Reconstruction in the Labeled Stochastic Block Model

Marc Lelarge and Laurent Massoulié and Jiaming Xu

***Abstract*—The labeled stochastic block model is a random graph model representing networks with community structure and interactions of multiple types. In its simplest form, it consists of two communities of approximately equal size, and the edges are drawn and labeled at random with probability depending on whether their two endpoints belong to the same community or not.**

It has been conjectured in [1] that this model exhibits a phase transition: reconstruction (i.e. identification of a partition positively correlated with the "true partition" into the underlying communities) would be feasible if and only if a model parameter exceeds a threshold.

We prove one half of this conjecture, i.e., reconstruction is impossible when below the threshold. In the converse direction, we introduce a suitably weighted graph. We show that when above the threshold by a specific constant, reconstruction is achieved by (1) minimum bisection, and (2) a spectral method combined with removal of nodes of high degree.

I. Introduction

A. Motivation

Community detection aims to identify underlying communities of similar characteristics in an overall population from the observation of pairwise interactions between individuals [2]. The stochastic block model, also known as *planted partition model*, is a popular random graph model for analyzing the community detection problem [3], [4], in which pairwise interactions are binary: an edge is either present or absent between two individuals. In its simplest form, the stochastic block model consists of two communities of approximately equal size, where the within-community edge is present at random with probability p; while the across-community edge is present with probability q. If $p > q$, it corresponds to assortative communities where interactions are more likely within rather than across communities; while $p < q$ corresponds to disassortative communities.

In practice, interactions can be of various types and these types reveal more information on the underlying communities than the mere existence of the interaction itself. For example, in recommender systems, interactions between users and items come with user ratings. Such ratings contain far more information than the interaction itself to characterize the user and item types. Similarly, protein-protein chemical interactions in biological networks can be exothermic and endothermic; email exchanges in a club may be formal or informal; friendship in social networks may be strong or weak. The labeled stochastic block model was recently proposed in [1] to capture rich interaction types. In this model interaction types are described by labels drawn from an arbitrary collection. In particular, for the simple two communities case, the within-community edge is labeled at random with distribution μ; while the across-community edge is labeled with a different distribution ν. In this context an important question is how to leverage the labeling information for detecting underlying communities.

B. Information-Scarce Regime

In this paper, we focus on the sparse labeled stochastic block model in which every node has a limited average degree, i.e., $p, q = O(1/n)$, where n is the number of nodes. It corresponds to the information-scarce regime where only $O(n)$ edges and labels are observed in total[1]. This regime is of practical interest, arising in several contexts. For example, in recommender systems, users only give ratings to few items; in biological networks, only few protein-protein interactions are observed due to cost constraints; in social networks, a person only has a limited number of friends.

For the stochastic block model in this information-scarce regime, there are $\Theta(n)$ isolated nodes, as in Erdős-Rényi random graphs with bounded average degree. For isolated nodes, it is impossible to determine their community membership and thus exact reconstruction of communities is impossible. Therefore, we resort to finding a partition into communities positively correlated to the true community partition (see Definition 1 below).

C. Main Results

Focusing on the two communities scenario, we show that a positively correlated reconstruction is fundamentally impossible when below a threshold. This establishes one half of the conjecture in [1]. In the positive direction, we establish the following results. We introduce a graph weighted by a suitable function of observed labels, on which we show that:

(1) Minimum bisection gives a positively correlated partition when above the threshold by a factor of $32 \ln 2$.

(2) A spectral method combined with removal of nodes of high degree gives a positively correlated partition when above the threshold by a constant factor.

Due to space constraints, we only provide proof sketches in the Appendix; detailed proofs are available in [5].

M. Lelarge is with INRIA-ENS. J. Xu is with University of Illinois at Urbana-Champaign - work performed while an intern with Technicolor. L. Massoulié is with Microsoft Research-Inria Joint Centre. Last modified: July 24, 2013.

[1] We also provide results for $p, q = O(\text{polylog}(n)/n)$ in Theorem 3.

D. Related Work

For the stochastic block model, most previous work focuses on the "dense" regime with an average degree diverging as the size of the graph n grows and "exact" reconstruction, such as [6], [7], [8]. McSherry [9] showed that the spectral method works as long as $p-q \geq \Omega(\sqrt{p \log n/n})$ with average degree as low as $\Omega(\log^6 n)$. Massoulié and Tomozei [10] reduced this lower bound on the average degree to $\Omega(\log n)$. More recently, it was shown in [11] that a matrix completion approach works when $p - q \geq \Omega(r\sqrt{p/n}\log^2 n)$ where the number of communities r could scale with n.

For the "sparse" regime with bounded average degrees, a sharp phase transition threshold for reconstruction (as defined in Definition 1) is conjectured in [4] by analyzing the belief propagation algorithm. The converse part of the conjecture was rigorously proved in [12]. In the converse direction, it is shown in [13] that a variant of spectral method gives a positively correlated partition when above the threshold by an unknown constant factor.

The labeled stochastic block was first proposed and studied in [1] and a new reconstruction threshold that incorporates the extra labeling information was conjectured. Simulations further indicate that the belief propagation algorithm works when above the threshold, but reconstruction algorithms that provably work are still unknown.

II. Model and Notation

This section formally defines the labeled stochastic block model with two symmetric communities and introduces the key notations and definitions used in the paper.

The labeled stochastic block model $\mathcal{G}(n, p, q, \mu, \nu)$ is a random graph with n nodes of $\{\pm 1\}$ types and $\{\ell \in \mathcal{L}\}$-labeled edges, where $\mathcal{L}$ is a finite set of labels. To generate a particular realization (G, L, σ), first assign type $\sigma_u \in \{\pm 1\}$ to each node u uniformly and independently at random. Then, for every node pair (u, v), independently of everything else, draw an edge between u and v with probability p if $\sigma_u = \sigma_v$ and with probability q otherwise. Finally, every edge $e = (u, v)$ is labeled as $\ell \in \mathcal{L}$ independently at random with probability $\mu(\ell)$ if $\sigma_u = \sigma_v$ and with probability $\nu(\ell)$ otherwise.

When $\mu = \nu$, it reduces to the classical stochastic block model without labels. This paper focuses on the sparse case where $p = a/n$ and $q = b/n$ for two fixed constants a and b, and the goal is to reconstruct the true underlying types of nodes σ by observing the graph structure G and the labels on edges L.

It is known that in the sparse graph, there are $\Theta(n)$ isolated nodes whose types clearly cannot be recovered accurately. Therefore, our goal is to reconstruct a type assignment which is positively correlated to the true type assignment. More formally, we adopt the following definition.

Definition 1. *A type assignment $\hat{\sigma}$ is said to be correlated with the true type assignment σ if a.a.s.*

$$Q(\sigma, \hat{\sigma}) := \frac{1}{2} - \frac{1}{n} \min\{d(\sigma, \hat{\sigma}), d(\sigma, -\hat{\sigma})\} > 0, \tag{1}$$

where d is the Hamming distance. Q is called the Overlap.

The shorthand a.a.s. denotes *asymptotically almost surely*. A sequence of events A_n holds a.a.s. if the probability of A_n converges to 1 as $n \to \infty$.

Define τ as

$$\tau = \frac{a+b}{2} \sum_{\ell \in \mathcal{L}} \frac{a\mu(\ell) + b\nu(\ell)}{a+b} \left(\frac{a\mu(\ell) - b\nu(\ell)}{a\mu(\ell) + b\nu(\ell)} \right)^2. \tag{2}$$

It was conjectured in [1] that τ is the threshold for positively correlated reconstruction.

Conjecture 1. (i) *If $\tau > 1$, then it is possible to find a type assignment correlated with the true assignement a.a.s.*

(ii) *If $\tau < 1$, then it is not possible to find a type assignment correlated with the true assignement a.a.s.*

In this paper, we prove (ii) and propose a simple spectral algorithm able to find a type assignment correlated with the true assignment for τ big enough.

III. Minimum Bisection

To recover the community partition, a natural way is via maximum likelihood estimation, where the log-likelihood can be written as:

$$\log \mathbb{P}(G, L|\sigma) \propto \sum_{(u,v) \in E(G)} \log \frac{a\mu(L_{uv})}{b\nu(L_{uv})} \sigma_u \sigma_v + \log \left(\frac{1 - a/n}{1 - b/n} \right) \sum_{(u,v) \notin E(G)} \sigma_u \sigma_v$$

Using the constraint $\sum_u \sigma_u = 0$, maximum log likelihood estimation reduces to

$$\max_{\sigma} \sum_{(u,v) \in E(G)} \left(\log \frac{a\mu(L_{uv})}{b\nu(L_{uv})} + \log \frac{1 - a/n}{1 - b/n} \right) \sigma_u \sigma_v$$
$$\text{s.t.} \sum_u \sigma_u = 0, \ \sigma_u \in \{\pm 1\},$$

If we ignore the $o(n)$ term in the sum, this is equivalent to the minimum bisection on the weighted graph with a specific weight function $w(\ell) = \log \frac{a\mu(\ell)}{b\nu(\ell)}$. For a general weighing function $w : \mathcal{L} \to \mathbb{R}$, the minimum bisection finds the balanced bipartite subgraph in G with the minimum weighted cut, i.e.,

$$\min_{\sigma} \sum_{(u,v): \sigma_u \neq \sigma_v} W_{uv}$$
$$\text{s.t.} \sum_u \sigma_u = 0, \ \sigma_u \in \{\pm 1\}, \tag{3}$$

where $W_{uv} = A_{uv} w(L_{uv})$ and A is the adjacency matrix of G.

Theorem 1. *Assume the technical condition: $\sum_\ell a\mu(\ell) w^2(\ell), \sum_\ell b\nu(\ell) w^2(\ell) > 8 \ln 2$. Then if*

$$\frac{[\sum_\ell (a\mu(\ell) - b\nu(\ell)) w(\ell)]^2}{\sum_\ell (a\mu(\ell) + b\nu(\ell)) w^2(\ell)} > 64 \ln 2, \tag{4}$$

solutions of the minimum bisection (3) are positively correlated to the true type assignment $\sigma^\star$ a.a.s. Moreover, the left hand side of (4) is maximized when $w(\ell) = \frac{a\mu(\ell)-b\nu(\ell)}{a\mu(\ell)+b\nu(\ell)}$, in which case (4) reduces to $\tau > 32\ln 2$.

IV. Spectral Method

The minimum bisection is NP-hard [14]. In this section, we present a simple spectral algorithm based on the matrix W introduced above (see [15] for a similar approach in the unlabeled case). We show that this algorithm finds an assignment correlated with the true assignment provided τ is large enough.

First note that we have $\mathbb{E}[W|\sigma] = \frac{\alpha}{n}\mathbf{1}\mathbf{1}^\top + \frac{\beta}{n}\sigma\sigma^\top - \frac{\alpha+\beta}{n}\mathbf{I}$ with

$$\alpha = \frac{1}{2}\sum_\ell w(\ell)(a\mu(\ell) + b\nu(\ell)),$$

$$\beta = \frac{1}{2}\sum_\ell w(\ell)(a\mu(\ell) - b\nu(\ell)). \tag{5}$$

The term $\frac{\alpha+\beta}{n}\mathbf{I}$ is irrelevant to the main results (thanks to Weyl's perturbation theorem) and neglected for simplicity. Let $D = W - \frac{\alpha}{n}\mathbf{1}\mathbf{1}^\top$ and then $\mathbb{E}[D|\sigma] = \frac{\beta}{n}\sigma\sigma^\top$ has rank one with singular value β. Hence, it makes sense to define $\hat{D}$ as the best rank-1 approximation of the matrix D. In other words, if $D = \sum_i v_i x_i y_i^\top$ is the singular value decomposition of D with singular values $v_1 \geq v_2 \geq \dots$, we define $\hat{D} = v_1 x_1 y_1^\top$. Then if the matrix D is "close" to its mean $\mathbb{E}[D|\sigma]$, we expect v_1 to be "close" to β and $\text{sign}(x_1)$ to be correlated with σ, where $\text{sign}(x)$ gives the sign of x componentwise. Note that D is very similar to the modularity matrix defined in [16] and thus dividing the vertices into two communities according to $\text{sign}(x_1)$ can be seen as an algorithm to maximize the modularity. Unfortunately, in the sparse stochastic block model, there are vertices of degree $\Omega(\frac{\log n}{\log\log n})$ and thus the largest singular value of W could reach $\Omega(\sqrt{\frac{\log n}{\log\log n}})$ which is much higher than β.

In order to take care of this issue, we begin with a preliminary step to "clean" the spectrum of W: we remove all vertices in the graph with degree larger than $\frac{3}{2}\frac{a+b}{2}$. To summarize, for a given weight function $w(\ell)$, our algorithm $\mathrm{Spectral-Reconstruction}$ has the following structure:

1) Remove vertices with degree larger than $\frac{3}{2}\frac{a+b}{2}$ and assign a random type to these vertices.
2) Define W' by setting to zero the rows and columns of W corresponding to vertices removed.
3) Let $\hat{x}$ be the left-singular vector associated with the largest singular value of $D' = W' - \frac{\alpha}{n}\mathbf{1}\mathbf{1}^\top$. Output $\text{sign}(\hat{x})$ for the types of the remaining vertices.

Observe that step 3) can be seen as a relaxation of the minimum bisection (3) by replacing the integer constraint with the unit-norm constraint and relaxing the hard constraint $\sum_u \sigma_u = 0$ to be a regularized term in the objective function. $\mathrm{Spectral-Reconstruction}$ needs estimates of α and $a+b$, which can be well approximated by $\frac{1}{n}\mathbf{1}^\top W\mathbf{1}$ and $\frac{2}{n}\mathbf{1}^\top A\mathbf{1}$

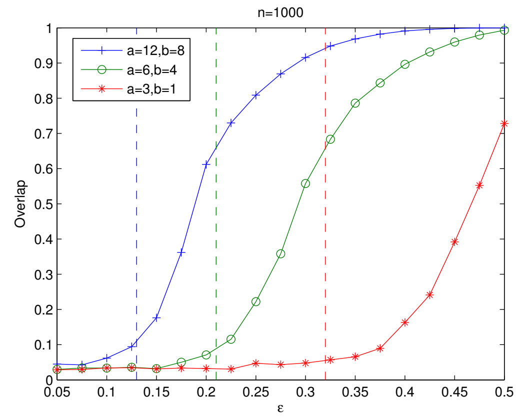

Figure 1. The overlap Q against ϵ from 0.05 to 0.5.

respectively. To simplify the analysis, we will assume that the exact value of α and $a+b$ is known.

Theorem 2. *Assume $|w(\ell)| \leq 1$ for all ℓ. There exists a universal constant C (i.e. not depending on a, b, μ or ν) such that if $\beta > C\sqrt{a+b}$, where β is defined in (5) then* $\mathrm{Spectral-Reconstruction}$ *outputs a type assignment correlated with the true assignment a.a.s. In the particular case, where $w(\ell) = \frac{a\mu(\ell)-b\nu(\ell)}{a\mu(\ell)+b\nu(\ell)}$, the condition $\beta > C\sqrt{a+b}$ reduces to $\tau > C'\sqrt{a+b}$.*

Compared to point (i) in the Conjecture 1, our result does not give the right order of magnitude when a and b are large. Indeed, we are able to improve it if we allow a and b to grow with n.

Theorem 3. *Assume that $\min(a,b) = \Omega(\log^6 n)$. If*

$$\frac{[\sum_\ell (a\mu(\ell) - b\nu(\ell))w(\ell)]^2}{\sum_\ell (a\mu(\ell) + b\nu(\ell))w^2(\ell)} > 128, \tag{6}$$

then $\mathrm{Spectral-Reconstruction}$ *outputs a type assignment correlated with the true assignment a.a.s. Moreover, the left hand side of (6) is maximized when $w(\ell) = \frac{a\mu(\ell)-b\nu(\ell)}{a\mu(\ell)+b\nu(\ell)}$, in which case (6) reduces to $\tau > 64$. With this choice of $w(\ell)$, as soon as $\tau \to \infty$,* $\mathrm{Spectral-Reconstruction}$ *outputs the true assignment for all vertices except $o(n)$ a.a.s.*

Note that in the regime $\min(a,b) = \Omega(\log^6 n)$, the degrees are very concentrated and step 1) of the algorithm can be removed without harm. The simulation results, depicted in Fig. 1, further indicate that $\mathrm{Spectral-Reconstruction}$ leaving out step 1) outputs a positively correlated assignment when above the threshold. In the simulation, we assume for simplicity only two labels: r and b, and define $\mu(r) = 0.5+\epsilon$ and $\nu(r) = 0.5-\epsilon$. We generate the graph from the labeled stochastic block model with $n = 1000$ nodes for various a, b, ϵ. Fix a, b, we plot the overlap Q against ϵ and indicate the threshold $\tau = 1$ as a vertical dash line. All plotted values are averages over 100 trials.

Note that our algorithm is most efficient when the parameters (a, b, μ and ν) of the model are known as the optimal weight function depends on these parameters. In the case where the labels are uninformative, i.e. $\mu = \nu$, our algorithm is very simple, does not require to know the values a and b and is (in the range of Theorem 3) the algorithm with the best performance guarantees (see [11]).

V. Converse Result

This section proves part (ii) of Conjecture 1. In particular, we show that when $\tau < 1$, asymptotically it is impossible to tell whether any two nodes are more likely to belong to the same community. Since reconstructing a positively correlated type assignment is harder than telling whether any two nodes are more likely to belong to the same community, it further implies that reconstructing a positively correlated type assignment is fundamentally impossible.

Theorem 4. *If $\tau < 1$, then for any fixed vertices ρ and v,*

$$\mathbb{P}_n(\sigma_\rho = +1 | G, L, \sigma_v = +1) \to 1/2 \text{ a.a.s.} \tag{7}$$

Theorem 4 is related to the Ising spin model in the statistical physics [17], [18], and it essentially says that there is no long range correlation in the type assignment when $\tau < 1$. The main idea in the proof of Theorem 4 is borrowed from [12] and works as follows: (1) pick any two fixed vertices ρ, v and consider the local neighborhood of ρ up to distance $O(\log(n))$. The vertex v lies outside of the local neighborhood of ρ a.a.s.. (2) conditional on the type assignment at the boundary of the local neighborhood, σ_ρ is asymptotically independent with σ_v. (3) the local neighborhood of ρ looks like a Markov process on a labeled Galton-Watson tree rooted at ρ. (4) For the Markov process on the labeled Galton-Watson tree, the types of leaves provide no information about the type of the root ρ when the depth of tree goes to infinity.

VI. Conclusion

Our results show that when $\tau < 1$ it is fundamentally impossible to give a positively correlated reconstruction; when $\tau > 1$, the labeling information can be effectively exploited through the suitably weighted graph. An interesting future work is to prove the positive part of Conjecture 1.

VII. Acknowledgement

J. X. would like to thank Y. Chen and B. Hajek for helpful conversations. M. L. acknowledges the support of the French Agence Nationale de la Recherche (ANR) under reference ANR-11-JS02-005-01 (GAP project). J. X. acknowledges the support of the National Science Foundation under Grant ECCS 10-28464.

Appendix

A. Proof of Theorem 1

Let $m(\sigma) = |\{u : \sigma_u = +1, \sigma^\star_u = -1\}|$ and $\epsilon > 0$ be an arbitrarily small constant. To prove the theorem, by the definition of positively correlated reconstruction, it suffices to show that for all σ with $\frac{n}{4}(1-\epsilon) \le m(\sigma) \le \frac{n}{4}$,

$$\sum_{\substack{(u,v):\sigma_u \neq \sigma_v, \\ \sigma^\star_u = \sigma^\star_v}} W_{uv} - \sum_{\substack{(u,v):\sigma_u = \sigma_v, \\ \sigma^\star_u \neq \sigma^\star_v}} W_{uv} := Y_1(\sigma) - Y_2(\sigma) > 0.$$

Observe that $Y_1(\sigma)$ is a sum of $2m(n/2 - m)$ i.i.d. random variables whose value is $w(\ell)$ with probability $\frac{a}{n}\mu(\ell)$; $Y_2(\sigma)$ is a sum of $2m(n/2-m)$ i.i.d. random variables whose value is $w(\ell)$ with probability $\frac{b}{n}\nu(\ell)$. Using the Chernoff bound and an union bound, we can show that a.a.s. for all σ with $\frac{n}{4}(1-\epsilon) \le m(\sigma) \le \frac{n}{4}$, $Y_1(\sigma) - Y_2(\sigma) > 0$, if (4) is satisfied. By Cauchy-Schwartz inequality,

$$\left(\sum_\ell (a\mu(\ell) - b\nu(\ell)) w(\ell) \right)^2 \le 2\tau \sum_\ell (a\mu(\ell) + b\nu(\ell)) w^2(\ell)$$

with equality achieved when $w(\ell) = \frac{a\mu(\ell) - b\nu(\ell)}{a\mu(\ell) + b\nu(\ell)}$. This completes the proof.

B. Proof of Theorem 2

Let C be a universal constant whose numerical value may change in the proof. Let $\|X\|$ and $\|X\|_F$ denote the spectral norm and Frobenius norm of a matrix X respectively.

A simple calculation shows:

Lemma 1. *For any $x \in \{-1,+1\}^n$ and $y \in \mathbb{R}^n$ with $\|y\| = 1$, we have*

$$d(x, \text{sign}(y)) \le n \| \frac{x}{\sqrt{n}} - y \|^2.$$

Applying this lemma with σ and $\hat{x}$ gives:

$$\frac{1}{n} d(\sigma, \pm\text{sign}(\hat{x})) \le \| \frac{\sigma}{\sqrt{n}} \pm \hat{x} \|^2.$$

Lemma 2. *Let $M = \alpha x x^\top$ and $M' = \beta y y^\top$, with $\alpha, \beta \in \mathbb{R}$, $\|x\| = \|y\| = 1$ and $x^\top y \ge 0$. Then we have*

$$\|x - y\| \le \frac{\sqrt{2}}{\max\{|\alpha|, |\beta|\}} \|M - M'\|.$$

Define $\hat{D}'$ as the best rank-1 approximation of the matrix D'. Applying this lemma with $\hat{D}'$ and $\mathbb{E}[D|\sigma]$ gives:

$$\min\{\| \frac{\sigma}{\sqrt{n}} - \hat{x} \|, \| \frac{\sigma}{\sqrt{n}} + \hat{x} \|\} \le \frac{\sqrt{2}}{\beta} \|\hat{D}' - \mathbb{E}[D|\sigma]\|.$$

Hence, we have

$$\frac{1}{n} \min\{d(\sigma, \text{sign}(\hat{x}), d(\sigma, -\text{sign}(\hat{x})\} \le \frac{2}{\beta^2} \|\hat{D}' - \mathbb{E}[D|\sigma]\|^2. \tag{8}$$

Also, a simple derivation yields

$$\|\hat{D}' - \mathbb{E}[D|\sigma]\| \le 2\|W' - \mathbb{E}[W|\sigma]\|. \tag{9}$$

The following lemma bounds the quantity $\|W' - \mathbb{E}[W|\sigma]\|$. Its proof is similar to the proof of Lemma 3.2 in [19].

Lemma 3. *For any $\sigma \in \{\pm 1\}^n$, there exists an universal constant C such that*

$$\|W' - \mathbb{E}[W|\sigma]\| \leq C\sqrt{a+b}, \quad a.a.s. \qquad (10)$$

Combining this lemma with (8) and (9), we get

$$\frac{1}{n}\min\{d(\sigma, \text{sign}(\hat{x}), d(\sigma, -\text{sign}(\hat{x})\} \leq C\frac{a+b}{\beta^2},$$

and the theorem follows.

C. Proof of Theorem 3

The proof follows the same steps as for Theorem 2, except that we are able to strengthen Lemma 3 thanks to a result of Vu [20]. Note that the variance of the elements of W is upper bounded by $\frac{1}{n}\sum_\ell w^2(\ell)\left(a\mu(\ell) + b\nu(\ell)\right)$ so that by Theorem 1.4 in [20], we get

Lemma 4. *Under the conditions of Theorem 3, we have*

$$\|W - \mathbb{E}[W|\sigma]\| \leq 2\sqrt{\sum_\ell w^2(\ell)\left(a\mu(\ell) + b\nu(\ell)\right)} \quad a.a.s.$$

D. Proof of Theorem 4

Consider a Galton-Watson tree T with Poisson offspring distribution with mean $\frac{a+b}{2}$. The type of the root ρ is chosen from $\{\pm 1\}$ uniformly at random. Each child has the same type as its parent with probability $\frac{a}{a+b}$ and different type otherwise. Every edge (u, v) is labeled at random with distribution μ if $\sigma_u = \sigma_v$ and ν otherwise. Let T_R denote the Galton-Watson tree T up to depth R and ∂T_R denote the set of leaves of T_R. Let G_R denote the induced subgraph of G up to distance R from ρ and ∂G_R be the set of nodes at distance R from ρ.

The following lemma similar to Proposition 4.2 in [12] establishes a coupling between the local neighborhood of ρ and the labeled Galton-Watson tree rooted at ρ.

Lemma 5. *Let $R = R(n) = \lfloor \frac{\log n}{10\log(2(a+b))} \rfloor$, then there exists a coupling such that a.a.s.*

$$(G_R, L_{G_R}, \sigma_{G_R}) = (T_R, L_{T_R}, \sigma_{T_R}),$$

where L_{G_R} and σ_{G_R} denote the labels and types on the subgraph G_R respectively.

To ease notation, we omit the shorthand a.a.s. in the sequel. To prove Theorem 4, it suffices to show that $\text{Var}(\sigma_\rho|G, L, \sigma_v) \to 1$. By the monotonicity property of conditional variance, it further reduces to show that $\text{Var}(\sigma_\rho|G, L, \sigma_v, \sigma_{\partial G_R}) \to 1$. Let R be as in Lemma 5, then $G_R = o(\sqrt{n})$ and thus $v \notin G_R$. Lemma 4.7 in [12] shows that σ_ρ is asymptotically independent with σ_v conditionally on $\sigma_{\partial G_R}$. Hence,

$$\text{Var}(\sigma_\rho|G, L, \sigma_v, \sigma_{\partial G_R}) \to \text{Var}(\sigma_\rho|G, L, \sigma_{\partial G_R}).$$

Also, note that

$$\text{Var}(\sigma_\rho|G, L, \sigma_{\partial G_R}) = \text{Var}(\sigma_\rho|G_R, L_{G_R}, \sigma_{\partial G_R}).$$

Lemma 5 implies that

$$\text{Var}(\sigma_\rho|G_R, L_{G_R}, \sigma_{\partial G_R}) \to \text{Var}(\sigma_\rho|T_R, L_{T_R}, \sigma_{\partial T_R}).$$

For labeled Galton-Watson tree, it was shown in [1] that if $\tau < 1$, the types of the leaves provide no information about the type of root when the depth $R \to \infty$, i.e.,

$$\mathbb{P}(\sigma_\rho = +1|T, L, \sigma_{\partial T_R}) \to \frac{1}{2}.$$

Hence, $\text{Var}(\sigma_\rho|T_R, L_{T_R}, \sigma_{\partial T_R}) \to 1$ and the theorem follows.

References

[1] S. Heimlicher, M. Lelarge, and L. Massoulié, "Community detection in the labelled stochastic block model," Nov. 2012, available at: http://arxiv.org/abs/1209.2910.

[2] S. Fortunato, "Community detection in graphs," Jan. 2010, available at: http://arxiv.org/abs/0906.0612.

[3] K. Rohe, S. Chatterjee, and B. Yu, "Spectral clustering and the high-dimensional stochastic blockmodel," *The Annals of Statistics*, vol. 39, no. 4, pp. 1878–1915, 2011.

[4] A. Decelle, F. Krzakala, C. Moore, and L. Zdeborova, "Asymptotic analysis of the stochastic block model for modular networks and its algorithmic applications," *Physics Review E*, vol. 84:066106, 2011.

[5] M. Lelarge, L. Massoulié, and J. Xu, "Reconstruction in the labelled stochastic block model," May. 2013, Preprint.

[6] M. Dyer and A. Frieze, "The solution of some random NP-hard problems in polynomial expected time," *Journal of Algorithms*, vol. 10, no. 4, pp. 451 – 489, 1989.

[7] M. Jerrum and G. B. Sorkin, "The metropolis algorithm for graph bisection," *Discrete Applied Mathematics*, vol. 82, no. 1-3, pp. 155 – 175, 1998.

[8] A. Condon and R. M. Karp, "Algorithms for graph partitioning on the planted partition model," *Random Struct. Algorithms*, vol. 18, no. 2, pp. 116–140, Mar 2001.

[9] F. McSherry, "Spectral partitioning of random graphs," in *42nd IEEE Symposium on Foundations of Computer Science*, Oct. 2001, pp. 529 – 537.

[10] L. Massoulié and D.-C. Tomozei, "Distributed user profiling via spectral methods," Sep. 2011, available at: http://arxiv.org/abs/1109.3318.

[11] Y. Chen, S. Sanghavi, and H. Xu, "Clustering sparse graphs," Oct. 2012, available at: http://arxiv.org/abs/1210.3335.

[12] E. Mossel, J. Neeman, and A. Sly, "Stochastic block models and reconstruction," Feb. 2012, available at: http://arxiv.org/abs/1202.1499.

[13] A. Coja-oghlan, "Graph partitioning via adaptive spectral techniques," *Comb. Probab. Comput.*, vol. 19, no. 2, 2010.

[14] M. R. Garey, D. S. Johnson, and L. Stockmeyer, "Some simplified NP-complete graph problems," *Theoret. Comput. Sci.*, vol. 1, no. 3, pp. 237–267, 1976.

[15] A. Coja-Oghlan, "A spectral heuristic for bisecting random graphs," *Random Structures Algorithms*, vol. 29, no. 3, pp. 351–398, 2006. [Online]. Available: http://dx.doi.org/10.1002/rsa.20116

[16] M. E. J. Newman, "Modularity and community structure in networks," *Proceedings of the National Academy of Sciences*, vol. 103, no. 23, pp. 8577–8582, 2006.

[17] W. Evans, C. Kenyon, Y. Peres, and L. J. Schulman, "Broadcasting on trees and the ising model," *The Annals of Applied Probability*, vol. 10, no. 2, pp. 410–433, 2000.

[18] E. Mossel, "Survey - information flows on trees," *DIMACS series in discrete mathematics and theoretical computer science*, pp. 155–170, 2004.

[19] R. Keshavan, A. Montanari, and S. Oh, "Matrix completion from a few entries," *IEEE Transactions on Information Theory*, vol. 56, no. 6, pp. 2980 –2998, June 2010.

[20] V. H. Vu, "Spectral norm of random matrices," *Combinatorica*, vol. 27, no. 6, pp. 721–736, 2007. [Online]. Available: http://dx.doi.org/10.1007/s00493-007-2190-z