

Agterberg, Joshua; Park, Youngser; Larson, Jonathan; White, Christopher; Priebe, Carey E.; Lyzinski, Vince

Vertex nomination, consistent estimation, and adversarial modification. (English)

Zbl 1448.62087

Electron. J. Stat. 14, No. 2, 3230-3267 (2020).

Summary: Given a pair of graphs G_1 and G_2 and a vertex set of interest in G_1 , the vertex nomination (VN) problem seeks to find the corresponding vertices of interest in G_2 (if they exist) and produce a rank list of the vertices in G_2 , with the corresponding vertices of interest in G_2 concentrating, ideally, at the top of the rank list. In this paper, we define and derive the analogue of Bayes optimality for VN with multiple vertices of interest, and we define the notion of maximal consistency classes in vertex nomination. This theory forms the foundation for a novel VN adversarial contamination model, and we demonstrate with real and simulated data that there are VN schemes that perform effectively in the uncontaminated setting, and adversarial network contamination adversely impacts the performance of our VN scheme. We further define a network regularization method for mitigating the impact of the adversarial contamination, and we demonstrate the effectiveness of regularization in both real and synthetic data.

MSC:

62H22 Probabilistic graphical models

62H12 Estimation in multivariate analysis

68T05 Learning and adaptive systems in artificial intelligence

05C90 Applications of graph theory

Keywords:

statistics; random graphs; networks; adversarial machine learning; vertex nomination

Software:

mclust

Full Text: [DOI](#) [Euclid](#)

References:

- [1] J. Arroyo-Reli3n, D. Kessler, E. Levina, and S. Taylor. Network classification with applications to brain connectomics., *The Annals of Applied Statistics*, 13(3) :1648-1677, 09 2019. · [Zbl 1433.62314](#) · [doi:10.1214/19-AOAS1252](#)
- [2] A. Athreya, D. Fishkind, K. Levin, V. Lyzinski, Y. Park, Y. Qin, D. Sussman, M. Tang, J. Vogelstein, and C. Priebe. Statistical inference on random dot product graphs: A survey., *Journal of Machine Learning Research*, 18, 09 2017. · [Zbl 06982982](#)
- [3] A. Athreya, C. E. Priebe, M. Tang, V. Lyzinski, D.J. Marchette, and D.L. Sussman. A limit theorem for scaled eigenvectors of random dot product graphs., *Sankhya A*, 1-18, 2015. · [Zbl 1338.62061](#) · [doi:10.1007/s13171-015-0071-x](#)
- [4] P. J. Bickel and A. Chen. A nonparametric view of network models and Newman-Girvan and other modularities., *Proceedings of the National Academy of Sciences, USA*, 106 :21068-21073, 2009. · [Zbl 1359.62411](#) · [doi:10.1073/pnas.0907096106](#)
- [5] P. J. Bickel, D. Choi, X. Chang, and H. Zhang. Asymptotic normality of maximum likelihood and its variational approximation for stochastic blockmodels., *The Annals of Statistics*, 41(4) :1922-1943, 2013. · [Zbl 1292.62042](#) · [doi:10.1214/13-AOS1124](#)
- [6] V. D. Blondel, J.-L. Guillaume, R. Lambiotte, and E. Lefebvre. Fast unfolding of communities in large networks., *Journal of Statistical Mechanics: Theory and Experiment*, 2008. · [Zbl 1459.91130](#) · [doi:10.1088/1742-5468/2008/10/P10008](#)
- [7] E. Bullmore and O. Sporns. Complex brain networks: graph theoretical analysis of structural and functional systems., *Nature Reviews Neuroscience*, 10(3):186, 2009.
- [8] T. T. Cai and X. Li. Robust and computationally feasible community detection in the presence of arbitrary outlier nodes., *The Annals of Statistics*, 43(3) :1027-1059, 2015. · [Zbl 1328.62381](#) · [doi:10.1214/14-AOS1290](#)
- [9] J. Cape, M. Tang, and C. E. Priebe. The two-to-infinity norm and singular subspace geometry with applications to high-dimensional statistics., *The Annals of Statistics*, 47(5) :2405-2439, 2019. · [Zbl 07114917](#) · [doi:10.1214/18-AOS1752](#)
- [10] S. Chatterjee. Matrix estimation by universal singular value thresholding., *The Annals of Statistics*, 43(1):177-214, 2014. · [Zbl 1308.62038](#) · [doi:10.1214/14-AOS1272](#)
- [11] L. Chen, C. Shen, J. T. Vogelstein, and C. E. Priebe. Robust vertex classification., *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 38(3):578-590, 2016.

- [12] D. Conte, P. Foggia, C. Sansone, and M. Vento. Thirty years of graph matching in pattern recognition., *International Journal of Pattern Recognition and Artificial Intelligence*, 18(03):265-298, 2004.
- [13] G. Coppersmith. Vertex nomination., *Wiley Interdisciplinary Reviews: Computational Statistics*, 6(2):144-153, 2014.
- [14] G. A. Coppersmith and C. E. Priebe. Vertex nomination via content and context., *arXiv preprint arXiv:1201.4118*, 2012.
- [15] H. Dai, H. Li, T. Tian, X. Huang, L. Wang, J. Zhu, and L. Song. Adversarial attack on graph structured data. In Jennifer Dy and Andreas Krause, editors, *Proceedings of the 35th International Conference on Machine Learning*, volume 80 of *Proceedings of Machine Learning Research*, pages 1115-1124, Stockholmsmässan, Stockholm Sweden, 10-15 Jul 2018. PMLR.
- [16] D. Edge, J. Larson, M. Mobius, and C. White. Trimming the hairball: Edge cutting strategies for making dense graphs usable. In, *2018 IEEE International Conference on Big Data (Big Data)*, pages 3951-3958. IEEE, 2018.
- [17] D. E. Fishkind, V. Lyzinski, H. Pao, L. Chen, and C. E. Priebe. Vertex nomination schemes for membership prediction., *The Annals of Applied Statistics*, 9(3) :1510-1532, 2015. · [Zbl 1454.62180](#) · [doi:10.1214/15-AOAS834](#)
- [18] P. Foggia, G. Percannella, and M. Vento. Graph matching and learning in pattern recognition in the last 10 years., *International Journal of Pattern Recognition and Artificial Intelligence*, 28(01) :1450001, 2014.
- [19] C. Fraley and A. E. Raftery. Mclust: Software for model-based cluster analysis., *Journal of Classification*, 16(2):297-306, 1999. · [Zbl 0951.91500](#) · [doi:10.1007/s003579900058](#)
- [20] K. J. Gile and M. S. Handcock. 7. Respondent-driven sampling: An assessment of current methodology., *Sociological Methodology*, 40(1):285-327, 2010.
- [21] M. Girvan and M. E. J. Newman. Community structure in social and biological networks., *Proceedings of the National Academy of Sciences*, 99(12) :7821-7826, 2002. · [Zbl 1032.91716](#) · [doi:10.1073/pnas.122653799](#)
- [22] D. D. Heckathorn. Respondent-driven sampling: a new approach to the study of hidden populations., *Social Problems*, 44(2):174-199, 1997.
- [23] P. W. Holland, K. B. Laskey, and S. Leinhardt. Stochastic blockmodels: First steps., *Social Networks*, 5(2):109-137, 1983.
- [24] L. Huang, A. D. Joseph, B. Nelson, B. I. P. Rubinstein, and J. D. Tygar. Adversarial machine learning. In, *Proceedings of the 4th ACM Workshop on Security and Artificial Intelligence*, pages 43-58. ACM, 2011.
- [25] C. M. Le, E. Levina, and R. Vershynin. Concentration and regularization of random graphs., *Random Structures & Algorithms*, 51(3):538-561, 2017. · [Zbl 1373.05179](#) · [doi:10.1002/rsa.20713](#)
- [26] D. S. Lee and C. E. Priebe. Bayesian vertex nomination., *arXiv preprint arXiv:1205.5082*, 2012.
- [27] V. Lyzinski, K. Levin, and C. E. Priebe. On consistent vertex nomination schemes., *Journal of Machine Learning Research*, to appear, 2019. · [Zbl 07064049](#)
- [28] V. Lyzinski, D. L. Sussman, M. Tang, A. Athreya, and C. E. Priebe. Perfect clustering for stochastic blockmodel graphs via adjacency spectral embedding., *Electronic Journal of Statistics*, 8 :2905-2922, 2014. · [Zbl 1308.62131](#) · [doi:10.1214/14-EJS978](#)
- [29] V. Lyzinski, M. Tang, A. Athreya, Y. Park, and C. E. Priebe. Community detection and classification in hierarchical stochastic blockmodels., *IEEE Transactions on Network Science and Engineering*, 4(1):13-26, 2017.
- [30] D. Marchette, C. E. Priebe, and G. Coppersmith. Vertex nomination via attributed random dot product graphs. In, *Proceedings of the 57th ISI World Statistics Congress*, volume 6, page 16, 2011.
- [31] S. Maslov and K. Sneppen. Specificity and stability in topology of protein networks., *Science*, 296 (5569):910-913, 2002.
- [32] R. Mastrandrea, J. Fournet, and A. Barrat. Contact patterns in a high school: a comparison between data collected using wearable sensors, contact diaries and friendship surveys., *PLoS One*, 10(9): e0136497, 2015.
- [33] R. Milo, S. Shen-Orr, S. Itzkovitz, N. Kashtan, D. Chklovskii, and U. Alon. Network motifs: simple building blocks of complex networks., *Science*, 298 (5594):824-827, 2002.
- [34] M. E. J. Newman. Modularity and community structure in networks., *Proceedings of the National Academy of Sciences*, 103(23) :8577-8582, 2006.
- [35] M. E. J. Newman, D. J. Watts, and S. H. Strogatz. Random graph models of social networks., *Proceedings of the National Academy of Sciences*, 99(suppl 1) :2566-2572, 2002. · [Zbl 1114.91362](#) · [doi:10.1073/pnas.012582999](#)
- [36] N. Papernot, P. McDaniel, S. Jha, M. Fredrikson, Z. B. Celik, and A. Swami. The limitations of deep learning in adversarial settings. In, *2016 IEEE European Symposium on Security and Privacy (EuroS&P)*, pages 372-387. IEEE, 2016.
- [37] H. G. Patsolic, Y. Park, V. Lyzinski, and C. E. Priebe. Vertex nomination via local neighborhood matching., *arXiv preprint arXiv:1705.00674*, 2017.
- [38] T. Qin and K. Rohe. Regularized spectral clustering under the degree-corrected stochastic blockmodel., *Advances in Neural Information Processing Systems*, 2013.
- [39] K. Rohe, S. Chatterjee, and B. Yu. Spectral clustering and the high-dimensional stochastic blockmodel., *The Annals of Statistics*, 39 :1878-1915, 2011. · [Zbl 1227.62042](#) · [doi:10.1214/11-AOS887](#)
- [40] P. H. Schönemann. A generalized solution of the orthogonal procrustes problem., *Psychometrika*, 31(1):1-10, 1966. · [Zbl 0147.19401](#)
- [41] J. Scott., *Social Network Analysis*. Sage, 2017.
- [42] O. Sporns. Graph theory methods: applications in brain networks., *Dialogues in Clinical Neuroscience*, 20(2):111, 2018.
- [43] D. L. Sussman, M. Tang, and C. E. Priebe. Consistent latent position estimation and vertex classification for random dot product graphs., *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 36(1):48-57, 2014.
- [44] M. Tang, A. Athreya, D. L. Sussman, V. Lyzinski, Y. Park, and C. E. Priebe. A semiparametric two-sample hypothesis

testing problem for random graphs., *Journal of Computational and Graphical Statistics*, 26(2):344-354, 2017. · [Zbl 1450.62040](#) · [doi:10.3150/15-BEJ789](#)

- [45] M. Tang, A. Athreya, D. L. Sussman, V. Lyzinski, and C. E. Priebe. A nonparametric two-sample hypothesis testing problem for random dot product graphs., *Bernoulli*, 23(3) :1599-1630, 2017. · [Zbl 1450.62040](#) · [doi:10.3150/15-BEJ789](#)
- [46] J. T. Vogelstein, W. G. Roncal, R. J. Vogelstein, and C. E. Priebe. Graph classification using signal-subgraphs: Applications in statistical connectomics., *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 35(7) :1539-1551, 2013.
- [47] J. Yan, X. Yin, W. Lin, C. Deng, H. Zha, and X. Yang. A short survey of recent advances in graph matching. In, *Proceedings of the 2016 ACM on International Conference on Multimedia Retrieval*, pages 167-174. ACM, 2016.
- [48] J. Yoder, L. Chen, H. Pao, E. Bridgeford, K. Levin, D. E. Fishkind, C. E. Priebe, and V. Lyzinski. Vertex nomination: The canonical sampling and the extended spectral nomination schemes., *arXiv preprint arXiv:1802.04960* , 2018. · [Zbl 07178860](#) · [doi:10.1016/j.csda.2020.106916](#)
- [49] M. Zhu and A. Ghodsi. Automatic dimensionality selection from the scree plot via the use of profile likelihood., *Computational Statistics & Data Analysis*, 51(2):918-930, 2006. · [Zbl 1157.62429](#) · [doi:10.1016/j.csda.2005.09.010](#)
- [50] D.

This reference list is based on information provided by the publisher or from digital mathematics libraries. Its items are heuristically matched to zbMATH identifiers and may contain data conversion errors. It attempts to reflect the references listed in the original paper as accurately as possible without claiming the completeness or perfect precision of the matching.