PARTIAL RECOVERY ON RANDOM REGULAR GRAPHS

DYLAN AIREY AND STEVEN DAMELIN

Abstract. The problem of community detection from noisy observations of interactions between individuals has important applications to diverse areas such as biology, sociology, and computer science. This problem is particularly interesting when the number of interactions is on the order of the size of the populations, that is the graph where two interacting individuals are connected by an edge is sparse. One model which has been studied extensively is the labelled stochastic block model. In this model a planted partition is used to generate a random graph with a random edge labelling. In the sparse case the goal is to recover communities which are correlated with the planted partition, that is partially recover the planted partition. Over the course of a number of papers a threshold has been established where partial recovery transitions from impossible to possible. We aim to strengthen these results and find the expected optimal overlap with the planted partition in the regime where partial recovery is possible.

1. Introduction

Techniques from statistical mechanics, in particular the theory of spin glasses, have recently been very effective in studying many problems in discrete mathematics and computer science. One such problem is that of community detection. Given two equally sized populations, say of size $\frac{n}{2}$, connect each pair of individuals in the same population by an edge with probability $\frac{a}{n}$ and each pair of individuals in different populations with probability $\frac{b}{n}$ where $0 \leq a, b \leq n$. This generates a random graph on the populations called the **stochastic block model** and one would like to be able to recover the two populations as $n \to \infty$.

When $a, b \in O(1)$ the average degree remains bounded and we are in the **sparse regime**. Here we cannot hope to fully recover the two populations. Instead the best we can do is to recover two populations which are positively correlated with the true populations. We call this **partial recovery**. E. Mossel, J. Neeman, and A. Sly proved in [4] that partial recovery is possible asymptotically almost surely (a.a.s.) for the stochastic block model if

$$(a-b)^2 > 2(a+b)$$

and impossible if

$$(a-b)^2 < 2(a+b).$$

This threshold matches that predicted by A. Decelle, F. Krzakala, C. Moore, and L. Zdeborová in [1] which was established through sophisticated but non-rigorous predictions from statistical mechanics. They used the cavity method to argue that a belief propagation (BP) algorithm will partially recover the true populations when possible and gave strong numerical evidence for their conjectures. Prior to [4], A. Coja-Oghlan in [2] used spectral algorithms to show that partial recovery is possible if $(a-b)^2 > C(a+b)$ for some sufficiently large constant C.

A generalization of the stochastic block model called the **labelled stochastic block model** introduces a random $\{0,1\}$ labelling on the edges. Given $0 \le a, b \le n$ and $\epsilon \in [0, 1/2]$ generate the random graph structure on the two populations as in the stochastic block model and now label each edge connecting two individuals in the same population with 1 with probability ϵ and 0 with probability $1-\epsilon$ and label an edge connecting two individuals in different populations with a 0 with probability ϵ and 1 with probability $1-\epsilon$. The goal now is to recover the two populations given the graph as well as the edge labels.

In the dense regime, E. Abbe, A. Bandeira, A. Bracher, and A. Singer in [3] showed in the case $a = b$ that one can exactly recover the two populations a.a.s. if

$$\frac{a}{\log n} > \frac{2}{(1-2\epsilon)^2} + o\left(\frac{1}{(1-2\epsilon)^2}\right)$$

but that exact recovery is impossible if the inequality is reversed.

In the sparse regime M. Lelarge, L. Massoulié, and J. Xu showed in [9] that partial recovery in the labelled stochastic block model is impossible a.a.s. if

$$\tau := \frac{1}{2}\left(\frac{(a(1-\epsilon)-b\epsilon)^2}{a(1-\epsilon)+b\epsilon} + \frac{(a\epsilon - b(1-\epsilon))^2}{a\epsilon + b(1-\epsilon)}\right) < 1$$

and F. Krzakala, M. Lelarge, A. Saade, and L. Zdeborová in [5] in the case where $a = b$ showed that partial recovery is possible a.a.s. if

$$\tau > 1.$$

Note in the case that $\epsilon = \frac{1}{2}$ we have the unlabelled stochastic block model and $\tau = \frac{1}{2} n \frac{(a-b)^2}{a+b}$ giving the same threshold as [4]. The proof in [5] relied on spectral analysis of the non-backtracking operator and the Bethe-Hessian operator as opposed to the standard Laplacian of the random graph. Additionally S. Heimlicher, M. Lelarge, and L. Massoulié in [11] give numerical evidence that belief propagation (BP) with random initial conditions partially recovers the true partition when $\tau > 1$. The authors in [5] also numerically test their spectral algorithms and find that they on average achieve an overlap with the true partition slightly less than that achieved by BP. It is conjectured that BP in fact achieves the optimal overlap. We would like to understand the expected optimal overlap with the true partition.

2. Optimal overlap with the planted partition

Fix a set of vertices $V = [n]$ and an integer $d > 2$. Instead of the random Erdös-Renyi graph we work with the **random regular graph**, that is we pick a d-regular graph with vertex set V uniformly at random from all such d-regular graphs. This uniform sampling is difficult to do in general so we instead sample a d-regular graph with the **configuration model**. Under this model one place d labelled half-edges on each vertex and then selects a perfect matching on the half-edges uniformly at random. Note that we must have nd even for there to be a perfect matching. We make this assumption from now on. It is possible to create non-simple graphs through this procedure but the probability that the graph generated is simple remains bounded above 0 as $n \to \infty$. Thus these two models for generating d-regular graphs share probability 0 events. The configuration model lends itself well to combinatorial arguments. For example the total number of perfect matchings on the nd half-edges is given by the double factorial

$$(nd-1)!! := (nd-1)\times(nd-3)\times\cdots\times 3\times 1.$$

For a d-regular graph $G=(V,E)$ and vertex configuration $\sigma^V\in\{0,1\}^V$ the probability of an edge configuration $\sigma^E\in\{0,1\}^E$ is

$$\mathbb{P}^G(\sigma^E|\sigma^V)=\prod_{(v,w)\in E}\epsilon^{\sigma_v\oplus\sigma_w\oplus\sigma_{(v,w)}}(1-\epsilon)^{1\oplus\sigma_v\oplus\sigma_w\oplus\sigma_{(v,w)}}.$$

Note that this is a **spin glass** which has been considered in [10] where the couplings between vertex variables are correlated with σ^V. Equivalently we can write

$$\sigma^E = B_G\sigma^V\oplus\xi^E$$

where $\xi^E=(\xi_e)_{e\in E}$ is a sequence of iid $Bernoulli(\epsilon)$ distributed random variables and B_G is the $|V|\times|E|$ incidence matrix of G. From now on we will assume that each vertex configuration is equally likely, that is $\mathbb{P}^G(\sigma^V)=\frac{1}{2^n}$. A natural candidate for an estimate of the true populations (now thought of as an element of $\{0,1\}^V$) is the maximum a posteriori (MAP) estimator. Recall that given an edge configuration σ^E the MAP estimator is given by

$$\mathrm{MAP}(\sigma^E)=\underset{\sigma^V\in\{0,1\}^V}{\arg\max}\ \mathbb{P}(\sigma^V|\sigma^E).$$

Under the assumption of uniform priors this is equivalent to the maximum likelihood (ML) estimator

$$\mathrm{ML}(\sigma^E)=\underset{\sigma^V\in\{0,1\}^V}{\arg\max}\ \mathbb{P}(\sigma^E|\sigma^V).$$

Note the ML estimator is the vertex configuration $\tilde\sigma^V$ such that for each $\sigma^V\neq\tilde\sigma^V,\tilde\sigma^V\oplus 1^V$

$$d_H(B_G\tilde\sigma^V,\sigma^E)<d_H(B_G\sigma^V,\sigma^E)$$

where d_H is the Hamming metric on $\{0,1\}^E$. Equivalently this is the ground state of the Hamiltonian of the spin glass. Define $D_G(x^V)=\mathrm{ML}^{-1}(x^V)$ and

$$D_G(0^V,p)=\bigcup_{\substack{x^V\in\{0,1\}^V\\ \sum_v x_v=np}} D_G(x^V).$$

Then partial recovery is possible a.a.s. if and only if

$$\lim_{p\to 1/2}\lim_{n\to\infty}\mathbb{P}\left(\xi^E\in\bigcup_{q\le p}D_G(0^V,q)\right)=1.$$

Our goal is to determine the minimum value p_0 such that

$$\lim_{p\to p_0}\lim_{n\to\infty}\mathbb{P}\left(\xi^E\in\bigcup_{q\le p}D_G(0^V,q)\right)=1.$$

The value of $\frac{1}{nd/2}\sum_e \xi_e$ concentrates about ϵ. Thus we partition $D_G(0^V, p)$ based on edge density by defining

$$D_G(0^V, p, \alpha) = \left\{ \sigma^E \in D_G(0^V, p) : \sum_e \sigma_e = \frac{nd\alpha}{2} \right\}.$$

Then

$$\begin{aligned}
\mathbb{P}\left[\xi^E \in D_G(0^V, p)\right] &= \sum_{k=0}^{nd/2} \mathbb{P}\left[\xi^E \in D_G\left(0^V, p, \frac{2k}{nd}\right)\right] \\
&= \sum_{k=0}^{nd/2} \mathbb{P}\left[\xi^E \in D_G\left(0^V, p, \frac{2k}{nd}\right) \Big| \sum_e \xi_e = k\right] \mathbb{P}\left[\sum_e \xi_e = k\right] \\
&= \sum_{k=0}^{nd/2} \frac{\mathbb{E}\left[\# D_G\left(0^V, p, \frac{2k}{nd}\right)\right]}{\binom{nd/2}{k}} \binom{nd/2}{k} \epsilon^k (1-\epsilon)^{nd/2-k} \\
&= \sum_{k=0}^{nd/2} e^{n\Phi\left(p, \frac{2k}{nd}\right) + o(n)} \epsilon^k (1-\epsilon)^{nd/2-k}
\end{aligned}$$

where we have defined

$$\Phi(p, \alpha) := \lim_n \frac{1}{n} \log \mathbb{E}\left[\# D_G(0^V, p, \alpha)\right].$$

This sum is then

$$\begin{aligned}
&= \sum_\alpha \exp n \left(\Phi(p, \alpha) + \frac{d}{2}\alpha \log \epsilon + \frac{d}{2}(1-\alpha) \log 1 - \epsilon + o(1) \right) \\
&= \sum_\alpha \exp n \left(\Phi(p, \alpha) - \frac{d}{2} H(\alpha) - \frac{d}{2} D(\alpha || \epsilon) + o(1) \right).
\end{aligned}$$

Since this sum is over $O(n)$ terms a necessary condition for $\lim_{p \to p_0} \lim_{n \to \infty} \mathbb{P}\left(\xi^E \in \bigcup_{q<p} D_G(0^V, q)\right) = 1$ is

$$\limsup_{p \to p_0} \sup_\alpha \Phi(p, \alpha) - \frac{d}{2} H(\alpha) - \frac{d}{2} D(\alpha || \epsilon) = 0.$$

Unfortunately computing $\Phi(p, \alpha)$ is difficult. The set $D_G(0^V, p, \alpha)$ is defined in terms of a global minimizer so the uniform measure on $D_G(0^V, p, \alpha)$ cannot be expressed as a factor model. To avoid these difficulties we approximate $D_G(0^V, p, \alpha)$ by larger sets $D_G^{(r)}(0^V, p, \alpha)$ with $r \in \mathbb{N}$ where

$$\begin{aligned}
D_G^{(r)}(0^V, p, \alpha) = \bigcup_{\substack{\sigma^V \in \{0,1\}^V \\ \sum_v \sigma_v = np}} \Bigg\{ \sigma^E \in \{0,1\}^E : &\sum_{(u,w) \in E} \sigma_{(u,w)} = \frac{nd\alpha}{2}, \forall v \in V, \forall S \subseteq B_r(v) \\
&\sum_{(u,w) \in E(S, V \setminus S)} \sigma_{(u,w)} \oplus \sigma_u \oplus \sigma_w \le \frac{1}{2} \# E(S, V \setminus S) \Bigg\}.
\end{aligned}$$

An edge configuration is in $D_G^{(r)}(0^V, p, \alpha)$ if ML decoding decides for some σ^V with $\sum_v \sigma_v = np$ over any node configuration given by perturbing σ^V on a ball of radius r.

We have

$$D_G(0^V, p, \alpha) \subseteq D_G^{(r)}(0^V, p, \alpha)$$

so

$$\Phi_r(p, \alpha) := \lim_n \frac{1}{n} \log \mathbb{E}\left[D_G^{(r)}(0^V, p, \alpha)\right] \geq \Phi(p, \alpha).$$

3. Computing $\Phi_r(p, \alpha)$

A basic but extremely important fact about random regular graphs is that they **locally converge** to the d-regular tree T_d. That is, if one selects uniformly at random a vertex ρ from a random regular graph G, then for r fixed, with probability tending to 1 as $n \to \infty$ the r-radius neighborhood of ρ will be a tree. This will allow us to compute $\Phi_r(p, \alpha)$ as the supremum over the measure-theoretic entropies of Markov chains on T_d which are compatible with $D_G^{(r)}(0^V, p, \alpha)$.

Let $T_d(k)$ be the radius r neighborhood of some node in T_d and let $E(T_d(k))$ be the set of edges in $T_d(r)$. Define

$$\mathcal{X} = \{(x_v, y_v) \in \{0,1\}^{T_d(r)} \times \{0,1\}^{E(T_d(r))} : \forall S \subseteq T_d(r)$$
$$\sum_{(u,w) \in E(S, T_d(r) \setminus S)} x_u \oplus x_w \oplus x_{(u,w)} \leq \frac{1}{2} \# E(S, T_d(r) \setminus S)\}$$

Let $\phi : \{0,1\}^V \times \{0,1\}^E \to \left(\{0,1\}^{T_d(r)} \times \{0,1\}^{E(T_d(r))}\right)^V$ be the map where we use the labelling on the half-edges of G to identify $v \in V$ with the root $\rho \in T_d(r)$ and $\partial^i v$ with $\partial^i \rho$ for each $1 \leq i \leq r$ and then map the vertex and edge labels on $B_r(v)$ to their corresponding vertices and edges in $T_d(r)$. Given $\sigma^V \in \{0,1\}^V$ and $\sigma^E \in \{0,1\}^E$ consider the $d+1$-tuple of empirical measures $(\dot{\pi}, \hat{\pi}_1, \cdots, \hat{\pi}_d)$ where $\dot{\pi}$ is a measure on $\{0,1\}^{T_d(r)} \times \{0,1\}^{E(T_d(r))}$ and $\hat{\pi}_i$ is a measure on $\left(\{0,1\}^{T_d(r)} \times \{0,1\}^{E(T_d(r))}\right)^2$ for each $1 \leq i \leq d$ defined by

$$\dot{\pi}(\omega) := \frac{1}{n} \sum_{v \in V} \mathbb{1}_{\phi(\sigma^V, \sigma^E)_v = \omega}$$
$$\hat{\pi}(\omega, \tau) := \frac{1}{n} \sum_{v \in V} \mathbb{1}_{\phi(\sigma^V, \sigma^E)_v = \omega, \phi(\sigma^V, \sigma^E)_{h_i(v)} = \tau}$$

where $h_i(v)$ is the vertex in V which is matched with the half-edge i of v. Note that

$$\dot{\pi}(\omega) = \sum_\tau \hat{\pi}_i(\omega, \tau) = \sum_\tau \hat{\pi}_i(\tau, \omega) \tag{1}$$

for every $1 \leq i \leq d$.

Let Δ be the set of $d+1$-tuples of measures which satisfy (1) and where $\dot{\pi}$ is supported on $\mathcal{X}$ and $\hat{\pi}_i$ is supported in $\mathcal{X} \times \mathcal{X}$. Let $\Delta_{p,\alpha} \subset \Delta$ be the subset of $d+1$-tuples which satisfy

$$\int \omega_\rho d\dot{\pi}(\omega) = p$$

$$\int \sum_{v\in\partial\rho} \omega_{(\rho,v)} d\dot{\pi}(\omega) = 2\alpha.$$

For $\pi = (\dot{\pi}, \hat{\pi}_i)$ let $Z(\pi)$ be the number of (σ^V, σ^E) which have empirical measures π. Then

$$\mathbb{E}\left[\# D_G^{(r)}(0^V, p, \alpha)\right] = e^{o(1)} \sum_{\pi\in\Delta_{p,\alpha}} \mathbb{E}\left[Z(\pi)\right]$$

where the $e^{o(1)}$ appears since there can be a non-zero but still vanishingly small fraction of the vertices with $B_r(v)$ having a cycle. We have

$$\begin{aligned}\mathbb{E}\left[Z(\pi)\right] = &\left(n! \prod_{\omega} \frac{\mathbb{1}_{\omega\in\mathcal{X}}}{(n\dot{\pi}(\omega))!}\right) \times \left(\prod_{i=1}^{d}\prod_{\omega} (n\dot{\pi}(\omega))! \prod_{\tau\in\mathcal{X}} \frac{\varphi_i(\omega,\tau)}{(n\hat{\pi}_i(\omega,\tau))!}\right) \\ &\times \frac{1}{nd-1!!}\left(\prod_{\omega}\left(\prod_{\tau\neq\omega}\left(n\sum_i \hat{\pi}_i(\omega,\tau)\right)!\right)^{1/2}\left(n\sum_i \hat{\pi}_i(\omega,\omega)\right)!!\right)\end{aligned}$$

where $\varphi_i(\omega,\tau)$ is the indicator that there is a $\sigma \in \{0,1\}^{T_d}$ such that $\sigma|_{B_r(\rho)} = \omega$ and $\sigma|_{B_r(h_i(\rho))} = \tau$. Then

$$\Phi(\pi) := \lim_n \frac{1}{n}\mathbb{E}\left[Z(\pi)\right] = (1-d)H(\dot{\pi}) + \sum_{i=1}^{d} H(\hat{\pi}_i) - \frac{d}{2}H\left(\frac{1}{d}\sum_{i=1}^{d}\hat{\pi}_i\right).$$

where H is the standard Shannon entropy. This implies

$$\Phi_r(p,\alpha) = \sup_{\pi\in\Delta_{p,\alpha}} \Phi(\pi).$$

Taking the supremum over this hyperplane in Δ could be difficult so we introduce two Lagrange multipliers $\log\dot{\lambda}$ and $\log\hat{\lambda}$ to get

$$\begin{aligned}\Phi_r(p,\alpha) &= \sup_{\pi\in\Delta} \Phi(\pi) + \log\dot{\lambda}\int \omega_\rho d\dot{\pi}(\omega) + \frac{1}{2}\log\hat{\lambda}\int\sum_{v\in\partial\rho}\omega_{(\rho,v)}d\dot{\pi}(\omega) \\ &=: \sup_{\pi\in\Delta} \Phi^{\dot{\lambda},\hat{\lambda}}(\pi)\end{aligned}$$

where we will choose the values of the Langrange multipliers to ensure the maximizer lies in $\Delta_{p,\alpha}$.

3.1. **Finding the maximizers of $\Phi^{\dot{\lambda},\hat{\lambda}}(\pi)$.** The goal now is to apply the ideas from spin glass theory to show that a maximizer of $\Phi^{\dot{\lambda},\hat{\lambda}}(\pi)$ is the fixed point of a belief propagation algorithm. Such techniques have been employed successfully by J. Ding, A. Sly, and N. Sun in [6, 7, 8]. It is likely that these methods will be significantly harder in our case: the space of spins considered in these other papers have size on the order of d^2 while the space of spins for us when $d = 4$ and $r = 3$ is already of order 2^{33}. We hope to dramatically reduce the dimensionality by characterizing some limiting behavior as $r \to \infty$ but currently have no solid evidence that this will be successful. Another difficulty is that $\mathcal{X}$ and the transition function $\varphi : \mathcal{X} \to \mathcal{X}$ have a more intricate structure than the spins in these other papers and we would likely need a good combinatorial description of $\mathcal{X}$ and φ in order to make further progress. Unfortunately given the high dimensionality finding the maximizer with a computer is not feasible even for $r = 4$.

References

1. C. Moore L. Zdeborová A. Decelle, F. Krzakala, *Asymptotic analysis of the stochastic block model for modular networks and its algorithmic applications.*
2. A. Coja-Oghlan, *Graph partitioning via adaptive spectral techniques.*
3. A. Bracher A. Singer E. Abbe, A. Bandeira, *Decoding binary node labels from censored edge measurements: phase transition and efficient recovery*, IEEE Transactions on Network Science and Engineering (2014).
4. A. Sly E. Mossel, J. Neeman, *Reconstruction and estimation in the planted partition model.*
5. A. Saade L. Zdeborová F. Krzakala, M. Lelarge, *Spectral detection in the censored block model.*
6. N. Sun J. Ding, A. Sly, *Proof of the satisfiability conjecture for large k*, arXiv (2014).
7. ——, *Satisfiability threshold for random regular nae-sat*, Proceedings of the ACM Symposium on Theory of Computing (2014).
8. ——, *Maximum independent sets on random regular graphs*, arXiv (2015).
9. J. Xu M. Lelarge, L. Massoulié, *Reconstruction in the labeled stochastic block model.*
10. G. Parisi M. Mézard and M.A. Virasoro, *Spin glass theory and beyond.*
11. L. Massoulié S. Heimlicher, M. Lelarge, *Community detection in the labelled stochastic block model.*