

An Efficient Algorithm for Exact Recovery of Vertex Variables from Edge Measurements

Afonso S. Bandeira, MIT

This installment of Computer's series highlighting the work published in IEEE Computer Society journals comes from IEEE Transactions on Network Science and Engineering.

What do community detection in complex networks, structure from motion in computer vision, and multiple-image registration have in common? They can each be formulated as an inverse problem on a graph. More specifically, we can associate each data unit—such as a node label, photo, or rotated image—to a graph node, and each pairwise measurement to an edge connecting the two nodes. In doing so, the problem becomes one of estimating for each node an unknown variable, such as community membership or rotation, from relative information between pairs of such variables. It's useful to think of node variables as taking values in a group G and relative measurements as revealing information about group ratios $g_i (g_j)^{-1} \in G$.

Although much literature proposes algorithmic approaches to this problem, we still don't understand it from an information-theoretical or computational average-case-complexity viewpoint. In "Decoding Binary Node Labels from Censored Edge Measurements: Phase Transition

and Efficient Recovery" (*IEEE Trans. Network Science and Eng.*, vol. 1, no. 1, 2014, pp. 10–22), my colleagues and I moved toward a better understanding by treating a simple, yet crucial, instance of this issue—the setting in which node labels take only two different values (that is, G is the group of two elements).¹

In this setting, it's best to view the problem as community detection on an observation graph $G = (V, E)$. The unknown vertex labels x^V represent community memberships and the edge measurements y^E are noisy indications of whether node pairs belong to the same cluster. More precisely, given a noise level $\varepsilon < 1/2$, for each edge (i, j) , $y^E_{i,j}$ indicates whether i and j belong to the same community or not, making an error with probability ε for each edge, and independently to errors in other edges. The goal is then to determine for which graphs G and values of ε we can recover x^V given y^E , and whether this inverse problem can be solved efficiently. Note that x^V can only be recovered up to a global flip, corresponding to a cluster relabeling. Indeed, because only relative

information is available, the measurements are invariant to relabeling the clusters.

This model is closely related to the popular stochastic block model for two communities. In that model, a random graph is drawn on vertices belonging to two or more communities from a distribution with independent edges and probability p if between two vertices of the same community and q if otherwise. The objective is to recover the community memberships. The main difference between the models is that in the stochastic block model, every node pair provides information (the nonexistence of an edge is itself information), whereas in our model node pairs not connected in G don't provide information.

Despite the models' differences, we've successfully adapted the techniques discussed in our paper¹ to a stochastic block model setting.²

Our paper's main contribution was demonstrating that, if we consider the observation graph an Erdős-Rényi graph (a random graph in which each node is connected by an edge, independently and with probability p), with average degree $(n-1)p$, exact recovery of x^V is possible with high probability if and only if

$$\alpha = np / \log(n) > 2 / (1 - 2\varepsilon)^2 + o(2 / (1 - 2\varepsilon)^2).$$

If $\alpha < 1$, then we know with high probability that the observation graph contains isolated nodes. This renders recovery impossible even in a noiseless

setting, because there would be nodes with no available information.

On the algorithmic side, we propose the use of a semidefinite programming relaxation-based algorithm. Duality and estimates on spectral norms of certain random matrices show that this efficient algorithmic approach exactly recovers x^V with high probability at regimes very close to the information-theoretical limit.³ Remarkably, with the use of sharper estimates on the spectrum of random matrices, these results have since improved.^{4,5}

The techniques in our paper have since been adapted to community detection with more than two communities.^{5–8} Furthermore, although our work focused on exactly recovering x^V , more recent studies have solved one of the open problems we posed in our paper: how to partially recover $x^{V,10}$. 

REFERENCES

1. E. Abbe et al., "Decoding Binary Node Labels from Censored Edge Measurements: Phase Transition and Efficient Recovery," *IEEE Trans. Network Science and Eng.*, vol. 1, no. 1, 2014, pp. 10–22.
2. E. Abbe, A.S. Bandeira, and G. Hall, "Exact Recovery in the Stochastic Block Model," 2014; <http://arxiv.org/abs/1405.3267>.
3. J.A. Tropp, "User-Friendly Tail Bounds for Sums of Random Matrices," *Foundations of Computational Mathematics*, vol. 12, no. 4, 2012, pp. 389–434.
4. A.S. Bandeira, "Random Laplacian Matrices and Convex Relaxations," 2015; <http://arxiv.org/abs/1504.03987>.
5. B. Hajek, Y. Wu, and J. Xu, "Achieving Exact Cluster Recovery Threshold via Semidefinite Programming: Extensions," 2015; <http://arxiv.org/abs/1502.07738>.
6. N. Agarwal et al., "Multisection in the Stochastic Block Model Using Semidefinite Programming," 2015; <http://arxiv.org/abs/1507.02323>.
7. W. Perry and A.S. Wein, "A Semidefinite Program for Unbalanced Multisection in the Stochastic Block Model," 2015; <http://arxiv.org/abs/1507.05605>.
8. E. Abbe and C. Sandon, "Community Detection in General Stochastic Block Models: Fundamental Limits and Efficient Recovery Algorithms," 2015; <http://arxiv.org/abs/1503.00609>.
9. P. Chin, A. Rao, and V. Vu, "Stochastic Block Model and Community Detection in the Sparse Graphs: A Spectral Algorithm with Optimal Rate of Recovery," 2015; <http://arxiv.org/abs/1501.05021>.
10. A. Saade et al., "Spectral Detection in the Censored Block Model," 2015; <http://arxiv.org/abs/1502.00163>.

AFONSO S. BANDEIRA is an instructor and postdoctoral fellow in MIT's Department of Mathematics. Contact him at bandeira@mit.edu.



Selected CS articles and columns are also available for free at <http://ComputingNow.computer.org>.



handles the details *so you don't have to!*

- Professional management and production of your publication
- Inclusion into the IEEE Xplore and CSDL Digital Libraries
- Access to CPS Online: Our Online Collaborative Publishing System
- Choose the product media type that works for your conference:
Books, CDs/DVDs, USB Flash Drives, SD Cards, and Web-only delivery!

Contact CPS for a Quote Today!

www.computer.org/cps or cps@computer.org



IEEE  computer society