

Lecture 3: Block models & information-computation gaps

In this lecture we give another simple application of the sum-of-squares paradigm, to the problem of recovering communities in stochastic block models with average degrees $\Omega(\log n)$. Sparse stochastic block models exhibit a notorious information-computation gap, and we describe the gap as well as give a heuristic derivation of the threshold. Some bibliographic remarks will be deferred to the end.

These notes have not been reviewed with the same scrutiny applied to formal publications. There may be errors.

1 Stochastic block models

Definition 1.1. The *stochastic block model* is a family of distributions over graphs, parametrized by integers n (the number of vertices) and k (the number of communities), and probabilities p_{in} (internal edge probability) and p_{out} (external edge probability).

One samples $G \sim \mathcal{G}(n, k, p_{\text{in}}, p_{\text{out}})$ as follows:

1. Identify the vertex set $V(G)$ with $[n]$.
2. Sample a uniform partition of $[n]$ into k sets or “communities” of equal size. Let $X \in \{0, 1\}^{n \times k}$ represent this partition, with $X_{ic} = 1$ [vertex $i \in$ community c].
3. For each pair $i \neq j \in [n]$ independently, add edge (i, j) to $E(G)$ with probability p_{in} if i, j are in the same community, and with probability p_{out} otherwise.

In this lecture, we will focus on the case when $k = \Theta(1)$ and $\min(p_{\text{in}}n, p_{\text{out}}n) = \Omega(1)$. Some of what we prove can be extended to other regimes too.

These simple models are employed in multiple contexts; they are basic models for social networks, in which each partition can model a community of people (people are more likely to know other people from the same community). $\mathcal{G}(n, k, p_{\text{in}}, p_{\text{out}})$ is a primitive model for a social network, to be sure, but network science is hard, and if you want to evaluate a clustering algorithm (or some other sort of algorithm), it’s not a bad place to start. Block models can also model physical systems, in which we think of each partition as a type of particle, and particles of the same type are either attracting ($p_{\text{in}} > p_{\text{out}}$) or repulsive ($p_{\text{in}} < p_{\text{out}}$).

Remark 1.2. Block models are sometimes defined more flexibly; the model above is sometimes referred to as the “symmetric block model.” In the most general case, we allow communities of different sizes, and distinct probabilities between each pair of communities, $\{p_{a,b}\}_{a,b \in [k]}$.

Even in the symmetric block model, sometimes the partition X is chosen so that the community of each $i \in [n]$ is chosen from $\text{Unif}([k])$ independently, so that the parts in the partition are not even. This is sometimes more convenient for calculations, and when $k = \Theta(1)$, the differences between the models are usually qualitatively negligible.

Algorithmic questions on block models. There are two algorithmic questions typically associated with block models:

1. **Recovery:** given $G \sim \mathcal{G}(n, k, p_{\text{in}}, p_{\text{out}})$, recover X (either exactly or approximately).

2. **Detection:** you are given a graph G , and you wish to hypothesis test between the two hypotheses: $G \sim \mathcal{H}_0 = \mathcal{G}(n, \frac{1}{k}p_{\text{in}} + \frac{k-1}{k}p_{\text{out}})$, or $G \sim \mathcal{H}_1 = \mathcal{G}(n, k, p_{\text{in}}, p_{\text{out}})$. In the null hypothesis \mathcal{H}_0 , there is no community structure.

In what follows, we will give an SoS algorithm for recovery in the regime $p_{\text{in}}, p_{\text{out}} = \Omega(\frac{\log n}{n})$, and we will discuss a notorious information-computation gap for detection in the regime $p_{\text{in}}, p_{\text{out}} = O(\frac{1}{n})$.

2 Sum-of-squares paradigm for recovery in block models

In the recovery problem, we wish to recover the partition X which maximizes $\Pr[X | G]$, the *maximum likelihood estimator*. Here, the maximum likelihood estimate is not convenient to work with, and in particular the constraint that X is the maximum likelihood estimator cannot be easily expressed as a low-degree polynomial constraint. Instead, we'll consider the *maximum a-posteriori estimator*, or the *MAP estimator*,

$$Y^* = \underset{Y \text{ balanced } k\text{-partition}}{\arg \max} \Pr[G | Y].$$

We'll prove that if G is sampled from partition X , then with high probability $(Y^*)(Y^*)^\top$ is close to XX^\top . Further, we'll give a sum-of-squares proof of this fact, which will give us an algorithm.

First, it will be useful for us to better understand the maximizer Y^* . The following characterization of the maximizer is easier to encode in a low-degree polynomial system.

Lemma 2.1. *The balanced k -way partition Y^* maximizing $\Pr[G | Y]$ is also the balanced k -way partition maximizing the function $\langle A_G, YY^\top \rangle \cdot \log \frac{p_{\text{in}}(1-p_{\text{out}})}{p_{\text{out}}(1-p_{\text{in}})}$, or equivalently,*

$$Y^* = \begin{cases} \arg \max_Y \langle A_G, YY^\top \rangle & \frac{p_{\text{in}}}{1-p_{\text{in}}} > \frac{p_{\text{out}}}{1-p_{\text{out}}} \\ \arg \min_Y \langle A_G, YY^\top \rangle & \frac{p_{\text{in}}}{1-p_{\text{in}}} < \frac{p_{\text{out}}}{1-p_{\text{out}}} \end{cases}.$$

Proof. For $i \in [n]$, let $Y_i \in \{0, 1\}^k$ be such that $Y_i(c) = Y_{ic}$. Note that for $i, j \in [n]$, $\langle Y_i, Y_j \rangle = \mathbf{1}[i, j \text{ in same community}]$. Now since the edges of G are independent conditioned on Y , we can write,

$$\Pr[G | Y] = \prod_{i, j \in [n]} \left(p_{\text{in}}^{\langle Y_i, Y_j \rangle} \cdot p_{\text{out}}^{1-\langle Y_i, Y_j \rangle} \right)^{\mathbf{1}[(i, j) \in E(G)]} \left((1-p_{\text{in}})^{\langle Y_i, Y_j \rangle} (1-p_{\text{out}})^{1-\langle Y_i, Y_j \rangle} \right)^{\mathbf{1}[(i, j) \notin E(G)]}$$

Because \log is a strictly increasing function, the maximizer of $\Pr[G | Y]$ is the same as the maximizer of $\log \Pr[G | Y]$. Now, applying the logarithm and simplifying the above, we can write,

$$\begin{aligned} \log \Pr[G | Y] &= \sum_{(i, j) \in E(G)} \langle Y_i, Y_j \rangle \log \frac{p_{\text{in}}}{p_{\text{out}}} + \sum_{(i, j) \notin E(G)} \langle Y_i, Y_j \rangle \log \frac{1-p_{\text{in}}}{1-p_{\text{out}}} \\ &\quad + |E(G)| \log p_{\text{out}} + \left(\binom{n}{2} - |E(G)| \right) \log(1-p_{\text{out}}). \end{aligned}$$

The quantity on the second line does not depend on Y at all, and so we can call it $\text{stuff}(G)$. Expressing the above as a matrix-matrix inner product,

$$= \frac{1}{2} \log \frac{1-p_{\text{in}}}{1-p_{\text{out}}} \cdot \langle YY^\top, J \rangle + \frac{1}{2} \log \frac{p_{\text{in}}(1-p_{\text{out}})}{p_{\text{out}}(1-p_{\text{in}})} \cdot \langle YY^\top, A_G \rangle + \text{stuff}(G),$$

for J the $n \times n$ all-1's matrix, and A_G the adjacency matrix of G .¹ We can simplify even further with the following claim:

¹Really to be absolutely correct we should subtract the diagonal of J , since we don't consider the $i = j$ terms. However the diagonal of YY^\top is just $\mathbb{1}$, so this will not matter, and you can check that this does not affect the overall argument.

Claim 2.2. For any balanced k -way partition Y of $[n]$, $\langle YY^\top, J \rangle = \frac{n^2}{k}$.

Proof of Claim 2.2. We can manipulate the expression algebraically, to start:

$$\langle YY^\top, J \rangle = \langle YY^\top, \mathbf{1}\mathbf{1}^\top \rangle = \text{Tr}(YY^\top \mathbf{1}\mathbf{1}^\top) = \|\mathbf{1}^\top Y\|_2^2.$$

Now, for each column Y^c of Y , we have that $\langle \mathbf{1}, Y^c \rangle = \frac{n}{k}$ by the balanced assumption. This yields $\langle YY^\top, J \rangle = \left(\frac{n}{k}\right)^2 \cdot k$ as desired. \square

Applying Claim 2.2, we can push the $\langle YY^\top, J \rangle$ term into $\text{stuff}(G)$, creating a term $\text{stuff}'(G)$ which again does not depend on the specific choice of balanced partition Y . So,

$$\log \Pr[G | Y] = \log \frac{p_{\text{in}}(1 - p_{\text{out}})}{p_{\text{out}}(1 - p_{\text{in}})} \cdot \langle YY^\top, A_G \rangle + \text{stuff}'(G),$$

from which the conclusion follows. \square

2.1 Algorithm from the SoS paradigm

We will write a polynomial system which searches for Y^* . Our variables will be $Y \in \mathbb{R}^{n \times k}$, and our axioms will enforce that Y corresponds to a balanced k -way partition:

$$\mathcal{A} = \{Y_{ic}^2 = Y_{ic}\}_{i \in [n], c \in [k]} \cup \left\{ (\mathbf{1}_n^\top Y)_c = \frac{n}{k} \right\}_{c \in [k]} \cup \{ \langle Y_i, \mathbf{1}_k \rangle = 1 \}_{i \in [n]}.$$

Our polynomial optimization program will be

$$\max \langle YY^\top, A_G \rangle \text{ s.t. } \mathcal{A}.$$

The theorem below is not optimal in its dependence on k ; see [HWX16, Ban18] for a sharper analysis of the performance of SDPs for exact recovery in block models with $k = 2$, and [AS15] for an information-theoretically optimal efficient algorithm.

Theorem 2.3. Suppose $p_{\text{in}} = \alpha \frac{\log n}{n}$, $p_{\text{out}} = \beta \frac{\log n}{n}$ satisfy $p_{\text{in}} > p_{\text{out}}$ ², for $\alpha, \beta = \Omega(1)$. Then with high probability over the choice of G ,

$$\mathcal{A} \vdash_2 \left\langle \frac{\tilde{\mathbf{E}}[YY^\top]}{\|\tilde{\mathbf{E}}[YY^\top]\|_F}, \frac{XX^\top}{\|XX^\top\|_F} \right\rangle \geq 1 - \frac{\sqrt{k\alpha} + \sqrt{k^3\beta}}{\alpha - \beta} \cdot \sqrt{\frac{k}{\log n}}.$$

This is to say that, if α and β are sufficiently separated compared to $\sqrt{k\alpha} + \sqrt{\beta}$ and $\frac{k}{\log n}$, we know that $\tilde{\mathbf{E}}[YY^\top]$ must be closely aligned with the block matrix XX^\top which defines the communities, so the communities can be recovered up to some error (depending on ϵ). With more work one can actually show that an SDP-based algorithm recovers X with *no error* with high probability; again see [HWX16, Ban18].

Proof. We want to use the fact that the objective value, $\langle YY^\top, A_G \rangle$ is large to conclude that YY^\top should be close to XX^\top . To that end, we will split A_G into a sum of its expectation and a random matrix representing its deviation,

$$A_G = \mathbf{E}[A_G] + \Delta = p_{\text{out}} \cdot J_n + (p_{\text{in}} - p_{\text{out}}) \cdot XX^\top + \Delta.$$

²The opposite case can be handled by a near-identical proof

Now we can write,

$$\begin{aligned}
\langle YY^\top, A_G \rangle &= p_{\text{out}} \cdot \langle YY^\top, J \rangle + (p_{\text{in}} - p_{\text{out}}) \langle YY^\top, XX^\top \rangle + \langle YY^\top, \Delta \rangle \\
&\leq p_{\text{out}} \cdot \langle YY^\top, J \rangle + (p_{\text{in}} - p_{\text{out}}) \langle YY^\top, XX^\top \rangle + \sum_{c \in [k]} \|Y^c\|^2 \cdot \lambda_{\max}(\Delta) \\
&= p_{\text{out}} \cdot \frac{n^2}{k} + (p_{\text{in}} - p_{\text{out}}) \langle YY^\top, XX^\top \rangle + n \cdot \lambda_{\max}(\Delta),
\end{aligned}$$

where the inequality was a degree-2 sum-of-squares inequality (since $\Delta \leq \lambda_{\max}(\Delta) \cdot \mathbb{1}$) and the final equality uses \mathcal{A} to conclude that $\|Y^c\|^2 = \langle Y^c, \mathbf{1}_n \rangle = \frac{n}{k}$ for each $c \in [k]$, and also since the proof of [Claim 2.2](#) is an SoS proof modulo the axioms $\mathbf{1}^\top Y^c = \frac{n}{k}$ in \mathcal{A} . So we conclude that

$$\mathcal{A} \vdash_2 (p_{\text{in}} - p_{\text{out}}) \langle YY^\top, XX^\top \rangle \geq \langle YY^\top, A_G \rangle - p_{\text{out}} \frac{n^2}{k} - n \lambda_{\max}(\Delta). \quad (1)$$

Now, we want to understand what this implies about $\tilde{\mathbf{E}}[YY^\top]$ and XX^\top . Clearly \mathcal{A} is feasible (so long as $\frac{n}{k}$ is an integer), so a pseudoexpectation operator respecting \mathcal{A} must exist. Further, since the pseudoexpectation is a relaxation of a moment operator over the set of solutions to \mathcal{A} , we have that the maximizing (recall $p_{\text{in}} > p_{\text{out}}$) pseudoexpectation satisfies

$$\langle \tilde{\mathbf{E}}[YY^\top], A_G \rangle \geq \langle XX^\top, A_G \rangle \approx \mathbf{E}_{G|X}[\langle XX^\top, A_G \rangle] = p_{\text{in}} \cdot \frac{n^2}{k}, \quad (2)$$

since $\langle XX^\top, A_G \rangle$ is a sum of independent random variables conditioned on X , so we expect it to concentrate quite well around its mean (so long as $p_{\text{in}} n^2 / k \gg 1$). We'll content ourselves with pretending this \approx is an equality, keeping in mind that it is actually only true up to $(1 \pm o(1))$ factors.

Hence combining (1) and (2),

$$\langle \tilde{\mathbf{E}}[YY^\top], XX^\top \rangle \geq \frac{n^2}{k} - \frac{n \lambda_{\max}(\Delta)}{p_{\text{in}} - p_{\text{out}}} \quad (3)$$

Also, since XX^\top is a 0/1 valued matrix, $\|XX^\top\|_F^2 = \langle J, XX^\top \rangle = \frac{n^2}{k}$ (the final equality is by [Claim 2.2](#)). Our axioms imply, $\mathcal{A} \vdash_2 \|YY^\top\|_F^2 = \langle J, YY^\top \rangle = \frac{n^2}{k}$, since the axioms $Y_{ic}^2 = Y_{ic}$ and $\langle Y_i, \mathbf{1}_k \rangle = 1$ imply that $\frac{n^2}{k} = \langle J, YY^\top \rangle = \|YY^\top\|_F^2$. Since $\tilde{\mathbf{E}}[a^2] \geq \tilde{\mathbf{E}}[a]^2$ is a degree-2 sum-of-squares inequality, $\|\tilde{\mathbf{E}}[YY^\top]\|_F^2 \leq \tilde{\mathbf{E}}\|YY^\top\|_F^2 = \frac{n^2}{k}$. Using this information with (3),

$$\left\langle \frac{\tilde{\mathbf{E}}[YY^\top]}{\|\tilde{\mathbf{E}}[YY^\top]\|_F}, \frac{XX^\top}{\|XX^\top\|_F} \right\rangle \geq 1 - \frac{k \lambda_{\max}(\Delta)}{n(p_{\text{in}} - p_{\text{out}})}.$$

So we get some bound on the error, in terms of $\lambda_{\max}(\Delta)$.

What is $\lambda_{\max}(\Delta)$? Δ is a block matrix with $k \times k$ blocks $\{\Delta_{a,b}\}_{a,b \in [k]}$ of size $\frac{n}{k} \times \frac{n}{k}$ each, where the blocks are indexed by pairs of communities. the entries of $\Delta_{a,b}$ are independent (up to symmetry, $\Delta_{a,b} = \Delta_{b,a}^\top$), each entry distributed like $\text{Ber}(p_{\text{in}}) - p_{\text{in}}$ if $a = b$ and like $\text{Ber}(p_{\text{out}}) - p_{\text{out}}$ if $a \neq b$. If the entries were all identically distributed as $\text{Ber}(p) - p$ with $p = \Omega(\frac{\log n}{n})$, we would have that $\|\Delta\| \leq 2\sqrt{p(1-p)n} = (1 + o(1)) \cdot 2\sqrt{\text{Var}(\text{Ber}(p))n}$ with high probability. To get a very loose bound, we can write Δ as a sum of its blocks and apply the triangle inequality for the operator norm and get that with high probability,

$$\lambda_{\max}(\Delta) \leq \|\Delta\| \leq \sum_{a,b \in [k]} \|\Delta_{a,b}\| \leq k \cdot \sqrt{p_{\text{in}}(1-p_{\text{in}})\frac{n}{k}} + k(k-1) \cdot \sqrt{p_{\text{out}}(1-p_{\text{out}})\frac{n}{k}} \leq \sqrt{p_{\text{in}}nk} + \sqrt{p_{\text{out}}nk^3}.$$

This bound is not tight, but it is sufficient to get us something nontrivial. Hence, parametrizing $p_{\text{in}} = \alpha \frac{\log n}{n}$ and $p_{\text{out}} = \beta \frac{\log n}{n}$, we have

$$\frac{k\lambda_{\max}(\Delta)}{n(p_{\text{in}} - p_{\text{out}})} \leq \frac{k}{n(p_{\text{in}} - p_{\text{out}})} \cdot (\sqrt{p_{\text{in}}nk} + \sqrt{p_{\text{out}}nk^3}) = \frac{\sqrt{k\alpha} + \sqrt{\beta k^3}}{(\alpha - \beta)} \cdot \sqrt{\frac{k}{\log n}}.$$

This completes the proof. \square

3 An information-computation gap in the sparse regime

The problem of *detection* in stochastic block models has an *information-computation gap* when $p_{\text{in}}, p_{\text{out}} = \Theta(\frac{1}{n})$. Let $p_{\text{in}} = \frac{a}{n}$ and $p_{\text{out}} = \frac{b}{n}$, and denote also the average degree $d_{\text{avg}} = \frac{1}{k}a + \frac{k-1}{k}b$. As is intuitively clear, the problem of hypothesis testing between $\mathcal{H}_0 = \mathcal{G}(n, p_{\text{avg}})$ (for $p_{\text{avg}} = \frac{1}{k}p_{\text{in}} + \frac{k-1}{k}p_{\text{out}}$) and $\mathcal{H}_1 = \mathcal{G}(n, k, p_{\text{in}}, p_{\text{out}})$ gets easier as $p_{\text{in}}, p_{\text{out}}$ (or a, b) get further apart.

We'll associate the following *signal-to-noise ratio* parameter with this problem:

$$s = \left(\frac{a-b}{k}\right)^2 \cdot \frac{1}{d_{\text{avg}}}.$$

This may look unnatural at first, but below we'll explain where this quantity comes from.

The parameter setting where $s = 1$ is called the *Kesten-Stigum threshold* (or KS-threshold). For all $k \geq 2$, the best detection algorithms known to date succeed if and only if $s > 1$, or “above the Kesten-Stigum threshold” [AS16a, BMR21].

In the special case $k = 2$, the KS threshold $\tau_{\text{KS}} = 1$ and the information-theoretic threshold τ_{info} below which (when $s < \tau_{\text{info}}$) $d_{\text{TV}}(\mathcal{H}_0, \mathcal{H}_1) = o(1)$ are known to coincide [MNS15, MNS18, Mas14]. However, the information theoretic threshold is known to scale with k as $\tau_{\text{info}} = \Theta(\frac{\log k}{k})$ [BMNN16]. Hence, for k large enough, for $\Theta(\frac{\log k}{k}) < s \leq 1$, the detection problem is solvable information-theoretically, but we don't know computationally efficient algorithms that do so.

3.1 Heuristic justification for the KS threshold

The signal-to-noise parameter s and the threshold $s = 1$ come from an information-theoretic lower bound for solving the related *root recovery* problem in a *broadcast process* on a random tree, generated in such a manner that this random tree can be coupled with a breadth-first-search starting at an arbitrary vertex in $G \sim \mathcal{G}(n, k, p_{\text{in}}, p_{\text{out}})$.

Breadth-first-search and/or a broadcast process on a tree. Imagine doing breadth-first search (BFS) up to depth t starting at some node $v \in V(G)$. We can couple this BFS with the sampling of G itself, for any constant t , as follows:

1. Start at the root, v .
2. Assign v a community label $\in [k]$ uniformly at random.
3. Push v onto a queue.
4. While the queue is non-empty, or until the vertex at the head of the queue is of distance t from the root v :
 - (a) Remove the leading vertex u from the queue.

- (b) Independently sample $\text{Pois}(\frac{1}{k}a)$ children of u with the same label c as u , and $\text{Pois}(\frac{1}{k}b)$ children with label c' for each $c' \in [k] \setminus \{c\}$.
- (c) Put each of these children into the queue.

One can show that for t constant, this process can be coupled with a breadth-first search in G which succeeds with high probability, by using the fact that $\text{Bin}(n, \frac{\lambda}{n})$ is well-approximated by $\text{Pois}(\lambda)$ as $n \rightarrow \infty$. This is a classic style of coupling, see for example [AS16b], Chapter 11.

Now, this can also be viewed as a *broadcast process* on a *Galton-Watson tree* of depth at most t . The idea is that each node samples $\text{Pois}(d_{\text{avg}})$ children independently, then “broadcasts” a label to each of its children, keeping its own label with probability $\frac{a}{d_{\text{avg}}}$ and choosing one of the $k - 1$ other labels uniformly at random with the remaining probability. By Poisson splitting, this is equivalent to the process described above.

Root reconstruction and a heuristic lower bound. A natural question in the context of a broadcast process on a tree is the following: given the labels at the leaves of the tree, can we get a reasonable estimate for the label of the root? The KS threshold arises as the threshold at which root reconstruction becomes information-theoretically possible [KS66, EKPS00]. Below we’ll prove one direction: we’ll show that when $s > 1$, root reconstruction is possible with a canonical estimator.

Theorem 3.1. *Let X_t be the number of vertices with label c at depth t in the tree (so $X_0 = 1[\text{root has label } c]$). Then*

$$\lim_{t \rightarrow \infty} \frac{\mathbf{E}[X_t \mid X_0 = 1] - \mathbf{E}[X_t \mid X_0 = 0]}{\sqrt{\text{Var}[X_t]}} = \frac{1}{k} s^{t/2}.$$

In particular if $s > 1$, one can use the value of X_t for t large enough to estimate whether $X_0 = 1$ (equivalently, if the root has label c) by checking if X_t is closer to $\mathbf{E}[X_t \mid X_0 = 1]$ or $\mathbf{E}[X_t \mid X_0 = 0]$, with the bound on the variance furnishing concentration via the second moment method. When $s \leq 1$, the difference in expectations is overwhelmed by the variance.

Before we dive into the proof, what does root reconstruction have to do with detection in block models? This has to do with the performance of the *Belief Propagation Algorithm* (BP), which precisely computes the marginal distribution of the root vertex in a broadcast process on a tree (given access to the leaf vertex labels, and to the structure of the tree). If we are in the regime $s > 1$ where root reconstruction is possible, then Belief Propagation also works for reconstructing the root, and in fact it achieves the statistically optimal estimation rate.

Though Belief Propagation is designed to work on trees, people can (and do) run it on graphs with cycles; it seems to work okay on sparse graphs, and in some cases one can even prove formally that some version of it works [KMM⁺13, AS16a, BLM15]. The reason $s > 1$ might suggest that BP works is that if you choose a random partition of the vertices of G to start, then patch it up using BP as though each vertex were the root of a depth- t tree (in sparse random graphs, most vertices’ depth- t neighborhoods are trees), you get nontrivial information about some fraction of the vertices’ labels. Alternatively if $s \leq 1$ and BP fails, then this could be taken as evidence that any polynomial-time algorithm will fail, if you believe that BP is optimal. This is discussed further in, for example, these sources [DKMZ11, M⁺17, LM21].

It is surprising that such imprecise heuristic reasoning seems to predict the limits of performance not only of BP algorithms, but also of dramatically different algorithms like SoS/SDPs [MS16, BMR21] and low-degree polynomial estimators [HS17, BBK⁺21]. This apparent universality across algorithms is a theme we’ll return to throughout the class.

Now, we prove that X_t is a reasonable estimator for whether $X_0 = 1$.

Proof of Theorem 3.1. Let X_t be the number of vertices of label c at depth t , and let Y_t be the number of vertices of label not c at depth t . By definition of our process,

$$\begin{aligned}\mathbf{E}[X_t | X_{t-1}, Y_{t-1}] &= \frac{1}{k}a \cdot X_{t-1} + \frac{1}{k}b \cdot Y_{t-1} \\ \mathbf{E}[Y_t | X_{t-1}, Y_{t-1}] &= \frac{k-1}{k}b \cdot X_{t-1} + \left(\frac{a}{b} + \frac{k-2}{k}b\right) \cdot Y_{t-1}.\end{aligned}$$

Encoding this linear-algebraically for convenience, we have that

$$\mathbf{E} \begin{bmatrix} X_t & Y_t \end{bmatrix} = \begin{bmatrix} X_{t-1} & Y_{t-1} \end{bmatrix} A, \quad \text{for} \quad A = \begin{bmatrix} \frac{1}{k}a & \frac{k-1}{k}b \\ \frac{1}{k}b & \frac{1}{k}a + \frac{k-2}{k}b \end{bmatrix}.$$

So A is like the "transition matrix" for the expected number of vertices of each type, with the first row/column corresponding to X and the second row/column corresponding to Y . One can check that A has eigenvalues $\frac{a-b}{k}$ and $d_{\text{avg}} = \frac{a+(k-1)b}{k}$, with corresponding left-eigenvectors $[1 \ -1]$ and $[1 \ k-1]$ respectively.

Now, notice that

$$\mathbf{E}[X_t | X_0 = 1] - \mathbf{E}[X_t | X_0 = 0] = e_1^\top A^t e_1 - e_2^\top A^t e_1 = (e_1 - e_2)^\top A^t e_1 = \left(\frac{a-b}{k}\right)^t,$$

because $e_1 - e_2$ is the left eigenvector of A with eigenvalue $(a-b)/k$.

We now compute $\text{Var}[X_t]$. This is made simple if we consider $\text{Var}[X_t | X_{t-1}, Y_{t-1}]$. Conditioned on X_{t-1}, Y_{t-1} , X_t is a sum of independent Poisson random variables, one for the number of children of community c for each node in the $t-1$ th level. By independence, the variance of the sum is simply the sum of the variances. Further, for $Q \sim \text{Pois}(\lambda)$, $\text{Var}[Q] = \lambda$. This implies that

$$\text{Var}[X_t | X_{t-1}, Y_{t-1}] = \frac{1}{k}aX_{t-1} + \frac{1}{k}bY_{t-1} = \mathbf{E}[X_t | X_{t-1}, Y_{t-1}],$$

And hence, we can compute

$$\text{Var}[X_t] = \mathbf{E}[X_t] = \begin{bmatrix} \frac{1}{k} & \frac{k-1}{k} \end{bmatrix} A^t e_1 = \frac{1}{k}d_{\text{avg}}^t,$$

where we have used that $\Pr[X_0 = 1] = \frac{1}{k}$, and that $[1 \ k-1]$ is the eigenvector of A corresponding to the eigenvalue d_{avg} .

Combining the above,

$$\lim_{t \rightarrow \infty} \frac{\mathbf{E}[X_t | X_0 = 1] - \mathbf{E}[X_t | X_0 = 0]}{\sqrt{\text{Var}[X_t]}} = \lim_{t \rightarrow \infty} \frac{1}{k} \left(\frac{a-b}{k\sqrt{d_{\text{avg}}}} \right)^t,$$

and the quantity on the right can be re-written as $\frac{1}{k}s^{t/2}$, as desired. \square

3.2 Weak recovery

In addition to the question of detection in the sparse regime, one can also ask about recovery. It's not hard to convince yourself that exact recovery is impossible, even when detection is possible: since $p_{\text{in}}, p_{\text{out}} = \Theta(\frac{1}{n})$, a constant fraction of the vertices will be isolated, and there is no hope in recovering their community membership. Still, you could hope to recover the partitions with accuracy better than a uniform random assignment. This algorithmic task is called *weak recovery*. A very recent paper [YP22] gives the

information-theoretically optimal error rate for weak recovery everywhere above the KS-threshold, for the special case $k = 2$.

An interesting open problem is whether this algorithm can be matched by SoS algorithms; one advantage of SoS algorithms and SDPs is that they typically enjoy strong robustness guarantees, see e.g. [DdNS22]. Robust SDP algorithms are known to achieve nontrivial weak recovery above the KS threshold, but it is open whether they achieve the optimal rates. The work of Banks, Mohanty, and Raghavendra [BMR21] presents an intriguing approach.

4 Conclusion

Bibliographic remarks. There is a vast body of literature on the stochastic block model, and I won't do it justice here; the surveys [Abb17, M⁺17] are nice resources. See also the specific reference pointers throughout the notes above, though there are many references that I have left out.

Contact. Comments are welcome at tselil@stanford.edu.

References

- [Abb17] Emmanuel Abbe. Community detection and stochastic block models: recent developments. *The Journal of Machine Learning Research*, 18(1):6446–6531, 2017. [8](#)
- [AS15] Emmanuel Abbe and Colin Sandon. 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*, pages 670–688. IEEE, 2015. [3](#)
- [AS16a] Emmanuel Abbe and Colin Sandon. Achieving the ks threshold in the general stochastic block model with linearized acyclic belief propagation. *Advances in Neural Information Processing Systems*, 29, 2016. [5](#), [6](#)
- [AS16b] Noga Alon and Joel H Spencer. *The probabilistic method*. John Wiley & Sons, 2016. [6](#)
- [Ban18] Afonso S Bandeira. Random laplacian matrices and convex relaxations. *Foundations of Computational Mathematics*, 18(2):345–379, 2018. [3](#)
- [BBK⁺21] Afonso S Bandeira, Jess Banks, Dmitriy Kunisky, Christopher Moore, and Alex Wein. Spectral planting and the hardness of refuting cuts, colorability, and communities in random graphs. In *Conference on Learning Theory*, pages 410–473. PMLR, 2021. [6](#)
- [BLM15] Charles Bordenave, Marc Lelarge, and Laurent Massoulié. Non-backtracking spectrum of random graphs: community detection and non-regular Ramanujan graphs. In *2015 IEEE 56th Annual Symposium on Foundations of Computer Science*, pages 1347–1357. IEEE, 2015. [6](#)
- [BMNN16] Jess Banks, Cristopher Moore, Joe Neeman, and Praneeth Netrapalli. Information-theoretic thresholds for community detection in sparse networks. In *Conference on Learning Theory*, pages 383–416. PMLR, 2016. [5](#)

- [BMR21] Jess Banks, Sidhanth Mohanty, and Prasad Raghavendra. Local statistics, semidefinite programming, and community detection. In *Proceedings of the 2021 ACM-SIAM Symposium on Discrete Algorithms (SODA)*, pages 1298–1316. SIAM, 2021. [5](#), [6](#), [8](#)
- [DdNS22] Jingqiu Ding, Tommaso d’Orsi, Rajai Nasser, and David Steurer. Robust recovery for stochastic block models. In *62nd Annual Symposium on Foundations of Computer Science (FOCS 2021)*, 2022. [8](#)
- [DKMZ11] Aurelien Decelle, Florent Krzakala, Cristopher Moore, and Lenka Zdeborová. Asymptotic analysis of the stochastic block model for modular networks and its algorithmic applications. *Physical Review E*, 84(6):066106, 2011. [6](#)
- [EKPS00] William Evans, Claire Kenyon, Yuval Peres, and Leonard J Schulman. Broadcasting on trees and the ising model. *Annals of Applied Probability*, pages 410–433, 2000. [6](#)
- [HS17] Samuel B Hopkins and David Steurer. Efficient bayesian estimation from few samples: community detection and related problems. In *2017 IEEE 58th Annual Symposium on Foundations of Computer Science (FOCS)*, pages 379–390. IEEE, 2017. [6](#)
- [HWX16] Bruce Hajek, Yihong Wu, and Jiaming Xu. Achieving exact cluster recovery threshold via semidefinite programming. *IEEE Transactions on Information Theory*, 62(5):2788–2797, 2016. [3](#)
- [KMM⁺13] Florent Krzakala, Cristopher Moore, Elchanan Mossel, Joe Neeman, Allan Sly, Lenka Zdeborová, and Pan Zhang. Spectral redemption in clustering sparse networks. *Proceedings of the National Academy of Sciences*, 110(52):20935–20940, 2013. [6](#)
- [KS66] Harry Kesten and Bernt P Stigum. Additional limit theorems for indecomposable multidimensional galton-watson processes. *The Annals of Mathematical Statistics*, 37(6):1463–1481, 1966. [6](#)
- [LM21] Siqi Liu and Sidhanth Mohanty. On statistical inference when fixed points of belief propagation are unstable. In *IEEE Symposium on Foundations of computer science*, 2021. [6](#)
- [M⁺17] Cristopher Moore et al. The computer science and physics of community detection: Landscapes, phase transitions, and hardness. *Bulletin of EATCS*, 1(121), 2017. [6](#), [8](#)
- [Mas14] Laurent Massoulié. Community detection thresholds and the weak Ramanujan property. In *Proceedings of the forty-sixth annual ACM symposium on Theory of computing*, pages 694–703, 2014. [5](#)
- [MNS15] Elchanan Mossel, Joe Neeman, and Allan Sly. Reconstruction and estimation in the planted partition model. *Probability Theory and Related Fields*, 162(3):431–461, 2015. [5](#)
- [MNS18] Elchanan Mossel, Joe Neeman, and Allan Sly. A proof of the block model threshold conjecture. *Combinatorica*, 38(3):665–708, 2018. [5](#)
- [MS16] Andrea Montanari and Subhabrata Sen. Semidefinite programs on sparse random graphs and their application to community detection. In *Proceedings of the forty-eighth annual ACM symposium on Theory of Computing*, pages 814–827, 2016. [6](#)

[YP22] Qian Yu and Yury Polyanskiy. Ising model on locally tree-like graphs: Uniqueness of solutions to cavity equations. 2022. [7](#)