networks. *Ann. Appl. Stat.* **11** 919–948. MR3693552 <https://doi.org/10.1214/17-AOAS1023>
- [17] PAPADOGEORGOU, G., CHOIRAT, C. and ZIGLER, C. M. (2019). Adjusting for unmeasured spatial confounding with distance adjusted propensity score matching. *Biostatistics* **20** 256–272. MR3922132 <https://doi.org/10.1093/biostatistics/kxx074>
- [18] PAPADOGEORGOU, G., MEALLI, F. and ZIGLER, C. M. (2019). Causal inference with interfering units for cluster and population level treatment allocation programs. *Biometrics* **75** 778–787.
- [19] PEREZ-HEYDRICH, C., HUDGENS, M. G., HALLORAN, M. E., CLEMENS, J. D., ALI, M. and EMCH, M. E. (2014). Assessing effects of cholera vaccination in the presence of interference. *Biometrics* **70** 734–744. MR3261790 <https://doi.org/10.1111/biom.12184>
- [20] ROSENBAUM, P. R. (2007). Interference between units in randomized experiments. *J. Amer. Statist. Assoc.* **102** 191–200. MR2345537 <https://doi.org/10.1198/016214506000001112>
- [21] RUBIN, D. B. (1990). Comment on J. Neyman and causal inference in experiments and observational studies: “On the application of probability theory to agricultural experiments. Essay on principles. Section 9” [Ann. Agric. Sci. **10** (1923), 1–51]. *Statist. Sci.* **5** 472–480. MR1092987
- [22] SOBEL, M. E. (2006). What do randomized studies of housing mobility demonstrate?: Causal inference in the face of interference. *J. Amer. Statist. Assoc.* **101** 1398–1407. MR2307573 <https://doi.org/10.1198/016214506000000636>
- [23] TCHETGEN TCHETGEN, E. J. and VANDERWEELE, T. J. (2012). On causal inference in the presence of interference. *Stat. Methods Med. Res.* **21** 55–75. MR2867538 <https://doi.org/10.1177/0962280210386779>
- [24] VERBITSKY-SAVITZ, N. and RAUDENBUSH, S. W. (2012). Causal inference under interference in spatial settings: A case study evaluating community policing program in Chicago. *Epidemiol. Methods* **1** 107–130.
- [25] WARD, J. H. JR. (1963). Hierarchical grouping to optimize an objective function. *J. Amer. Statist. Assoc.* **58** 236–244. MR0148188
- [26] ZIGLER, C. M., DOMINICI, F. and WANG, Y. (2012). Estimating causal effects of air quality regulations using principal stratification for spatially correlated multivariate intermediate outcomes. *Biostatistics* **13** 289–302.

A General Framework for Vecchia Approximations of Gaussian Processes

Matthias Katzfuss and Joseph Guinness

Abstract. Gaussian processes (GPs) are commonly used as models for functions, time series, and spatial fields, but they are computationally infeasible for large datasets. Focusing on the typical setting of modeling data as a GP plus an additive noise term, we propose a generalization of the Vecchia (*J. Roy. Statist. Soc. Ser. B* **50** (1988) 297–312) approach as a framework for GP approximations. We show that our general Vecchia approach contains many popular existing GP approximations as special cases, allowing for comparisons among the different methods within a unified framework. Representing the models by directed acyclic graphs, we determine the sparsity of the matrices necessary for inference, which leads to new insights regarding the computational properties. Based on these results, we propose a novel sparse general Vecchia approximation, which ensures computational feasibility for large spatial datasets but can lead to considerable improvements in approximation accuracy over Vecchia’s original approach. We provide several theoretical results and conduct numerical comparisons. We conclude with guidelines for the use of Vecchia approximations in spatial statistics.

Key words and phrases: Computational complexity, covariance approximation, directed acyclic graphs, large datasets, sparsity, spatial statistics.

REFERENCES

- BANERJEE, S., CARLIN, B. P. and GELFAND, A. E. (2015). *Hierarchical Modeling and Analysis for Spatial Data*, 2nd ed. *Monographs on Statistics and Applied Probability* **135**. CRC Press, Boca Raton, FL. [MR3362184](#)
- BANERJEE, S., GELFAND, A. E., FINLEY, A. O. and SANG, H. (2008). Gaussian predictive process models for large spatial data sets. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **70** 825–848. [MR2523906](#) <https://doi.org/10.1111/j.1467-9868.2008.00663.x>
- CRESSIE, N. and DAVIDSON, J. L. (1998). Image analysis with partially ordered Markov models. *Comput. Statist. Data Anal.* **29** 1–26. [MR1665674](#) [https://doi.org/10.1016/S0167-9473\(98\)00052-8](https://doi.org/10.1016/S0167-9473(98)00052-8)
- CRESSIE, N. and JOHANNESSON, G. (2008). Fixed rank kriging for very large spatial data sets. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **70** 209–226. [MR2412639](#) <https://doi.org/10.1111/j.1467-9868.2007.00633.x>
- CRESSIE, N. and WIKLE, C. K. (2011). *Statistics for Spatio-Temporal Data*. *Wiley Series in Probability and Statistics*. Wiley, Hoboken, NJ. [MR2848400](#)
- DATTA, A., BANERJEE, S., FINLEY, A. O. and GELFAND, A. E. (2016a). Hierarchical nearest-neighbor Gaussian process models for large geostatistical datasets. *J. Amer. Statist. Assoc.* **111** 800–812. [MR3538706](#) <https://doi.org/10.1080/01621459.2015.1044091>
- DATTA, A., BANERJEE, S., FINLEY, A. O. and GELFAND, A. E. (2016b). On nearest-neighbor Gaussian process models for massive spatial data. *Wiley Interdiscip. Rev.: Comput. Stat.* **8** 162–171. [MR3544254](#) <https://doi.org/10.1002/wics.1383>
- DATTA, A., BANERJEE, S., FINLEY, A. O., HAMM, N. A. S. and SCHAAP, M. (2016c). Nonseparable dynamic nearest neighbor Gaussian process models for large spatio-temporal data with an application to particulate matter analysis. *Ann. Appl. Stat.* **10** 1286–1316. [MR3553225](#) <https://doi.org/10.1214/16-AOAS931>
- DU, J., ZHANG, H. and MANDREKAR, V. S. (2009). Fixed-domain asymptotic properties of tapered maximum likelihood estimators. *Ann. Statist.* **37** 3330–3361. [MR2549562](#) <https://doi.org/10.1214/08-AOS676>
- EIDSVIK, J., SHABY, B. A., REICH, B. J., WHEELER, M. and NIEMI, J. (2014). Estimation and prediction in spatial models with block composite likelihoods. *J. Comput. Graph. Statist.* **23** 295–315. [MR3215812](#) <https://doi.org/10.1080/10618600.2012.760460>
- EUBANK, R. L. and WANG, S. (2002). The equivalence between the Cholesky decomposition and the Kalman filter. *Amer. Statist.* **56** 39–43. [MR1939395](#) <https://doi.org/10.1198/000313002753631349>
- FINLEY, A. O., SANG, H., BANERJEE, S. and GELFAND, A. E. (2009). Improving the performance of predictive process modeling for large datasets. *Comput. Statist. Data Anal.* **53** 2873–2884. [MR2667597](#) <https://doi.org/10.1016/j.csda.2008.09.008>
- FINLEY, A. O., DATTA, A., COOK, B. C., MORTON, D. C., ANDERSEN, H. E. and BANERJEE, S. (2019). Efficient algorithms for Bayesian nearest neighbor Gaussian processes. *J. Comput. Graph. Statist.* **28** 401–414.

- FURRER, R., GENTON, M. G. and NYCHKA, D. (2006). Covariance tapering for interpolation of large spatial datasets. *J. Comput. Graph. Statist.* **15** 502–523. MR2291261 <https://doi.org/10.1198/106186006X132178>
- FURRER, R. and SAIN, S. R. (2010). spam: A sparse matrix R package with emphasis on MCMC methods for Gaussian Markov random fields. *J. Stat. Softw.* **36** 1–25.
- GERBER, F., FURRER, R., SCHAEPMAN-STRUB, G., DE JONG, R. and SCHAEPMAN, M. E. (2018). Predicting missing values in spatio-temporal satellite data. *IEEE Trans. Geosci. Remote Sens.* **56** 2841–2853.
- GRAMACY, R. B. and APLEY, D. W. (2015). Local Gaussian process approximation for large computer experiments. *J. Comput. Graph. Statist.* **24** 561–578. MR3357395 <https://doi.org/10.1080/10618600.2014.914442>
- GRAMACY, R. B. and LEE, H. K. H. (2012). Cases for the nugget in modeling computer experiments. *Stat. Comput.* **22** 713–722. MR2909617 <https://doi.org/10.1007/s11222-010-9224-x>
- GUHANIYOGI, R. and BANERJEE, S. (2018). Meta-kriging: Scalable Bayesian modeling and inference for massive spatial datasets. *Technometrics* **60** 430–444. MR3878099 <https://doi.org/10.1080/00401706.2018.1437474>
- GUINNESS, J. (2018). Permutation and grouping methods for sharpening Gaussian process approximations. *Technometrics* **60** 415–429. MR3878098 <https://doi.org/10.1080/00401706.2018.1437476>
- HEATON, M. J., DATTA, A., FINLEY, A. O., FURRER, R., GUINNESS, J., GUHANIYOGI, R., GERBER, F., GRAMACY, R. B., HAMMERLING, D. et al. (2019). A case study competition among methods for analyzing large spatial data. *J. Agric. Biol. Environ. Stat.* **24** 398–425.
- HIGDON, D. (1998). A process-convolution approach to modelling temperatures in the North Atlantic Ocean. *Environ. Ecol. Stat.* **5** 173–190.
- HUANG, H. and SUN, Y. (2018). Hierarchical low rank approximation of likelihoods for large spatial datasets. *J. Comput. Graph. Statist.* **27** 110–118. MR3788305 <https://doi.org/10.1080/10618600.2017.1356324>
- JUREK, M. and KATZFUSS, M. (2018). Multi-resolution filters for massive spatio-temporal data. Available at [arXiv:1810.04200](https://arxiv.org/abs/1810.04200).
- KALMAN, R. E. (1960). A new approach to linear filtering and prediction problems. *Trans. ASME Ser. D. J. Basic Eng.* **82** 35–45. MR3931993
- KATZFUSS, M. (2017). A multi-resolution approximation for massive spatial datasets. *J. Amer. Statist. Assoc.* **112** 201–214. MR3646566 <https://doi.org/10.1080/01621459.2015.1123632>
- KATZFUSS, M. and CRESSIE, N. (2011). Spatio-temporal smoothing and EM estimation for massive remote-sensing data sets. *J. Time Series Anal.* **32** 430–446. MR2841794 <https://doi.org/10.1111/j.1467-9892.2011.00732.x>
- KATZFUSS, M. and GONG, W. (2020). A class of multi-resolution approximations for large spatial datasets. *Statist. Sinica* **30** 2203–2226.
- KATZFUSS, M., GUINNESS, J., GONG, W. and ZILBER, D. (2020). Vecchia approximations of Gaussian-process predictions. *J. Agric. Biol. Environ. Stat.* **25** 383–414.
- KAUFMAN, C. G., SCHERVISH, M. J. and NYCHKA, D. W. (2008). Covariance tapering for likelihood-based estimation in large spatial data sets. *J. Amer. Statist. Assoc.* **103** 1545–1555. MR2504203 <https://doi.org/10.1198/016214508000000959>
- LAURITZEN, S. L. (1996). *Graphical Models*. Oxford Statistical Science Series **17**. The Clarendon Press, Oxford University Press, New York. MR1419991
- LINDGREN, F., RUE, H. and LINDSTRÖM, J. (2011). An explicit link between Gaussian fields and Gaussian Markov random fields: The stochastic partial differential equation approach. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **73** 423–498. MR2853727 <https://doi.org/10.1111/j.1467-9868.2011.00777.x>
- MINDEN, V., DAMLE, A., HO, K. L. and YING, L. (2017). Fast spatial Gaussian process maximum likelihood estimation via skeletonization factorizations. *Multiscale Model. Simul.* **15** 1584–1611. MR3720352 <https://doi.org/10.1137/17M1116477>
- NYCHKA, D., BANDYOPADHYAY, S., HAMMERLING, D., LINDGREN, F. and SAIN, S. (2015). A multiresolution Gaussian process model for the analysis of large spatial datasets. *J. Comput. Graph. Statist.* **24** 579–599. MR3357396 <https://doi.org/10.1080/10618600.2014.914946>
- QUINONERO-CANDELA, J. and RASMUSSEN, C. E. (2005). A unifying view of sparse approximate Gaussian process regression. *J. Mach. Learn. Res.* **6** 1939–1959. MR2249877
- RASMUSSEN, C. E. and WILLIAMS, C. K. I. (2006). *Gaussian Processes for Machine Learning*. Adaptive Computation and Machine Learning. MIT Press, Cambridge, MA. MR2514435
- RAUCH, H. E., TUNG, F. and STRIEBEL, C. T. (1965). Maximum likelihood estimates of linear dynamic systems. *AIAA J.* **3** 1445–1450. MR0181489 <https://doi.org/10.2514/3.3166>
- RUE, H. and HELD, L. (2005). *Gaussian Markov Random Fields: Theory and Applications*. Monographs on Statistics and Applied Probability **104**. CRC Press/CRC, Boca Raton, FL. MR2130347 <https://doi.org/10.1201/9780203492024>
- RUE, H. and HELD, L. (2010). Discrete spatial variation. In *Handbook of Spatial Statistics*. Chapman & Hall/CRC Handb. Mod. Stat. Methods 171–200. CRC Press, Boca Raton, FL. MR2730942 <https://doi.org/10.1201/9781420072884-c12>
- RÜTIMANN, P. and BÜHLMANN, P. (2009). High dimensional sparse covariance estimation via directed acyclic graphs. *Electron. J. Stat.* **3** 1133–1160. MR2566184 <https://doi.org/10.1214/09-EJS534>
- SANG, H. and HUANG, J. Z. (2012). A full scale approximation of covariance functions for large spatial data sets. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **74** 111–132. MR2885842 <https://doi.org/10.1111/j.1467-9868.2011.01007.x>
- SANG, H., JUN, M. and HUANG, J. Z. (2011). Covariance approximation for large multivariate spatial data sets with an application to multiple climate model errors. *Ann. Appl. Stat.* **5** 2519–2548. MR2907125 <https://doi.org/10.1214/11-AOAS478>
- SHABY, B. A. (2014). The open-faced sandwich adjustment for MCMC using estimating functions. *J. Comput. Graph. Statist.* **23** 853–876. MR3224659 <https://doi.org/10.1080/10618600.2013.842174>
- SNELSON, E. and GHAHRAMANI, Z. (2007). Local and global sparse Gaussian process approximations. In *Proceedings of the Eleventh International Conference on Artificial Intelligence and Statistics (AISTATS)* 524–531.
- STEIN, M. L. (2011). 2010 Rietz Lecture: When does the screening effect hold? *Ann. Statist.* **39** 2795–2819. MR3012392 <https://doi.org/10.1214/11-AOS909>
- STEIN, M. L. (2014). Limitations on low rank approximations for covariance matrices of spatial data. *Spat. Stat.* **8** 1–19. MR3326818 <https://doi.org/10.1016/j.spasta.2013.06.003>
- STEIN, M. L., CHI, Z. and WELTY, L. J. (2004). Approximating likelihoods for large spatial data sets. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **66** 275–296. MR2062376 <https://doi.org/10.1046/j.1369-7412.2003.05512.x>
- SUN, Y. and STEIN, M. L. (2016). Statistically and computationally efficient estimating equations for large spatial datasets. *J. Comput. Graph. Statist.* **25** 187–208. MR3474043 <https://doi.org/10.1080/10618600.2014.975230>
- TOLEDO, S. (2007). Lecture notes on combinatorial preconditioners, Chapter 3. Available at <http://www.tau.ac.il/~stoledo/Support/chapter-direct.pdf>.

- VARIN, C., REID, N. and FIRTH, D. (2011). An overview of composite likelihood methods. *Statist. Sinica* **21** 5–42. [MR2796852](#)
- VECCHIA, A. V. (1988). Estimation and model identification for continuous spatial processes. *J. Roy. Statist. Soc. Ser. B* **50** 297–312. [MR0964183](#)
- WIKLE, C. K. and CRESSIE, N. (1999). A dimension-reduced approach to space-time Kalman filtering. *Biometrika* **86** 815–829. [MR1741979](#) <https://doi.org/10.1093/biomet/86.4.815>
- ZHANG, H. (2012). Asymptotics and computation for spatial statistics. In *Advances and Challenges in Space-Time Modelling of Natural Events* 239–252. Springer, Berlin.
- ZHANG, B., SANG, H. and HUANG, J. Z. (2019). Smoothed full-scale approximation of Gaussian process models for computation of large spatial data sets. *Statist. Sinica* **29** 1711–1737. [MR3970332](#)

On General Notions of Depth for Regression

Yijun Zuo

Abstract. Depth notions in location have generated tremendous attention in the literature. In fact, data depth and its applications remain as one of the most active research topics in statistics over the last three decades. Most favored notions of depth in location include Tukey (In *Proceedings of the International Congress of Mathematicians (Vancouver, B.C., 1974)*, Vol. 2 (1975) 523–531) half-space depth (HD), Liu (*Ann. Statist.* **18** (1990) 405–414) simplicial depth and projection depth (PD) (Stahel (1981) and Donoho (1982), Liu (In *L_1 -Statistical Analysis and Related Methods (Neuchâtel, 1992)* (1992) 279–294 North-Holland), Zuo and Serfling (*Ann. Statist.* **28** (2000) 461–482) and (ZS00) and Zuo (*Ann. Statist.* **31** (2003) 1460–1490)), among others.

Depth notions in regression have also been proposed sporadically, nevertheless. The regression depth (RD) of Rousseeuw and Hubert (*J. Amer. Statist. Assoc.* **94** (1999) 388–433) (RH99), the most famous, exemplifies a direct extension of Tukey HD to regression. Other notions include Carrizosa (*J. Multivariate Anal.* **58** (1996) 21–26) and the ones proposed in this article via modifying a functional in Maronna and Yohai (*Ann. Statist.* **21** (1993) 965–990) (MY93). Is there any relationship between Carrizosa depth and the RD of RH99? Do these depth notions possess desirable properties? What are the desirable properties? Can existing notions serve well as depth notions in regression? These questions remain open.

The major objectives of the article include (i) revealing the connection between Carrizosa depth and RD of RH99; (ii) expanding location depth evaluating criteria in ZS00 for regression depth notions; (iii) examining the existing regression notions with respect to the gauges; and (iv) proposing the regression counterpart of the eminent location projection depth.

Key words and phrases: Depth, unfitness, linear regression, maximum depth regression estimating functionals, robustness.

REFERENCES

- ADROVER, J. G., MARONNA, R. A. and YOHAI, V. J. (2002). Relationships between maximum depth and projection regression estimates. *J. Statist. Plann. Inference* **105** 363–375. [MR1910058](https://doi.org/10.1016/S0378-3758(01)00264-6) [https://doi.org/10.1016/S0378-3758\(01\)00264-6](https://doi.org/10.1016/S0378-3758(01)00264-6)
- AGOSTINELLI, C. and ROMANAZZI, M. (2011). Local depth. *J. Statist. Plann. Inference* **141** 817–830. [MR2732952](https://doi.org/10.1016/j.jspi.2010.08.001) <https://doi.org/10.1016/j.jspi.2010.08.001>
- BAI, Z.-D. and HE, X. (1999). Asymptotic distributions of the maximal depth estimators for regression and multivariate location. *Ann. Statist.* **27** 1616–1637. [MR1742502](https://doi.org/10.1214/aos/1017939144) <https://doi.org/10.1214/aos/1017939144>
- BEATON, A. E. and TUKEY, J. W. (1974). The fitting of power series, meaning polynomials, illustrated on band-spectroscopic data. *Technometrics* **16** 147–185.
- CAPLIN, A. and NALEBUFF, B. (1988). On 64%-majority rule. *Econometrica* **56** 787–814. [MR0951758](https://doi.org/10.2307/1912699) <https://doi.org/10.2307/1912699>
- CAPLIN, A. and NALEBUFF, B. (1991a). Aggregation and social choice: A mean voter theorem. *Econometrica* **59** 1–23. [MR1085582](https://doi.org/10.2307/2938238) <https://doi.org/10.2307/2938238>
- CAPLIN, A. and NALEBUFF, B. (1991b). Aggregation and imperfect competition: On the existence of equilibrium. *Econometrica* **59** 25–59. [MR1085583](https://doi.org/10.2307/2938239) <https://doi.org/10.2307/2938239>
- CASCOS, I. and MOLCHANOV, I. (2007). Multivariate risks and depth-trimmed regions. *Finance and Stochastics* **11** 373–397. [MR2322918](https://doi.org/10.1007/s00780-007-0043-7) <https://doi.org/10.1007/s00780-007-0043-7>
- CARRIZOSA, E. (1996). A characterization of halfspace depth. *J. Multivariate Anal.* **58** 21–26. [MR1408126](https://doi.org/10.1006/jmva.1996.0037) <https://doi.org/10.1006/jmva.1996.0037>
- CHAKRABORTY, A. and CHAUDHURI, P. (2014). The spatial distribution in infinite dimensional spaces and related quantiles and depths. *Ann. Statist.* **42** 1203–1231. [MR3224286](https://doi.org/10.1214/14-AOS1226) <https://doi.org/10.1214/14-AOS1226>
- CHEBANA, F. and OUARDA, T. B. M. J. (2008). Depth and homogeneity in regional flood frequency analysis. *Water Resour. Res.* **44**

- W11422. <https://doi.org/10.1029/2007WR006771>
- CHEBANA, F. and OUARDA, T. B. M. J. (2011). Depth-based multivariate descriptive statistics with hydrological applications. *J. Geophys. Res.* **116** D10120. <https://doi.org/10.1029/2010JD015338>
- CHEN, M., GAO, C. and REN, Z. (2018). Robust covariance and scatter matrix estimation under Huber's contamination model. *Ann. Statist.* **46** 1932–1960. MR3845006 <https://doi.org/10.1214/17-AOS1607>
- CHERNOZHUKOV, V., GALICHON, A., HALLIN, M. and HENRY, M. (2017). Monge–Kantorovich depth, quantiles, ranks and signs. *Ann. Statist.* **45** 223–256. MR3611491 <https://doi.org/10.1214/16-AOS1450>
- CLAESKENS, G., HUBERT, M., SLAETS, L. and VAKILI, K. (2014). Multivariate functional halfspace depth. *J. Amer. Statist. Assoc.* **109** 411–423. MR3180573 <https://doi.org/10.1080/01621459.2013.856795>
- CUI, X., LIN, L. and YANG, G. (2008). An extended projection data depth and its applications to discrimination. *Comm. Statist. Theory Methods* **37** 2276–2290. MR2526679 <https://doi.org/10.1080/03610920701858396>
- DANG, X. and SERFLING, R. (2010). Nonparametric depth-based multivariate outlier identifiers, and masking robustness properties. *J. Statist. Plann. Inference* **140** 198–213. MR2568133 <https://doi.org/10.1016/j.jspi.2009.07.004>
- DONOHO, D. L. (1982). Breakdown properties of multivariate location estimators. Ph.D. qualifying paper, Harvard Univ.
- DONOHO, D. L. and GASKO, M. (1992). Breakdown properties of location estimates based on halfspace depth and projected outlyingness. *Ann. Statist.* **20** 1803–1827. MR1193313 <https://doi.org/10.1214/aos/1176348890>
- DYCKERHOFF, R. (2004). Data depths satisfying the projection property. *Allg. Stat. Arch.* **88** 163–190. MR2074729 <https://doi.org/10.1007/s101820400167>
- EISENHAEUER, J. G. (2003). Regression through the origin. *Teach. Stat.* **25** 76–80.
- FEBRERO, M., GALEANO, P. and GONZÁLEZ-MANTEIGA, W. (2008). Outlier detection in functional data by depth measures, with application to identify abnormal NO_x levels. *Environmetrics* **19** 331–345. MR2440036 <https://doi.org/10.1002/env.878>
- GHOSH, A. K. and CHAUDHURI, P. (2005). On data depth and distribution-free discriminant analysis using separating surfaces. *Bernoulli* **11** 1–27. MR2121452 <https://doi.org/10.3150/bj/1110228239>
- GIJBELS, I. and NAGY, S. (2017). On a general definition of depth for functional data. *Statist. Sci.* **32** 630–639. MR3730526 <https://doi.org/10.1214/17-STS625>
- HALLIN, M., PAINDAVEINE, D. and ŠIMAN, M. (2010). Multivariate quantiles and multiple-output regression quantiles: From L_1 optimization to halfspace depth. *Ann. Statist.* **38** 635–669. MR2604670 <https://doi.org/10.1214/09-AOS723>
- HAMPEL, F. R. (1974). The influence curve and its role in robust estimation. *J. Amer. Statist. Assoc.* **69** 383–393. MR0362657
- HOBERG, R. (2000). Cluster analysis based on data depth. In *Data Analysis, Classification and Related Methods* (H. Kiers, J. P. Rasson, P. Groenen and M. Schader, eds.) 17–22. Springer, Berlin.
- HUBER, P. J. (1964). Robust estimation of a location parameter. *Ann. Math. Stat.* **35** 73–101. MR0161415 <https://doi.org/10.1214/aoms/1177703732>
- HUBER, P. J. (1973). Robust regression: Asymptotics, conjectures and Monte Carlo. *Ann. Statist.* **1** 799–821. MR0356373
- HUBERT, M., ROUSSEEUW, P. J. and SEGHAERT, P. (2015). Multivariate functional outlier detection. *Stat. Methods Appl.* **24** 177–202. MR3376852 <https://doi.org/10.1007/s10260-015-0297-8>
- HUBERT, M., ROUSSEEUW, P. J. and VAN AELST, S. (2001). Similarities between location depth and regression depth. In *Statistics in Genetics and in the Environmental Sciences* (Ascona, 1999) (L. Fernholz, S. Morgenthaler and W. Stahel, eds.). Trends Math. 159–172. Birkhäuser, Basel. MR1843176
- JAECKEL, L. A. (1972). Estimating regression coefficients by minimizing the dispersion of the residuals. *Ann. Math. Stat.* **43** 1449–1458. MR0348930 <https://doi.org/10.1214/aoms/1177692377>
- JÖRNSTEN, R. (2004). Clustering and classification based on the L_1 data depth. *J. Multivariate Anal.* **90** 67–89. MR2064937 <https://doi.org/10.1016/j.jmva.2004.02.013>
- JUREČKOVÁ, J. (1971). Nonparametric estimate of regression coefficients. *Ann. Math. Stat.* **42** 1328–1338. MR0295487 <https://doi.org/10.1214/aoms/1177693245>
- KOENKER, R. and BASSETT, G. JR. (1978). Regression quantiles. *Econometrica* **46** 33–50. MR0474644 <https://doi.org/10.2307/1913643>
- KOSHEVOY, G. and MOSLER, K. (1997). Zonoid trimming for multivariate distributions. *Ann. Statist.* **25** 1998–2017. MR1474078 <https://doi.org/10.1214/aos/1069362382>
- KOUL, H. L. (1970). Some convergence theorems for ranks and weighted empirical cumulatives. *Ann. Math. Stat.* **41** 1768–1773. MR0267631 <https://doi.org/10.1214/aoms/1177696824>
- KOUL, H. L. (1971). Asymptotic behavior of a class of confidence regions based on ranks in regression. *Ann. Math. Stat.* **42** 466–476. MR0288896 <https://doi.org/10.1214/aoms/1177693398>
- LANGE, T., MOSLER, K. and MOZHAROVSKIY, P. (2014). Fast nonparametric classification based on data depth. *Statist. Papers* **55** 49–69. MR3152767 <https://doi.org/10.1007/s00362-012-0488-4>
- LI, J., CUESTA-ALBERTOS, J. A. and LIU, R. Y. (2012). DD-classifier: Nonparametric classification procedure based on DD-plot. *J. Amer. Statist. Assoc.* **107** 737–753. MR2980081 <https://doi.org/10.1080/01621459.2012.688462>
- LI, J. and LIU, R. Y. (2004). New nonparametric tests of multivariate locations and scales using data depth. *Statist. Sci.* **19** 686–696. MR2185590 <https://doi.org/10.1214/088342304000000594>
- LIU, R. Y. (1990). On a notion of data depth based on random simplices. *Ann. Statist.* **18** 405–414. MR1041400 <https://doi.org/10.1214/aos/1176347507>
- LIU, R. Y. (1992). Data depth and multivariate rank tests. In L_1 -Statistical Analysis and Related Methods (Neuchâtel, 1992) (Y. Dodge, ed.) 279–294. North-Holland, Amsterdam. MR1214839
- LIU, R. Y. (1995). Control charts for multivariate processes. *J. Amer. Statist. Assoc.* **90** 1380–1387. MR1379481
- LIU, R. Y., PARELIUS, J. M. and SINGH, K. (1999). Multivariate analysis by data depth: Descriptive statistics, graphics and inference. *Ann. Statist.* **27** 783–858. MR1724033 <https://doi.org/10.1214/aos/1018031260>
- LIU, R. Y. and SINGH, K. (1993). A quality index based on data depth and multivariate rank tests. *J. Amer. Statist. Assoc.* **88** 252–260. MR1212489
- LÓPEZ-PINTADO, S. and ROMO, J. (2009). On the concept of depth for functional data. *J. Amer. Statist. Assoc.* **104** 718–734. MR2541590 <https://doi.org/10.1198/jasa.2009.0108>
- MAJUMDAR, S. and CHATTERJEE, S. (2018). Non-convex penalized multitask regression using data depth-based penalties. *Stat* **7** e174. MR3769054 <https://doi.org/10.1002/sta4.174>
- MARONNA, R. A., MARTIN, R. D. and YOHAI, V. J. (2006). *Robust Statistics: Theory and Methods*. Wiley Series in Probability and Statistics. Wiley, Chichester. MR2238141 <https://doi.org/10.1002/0470010940>
- MARONNA, R. A. and YOHAI, V. J. (1993). Bias-robust estimates of regression based on projections. *Ann. Statist.* **21** 965–990. MR1232528 <https://doi.org/10.1214/aos/1176349160>

- MIZERA, I. (2002). On depth and deep points: A calculus. *Ann. Statist.* **30** 1681–1736. MR1969447 <https://doi.org/10.1214/aos/1043351254>
- MIZERA, I. and MÜLLER, C. H. (2004). Location-scale depth. *J. Amer. Statist. Assoc.* **99** 949–989. MR2109488 <https://doi.org/10.1198/016214504000001312>
- MOSLER, K. (2002). *Multivariate Dispersion, Central Regions and Depth: The Lift Zonoid Approach. Lecture Notes in Statistics* **165**. Springer, Berlin. MR1913862 <https://doi.org/10.1007/978-1-4613-0045-8>
- MOSLER, K. and BAZOVKIN, P. (2014). Stochastic linear programming with a distortion risk constraint. *OR Spectrum* **36** 949–969. MR3258501 <https://doi.org/10.1007/s00291-014-0372-9>
- MOSLER, K. and HOBBERG, R. (2006). Data analysis and classification with the zonoid depth. In *Data Depth: Robust Multivariate Analysis, Computational Geometry and Applications* (R. Liu, R. Serfling and D. Souvaine, eds.). DIMACS Ser. Discrete Math. Theoret. Comput. Sci. **72** 49–59. Amer. Math. Soc., Providence, RI. MR2343112
- NIETO-REYES, A. and BATTEY, H. (2016). A topologically valid definition of depth for functional data. *Statist. Sci.* **31** 61–79. MR3458593 <https://doi.org/10.1214/15-STS532>
- PAINDAVEINE, D. and VAN BEVER, G. (2013). From depth to local depth: A focus on centrality. *J. Amer. Statist. Assoc.* **108** 1105–1119. MR3174687 <https://doi.org/10.1080/01621459.2013.813390>
- PAINDAVEINE, D. and VAN BEVER, G. (2015). Nonparametrically consistent depth-based classifiers. *Bernoulli* **21** 62–82. MR3322313 <https://doi.org/10.3150/13-BEJ561>
- PAINDAVEINE, D. and VAN BEVER, G. (2018). Halfspace depths for scatter, concentration and shape matrices. *Ann. Statist.* **46** 3276–3307. MR3852652 <https://doi.org/10.1214/17-AOS1658>
- POLLARD, D. (1984). *Convergence of Stochastic Processes. Springer Series in Statistics*. Springer, New York. MR0762984 <https://doi.org/10.1007/978-1-4612-5254-2>
- PORTNOY, S. (2003). Censored regression quantiles. *J. Amer. Statist. Assoc.* **98** 1001–1012. MR2041488 <https://doi.org/10.1198/016214503000000954>
- PORTNOY, S. (2012). Nearly root- n approximation for regression quantile processes. *Ann. Statist.* **40** 1714–1736. MR3015041 <https://doi.org/10.1214/12-AOS1021>
- ROUSSEEUW, P. J. (1984). Least median of squares regression. *J. Amer. Statist. Assoc.* **79** 871–880. MR0770281
- ROUSSEEUW, P. J. and HUBERT, M. (1999). Regression depth. *J. Amer. Statist. Assoc.* **94** 388–433. MR1702314 <https://doi.org/10.2307/2670155>
- ROUSSEEUW, P. J. and LEROY, A. M. (1987). *Robust Regression and Outlier Detection. Wiley Series in Probability and Mathematical Statistics: Applied Probability and Statistics*. Wiley, New York. MR0914792 <https://doi.org/10.1002/0471725382>
- ROUSSEEUW, P. J. and STRUYF, A. (2004). Characterizing angular symmetry and regression symmetry. *J. Statist. Plann. Inference* **122** 161–173. MR2057920 <https://doi.org/10.1016/j.jspi.2003.06.015>
- RUPPERT, D. and CARROLL, R. J. (1980). Trimmed least squares estimation in the linear model. *J. Amer. Statist. Assoc.* **75** 828–838. MR0600964
- SERFLING, R. (2006). Depth functions in nonparametric multivariate inference. In *Data Depth: Robust Multivariate Analysis, Computational Geometry and Applications. DIMACS Ser. Discrete Math. Theoret. Comput. Sci.* **72** 1–16. Amer. Math. Soc., Providence, RI. MR2343109
- SERFLING, R. (2019). Perspectives on depth functions on general data spaces, with consideration of the Tukey, projection, spatial, ‘density’, ‘local’, and ‘contour’ depths. Preprint.
- SERFLING, R. and WANG, S. (2014). General foundations for studying masking and swamping robustness of outlier identifiers. *Stat. Methodol.* **20** 79–90. MR3205723 <https://doi.org/10.1016/j.stamet.2013.08.004>
- STAHEL, W. A. (1981). Robuste Schätzungen: Infinitesimale Optimalität und Schätzungen Von Kovarianzmatrizen. Ph.D. dissertation, ETH, Zurich.
- TUKEY, J. W. (1975). Mathematics and the picturing of data. In *Proceedings of the International Congress of Mathematicians (Vancouver, B.C., 1974)*, Vol. 2 (R. D. James, ed.) 523–531. MR0426989
- VAN AELST, S. and ROUSSEEUW, P. J. (2000). Robustness of deepest regression. *J. Multivariate Anal.* **73** 82–106. MR1766122 <https://doi.org/10.1006/jmva.1999.1870>
- VARDI, Y. and ZHANG, C.-H. (2000). The multivariate L_1 -median and associated data depth. *Proc. Natl. Acad. Sci. USA* **97** 1423–1426. MR1740461 <https://doi.org/10.1073/pnas.97.4.1423>
- VELASCO-FORERO, S. and ANGULO, J. (2011). Mathematical morphology for vector images using statistical depth. In *Mathematical Morphology and Its Applications to Image and Signal Processing* (P. Soille, M. Pesaresi and G. K. Ouzounis, eds.) 355–366. Springer, Berlin.
- VELASCO-FORERO, S. and ANGULO, J. (2012). Random projection depth for multivariate mathematical morphology. *IEEE J. Sel. Top. Signal Process.* **6** 753–763.
- WANG, S. and SERFLING, R. (2015). On masking and swamping robustness of leading nonparametric outlier identifiers for univariate data. *J. Statist. Plann. Inference* **162** 62–74. MR3323104 <https://doi.org/10.1016/j.jspi.2015.02.002>
- WANG, S. and SERFLING, R. (2018). On masking and swamping robustness of leading nonparametric outlier identifiers for multivariate data. *J. Multivariate Anal.* **166** 32–49. MR3799633 <https://doi.org/10.1016/j.jmva.2018.02.003>
- WU, M. and ZUO, Y. (2008). Trimmed and Winsorized standard deviations based on a scaled deviation. *J. Nonparametr. Stat.* **20** 319–335. MR2436381 <https://doi.org/10.1080/10485250802036909>
- WU, M. and ZUO, Y. (2009). Trimmed and Winsorized means based on a scaled deviation. *J. Statist. Plann. Inference* **139** 350–365. MR2474011 <https://doi.org/10.1016/j.jspi.2008.03.039>
- YEH, A. B. and SINGH, K. (1997). Balanced confidence regions based on Tukey’s depth and the bootstrap. *J. Roy. Statist. Soc. Ser. B* **59** 639–652. MR1452031 <https://doi.org/10.1111/1467-9868.00088>
- ZUO, Y. (1998). *Contributions to the Theory and Applications of Statistical Depth Functions*. ProQuest LLC, Ann Arbor, MI. Ph.D. thesis, Univ. Texas at Dallas. MR2697586
- ZUO, Y. (2003). Projection-based depth functions and associated medians. *Ann. Statist.* **31** 1460–1490. MR2012822 <https://doi.org/10.1214/aos/1065705115>
- ZUO, Y. (2004). Robustness of weighted L^p -depth and L^p -median. *Allg. Stat. Arch.* **88** 215–234. MR2074731 <https://doi.org/10.1007/s101820400169>
- ZUO, Y. (2009). Data depth trimming procedure outperforms the classical t (or T^2) one. *J. Probab. Stat.* <https://doi.org/10.1155/2009/373572>
- ZUO, Y. (2010). Is the t confidence interval $\bar{X} \pm t_\alpha(n-1)s/\sqrt{n}$ optimal? *Amer. Statist.* **64** 170–173. MR2757009 <https://doi.org/10.1198/tast.2010.09021>
- ZUO, Y. (2018). Robustness of deepest projection regression depth functional. *Statist. Papers.* <https://doi.org/10.1007/s00362-019-01129-4>
- ZUO, Y. (2019a). Asymptotics for the maximum regression depth estimator. Available at [arXiv:1809.09896](https://arxiv.org/abs/1809.09896).
- ZUO, Y. (2019b). Computation of projection regression depth and its induced median. Available at [arXiv:1905.11846](https://arxiv.org/abs/1905.11846).
- ZUO, Y. and SERFLING, R. (2000). General notions of statistical depth function. *Ann. Statist.* **28** 461–482. MR1790005 <https://doi.org/10.1214/aos/1016218226>

A Conversation with Tze Leung Lai

Ying Lu, Dylan S. Small and Zhiliang Ying

Abstract. This conversation began in June 2015 in the Department of Statistics at Columbia University during Lai’s visit to his alma mater where he celebrated his seventieth birthday. It continued in the subsequent years at Columbia and Stanford. Lai was born on June 28, 1945, in Hong Kong, where he grew up and attended The University of Hong Kong, receiving his B.A. degree (First Class Honors) in Mathematics in 1967. He went to Columbia University in 1968 for graduate study in statistics and received his Ph.D. degree in 1971. He stayed on the faculty at Columbia and was appointed Higgins Professor of Mathematical Statistics in 1986. A year later he moved to Stanford, where he is currently Ray Lyman Wilbur Professor of Statistics, and by courtesy, also of Biomedical Data Science and Computational and Mathematical Engineering. He is a fellow of the Institute of Mathematical Statistics, the American Statistical Association and an elected member of Academia Sinica in Taiwan. He was the third recipient of the COPSS Award which he won in 1983. He has been married to Letitia Chow since 1975, and they have two sons and two grandchildren.

Key words and phrases: Hong Kong, Columbia University, Stanford University, statistics, biostatistics, sequential experimentation, mathematical finance.

REFERENCES

- BHAT (1982). A randomized trial of propranolol in patients with acute myocardial infarction: I. mortality results. *J. Am. Med. Assoc.* **247** 1707–1714.
- CHEN, G. J., LAI, T. L. and WEI, C. Z. (1981). Convergence systems and strong consistency of least squares estimates in regression models. *J. Multivariate Anal.* **11** 319–333. [MR0629791](https://doi.org/10.1016/0047-259X(81)90078-6)
- FELLER, W. (1971). *An Introduction to Probability Theory and Its Applications. Vol. II.* Wiley, New York. [MR0270403](https://doi.org/10.1016/0047-259X(81)90078-6)
- KOUL, H., SUSARLA, V. and VAN RYZIN, J. (1981). Regression analysis with randomly right-censored data. *Ann. Statist.* **9** 1276–1288. [MR0630110](https://doi.org/10.1016/0047-259X(81)90078-6)
- LAI, T. L. (2009). Martingales in sequential analysis and time series, 1945–1985. *J. Électron. Hist. Probab. Stat.* **5** 1–31. [MR2520670](https://doi.org/10.1016/0047-259X(81)90078-6)
- LAI, T. L. and ROBBINS, H. (1979). Adaptive design and stochastic approximation. *Ann. Statist.* **7** 1196–1221. [MR0550144](https://doi.org/10.1016/0047-259X(81)90078-6)
- LAI, T. L. and ROBBINS, H. (1985). Asymptotically efficient adaptive allocation rules. *Adv. in Appl. Math.* **6** 4–22. [MR0776826](https://doi.org/10.1016/0196-8858(85)90002-8)
- LAI, T. L., ROBBINS, H. and WEI, C. Z. (1978). Strong consistency of least squares estimates in multiple regression. *Proc. Natl. Acad. Sci. USA* **75** 3034–3036. [MR0518953](https://doi.org/10.1073/pnas.75.7.3034)
- LAI, T. L., ROBBINS, H. and WEI, C. Z. (1979). Strong consistency of least squares estimates in multiple regression. II. *J. Multivariate Anal.* **9** 343–361. [MR0548786](https://doi.org/10.1016/0047-259X(79)90093-9)
- LAI, T. L. and SIEGMUND, D. (1977). A nonlinear renewal theory with applications to sequential analysis. I. *Ann. Statist.* **5** 946–954. [MR0445599](https://doi.org/10.1016/0047-259X(79)90093-9)
- LAI, T. L. and SIEGMUND, D. (1979). A nonlinear renewal theory with applications to sequential analysis. II. *Ann. Statist.* **7** 60–76. [MR0515684](https://doi.org/10.1016/0047-259X(79)90093-9)
- LAI, T. L. and WEI, C. Z. (1982a). Asymptotic properties of projections with applications to stochastic regression problems. *J. Multivariate Anal.* **12** 346–370. [MR0666011](https://doi.org/10.1016/0047-259X(82)90071-9)
- LAI, T. L. and WEI, C. Z. (1982b). A law of the iterated logarithm for double arrays of independent random variables with applications to regression and time series models. *Ann. Probab.* **10** 320–335. [MR0647507](https://doi.org/10.1016/0047-259X(82)90071-9)
- LAI, T. L. and WEI, C. Z. (1982c). Least squares estimates in stochastic regression models with applications to identification and control of dynamic systems. *Ann. Statist.* **10** 154–166. [MR0642726](https://doi.org/10.1016/0047-259X(82)90071-9)
- LAI, T. L. and WEI, C. Z. (1983a). Asymptotic properties of general autoregressive models and strong consistency of least-squares estimates of their parameters. *J. Multivariate Anal.* **13** 1–23. [MR0695924](https://doi.org/10.1016/0047-259X(83)90002-7)

Ying Lu is Professor of Biomedical Data Science, and by courtesy, of Radiology and Health Research and Policy, Stanford University, Stanford, California 94305, USA (e-mail: yylu1@stanford.edu). Dylan Small is the Class of 1965 Wharton Professor of Statistics and Co-director of the Center for Causal Inference, University of Pennsylvania, Philadelphia, Pennsylvania 19104, USA (e-mail: dsmall@wharton.upenn.edu). Zhiliang Ying is Professor, Department of Statistics, Columbia University, New York, New York 10027, USA (e-mail: zying@stat.columbia.edu).

- LAI, T. L. and WEI, C. Z. (1983b). Lacunary systems and generalized linear processes. *Stochastic Process. Appl.* **14** 187–199. [MR0679672](#) [https://doi.org/10.1016/0304-4149\(83\)90071-6](https://doi.org/10.1016/0304-4149(83)90071-6)
- LAI, T. L. and WEI, C. Z. (1985). Orthonormal Banach systems with applications to linear processes. *Z. Wahrsch. Verw. Gebiete* **70** 381–393. [MR0803679](#) <https://doi.org/10.1007/BF00534870>
- LAI, T. L. and WEI, C. Z. (1986a). On the concept of excitation in least squares identification and adaptive control. *Stochastics* **16** 227–254. [MR0837613](#) <https://doi.org/10.1080/17442508608833375>
- LAI, T. L. and WEI, C. Z. (1986b). Extended least squares and their applications to adaptive control and prediction in linear systems. *IEEE Trans. Automat. Control* **31** 898–906. [MR0855543](#) <https://doi.org/10.1109/TAC.1986.1104138>
- LAI, T. L. and YING, Z. (1991a). Recursive identification and adaptive prediction in linear stochastic systems. *SIAM J. Control Optim.* **29** 1061–1090. [MR1110087](#) <https://doi.org/10.1137/0329058>
- LAI, T. L. and YING, Z. (1991b). Parallel recursive algorithms in asymptotically efficient adaptive control of linear stochastic systems. *SIAM J. Control Optim.* **29** 1091–1127. [MR1110088](#) <https://doi.org/10.1137/0329059>
- LAI, T. L., YING, Z. and ZHENG, Z. K. (1995). Asymptotic normality of a class of adaptive statistics with applications to synthetic data methods for censored regression. *J. Multivariate Anal.* **52** 259–279. [MR1323333](#) <https://doi.org/10.1006/jmva.1995.1013>
- LEONG, Y. K. (2019). From accidental statistician to interdisciplinary statistician who combines theory with practice. Imprints, Institute of Mathematical Sciences, National University of Singapore.
- SHEN, M., TSANG, K. W. and WONG, S. P. S. (2016). Conversations with Tze Leung Lai. *ICSA Bull.* **28** 29–30.
- TSIATIS, A. A. (1981). The asymptotic joint distribution of the efficient scores test for the proportional hazards model calculated over time. *Biometrika* **68** 311–315. [MR0614968](#) <https://doi.org/10.1093/biomet/68.1.311>

INSTITUTE OF MATHEMATICAL STATISTICS

(Organized September 12, 1935)

The purpose of the Institute is to foster the development and dissemination of the theory and applications of statistics and probability.

IMS OFFICERS

President: Regina Y. Liu, Department of Statistics, Rutgers University, Piscataway, New Jersey 08854-8019, USA

President-Elect: Krzysztof Burdzy, Department of Mathematics, University of Washington, Seattle, Washington 98195-4350, USA

Past President: Susan Murphy, Department of Statistics, Harvard University, Cambridge, Massachusetts 02138-2901, USA

Executive Secretary: Edsel Peña, Department of Statistics, University of South Carolina, Columbia, South Carolina 29208-001, USA

Treasurer: Zhengjun Zhang, Department of Statistics, University of Wisconsin, Madison, Wisconsin 53706-1510, USA

Program Secretary: Ming Yuan, Department of Statistics, Columbia University, New York, NY 10027-5927, USA

IMS EDITORS

The Annals of Statistics. *Editors:* Richard J. Samworth, Statistical Laboratory, Centre for Mathematical Sciences, University of Cambridge, Cambridge, CB3 0WB, UK. Ming Yuan, Department of Statistics, Columbia University, New York, NY 10027, USA

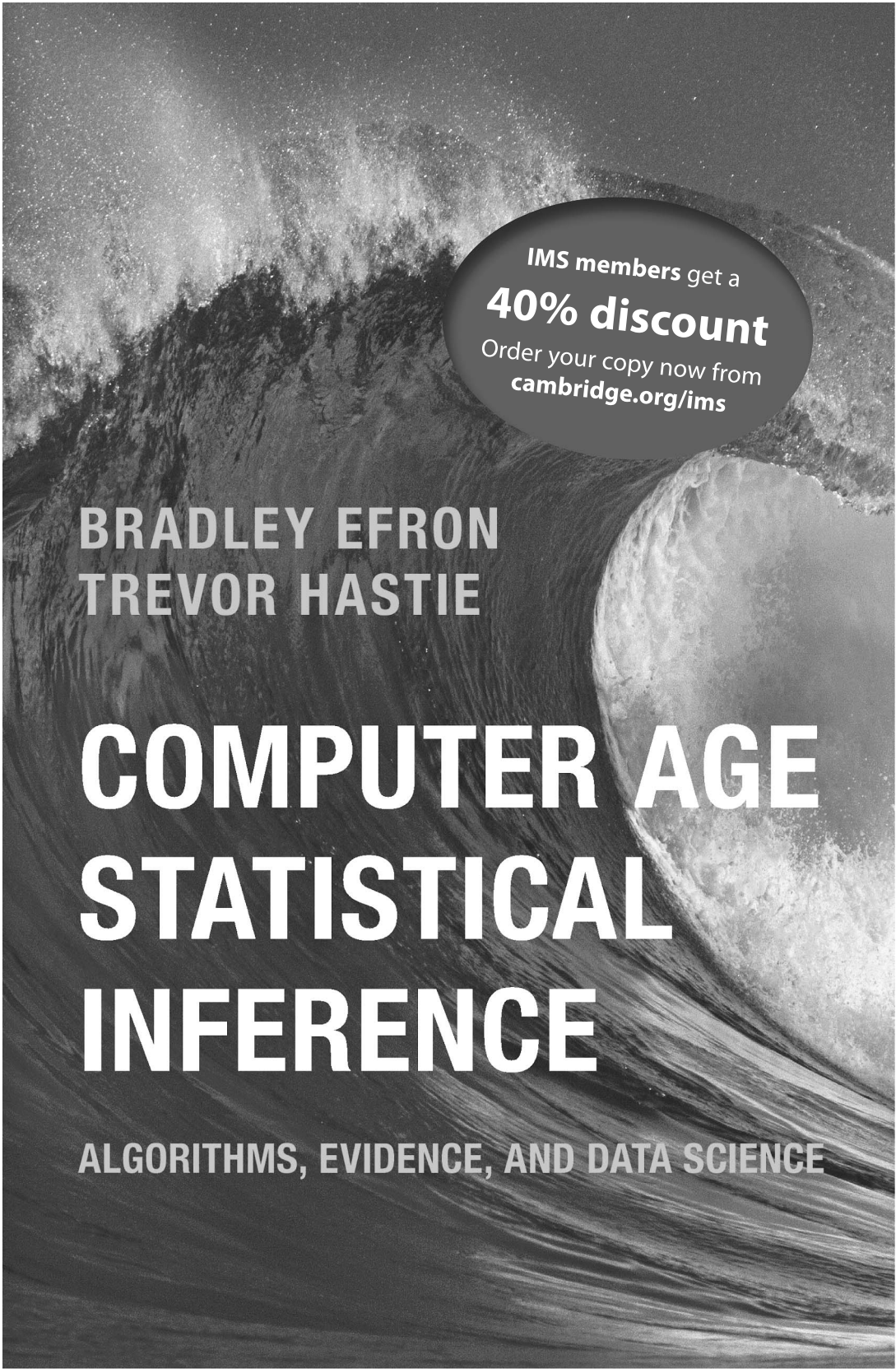
The Annals of Applied Statistics. *Editor-in-Chief:* Karen Kafadar, Department of Statistics, University of Virginia, Heidelberg Institute for Theoretical Studies, Charlottesville, VA 22904-4135, USA

The Annals of Probability. *Editor:* Amir Dembo, Department of Statistics and Department of Mathematics, Stanford University, Stanford, California 94305, USA

The Annals of Applied Probability. *Editors:* François Delarue, Laboratoire J. A. Dieudonné, Université de Nice Sophia-Antipolis, France-06108 Nice Cedex 2. Peter Friz, Institut für Mathematik, Technische Universität Berlin, 10623 Berlin, Germany and Weierstrass-Institut für Angewandte Analysis und Stochastik, 10117 Berlin, Germany

Statistical Science. *Editor:* Sonia Petrone, Department of Decision Sciences, Università Bocconi, 20100 Milano MI, Italy

The IMS Bulletin. *Editor:* Vlada Limic, UMR 7501 de l'Université de Strasbourg et du CNRS, 7 rue René Descartes, 67084 Strasbourg Cedex, France



IMS members get a
40% discount
Order your copy now from
cambridge.org/ims

BRADLEY EFRON
TREVOR HASTIE

COMPUTER AGE STATISTICAL INFERENCE

ALGORITHMS, EVIDENCE, AND DATA SCIENCE