

Stochastic Block Models

1 Block Models

Definition 1 (Vanilla SBM). *Let n be a positive integer (the number of vertices). The pair (X, G) is drawn under $VSBM(n, q_{in}, q_{out})$ if X is an n -dimensional random vector with i.i.d. components distributed under $Rad(1/2)$ in the Bernoulli model, and X is drawn uniformly at random with the constraint $1^T X = 0$, in the uniform or strictly balanced model, and G is an n -vertex simple graph where vertices i and j are connected with probability p if $X_i = X_j$ and with probability q if $X_i \neq X_j$, independently of other pairs of vertices. We also define the community sets by $\Omega_i = \Omega_i(X) := \{v \in [n] : X_v = i\}, i \in \{-1, 1\}$. Thus the distribution of G given X , where $G = ([n], E(G))$ is defined as follows; for $x \in \{-1, 1\}^n$ and $y \in \{0, 1\}^{\binom{n}{2}}$,*

$$\mathbb{P}(G = g | X = x) = q_{in}^{E_{in}} (1 - q_{in})^{E_{in}^C} q_{out}^{E_{out}} (1 - q_{out})^{E_{out}^C},$$

where

$$\begin{aligned} E_{in} &= \sum_{\substack{u < v \\ X_u = X_v}} \mathbb{1}(y_{uv} = 1) \\ E_{in}^C &= \binom{|\Omega_1|}{2} + \binom{|\Omega_{-1}|}{2} - E_{in} \\ E_{out} &= \sum_{\substack{u < v \\ X_u \neq X_v}} \mathbb{1}(y_{uv} = 1) \\ E_{out}^C &= |\Omega_1| |\Omega_{-1}| - E_{out}. \end{aligned}$$

Recall (Rademacher distribution) : $Y \sim Rad(\varepsilon)$ if $\mathbb{P}(Y = 1) = \varepsilon$ and $\mathbb{P}(Y = -1) = 1 - \varepsilon$.

Definition 2 (General SBM). *Let n be a positive integer (the number of vertices), k be a positive integer (the number of communities), $p = (p_1, \dots, p_k)$ be a probability vector on $[k] := \{1, \dots, k\}$ (the prior on the k communities) and W be a $k \times k$ symmetric matrix with entries in $[0, 1]$ (the connectivity probabilities). The pair (X, G) is drawn under $SBM(n, p, W)$ if X is an n -dimensional random vector with i.i.d. components distributed under p , and G is an n -vertex simple graph where vertices i and j are connected with probability W_{X_i, X_j} , independently of other pairs of vertices. We also define the community sets by $\Omega_i = \Omega_i(X) := \{v \in [n] : X_v = i\}, i \in [k]$. Thus the distribution of G given X , where $G = ([n], E(G))$ is defined as follows; for $x \in [k]^n$ and $y \in \{0, 1\}^{\binom{n}{2}}$*

$$\mathbb{P}(G = g | X = x) = \prod_{1 \leq u, v \leq n} W_{X_u, X_v}^{y_{uv}} (1 - W_{X_u, X_v})^{1 - y_{uv}}.$$

Remark 1. If $W = p\mathbb{1}$, where $\mathbb{1}$ is a $k \times k$ matrix such that all entries are 1, then SBM is equivalent to the Erdős-Rényi model.

Remark 2. For Vanilla SBM, $W = \begin{pmatrix} q_{in} & q_{out} \\ q_{out} & q_{in} \end{pmatrix}$

2 Some topology of SBM

2.1 Probability of an edge

$$\begin{aligned} \mathbb{P}(E_{1,2} = 1) &= \sum_{x_1, x_2} \mathbb{P}(E_{1,2} = 1 | X_1 = x_1, X_2 = x_2) p(x_1) p(x_2) \\ &= \sum_{x_1, x_2} W_{x_1, x_2} p(x_1) p(x_2) = p^T W p. \end{aligned}$$

For VSBM, $\mathbb{P}(E_{1,2} = 1) = \frac{q_{in} + q_{out}}{2}$.

2.2 Expected degree

$$\begin{aligned} \mathbb{E}[deg(1)] &= \sum_{v=2}^n \mathbb{P}(E_{1,v} = 1) = \sum_{v=2}^n \sum_{j=1}^k \mathbb{P}(E_{1,v} = 1 | X_v = j) p_j \\ &= (n-1) \sum_{j=1}^k W_{X_1, j} p_j \sim n(Wp)_{X_1} \end{aligned}$$

For VSBM, $\mathbb{E}[deg(1)] \sim \frac{n(q_{in} + q_{out})}{2}$.

2.3 Community of a neighbour

$$\begin{aligned} \mathbb{P}(X_2 = x_2 | E_{1,2} = 1, X_1 = x_1) &= \frac{\mathbb{P}(E_{1,2} = 1 | X_1 = x_1, X_2 = x_2) p(x_2)}{\sum_{x_2} \mathbb{P}(E_{1,2} = 1 | X_1 = x_1, X_2 = x_2) p(x_2)} \\ &= \frac{W_{x_1, x_2} p(x_2)}{\sum_{x_2} W_{x_1, x_2} p(x_2)}. \end{aligned}$$

Define $M := WP$, where $P = diag(p)$ (the square diagonal matrix with the vector p on the diagonal). Let $\hat{M}$ be M with normalized rows, i.e. $\hat{M} = diag(M\mathbb{1})^{-1}M$. Then,

$$\mathbb{P}(X_2 = x_2 | E_{1,2} = 1, X_1 = x_1) = \hat{M}(x_1, x_2),$$

that is the probability that 2 is in community x_2 , given that one of his neighbors is in community x_1 .

$$\text{For VSBM, } \mathbb{P}(X_2 = x_2 | E_{1,2} = 1, X_1 = x_1) = \begin{cases} \frac{q_{in}}{q_{in} + q_{out}} & \text{if } x_1 = x_2 \\ \frac{q_{out}}{q_{in} + q_{out}} & \text{if } x_1 \neq x_2 \end{cases}$$

2.4 Cycles

Let us fix m vertices in $[n]$, and let us compute the probability that these m vertices form a cycle of length m

$$\begin{aligned}\mathbb{P}(m\text{-cycle}) &= \sum_{x_1, \dots, x_m} W_{x_1, x_2} W_{x_2, x_3} \dots W_{x_m, x_1} p(x_1) p(x_2) \dots p(x_m) \\ &= \sum_{x_1, \dots, x_m} \prod_{i=1}^m W_{x_i, x_{i+1}} p(x_{i+1}) = \text{Tr}((WP)^m),\end{aligned}$$

where we assumed that $m+1=1$ and where $P = \text{diag}(p)$ (the square diagonal matrix with the vector p on the diagonal).

If $W = \frac{Q}{n}$, where Q is a matrix that does not scale with n , then the expected number of m -cycles is given by

$$\mathbb{E}[\text{nb. } m\text{-cycles}] = \binom{n}{m} \frac{m!}{2m} \text{Tr}((WP)^m) \sim \frac{n^m}{2m} \text{Tr}((WP)^m).$$

For VSBM, $\text{Tr} \left(\frac{1}{2} \begin{pmatrix} q_{in} & q_{out} \\ q_{out} & q_{in} \end{pmatrix} \right)^m = \left(\frac{q_{in} + q_{out}}{2} \right)^m + \left(\frac{q_{in} - q_{out}}{2} \right)^m.$

3 Recovery requirements

The goal of community detection is to recover the labels X by observing G , up to some level of accuracy.

Definition 3. Let $(X, G) \sim \text{VSBM}(n, p, q)$. An algorithm $\hat{X}(G)$ solves reconstruction or recovery with accuracy $\alpha \in [0, 1]$ if whp

$$\frac{1}{n} \left| \langle X, \hat{X}(G) \rangle \right| \geq \alpha.$$

We say that the LHS is the agreement between X and $\hat{X}(G)$.

Remark 3. If $X, \hat{X}(G) \in \{-1, 1\}^n$, then $\frac{1}{n} \left| \langle X, \hat{X}(G) \rangle \right| \in [0, 1]$. If $\frac{1}{n} \left| \langle X, \hat{X}(G) \rangle \right| = 1$, then X and $\hat{X}(G)$ either agree or disagree perfectly, and in both cases $\hat{X}(G)$ recovered the true partition. If $\frac{1}{n} \left| \langle X, \hat{X}(G) \rangle \right| = 0$, then the partition recovered by $\hat{X}(G)$ is completely off.

Depending on the accuracy α that we require, we have different terminologies:

α	"name"
$\Omega(1)$	weak recovery
$1 - o(1)$	almost exact recovery
1	exact recovery

where $\Omega(1)$ denotes some quantity that is bounded away from 0 as $n \rightarrow \infty$. Note that an accuracy $\alpha = o(1)$ can be obtained by random guessing. In other words, exact recovery requires the entire partition to be correctly recovered, almost exact recovery allows for a vanishing fraction of misclassified vertices, and weak recovery allows for a non-trivial fraction of misclassified vertices.

4 Exact Recovery

Recall that an algorithm $\hat{X}(G)$ solves exact recovery if $\mathbb{P}\{\frac{1}{n}|\langle X, \hat{X}(G) \rangle| = 1\} = 1 - o(1)$, i.e. if it correctly recovers the entire partition with high probability.

A natural starting point is to resolve the estimation of X from the noisy observation G by taking the Maximum A Posteriori estimator, i.e.

$$\hat{X}_{MAP}(G) = \arg \max_{x \in \{-1, 1\}^n} \mathbb{P}(X = x|G)$$

This is the estimator that minimizes the probability of error, that is given by

$$\begin{aligned} \mathbb{P}_e &:= \mathbb{P}(\hat{X}_{MAP}(G) \neq X) \\ &= \sum_g \mathbb{P}(\hat{X}_{MAP}(g) \neq X|G = g)\mathbb{P}(G = g) \\ &= 1 - \sum_g \max_{x \in \{-1, 1\}^n} \mathbb{P}(X = x|G = g)\mathbb{P}(G = g), \end{aligned}$$

and this is minimized if $\mathbb{P}(X = x|G = g)$ is maximized for every realization g of G .

Lemma 1. *In Uniform-VSBM(n, q_{in}, q_{out}), if $q_{in} \geq q_{out}$ then MAP is equivalent to finding the min-bisection of G , i.e. a balanced partition with the least number of crossing edges.*

Proof.

$$\begin{aligned} \mathbb{P}(G = g|X = x) &= q_{in}^{E_{in}(g, x)} (1 - q_{in})^{2\binom{n}{2} - E_{in}(g, x)} q_{out}^{E_{out}(g, x)} (1 - q_{out})^{\binom{n}{2} - E_{out}(g, x)} \\ &\propto \left(\frac{q_{in}}{1 - q_{in}}\right)^{E_{in}(g, x)} \left(\frac{q_{out}}{1 - q_{out}}\right)^{E_{out}(g, x)} \\ &\propto \left(\frac{q_{out}(1 - q_{in})}{q_{in}(1 - q_{out})}\right)^{E_{out}(g, x)}, \end{aligned}$$

since $E_{in}(x, g) + E_{out}(x, g) = E(g)$, that does not depend on x . Assuming $q_{in} \geq q_{out}$, we have $\frac{q_{out}(1 - q_{in})}{q_{in}(1 - q_{out})} \leq 1$, thus

$$\arg \max_{x: x^T \mathbf{1} = 0} \mathbb{P}(X = x|G = g) \equiv \arg \max_{x: x^T \mathbf{1} = 0} \mathbb{P}(G = g|X = x) \quad (1)$$

$$= \arg \min_{x: x^T \mathbf{1} = 0} E_{out}(x, g). \quad (2)$$

Note that (1) holds because $\mathbb{P}(X = x)$ is the same for all equal size partitions. It follows that the MAP is equivalent to the Maximum Likelihood estimator. $\square$

Since MAP minimizes the probability of making an error for the reconstruction of the entire partition X , it minimizes the error probability for exact recovery. Thus, if MAP fails in solving exact recovery, no other algorithm can succeed.

Lemma 2. *Exact recovery is solvable whp if and only if the min-bisection of G is equal to the true partition (or planted partition) X whp.*

Theorem 1. *Exact recovery in $VSBM(n, \frac{a \log n}{n}, \frac{b \log n}{n})$ is*

1. *solvable if $\frac{a+b}{2} - \sqrt{ab} > 1$;*
2. *not solvable if $\frac{a+b}{2} - \sqrt{ab} < 1$.*

Proof. part 2): see monography (p. 43-48). □

4.1 Spectral Methods

Assume that $\frac{a+b}{2} - \sqrt{ab} > 1$, how do we actually compute the $\hat{X}_{MAP}$ (i.e. the min-bisection) estimator?

Naively, we could look at all the bisections of G and pick the one that minimizes the number of cross-edges. It turns out that this problem is NP-hard, because the number of possible bisections of a graph with n vertices is $\asymp 2^n$. But can we still solve exact recovery efficiently (i.e. in polynomial time) whp when $\frac{a+b}{2} - \sqrt{ab} > 1$?

Let S_{typ} be a typical set, i.e. $\mathbb{P}(G \in S_{typ}) = 1 - o(1)$. Then,

$$\mathbb{P}(\hat{X}_{MAP}(G) = X) = \mathbb{P}(\hat{X}_{MAP}(G) = X | G \in S_{typ}) \mathbb{P}(G \in S_{typ}) \quad (3)$$

$$+ \mathbb{P}(\hat{X}_{MAP}(G) = X | G \notin S_{typ}) \mathbb{P}(G \notin S_{typ}) \quad (4)$$

$$= \mathbb{P}(\hat{X}_{MAP}(G) = X | G \in S_{typ}) + o(1). \quad (5)$$

In words, in order to solve exact recovery with high probability, we can ignore atypical graphs. The Spectral Algorithm uses this relaxation.

Theorem 2. *If $\frac{a+b}{2} - \sqrt{ab} > 1$, then the Spectral Algorithm solves exact recovery.*

Spectral Algorithm for SBM:

Let A_G be the adjacency matrix of the graph G . We can express A_G as a perturbation of its expected value, i.e.

$$A_G = \mathbb{E}[A_G] + (A_G - \mathbb{E}[A_G]). \quad (6)$$

If A_G is indexed such that the first $n/2$ rows and columns are in the same community, then

$$\mathbb{E}[A_G] = \begin{pmatrix} q_{in}^{n/2 \times n/2} & q_{out}^{n/2 \times n/2} \\ q_{out}^{n/2 \times n/2} & q_{in}^{n/2 \times n/2} \end{pmatrix}, \quad (7)$$

where $q_{in}^{n/2 \times n/2}$ is the $n/2 \times n/2$ matrix with all entries equal to q_{in} . Here, in order to have q_{in} on the diagonal of $\mathbb{E}[A_G]$, we assumed that for each vertex i in G , $\mathbb{P}(E_{i,i} = 1) = q_{in}$, i.e. each vertex contains a self-loop with probability q_{in} . M has rank 2, and therefore it has two non-zero eigenvalues. The eigenvalues and the corresponding normalized eigenvectors of $\mathbb{E}[A_G]$ are

$$\lambda_1 = \frac{q_{in} + q_{out}}{2} n, \quad \phi_1 = \frac{1}{\sqrt{n}} \mathbf{1}^n; \quad (8)$$

$$\lambda_2 = \frac{q_{in} - q_{out}}{2} n, \quad \phi_2 = \frac{1}{\sqrt{n}} \begin{pmatrix} \mathbf{1}^{n/2} \\ -\mathbf{1}^{n/2} \end{pmatrix}; \quad (9)$$

$$\lambda_3 = \dots = \lambda_n = 0, \quad \langle \phi_3, \dots, \phi_n \rangle = \langle \phi_1, \phi_2 \rangle^\perp. \quad (10)$$

Note that ϕ_2 takes opposite signs on each community. If the noise $Z = A_G - \mathbb{E}[A_G]$ does not disrupt the first two eigenvectors of A_G to be somewhat aligned with those of $\mathbb{E}[A_G]$, then we can exploit the second eigenvector of A_G to obtain a plausible reconstruction.

Denote by $\tilde{\phi}_2$ the eigenvector corresponding to the second largest eigenvalue of A_G . Define for $i \in [n]$

$$\hat{X}_{spec}(i) := \begin{cases} +1 & \text{if } \tilde{\phi}_2(i) \geq 0, \\ -1 & \text{if } \tilde{\phi}_2(i) < 0. \end{cases} \quad (11)$$

Theorem 3. $\hat{X}_{spec} \equiv \hat{X}_{MAP}$ with high probability if $\frac{a+b}{2} - \sqrt{ab} > 1$.

The proof of this Theorem uses some results of spectral theory, such as the Weyl's Theorem and the Davis-Kahan Theorem.

5 Weak Recovery

Recall that an algorithm $\hat{X}(G)$ solves weak recovery if $\mathbb{P}\{\frac{1}{n} |\langle X, \hat{X}(G) \rangle| \geq \Omega(1)\} = 1 - o(1)$, i.e. if it performs better than any trivial algorithm (such as random guessing).

Theorem 4. *Weak recovery in $VSBM(n, \frac{a}{n}, \frac{b}{n})$ is*

1. *solvable if $\frac{a+b}{2} - \frac{2ab}{a+b} > 1$;*
2. *not solvable if $\frac{a+b}{2} - \frac{2ab}{a+b} < 1$.*

5.1 Broadcasting on Trees

Let T_t be a d -regular tree, i.e. each vertex has exactly d descendants, where $d \in \mathbb{Z}_+$. Let $X^{(0)} \sim \text{Rad}(1/2)$ be the root bit, that is drawn uniformly at random in $\{-1, 1\}$. Assume that on each branch of the tree the incoming bit is flipped with probability $\varepsilon \in [0, 1]$ independently from the other branches, and let $X^{(t)} = (X_1^{(t)}, \dots, X_d^{(t)})$ be the bits received at depth t in this tree. For instance the distribution of $X^{(1)}$ given $X^{(0)}$ will be

$$\mathbb{P}_{X_1^{(1)}, \dots, X_d^{(1)} | X^{(0)}}(x_1^{(1)}, \dots, x_d^{(1)} | x^{(0)}) = \prod_{i=1}^d (1 - \varepsilon)^{\mathbb{1}(x_i^{(1)} = x^{(0)})} \varepsilon^{\mathbb{1}(x_i^{(1)} \neq x^{(0)})} \quad (12)$$

The problem consists of broadcasting a bit from the root of the tree down to its leaves and trying to guess back the root bit from the leaf bits at large depth. We call this problem broadcasting on a d -regular tree with flip probability ε , and we denote it by $\text{BOT}(d, \varepsilon)$.

To define weak recovery in this context, note that $\mathbb{E}(X^{(0)} | X^{(t)})$ is a random variable that gives the difference between the probability that $X^{(0)} = 1$ and the probability that $X^{(0)} = -1$ given the leaf bits, as a function of the leaf bits $X^{(t)}$. If $\mathbb{E}(X^{(0)} | X^{(t)}) = 0$, then $\mathbb{P}(X^{(0)} = 1 | X^{(t)}) = \mathbb{P}(X^{(0)} = -1 | X^{(t)})$, and thus the leaf bits provide no useful information about the root. We are interested in understanding whether this takes place in the limit of large t or not.

Definition 4. *Weak recovery (or reconstruction) is solvable in broadcasting on a regular tree if $\lim_{t \rightarrow \infty} \mathbb{E}|\mathbb{E}(X^{(0)}|X^{(t)})| > 0$.*

We can give another definition of reconstruction on trees based on an hypothesis testing formulation. Let $P_H(x^{(t)}) := \mathbb{P}_{X^{(t)}|X^{(0)}}(x^{(t)}|H)$, with $H \in \{-1, 1\}$. Upon observing $X^{(t)} = x^{(t)}$, we declare H such that $P_H(x^{(t)})$ is maximal. Then the probability of error is given by

$$P_e = \frac{1}{2} \sum_{x^{(t)}} P_{-1}(x^{(t)}) \wedge P_1(x^{(t)}), \quad (13)$$

where $P_{-1}(x^{(t)}) \wedge P_1(x^{(t)}) = \min\{P_{-1}(x^{(t)}), P_1(x^{(t)})\}$.

Notice that $\sum_{x^{(t)}} |P_{-1}(x^{(t)}) - P_1(x^{(t)})| + 2 \sum_{x^{(t)}} P_{-1}(x^{(t)}) \wedge P_1(x^{(t)}) = 2$, thus

$$P_e = \frac{1}{2} - \frac{1}{2} TV(P_{-1}, P_1), \quad (14)$$

where $TV(P_{-1}, P_1) := \frac{1}{2} \sum_{x^{(t)}} |P_{-1}(x^{(t)}) - P_1(x^{(t)})|$ denotes the Total Variation distance between P_{-1} and P_1 .

Definition 5. *Weak recovery in $BOT(d, \varepsilon)$ is solvable if $P_e = \frac{1}{2} - \Omega(1)$, where $\Omega(1)$ denotes some quantity that is bounded away from 0 as $t \rightarrow \infty$.*

Remark 4. Definition 4 and 5 are equivalent, because

$$\mathbb{E} \left| \mathbb{E}(X^{(0)}|X^{(t)}) \right| = \sum_{x^{(t)}} \left| \mathbb{P}_{X^{(0)}|X^{(t)}}(1|x^{(t)}) - \mathbb{P}_{X^{(0)}|X^{(t)}}(-1|x^{(t)}) \right| \mathbb{P}_{X^{(t)}}(x^{(t)}) \quad (15)$$

$$= \sum_{x^{(t)}} \left| \mathbb{P}_{X^{(0)}, X^{(t)}}(1, x^{(t)}) - \mathbb{P}_{X^{(0)}, X^{(t)}}(-1, x^{(t)}) \right| \quad (16)$$

$$= \sum_{x^{(t)}} \left| \mathbb{P}_{X^{(0)}}(1) P_1(x^{(t)}) - \mathbb{P}_{X^{(0)}}(-1) P_{-1}(x^{(t)}) \right| \quad (17)$$

$$= \frac{1}{2} \sum_{x^{(t)}} \left| P_1(x^{(t)}) - P_{-1}(x^{(t)}) \right| = TV(P_1, P_{-1}), \quad (18)$$

thus $\mathbb{E} \left| \mathbb{E}(X^{(0)}|X^{(t)}) \right| = \Omega(1) \iff TV(P_1, P_{-1}) = \Omega(1) \iff P_e = \frac{1}{2} - \Omega(1)$.

The first result is due to Kesten-Stigum.

Theorem 5 (Kesten-Stigum). *Weak recovery in $BOT(d, \varepsilon)$ is*

1. *solvable if $d(1 - 2\varepsilon)^2 > 1$;*
2. *not solvable if $d(1 - 2\varepsilon)^2 \leq 1$.*

Proof of the positive part of Theorem 5.

Following Definition 5 we want to show that $TV(P_1, P_{-1}) = \Omega(1)$ if $d(1 - 2\varepsilon)^2 > 1$. Let $\Delta^{(t)} = \sum_{i=1}^d X_i^{(t)} = \#\{+1 \text{ at depth } t\} - \#\{-1 \text{ at depth } t\}$ be the difference variable at

depth t . Note that $\Delta^{(t)} > 0$ means that the majority of the leaves at depth t is in community +1, and $\Delta^{(t)} < 0$ implies that the majority of the leaves at depth t is in community -1. Hence, upon observing $\Delta^{(t)} = \delta^{(t)}$, we may declare $H = \text{sign}(\delta^{(t)})$, by majority vote. Let $Q_H^{(t)}(\delta^{(t)}) := \mathbb{P}_{\Delta^{(t)}|X^{(0)}}(\delta^{(t)}|H)$, $H \in \{-1, 1\}$. We compute the first and the second moment of these distributions. If there are x bits of value 1 and y bits of value -1 at generation t , then $\Delta^{(t+1)}$ would be the sum of $xd \text{Rad}(1 - \varepsilon)$ and $yd \text{Rad}(\varepsilon)$ (all independent), and since the expectation of $\text{Rad}(1 - \varepsilon)$ is $1 - 2\varepsilon$, we have

$$\mathbb{E}(\Delta^{(t+1)}|\Delta^{(t)}) = dx(1 - 2\varepsilon) + dy(2\varepsilon - 1) \quad (19)$$

$$= d(1 - 2\varepsilon)(x - y) \quad (20)$$

$$= d(1 - 2\varepsilon)\Delta^{(t)}. \quad (21)$$

Then,

$$\mathbb{E}(\Delta^{(t+1)}|X^{(0)}) = \mathbb{E}(\mathbb{E}(\Delta^{(t+1)}|\Delta^{(t)})|X^{(0)}) \quad (22)$$

$$= d(1 - 2\varepsilon)\mathbb{E}(\Delta^{(t)}|X^{(0)}) \quad (23)$$

$$= d^{t+1}(1 - 2\varepsilon)^{t+1}X^{(0)}. \quad (24)$$

We now look at the second moment.

$$\text{Var}(\Delta^{(t+1)}|\Delta^{(t)}) = \text{Var}\left(\sum_{i=1}^{dx} \text{Rad}(1 - \varepsilon) + \sum_{j=1}^{dy} \text{Rad}(\varepsilon)\right) \quad (25)$$

$$= \sum_{i=1}^{dx} \text{Var}(\text{Rad}(1 - \varepsilon)) + \sum_{j=1}^{dy} \text{Var}(\text{Rad}(\varepsilon)) \quad (26)$$

$$= dx(4\varepsilon(1 - \varepsilon)) + dy(4\varepsilon(1 - \varepsilon)) \quad (27)$$

$$= d^{t+1}4\varepsilon(1 - \varepsilon), \quad (28)$$

since $x + y = d^t$. Hence,

$$\mathbb{E}((\Delta^{(t+1)})^2|\Delta^{(t)}) = \text{Var}(\Delta^{(t+1)}|\Delta^{(t)}) + (\mathbb{E}(\Delta^{(t+1)}|\Delta^{(t)}))^2 \quad (29)$$

$$= d^{t+1}4\varepsilon(1 - \varepsilon) + d^2(1 - 2\varepsilon)^2(\Delta^{(t)})^2 \quad (30)$$

$$= d^{t+1}v + ds(\Delta^{(t)})^2, \quad (31)$$

where we denoted $v := 4\varepsilon(1 - \varepsilon)$ and $s := d(1 - 2\varepsilon)^2$.

Then,

$$\mathbb{E}((\Delta^{(t+1)})^2|X^{(0)}) = \mathbb{E}(\mathbb{E}((\Delta^{(t+1)})^2|\Delta^{(t)})|X^{(0)}) \quad (32)$$

$$= d^{t+1}v + ds\mathbb{E}((\Delta^{(t)})^2|X^{(0)}) \quad (33)$$

$$= d^{t+1}v + d^{t+1}vs + (ds)^2\mathbb{E}((\Delta^{(t-1)})^2|X^{(0)}) \quad (34)$$

$$= d^{t+1}v + d^{t+1}vs + d^{t+1}vs^2 + (ds)^3\mathbb{E}((\Delta^{(t-2)})^2|X^{(0)}) \quad (35)$$

$$= d^{t+1}v(1 + s + s^2 + \dots + s^t) + (ds)^{t+1} \quad (36)$$

$$= d^{t+1}v \left(\frac{s^{t+1} - 1}{s - 1} \right) + (ds)^{t+1} \quad (37)$$

$$= (ds)^{t+1} \left(\frac{v}{s - 1} + 1 \right) - \frac{d^{t+1}v}{s - 1} \quad (38)$$

$$= (ds)^{t+1} \left(\frac{v}{s - 1} + 1 \right) (1 + o_t(1)) \quad (39)$$

$$= d^{2(t+1)}(1 - 2\varepsilon)^{2(t+1)} \left(\frac{4\varepsilon(1 - \varepsilon)}{d(1 - 2\varepsilon)^2 - 1} + 1 \right) (1 + o_t(1)), \quad (40)$$

where in (37) and in (39) we used the hypothesis that $s > 1$.

By applying Lemma 3 we obtain

$$2TV(Q_{-1}^{(t)}, Q_1^{(t)}) \geq \frac{4(ds)^t}{2(ds)^t(\frac{v}{s-1} + 1)} = \Omega(1). \quad (41)$$

It can be shown (by Data Processing Inequality) that $TV(Q_1^{(t)}, Q_{-1}^{(t)}) \leq TV(P_1^{(t)}, P_{-1}^{(t)})$, thus $TV(Q_1^{(t)}, Q_{-1}^{(t)}) = \Omega(1) \implies TV(P_1^{(t)}, P_{-1}^{(t)}) = \Omega(1)$. $\square$

Lemma 3.

$$2TV(Q_{-1}, Q_1) \geq \frac{(m_1 - m_{-1})^2}{m_1^{(2)} + m_{-1}^{(2)}}, \quad (42)$$

where $m_i = \mathbb{E}[\Delta^{(t)}|X^{(0)} = i]$ and $m_i^{(2)} = \mathbb{E}[(\Delta^{(t)})^2|X^{(0)} = i]$, for $i = -1, 1$.

Proof.

$$m_1 - m_{-1} = \sum_{\delta \in \mathbb{Z}} \delta(Q_1(\delta) - Q_{-1}(\delta)) \quad (43)$$

$$= \sum_{\delta \in \mathbb{Z}} \frac{Q_1(\delta) - Q_{-1}(\delta)}{\sqrt{Q_1(\delta) + Q_{-1}(\delta)}} \delta \sqrt{Q_1(\delta) + Q_{-1}(\delta)} \quad (44)$$

$$\leq \sqrt{\sum_{\delta \in \mathbb{Z}} \frac{(Q_1(\delta) - Q_{-1}(\delta))^2}{Q_1(\delta) + Q_{-1}(\delta)}} \sqrt{\sum_{\delta \in \mathbb{Z}} \delta^2(Q_1(\delta) + Q_{-1}(\delta))} \quad (45)$$

$$\leq \sqrt{\sum_{\delta \in \mathbb{Z}} |Q_1(\delta) - Q_{-1}(\delta)|} \sqrt{\sum_{\delta \in \mathbb{Z}} \delta^2(Q_1(\delta) + Q_{-1}(\delta))} \quad (46)$$

$$\leq \sqrt{2TV(Q_{-1}, Q_1)} \sqrt{m_1^{(2)} + m_{-1}^{(2)}}, \quad (47)$$

where (45) follows from Cauchy-Schwarz inequality and (46) follows because $\frac{|Q_1(\delta) - Q_{-1}(\delta)|}{Q_1(\delta) + Q_{-1}(\delta)} \leq 1$ for any δ . $\square$

Intuition for the negative part of Theorem 5.

Let $I(X^{(0)}; X^{(t)})$ be the Mutual Information between $X^{(0)}$ and $X^{(t)}$. One can show that

$$I(X^{(0)}; X^{(t)}) \leq d^t(1 - 2\varepsilon)^{2t}, \quad (48)$$

that is vanishing as $t \rightarrow \infty$, if $d(1 - 2\varepsilon)^2 < 1$. Intuitively, the mutual information between two variables is a measure of dependency. For instance, $I(X; Y) = 0 \iff X$ and Y are independent. Thus, (48) that in the limit of $t \rightarrow \infty$, the communities of the leaves gives no information about the community of the root, hence weak recovery is not solvable. $\square$

It has been shown that the result of Theorem 5 is still true if instead of a d -regular tree we take a Galton-Watson tree of Poisson(d) offspring.

Theorem 6. *In the broadcasting model with a Galton-Watson tree of Poisson(d) offspring and flip probability ε , weak recovery is solvable if and only if*

$$d(1 - 2\varepsilon)^2 > 1$$

5.2 Negative part of Theorem 4

Lemma 4. *If $\mathbb{P}(X_2 = +1 | G, X_1 = +1) \rightarrow \frac{1}{2}$ almost surely, then weak recovery is not solvable.*

In words, if the community of vertex 1 does not give any information about the community of vertex 2 then weak recovery will not be solvable. Here the role of vertex 1 and 2 is arbitrary, it could be any pair of vertices. The proof of this lemma requires some thoughts.

To connect weak recovery on VSBM with reconstruction on BOT we need two facts that follow from the homeworks:

1. $d_G(1, 2) = \Omega(\log n)$ with high probability, where $d_G(1, 2)$ is the length of the shortest path from vertex 1 to vertex 2 in G .
2. $B(1, \log \log n) = \{v' \in [n] : d_G(1, v') \leq \log \log n\}$ is a tree with high probability. In fact, one can show that $|B(1, \log \log n) - GW(\text{Pois}(\frac{a+b}{2}))|_{TV} \rightarrow 0$ as $n \rightarrow \infty$, where $GW(\text{Pois}(\frac{a+b}{2}))$ is the Galton-Watson tree with offspring distribution $\text{Pois}(\frac{a+b}{2})$.

To establish the negative part of Theorem 4, it is sufficient to argue that, if it is impossible to weakly recover a single vertex when a genie reveals all the leaves at such a depth, it must be impossible to solve weak recovery. In fact, consider $\mathbb{P}\{X_u = x_u | G = g, X_v = x_v\}$, the posterior distribution given the graph and an arbitrary vertex revealed (here u and v are arbitrary and chosen before the graph is drawn). With high probability, these vertices will not be within small graph-distance of each other, e.g. at distance $\Omega(\log n)$ with high probability (fact 1), and one can open a small neighborhood around u of diverging but small enough depth (e.g., $\log \log n$). Now reveal not only the value of X_v but in fact all

the values at the boundary of this neighborhood. This is an easier problem since the neighborhood is a tree with high probability (fact 2) and since there is approximately a Markov relationship between these boundary vertices and the original X_v (note that ‘approximately’ is used here since there is a negligible effect due to non-edges). We are now back to the broadcasting problem on trees discussed above, and the requirement $d(1 - 2\varepsilon)^2 \leq 1$ gives the theorem’s bound (since $d = (a + b)/2$ and $\varepsilon = b/(a + b)$).