A Bound of Sample Loss and Gradient Change Wrt. Parameters

In this section, we prove the maximum change of sample loss and maximum change of gradient of loss function with respect to parameters $\{\Theta, W\}$ of spectral GNNs in Eq. (1). These two properties are widely used in the subsequent analysis.

Based on Assumption 1, we have the following lemmas.

Lemma 17 (Bound of Loss function to Parameters)**.** *Under Assumption 1, given a loss function ℓ and a spectral GNN, for parameters $\bar{\Theta}, \bar{W}, \Theta', W'$ and any node v_i with truth class y_i we have*

$$\|\ell(y_i, \hat{y}_i|_{\Theta=\bar{\Theta}, W=\bar{W}}) - \ell(y_i, \hat{y}_i|_{\Theta', W'})\|_F \leq \alpha_1 \sqrt{\|\bar{\Theta} - \Theta'\|_F^2 + \|\bar{W} - W'\|_F^2}$$

where $\alpha_1 = Lip(\ell)Lip(\Psi)$.

Proof. Under Assumption 1, we have

$$\|\ell(y_i, \hat{y}_i|_{\tau=\bar{\tau}}) - \ell(y_i, \hat{y}_i|_{\tau=\tau'})\| \leq Lip(\ell)\|\hat{y}_i|_{\tau=\bar{\tau}} - \hat{y}_i|_{\tau=\tau'}\|_F$$

and

$$\|Lip(\ell)\|\hat{y}_i|_{\tau=\bar{\tau}} - \hat{y}_i|_{\tau=\tau'}\|_F \leq Lip(\Psi)\|\bar{\tau} - \tau'\|_F.$$

This leads to

$$\|\ell(y_i, \hat{y}_i|_{\tau=\bar{\tau}}) - \ell(y_i, \hat{y}_i|_{\tau=\tau'})\| \leq Lip(\ell)Lip(\Psi)\|\bar{\tau} - \tau'\|_F.$$

□

Lemma 18 (Bound of Gradient to Parameters)**.** *Under Assumption 1, Assumption 2, for parameters $\bar{\Theta}, \bar{W}, \Theta', W'$ of a spectral GNN, the following holds for any node v_i with truth class y_i*

$$\|\nabla\ell(y_i, \hat{y}_i|_{\Theta=\bar{\Theta}, W=\bar{W}}) - \nabla\ell(y_i, \hat{y}_i|_{\Theta', W'})\|_F \leq \alpha_2 \sqrt{\|\bar{\Theta} - \Theta'\|_F^2 + \|\bar{W} - W'\|_F^2}$$

where $\alpha_2 = (Smt(\Psi)\beta_1 + Smt(\ell)Lip(\Psi)\beta_2)$.

Proof. Since we know that

$$\nabla\ell(y_i, \hat{y}_i|_{\tau=\bar{\tau}}) = \nabla_{\hat{y}_i}\ell(y, \hat{y}_i)|_{\tau=\bar{\tau}} \cdot \nabla\hat{y}_i|_{\tau=\bar{\tau}}$$

and

$$\nabla\ell(y_i, \hat{y}_i|_{\tau=\tau'}) = \nabla_{\hat{y}_i}\ell(y, \hat{y}_i)|_{\tau=\tau'} \cdot \nabla\hat{y}_i|_{\tau=\tau'}.$$

this gives

$$\begin{aligned}\nabla\ell(y_i, \hat{y}_i|_{\tau=\bar{\tau}}) - \nabla\ell(y_i, \hat{y}_i|_{\tau=\tau'}) = {} & \nabla_{\hat{y}_i}\ell(y, \hat{y}_i)|_{\tau=\bar{\tau}}(\nabla\hat{y}_i|_{\tau=\bar{\tau}} - \nabla\hat{y}_i|_{\tau=\tau'}) \\ & + (\nabla_{\hat{y}_i}\ell(y, \hat{y}_i)|_{\tau=\bar{\tau}} - \nabla_{\hat{y}_i}\ell(y, \hat{y}_i)|_{\tau=\tau'})\,\nabla\hat{y}_i|_{\tau=\tau'}.\end{aligned}$$

Hence, we obtain the following

$$\begin{aligned}\|\nabla\ell(y_i, \hat{y}_i|_{\tau=\bar{\tau}}) - \nabla\ell(y_i, \hat{y}_i|_{\tau=\tau'})\|_F \leq {} & \|\nabla_{\hat{y}_i}\ell(y, \hat{y}_i)|_{\tau=\bar{\tau}}\|_F \cdot \|\nabla\hat{y}_i|_{\tau=\bar{\tau}} - \nabla\hat{y}_i|_{\tau=\tau'}\|_F \\ & + \|\nabla_{\hat{y}_i}\ell(y, \hat{y}_i)|_{\tau=\bar{\tau}} - \nabla_{\hat{y}_i}\ell(y, \hat{y}_i)|_{\tau=\tau'}\|_F \cdot \|\nabla\hat{y}_i|_{\tau=\tau'}\|_F.\end{aligned} \tag{5}$$

Under Assumption 1and Assumption 2, we have:

$$\begin{aligned}\|\nabla\hat{y}_i|_{\tau=\bar{\tau}} - \nabla\hat{y}_i|_{\tau=\tau'}\|_F &\leq Smt(\Psi)\|\bar{\tau} - \tau'\|_F \\ \|\nabla_{\hat{y}_i}\ell(y, \hat{y}_i)|_{\tau=\bar{\tau}}\|_F &\leq \beta_1.\end{aligned} \tag{6}$$

Under Assumption 1, we have

$$\begin{aligned}\|\nabla_{\hat{y}_i}\ell(y, \hat{y}_i)|_{\tau=\bar{\tau}} - \nabla_{\hat{y}_i}\ell(y, \hat{y}_i)|_{\tau=\tau'}\|_F &\leq Smt(\ell)\|\hat{y}_i|_{\tau=\bar{\tau}} - \hat{y}_i|_{\tau=\tau'}\|_F \\ &\leq Smt(\ell)Lip(\Psi)\|\bar{\tau} - \tau'\|_F.\end{aligned} \tag{7}$$

Under Assumption 2, we have:

$$\|\nabla\hat{y}_i|_{\tau=\tau'}\|_F \leq \beta_2. \tag{8}$$

Substitute Eq. (6), Eq. (7), Eq. (8) into Eq. (5), we have

$$\begin{aligned}\|\nabla\ell(y_i, \hat{y}_i|_{\tau=\bar{\tau}}) - \nabla\ell(y_i, \hat{y}_i|_{\tau=\tau'})\|_F &\leq Smt(\Psi)\|\bar{\tau} - \tau'\|_F \cdot \beta_1 + Smt(\ell)Lip(\Psi)\|\bar{\tau} - \tau'\|_F \cdot \beta_2 \\ &= (Smt(\Psi)\beta_1 + Smt(\ell)Lip(\Psi)\beta_2)\,\|\bar{\tau} - \tau'\|_F\end{aligned}$$

□

B UNIFORM TRANSDUCTIVE STABILITY OF SPECTRAL GNNS

Theorem 6 (Stability and Gradient Norm)**.** *Let Ψ be a spectral GNN trained using gradient descent for T iterations with a learning rate η on a training dataset S_m, and evaluated on a testing set $\mathcal{D}_u$. Under Assumption 1, for all iterations $t \in [1, T]$ and any sample (x_i, y_i) in S_m or $\mathcal{D}_u$, if the gradient norm satisfies $\|\nabla \ell(y_i, \hat{y}_i|_{\Theta^t, W^t})\|_F \leq \beta$, where $\{\Theta^t, W^t\}$ are the parameters at the t-th iteration, then Ψ satisfies γ-uniform transductive stability with:*

$$\gamma = r\beta, \quad r = \frac{2\eta\alpha_1}{m} \sum_{t=1}^{T} (1 + \eta\alpha_2)^{t-1},$$

where $\alpha_1 = Lip(\ell) \cdot Lip(\Psi)$ and $\alpha_2 = Smt(\Psi)\beta_1 + Smt(\ell)Lip(\Psi)\beta_2$.

Proof. We denote $\tau = [\Theta; W]$ as the concatenation of parameters Θ, W. According to Lemma 17, Lemma 18, we have

$$\|\ell(y_i, \hat{y}_i|_\tau) - \ell(y_i, \hat{y}_i|_{\tau'})\|_F \leq \alpha_1 \|\tau - \tau'\|_F$$

where $\alpha_1 = Lip(\ell)Lip(\Psi)$. and

$$\|\nabla\ell(y_i, \hat{y}_i|_\tau) - \nabla\ell(y_i, \hat{y}_i|_{\tau'})\|_F \leq \alpha_2 \|\tau - \tau'\|_F$$

where $\alpha_2 = (Smt(\Psi)\beta_1 + Smt(\ell)Lip(\Psi)\beta_2)$. The updation rule of gradient descent is:

$$\tau^{t+1} = \tau^t - \eta\nabla\mathcal{L}_{S_m}(\tau^t)$$
$$\tau_{ij}^{t+1} = \tau_{ij}^t - \eta\nabla\mathcal{L}_{S_m^{ij}}(\tau_{ij}^t)$$

where

$$\mathcal{L}_{S_m}(\tau^t) = \frac{1}{m}\sum_{r=1}^{m} \ell(y_r, \hat{y}_r|_{\tau^t})$$
$$\mathcal{L}_{S_m^{ij}}(\tau_{ij}^t) = \frac{1}{m}\sum_{r=1}^{m} \ell(y_r, \hat{y}_r|_{\tau_{ij}^t})$$

are the empirical loss on training dataset S_m and S_m^{ij} respectively. The difference between empirical loss is:

$$\mathcal{L}_{S_m^{ij}}(\tau_{ij}^t) - \mathcal{L}_{S_m}(\tau^t) = \frac{1}{m}\left[\sum_{r=1, r\neq i,j}^{m} \left(\ell(y_r, \hat{y}_r|_{\tau_{ij}^t}) - \ell(y_r, \hat{y}_r|_{\tau^t})\right) + \ell(y_j, \hat{y}_j|_{\tau_{it}^t}) - \ell(y_i, \hat{y}_i|_{\tau^t})\right].$$

We derive the parameter difference:

$$\begin{aligned}
&\|\tau_{ij}^{t+1} - \tau^{t+1}\|_F = \left\|\tau_{ij}^t - \eta\nabla\mathcal{L}_{S_m^{ij}}(\tau_{ij}^t) - \tau^t + \eta\nabla\mathcal{L}_{S_m}(\tau^t)\right\|_F \\
&\leq \|\tau_{ij}^t - \tau^t\|_F + \eta\|\nabla(\mathcal{L}_{S_m}(\tau^t) - \mathcal{L}_{S_m^{ij}}(\tau_{ij}^t))\|_F \\
&= \|\tau_{ij}^t - \tau^t\|_F + \frac{\eta}{m}\left\|\nabla\left[\sum_{\substack{r=1\\r\neq i,j}}^{m} \left(\ell(y_r, \hat{y}_r|_{\tau_{ij}^t}) - \ell(y_r, \hat{y}_r|_{\tau^t})\right) + \ell(y_j, \hat{y}_j|_{\tau_{ij}^t}) - \ell(y_i, \hat{y}_i|_{\tau^t})\right]\right\|_F \\
&\leq \|\tau_{ij}^t - \tau^t\|_F + \frac{\eta}{m}\left\|\sum_{\substack{r=1\\r\neq i,j}}^{m} \alpha_2\|\tau_{ij}^t - \tau^t\|_F + \nabla\left[\ell(y_j, \hat{y}_j|_{\tau_{ij}^t}) - \ell(y_i, \hat{y}_i|_{\tau^t})\right]\right\|_F \quad (Assumption\ 1) \\
&\leq \|\tau_{ij}^t - \tau^t\|_F + \frac{\eta}{m}(m-1)\alpha_2\|\tau_{ij}^t - \tau^t\|_F + \frac{\eta}{m}\left\|\nabla\left[\ell(y_j, \hat{y}_j|_{\tau_{ij}^t}) - \ell(y_i, \hat{y}_i|_{\tau^t})\right]\right\|_F \\
&\leq \|\tau_{ij}^t - \tau^t\|_F + \frac{\eta}{m}(m-1)\alpha_2\|\tau_{ij}^t - \tau^t\|_F + \frac{2\eta\beta}{m} \quad (Theorem\ 13) \\
&= \left(1 + \frac{m-1}{m}\eta\alpha_2\right)\|\tau_{ij}^t - \tau^t\|_F + \frac{2\eta\beta}{m} \\
&\leq (1 + \eta\alpha_2)\|\tau_{ij}^t - \tau^t\|_F + \frac{2\eta\beta}{m}
\end{aligned}$$

After T iterations, we obtain

$$\begin{aligned}\left\|\tau_{ij}^T-\tau^T\right\|_F &\leq (1+\eta\alpha_2)\left\|\tau_{ij}^{T-1}-\tau^{T-1}\right\|_F+\frac{2\eta\beta}{m}\\ &\leq (1+\eta\alpha_2)[(1+\eta\alpha_2)\left\|\tau_{ij}^{T-2}-\tau^{T-2}\right\|_F+\frac{2\eta\beta}{m}]\\ &\leq (1+\eta\alpha_2)^T\left\|\tau_{ij}^0-\tau^0\right\|_F+\sum_{t=1}^{T}(1+\eta\alpha_2)^{t-1}\frac{2\eta\beta}{m}\\ &=\sum_{t=1}^{T}(1+\eta\alpha_2)^{t-1}\frac{2\eta\beta}{m}\end{aligned}$$

If the loss function ℓ is α_1 Lipschitz continuous, then for the loss function ℓ on any sample (x_i, y_i) with parameter $\tau^T=[\Theta^T;W^T]$, $\tau_{ij}^T=[\Theta_{ij}^T;W_{ij}^T]$, we have:

$$\begin{aligned}\left|\ell(\hat{y}_i,y_i;\tau^T)-\ell(\hat{y}_i,y_i;\tau_{ij}^T)\right| &\leq \alpha_1\left|\tau^T-\tau_{ij}^T\right|\\ &\leq \alpha_1\sum_{t=1}^{T}(1+\eta\alpha_2)^{t-1}\frac{2\eta\beta}{m}\end{aligned}$$

□

C UNIFORM TRANSDUCTIVE STABILITY ON GENERAL MULTI-CLASS CSBM

We derive the uniform transductive stability of spectral GNNs defined in Eq. (1) on graphs generated by $G\sim cSBM(n,f,\Pi,Q)$. Then we discuss how the non-linear feature transformation function affect the stability.

We first give a brief introduction to inequalities and lemmas used in this proof.

Lemma 19 (Jensen's Inequality)**.** *Let X be an arbitrary random variable, and let $f:\mathbb{R}^1\to\mathbb{R}^1$ be a convex function such that $\mathbb{E}[f(X)]$ is finite. Then $f(\mathbb{E}[f(X)])\leq\mathbb{E}[f(X)]$.*

Lemma 20 (Markov's Inequality)**.** *If X is a non-negative random variable, then for all $a>0$,*

$$P(X\geq a)\leq\frac{\mathbb{E}[X]}{a}.$$

That is, the probability that X exceeds any given value a is no more than the expectation of X divided by a.

Remark. Lemma 19, Lemma 20 are important inequalities about a variable and its expectation. Details can be found in (Evans & Rosenthal, 2004).

Lemma 21 (Cauchy-Schwarz Inequality (Arfken et al., 2011))**.**

$$(\sum_{k=1}^{n}a_kb_k)^2\leq(\sum_{k=1}^{n}a_k^2)(\sum_{k=1}^{n}b_k^2)$$

The square of the ℓ_2-norm of product of two vectors is smaller than the product of the square of ℓ_2-norm of each vector.

Lemma 22 (Trace and Frobenius Norm)**.** *For any matrix $A\in\mathbb{R}^{n\times n}$, the relation between its trance and its Frobenius norm is*

$$Tr(A)\leq\sqrt{n}\cdot\|A\|_F.$$

Proof.

$$Tr(A)=\sum_{i=1}^{n}a_{ii}\leq\sum_{i=1}^{n}|a_{ii}|\leq\sqrt{n\cdot\sum_{i=1}^{n}|a_{ii}|^2}\quad(Lemma\ 21)=\sqrt{n}\cdot\sqrt{\sum_{i=1}^{n}a_{ii}^2}=\|A\|_F$$

□

Lemma 23 (Partial Derivatives)**.** *For spectral graph neural networks* $\hat{Y} = softmax(\sum_{k=0}^{K} \theta_k \tilde{A}^K XW)$*, node feature matrix* $X \in \mathbb{R}^{n \times f}$ *and ground truth node label matrix* $Y \in \mathbb{R}^{n \times C}$*, the cross-entropy loss of sample* (x_i, y_i) *is* $\ell(\hat{y}_i, y_i; \Theta, W) = -\sum_{c=1}^{C} Y_{ic} \log\left(\hat{Y}_{ic}\right)$*, then the partial derivative of* $\ell(\hat{y}_i, y_i; \Theta, W)$ *with respect to* θ_k *and* W *is*

$$\frac{\partial \ell(\hat{y}_i, y_i; \Theta, W)}{\partial \theta_k} = \sum_{c=1}^{C} \left(\hat{Y}_{ic} - Y_{ic}\right) \left(\tilde{A}^k XW\right)_{ic}$$

and

$$\frac{\partial \ell(\hat{y}_i, y_i; \Theta, W)}{\partial W_{pq}} = \left(\hat{Y}_{iq} - Y_{iq}\right) \left(\sum_{k=0}^{K} \theta_k \tilde{A}^k X\right)_{ip}$$

Proof. We begin with

$$Z = \sum_{k=0}^{K} \theta_k \tilde{A}^k XW, \qquad \hat{Y}_{ic} = \frac{e^{Z_{ic}}}{\sum_{c'=1}^{C} e^{Z_{ic'}}}, \qquad \ell(\hat{y}_i, y_i; \Theta, W) = -\sum_{c=1}^{C} Y_{ic} \log(\hat{Y}_{ic})$$

Then, we have

$$\frac{\partial \ell(\hat{y}_i, y_i; \Theta, W)}{\partial \hat{Y}_{ic}} = -\sum_{c=1}^{C} \frac{Y_{ic}}{\hat{Y}_{ic}},$$

$$\frac{\partial \hat{Y}_{ic}}{\partial Z_{ic'}} = \hat{Y}_{ic}(\delta_{cc'} - \hat{Y}_{ic'}),$$

where δ_{cq} is the Kronecker delta, which is 1 if $c = c'$ and 0 otherwise.

(1) Gradient w.r.t. θ_k

We have

$$\frac{\partial Z_{ic}}{\partial \theta_k} = (\tilde{A}^k XW)_{ic}$$

By the chain rule of gradient, we have:

$$\begin{aligned}
\frac{\partial \ell(\hat{y}_i, y_i; \Theta, W)}{\partial \theta_k} &= -\sum_{c=1}^{C} \frac{\ell(\hat{y}_i, y_i; \Theta, W)}{\partial \hat{Y}_{ic}} \cdot \left(\sum_{c'=1}^{C} \frac{\partial \hat{Y}_{ic}}{\partial Z_{ic'}} \cdot \frac{\partial Z_{ic'}}{\partial \theta_k}\right) \\
&= -\sum_{c=1}^{C} \frac{Y_{ic}}{\hat{Y}_{ic}} \cdot \left(\sum_{c'=1}^{C} \hat{Y}_{ic} \left(\delta_{cc'} - \hat{Y}_{ic'}\right) \cdot \left(\tilde{A}^k XW\right)_{ic'}\right) \\
&= -\sum_{c=1}^{C} Y_{ic} \cdot \left(\sum_{c'=1}^{C} \left(\delta_{cc'} - \hat{Y}_{ic'}\right) \cdot \left(\tilde{A}^k XW\right)_{ic'}\right) \\
&= -\sum_{c=1}^{C} Y_{ic} \cdot \left(\left(\tilde{A}^k XW\right)_{ic} - \sum_{c'=1}^{C} \hat{Y}_{ic'} \left(\tilde{A}^k XW\right)_{ic'}\right) \\
&= -\sum_{c=1}^{C} Y_{ic} \left(\tilde{A}^k XW\right)_{ic} + \sum_{c'=1}^{C} \hat{Y}_{ic'} \left(\tilde{A}^k XW\right)_{ic'} \\
&= \sum_{c=1}^{C} \left(\hat{Y}_{ic} - Y_{ic}\right) \left(\tilde{A}^k XW\right)_{ic}
\end{aligned}$$

(2) Gradient w.r.t. W

Based on the following

$$Z_{ic} = \sum_{k=0}^{K} \theta_k \sum_{j=1}^{n} (\tilde{A}^k)_{ij} \sum_{r=1}^{f} X_{jr} W_{rc},$$

we have

$$\frac{\partial Z_{ic}}{\partial W_{pq}} = \sum_{k=0}^{K} \theta_k \sum_{j=1}^{n} (\tilde{A}^k)_{ij} X_{jp} \delta_{cq} = \delta_{cq} \sum_{k=0}^{K} \theta_k \left(\tilde{A}^k X\right)_{ip}$$

where δ_{cq} is the Kronecker delta, which is 1 if $c = q$ and 0 otherwise. Then, by the chain rule of gradient, we have:

$$\begin{aligned}
\frac{\partial \ell(\hat{y}_i, y_i; \Theta, W)}{\partial W_{pq}} &= -\sum_{c=1}^{C} \frac{\ell(\hat{y}_i, y_i; \Theta, W)}{\partial \hat{Y}_{ic}} \cdot \left(\sum_{c'=1}^{C} \frac{\partial \hat{Y}_{ic}}{\partial Z_{ic'}} \cdot \frac{\partial Z_{ic'}}{\partial W_{pq}} \right) \\
&= -\sum_{c=1}^{C} \frac{Y_{ic}}{\hat{Y}_{ic}} \cdot \left(\sum_{c'=1}^{C} \hat{Y}_{ic} \left(\delta_{cc'} - \hat{Y}_{ic'} \right) \cdot \left(\delta_{c'q} \sum_{k=0}^{K} \theta_k \left(\tilde{A}^k X\right)_{ip} \right) \right) \\
&= -\sum_{c=1}^{C} Y_{ic} \cdot \left(\sum_{c'=1}^{C} \left(\delta_{cc'} - \hat{Y}_{ic'} \right) \cdot \left(\delta_{c'q} \sum_{k=0}^{K} \theta_k \left(\tilde{A}^k X\right)_{ip} \right) \right) \\
&= -\sum_{c=1}^{C} Y_{ic} \cdot \left(\left(\delta_{cq} \sum_{k=0}^{K} \theta_k \left(\tilde{A}^k X\right)_{ip} \right) - \sum_{c'=1}^{C} \hat{Y}_{ic'} \left(\delta_{c'q} \sum_{k=0}^{K} \theta_k \left(\tilde{A}^k X\right)_{ip} \right) \right) \\
&= -\sum_{c=1}^{C} Y_{ic} \left(\delta_{cq} \sum_{k=0}^{K} \theta_k \left(\tilde{A}^k X\right)_{ip} \right) + \sum_{c'=1}^{C} \hat{Y}_{ic'} \left(\delta_{c'q} \sum_{k=0}^{K} \theta_k \left(\tilde{A}^k X\right)_{ip} \right) \\
&= \sum_{c=1}^{C} \left(\hat{Y}_{ic} - Y_{ic} \right) \left(\delta_{cq} \sum_{k=0}^{K} \theta_k \left(\tilde{A}^k X\right)_{ip} \right) \\
&= \sum_{c=1}^{C} \sum_{k=0}^{K} \theta_k \delta_{cq} \left(\hat{Y}_{ic} - Y_{ic} \right) \left(\tilde{A}^k X\right)_{ip} \\
&= \left(\hat{Y}_{iq} - Y_{iq} \right) \left(\sum_{k=0}^{K} \theta_k \tilde{A}^k X \right)_{ip}
\end{aligned}$$

□

Theorem 8. *Consider a spectral GNN Ψ with polynomial order K trained using full-batch gradient descent for T iterations with a learning rate η on a training dataset S_m sampled from a graph $G \sim cSBM(n, f, \Pi, Q)$ with average node degree $d \ll n$. When $n \to \infty$ and $K \ll n$, under Assumptions 1, 2, and 4, for any node v_i, $i \in [n]$, and for a constant $\epsilon \in (0, 1)$, with probability at least $1 - \epsilon$, Ψ satisfies γ-uniform transductive stability, where $\gamma = r\beta$ and*

$$\begin{aligned}
\beta = \frac{1}{\epsilon} \Bigg[& O\left(\mathbb{E}\left[\|\hat{y}_i - y_i\|_F^2 \right] \right) + O\left(\|\pi_{y_i}^\top \pi_{y_i} + \Sigma_{y_i}\|_F \right) \\
& + O\left(\sum_{k=1}^{K} \sum_{j=1}^{n} \mathbb{E}[A_{ij}^k] \left\| \sum_{t=1}^{n} \mathbb{E}[A_{it}^k] \pi_{y_j}^\top \pi_{y_t} + \mathbb{E}[A_{ij}^k] \Sigma_{y_j} \right\|_F \right) \Bigg].
\end{aligned}$$

Proof. Any spectral GNNs in Eq. (1) with linear feature transformation function, and polynomial basis expanded on normalized graph matrix can be transformed into the format:

$$\hat{Y} = softmax(\sum_{k=0}^{K} \theta_k \tilde{A}^k X W) \tag{9}$$

where $\tilde{A} = D^{-\frac{1}{2}} A D^{-\frac{1}{2}}$ is the normalized graph adjacency matrix, D is the diagonal degree matrix. We denotes $Y \in \mathbb{R}^{n \times C}$ as the ground truth node label matrix.

(1) Walk counting

According to Definition 7, we have

$$\mathbb{E}[A_{ij}^k] = \sum_{p \in P_{ij}^k} \prod_{(v,v') \in p} Q_{yy'}$$

(2) Feature expectation

Since we have $G \sim cSBM(n, f, \Pi, Q)$, node classes have a uniform prior $y_i \sim \mathcal{U}(1, C)$. Thus,

$$\begin{aligned}\mathbb{E}\left[XW\right]_{ij} &= \frac{1}{n}\sum_{u=1}^{n}(\pi_{y_u}W)_j \\ &= \frac{1}{n}\sum_{u=1}^{n}\sum_{c=1}^{C}p(y_u=c)(\pi_c W)_j \\ &= \frac{1}{n}\sum_{u=1}^{n}\sum_{c=1}^{C}\frac{1}{C}(\pi_c W)_j \\ &= \frac{1}{C}\sum_{c=1}^{C}(\pi_c W)_j\end{aligned} \tag{10}$$

When $k \geq 1$, we have

$$\begin{aligned}\mathbb{E}[(\tilde{A}^k XW)_{ij}] &= \mathbb{E}\left[\tilde{A}^k_{i:}\right]\mathbb{E}\left[(XW)_{:j}\right] \\ &= \sum_{s=1}^{n}\mathbb{E}\left[\tilde{A}^k_{is}\right]\mathbb{E}\left[(XW)_{sj}\right] \\ &= \sum_{s=1}^{n}\mathbb{E}\left[\tilde{A}^k_{is}\right]\cdot\frac{1}{C}\sum_{c=1}^{C}(\pi_c W)_j\end{aligned}$$

when $k = 0$, we have

$$\begin{aligned}\mathbb{E}[(IXW)_{ij}] &= \mathbb{E}\left[(XW)_{ij}\right] \\ &= \frac{1}{C}\sum_{c=1}^{C}(\pi_c W)_j\end{aligned}$$

Thus,

$$\mathbb{E}[(\tilde{A}^k XW)_{ij}] = \begin{cases}\frac{1}{C}\sum_{c=1}^{C}(\pi_c W)_j, & k=0 \\ \sum_{s=1}^{n}\mathbb{E}\left[\tilde{A}^k_{is}\right]\cdot\frac{1}{C}\sum_{c=1}^{C}(\pi_c W)_j, & k\geq 1\end{cases} \tag{11}$$

(3) Gradient Norm.

The gradient norm can be relaxed as:

$$\begin{aligned}&\mathbb{E}\left[\|\nabla\ell(\hat{y}_i, y_i; \Theta, W)\|_F\right] \leq \mathbb{E}\left[\|\nabla\ell(\hat{y}_i, y_i; \Theta, W)\|_{\ell_1}\right] \\ &\qquad = \sum_{k=0}^{K}\mathbb{E}\left[\|\frac{\partial\ell(\hat{y}_i, y_i; \Theta, W)}{\partial\theta_k}\|_{\ell_1}\right] + \mathbb{E}\left[\|\frac{\partial\ell(\hat{y}_i, y_i; \Theta, W)}{\partial W}\|_{\ell_1}\right]\end{aligned} \tag{12}$$

According to Eq. (9) and Lemma 23, we get the partial derivatives $\frac{\partial\ell(\hat{y}_i, y_i; \Theta, W)}{\partial\theta_k}$ and $\frac{\partial\ell(\hat{y}_i, y_i; \Theta, W)}{\partial W_{pq}}$.

Specially, when $m = 1$, we get the partial derivatives of empirical loss on training sample (x_i, y_i)

$$\frac{\partial\ell(\hat{y}_i, y_i; \Theta, W)}{\partial\theta_k} = \sum_{c=1}^{C}\left(\hat{Y}_{ic} - Y_{ic}\right)\left(\tilde{A}^k XW\right)_{ic} \tag{13}$$

$$\frac{\partial\ell(\hat{y}_i, y_i; \Theta, W)}{\partial W_{pq}} = \left(\hat{Y}_{iq} - Y_{iq}\right)\left(\sum_{k=0}^{K}\theta_k\tilde{A}^k X\right)_{ip} \tag{14}$$

Thus, we have:

$$\begin{aligned}
\mathbb{E}\left[\|\frac{\partial \ell(\hat{y}_i, y_i; \Theta, W)}{\partial \theta_k}\|_{\ell_1}\right] &= \mathbb{E}\left[|\sum_{c=1}^{C}\left(\hat{Y}_{ic} - Y_{ic}\right)\left(\tilde{A}^k XW\right)_{ic}|\right] \\
&\leq \sum_{c=1}^{C}\mathbb{E}\left[|\left(\hat{Y}_{ic} - Y_{ic}\right)\left(\tilde{A}^k XW\right)_{ic}|\right] \\
&= \sum_{c=1}^{C}\mathbb{E}\left[|\left(\hat{Y}_{ic} - Y_{ic}\right)| \cdot |\left(\tilde{A}^k XW\right)_{ic}|\right] \\
&\leq \sum_{c=1}^{C}\frac{1}{2}\left(\mathbb{E}\left[\left(\hat{Y}_{ic} - Y_{ic}\right)^2\right] + \mathbb{E}\left[\left(\tilde{A}^k XW\right)_{ic}^2\right]\right), \quad (Lemma\ 28) \\
&= \frac{1}{2}\left(\mathbb{E}\left[\|\hat{y}_i - y_i\|_F^2\right] + \mathbb{E}\left[\|\tilde{A}^k_{i:}XW\|_F^2\right]\right)
\end{aligned} \tag{15}$$

and

$$\begin{aligned}
\mathbb{E}\left[\|\frac{\partial \ell(\hat{y}_i, y_i; \Theta, W)}{\partial W}\|_{\ell_1}\right] &= \sum_{p=1}^{f}\sum_{q=1}^{C}\mathbb{E}\left[\|\frac{\partial \ell(\hat{y}_i, y_i; \Theta, W)}{\partial W_{pq}}\|_{\ell_1}\right] \\
&= \sum_{p=1}^{f}\sum_{q=1}^{C}\mathbb{E}\left[|\left(\hat{Y}_{iq} - Y_{iq}\right)\left(\sum_{k=0}^{K}\theta_k \tilde{A}^k X\right)_{ip}|\right] \\
&\leq \sum_{p=1}^{f}\sum_{k=0}^{K}|\theta_k|\left(\sum_{q=1}^{C}\mathbb{E}\left[|\left(\hat{Y}_{iq} - Y_{iq}\right)| \cdot |\left(\tilde{A}^k X\right)_{ip}|\right]\right) \\
&\leq \sum_{p=1}^{f}\sum_{k=0}^{K}|\theta_k|\left(\mathbb{E}\left[\sum_{q=1}^{C}\left(\hat{Y}_{iq} - Y_{iq}\right)^2\right] + \mathbb{E}\left[\sum_{q=1}^{C}\left(\tilde{A}^k X\right)_{ip}^2\right]\right), \quad (Lemma\ 28) \\
&= \sum_{p=1}^{f}\sum_{k=0}^{K}|\theta_k|\left(\mathbb{E}\left[\|\hat{y}_i - y_i\|_F^2\right] + C\mathbb{E}\left[\left(\tilde{A}^k X\right)_{ip}^2\right]\right) \\
&= \sum_{k=0}^{K}|\theta_k|\left(f \cdot \mathbb{E}\left[\|\hat{y}_i - y_i\|_F^2\right] + C\mathbb{E}\left[\|\tilde{A}^k_{i:}X\|_F^2\right]\right)
\end{aligned} \tag{16}$$

(4) Expectation $\mathbb{E}\left[\|\tilde{A}^k_{i:}XW\|_F^2\right]$ and $\mathbb{E}\left[\|\tilde{A}^k_{i:}X\|_F^2\right]$

For sparse graphs G and its adjacency matrix A, when $d \ll n$ and $k \ll n$, A^k_{ia}, A^k_{ib} can be treated as independent variables due to following reasons: (1) The overlap between walks of different lengths is limited due to the sparsity; (2) there exist k-length walk between two nodes is rare event when $k \ll n$ and the joint occurrences of two rare event can be neglected. (3) when $d \ll n$, the variance of A^k_{ij} can be neglected compared with $\left(\mathbb{E}\left[A^k_{ij}\right]\right)^2$. Thus, by Eq. (11), we have the following for the case $k \geq 1$:

$$
\begin{aligned}
\mathbb{E}[\|\tilde{A}^k_{i:}XW\|_F^2] &= \mathbb{E}\left[\sum_{c=1}^{C}\left(\sum_{s=1}^{n}\tilde{A}^k_{is}\left(XW\right)_{sc}\right)^2\right] \\
&= \mathbb{E}\left[\sum_{c=1}^{C}\sum_{s=1}^{n}\sum_{t=1}^{n}\tilde{A}^k_{is}\tilde{A}^k_{it}\left(XW\right)_{sc}\left(XW\right)_{tc}\right] \\
&= \sum_{c=1}^{C}\sum_{s,t=1}^{n}\mathbb{E}\left[\tilde{A}^k_{is}\tilde{A}^k_{it}\left(XW\right)_{sc}\left(XW\right)_{tc}\right] \\
&= \sum_{c=1}^{C}\sum_{s,t=1}^{n}\mathbb{E}\left[\tilde{A}^k_{is}\right]\cdot\mathbb{E}\left[\tilde{A}^k_{it}\right]\cdot\mathbb{E}\left[\left(XW\right)_{sc}\left(XW\right)_{tc}\right] \\
&= \sum_{c=1}^{C}\sum_{s=1}^{n}\mathbb{E}\left[\tilde{A}^k_{is}\right]\Bigg[\sum_{t=1,t\neq s}\mathbb{E}\left[\tilde{A}^k_{it}\right]\cdot\mathbb{E}\left[\left(XW\right)_{sc}\left(XW\right)_{tc}\right] \\
&\quad + \mathbb{E}\left[\tilde{A}^k_{is}\right]\cdot\mathbb{E}\left[\left(XW\right)^2_{sc}\right]\Bigg] \\
&= \frac{1}{d^{2k}}\sum_{c=1}^{C}\sum_{s=1}^{n}\mathbb{E}\left[\tilde{A}^k_{is}\right]\Bigg[\sum_{t=1,t\neq s}^{n}\mathbb{E}\left[\tilde{A}^k_{it}\right]\cdot\left(\pi_{y_s}W\right)_c\cdot\left(\pi_{y_t}W\right)_c \\
&\quad + \mathbb{E}\left[\tilde{A}^k_{is}\right]\cdot W_{:c}^\top\left(\pi_{y_s}^\top\pi_{y_s}+\Sigma_{y_s}\right)W_{:c}\Bigg]
\end{aligned}
$$

when $k=0$:

$$
\begin{aligned}
\mathbb{E}\left[\|\tilde{A}^k_{i:}XW\|_F^2\right] &= \mathbb{E}\left[\|X_{i:}W\|_F^2\right] \\
&= \mathbb{E}\left[\sum_{c=1}^{C}\left(XW\right)^2_{ic}\right] \\
&= \sum_{c=1}^{C}W_{:c}^\top\left(\pi_{y_i}^\top\pi_{y_i}+\Sigma_{y_i}\right)W_{:c}
\end{aligned}
$$

Thus, we have

$$
\mathbb{E}\left[\|\tilde{A}^k_{i:}XW\|_F^2\right] = \begin{cases} \sum_{c=1}^{C}W_{:c}^\top\left(\pi_{y_i}^\top\pi_{y_k}+\Sigma_{y_i}\right)W_{:c}, k=0 \\ \frac{1}{d^{2k}}\sum_{c=1}^{C}\sum_{s=1}^{n}\mathbb{E}\left[\tilde{A}^k_{is}\right]\Big[\sum_{t=1,t\neq s}^{n}\mathbb{E}\left[\tilde{A}^k_{it}\right]\cdot\left(\pi_{y_s}W\right)_c\cdot\left(\pi_{y_t}W\right)_c \\ \quad +\mathbb{E}\left[\tilde{A}^k_{is}\right]\cdot W_{:c}^\top\left(\pi_{y_s}^\top\pi_{y_s}+\Sigma_{y_s}\right)W_{:c}\Big], k\geq 1 \end{cases} \tag{17}
$$

Similarly, by Eq. (10), we have

$$
\mathbb{E}\left[\|\tilde{A}^k_{i:}X\|_F^2\right] = \begin{cases} \sum_{c=1}^{C}I_{:c}^\top\left(\pi_{y_i}^\top\pi_{y_k}+\Sigma_{y_i}\right)I_{:c}, k=0 \\ \frac{1}{d^{2k}}\sum_{q=1}^{f}\sum_{s=1}^{n}\mathbb{E}\left[\tilde{A}^k_{is}\right]\Big[\sum_{t=1,t\neq s}^{n}\mathbb{E}\left[\tilde{A}^k_{it}\right]\cdot\pi_{y_s,q}\cdot\pi_{y_t,q} \\ \quad +\mathbb{E}\left[\tilde{A}^k_{is}\right]\cdot I_{:q}^\top\left(\pi_{y_s}^\top\pi_{y_s}+\Sigma_{y_s}\right)I_{:q}\Big], k\geq 1 \end{cases} \tag{18}
$$

By putting Eq. (17) into Eq. (15), Eq. (18) into Eq. (16), and Eq. (15), Eq. (16) into Eq. (12), we have the following

$$
\begin{aligned}
&\mathbb{E}\left[\|\nabla \ell(\hat{y}_i, y_i; \Theta, W)\|_F\right] \leq \frac{1}{2}\left(\mathbb{E}\left[\|\hat{y}_i - y_i\|_F^2\right] + \sum_{c=1}^{C} W_{:c}^\top \left(\pi_{y_i}^\top \pi_{y_i} + \Sigma_{y_i}\right) W_{:c}\right) \\
&+ \sum_{k=1}^{K} \frac{1}{2}\Bigg[\mathbb{E}\left[\|\hat{y}_i - y_i\|_F^2\right] + \frac{1}{d^{2k}} \sum_{c=1}^{C}\sum_{s=1}^{n} \mathbb{E}\left[\tilde{A}_{is}^k\right] \\
&\quad \cdot \left[\sum_{t=1, t\neq s}^{n} \mathbb{E}\left[\tilde{A}_{it}^k\right] \cdot \left(\pi_{y_s} W\right)_c \cdot \left(\pi_{y_t} W\right)_c + \mathbb{E}\left[\tilde{A}_{is}^k\right] \cdot W_{:c}^\top \left(\pi_{y_s}^\top \pi_{y_s} + \Sigma_{y_s}\right) W_{:c}\right]\Bigg] \\
&+ |\theta_0| \left(f \cdot \mathbb{E}\left[\|\hat{y}_i - y_i\|_F^2\right] + C \sum_{c=q}^{f} I_{:q}^\top \left(\pi_{y_i}^\top \pi_{y_i} + \Sigma_{y_i}\right) I_{:q}\right) \\
&+ \sum_{k=1}^{K} |\theta_k| \Bigg[f \cdot \mathbb{E}\left[\|\hat{y}_i - y_i\|_F^2\right] + C \frac{1}{d^{2k}} \sum_{c=1}^{C}\sum_{s=1}^{n} \mathbb{E}\left[\tilde{A}_{is}^k\right] \\
&\quad \cdot \left[\sum_{t=1, t\neq s}^{n} \mathbb{E}\left[\tilde{A}_{it}^k\right] \cdot \pi_{y_s,q} \cdot \pi_{y_t,q} + \mathbb{E}\left[\tilde{A}_{is}^k\right] \cdot I_{:q}^\top \left(\pi_{y_s}^\top \pi_{y_s} + \Sigma_{y_s}\right) I_{:q}\right]\Bigg] \\
&= \left(\frac{K+1}{2} + f \sum_{k=0}^{K} |\theta_k|\right) \mathbb{E}\left[\|\hat{y}_i - y_i\|_F^2\right] \\
&\quad + \frac{1}{2} \sum_{c=1}^{C} W_{:c}^\top \left(\pi_{y_i}^\top \pi_{y_i} + \Sigma_{y_i}\right) W_{:c} + |\theta_0| C \sum_{c=1}^{C} I_{:c}^\top \left(\pi_{y_i}^\top \pi_{y_i} + \Sigma_{y_i}\right) I_{:c} \\
&\quad + \sum_{k=1}^{K} \frac{1}{d^{2k}} \sum_{c=1}^{C}\sum_{s=1}^{n} \mathbb{E}\left[\tilde{A}_{is}^k\right] \\
&\qquad \cdot \left[\sum_{t=1, t\neq s}^{n} \mathbb{E}\left[\tilde{A}_{it}^k\right] \cdot \left(\pi_{y_s} W\right)_c \cdot \left(\pi_{y_t} W\right)_c + \mathbb{E}\left[\tilde{A}_{is}^k\right] \cdot W_{:c}^\top \left(\pi_{y_s}^\top \pi_{y_s} + \Sigma_{y_s}\right) W_{:c}\right] \\
&\quad + \sum_{k=1}^{K} \frac{C}{d^{2k}} |\theta_k| \sum_{q=1}^{f}\sum_{s=1}^{n} \mathbb{E}\left[\tilde{A}_{is}^k\right] \\
&\qquad \cdot \left[\sum_{t=1, t\neq s}^{n} \mathbb{E}\left[\tilde{A}_{it}^k\right] \cdot \pi_{y_s,q} \cdot \pi_{y_t,q} + \mathbb{E}\left[\tilde{A}_{is}^k\right] \cdot I_{:q}^\top \left(\pi_{y_s}^\top \pi_{y_s} + \Sigma_{y_s}\right) I_{:q}\right] \\
&= \left(\frac{K+1}{2} + f \sum_{k=0}^{K} |\theta_k|\right) \mathbb{E}\left[\|\hat{y}_i - y_i\|_F^2\right] \\
&\quad + \frac{1}{2} \sum_{c=1}^{C} W_{:c}^\top \left(\pi_{y_i}^\top \pi_{y_i} + \Sigma_{y_i}\right) W_{:c} + |\theta_0| C \sum_{c=1}^{C} I_{:c}^\top \left(\pi_{y_i}^\top \pi_{y_i} + \Sigma_{y_i}\right) I_{:c} \\
&\quad + \sum_{k=1}^{K} \frac{1}{d^{2k}} \sum_{c=1}^{C}\sum_{s=1}^{n} \mathbb{E}\left[\tilde{A}_{is}^k\right] \left[W_{:c}^\top \left(\sum_{\substack{t=1 \\ t\neq s}}^{n} \mathbb{E}\left[\tilde{A}_{it}^k\right] \pi_{y_s}^\top \pi_{y_t} + \mathbb{E}\left[\tilde{A}_{is}^k\right] \left(\pi_{y_s}^\top \pi_{y_s} + \Sigma_{y_s}\right)\right) W_{:c}\right] \\
&\quad + \sum_{k=1}^{K} \frac{C|\theta_k|}{d^{2k}} \sum_{q=1}^{f}\sum_{s=1}^{n} \mathbb{E}\left[\tilde{A}_{is}^k\right] \left[\sum_{\substack{t=1 \\ t\neq s}}^{n} \mathbb{E}\left[\tilde{A}_{it}^k\right] \pi_{y_s,q} \pi_{y_t,q} + \mathbb{E}\left[\tilde{A}_{is}^k\right] I_{:q}^\top \left(\left(\pi_{y_s}^\top \pi_{y_s} + \Sigma_{y_s}\right)\right) I_{:q}\right]
\end{aligned}
$$

We further simplify it and relax it under Assumption 4 that:

$$
\begin{aligned}
\mathbb{E}\left[\|\nabla \ell(\hat{y}_i, y_i; \Theta, W)\|_F\right] \leq & \left(\frac{K+1}{2} + f\sum_{k=0}^{K} B_\Theta\right) \mathbb{E}\left[\|\hat{y}_i - y_i\|_F^2\right] \\
& + \frac{1}{2} Tr\left(W^T\left(\pi_{y_i}^\top \pi_{y_i} + \Sigma_{y_i}\right) W\right) + B_\Theta C Tr\left(\pi_{y_i}^\top \pi_{y_i} + \Sigma_{y_i}\right) \\
& + \sum_{k=1}^{K} \frac{1}{d^{2k}} \sum_{s=1}^{n} \mathbb{E}\left[\tilde{A}_{is}^k\right] Tr\left(\sum_{\substack{t=1 \\ t\neq s}}^{n} \mathbb{E}\left[\tilde{A}_{it}^k\right] \pi_{y_s}^\top \pi_{y_t} + \mathbb{E}\left[\tilde{A}_{is}^k\right]\left(\pi_{y_s}^\top \pi_{y_s} + \Sigma_{y_s}\right)\right) \\
& + \sum_{k=1}^{K} \frac{C B_\Theta}{d^{2k}} \sum_{s=1}^{n} \mathbb{E}\left[\tilde{A}_{is}^k\right] \left[\sum_{\substack{t=1 \\ t\neq s}}^{n} \mathbb{E}\left[\tilde{A}_{it}^k\right] Tr\left(\pi_{y_s}^\top \pi_{y_t}\right) + \mathbb{E}\left[\tilde{A}_{is}^k\right] Tr\left(\left(\pi_{y_s}^\top \pi_{y_s} + \Sigma_{y_s}\right)\right)\right] \\
\leq & \left(\frac{K+1}{2} + f B_\Theta (K+1)\right) \mathbb{E}\left[\|\hat{y}_i - y_i\|_F^2\right] \\
& + \left(\frac{B_W^2}{2} + B_\Theta C\right) Tr\left(\pi_{y_i}^\top \pi_{y_i} + \Sigma_{y_i}\right) \\
& + \sum_{k=1}^{K} \frac{1 + C B_\Theta}{d^{2k}} \sum_{j=1}^{n} \mathbb{E}\left[A_{ij}^k\right] \mathrm{Tr}\left(\sum_{\substack{t=1 \\ t\neq j}}^{n} \mathbb{E}\left[A_{it}^k\right] \pi_{y_j}^\top \pi_{y_t} + \mathbb{E}\left[A_{ij}^k\right]\left(\pi_{y_j}^\top \pi_{y_j} + \Sigma_{y_j}\right)\right)
\end{aligned} \tag{19}
$$

With Lemma 22, we rewrite it as

$$
\begin{aligned}
\mathbb{E}\left[\|\nabla \ell(\hat{y}_i, y_i; \Theta, W)\|_F\right] \leq & O\left(\mathbb{E}\left[\|\hat{y}_i - y_i\|_F^2\right]\right) + O\left(\|\pi_{y_i}^\top \pi_{y_i} + \Sigma_{y_i}\|_F\right) \\
& + O\left(\sum_{k=1}^{K}\sum_{j=1}^{n} \mathbb{E}\left[A_{ij}^k\right] \| \sum_{t=1}^{n} \mathbb{E}\left[A_{it}^k\right] \pi_{y_j}^\top \pi_{y_t} + \mathbb{E}\left[A_{ij}^k\right] \Sigma_{y_j} \|_F\right)
\end{aligned} \tag{20}
$$

(5) Concentration Bound.

By Jensen's inequality (Lemma 19), we have:

$$
\mathbb{E}[\|\nabla \ell(\hat{y}_i, y_i; \Theta, W)\|_F]^2 \leq \mathbb{E}[\|\nabla \ell(\hat{y}_i, y_i; \Theta, W)\|_F^2]
$$

i.e.,

$$
\mathbb{E}[\|\nabla \ell(\hat{y}_i, y_i; \Theta, W)\|_F] \leq \sqrt{\mathbb{E}[\|\nabla \ell(\hat{y}_i, y_i; \Theta, W)\|_F^2]} \tag{21}
$$

By Markov's inequality (Lemma 20), for a positive constant a, we have:

$$
\mathbb{P}(\|\nabla \ell(\hat{y}_i, y_i; \Theta, W)\|_F \geq a) \leq \frac{\mathbb{E}[\|\nabla \ell(\hat{y}_i, y_i; \Theta, W)\|_F]}{a} = \epsilon \tag{22}
$$

solving for a:

$$
a = \frac{\mathbb{E}[\|\nabla \ell(\hat{y}_i, y_i; \Theta, W)\|_F]}{\epsilon} \tag{23}
$$

Therefore, combining Eq. (20), Eq. (21), Eq. (22), Eq. (23), with probability at least $1 - \epsilon$, we have

$$
\|\nabla \ell(\hat{y}_i, y_i; \Theta, W)\|_F \leq \beta = \frac{1}{\epsilon} \mathbb{E}\left[\|\nabla \ell(\hat{y}_i, y_i; \Theta, W)\|_F\right]
$$

When $\|\nabla \ell(\hat{y}_i, y_i; \Theta, W)\|_F \leq \beta$, according to Theorem 6, spectral GNNs on graphs $G \sim cSBM(n, f, \Pi, Q)$ has γ- uniform transductive stability, we rewrite it in big O format:

$$
\begin{aligned}
\gamma = r\beta; \beta = \frac{1}{\epsilon} \Bigg[& O\left(\mathbb{E}\left[\|\hat{y}_i - y_i\|_F^2\right]\right) + O\left(\|\pi_{y_i}^\top \pi_{y_i} + \Sigma_{y_i}\|_F\right) \\
& + O\left(\sum_{k=1}^{K}\sum_{j=1}^{n} \mathbb{E}\left[A_{ij}^k\right] \| \sum_{t=1}^{n} \mathbb{E}\left[A_{it}^k\right] \pi_{y_j}^\top \pi_{y_t} + \mathbb{E}\left[A_{ij}^k\right] \Sigma_{y_j} \|_F\right) \Bigg],
\end{aligned}
$$

where r is the same as that in Theorem 6.

□

D GENERALIZATION ERROR BOUND OF SPECTRAL GNNS

We derive the generalization error bound of spectral GNNs based on uniform transductive stability. Then we analyze how training sample number affect the generalization error bound.

We first introduce two lemmas for this proof.

Lemma 24 (Inequality for permutation (El-Yaniv & Pechyony, 2006)). *Let Z be a random permutation vector. Let $f(Z)$ be an (m,q)-symmetric permutation function satisfying $\|f(Z)-f(Z^{ij})\|\leq\beta$ for all $i\in I_1^m, j\in I_{m+1}^{m+q}$. Let $H_2(n)\triangleq\sum_{i=1}^n\frac{1}{i^2}$ and $\omega(m,q)\triangleq q^2\left(H_2(m+q)-H_2(q)\right)$. Then*

$$\mathbb{P}\left(f(Z)-\mathbb{E}\left[f(Z)\right]\geq\epsilon\right)\leq\exp\left(-\frac{\epsilon^2}{2\beta^2\Omega(m,q)}\right)$$

Lemma 25 (Risk and uniform stability (El-Yaniv & Pechyony, 2006)). *Give any training set S_m and test set $\mathcal{D}_u$, we have:*

$$\mathbb{E}\left[\mathcal{L}_{\mathcal{D}_u}(\Theta,W)-\mathcal{L}_{S_m}(\Theta,W)\right]=\mathbb{E}\left[\Delta(i,j,i,i)\right]\quad i\in I_1^m, j\in I_{m+1,m+q}$$

where $\Delta(i,j,i,i)$ is the loss change of sample (x_i,y_i) when model is trained on two datasets exchange sample x_i,y_i in training dataset and sample (x_j,y_j) in testing dataset.

Theorem 9 (Generalization Error Bound). *Let $H_2(n)\triangleq\sum_{i=1}^n\frac{1}{i^2}$ and $\Omega(m,n-m)\triangleq(n-m)^2\left(H_2(n)-H_2(n-m)\right)$. For $\epsilon\in(0,1)$, if a spectral GNN is γ-uniform transductive stability with probability $1-\epsilon$, then under Assumption 3, for $\delta\in(0,1)$, with probability at least $(1-\delta)(1-\epsilon)$, the generalization error $\mathcal{L}_{\mathcal{D}_u}(\Theta,W)-\mathcal{L}_{S_m}(\Theta,W)$ is upper-bounded by:*

$$\gamma+\left(2\gamma+\left(\frac{1}{n-m}+\frac{1}{m}\right)\left(B_\ell-\gamma\right)\right)\sqrt{2\Omega(m,n-m)\log\frac{1}{\delta}}. \tag{3}$$

Proof. Let $\Delta(i,j,s,t)\triangleq\ell(\hat{y}_t,y_t;\Theta_{ij}^T,W_{ij}^T)-\ell(\hat{y}_s,y_s;\Theta^T,W^T)$, where Θ_{ij}^T,W_{ij}^T are model parameters trained on dataset S_m^{ij} for T iterations and Θ^T,W^T are model parameters trained on dataset S_m.

We first derive a bound on the permutation stability of function $f(S_m,\mathcal{D}_u)\triangleq\mathcal{L}_{\mathcal{D}_u}(\Theta,W)-\mathcal{L}_{S_m}(\Theta,W)$:

$$\begin{aligned}&\|\left(\mathcal{L}_{\mathcal{D}_u}(\Theta,W)-\mathcal{L}_{S_m}(\Theta,W)\right)-\left(\mathcal{L}_{\mathcal{D}_u}(\Theta^{ij},W^{ij})-\mathcal{L}_{S_m}(\Theta^{ij},W^{ij})\right)\|\leq\\&\frac{1}{q}\sum_{r=m+1,r\neq j}^{m+q}\|\Delta(i,j,r,r)\|+\frac{1}{q}\|\Delta(i,j,i,j)\|+\frac{1}{m}\sum_{r=1,r\neq i}^{m}\|\Delta(i,j,r,r)\|+\frac{1}{m}\|\Delta(i,j,j,i)\|\end{aligned} \tag{24}$$

where $q=n-m$.

According to Definition 5, Assumption 3 and Theorem 6, we have

$$\max_{1\leq r\leq m+q}\|\Delta(i,j,r,r)\|\leq\gamma=\alpha_1\sum_{t=1}^{T}(1+\eta\alpha_2)^{t-1}\frac{2\eta\beta}{m}$$

and Eq. (24) is bounded by

$$\begin{aligned}&\|\left(\mathcal{L}_{\mathcal{D}_u}(\Theta,W)-\mathcal{L}_{S_m}(\Theta,W)\right)-\left(\mathcal{L}_{\mathcal{D}_u}(\Theta^{ij},W^{ij})-\mathcal{L}_{S_m}(\Theta^{ij},W^{ij})\right)\|\\&\leq\frac{q-1}{q}\gamma+\frac{1}{q}B_\ell+\frac{m-1}{m}\gamma+\frac{1}{m}B_\ell\\&=\left(\frac{q-1}{q}+\frac{m-1}{m}\right)\gamma+\left(\frac{1}{q}+\frac{1}{m}\right)B_\ell\end{aligned}$$

Let $\tilde{\beta} = \left(\frac{q-1}{q} + \frac{m-1}{m}\right)\gamma + \left(\frac{1}{q} + \frac{1}{m}\right)B_\ell$. Then, the function $f(S_m, \mathcal{D}_u) = \mathcal{L}_{\mathcal{D}_u}(\Theta, W) - \mathcal{L}_{S_m}(\Theta, W)$ has transductive stability $\tilde{\beta}$. Apply Lemma 24 to $f(S_m, \mathcal{D}_u)$, equating the bound to δ

$$\exp\left(-\frac{\epsilon^2}{2\tilde{\beta}^2\Omega(m,q)}\right) = \delta$$

we get

$$\epsilon = \tilde{\beta}\sqrt{2\Omega(m,q)\log\frac{1}{\delta}}$$

Therefore, we obtain that the probability at least $1-\delta$ that

$$\mathcal{L}_{\mathcal{D}_u}(\Theta, W) - \mathcal{L}_{S_m}(\Theta, W) - \mathbb{E}\left[\mathcal{L}_{\mathcal{D}_u}(\Theta^{ij}, W^{ij}) - \mathcal{L}_{S_m}(\Theta^{ij}, W^{ij})\right] \leq \tilde{\beta}\sqrt{2\Omega(m,q)\log\frac{1}{\delta}} \quad (25)$$

According to Lemma 25 and Theorem 6, for $1 \leq i \leq m, m+1 \leq j \leq n$, we have

$$\mathbb{E}\left[\mathcal{L}_{\mathcal{D}_u}(\Theta^{ij}, W^{ij}) - \mathcal{L}_{S_m}(\Theta^{ij}, W^{ij})\right] = \mathbb{E}\left[\Delta(i,j,i,i)\right] \leq \gamma \quad (26)$$

Substitute Eq. (26) into Eq. (25), we get:

$$\mathcal{L}_{\mathcal{D}_u}(\Theta, W) \leq \mathcal{L}_{S_m}(\Theta, W) + \gamma + \tilde{\beta}\sqrt{2\Omega(m,q)\log\frac{1}{\delta}}$$

We rewrite it as

$$\mathcal{L}_{\mathcal{D}_u}(\Theta, W) - \mathcal{L}_{S_m}(\Theta, W) \leq \gamma + \left(2\gamma + \left(\frac{1}{n-m} + \frac{1}{m}\right)(B_\ell - \gamma)\right)\sqrt{2\Omega(m, n-m)\log\frac{1}{\delta}}$$

□

Lemma 10. *Consider a spectral GNN trained with m samples as $n \to \infty$. As the sample size m increases, the generalization error bound decreases at the rate $O(1/m) + O\big(\sqrt{2\log(1/\delta)/m}\big)$.*

Proof. (1) $\frac{1}{n-m}$ is neglectable compared with $\frac{1}{m}$

As $m < n$, we have $m = o(n)$.

$\frac{m}{n-m} = \frac{m}{n} \cdot \frac{1}{1-\frac{m}{n}}$ when $n \to \infty$, we have $\frac{m}{n} \to 0$ and $\frac{1}{1-\frac{m}{n}} \to 1$ as $m = o(n)$. Therefore,

$$\lim_{n\to\infty}\frac{m}{n-m} = 0, \lim_{n\to\infty}\frac{\frac{1}{n-m}}{\frac{1}{m}} = 0;$$

which indicates

$$\frac{1}{n-m} = o(\frac{1}{m})$$

(2) $\Omega(m, n-m)$ increase with m

$\Omega(m, n-m) = (n-m)^2\left(H_2(n) - H_2(n-m)\right)$ and $H_2(k) = \sum_{i=1}^{k}\frac{1}{i^2}$. So

$$H_2(n) - H_2(n-m) = \sum_{i=n-m+1}^{n}\frac{1}{i^2}$$

As

$$m \cdot \frac{1}{n^2} \leq \sum_{i=n-m+1}^{n}\frac{1}{i^2} \leq m \cdot \frac{1}{(n-m)^2},$$

we have

$$m \cdot \frac{1}{n^2} \leq H_2(n) - H_2(n-m) \leq m \cdot \frac{1}{(n-m)^2}.$$

Therefore,

$$(n-m)^2 \cdot m \cdot \frac{1}{n^2} \leq \Omega(m, n-m) \leq (n-m)^2 \cdot m \cdot \frac{1}{(n-m)^2},$$

i.e.,

$$\frac{m(n-m)^2}{n^2} \leq \Omega(m, n-m) \leq m$$

Thus,

$$\Omega(m, n-m) = O(m)$$

As $\gamma = O(\frac{1}{m})$, we have

$$\begin{aligned}
&\gamma + \left(2\gamma + \left(\frac{1}{n-m} + \frac{1}{m}\right)(B_\ell - \gamma)\right)\sqrt{2\Omega(m, n-m)\log\frac{1}{\delta}} \\
&= O(\frac{1}{m}) + \left(O(\frac{1}{m}) + \left(o(\frac{1}{m}) + \frac{1}{m}\right)\left(B_\ell - O(\frac{1}{m})\right)\right)\sqrt{2O(m)\log\frac{1}{\delta}} \\
&= O(\frac{1}{m}) + B_\ell O(\frac{1}{m})O(m^{1/2})\sqrt{2\log\frac{1}{\delta}} \\
&= O\left(\frac{1}{m} + B_\ell\sqrt{\frac{2\log(\frac{1}{\delta})}{m}}\right)
\end{aligned}$$

Thus, the total effect on bound is: $O\left(\frac{1}{m} + O(\sqrt{\frac{2\log(\frac{1}{\delta})}{m}}\right)$.

□

Proposition 11. *For a spectral GNN $\Psi_{\tilde{\sigma}}$ with a non-linear feature transformation function $f_W(X) = \tilde{\sigma}(XW)$, assume the gradient norm bound β in Theorem 9 is the same for Ψ and $\Psi_{\tilde{\sigma}}$. If Lip$(\tilde{\sigma}) \leq 1$ and Smt$(\tilde{\sigma}) \leq 1$, then $\gamma_{\tilde{\sigma}} \leq \gamma$, where $\gamma_{\tilde{\sigma}}$ is the stability of $\Psi_{\tilde{\sigma}}$.*

Proof. We consider spectral GNN Ψ:

$$\Psi(M, X) = \sigma(\sum_{k=0}^{K} \tilde{A}^k XW)$$

and spectral GNN $\Psi_{\tilde{\sigma}}$:

$$\Psi_{\tilde{\sigma}}(M, X) = \sigma(\sum_{k=0}^{K} \tilde{\sigma}\left(\tilde{A}^k XW\right))$$

(1) Lipschitz Constant:

For any two sets of parameters (Θ_1, W_1) and (Θ_2, W_2):

$$
\begin{aligned}
&\|\Psi_{\tilde{\sigma}}(\Theta_1, W_1) - \Psi_{\tilde{\sigma}}(\Theta_2, W_2)\| \\
&= \|\sigma(\sum_{i=0}^{K} \theta_{1k}\tilde{\sigma}(\tilde{A}^k X W_1)) - \sigma(\sum_{i=0}^{K} \theta_{2k}\tilde{\sigma}(\tilde{A}^k X W_2))\| \\
&\leq Lip(\sigma)\|\sum_{i=0}^{K} \theta_{1k}\tilde{\sigma}(\tilde{A}^k X W_1) - \sum_{i=0}^{K} \theta_{2k}\tilde{\sigma}(\tilde{A}^k X W_2)\| \\
&\leq Lip(\sigma)\|\sum_{i=0}^{K} (\theta_{1k} - \theta_{2k})\tilde{\sigma}(\tilde{A}^k X W_1) + \sum_{i=0}^{K} \theta_{2k}(\tilde{\sigma}(\tilde{A}^k X W_1) - \tilde{\sigma}(\tilde{A}^k X W_2))\| \\
&\leq Lip(\sigma)(\|\sum_{i=0}^{K} (\theta_{1k} - \theta_{2k})\tilde{\sigma}(\tilde{A}^k X W_1)\| + \|\sum_{i=0}^{K} \theta_{2k}(\tilde{\sigma}(\tilde{A}^k X W_1) - \tilde{\sigma}(\tilde{A}^k X W_2))\|) \\
&\leq Lip(\sigma)(\|\Theta_1 - \Theta_2\|_F \cdot \max_k \|\tilde{\sigma}(\tilde{A}^k X W_1)\|_2 + \|\Theta_2\|_F \cdot Lip(\tilde{\sigma}) \cdot \max_k \|\tilde{A}^k X (W_1 - W_2)\|_2)
\end{aligned}
$$

Since $Lip(\tilde{\sigma}) \leq 1$, we have:

$$\|\Psi_{\tilde{\sigma}}(\Theta_1, W_1) - \Psi_{\tilde{\sigma}}(\Theta_2, W_2)\| \leq Lip(\sigma)(\|\Theta_1 - \Theta_2\|_F \cdot C_1 + \|\Theta_2\|_F \cdot \|W_1 - W_2\|_F \cdot C_2)$$

where C_1, C_2 are constants depending on $X, \tilde{A}$.

The right hand side is identical to the bound we get for Ψ without activation function. Therefore, $Lip(\Psi_{\tilde{\sigma}}) \leq Lip(\Psi)$.

(2) Smoothness Constant:

We first get partial derivatives of Ψ:

$$
\begin{aligned}
\frac{\partial \Psi}{\partial \theta_k} &= \nabla\sigma(\sum_{i=0}^{K} \theta_i \tilde{A}^i X W) \cdot \tilde{A}^k X W \\
\frac{\partial \Psi_{\tilde{\sigma}}}{\partial \theta_k} &= \nabla\sigma(\sum_{i=0}^{K} \theta_i \tilde{\sigma}(\tilde{A}^i X W)) \cdot \tilde{\sigma}(\tilde{A}^k X W)
\end{aligned}
$$

Partial derivatives of $\Psi_{\tilde{\sigma}}$ are:

$$
\begin{aligned}
\frac{\partial \Psi}{\partial W} &= \nabla\sigma(\sum_{i=0}^{K} \theta_i \tilde{A}^i X W) \cdot \sum_{i=0}^{K} \theta_i \tilde{A}^i X \\
\frac{\partial \Psi_{\tilde{\sigma}}}{\partial W} &= \nabla\sigma(\sum_{i=0}^{K} \theta_i \tilde{\sigma}(\tilde{A}^i X W)) \cdot \sum_{i=0}^{K} \theta_i \nabla\tilde{\sigma}(\tilde{A}^i X W) \cdot \tilde{A}^i X
\end{aligned}
$$

The Lipschitz constant of these gradients determine the smoothness. For $\Psi_{\tilde{\sigma}}$, the additional $\tilde{\sigma}$ and $\nabla\tilde{\sigma}$ terms do not increase the Lipschitz constant of the gradient as $Lip(\tilde{\sigma}) \leq 1, Smt(\tilde{\sigma}) \leq 1$.

1) $\tilde{\sigma}$ is 1-Lipschitz, so it doesn't increase the difference between inputs. 2) $\nabla\tilde{\sigma}$ is bounded by 1 (since $Smt(\tilde{\sigma}) \leq 1$), so it doesn't amplify the gradient.

Therefore, the Lipschitz constant of the gradient of $\Psi_{\tilde{\sigma}}$ is at most equal to that of Ψ, i.e., :

$$Smt(\Psi_{\tilde{\sigma}}) \leq Smt(\Psi)$$

(3) stability $\gamma_{\tilde{\sigma}}$

According to Theorem 6, we have $\alpha_1 = Lip(\ell) \cdot Lip(\Psi)$ and $\alpha_2 = Smt(\Psi)\beta_1 + Smt(\ell)Lip(\Psi)\beta_2$. Thus, we have a smaller $\alpha_{1\tilde{\sigma}}, \alpha_{2\tilde{\sigma}}$ as $Lip(\Psi_{\tilde{\sigma}}) \leq Lip(\Psi)$ and $\Psi_{\tilde{\sigma}}) \leq Smt(\Psi)$. Then, we have $r_{\tilde{\sigma}} \leq r$.

As β is the same for $\Psi_{\tilde{\sigma}}$ and Ψ and $\gamma_{\tilde{\gamma}} = \beta r_{\tilde{\sigma}}, \gamma = \beta r$, we have

$$\gamma_{\tilde{\sigma}} \leq \gamma$$

□

E UNIFORM TRANSDUCTIVE STABILITY ON CSBM

We show the uniform transductive stability of spectral GNNs of architecture in Eq. (1) on graphs generated by $G \sim cSBM(n, f, \mu, u, \lambda, d)$. Theorem 13 is a specialized form of Theorem 8. The data model is specialized to be nodes of binary classes and Gaussian node features.

We first introduce lemmas that are closely related to calculating node features after graph convolution in Appendix E.1. Then we give the expectation and variance of element A_{ij}^k in adjacency matrix and the expectation and variance of node features after graph convolution in Appendix E.2. Based on that, we derive the transductive stability of spectral GNNs on specialized data model in Appendix E.3.

E.1 LEMMAS FOR THEOREM 13

Lemma 26 (Poisson Limit Theorem (Durrett, 2019)). *For each n, let $X_{n,m}, 1 \leq m \leq n$ be independent random variables with $\mathbb{P}[X_{n,m} = 1] = p_{n,m}, \mathbb{P}[X_{n,m} = 0] = 1 - p_{n,m}$. Suppose (1) $\sum_{m=1}^{n} p_{n,m} \to \lambda \in (0, \infty)$ and (2) $\max_{1 \leq m \leq n} p_{n,m} \to 0$, if $S_n = \sum_{m=1}^{n} X_{n,m}$, then S_n obeys the Gaussian distribution $Poisson(\lambda)$.*

Remark. The Poisson limit theorem is also know as law of rare events. It states that the total number of events will follow a Poisson distribution if the probability of occurrence of an event is small in every trail and it may occur in a large number of trials. More details can be found in (Durrett, 2019).

Lemma 27 (Binomial coefficient approximation). *When $n \gg k$, the binomial coefficient $\binom{n}{k} = \frac{n^k}{k!}$.*

Proof.

$$\begin{aligned}
\binom{n}{k} &= \frac{n!}{k!(n-k)!} \\
&= \frac{n \cdot (n-1) \cdot (n-2) \cdot \ldots \cdot (n-k+1) \cdot (n-k)!}{k! \cdot (n-k)!} \\
&= \frac{n \cdot (n-1) \cdot (n-2) \cdot \ldots \cdot (n-k+1) \cdot (n-k)!}{k! \cdot (n-k)!} \\
&= \frac{n \cdot (n-1) \cdot (n-2) \cdot \ldots \cdot (n-k+1)}{k!} \\
&= \frac{n^k}{k!}, \quad n \gg k
\end{aligned}$$

□

Lemma 28 (Expecatations of $\mathbb{E}[AB]$). *For any two random variable A, B, we have the expectations*

$$\mathbb{E}[AB] \leq \frac{1}{2}\mathbb{E}[A^2] + \frac{1}{2}\mathbb{E}[B^2]$$

Proof. Define a function $f(t)$ for any real number t:

$$f(t) = \mathbb{E}[(\frac{1}{\sqrt{2}}A - \frac{t}{\sqrt{2}}B)^2]$$

Since this is an expectation of a squared term, we have $f(t) \geq 0$ for any real t.

Expand $f(t)$:

$$f(t) = \mathbb{E}[\frac{1}{2}A^2 - tAB + \frac{t^2}{2}B^2] = \frac{1}{2}\mathbb{E}[A^2] - t\mathbb{E}[AB] + \frac{t^2}{2}\mathbb{E}[B^2]$$

set $t = 1$, we have

$$\frac{1}{2}\mathbb{E}[A^2] - \mathbb{E}[AB] + \frac{1}{2}\mathbb{E}[B^2] \geq 0$$

i.e.,

$$\mathbb{E}[AB] \leq \frac{1}{2}\mathbb{E}[A^2] + \frac{1}{2}\mathbb{E}[B^2]$$

□

Lemma 29 (Monotonicity of $g(\lambda) = \left(\left(d+\lambda\sqrt{d}\right)^k - \left(d-\lambda\sqrt{d}\right)^k\right)^2$). *The function* $g(\lambda) = \left(\left(d+\lambda\sqrt{d}\right)^k - \left(d-\lambda\sqrt{d}\right)^k\right)^2, \lambda \in [-\sqrt{d}, \sqrt{d}]$

- *monotonously increases on* $\lambda \in [0, \sqrt{d}]$*;*
- *monotonously decreases on* $\lambda \in [-\sqrt{d}, 0]$*;*
- *achieves the minimum value when* $\lambda = 0$*.*

Proof. First, observe that $g(\lambda)$ is a symmetry function. Thus, we only need to analyze its behaviour for $\lambda \geq 0$ and then mirror the results for $\lambda < 0$.

Define:

$$A = d + \lambda\sqrt{d}, \quad B = d - \lambda\sqrt{d}$$

So, the function becomes:

$$g(\lambda) = (A^k - B^k)^2$$

Compute the derivative $g'(\lambda)$:

$$g'(\lambda) = 2k\sqrt{d}(A^k - B^k)(A^{k-1} + B^{k-1})$$

When $\lambda \geq 0$, $A \geq B \geq 0$, both $(A^k - B^k)$ and $(A^{k-1} + B^{k-1})$ are non-negative, thus, $g'(\lambda) \geq 0$. This indicates that $g(\lambda)$ monotonously increases on $\lambda \in [0, \sqrt{d}]$.

Due to the even symmetry of $g(\lambda)$, $g(\lambda)$ monotonously decreases on $\lambda \in [-\sqrt{d}, 0]$.

□

Lemma 30 (Monotonicity of $g(\lambda) = \sum_{s=1}^{k} \left(d+\lambda\sqrt{d}\right)^{k-s} \left(d-\lambda\sqrt{d}\right)^s$). *The function* $g(\lambda) = \sum_{s=1}^{k} \left(d+\lambda\sqrt{d}\right)^{k-s} \left(d-\lambda\sqrt{d}\right)^s, \lambda \in [-\sqrt{d}, \sqrt{d}]$

- *monotonously decreases on* $\lambda \in [0, \sqrt{d}]$*;*
- *monotonously increases on* $\lambda \in [-\sqrt{d}, 0]$*;*
- *achieves the maximum value at* $\lambda = 0$*.*

Proof. $g(\lambda)$ can be rewritten as

$$g(\lambda) = (2d)^k - \left(d+\lambda\sqrt{d}\right)^k - \left(d-\lambda\sqrt{d}\right)^k$$

compute the first derivative of $g(\lambda)$ with respect to λ:

$$g'(\lambda) = k\sqrt{d}\left[\left(d-\lambda\sqrt{d}\right)^{k-1} - \left(d+\lambda\sqrt{d}\right)^{k-1}\right]$$

- when $\lambda > 0$, we have $\left(d-\lambda\sqrt{d}\right) < \left(d+\lambda\sqrt{d}\right)$, i.e., $g'(\lambda) < 0$, thus, $g(\lambda)$ is strictly decreasing;
- when $\lambda < 0$, we have $\left(d-\lambda\sqrt{d}\right) > \left(d+\lambda\sqrt{d}\right)$, i.e., $g'(\lambda) > 0$, thus, $g(\lambda)$ is strictly increasing;
- when $\lambda = 0$, $g'(\lambda) > 0$, $g(\lambda)$ achieves the maximum value.

□

E.2 Expectation and Variance of A_{ij}^k and $\left(\tilde{A}^k XW\right)_{ij}$

Theorem 31 (Expectation and variance of A_{ij}^k)**.** *Given a graph generated by $G \sim cSBM(n, f, \mu, u, \lambda, d)$. When $n \to \infty, d \ll n, 2 \le k \le k^2 \ll n$, the k-length walk connecting node v_i, v_j obeys Poisson distribution $Poisson(\rho')$, where*

$$\rho' = \begin{cases} \rho_= = \frac{(k-1)!}{n \cdot 2^{k-1}} \sum_{a=2}^{k+1} O\left(\sum_{s=\min(2,2(a-2),2(k+1-a))}^{\min(2(a-1),2(k+1-a))} c_{in}^{k-s} \cdot c_{out}^{s} \right), & \text{if} \quad y_i = y_j \\ \rho_\neq = \frac{(k-1)!}{n \cdot 2^{k-1}} \sum_{a=1}^{k} O\left(\sum_{s=1}^{\min(2a-1,2(k-a)+1)} c_{in}^{k-s} \cdot c_{out}^{s} \right), & \text{if} \quad y_i \neq y_j \end{cases}$$

and expectation is $\mathbb{E}\left[A_{ij}^k\right] = \rho'$, variance is $\mathbb{V}\left[A_{ij}^k\right] = \rho'$.

When $k = 1$, the walk connecting node v_i, v_j obeys Bernoulli distribution $Ber(p)$, where

$$p = \begin{cases} p_= = \frac{c_{in}}{n}, & \text{if} \quad y_i = y_j \\ p_\neq = \frac{c_{out}}{n}, & \text{if} \quad y_i \neq y_j \end{cases}$$

and expectation is $\mathbb{E}\left[A_{ij}^k\right] = p$, variance is $\mathbb{V}\left[A_{ij}^k\right] = p(1-p)$.

Proof. According to Definition 7, we have

$$\mathbb{E}[A_{ij}^k] = \sum_{p \in P_{ij}^k} \prod_{(v,v') \in p} Q_{yy'}$$

When $C = 2$, we have

$$Q_{yy'} = \begin{cases} \frac{c_{in}}{n}, & \text{if} \quad y = y' \\ \frac{c_{out}}{n}, & \text{if} \quad y \neq y' \end{cases} \tag{27}$$

Case 1: $y_i = y_j$ and $k \ge 2$

For nodes v_i and v_j sharing the same class y_i, we consider walks of length k that include a nodes sharing the class y_i and $k + 1 - a$ nodes with different classes.

As we start at v_i and end at v_j, both with class y_i, we need to choose $a - 2$ nodes from the same cluster and $k - a$ nodes from the other cluster.

The total number of ways to arrange these nodes in a walk is $(k-1)!$ as we have $k-1$ positions to fill. The probability of each edge depends on whether it's connecting same-class or different-class nodes.

Number of ways to choose the nodes:

- Choose $a - 2$ nodes from $(\frac{n}{2} - 2)$ nodes in the same cluster: $\binom{\frac{n}{2}-2}{a-2}$;
- Choose $k - a + 1$ nodes from $\frac{n}{2}$ nodes in the other cluster: $\binom{\frac{n}{2}}{k-a+1}$.

Number of ways to arrange these nodes: $(k-1)!$.

Consider the class change of the k-length walk, we denote s the number of walk of class change, when $2a \ge k+1$, we have $s_{min} = \min(2, 2(k+1-a)), s_{max} = 2(k+1-a)$; when $2a \le k+1$, we have $s_{min} = \min(2, 2(a-2)), s_{max} = 2(a-1)$.

Therefore, the probability that there are walk of length k and a nodes on walk sharing same class with v_i is:

$$p_k^a(v_i, v_j \mid y_i = y_j) =$$
$$\begin{cases} \binom{\frac{n}{2}-2}{a-2} \cdot \binom{\frac{n}{2}}{k-a+1} \cdot (k-1)! \cdot \left(\sum_{s=\min(2,2(k+1-a))}^{2(k+1-a)} \left(\frac{c_{in}}{n}\right)^{k-s} \cdot \left(\frac{c_{out}}{n}\right)^{s} \right), \text{if} \quad 2a \ge k+1; \\ \binom{\frac{n}{2}-2}{a-2} \cdot \binom{\frac{n}{2}}{k-a+1} \cdot (k-1)! \cdot \left(\sum_{s=\min(2,2(a-2))}^{2(a-1)} \left(\frac{c_{in}}{n}\right)^{k-s} \cdot \left(\frac{c_{out}}{n}\right)^{s} \right), \text{if} \quad 2a < k+1; \end{cases}$$

The probability that there are walk of length k connecting node v_i, v_j and $y_i = y_j$ is:

$$
\begin{aligned}
&p_k(v_i, v_j | y_i = y_j) = \\
&\sum_{a=2}^{\frac{k+1}{2}} \binom{\frac{n}{2}-2}{a-2} \cdot \binom{\frac{n}{2}}{k-a+1} \cdot (k-1)! \cdot \left(\sum_{s=\min(2,2(a-2))}^{2(a-1)} \left(\frac{c_{in}}{n}\right)^{k-s} \cdot \left(\frac{c_{out}}{n}\right)^{s} \right) \\
&+ \sum_{\frac{k+1}{2}}^{k+1} \binom{\frac{n}{2}-2}{a-2} \cdot \binom{\frac{n}{2}}{k-a+1} \cdot (k-1)! \cdot \left(\sum_{s=\min(2,2(k+1-a))}^{2(k+1-a)} \left(\frac{c_{in}}{n}\right)^{k-s} \cdot \left(\frac{c_{out}}{n}\right)^{s} \right)
\end{aligned} \tag{28}
$$

When $k \ll n$, using Lemma 27, binomial coefficients

$$
\binom{\frac{n}{2}-2}{a-2} = \frac{\left(\frac{n}{2}-2\right)^{a-2}}{(a-2)!}
$$

$$
\binom{\frac{n}{2}}{k-a+1} = \frac{\left(\frac{n}{2}\right)^{k-a+1}}{(k-a+1)!}
$$

Then,

$$
\begin{aligned}
\binom{\frac{n}{2}-2}{a-2} \cdot \binom{\frac{n}{2}}{k-a+1} \cdot (k-1)! &= \frac{\left(\frac{n}{2}-2\right)^{a-2}}{(a-2)!} \cdot \frac{\left(\frac{n}{2}\right)^{k-a+1}}{(k-a+1)!} \cdot (k-1)! \\
&= O\left(\left(\frac{n}{2}\right)^{k-1} \cdot \binom{k-1}{a-2} \right)
\end{aligned}
$$

Then we can simplify Eq. (28) to

$$
\begin{aligned}
p_k(v_i, v_j | y_i = y_j) &= \sum_{a=2}^{\frac{k+1}{2}} O\left(\left(\frac{n}{2}\right)^{k-1} \cdot \binom{k-1}{a-2} \right) \cdot \left(\sum_{s=\min(2,2(a-2))}^{2(a-1)} \left(\frac{c_{in}}{n}\right)^{k-s} \cdot \left(\frac{c_{out}}{n}\right)^{s} \right) \\
&\quad + \sum_{\frac{k+1}{2}}^{k+1} O\left(\left(\frac{n}{2}\right)^{k-1} \cdot \binom{k-1}{a-2} \right) \cdot \left(\sum_{s=\min(2,2(k+1-a))}^{2(k+1-a)} \left(\frac{c_{in}}{n}\right)^{k-s} \cdot \left(\frac{c_{out}}{n}\right)^{s} \right) \\
&= \frac{1}{n \cdot 2^{k-1}} \sum_{a=2}^{\frac{k+1}{2}} O\left(\binom{k-1}{a-2} \cdot \left(\sum_{s=\min(2,2(a-2))}^{2(a-1)} c_{in}^{k-s} \cdot c_{out}^{s} \right) \right) \\
&\quad + \frac{1}{n \cdot 2^{k-1}} \sum_{\frac{k+1}{2}}^{k+1} O\left(\binom{k-1}{a-2} \cdot \left(\sum_{s=\min(2,2(k+1-a))}^{2(k+1-a)} c_{in}^{k-s} \cdot c_{out}^{s} \right) \right) \\
&= \frac{(k-1)!}{n \cdot 2^{k-1}} \sum_{a=2}^{k+1} O\left(\sum_{s=\min(2,2(a-2),2(k+1-a))}^{\min(2(a-1),2(k+1-a))} c_{in}^{k-s} \cdot c_{out}^{s} \right)
\end{aligned} \tag{29}
$$

Case 2: $y_i \neq y_j$ **and** $k \geq 2$

For node v_i, v_j, when they have different classes $y_i \neq y_j$, we count the walk number that the walk has length k and there are a nodes of same class of y_i and $k+1-a$ nodes of different class of y_i.

We need to choose $a-1$ nodes from the same cluster as v_i and $k-a$ nodes from the cluster of v_j as $y_i \neq y_j$.

The total number of ways to arrange these nodes in a walk is $(k-2)!$ as we have $k-2$ positions to fill.

Number of ways to choose the nodes:

- Choose $a-1$ nodes from $(\frac{n}{2}-1)$ nodes in the same cluster as v_i: $\binom{\frac{n}{2}-1}{a-1}$;

- Choose $k-a$ nodes from $(\frac{n}{2}-1)$ nodes in the same cluster as v_j: $\binom{\frac{n}{2}-1}{k-a}$.

Number of ways to arrange these nodes: $(k-1)!$.

Consider the class change of the k-length walk, we denote s the number of walk of class change, when $2a \geq k+1$, we have $s_{min}=1, s_{max}=2(k-a)+1$; when $2a \leq k+1$, we have $s_{min}=1, s_{max}=2a-1$.

Therefore, the probability that there are walk of length k and a nodes on walk sharing same class with v_i is

$$p_k^a(v_i,v_j|y_i \neq y_j) = \begin{cases} \binom{\frac{n}{2}-1}{a-1}\cdot\binom{\frac{n}{2}-1}{k-a}\cdot(k-1)!\cdot\left(\sum_{s=1}^{2(k-a)+1}\left(\frac{c_{in}}{n}\right)^{k-s}\cdot\left(\frac{c_{out}}{n}\right)^s\right), & \text{if} \quad 2a \geq k+1 \\ \binom{\frac{n}{2}-1}{a-1}\cdot\binom{\frac{n}{2}-1}{k-a}\cdot(k-1)!\cdot\left(\sum_{s=1}^{2a-1}\left(\frac{c_{in}}{n}\right)^{k-s}\cdot\left(\frac{c_{out}}{n}\right)^s\right), & \text{if} \quad 2a < k+1 \end{cases}$$

The probability that there are walk of length k connecting node v_i, v_j and $y_i \neq y_j$ is

$$\begin{aligned} &p_k(v_i,v_j|y_i \neq y_j) = \\ &\sum_{a=1}^{\frac{k+1}{2}}\binom{\frac{n}{2}-1}{a-1}\cdot\binom{\frac{n}{2}-1}{k-a}\cdot(k-1)!\cdot\left(\sum_{s=1}^{2a-1}\left(\frac{c_{in}}{n}\right)^{k-s}\cdot\left(\frac{c_{out}}{n}\right)^s\right) \\ &+\sum_{a=\frac{k+1}{2}}^{k}\binom{\frac{n}{2}-1}{a-1}\cdot\binom{\frac{n}{2}-1}{k-a}\cdot(k-1)!\cdot\left(\sum_{s=1}^{2(k-a)+1}\left(\frac{c_{in}}{n}\right)^{k-s}\cdot\left(\frac{c_{out}}{n}\right)^s\right) \end{aligned} \tag{30}$$

When $k \ll n$, using Lemma 27, we have $\binom{\frac{n}{2}-1}{a-1} = \frac{(\frac{n}{2}-1)^{a-1}}{(a-1)!}, \binom{\frac{n}{2}-1}{k-a} = \frac{(\frac{n}{2}-1)^{k-a}}{(k-a)!}$.

Then:

$$\begin{aligned} \binom{\frac{n}{2}-1}{a-1}\cdot\binom{\frac{n}{2}-1}{k-a}\cdot(k-1)! &= \frac{(\frac{n}{2}-1)^{a-1}}{(a-1)!}\cdot\frac{(\frac{n}{2}-1)^{k-a}}{(k-a)!}\cdot(k-1)! \\ &= \left(\frac{n}{2}-1\right)^{k-1}\cdot\binom{k-1}{a-1} \end{aligned}$$

We simplify Eq. (30) to

$$\begin{aligned} p_k(v_i,v_j|y_i \neq y_j) =& \sum_{a=1}^{\frac{k+1}{2}}\left(\frac{n}{2}-1\right)^{k-1}\cdot\binom{k-1}{a-1}\cdot\left(\sum_{s=1}^{2a-1}\left(\frac{c_{in}}{n}\right)^{k-s}\cdot\left(\frac{c_{out}}{n}\right)^s\right) \\ &+\sum_{a=\frac{k+1}{2}}^{k}\left(\frac{n}{2}-1\right)^{k-1}\cdot\binom{k-1}{a-1}\cdot\left(\sum_{s=1}^{2(k-a)+1}\left(\frac{c_{in}}{n}\right)^{k-s}\cdot\left(\frac{c_{out}}{n}\right)^s\right) \\ &= \frac{1}{n\cdot 2^{k-1}}\sum_{a=1}^{\frac{k+1}{2}}O\left(\binom{k-1}{a-1}\cdot\left(\sum_{s=1}^{2a-1}c_{in}^{k-s}\cdot c_{out}^s\right)\right) \\ &+\frac{1}{n\cdot 2^{k-1}}\sum_{a=\frac{k+1}{2}}^{k}O\left(\binom{k-1}{a-1}\cdot\left(\sum_{s=1}^{2(k-a)+1}c_{in}^{k-s}\cdot c_{out}^s\right)\right) \\ &= \frac{(k-1)!}{n\cdot 2^{k-1}}\sum_{a=1}^{k}O\left(\sum_{s=1}^{\min(2a-1,2(k-a)+1)}c_{in}^{k-s}\cdot c_{out}^s\right) \end{aligned} \tag{31}$$

Case 3: $k=1$
When $k=1$, $A^k=A$,

$$\mathbb{E}\left[A_{ij}\right] = \begin{cases} \frac{c_{in}}{n}; \text{if} \quad y_i = y_j \\ \frac{c_{out}}{n}; \text{if} \quad y_i \neq y_j \end{cases}$$

In the following, we prove that when a graph is sparse and k is small, A_{ij}^k can be modeled with Poisson Distribution.

(1) when a sparse graph contains a large number of nodes $n \to \infty, d \ll n$, the potential k-length walk has a low probability of existing; (2) when $k \ll n$, the dependence between two different k-length walks is negligible; (3) the number of potential k-length walk is large $n^{k-1}, n \to \infty$. Thus, according to Lemma 26, the k-length walk connecting node v_i, v_j obeys the Poisson distribution $Poisson(\rho')$ when $k \geq 2$ where

$$\rho' = \begin{cases} \rho_{=} = \frac{(k-1)!}{n \cdot 2^{k-1}} \sum_{a=2}^{k+1} O\left(\sum_{s=\min(2,2(a-2),2(k+1-a))}^{\min(2(a-1),2(k+1-a))} c_{in}^{k-s} \cdot c_{out}^{s} \right), & \text{if} \quad y_i = y_j \\ \rho_{\neq} = \frac{(k-1)!}{n \cdot 2^{k-1}} \sum_{a=1}^{k} O\left(\sum_{s=1}^{\min(2a-1,2(k-a)+1)} c_{in}^{k-s} \cdot c_{out}^{s} \right), & \text{if} \quad y_i \neq y_j \end{cases}$$

When $k = 1$, $p(v_i, v_j)$ obeys the Bernoulli distribution $Ber(p)$ that

$$p = \begin{cases} \frac{c_{in}}{n}, & \text{if} \quad y_i = y_j \\ \frac{c_{out}}{n}, & \text{if} \quad y_i \neq y_j \end{cases}$$

□

Theorem 32 (Expectation and variance of $(\tilde{A}^k XW)_{ij}$)**.** *Given a graph generated by $G \sim cSBM(n, f, \mu, u, \lambda, d)$. The input node feature matrix is X and the normalized adjacency matrix is $\tilde{A}$. The k-th power matrix $\tilde{A}^k$ is applied to obtain a new feature matrix $\tilde{A}^k XW$, then the expectation and the variance of $(\tilde{A}^k XW)_{ij}$ are as follows:*

For $k = 1$:

$$\mathbb{E}\left[(\tilde{A}^k XW)_{ij}\right] = \frac{1}{2d}\sqrt{\frac{\mu}{n}}\left(c_{in} - c_{out}\right) y_i u W_{:j}$$

$$\mathbb{V}\left[(\tilde{A}^k XW)_{ij}\right] = \frac{1}{2 \cdot d^2}\left(d - \frac{c_{in}^2 + c_{out}^2}{n}\right) \cdot \left(\frac{\mu}{n}\left(uW_{:j}\right)^2 + \frac{||W_{:j}||_2^2}{f}\right)$$

For $k \geq 2$:

$$\mathbb{E}\left[(\tilde{A}^k XW)_{ij}\right] = \frac{(k-1)!}{d^k \cdot 2^{k-1}} O\left(c_{in}^k - c_{out}^k\right) \sqrt{\frac{\mu}{n}} y_i u W_{:j}$$

$$\mathbb{V}\left[(\tilde{A}^k XW)_{ij}\right] = \frac{(k-1)!}{d^{2k} \cdot 2^k}\left(\sum_{a=2}^{k+1} O\left(\sum_{s=\min(2,2(a-2),2(k+1-a))}^{\min(2(a-1),2(k+1-a))} c_{in}^{k-s} \cdot c_{out}^{s}\right) + \sum_{a=1}^{k} O\left(\sum_{s=1}^{\min(2a-1,2(k-a)+1)} c_{in}^{k-s} \cdot c_{out}^{s}\right)\right)\left(\frac{\mu}{n}\left(uW_{:j}\right)^2 + \frac{||W_{:j}||_2^2}{f}\right)$$

Proof. Given that the node feature x_i for node v_i, generated by a conditional Stochastic Block Model (cSBM) conditioned on u and node class y_i, is distributed as:

$$x_i \sim \mathcal{N}\left(\sqrt{\frac{\mu}{n}} y_i u, \frac{I_f}{f}\right)$$

For a linear transformation matrix W, the transformed node feature is given by:

$$x_i W \sim \mathcal{N}\left(\sqrt{\frac{\mu}{n}} y_i u W, \frac{W^T W}{f}\right)$$

Feature after transformation with W and propagation with $\tilde{A}^k$ is

$$\begin{aligned}\left(\tilde{A}^k XW\right)_{ij} &= \sum_{r=1}^{n} \tilde{A}^k_{ir} (XW)_{rj} \\ &= \sum_{r=1}^{n} \tilde{A}^k_{ir} \left(\sqrt{\frac{\mu}{n}} y_r u W_{:j} + \frac{\epsilon_r W_{:j}}{\sqrt{f}}\right) \\ &= \sum_{r=1}^{n} \tilde{A}^k_{ir} \sqrt{\frac{\mu}{n}}\, y_r u W_{:j}\end{aligned}$$

and

$$\mathbb{E}\left[\left(\tilde{A}^k XW\right)_{ij}\right] = \sqrt{\frac{\mu}{n}} \left(\sum_{r=1}^{n} \mathbb{E}\left[\tilde{A}^k_{ir}\right] y_r\right) u W_{:j} \tag{32}$$

We now derive the expectation $\mathbb{E}[A^k_{ij}]$ of the adjacency matrix A raised to the power k.

1. Expectation $\mathbb{E}\left[\left(\tilde{A}^k XW\right)_{ij}\right]$ when $k \geq 2$

Two clusters generated by cSBM are in equal size. According to Theorem 31, we have

$$\begin{aligned}\mathbb{E}\left[\left(\tilde{A}^k XW\right)_{ij}\right] &= \sqrt{\frac{\mu}{n}} \left(\sum_{r=1}^{n} \mathbb{E}\left[\tilde{A}^k_{ir}\right] y_r\right) u W_{:j} \\ &= \frac{1}{d^k} \sqrt{\frac{\mu}{n}} \left(\sum_{r=1}^{n} \left(\mathbb{E}\left[A^k_{ir} | y_i = y_r\right] + \mathbb{E}\left[A^k_{ir} | y_i \neq y_r\right]\right) y_r\right) u W_{:j} \\ &= \frac{1}{d^k} \sqrt{\frac{\mu}{n}} \Bigg(\sum_{r=1}^{n} \Bigg(\frac{(k-1)!}{n \cdot 2^{k-1}} \sum_{a=2}^{k+1} O\left(\sum_{s=\min(2,2(a-2),2(k+1-a))}^{\min(2(a-1),2(k+1-a))} c_{in}^{k-s} \cdot c_{out}^{s} \right) \\ &\quad + \frac{(k-1)!}{n \cdot 2^{k-1}} \sum_{a=1}^{k} O\left(\sum_{s=1}^{\min(2a-1,2(k-a)+1)} c_{in}^{k-s} \cdot c_{out}^{s} \right) \Bigg) y_r \Bigg) u W_{:j} \\ &= \frac{(k-1)!}{d^k \cdot 2^{k-1}} O\Bigg(\sum_{a=2}^{k+1} \sum_{s=\min(2,2(a-2),2(k+1-a))}^{\min(2(a-1),2(k+1-a))} c_{in}^{k-s} \cdot c_{out}^{s} \\ &\quad - \sum_{a=1}^{k} \sum_{s=1}^{\min(2a-1,2(k-a)+1)} c_{in}^{k-s} \cdot c_{out}^{s} \Bigg) \sqrt{\frac{\mu}{n}} y_i u W_{:j} \\ &= \frac{(k-1)!}{d^k \cdot 2^{k-1}} O\left(c_{in}^{k} - c_{out}^{k}\right) \sqrt{\frac{\mu}{n}} y_i u W_{:j}\end{aligned}$$

2. Variance $\mathbb{E}\left[\left(\tilde{A}^k XW\right)_{ij}\right]$ when $k \geq 2$

The variance of new feature X'_{ij} given u, Y can be expressed as:

$$
\begin{aligned}
&\mathbb{V}\left[(\tilde{A}^k XW)_{ij}\right] = \mathbb{V}\left[\sum_{r=1}^{n} \tilde{A}^k_{ir}(\sqrt{\frac{\mu}{n}} y_r u W_{:j} + \frac{\epsilon_r W_{:j}}{\sqrt{f}})\right] \\
&= \sum_{r=1}^{n} \mathbb{V}\left[\tilde{A}^k_{ir}(\sqrt{\frac{\mu}{n}} y_r u W_{:j} + \frac{\epsilon_r W_{:j}}{\sqrt{f}})\right], \quad \text{feature dimension independent} \\
&= \sum_{r=1}^{n}\left[\mathbb{E}\left[(\tilde{A}^k)^2_{ir}\right]\mathbb{E}\left[(\sqrt{\frac{\mu}{n}} y_r u W_{:j} + \frac{\epsilon_r W_{:j}}{\sqrt{f}})^2\right] - \left(\mathbb{E}\left[\tilde{A}^k_{ir}\right]\right)^2\left(\mathbb{E}\left[\sqrt{\frac{\mu}{n}} y_r u W_{:j} + \frac{\epsilon_r W_{:j}}{\sqrt{f}}\right]\right)^2\right] \\
&= \sum_{r=1}^{n}\left[\mathbb{E}\left[(\tilde{A}^k)^2_{ir}\right]\left(\left(\sqrt{\frac{\mu}{n}} y_r u W_{:j}\right)^2 + \frac{||W_{:j}||_2^2}{f}\right) - \left(\mathbb{E}\left[\tilde{A}^k_{ir}\right]\right)^2(\mathbb{E}\left[\sqrt{\frac{\mu}{n}} y_r u W_{:j} + \frac{\epsilon_r W_{:j}}{\sqrt{f}}\right])^2\right] \\
&= \sum_{r=1}^{n}\left[\mathbb{E}\left[(\tilde{A}^k)^2_{ir}\right]\left(\left(\sqrt{\frac{\mu}{n}} y_r u W_{:j}\right)^2 + \frac{||W_{:j}||_2^2}{f}\right) - \left(\mathbb{E}\left[\tilde{A}^k_{ir}\right]\right)^2\left(\sqrt{\frac{\mu}{n}} y_r u W_{:j}\right)^2\right] \\
&= \sum_{r=1}^{n}\left[\left(\left(\mathbb{E}\left[\tilde{A}^k_{ir}\right]\right)^2 + \mathbb{V}\left[\tilde{A}^k_{ir}\right]\right)\cdot\left(\left(\sqrt{\frac{\mu}{n}} y_r u W_{:j}\right)^2 + \frac{||W_{:j}||_2^2}{f}\right)\right. \\
&\quad \left. - \left(\mathbb{E}\left[\tilde{A}^k_{ir}\right]\right)^2\left(\sqrt{\frac{\mu}{n}} y_r u W_{:j}\right)^2\right] \\
&= \frac{1}{d^{2k}}\sum_{r=1}^{n}\left[\left(\left(\mathbb{E}\left[A^k_{ir}\right]\right)^2 + \mathbb{V}\left[A^k_{ir}\right]\right)\cdot\left(\frac{\mu}{n}(uW_{:j})^2 + \frac{||W_{:j}||_2^2}{f}\right)\right. \\
&\quad \left. - \left(\mathbb{E}\left[A^k_{ir}\right]\right)^2\frac{\mu}{n}(uW_{:j})^2\right] \\
&= \frac{1}{d^{2k}}\sum_{r=1}^{n}\left[\left(\mathbb{E}\left[A^k_{ir}\right]\right)^2\cdot\frac{||W_{:j}||_2^2}{f} + \mathbb{V}\left[A^k_{ir}\right]\cdot\left(\frac{\mu}{n}(uW_{:j})^2 + \frac{||W_{:j}||_2^2}{f}\right)\right] \\
&= \frac{1}{d^{2k}}\frac{n}{2}\left(\left(\mathbb{E}\left[A^k_{ir}|y_i = y_r\right]\right)^2 + \left(\mathbb{E}\left[A^k_{ir}|y_i \neq y_r\right]\right)^2\right)\cdot\frac{||W_{:j}||_2^2}{f} \\
&\quad + \frac{1}{d^{2k}}\frac{n}{2}\left(\mathbb{V}\left[A^k_{ir}|y_i = y_r\right] + \mathbb{V}\left[A^k_{ir}|y_i \neq y_r\right]\right)\cdot\left(\frac{\mu}{n}(uW_{:j})^2 + \frac{||W_{:j}||_2^2}{f}\right)
\end{aligned}
\tag{33}
$$

According to Theorem 31, when $k \geq 2$, we have

$$
\begin{aligned}
\left(\mathbb{E}\left[A^k_{ij}|y_i = y_j\right]\right)^2 &= \left(\frac{(k-1)!}{n\cdot 2^{k-1}}\sum_{a=2}^{k+1} O\left(\sum_{s=\min(2,2(a-2),2(k+1-a))}^{\min(2(a-1),2(k+1-a))} c_{in}^{k-s}\cdot c_{out}^{s}\right)\right)^2 \\
\left(\mathbb{E}\left[A^k_{ij}|y_i \neq y_j\right]\right)^2 &= \left(\frac{(k-1)!}{n\cdot 2^{k-1}}\sum_{a=1}^{k} O\left(\sum_{s=1}^{\min(2a-1,2(k-a)+1)} c_{in}^{k-s}\cdot c_{out}^{s}\right)\right)^2
\end{aligned}
$$

Two clusters generated by cSBM are in equal size. Then, Eq. (33) is written as:

$$
\begin{aligned}
\mathbb{V}\left[(\tilde{A}^k XW)_{ij}\right] &= \frac{1}{d^{2k}}\frac{n}{2}\left(\left(\mathbb{E}\left[A_{ir}^k|y_i=y_r\right]\right)^2+\left(\mathbb{E}\left[A_{ir}^k|y_i\neq y_r\right]\right)^2\right)\cdot\frac{||W_{:j}||_2^2}{f} \\
&+\frac{1}{d^{2k}}\frac{n}{2}\left(\mathbb{V}\left[A_{ir}^k|y_i=y_r\right]+\mathbb{V}\left[A_{ir}^k|y_i\neq y_r\right]\right)\cdot\left(\frac{\mu}{n}\left(uW_{:j}\right)^2+\frac{||W_{:j}||_2^2}{f}\right) \\
&= \frac{\left((k-1)!\right)^2}{n\cdot d^{2k}\cdot 2^{2k-1}}O\Bigg(\left(\sum_{a=2}^{k+1}O\left(\sum_{s=\min(2,2(a-2),2(k+1-a))}^{\min(2(a-1),2(k+1-a))}c_{in}^{k-s}\cdot c_{out}^s\right)\right)^2 \\
&+\left(\sum_{a=1}^{k}O\left(\sum_{s=1}^{\min(2a-1,2(k-a)+1)}c_{in}^{k-s}\cdot c_{out}^s\right)\right)^2\Bigg)\cdot\frac{||W_{:j}||_2^2}{f} \\
&+\frac{(k-1)!}{d^{2k}\cdot 2^k}\Bigg(\sum_{a=2}^{k+1}O\left(\sum_{s=\min(2,2(a-2),2(k+1-a))}^{\min(2(a-1),2(k+1-a))}c_{in}^{k-s}\cdot c_{out}^s\right) \\
&+\sum_{a=1}^{k}O\left(\sum_{s=1}^{\min(2a-1,2(k-a)+1)}c_{in}^{k-s}\cdot c_{out}^s\right)\Bigg)\left(\frac{\mu}{n}\left(uW_{:j}\right)^2+\frac{||W_{:j}||_2^2}{f}\right) \\
&= \frac{(k-1)!}{d^{2k}\cdot 2^k}\Bigg(\sum_{a=2}^{k+1}O\left(\sum_{s=\min(2,2(a-2),2(k+1-a))}^{\min(2(a-1),2(k+1-a))}c_{in}^{k-s}\cdot c_{out}^s\right) \\
&+\sum_{a=1}^{k}O\left(\sum_{s=1}^{\min(2a-1,2(k-a)+1)}c_{in}^{k-s}\cdot c_{out}^s\right)\Bigg)\left(\frac{\mu}{n}\left(uW_{:j}\right)^2+\frac{||W_{:j}||_2^2}{f}\right), \quad n\to\infty
\end{aligned}
$$

3. Expectation and variance of $\left(\tilde{A}^k XW\right)_{ij}$ when $k=1$

$$
\begin{aligned}
\mathbb{E}\left[(\tilde{A}XW)_{ij}\right] &= \sqrt{\frac{\mu}{n}}\left(\sum_{r=1}^{n}\mathbb{E}\left[\tilde{A}_{ir}\right]y_r\right)uW_{:j} \\
&= \frac{1}{d}\sqrt{\frac{\mu}{n}}\left(\sum_{r=1}^{n}\mathbb{E}\left[A_{ir}|y_i=y_r\right]y_i-\sum_{r=1}^{n}\mathbb{E}\left[A_{ir}|y_i\neq y_r\right]y_i\right)uW_{:j} \\
&= \frac{1}{d}\sqrt{\frac{\mu}{n}}\left(\frac{n}{2}\frac{c_{in}}{n}y_i-\frac{n}{2}\frac{c_{out}}{n}y_i\right)uW_{:j} \\
&= \frac{1}{2d}\sqrt{\frac{\mu}{n}}\left(c_{in}-c_{out}\right)y_i uW_{:j}
\end{aligned}
$$

when $k=1$, we have

$$
\begin{aligned}
\left(\mathbb{E}\left[A_{ij}^k|y_i=y_j\right]\right)^2 &= \left(\frac{c_{in}}{n}\right)^2 \\
\left(\mathbb{E}\left[A_{ij}^k|y_i\neq y_j\right]\right)^2 &= \left(\frac{c_{out}}{n}\right)^2
\end{aligned}
$$

Eq. (33) is written as:

$$
\begin{aligned}
\mathbb{V}\left[(\tilde{A}XW)_{ij}\right] &= \frac{1}{d^2}\frac{n}{2}\left(\left(\mathbb{E}\left[A^k_{ir}|y_i=y_r\right]\right)^2+\left(\mathbb{E}\left[A^k_{ir}|y_i\neq y_r\right]\right)^2\right)\cdot\frac{||W_{:j}||_2^2}{f}\\
&\quad+\frac{1}{d^2}\frac{n}{2}\left(\mathbb{V}\left[A^k_{ir}|y_i=y_r\right]+\mathbb{V}\left[A^k_{ir}|y_i\neq y_r\right]\right)\cdot\left(\frac{\mu}{n}\left(uW_{:j}\right)^2+\frac{||W_{:j}||_2^2}{f}\right)\\
&= \frac{1}{d^2}\frac{n}{2}\left(\left(\frac{c_{in}}{n}\right)^2+\left(\frac{c_{out}}{n}\right)^2\right)\cdot\frac{||W_{:j}||_2^2}{f}\\
&\quad+\frac{1}{d^2}\frac{n}{2}\left(\frac{c_{in}}{n}\left(1-\frac{c_{in}}{n}\right)+\frac{c_{out}}{n}\left(1-\frac{c_{out}}{n}\right)\right)\cdot\left(\frac{\mu}{n}\left(uW_{:j}\right)^2+\frac{||W_{:j}||_2^2}{f}\right)\\
&= \frac{1}{2n\cdot d^2}\left(c_{in}^2+c_{out}^2\right)\cdot\frac{||W_{:j}||_2^2}{f}\\
&\quad+\frac{1}{2\cdot d^2}\left(d-\frac{c_{in}^2+c_{out}^2}{n}\right)\cdot\left(\frac{\mu}{n}\left(uW_{:j}\right)^2+\frac{||W_{:j}||_2^2}{f}\right)\\
&= \frac{1}{2\cdot d^2}\left(d-\frac{c_{in}^2+c_{out}^2}{n}\right)\cdot\left(\frac{\mu}{n}\left(uW_{:j}\right)^2+\frac{||W_{:j}||_2^2}{f}\right),\quad n\to\infty
\end{aligned}
$$

□

E.3 Uniform Transductive Stability on $G\sim cSBM(n,f,\mu,u,\lambda,d)$

We first give a lemma about the order of $\mathbb{E}\left[A^k_{ij}\right]$, which will be used in proof of Theorem 13.

Lemma 33 (order of $\mathbb{E}\left[A^k_{ij}\right]$)**.** *The order of* $\mathbb{E}\left[A^k_{ij}\right]$ *is* $O\left(\frac{k!\cdot d^k}{n\cdot 2^k}\right)$.

Proof. According to Theorem 31, $A^k_{ij}|y_i=y_j$ and $A^k_{ij}|y_i\neq y_j$ obeys different Poisson distributions. As

$$c_{in}^{k-s}\cdot c_{out}^s = O\left(d^k\right),$$

then,

$$
\begin{aligned}
\rho_{=} &= \frac{(k-1)!}{n\cdot 2^{k-1}}\sum_{a=2}^{k+1}O\left(\sum_{s=\min(2,2(a-2),2(k+1-a))}^{\min(2(a-1),2(k+1-a))}c_{in}^{k-s}\cdot c_{out}^s\right)\\
&= \frac{(k-1)!}{n\cdot 2^{k-1}}\sum_{a=2}^{k+1}O\left(\sum_{s=\min(2,2(a-2),2(k+1-a))}^{\min(2(a-1),2(k+1-a))}d^k\right)\\
&= \frac{(k-1)!}{n\cdot 2^{k-1}}\sum_{a=2}^{k+1}O\left(k\cdot d^k\right)\\
&= \frac{(k-1)!}{n\cdot 2^{k-1}}O\left(k^2\cdot d^k\right)\\
&= O\left(\frac{k!\cdot d^k}{n\cdot 2^k}\right)
\end{aligned}
$$

similarly, we have $\rho_{\neq}=O\left(\frac{k!\cdot d^k}{n\cdot 2^k}\right)$

□

According to Theorem 8, we prove Theorem 13 that a specific case that when graph $G\sim cSBM(n,f,\mu,u,\lambda,d)$.

Theorem 13. *Consider a spectral GNN* Ψ *parameterized by* Θ, W *trained using full-batch gradient descent for* T *iterations with a learning rate* η *on a training dataset containing* m *samples drawn from nodes on a graph* $G\sim cSBM(n,f,\mu,u,\lambda,d)$. *When* $n\to\infty$, $k\ll n$, *and* $d\ll n$, *under*

Assumptions 1, 2, and 4, for any node v_i on the graph, with probability at least $1-\epsilon$ for a constant $\epsilon \in (0,1)$, Ψ satisfies γ-uniform transductive stability, where $\gamma = r\beta$ and

$$\beta = \frac{1}{\epsilon}\left[O\left(\mathbb{E}\left[\|\hat{y}_i - y_i\|_F^2\right]\right) + O\left(\sum_{k=2}^{K}\left(\mathbb{E}\left[\left(A_{ij}^k \mid y_i = y_j\right)^2\right] + \mathbb{E}\left[\left(A_{ij}^k \mid y_i \neq y_j\right)^2\right]\right)\right)\right].$$

Proof. Any spectral GNNs in Eq. (1) with linear feature transformation function, and polynomial basis expanded on normalized graph matrix can be transformed into the format:

$$\hat{Y} = softmax(\sum_{k=0}^{K}\theta_k \tilde{A}^k XW) \tag{34}$$

where $\tilde{A} = D^{-\frac{1}{2}}AD^{-\frac{1}{2}}$ is the normalized graph adjacency matrix, D is the diagonal degree matrix. We denotes $Y \in \mathbb{R}^{n\times C}$ as the ground truth node label matrix.

When graph $G \sim cSBM(n, f, \mu, u, \lambda, d)$, the node feature

$$x_i \sim \mathcal{N}(y_i\sqrt{\mu/n}u, I_f/f)$$

Denote $B = XW$ and $S = BB^\top$, then we have

$$B_{ik} \sim \mathcal{N}(y_i\sqrt{\frac{\mu}{n}}uW_{:k}, \frac{\|W_{:k}\|_F^2}{f})$$

when $i \neq j$, B_{ik}, B_{jk} are independent, then

$$\begin{aligned}\mathbb{E}\left[S_{ij}\right] &= \sum_{k=1}^{C}\mathbb{E}\left[B_{ik}B_{kj}^\top\right]\\ &= \sum_{k=1}^{C} y_i y_j \frac{\mu}{n}\left(uW_{:k}\right)^2\\ &= y_i y_j \frac{\mu}{n}\|uW\|_F^2\end{aligned}$$

when $i = j$:

$$\mathbb{E}\left[S_{ii}\right] = \frac{\mu}{n}\|uW\|_F^2 + \frac{\|W\|_F^2}{f}$$

When node number $n \to \infty$, we have

$$\sum_{q=1,q\neq j}^{n}\mathbb{E}\left[S_{jq}\right] = \frac{n}{2}y_j^2\frac{\mu}{n}\|uW\|_F^2 + \frac{n}{2}y_j(-y_j)\frac{\mu}{n}\|uW\|_F^2 = 0$$

$$\begin{aligned}&\sum_{j=1}^{n}\sum_{q=1,q\neq j}^{n}\mathbb{E}\left[A_{ij}^k A_{iq}^k\right]\mathbb{E}\left[S_{jq}\right]\\ &= \frac{n^2}{4}\rho_{k=}^2\frac{\mu}{n}\|uW\|_F^2; \quad (y_i = y_j = y_q)\\ &+ \frac{n^2}{4}\rho_{k=}\rho_{k\neq} - \frac{\mu}{n}\|uW\|_F^2; \quad (y_i = y_j \neq y_q)\\ &+ \frac{n^2}{4}\rho_{k\neq}\rho_{k=} - \frac{\mu}{n}\|uW\|_F^2; \quad (y_i \neq y_j = y_q)\\ &+ \frac{n^2}{4}\rho_{k\neq}^2\frac{\mu}{n}\|uW\|_F^2; \quad (y_i = y_q \neq y_j)\\ &= \frac{n^2}{4}\cdot\frac{\mu}{n}\|uW\|_F^2\cdot\left(\rho_{k=}^2 - 2\rho_{k\neq}\rho_{k=} + \rho_{k\neq}^2\right)\\ &= \frac{n^2}{4}\cdot\frac{\mu}{n}\|uW\|_F^2\cdot\left(\rho_{k=} - \rho_{k\neq}\right)^2\end{aligned} \tag{35}$$

According to Theorem 31, when $k \geq 2$, $A_{ij}^k \sim Poisson(\rho_k')$, then

$$
\begin{aligned}
\mathbb{E}\left[\|\tilde{A}_{i:}^k XW\|_F^2\right] &= \mathbb{E}\left[\tilde{A}_{i:}^k XW\left(XW\right)^\top\left(\tilde{A}_{i:}^k\right)^\top\right] \\
&= \mathbb{E}\left[\tilde{A}_{i:}^k S\left(\tilde{A}_{i:}^k\right)^\top\right] \\
&= \mathbb{E}\left[\sum_{q=1}^n\sum_{j=1}^n\left(\tilde{A}_{ij}^k\tilde{A}_{iq}^k S_{jq}\right)\right] \\
&= \frac{1}{d^{2k}}\mathbb{E}\left[\sum_{q=1}^n\sum_{j=1}^n\left(A_{ij}^k A_{iq}^k S_{jq}\right)\right] \\
&= \frac{1}{d^{2k}}\sum_{q=1}^n\sum_{j=1}^n\mathbb{E}\left[A_{ij}^k A_{iq}^k\right]\mathbb{E}\left[S_{jq}\right] \\
&= \frac{1}{d^{2k}}\sum_{j=1}^n\mathbb{E}\left[\left(A_{ij}^k\right)^2\right]\mathbb{E}\left[S_{jj}\right] + \frac{1}{d^{2k}}\sum_{j=1}^n\sum_{q=1,q\neq j}^n\mathbb{E}\left[A_{ij}^k A_{iq}^k\right]\mathbb{E}\left[S_{jq}\right] \\
&= \frac{1}{d^{2k}}\frac{n}{2}\mathbb{E}\left[\left(A_{ij}^k\right)^2 \mid y_i = y_j\right]\mathbb{E}\left[S_{jj}\right] + \frac{1}{d^{2k}}\frac{n}{2}\mathbb{E}\left[\left(A_{ij}^k\right)^2 \mid y_i \neq y_j\right]\mathbb{E}\left[S_{jj}\right] \\
&\quad + \frac{1}{d^{2k}}\frac{n^2}{4}\cdot\frac{\mu}{n}\|uW\|_F^2\cdot\left(\rho_{k=}-\rho_{k\neq}\right)^2 \quad (Eq.\ (35)) \\
&= \frac{1}{d^{2k}}\frac{n}{2}\left(\rho_{k=}+\rho_{k=}^2+\rho_{k\neq}+\rho_{k\neq}^2\right)\left(\frac{\mu}{n}\|uW\|_F^2+\frac{\|W\|_F^2}{f}\right) \\
&\quad + \frac{1}{d^{2k}}\frac{n^2}{4}\cdot\frac{\mu}{n}\|uW\|_F^2\cdot\left(\rho_{k=}-\rho_{k\neq}\right)^2 \\
&= \frac{1}{2d^{2k}}\zeta_k\left(\mu\|uW\|_F^2+\frac{n\|W\|_F^2}{f}\right)+\frac{n\mu}{4d^{2k}}\|uW\|_F^2\cdot\left(\rho_{k=}-\rho_{k\neq}\right)^2
\end{aligned}
$$

where $\zeta_k = \rho_{k=}^2 + \rho_{k=} + \rho_{k\neq}^2 + \rho_{k\neq}$

When $k = 1$, $A_{ij} \sim Ber(p)$, then

$$
\begin{aligned}
\mathbb{E}\left[\|\tilde{A}_{i:}XW\|_F^2\right] &= \frac{1}{d^2}\frac{n}{2}\left(p_=^2+p_=(1-p_=)+p_{\neq}^2+p_{\neq}(1-p_{\neq})\right)\left(\frac{\mu}{n}\|uW\|_F^2+\frac{\|W\|_F^2}{f}\right) \\
&= \frac{1}{d^2}\frac{n}{2}\left(p_=+p_{\neq}\right)\left(\frac{\mu}{n}\|uW\|_F^2+\frac{\|W\|_F^2}{f}\right) \\
&= \frac{1}{d^2}\frac{n}{2}\frac{2d}{n}\left(\frac{\mu}{n}\|uW\|_F^2+\frac{\|W\|_F^2}{f}\right) \\
&= \frac{1}{d}\left(\frac{\mu}{n}\|uW\|_F^2+\frac{\|W\|_F^2}{f}\right)
\end{aligned}
$$

Substitute it into Eq. (15), we have

$$
\mathbb{E}\left[\left|\frac{\partial\ell(\hat{y}_i,y_i;\Theta,W)}{\partial\theta_k}\right|\right] =
$$

$$
\begin{cases}
\frac{1}{2}\left(\mathbb{E}\left[\|\hat{y}_i-y_i\|_F^2\right]+\left(\frac{\mu}{n}\|uW\|_F^2+\frac{\|W\|_F^2}{f}\right)\right), & \text{if} \quad k=0 \\
\frac{1}{2}\left(\mathbb{E}\left[\|\hat{y}_i-y_i\|_F^2\right]+\frac{1}{d}\left(\frac{\mu}{n}\|uW\|_F^2+\frac{\|W\|_F^2}{f}\right)\right), & \text{if} \quad k=1 \\
\frac{1}{2}\left(\mathbb{E}\left[\|\hat{y}_i-y_i\|_F^2\right]+\frac{1}{2d^{2k}}\zeta_k\left(\mu\|uW\|_F^2+\frac{n\|W\|_F^2}{f}\right)+\frac{n\mu}{4d^{2k}}\|uW\|_F^2\cdot\left(\rho_{k=}-\rho_{k\neq}\right)^2\right), & \text{if} \quad k\geq 2
\end{cases}
\tag{36}
$$

similarly, we have

$$\mathbb{E}\left[\|\tilde{A}_{i:}^{k}X\|_F^2\right] = \begin{cases} \frac{\mu}{n}\|u\|_F^2 + 1, & \text{if} \quad k=0 \\ \frac{1}{d}\left(\frac{\mu}{n}\|u\|_F^2 + 1\right), & \text{if} \quad k=1 \\ \frac{1}{2d^{2k}}\zeta_k\left(\mu\|u\|_F^2+1\right) + \frac{n\mu}{4d^{2k}}\|u\|_F^2 \cdot \left(\rho_{k=} - \rho_{k\neq}\right)^2, & \text{if} \quad k \geq 2 \end{cases}$$

Substitute it into Eq. (16), we have

$$\begin{aligned} \mathbb{E}\left[\|\frac{\partial \ell(\hat{y}_i, y_i; \Theta, W)}{\partial W}\|_{\ell_1}\right] &= |\theta_0|\left(f \cdot \mathbb{E}\left[\|\hat{y}_i - y_i\|_F^2\right] + C\left(\frac{\mu}{n}\|u\|_F^2 + 1\right)\right) \\ &+ |\theta_1|\left(f \cdot \mathbb{E}\left[\|\hat{y}_i - y_i\|_F^2\right] + C\frac{1}{d}\left(\frac{\mu}{n}\|u\|_F^2 + 1\right)\right) \\ &+ \sum_{k=2}^{K}\frac{1}{d^{2k}}|\theta_k|\left(f \cdot \mathbb{E}\left[\|\hat{y}_i - y_i\|_F^2\right] + C\frac{1}{2d^{2k}}\zeta_k\left(\mu\|u\|_F^2+1\right) + \frac{n\mu}{4d^{2k}}\|u\|_F^2 \cdot \left(\rho_{k=} - \rho_{k\neq}\right)^2\right) \end{aligned} \tag{37}$$

Substitute Eq. (36), Eq. (37) into Eq. (12), we have

$$\begin{aligned} \mathbb{E}\left[\|\nabla \ell(\hat{y}_i, y_i; \Theta, W)\|_F\right] &\leq \sum_{k=0}^{K}\mathbb{E}\left[\|\frac{\partial \ell(\hat{y}_i, y_i; \Theta, W)}{\partial \theta_k}\|_{\ell_1}\right] + \mathbb{E}\left[\|\frac{\partial \ell(\hat{y}_i, y_i; \Theta, W)}{\partial W}\|_{\ell_1}\right] \\ &= \frac{1}{2}\left(\mathbb{E}\left[\|\hat{y}_i - y_i\|_F^2\right] + \left(\frac{\mu}{n}\|uW\|_F^2 + \frac{\|W\|_F^2}{f}\right)\right) \\ &+ \frac{1}{2}\left(\mathbb{E}\left[\|\hat{y}_i - y_i\|_F^2\right] + \frac{1}{d}\left(\frac{\mu}{n}\|uW\|_F^2 + \frac{\|W\|_F^2}{f}\right)\right) \\ &+ \sum_{k=2}^{K}\frac{1}{2}\left(\mathbb{E}\left[\|\hat{y}_i - y_i\|_F^2\right] + \frac{1}{2d^{2k}}\zeta_k\left(\mu\|uW\|_F^2 + \frac{n\|W\|_F^2}{f}\right) + \frac{n\mu}{4d^{2k}}\|uW\|_F^2 \cdot \tilde{\zeta}_k^2\right) \\ &+ |\theta_0|\left(f \cdot \mathbb{E}\left[\|\hat{y}_i - y_i\|_F^2\right] + C\left(\frac{\mu}{n}\|u\|_F^2 + 1\right)\right) \\ &+ |\theta_1|\left(f \cdot \mathbb{E}\left[\|\hat{y}_i - y_i\|_F^2\right] + C\frac{1}{d}\left(\frac{\mu}{n}\|u\|_F^2 + 1\right)\right) \\ &+ \sum_{k=2}^{K}\frac{1}{d^{2k}}|\theta_k|\left(f \cdot \mathbb{E}\left[\|\hat{y}_i - y_i\|_F^2\right] + C\frac{1}{2d^{2k}}\zeta_k\left(\mu\|u\|_F^2+1\right) + \frac{n\mu}{4d^{2k}}\|u\|_F^2 \cdot \tilde{\zeta}_k^2\right) \end{aligned} \tag{38}$$

where $\zeta_k = \rho_{k=}^2 + \rho_{k=} + \rho_{k\neq}^2 + \rho_{k\neq}, \tilde{\zeta}_k = \rho_{k=} - \rho_{k\neq}$.

According to Lemma 33, when $n \to \infty$, we have

$$\begin{aligned} n\left(\tilde{\zeta}_k\right)^2 &= n\left(\rho_{=} - \rho_{\neq}\right)^2 \\ &= n\left(O\left(\frac{k! \cdot d^k}{n \cdot 2^k}\right)\right)^2 \\ &= nO\left(\frac{\left(k! \cdot d^k\right)^2}{n^2 \cdot 2^{2k}}\right) \\ &= O\left(\frac{\left(k! \cdot d^k\right)^2}{n \cdot 2^{2k}}\right) \\ &\to 0 \end{aligned}$$

Thus, $n\left(\tilde{\zeta}_k\right)^2$ can be neglected. Thus, we rewrite Eq. (38) as

$$
\begin{aligned}
\mathbb{E}\left[\|\nabla \ell(\hat{y}_i, y_i; \Theta, W)\|_F\right] &= \sum_{k=0}^{K} \mathbb{E}\left[\|\frac{\partial \ell(\hat{y}_i, y_i; \Theta, W)}{\partial \theta_k}\|_{\ell_1}\right] + \mathbb{E}\left[\|\frac{\partial \ell(\hat{y}_i, y_i; \Theta, W)}{\partial W}\|_{\ell_1}\right] \\
&= \frac{1}{2}\left(\mathbb{E}\left[\|\hat{y}_i - y_i\|_F^2\right] + \left(\frac{\mu}{n}\|uW\|_F^2 + \frac{\|W\|_F^2}{f}\right)\right) \\
&\quad + \frac{1}{2}\left(\mathbb{E}\left[\|\hat{y}_i - y_i\|_F^2\right] + \frac{1}{d}\left(\frac{\mu}{n}\|uW\|_F^2 + \frac{\|W\|_F^2}{f}\right)\right) \\
&\quad + \sum_{k=2}^{K} \frac{1}{2}\left(\mathbb{E}\left[\|\hat{y}_i - y_i\|_F^2\right] + \frac{1}{2d^{2k}}\zeta_k\left(\mu\|uW\|_F^2 + \frac{n\|W\|_F^2}{f}\right)\right) \\
&\quad + |\theta_0|\left(f \cdot \mathbb{E}\left[\|\hat{y}_i - y_i\|_F^2\right] + C\left(\frac{\mu}{n}\|u\|_F^2 + 1\right)\right) \\
&\quad + |\theta_1|\left(f \cdot \mathbb{E}\left[\|\hat{y}_i - y_i\|_F^2\right] + C\frac{1}{d^{2k-1}}\left(\frac{\mu}{n}\|u\|_F^2 + 1\right)\right) \\
&\quad + \sum_{k=2}^{K} \frac{1}{d^{2k}}|\theta_k|\left(f \cdot \mathbb{E}\left[\|\hat{y}_i - y_i\|_F^2\right] + C\frac{1}{2d^{2k}}\zeta_k\left(\mu\|u\|_F^2 + 1\right)\right) \\
&\leq \frac{1}{2}\left(\mathbb{E}\left[\|\hat{y}_i - y_i\|_F^2\right] + \left(\frac{\mu}{n}\|u\|_F^2 B_W^2 + \frac{B_W^2}{f}\right)\right) \\
&\quad + \frac{1}{2}\left(\mathbb{E}\left[\|\hat{y}_i - y_i\|_F^2\right] + \frac{1}{d}\left(\frac{\mu}{n}\|u\|_F^2 B_W^2 + \frac{B_W^2}{f}\right)\right) \\
&\quad + \sum_{k=2}^{K} \frac{1}{2}\left(\mathbb{E}\left[\|\hat{y}_i - y_i\|_F^2\right] + \frac{1}{2d^{2k}}\zeta_k\left(\mu\|u\|_F^2 B_W^2 + \frac{nB_W^2}{f}\right)\right) \\
&\quad + B_\Theta\left(f \cdot \mathbb{E}\left[\|\hat{y}_i - y_i\|_F^2\right] + C\left(\frac{\mu}{n}\|u\|_F^2 + 1\right)\right) \\
&\quad + B_\Theta\left(f \cdot \mathbb{E}\left[\|\hat{y}_i - y_i\|_F^2\right] + C\frac{1}{d}\left(\frac{\mu}{n}\|u\|_F^2 + 1\right)\right) \\
&\quad + \sum_{k=2}^{K} \frac{1}{d^{2k}}B_\Theta\left(f \cdot \mathbb{E}\left[\|\hat{y}_i - y_i\|_F^2\right] + C\frac{1}{2d^{2k}}\zeta_k\left(\mu\|u\|_F^2 + 1\right)\right) \\
&= \left(\frac{K+1}{2} + 2fB_\Theta + \sum_{k=2}^{K}\frac{f}{d^{2k}}B_\Theta\right)\mathbb{E}\left[\|\hat{y}_i - y_i\|_F^2\right] \\
&\quad + \left(1 + \frac{1}{d}\right)\left(\left(\frac{B_W^2}{2} + CB_\Theta\right)\frac{\mu}{n}\|u\|_F^2 + \frac{B_W^2}{2f} + CB_\Theta\right) \\
&\quad + \sum_{k=2}^{K}\frac{\zeta_k}{d^{2k}}\left(\left(\mu\|u\|_F^2 + \frac{n}{f}\right)\frac{B_W^2}{4} + \left(\mu\|u\|_F^2 + 1\right)\frac{B_\Theta}{d^{2k}}\right)
\end{aligned}
\tag{39}
$$

We write it in big O format:

$$
\mathbb{E}\left[\|\nabla \ell(\hat{y}_i, y_i; \Theta, W)\|_F\right] = O\left(\mathbb{E}\left[\|\hat{y}_i - y_i\|_F^2\right]\right) + O\left(\sum_{k=2}^{K}\zeta_k\right)
$$

where $\zeta_k = \mathbb{E}\left[\left(A_{ij}^k \mid y_i = y_j\right)^2\right] + \mathbb{E}\left[\left(A_{ij}^k \mid y_i \neq y_j\right)^2\right]$

After get the upper bound of the norm of gradient, according to Theorem 6, we have the uniform transductive stability of spectral GNNs on graphs $G \sim cSBM(n, f, \mu, u, \lambda, d)$ of two classes $C = 2$ in big O format that

$$\gamma = r\beta; \beta = \frac{1}{\epsilon}\left[O\left(\mathbb{E}\left[\|\hat{y}_i - y_i\|_F^2\right]\right) + O\left(\sum_{k=2}^{K}\left(\mathbb{E}\left[\left(A_{ij}^k \mid y_i = y_j\right)^2\right] + \mathbb{E}\left[\left(A_{ij}^k \mid y_i \neq y_j\right)^2\right]\right)\right)\right]$$

where r is the same as that in Theorem 6.

□

F Relation between Stability of Spectral GNNs, Node Class Distribution and Architecture of spectral GNNs

In this section, we fist derive the relation between parameter λ in cSBM and the edge homophilic ratio on graph. Then, we first study how the expected prediction error $\mathbb{E}\left[\|\hat{y}_i - y_i\|_F^2\right]$ and ζ_k changes with λ, K. Next, we discuss how λ, K affect the uniform transductive stability and generalization of spectral GNNs.

Proposition 12. *For a graph $G \sim cSBM(n, \mu, u, \lambda, d)$, the expected edge homophily ratio is:*

$$\mathbb{E}[H_{edge}] = \frac{d + \lambda\sqrt{d}}{2d}; \quad \mathbb{E}[H_{edge}] = \frac{c_{in}}{c_{in} + c_{out}}. \tag{4}$$

Proof. As graphs generated with cSBM contains two clusters of same size. Thus, there are $\frac{n}{2}$ nodes having same class.

The expected number of edges between nodes of the same class is:

$$\mathbb{E}[E_{\text{same}}] = \binom{\frac{n}{2}}{2} \cdot \frac{c_{in}}{n} = \frac{c_{in}(n-2)}{8}$$

The expected number of edges between nodes of different class is:

$$\mathbb{E}[E_{\text{diff}}] = \frac{n}{2} \cdot \frac{n}{2} \cdot \frac{c_{out}}{n} \cdot \frac{1}{2} = \frac{c_{out}n}{8}$$

Thus, the expectation of H_{edge} is the expected number of edges between nodes with the same class to the total expected number of edges that

$$\begin{aligned}
\mathbb{E}\left[H_{edge}\right] &= \mathbb{E}[E_{\text{same}}] + \mathbb{E}[E_{\text{diff}}] \\
&= \frac{\frac{c_{in}(n-2)}{8}}{\frac{c_{in}(n-2)}{8} + \frac{c_{out}n}{8}} \\
&= \frac{(d + \lambda\sqrt{d})(n-2)}{(d + \lambda\sqrt{d})(n-2) + (d - \lambda\sqrt{d})n} \\
&= \frac{d + \lambda\sqrt{d}}{2d}, n \to \infty.
\end{aligned}$$

We also get the expectation between the H_{edge} and c_{in}, c_{out} that

$$\begin{aligned}
\mathbb{E}\left[H_{edge}\right] &= \mathbb{E}[E_{\text{same}}] + \mathbb{E}[E_{\text{diff}}] \\
&= \frac{\frac{c_{in}(n-2)}{8}}{\frac{c_{in}(n-2)}{8} + \frac{c_{out}n}{8}} \\
&= \frac{c_{in}(n-2)}{c_{in}(n-2) + c_{out}n} \\
&= \frac{c_{in}}{c_{in} + c_{out}}, n \to \infty.
\end{aligned}$$

□

Theorem 14 ($\mathbb{E}\left[\|\hat{y}_i - y_i\|_F^2\right]$ and λ, K)**.** *Given a graph $G \sim cSBM(n, \mu, u, \lambda, d)$ and a spectral GNN of order K, $\mathbb{E}[\|\hat{y}_i - y_i\|_F^2]$ for any node v_i satisfies the following: it increases with $\lambda \in [-\sqrt{d}, 0]$, decreases with $\lambda \in [0, \sqrt{d}]$, and reaches its maximum at $\lambda = 0$; it increases with K if $\sum_{k=2}^{K} \theta_k \frac{(k-1)!}{2^{k-1}}$ grows more slowly than $\sum_{k=2}^{K} \theta_k^2 \frac{(k-1)!}{2^k}$ as K increases.*

Proof. Denote

$$Z = \sum_{k=0}^{K} \theta_k \tilde{A}^k XW; \qquad \hat{Y} = \text{softmax}(Z)$$

Then, for any node v_i with truth class y_i, we denote its prediction as

$$\hat{y}_i = \text{softmax}(Z_{i:})$$

For the binary classification $C = 2$, for a node with truth class $y_i = [1, 0]$, the predicted class

$$\hat{y}_i = [\hat{y}_1, \hat{y}_2] = \text{softmax}([Z_{i1}, Z_{i2}]) = [\sigma(Z_{i1} - Z_{i2}), 1 - \sigma(Z_{i1} - Z_{i2})]$$

where $\sigma(x) = \frac{1}{1+e^{-x}}$ is the sigmoid function.

We denote $z_i = Z_{i1} - Z_{i2}$, then

$$\hat{y}_i = [\sigma(z_i), 1 - \sigma(z_i)]$$

Thus,

$$\|\hat{y}_i - y_i\|_F^2 = (\sigma(z_i) - 1)^2 + (1 - \sigma(z_i))^2 = 2(1 - \sigma(z_i))^2$$

and

$$\mathbb{E}\left[\|\hat{y}_i - y_i\|_F^2\right] = 2\mathbb{E}\left[(1 - \sigma(z_i))^2\right]$$

As node feature $x_i \sim \mathcal{N}(y_i\sqrt{\mu/n}u, I_f/f)$, any linear combination of Gaussian variable is still Gaussian variable, then

$$z_i \sim \mathcal{N}(\mu_{z_i}, \omega_{z_i}^2)$$

where

$$\mu_{z_i} = \mathbb{E}\left[z_i\right] = \mathbb{E}\left[Z_{i1} - Z_{i2}\right] = \mathbb{E}\left[Z_{i1}\right] - \mathbb{E}\left[Z_{i2}\right]$$

As $c_{in} = d + \lambda\sqrt{d}, c_{out} = d - \lambda\sqrt{d}, \lambda \in [-\sqrt{d}, \sqrt{d}]$, we have

$$c_{in}^k - c_{out}^k = O(d^k); \quad c_{in}^k = O(d^k); \quad c_{out}^k = O(d^k) \tag{40}$$

As $u \sim \mathcal{L}(0, I_f), d \ll f$ and Θ, W are bounded according to Assumption 4, we analyze dominant terms in μ_{z_i} and $\omega_{z_i}^2$.

According to Theorem 32, we have the expectation of $(\tilde{A}^k XW)_{ij}$, then

$$\begin{aligned}
\mu_{z_i} = \mathbb{E}\left[Z_{i1}\right] - \mathbb{E}\left[Z_{i2}\right] &= \theta_0 \sqrt{\frac{\mu}{n}} y_i u \left(W_{:1} - W_{:2}\right) \\
&\quad + \theta_1 \frac{1}{2d} \sqrt{\frac{\mu}{n}} \left(c_{in} - c_{out}\right) y_i u \left(W_{:1} - W_{:2}\right) \\
&\quad + \sum_{k=2}^{K} \theta_k \frac{(k-1)!}{d^k \cdot 2^{k-1}} O\left(c_{in}^k - c_{out}^k\right) \sqrt{\frac{\mu}{n}} y_i u \left(W_{:1} - W_{:2}\right) \\
&= O\left(\sum_{k=2}^{K} \theta_k \frac{(k-1)!}{2^{k-1}}\right) \quad (Eq.\ (40))
\end{aligned} \tag{41}$$

As $\tilde{A}^k, X$ are independent and columns of X are independent, $\left(\sum_{k=0}^{K} \theta_k \tilde{A}^k X\right)_{ij}, \left(\sum_{k=0}^{K} \theta_k \tilde{A}^k X\right)_{it}$ are independent. According to Theorem 32, we

have the variance of $(\tilde{A}^k XW)_{ij}$. Then we have

$$\begin{aligned}
\omega_{z_i}^2 &= \mathbb{V}\left[Z_{i1} - Z_{i2}\right] \\
&= \mathbb{V}\left[\left(\sum_{k=0}^{K} \theta_k \tilde{A}^k X\right)_{i:} (W_{:1} - W_{:2})\right] \\
&= \mathbb{V}\left[\sum_{j=1}^{f} \left(\sum_{k=0}^{K} \theta_k \tilde{A}^k X\right)_{ij} (W_{j1} - W_{j2})\right] \\
&= \sum_{j=1}^{f} (W_{j1} - W_{j2})^2 \sum_{k=0}^{K} \theta_k^2 \mathbb{V}\left[\left(\tilde{A}^k X\right)_{ij}\right] \quad \text{(independent)} \\
&= \sum_{j=1}^{f} (W_{j1} - W_{j2})^2 \sum_{k=0}^{K} \theta_k^2 \Bigg[\frac{1}{2 \cdot d^2}\left(d - \frac{c_{in}^2 + c_{out}^2}{n}\right) \cdot \left(\frac{\mu}{n}\left(uW_{:j}\right)^2 + \frac{||W_{:j}||_2^2}{f}\right) \\
&\quad + \frac{(k-1)!}{d^{2k} \cdot 2^k}\Bigg(\sum_{a=2}^{k+1} O\left(\sum_{s=\min(2,2(a-2),2(k+1-a))}^{\min(2(a-1),2(k+1-a))} c_{in}^{k-s} \cdot c_{out}^{s}\right) \\
&\quad + \sum_{a=1}^{k} O\left(\sum_{s=1}^{\min(2a-1,2(k-a)+1)} c_{in}^{k-s} \cdot c_{out}^{s}\right)\Bigg)\left(\frac{\mu}{n}\left(uW_{:j}\right)^2 + \frac{||W_{:j}||_2^2}{f}\right)\Bigg] \\
&= O\left(\sum_{k=2}^{K} \theta_k^2 \frac{(k-1)!}{2^k}\right) \quad (Eq.\ (40))
\end{aligned} \tag{42}$$

1. $\mathbb{E}\left[\|\hat{y}_i - y_i\|_F^2\right]$ **and** λ**.**
According to Lemma 29 and Lemma 30, we know that

- μ_{z_i} monotonously decreases and $\omega_{z_i}^2$ monotonously increases on $\lambda \in [-\sqrt{d}, 0]$;
- μ_{z_i} monotonously increases and $\omega_{z_i}^2$ monotonously decreases on $\lambda \in [0, \sqrt{d}]$;
- μ_{z_i} achieves the minimum value and $\omega_{z_i}^2$ achieves the maximum value when $\lambda = 0$.

$$\mathbb{E}[(1 - \sigma(z_i))^2] = \int_{-\infty}^{\infty} (1 - \sigma(z_i))^2 \cdot \frac{1}{\sqrt{2\pi}\omega_{z_i}} e^{-\frac{(z-\mu_{z_i})^2}{2\omega_{z_i}^2}} dz_i \tag{43}$$

the integral decreases with μ_{z_i} and $\omega_{z_i}^2$, thus,

- $\mathbb{E}[(1 - \sigma(z_i))^2]$ increases on $\lambda \in [-\sqrt{d}, 0]$;
- $\mathbb{E}[(1 - \sigma(z_i))^2]$ decreases on $\lambda \in [0, \sqrt{d}]$;
- $\mathbb{E}[(1 - \sigma(z_i))^2]$ achieves the maximum value when $\lambda = 0$.

$\mathbb{E}\left[\|\hat{y}_i - y_i\|_F^2\right]$ has the same trend with $\mathbb{E}\left[(1 - \sigma(z_i))^2\right]$.

2. $\mathbb{E}\left[\|\hat{y}_i - y_i\|_F^2\right]$ **and** K**.**

we rewrite variable z that

$$z = \mu_{z_i} + \omega_{z_i} y,$$

where $y \sim \mathcal{N}(0, 1)$.

Thus, Eq. (43) is rewritten as

$$\mathbb{E}\left[(1 - \sigma(z_i))^2\right] = \int_{-\infty}^{\infty} \left(1 - \sigma\left(\mu_{z_i} + \omega_{z_i} y\right)\right)^2 \frac{1}{\sqrt{2\pi}} e^{-\frac{y^2}{2}} dy$$

(1) μ_{z_i} increases faster than $\omega_{z_i}^2$ when K increases

In this case, z will be dominated by μ_{z_i}, thus, we have

$$\begin{aligned}\mathbb{E}\left[(1-\sigma(z))^2\right] &= \int_{-\infty}^{\infty}\left(1-\sigma\left(\mu_{z_i}\right)\right)^2\frac{1}{\sqrt{2\pi}}e^{-\frac{y^2}{2}}dy \\ &= \left(1-\sigma\left(\mu_{z_i}\right)\right)^2 \\ &\leq 0.25\end{aligned}$$

(2) μ_{z_i} increases slower than $\omega_{z_i}^2$ when K increases

In this case, z will be dominated by $\omega_{z_i}y$, thus, we have

$$\begin{aligned}\mathbb{E}\left[(1-\sigma(z))^2\right] &= \int_{-\infty}^{\infty}\left(1-\sigma\left(\omega_{z_i}y\right)\right)^2\frac{1}{\sqrt{2\pi}}e^{-\frac{y^2}{2}}dy \\ &= \int_{-\infty}^{0}\left(1-0\right)\frac{1}{\sqrt{2\pi}}e^{-\frac{y^2}{2}}dy + \int_{0}^{\infty}\left(1-1\right)^2\frac{1}{\sqrt{2\pi}}e^{-\frac{y^2}{2}}dy \\ &= 0.5\end{aligned}$$

From above analysis, we know that when μ_{z_i} increases slower than $\omega_{z_i}^2$ when K increases, $\mathbb{E}\left[(1-\sigma(z))^2\right]$ tends to have a larger value towards 0.5 compared with the case that μ_{z_i} increases faster than $\omega_{z_i}^2$ when K increases with corresponding values less or equal to 0.25.

In brief, when μ_{z_i} increases slower than $\omega_{z_i}^2$ when K increases, $\mathbb{E}\left[\|\hat{y}_i - y_i\|_F^2\right]$ increases with K.

From Eq. (41), Eq. (42), we know the the dominant term of μ_{z_i} is $\sum_{k=2}^{K}\theta_k\frac{(k-1)!}{2^{k-1}}$, the dominant term of $\omega_{z_i}^2$ is $\sum_{k=2}^{K}\theta_k^2\frac{(k-1)!}{2^k}$. Thus, $\mathbb{E}\left[\|\hat{y}_i - y_i\|_F^2\right]$ increases with K if $\sum_{k=2}^{K}\theta_k\frac{(k-1)!}{2^{k-1}}$ grows slower than $\sum_{k=2}^{K}\theta_k^2\frac{(k-1)!}{2^k}$.

□

Theorem 15 (ζ_k and λ, K)**.** *Given a graph $G \sim cSBM(n,\mu,u,\lambda,d)$ and a spectral GNN of order K, ζ_k has the following properties: (1) it increases with $\lambda \in [-\sqrt{d},0]$, decreases with $\lambda \in [0,\sqrt{d}]$, and achieves its maximum value at $\lambda = 0$; (2) it increases with k as k grows, for $k \in [0,K]$.*

Proof. The proof of this theorem is incorporated into the proof of Proposition 16. See the proof of Proposition 16 for details. □

Proposition 16. *For a fixed K, γ-uniform transductive stability and generalization error bound strictly increase as λ moves from $-\sqrt{d}$ to 0, and decreases as λ moves from 0 to $\sqrt{d}$. For a fixed λ, if $\sum_{k=2}^{K}\theta_k\frac{(k-1)!}{2^{k-1}}$ grows more slowly than $\sum_{k=2}^{K}\theta_k^2\frac{(k-1)!}{2^k}$ as K increases, then γ-uniform transductive stability and generalization error bound increase with K.*

Proof. According to Theorem 6 and Theorem 13, the uniform stability of spectral GNNs depends on the upper bound of the gradient norm β, and

$$\begin{aligned}\beta = &\left(\frac{K+1}{2} + 2fB_\Theta + \sum_{k=2}^{K}\frac{f}{d^{2k}}B_\Theta\right)\mathbb{E}\left[\|\hat{y}_i - y_i\|_F^2\right] \\ &+ \left(1+\frac{1}{d}\right)\left(\left(\frac{B_W^2}{2}+CB_\Theta\right)\frac{\mu}{n}\|u\|_F^2 + \frac{B_W^2}{2f} + CB_\Theta\right) \\ &+ \sum_{k=2}^{K}\frac{\zeta_k}{d^{2k}}\left(\left(\mu\|u\|_F^2+\frac{n}{f}\right)\frac{B_W^2}{4} + \left(\mu\|u\|_F^2+1\right)\frac{B_\Theta}{d^{2k}}\right)\end{aligned}$$

where $\zeta_k = \rho_=^2 + \rho_= + \rho_\neq^2 + \rho_\neq$, and $\rho_=$ and $\rho_\neq$ are the parameters of distribution in Theorem 31.

Denote

$$\psi_y = \left(\frac{K+1}{2} + 2fB_\Theta + \sum_{k=2}^{K} \frac{f}{d^{2k}} B_\Theta\right);$$

$$\psi_1 = \sum_{k=2}^{K} \frac{\zeta_k}{d^{2k}} \left(\left(\mu\|u\|_F^2 + \frac{n}{f}\right)\frac{B_W^2}{4} + \left(\mu\|u\|_F^2 + 1\right)\frac{B_\Theta}{d^{2k}}\right).$$

We show that the terms $\mathbb{E}\left[\|\hat{y}_i - y_i\|_F^2\right], \psi_y$, and ψ_1 can all be affected by λ, K.

(1) **Term** $\mathbb{E}\left[\|\hat{y}_i - y_i\|_F^2\right]$

According to Theorem 14, the expected prediction error $\mathbb{E}\left[\|\hat{y}_i - y_i\|_F^2\right]$ strictly increases with $\lambda \in [-\sqrt{d}, 0]$ and decreases with $\lambda \in [0, \sqrt{d}]$. In addition, it increases with K when $\sum_{k=2}^{K} \theta_k \frac{(k-1)!}{2^{k-1}}$ grows slower than $\sum_{k=2}^{K} \theta_k^2 \frac{(k-1)!}{2^k}$.

(2) **Term** ψ_y

As $\psi_y = \left(\frac{K+1}{2} + \sum_{k=0}^{K} |\theta_k| f\right)$ which does not contain λ, the class distribution has no effect on ψ_y. It also increases with order K.

(3) **Terms** ψ_1

ψ_1 is closely related with $\zeta_k = \rho_=^2 + \rho_= + \rho_\neq^2 + \rho_\neq$. According to Theorem 31, we have

$$\begin{aligned}
\zeta_k &= \rho_=^2 + \rho_= + \rho_\neq^2 + \rho_\neq \\
&= \left(\frac{(k-1)!}{n \cdot 2^{k-1}} \sum_{a=2}^{k+1} O\left(\sum_{s=\min(2,2(a-2),2(k+1-a))}^{\min(2(a-1),2(k+1-a))} c_{in}^{k-s} \cdot c_{out}^{s}\right)\right)^2 \\
&\quad + \frac{(k-1)!}{n \cdot 2^{k-1}} \sum_{a=2}^{k+1} O\left(\sum_{s=\min(2,2(a-2),2(k+1-a))}^{\min(2(a-1),2(k+1-a))} c_{in}^{k-s} \cdot c_{out}^{s}\right) \\
&\quad + \left(\frac{(k-1)!}{n \cdot 2^{k-1}} \sum_{a=1}^{k} O\left(\sum_{s=1}^{\min(2a-1,2(k-a)+1)} c_{in}^{k-s} \cdot c_{out}^{s}\right)\right)^2 \\
&\quad + \frac{(k-1)!}{n \cdot 2^{k-1}} \sum_{a=1}^{k} O\left(\sum_{s=1}^{\min(2a-1,2(k-a)+1)} c_{in}^{k-s} \cdot c_{out}^{s}\right).
\end{aligned}$$

As $c_{in} = d + \lambda\sqrt{d}$ and $c_{out} = d - \lambda\sqrt{d}$, all four terms in ρ are in the form of $g(\lambda) = \sum_{s=1}^{k} \left(d + \lambda\sqrt{d}\right)^{k-s} \cdot \left(d - \lambda\sqrt{d}\right)^{s}$. According to Lemma 30, all functions in the form of $g(\lambda)$ strictly increase on $\lambda \in [-\sqrt{d}, 0]$ and decreases on $\lambda \in [0, \sqrt{d}]$. Since all the other elements in ψ_1 except ζ_k are positive, ψ_1 strictly increases on $\lambda \in [-\sqrt{d}, 0]$ and decreases on $\lambda \in [0, \sqrt{d}]$.

When k increases, ζ_k contains more items and thus ψ_1 increases with order K.

According to Proposition 12, we have

$$\lambda \in [0, \sqrt{d}] \Leftrightarrow H_{edge} \in [0.5, 1] \text{ and } \lambda \in [-\sqrt{d}, 0] \Leftrightarrow H_{edge} \in [0, 0.5].$$

According to Theorem 9, any factors affecting γ affect the generalization error bound. Thus, we conclude the following cases:

(a) uniform transductive stability γ, generalization error bound and λ

From the above analysis, we know that ϕ_y is not affected by λ, and terms $\mathbb{E}\left[\|\hat{y}_i - y_i\|_F^2\right]$, ψ_1, and ψ_2 strictly increase on $\lambda \in [-\sqrt{d}, 0]$ and decrease on $\lambda \in [0, \sqrt{d}]$. This shows that

the stability decreases and the generalization error bound increases when $H_{edge} \in (0, 0.5]$. The stability increases and the generalization error bound decreases when $H_{edge} \in [0, 5, 1)$. Spectral GNNs are stable and generalize well on strong homophilic and heterophilic graphs.

(b) uniform transductive stability γ, generalization error bound, and K

From the above analysis, we know that terms ϕ_y, ψ_1, ψ_2 increase with k. Thus, when $\sum_{k=2}^{K} \theta_k \frac{(k-1)!}{2^{k-1}}$ grows slower than $\sum_{k=2}^{K} \theta_k^2 \frac{(k-1)!}{2^k}$, the expected prediction error $\mathbb{E}\left[\|\hat{y}_i - y_i\|_F^2\right]$ increases with K and the uniform transductive stability γ also increases with K. Under this condition, the stability becomes worse and generalization error bound thus increases when K increase.

□

G Details of Experiments

G.1 Datasets

The statistical properties of real-world datasets, including the number of nodes, edges, feature dimensions, node classes, and edge homophily ratios, are summarized in Table 2 and Table 3. We use the directed and cleaned versions of the Chameleon and Squirrel datasets provided by (Platonov et al., 2023), where repeated nodes have been removed.

Statistics	Texas	Wisconsin	Cornell	Actor	Chameleon	Squirrel	Citeseer	Pubmed	Cora
# Nodes	183	251	183	7,600	890	2,223	3,327	19,717	2,708
# Edges	295	466	295	26,752	27,168	131,436	4,676	44,327	5,278
# Features	1,703	1,703	1,703	932	2,325	2,089	3,703	500	1,433
# Classes	5	5	5	5	5	6	5	7	
Edge Homophily	0.11	0.21	0.22	0.24	0.22	0.74	0.8	0.81	

Table 2: Statistics of real-world datasets.

Statistics	OGBN-Arxiv	OGBN-Products
# Nodes	169,343	2,449,029
# Edges	2,315,598	61,859,140
# Features	128	100
# Classes	40	47
Edge Homophily	0.65	0.81

Table 3: Statistics of OGBN datasets.

G.2 Spectral GNNs

We detail the spectral GNNs used in our experiments below. For a graph with adjacency matrix A, degree matrix D, and identity matrix I, we define the following matrices: the normalized Laplacian matrix $\hat{L} = I - D^{-1/2}AD^{-1/2}$, the shifted normalized Laplacian matrix $\tilde{L} = -D^{-1/2}AD^{-1/2}$, the normalized adjacency matrix $\tilde{A} = D^{-1/2}AD^{-1/2}$, and the normalized adjacency matrix with self-loops $\tilde{A}' = (D + I)^{-1/2}(A + I)(D + I)^{-1/2}$.

ChebNet (Defferrard et al., 2016): This model uses the Chebyshev basis to approximate a spectral filter:

$$\hat{Y} = \sum_{k=0}^{K} \theta_k T_k(\tilde{L}) f_W(X)$$

where X is the raw feature matrix, $\Theta = [\theta_0, \theta_1, \ldots, \theta_K]$ is the graph convolution parameter, W is the feature transformation parameter and $f_W(X)$ is usually a 2-layer MLP. $T_k(\tilde{L})$ is the k-th Chebyshev

basis expanded on the shifted normalized graph Laplacian matrix $\tilde{L}$ and is recursively calculated:

$$T_0(\tilde{L}) = I$$
$$T_1(\tilde{L}) = \tilde{L}$$
$$T_k(\tilde{L}) = 2\tilde{L}T_{k-1}(\tilde{L}) - T_{k-2}(\tilde{L})$$

ChebNetII (He et al., 2022): The model is formulated as

$$\hat{Y} = \frac{2}{K+2}\sum_{k=0}^{K}\sum_{j=0}^{K}\theta_j T_k(x_j)T_k(\tilde{L})f_W(X),$$

where X is the input feature matrix, W is the feature transformation parameter, $f_W(X)$ is usually a 2-layer MLP, $T_k(\cdot)$ is the k-th Chebyshev basis expanded on $\cdot$, $x_j = \cos\left(\left(j+1/2\right)\pi/\left(K+1\right)\right)$ is the j-th Chebyshev node, which is the root of the Chebyshev polynomials of the first kind with degree $K+1$, and θ_j is a learnable parameter. Graph convolution parameter in ChebNet is reparameterized with Chebyshev nodes and learnable parameters θ_j.

JacobiConv (Wang & Zhang, 2022): This model uses the Jacobi basis to approximate a filter as:

$$\hat{Y} = \sum_{k=0}^{K}\theta_k T_k^{a,b}(\tilde{A})f_W(X),$$

where X is the input feature matrix, $\Theta = [\theta_0, \theta_1, \ldots, \theta_K]$ is the graph convolution parameter, W is the feature transformation parameter and $f_W(X)$ is usually a 2-layer MLP. $T_k^{a,b}(\tilde{A})$ is the Jacobi basis on normalized graph adjacency matrix $\tilde{A}$ and is recursively calculated as

$$T_k^{a,b}(\tilde{A}) = I$$
$$T_k^{a,b}(\tilde{A}) = \frac{1-b}{2}I + \frac{a+b+2}{2}\tilde{A}$$
$$T_k^{a,b}(\tilde{A}) = \gamma_k\tilde{A}T_{k-1}^{a,b}(\tilde{A}) + \gamma_k' T_{k-1}^{a,b}(\tilde{A}) + \gamma_k'' T_{k-2}^{a,b}(\tilde{A})$$

where $\gamma_k = \frac{(2k+a+b)(2k+a+b-1)}{2k(k+a+b)}, \gamma_k' = \frac{(2k+a+b-1)(a^2-b^2)}{2k(k+a+b)(2k+a+b-2)}, \gamma_k'' = \frac{(k+1-1)(k+b-1)(2k+a+b)}{k(k+a+b)(2k+a+b-2)}$. a and b are hyper-parameters. Usually, grid search is used to find the optimal a and b values.

GPRGNN (Chien et al., 2021): This model uses the monomial basis to approximate a filter:

$$\hat{Y} = \sum_{k=0}^{K}\theta_k\tilde{A}'^k f_W(X)$$

where X is the input feature matrix, $\Theta = [\theta_0, \theta_1, \ldots, \theta_K]$ is the graph convolution parameter, W is the feature transformation parameter and $f_W(X)$ is usually a 2-layer MLP. $\tilde{A}'$ is the normalized adjacency matrix with self-loops.

BernNet (He et al., 2021): This model uses the Bernstein basis for approximation:

$$\hat{Y} = \sum_{k=0}^{K}\theta_k\frac{1}{2^K}\binom{K}{k}(2I-\hat{L})^{K-k}\hat{L}^k f_W(X)$$

where X is the input feature matrix, $\Theta = [\theta_0, \theta_1, \ldots, \theta_K]$ is the graph convolution parameter, W is the feature transformation parameter and $f_W(X)$ is usually a 2-layer MLP. $\hat{L}$ is the normalized Laplacian matrix.

G.3 HYPER-PARAMETER SETTINGS

All experiments were conducted on an NVIDIA RTX A6000 GPU with 48GB of memory.

We employ a two-layer Multi-Layer Perceptron (MLP) with a hidden layer size of 64 for the feature transformation function f_W, using ReLU as the activation function across all spectral GNN models.

Following (Tang & Liu, 2023a; Cong et al., 2021), the dropout rate and weight decay are set to 0.0. The Adam optimizer is used for optimization. Each experiment runs for a maximum of 300 iterations and is repeated 10 times to report the mean and variance of the results. A grid search is conducted to determine the best learning rate from $\{0.05, 0.01, 0.001\}$.

G.4 DETAILED EXPERIMENTAL RESULTS

H_{edge}	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
ChebNet	94.92±0.24	86.08±0.43	81.09±0.63	75.11±0.73	72.69±0.66	74.66±0.65	79.62±0.78	86.03±0.6	94.64±0.39
Acc Gap	5.08±0.24	13.92±0.41	18.91±0.57	24.89±0.72	27.3±0.62	25.34±0.68	20.38±0.74	13.97±0.61	5.36±0.41
Loss Gap	0.64±0.07	3.15±0.14	3.72±0.2	5.42±0.24	5.88±0.5	6.01±0.27	4.62±0.3	3.04±0.18	0.98±0.06
ChebNetII	92.19±0.51	85.03±0.58	79.83±0.43	77.55±0.64	77.34±0.54	77.7±0.57	78.22±0.73	83.68±0.41	91.43±0.48
Acc Gap	7.81±0.47	14.97±0.58	20.17±0.41	22.45±0.66	22.66±0.49	22.3±0.57	21.77±0.71	16.32±0.44	8.57±0.47
Loss Gap	0.66±0.07	1.84±0.11	3.55±0.21	4.77±0.26	4.86±0.13	4.64±0.21	4.23±0.33	2.14±0.17	0.72±0.05
JacobiConv	89.25±3.35	77.23±4.51	77.19±0.66	77.0±0.55	79.06±0.61	80.2±0.57	84.64±0.39	90.48±0.24	96.91±0.24
Acc Gap	10.71±2.86	22.73±4.36	22.8±0.67	23.0±0.54	20.94±0.61	19.8±0.6	15.36±0.41	9.51±0.24	3.09±0.25
Loss Gap	0.69±0.26	1.58±0.45	4.08±0.21	4.33±0.14	5.36±0.33	1.95±0.13	1.58±0.13	0.99±0.06	0.16±0.01
GPRGNN	90.33±0.57	87.06±0.64	81.71±0.41	77.03±0.47	77.23±0.65	79.52±0.59	82.72±0.52	89.25±0.5	96.45±0.18
Acc Gap	9.66±0.54	12.94±0.67	18.29±0.42	22.96±0.49	22.77±0.64	20.48±0.6	17.27±0.52	10.75±0.54	3.55±0.2
Loss Gap	1.42±0.08	2.21±0.14	3.27±0.2	4.72±0.19	5.17±0.13	4.7±0.25	3.7±0.47	2.4±0.32	1.05±0.11
BernNet	87.44±0.5	82.92±0.67	79.3±0.44	77.69±0.53	77.97±0.54	77.49±0.72	76.58±0.79	79.73±1.3	85.68±1.05
Acc Gap	12.55±0.5	17.08±0.76	20.7±0.44	22.31±0.54	22.03±0.55	22.51±0.64	23.41±0.8	20.27±1.39	14.32±1.06
Loss Gap	1.2±0.06	2.45±0.21	3.69±0.16	4.77±0.24	4.72±0.15	4.7±0.17	4.35±0.35	2.92±0.31	1.36±0.14

Table 4: Testing accuracy, accuracy gap, loss gap of spectral GNNs on synthetic datasets with edge homophilic ratio $H_{edge} \in [0.1, 0.9]$. Small accuracy and loss gaps imply good generalization capability.

Datasets	Texas	Wisconsin	Actor	Squirrel	Chameleon	Cornell	Citeseer	Pubmed	Cora
ChebNet	40.82±7.25	52.23±3.77	26.63±0.53	30.08±1.14	33.94±1.58	44.88±6.19	64.16±0.82	84.74±0.37	74.95±0.96
Acc Gap	59.18±6.94	47.77±3.92	73.26±0.54	69.92±1.28	66.06±1.52	55.12±5.95	35.82±0.75	15.25±0.37	25.05±0.92
Loss Gap	5.91±0.66	5.77±0.87	21.64±0.8	35.68±2.33	36.17±3.04	6.57±0.82	4.68±0.22	1.44±0.06	3.9±0.29
ChebNetII	77.55±5.71	74.38±3.08	27.94±0.36	28.1±1.82	38.45±1.63	73.69±5.12	65.85±0.52	84.7±0.3	74.0±0.8
Acc Gap	22.45±5.2	25.62±3.31	71.94±0.33	71.83±1.77	61.47±1.53	26.31±5.0	34.12±0.48	15.16±0.28	26.0±0.75
Loss Gap	1.1±0.27	1.39±0.32	20.16±0.76	27.56±2.88	19.33±1.68	1.7±0.3	2.66±0.09	1.13±0.09	2.14±0.09
JacobiConv	78.06±5.31	77.62±2.92	27.89±0.63	26.78±1.28	32.2±2.08	80.41±3.98	73.56±0.64	86.33±0.47	84.31±0.49
Acc Gap	21.94±5.41	22.38±2.85	71.97±0.66	50.85±11.88	63.82±9.46	19.59±4.18	26.41±0.65	10.87±1.45	15.69±0.5
Loss Gap	0.94±0.26	1.19±0.22	31.67±0.86	32.75±11.57	38.77±7.16	0.91±0.16	2.16±0.06	0.51±0.14	1.28±0.09
GPRGNN	46.84±6.22	72.08±3.23	26.29±0.65	29.91±1.19	34.28±1.58	61.33±6.12	72.89±0.62	85.42±0.4	84.37±0.51
Acc Gap	53.16±6.12	27.92±2.92	71.52±4.82	70.09±1.09	65.72±1.69	38.67±6.43	27.08±0.67	14.58±0.37	15.63±0.54
Loss Gap	3.35±0.83	1.6±0.31	29.22±2.69	35.34±5.58	29.88±2.22	2.2±0.53	3.32±0.16	1.24±0.09	1.54±0.1
BernNet	75.92±5.31	81.85±2.23	27.28±0.76	33.42±1.14	33.72±1.38	81.43±3.46	67.17±0.59	84.82±0.25	73.39±0.87
Acc Gap	24.08±5.41	18.15±2.16	72.61±0.71	66.58±1.11	66.28±1.33	18.57±3.57	32.8±0.57	14.95±0.45	26.61±0.87
Loss Gap	1.24±0.31	0.87±0.26	24.68±0.71	28.17±1.47	27.83±1.75	1.06±0.18	2.66±0.09	1.13±0.13	2.18±0.08

Table 5: Testing accuracy, accuracy gap, loss gap of spectral GNNs on real world datasets with edge homophilic ratio $H_{edge} \in [0.11, 0.81]$. Small accuracy and loss gaps imply good generalization capability.

Order K	1	2	3	4	5	6	7	8	9	10
ChebNet	87.31±0.3	89.11±0.31	88.48±0.49	84.19±0.9	71.3±3.0	79.58±0.52	80.77±0.62	76.21±0.51	82.94±0.48	86.08±0.41
Acc Gap	12.7±0.32	10.89±0.31	11.52±0.5	15.8±0.92	28.7±3.54	20.42±0.51	19.23±0.57	23.79±0.47	17.06±0.45	13.92±0.42
Loss Gap	2.2±0.09	1.76±0.07	1.9±0.14	2.84±0.27	7.2±1.45	3.88±0.2	3.08±0.21	3.79±0.26	3.8±0.11	3.15±0.14
ChebNetII	85.92±0.56	80.1±0.99	82.65±0.7	85.56±0.45	84.64±0.8	84.62±0.59	85.27±0.51	86.2±0.64	86.39±0.5	85.03±0.57
Acc Gap	14.07±0.53	19.9±1.02	17.35±0.73	14.44±0.45	15.36±0.87	15.38±0.6	14.73±0.5	13.79±0.6	13.61±0.49	14.97±0.58
Loss Gap	1.94±0.08	3.23±0.31	2.62±0.14	2.06±0.14	1.94±0.21	1.95±0.17	1.99±0.15	1.75±0.14	1.83±0.11	1.84±0.11
JacobiConv	77.44±0.67	80.51±0.48	49.44±1.12	39.85±1.91	48.81±2.65	47.73±7.63	60.29±7.48	67.53±7.95	68.03±9.15	77.23±4.79
Acc Gap	22.55±0.62	19.49±0.46	50.56±1.18	60.13±1.98	51.19±2.63	52.25±7.08	39.7±7.32	32.45±7.76	31.96±9.19	22.73±4.82
Loss Gap	5.72±0.19	5.8±0.26	8.81±0.79	12.63±1.22	7.3±1.01	8.23±1.77	4.98±1.23	3.42±1.39	3.33±1.32	1.58±0.48
GPRGNN	83.61±0.66	86.14±0.29	79.44±1.05	88.36±0.28	87.25±0.5	88.0±0.39	87.57±0.47	87.5±0.3	87.17±0.3	87.06±0.59
Acc Gap	16.39±0.69	13.86±0.29	20.56±1.06	11.63±0.29	12.76±0.49	12.01±0.32	12.43±0.48	12.49±0.33	12.84±0.29	12.94±0.68
Loss Gap	2.37±0.11	2.21±0.1	3.18±0.19	1.83±0.1	2.14±0.2	1.93±0.09	2.06±0.13	2.12±0.09	2.19±0.13	2.21±0.14
BernNet	82.76±0.72	81.14±0.41	81.21±0.57	81.47±0.6	81.77±0.66	82.11±0.75	82.32±0.88	82.55±0.84	82.8±0.81	82.92±0.79
Acc Gap	17.24±0.71	18.86±0.39	18.79±0.56	18.53±0.7	18.23±0.62	17.89±0.85	17.68±0.84	17.45±0.79	17.2±0.79	17.08±0.7
Loss Gap	2.45±0.17	3.02±0.11	2.95±0.21	2.84±0.2	2.75±0.21	2.65±0.21	2.59±0.22	2.54±0.2	2.49±0.21	2.45±0.21

Table 6: Testing accuracy, accuracy gap, loss gap of spectral GNNs on synthetic dataset of edge homophilic ratio $H_{edge} = 0.2$ when $K \in [1, 10]$. Small accuracy and loss gaps imply good generalization capability.

Order K	1	2	3	4	5	6	7	8	9	10
ChebNet	83.78±2.45	80.61±4.59	80.51±3.47	61.73±5.0	63.37±8.57	36.33±5.72	44.18±5.0	24.39±2.14	30.2±4.8	40.82±7.35
Acc Gap	16.22±2.45	19.39±4.8	19.49±3.78	38.27±5.0	36.63±7.86	63.67±6.12	55.82±5.0	75.61±2.24	69.8±5.0	59.18±7.15
Loss Gap	1.49±0.44	1.26±0.44	1.48±0.31	2.77±0.53	3.08±0.59	8.98±0.68	6.09±0.72	7.99±0.93	9.0±1.03	5.91±0.69
ChebNetII	80.41±3.98	75.41±5.72	76.53±4.29	76.53±4.59	76.94±5.0	78.78±5.61	78.88±5.2	77.45±4.9	76.94±5.72	77.55±5.51
Acc Gap	19.59±3.78	24.59±5.2	23.47±4.59	23.47±4.49	23.06±4.8	21.22±5.61	21.12±5.82	22.55±4.49	23.06±5.61	22.45±5.31
Loss Gap	0.74±0.14	1.2±0.44	1.15±0.29	1.28±0.3	1.23±0.33	1.11±0.29	1.16±0.26	1.21±0.29	1.24±0.27	1.1±0.27
JacobiConv	52.24±5.41	80.92±3.78	75.31±5.31	74.39±3.78	79.08±3.67	78.67±4.08	80.0±3.06	73.67±6.33	77.65±5.41	78.06±5.61
Acc Gap	47.76±5.31	19.08±3.98	24.69±5.0	25.61±3.67	20.92±3.47	21.33±3.67	20.0±3.06	26.33±6.84	22.35±5.1	21.94±5.41
Loss Gap	2.54±0.42	0.89±0.2	1.1±0.25	1.18±0.27	0.9±0.17	0.97±0.16	0.93±0.13	1.22±0.39	0.97±0.26	0.94±0.24
GPRGNN	53.88±4.8	49.18±5.1	46.73±5.82	45.82±6.64	46.12±5.41	45.61±5.2	46.43±4.59	46.12±5.0	47.55±4.8	46.84±6.22
Acc Gap	46.12±4.9	50.82±5.31	53.27±5.61	54.18±6.63	53.88±5.72	54.39±5.2	53.57±4.9	53.88±4.9	52.45±5.1	53.16±6.43
Loss Gap	2.6±0.44	3.21±0.53	3.5±0.67	3.6±0.63	3.58±0.63	3.51±0.64	3.47±0.48	3.44±0.61	3.22±0.73	3.35±0.83
BernNet	76.73±3.67	75.92±2.45	75.61±3.67	77.04±3.88	77.14±4.39	75.2±4.7	74.9±5.72	75.2±5.2	74.8±5.92	75.71±5.71
Acc Gap	23.27±3.67	24.08±2.65	24.39±3.57	22.96±3.98	22.86±4.29	24.8±4.69	25.1±5.2	24.8±5.61	25.2±6.02	24.29±5.61
Loss Gap	0.96±0.22	0.95±0.18	1.01±0.17	1.02±0.21	1.06±0.21	1.13±0.25	1.19±0.31	1.18±0.26	1.27±0.34	1.25±0.31

Table 7: Testing accuracy, accuracy gap, loss gap of spectral GNNs on Texas dataset of edge homophilic ratio $H_{edge} = 0.11$ when $K \in [1, 10]$. Small accuracy and loss gaps imply good generalization capability.