Graph-Based Semi-Supervised Learning through the Lens of *Safety* (Supplementary Material)

Shreyas Sheshadri[1*] **Avirup Saha**[2*] **Priyank Patel**[1] **Samik Datta**[3] **Niloy Ganguly**[2]

[1]Flipkart Internet Private Limited, India
[2]Indian Institute of Technology, Kharagpur, India
[3]Amazon, India
[*]*Equal contribution*

1 JACOBI ITERATIONS

The general objective of G-SSL with 2 classes and labels being the 2 dimensional standard basis $\{e_1, e_2\}$ as provided in (3) can be rewritten as

$$\min_{Y} Tr\left(S(Y - Y')(Y - Y')^T\right) + \gamma_1 Tr\left(Y^T LY\right) + \gamma_2 Tr\left((Y - \lambda\mathbf{u})(Y - \lambda\mathbf{u})^T\right) - \gamma_3 Tr\left(Y^T WY\right) \quad (1)$$
$$Y \in \{0,1\}^{N\times 2} \text{ and } Y1_2 = 1_n \text{ and } Y^T 1_n = \frac{n}{2}1_2$$

where $W = 2Y'Y'^T - \mathbf{1}_{\mathcal{L}}\mathbf{1}_{\mathcal{L}}^T - S$, $\mathbf{u} = u1_n e_1^T + (1-u)1_n e_2^T$; $u \in \mathbb{R}$ is the prior distribution over the labels. Upon differentiating the objective in (1) we get $S(Y - Y') + \gamma_1 LY + \gamma_2(Y - \lambda\mathbf{u}) - \gamma_3 WY = 0$. Now by Jacobi method, we get $(S + \gamma_1 D + \gamma_2 I)Y = Y' + \gamma_2\mathbf{u} + (\gamma_1 A + \gamma_3 W)Y$ i.e., all the diagonal matrices are in the LHS and also we use $SY' = Y'$. Therefore we get the following fixed point iterations

$$Y_{vl}^{t+1} = \frac{Y'_{vl} + \gamma_1 \sum_{(v,j)\in E} A_{vj} Y_{jl}^t + \gamma_2 \lambda \mathbf{u}_{vl} + \gamma_3 \sum_{j\neq v} W_{vj} Y_{jl}^t}{S_{vv} + \gamma_1 \sum_{(i,j)\in E} A_{ij} + \gamma_2} \quad (2)$$

where v is the node index and l is the label index. Now the condition for Jacobi convergence is that the matrix $S + \gamma_1 L + \gamma_2 I - \gamma_3 W$ should be strictly diagonally dominant. Using the definition of W we see that $\gamma_2 > \gamma_3|n_1 - n_2|$ or $\gamma_2 > \gamma_3 n_l$ where n_1, n_2 are the sizes of classes labelled e_1 and e_2 respectively.

2 PROOFS

2.1 PROOF OF THEOREM 1

Proof. Let $\hat{Y} = \operatorname{argmin}_Y Q(Y)$ (the argmin is taken over the domain of Y as given in (1)) and we know that $Q(\hat{Y}) \leq Q(Y^*)$ Therefore it suffices to show that $Q(Y^*) < Q(Y)$ holds with high probability, for all $Y \neq Y^*$ and with Y having more than s mistakes. First let us rewrite Q as follows

$$\begin{aligned} Q(Y) &= Tr\left(S(Y - Y^*)(Y - Y^*)^T\right) + \gamma_1 Tr\left(Y^T LY\right) + \gamma_2 Tr\left((Y - \lambda\mathbf{u})(Y - \lambda\mathbf{u})^T\right) - \gamma_3 Tr\left(Y^T WY\right) \\ &= \langle S, (Y - Y^*)(Y - Y^*)^T\rangle - \langle W', YY^T\rangle + \gamma_2(n(1-\lambda) + \lambda^2) \end{aligned}$$

where S is a diagonal matrix such that $S_{ii} = 1$ if $i \in \mathcal{L}$ and $S_{ii} = 0$ if $i \notin \mathcal{L}$, and $W' = \gamma_1(A - D) + \gamma_3 W$. Let us consider

$$\begin{aligned} Q(Y) - Q(Y^*) &= \langle S, (Y - Y^*)(Y - Y^*)^T\rangle + \langle W', YY^T - Y^*Y^{*T}\rangle \\ &= \langle W' - E[W'], Y^*Y^{*T} - YY^T\rangle + \langle E[W], Y^*Y^{*T} - YY^T\rangle \\ &\quad + \langle S - E[S], (Y - Y^*)(Y - Y^*)^T\rangle + \langle E[S], (Y - Y^*)(Y - Y^*)^T\rangle \end{aligned} \quad (3)$$

We observe that, $E[W'] = \gamma_1 \left(q1_n1_n^T + (p-q)Y^*Y^{*T} - \frac{(pn+qn(K-1))}{K} I \right) + \gamma_3 \epsilon^2 (Y^*Y^{*T} - I)$. Let $d(Y) = \langle Y^*Y^{*T}, Y^*Y^{*T} - YY^T \rangle$ and we define $m(Y) = \langle I, (Y - Y^*)(Y - Y^*)^T \rangle$ i.e., the number of mismatches/mistakes between Y and Y^*. Also $\langle I, Y^*Y^{*T} - YY^T \rangle = \langle 1_n 1_n^T, Y^*Y^{*T} - YY^T \rangle = 0$. Therefore we can write (3) as

$$\begin{aligned} Q(Y) - Q(Y^*) = &\langle W' - E[W'], Y^*Y^{*T} - YY^T \rangle + \langle S - E[S], (Y - Y^*)(Y - Y^*)^T \rangle \\ &+ (\gamma_1(p-q) + \gamma_3 \epsilon^2) d(Y) + \epsilon m(Y) \end{aligned}$$

First, we have $\langle D - E[D], Y^*Y^{*T} - YY^T \rangle = 0$, next we note that $A_{ij}|X - E[A_{ij}|X]$ is a Bernoulli random variable with expectation 0 and variance bounded by $\frac{1}{4}$. Similarly $S'_{ii}|X - E[S'_{ii}|X]$ is a Bernoulli random variable with expectation 0 and variance bounded by $\frac{1}{4}$. Similarly Y'_i is a Bernoulli random variable with expectation 0 and variance bounded by $\frac{1}{4}$ independent of X. So we bound the variance as

$$\begin{aligned} Var[Q(y^*) - Q(y)] &= 4 \sum_{i<j} Var\left[W'_{ij}|X\right] (Y_i^{*T}Y_j^*)^2 \left(Y_i^{*T}Y_j^* - Y_i^TY_j\right)^2 + \sum_i Var[S_{ii}] \|Y_i - Y_i^*\|_2^2 \\ &= 4 \sum_{i,j} \left(\gamma_1^2 Var[A_{ij}|X] + \gamma_3^2 Var\left[Y'_{ij}|X\right]\right) \times (Y_i^{*T}Y_j^*)^2 \left(Y_i^{*T}Y_j^* - Y_i^TY_j\right)^2 + \sum_i \|Y_i - Y_i^*\|_2^2 \\ &\leq (\gamma_1^2 + \gamma_3^2) d(Y) + \frac{m(y)}{4} \end{aligned} \tag{4}$$

Let $A(Y) = \langle W' - E[W'], Y^*Y^{*T} - YY^T \rangle + \langle S - E[S], (Y - Y^*)^T(Y - Y^*) \rangle$ and $B(Y) = (\gamma_1(p-q) + \gamma_3 \epsilon_l^2) d(Y) + \epsilon_l m(Y)$. Now it is easy to see if $A(Y) > -B(Y)$ then $Q(Y) - Q(Y^*) > 0$. By applying LCI Bernstein's inequality, with $M = 2$, and using (4) we have

$$\begin{aligned} P(A(Y) \leq -B(Y)) &\leq \exp\left(-\frac{B(Y)^2}{2(\gamma_1^2 + \gamma_3^2)d(Y) + \frac{m(Y)}{2} + \frac{4}{3}B(Y)} \right) \\ &\leq \exp\left(-\frac{\left(\gamma_1(p-q) + \gamma_3\epsilon_l^2 + \epsilon_l \frac{m(Y)}{d(Y)}\right)^2 d(Y)}{c_1 + (\frac{4}{3}\epsilon_l + \frac{1}{2})\frac{m(Y)}{d(Y)}} \right) \end{aligned} \tag{5}$$

where $c_1 = 2\gamma_1(\gamma_1 + \frac{2}{3}(p-q)) + 2\gamma_3(\gamma_3 + \frac{2}{3}\epsilon_l^2)$. Now for this to be valid, we require $B(Y) > 0$ for all $Y \neq Y^*$. Therefore, we observe that in the case $\epsilon_l = 0$, it is required that $\gamma_1 > 0$. Similarly if $p = q$, then $\gamma_3 > 0$ must hold.

$\mathbf{m(\hat{Y}) \neq n}$**.** We first point out that we restrict $s + 1 \leq m(Y) \leq n$. In the case where there are no revealed labels or $\epsilon_l = 0$ (where we are interested in only recovering the partition), $m(Y) \leq \frac{n}{2}$ (because we can always flip labels). However we see that if $m(\hat{Y}) = n$ then $\hat{Y} = \mathbf{1} - Y^*$. For this value of $\hat{Y}$, in the case where least one label is revealed, we find $Q(\hat{Y}) - Q(Y^*) \geq 1$, thereby we get a contradiction. Hence $m(Y) \leq n - 1$.

Simplifying the proof. Instead of relying on Lemma 1.1 from [Chen and Xu, 2014], we provide an alternate simpler proof. For the un-revealed case, we swap the labels and we consider the labeling (of the two ways possible) which results in the minimum $m(Y)$. Now we write $d(Y) = \sum_{i,j} d(Y)^{ij}$, where $d(Y)^{ij} = (Y_i^{*T}Y_j^*)\left(Y_i^{*T}Y_j^* - Y_i^TY_j\right)$ and $d(Y)^{ij}$ can only take values 0 and 1. Thus in the case $Y_i^* = Y_j^* = e_1$, if $d(Y)^{ij} = 1$ then $Y_i = e_1, Y_j = e_2$ and in the case $Y_i^* = Y_j^* = e_2$, if $d(Y)^{ij} = 1$ then $Y_i = e_2, Y_j = e_1$. Therefore, we see that

$$d(Y) = m(Y)(n - m(Y)) \tag{6}$$

Since $s < m(Y) < n$, in fact because of balanced classes $m(Y)$ is always even, thereby $d(Y) > 0$ and similarly for $0 < m(Y) < n$, $d(Y)$ attains minimum value $n - 1$. Similarly we see that the maximum value of $d(Y)$ is $\frac{n^2}{4}$. We also note for any integer a, if $0 \leq a \leq n$ then $\frac{2}{n} \geq \frac{\min(a, n-a)}{a(n-a)}$. Therefore we get the following bound

$$d(Y) \geq \frac{n}{2} \min(m(Y), n - m(Y)) \text{ and } \frac{1}{n-s} \leq \frac{m(y)}{d(Y)} \leq \frac{1}{2} \tag{7}$$

Now since we require $B(Y) > 0$ for all $Y \neq Y^*$ and Y being balanced, we can use (7) to obtain the following sufficient condition

$$(p - q) + \epsilon_l^2 + \frac{\epsilon_l}{n - s} > 0 \tag{8}$$

Using (7) in (5), we get the following

$$P\left(A(Y) \leq -B(Y)\right) \leq \exp\left(-\frac{\left(c_2 + \frac{\epsilon_l}{n-s}\right)^2 \frac{n}{2}\min\left(m(Y), n-m(Y)\right)}{c_1 + \frac{4}{3}\epsilon_l + \frac{1}{2}}\right) \tag{9}$$

where $c_1 = 2\gamma_1(\gamma_1 + \frac{2}{3}(p-q)) + 2\gamma_3(\gamma_3 + \frac{2}{3}\epsilon_l^2)$ (same as in (5)) and $c_2 = \gamma_1(p-q) + \gamma_3\epsilon_l^2$.

Using (6) in (9), we have

$$P\left(A(Y) \leq -B(Y)\right) \leq \exp\left(-6(1-\frac{s}{n})\min(m(Y), n-m(Y))\log n\right) \tag{10}$$

Union Bound Now to prove this holds for all $Y \neq Y^*$.

$$\begin{aligned}
P\left(\exists Y \middle| m(Y) > s \wedge A(Y) \leq -B(Y)\right) &= \sum_{m(Y)>s} P\left(A(Y) \leq -B(Y)\right) \\
&\leq \sum_{t=s+1}^{n-1} |\{m(Y)=t\}| \exp\left(-6(1-\frac{s}{n})\min(t, n-t)\log n\right) \\
&\quad \text{(from (10))} \\
&= \sum_{t=s+1}^{n-1} \binom{n}{t} \exp\left(-6(1-\frac{s}{n})\min(t, n-t)\log n\right) \\
&\leq \sum_{t=s+1}^{n-1} n^{\min(t,n-t)} n^{-3\min(t,n-t)} \\
&\quad \left(\text{from Lemma 1 and since } s < \frac{n}{2}\right) \\
&\leq \sum_{t=s+1}^{n-1} n^{-2\min(t,n-t)} \\
&\leq \frac{1}{n^2}\sum_{t=s+1}^{n-1} 1 \\
&\leq \frac{1}{n}
\end{aligned}$$

□

Lemma 1. $\binom{n}{k} \leq n^{\min(k,n-k)}$

Proof.

$$\begin{aligned}
&\binom{n}{k} \leq n^k \leq n^{n-k} \\
&\implies \binom{n}{k} \leq n^{\min(k,n-k)} \quad \text{for } 1 \leq k \leq \frac{n}{2} \\
&\binom{n}{k} = \binom{n}{n-k} \leq n^{n-k} \leq n^k \\
&\implies \binom{n}{k} \leq n^{\min(k,n-k)} \quad \text{for } \frac{n}{2} \leq k \leq n
\end{aligned}$$

□

LCI Bernstein Inequality We restate from [Ke and Honorio, 2019]. Consider a sequence of $\{x_i\}$ which is LCI given Y, with $E[x_i|Y] = 0$ with $Var[x_i|Y] \leq \nu^2$ for all Y and $|x_i| \leq M$ (almost surely). For any $\epsilon > 0$

$$P\left(\sum_i x_i \geq \epsilon\right) \leq \exp\left(-\frac{\epsilon^2/2}{\sum_i \nu^2 + M\epsilon/3}\right) \tag{11}$$

3 ESTABLISHING THE GENERALIZATION HIERARCHY AMONG GLM, MAG AND SBM IN THE SYMMETRIC 2-COMMUNITY SETTING

In this section we establish a generalization hierarchy among three network generative models: Graph Latent Model (GLM) described by [Ke and Honorio, 2019], Multiplicative Attributed Graph (MAG) model described by [Kim and Leskovec, 2012] and Stochastic Block Model (SBM) described by [Abbe et al., 2015] in the symmetric 2-community setting. The method we follow here is to use the terminology of the most general model, viz. GLM to describe an equivalent of the MAG model in a restricted setting of the model parameters, and then further restrict the parameters of this equivalent MAG model to describe an equivalent of SBM. This clearly shows that GLM subsumes MAG, which in turn subsumes SBM in the symmetric 2-community setting.

3.1 REDUCING SYMMETRIC 2-COMMUNITY GLM TO SYMMETRIC 2-COMMUNITY MAG

We briefly restate a concise version of the definition of GLM in the symmetric 2-community setting.

Definition 1 (Symmetric 2-community GLM)**.** *Let n be a positive even integer, $d \in \mathbb{Z}+$, $f : \mathbb{R}^d \times \mathbb{R}^d \to [0, 1]$ such that $f(x, x') = f(x', x)$, and P_{e_1}, P_{e_2} be two distributions with support on $\mathbb{R}^d$ where e_1 and e_2 denote the 2-dimensional standard basis vectors. In a GLM with parameters $(n, d, f, P_{e_1}, P_{e_2})$, the community label vector Y^* is an $n \times 2$-dimensional vector from $\mathcal{Y} = \{Y : Y \in \{0, 1\}^{n\times 2}, Y^{*T} 1_n = \frac{n}{2} 1_2\}$ (i.e. communities are balanced) such that the community label of vertex $u \in [n]$ is $Y_u^* \in \{e_1, e_2\}$. $X \in \mathbb{R}^{n\times d}$ is a random matrix such that for each $i \in [n]$, $x_i \in \mathbb{R}^d$ is randomly generated from $P_{Y_i^*}$. A random graph G is generated as follows. For each pair of vertices $u, v \in [n]$, (u, v) is an edge of G with probability $f(x_u, x_v)$.*

In the context of [Kim and Leskovec, 2012], Multiplicative Attributed Graph (MAG) is defined as follows:

Definition 2 (Symmetric 2-community MAG)**.** *In a graph having n nodes (where n is a positive even integer), let each node $u \in [n]$ in a graph have l attributes denoted by $a_i(u)$, $i \in [l]$, and each attribute has cardinality c_i, $i \in [l]$. Let there be l 'affinity matrices' $\theta_i \in [0, 1]^{c_i \times c_i}$, $i \in [l]$ such that the probability of an edge between two nodes u and v of the graph is given by $P[u, v] = \prod_{i=1}^{l} \theta_i[a_i(u), a_i(v)]$ (θ_i need not be stochastic). We impose the additional constraint that the nodes belong to two balanced communities labeled e_1 and e_2 which denote the 2-dimensional standard basis vectors. The community label vector Y^* is an $n \times 2$-dimensional vector from $\mathcal{Y} = \{Y : Y \in \{0, 1\}^{n\times 2}, Y^{*T} 1_n = \frac{n}{2} 1_2\}$ such that the community label of node u is $Y_u^* \in \{e_1, e_2\}$.*

Let $k = \max_i c_i$ be the maximum cardinality of any attribute in the graph. For the sake of reducing GLM to the Multiplicative Attributed Graph (MAG) model as defined by [Kim and Leskovec, 2012], we must constrain the domain $\mathcal{X}$ of X to be $[k]^{n\times d}$ which is evidently a subset of $\mathbb{R}^{n\times d}$ which is normally used for GLM. This represents a graph with n nodes, each of which has d attributes with at most k values each. Let us consider d fixed k-dimensional symmetric matrices $\theta_i \in [0, 1]^{k\times k}$, $i \in [d]$. If the i^{th} attribute has cardinality $c_i < k$, then $\theta_i[a, b] = 0 \; \forall a, b \in \{c_i + 1, c_i + 2, \ldots, k\}$. Using these θ_i as lookup tables, let us define the homophily function f of GLM as $f : [k]^d \times [k]^d \to [0, 1]$ such that for any two vectors $x_u, x_v \in [k]^d$,

$$f(x_u, x_v) = \prod_{i=1}^{d} \theta_i[x_{ui}, x_{vi}] \tag{12}$$

Since θ_i is symmetric $\forall i$, we see that

$$f(x_v, x_u) = \prod_{i=1}^{d} \theta_i[x_{vi}, x_{ui}] \tag{13}$$
$$= \prod_{i=1}^{d} \theta_i[x_{ui}, x_{vi}] \tag{14}$$
$$= f(x_u, x_v) \tag{15}$$

so the symmetricity requirement of f is satisfied. Then for each pair of vertices $u, v \in [n]$, (u, v) is an edge of G with probability $f(x_u, x_v) = \prod_{i=1}^{d} \theta_i[x_{ui}, x_{vi}]$. This reduces the symmetric 2-community GLM to a MAG with two symmetric communities (with $n/2$ nodes in each community) with the community assignment given by the vector Y^*.

3.2 REDUCING SYMMETRIC 2-COMMUNITY MAG TO SYMMETRIC 2-COMMUNITY SBM

Let us provide a brief definition of symmetric 2-community SBM.

Definition 3 (Symmetric 2-community SBM)**.** *In a graph having n nodes (where n is a positive even integer), let each node $u \in [n]$ have a community label $Y_u^* \in \{e_1, e_2\}$ where e_1 and e_2 denote the 2-dimensional standard basis vectors. The community label vector Y^* is therefore an $n \times 2$-dimensional vector from $\mathcal{Y} = \{Y : Y \in \{0,1\}^{n\times 2}, Y^{*T} 1_n = \frac{n}{2} 1_2\}$ i.e. the communities are balanced. Let $Q \in [0,1]^{2\times 2}$ be a symmetric matrix (called the assortativity matrix) such that $Q_{i,j}$ denotes the probability of an edge between any two nodes belonging to communities labeled e_i and e_j. The probability of an edge between two nodes u and v of the graph is therefore given by $P[u, v] = Y_u^{*T} Q Y_v^*$.*

For the sake of reducing the GLM-equivalent of the symmetric 2-community MAG as defined above further to a symmetric 2-community SBM, we must fix $k = 2$ and $d = 1$ and restrict the domain $\mathcal{X}$ of X further to $\{1,2\}^n$ (actually $\{1,2\}^{n\times 1}$ since $d = 1$, but we use scalar notation for simplicity) where $x_i = 1$ if $Y_i^* = e_1$ and $x_i = 2$ if $Y_i^* = e_2$. The domain of X for SBM, $\mathcal{X} = \{1,2\}^n$ is clearly a subset of the domain $[k]^{n\times d}$ as defined for X in the context of MAG.

Then since $d = 1$, we have a single matrix $\theta \in [0,1]^{2\times 2}$. Let us define $f : \{1,2\} \times \{1,2\} \to [0,1]$ such that

$$f(x_u, x_v) = \theta[x_u, x_v] \tag{16}$$

Then for each pair of vertices $u, v \in [n]$, (u, v) is an edge of G with probability $f(x_u, x_v) = \theta[x_u, x_v]$. This reduces the symmetric 2-community MAG model to an SBM with two symmetric communities (labeled e_1 and e_2 as per the community label vector Y^*, or equivalently $1, 2$ as per the node attribute values X, with $n/2$ nodes in each community) with the community assignment given by the vector Y^* or the node attribute values X. The assortativity matrix Q of the SBM is given by θ which is by definition symmetric.

References

Emmanuel Abbe, Afonso S Bandeira, and Georgina Hall. Exact recovery in the stochastic block model. IEEE Transactions on Information Theory, 62(1):471–487, 2015.

Yudong Chen and Jiaming Xu. Statistical-computational phase transitions in planted models: The high-dimensional setting. In International Conference on Machine Learning, pages 244–252, 2014.

Chuyang Ke and Jean Honorio. Exact recovery in the latent space model. arXiv preprint arXiv:1902.03099, 2019.

Myunghwan Kim and Jure Leskovec. Multiplicative attribute graph model of real-world networks. Internet mathematics, 8 (1-2):113–160, 2012.