
Deeper and Better: Mitigating Oversmoothing and Oversquashing in MPNNs via Local Riemannian Geometry

(Technical Appendix)

Anonymous Author(s)

Affiliation

Address

email

1 This is technical appendix of our submission, including theorems, proofs, derivations, algorithm and
2 experimental details. Note that, an elaboration on the Case Study is provided in another PDF entitled,
3 Further Details on Visualization.

4 Contents

5	A Notations	2
6	B Theorems, Proofs and Derivations	3
7	B.1 Proofs of Theorem 4.2	3
8	B.2 Proofs of Theorem 4.3	3
9	B.3 Proofs of Theorem 5.1	5
10	B.4 Proofs of Corollary 5.2	8
11	B.5 Proofs of Lemma 5.3	8
12	B.6 Proofs of Theorem 5.4	10
13	B.7 Proofs of Lemma 6.2	10
14	B.8 Proofs of Theorem 6.3	11
15	B.9 Derivation of GBN formalism	13
16	C Algorithm	15
17	D Datasets and Baselines	15
18	D.1 Datasets	15
19	D.2 Baselines	15
20	E Implementation Notes	17
21	E.1 Graph transfer	17
22	E.2 Hyperparameter settings	18

Table 1: Notation Table.

Notation	Description
\mathcal{M}	A Riemannian manifold
g	The Riemannian metric
\mathbf{x}	A point on Riemannian manifold or a vector
$\text{vol}_{\mathcal{M}}$	The volume form on \mathcal{M}
$\int_{\mathcal{M}} \cdot \text{vol}_{\mathcal{M}}(\mathbf{y})$	The Riemannian integral w.r.t. variable \mathbf{y}
Δ_g	The Laplace-Beltrami operator on \mathcal{M} w.r.t. g
grad	The Riemannian gradient
div	The Riemannian divergence
λ_i, ψ_i	The i -th eigenvalue and eigenfunction of the Laplace-Beltrami operator
∂_t	The partial derivative w.r.t. t
$u(t, \mathbf{x})$	The solution of the heat equation
$H(t, \mathbf{x}, \mathbf{y})$	The heat kernel
λ	The spectral gap of a closed manifold
λ^D	The spectral gap of a compact manifold with Dirichlet boundary
λ^N	The spectral gap of a compact manifold with Neumann boundary
$\mathcal{G} = (\mathcal{V}, \mathcal{E})$	A graph with node set \mathcal{V} and edge set \mathcal{E}
\mathbf{X}	The feature matrix
\mathbf{L}	The normalized graph Laplacian
\mathbf{A}	The graph adjacency matrix
φ	A MLP
ϕ	The initial value or function of the heat equation
u, v, v_i, v_j	A node in graph
ℓ, i, j, k, m	The lower index
\mathbf{I}	The identity matrix
\mathcal{S}	The $(n - 1)$ -dimensional submanifold of \mathcal{M}
Area	The volume of \mathcal{M} restrict to the $(n - 1)$ -dimensional submanifold \mathcal{S}
vol	The volume of \mathcal{M}
$h_{\mathcal{M}}$	The Cheeger's constant of \mathcal{M}
ξ	A small value variable
$\frac{\delta f[\phi]}{\delta \phi}$	The variational derivative of f w.r.t. ϕ
$\text{Dir}[f]$	The Dirichlet energy of function f
$\frac{\partial}{\partial n}$	The outward normal derivative
α, β	The coefficients of the Robin condition
$\varepsilon(\cdot)$	The radius function
$\mathcal{B}_{\varepsilon}(r)$	The closed ball with radius $\varepsilon(r)$
S	The interior node set of graph G
∂S	The boundary node set of graph G
\mathbf{D}	The diagonal matrix for the Jacobi method
\mathbf{V}	The lower diagonal matrix for the Jacobi method
\mathbf{U}	The upper diagonal matrix for the Jacobi method
$\mathbf{\Gamma}_{\partial S}$	The source term of nonhomogeneous Robin condition for boundary nodes
\mathbf{F}_S	The steady state of heat equation for interior nodes
$\mathbf{\Gamma}$	The concatenate for \mathbf{F}_S and $\mathbf{\Gamma}_{\partial S}$

24 **Mathematical Background.** (\mathcal{M}, g) denotes a n -dimensional manifold endowed with the Rie-
25 mannian metric g .¹ The volume form is written as $\text{vol}_{\mathcal{M}} = \sqrt{|\det g|} dx^1 \wedge \cdots \wedge dx^n$, where
26 $\mathbf{x} \in \mathbb{R}^n$ denotes local coordinates and x is the element of \mathbf{x} . A Hilbert (function) space
27 $L^2(\mathcal{M}) = \{f \in C^\infty(\mathcal{M}) | \int_{\mathcal{M}} f \text{vol}_{\mathcal{M}}\}$ consists of smooth L^2 integrable functions on \mathcal{M} . Let
28 $\Delta_g f = \text{div}(\text{grad } f)$ be the Laplace-Beltrami operator on \mathcal{M} w.r.t. g . Δ_g on a closed manifold has
29 discrete eigenvalues λ , and its eigenfunctions ψ form a basis of $L^2(\mathcal{M})$.

¹Throughout this paper, the terminology of manifold \mathcal{M} refers to a closed Riemannian manifold (compact and without boundary) unless otherwise specified.

30 B Theorems, Proofs and Derivations

31 B.1 Proofs of Theorem 4.2

32 **Theorem 4.2 (Spectral Gap and Gradient Vanishing).** Suppose $t \in [t_\ell, t_{\ell+1}]$ with $1 < \ell \leq K$,
 33 the upper bound of the functional derivative $\frac{\delta u_\ell[\phi]}{\delta \phi(\mathbf{y})}(t, \mathbf{x})$ is determined by the heat kernel $H(t, \mathbf{x}, \mathbf{y})$
 34 and the model constant C_{model} ,

$$\frac{\delta u_\ell[\phi]}{\delta \phi(\mathbf{y})}(t, \mathbf{x}) \leq C_{\text{model}} \cdot H(t, \mathbf{x}, \mathbf{y}) = C_{\text{model}} \cdot \sum_{i=0} e^{-\lambda_i t} \psi_i(\mathbf{x}) \psi_i(\mathbf{y}).$$

35

36 *Proof.* The algebraic operations are omitted for clarity², and the key steps of the derivation are shown
 37 as follows,

$$\begin{aligned} \frac{\delta u_\ell[\phi]}{\delta \phi(\mathbf{y})}(t, x) &= \int_{\mathcal{M}} \frac{\delta u_\ell[\phi_\ell]}{\delta \phi_\ell(z_\ell)}(t, x) \frac{\delta \phi_\ell[\phi]}{\delta \phi(\mathbf{y})} \text{vol}_{\mathcal{M}}(z_\ell) \\ &= \int_{\mathcal{M}} \frac{\delta u_\ell[\phi_\ell]}{\delta \phi_\ell(z_\ell)}(t, x) \frac{\delta G_{\theta_\ell}[u_{\ell-1}]}{\delta u_{\ell-1}(t_\ell, z_\ell)} \frac{\delta u_{\ell-1}[\phi]}{\delta \phi(\mathbf{y})}(t_\ell, z_\ell) \text{vol}_{\mathcal{M}}(z_\ell) \\ &= \int_{\mathcal{M}^\ell} \frac{\delta u_\ell[\phi_\ell]}{\delta \phi_\ell(z_\ell)}(t, x) \frac{\delta G_{\theta_\ell}[u_{\ell-1}]}{\delta u_{\ell-1}(t_\ell, z_\ell)} \left[\prod_{i=2}^{\ell-1} \frac{\delta u_i[\varphi_i]}{\delta \varphi_i(z_i)}(t_{i+1}, z_{i+1}) \frac{\delta G_{\theta_i}[u_i]}{\delta u_{i-1}(t_i, z_i)} \right] \\ &\quad \frac{\delta u_1[\phi_1]}{\delta \phi_1(z_1)}(t_2, z_2) \frac{\delta \phi_1[\phi]}{\delta \phi(\mathbf{y})}(z_2) \text{vol}_{\mathcal{M}^\ell}(\times_{i=1}^\ell z_i) \\ &= \int_{\mathcal{M}^\ell} H(t - t_\ell, x, z_\ell) \left[\prod_{i=2}^{\ell-1} H(t_{i+1} - t_i, z_{i+1}, z_i) \right] H(t_2 - t_1, z_2, y) \\ &\quad \left[\prod_{i=2}^\ell \frac{\delta G_{\theta_i}[u_{i-1}]}{\delta u_{i-1}(t_i, z_i)} \right] \text{vol}_{\mathcal{M}^\ell}(\times_{i=2}^\ell z_i) \\ &\leq H(t, x, y) \int_{\mathcal{M}^\ell} \left[\prod_{i=2}^\ell \frac{\delta G_{\theta_i}[u_{i-1}]}{\delta u_{i-1}(t_i, z_i)} \right] \text{vol}_{\mathcal{M}^\ell}(\times_{i=2}^\ell z_i) \\ &\leq C_{\text{model}} H(t, x, y) \end{aligned} \tag{1}$$

38 Since \mathcal{M} is compact, by the eigen-decomposition w.r.t $H(t, x, y)$, we have

$$\frac{\delta u_\ell[\phi]}{\delta \phi(\mathbf{y})}(t, x) \leq C_{\text{model}} H(t, x, y) = C_{\text{model}} \cdot \sum_{i=0}^\infty e^{-\lambda_i t} \psi_i(x) \psi_i(y) \tag{2}$$

39

□

40 B.2 Proofs of Theorem 4.3

41 **Theorem 4.3 (Spectral Gap and Dirichlet Energy).** Suppose $t \in [t_\ell, t_{\ell+1}]$ with $1 < \ell \leq K$, and
 42 the initial state $u_\ell(t_\ell, \mathbf{x}) = \phi_\ell(\mathbf{x})$ has an eigenvalue decomposition such that $\phi_\ell = \sum_{i=1}^\infty c_{\ell,i} \psi_i$, the
 43 Dirichlet energy $\text{Dir}[u_\ell(t, \cdot)] = \frac{1}{2} \int_{\mathcal{M}} |\text{grad } u_\ell(t, \mathbf{x})|^2 \text{vol}_{\mathcal{M}}$ decays exponentially w.r.t. to t as

$$\text{Dir}[u_\ell(t, \cdot)] \leq \frac{1}{2} w_{\text{model}} \left(\sum_{i=1}^\infty c_i^2 \lambda_i e^{-2\lambda_i t} + \sum_{i=1, \{j_k \neq i\}}^\infty c_{j_2}^2 \lambda_i e^{-2[\lambda_i(t-t_\ell) + \sum_{k=2}^\ell \lambda_{j_k}(t_k - t_{k-1})]} \right).$$

44 where ψ_i denotes eigenfunctions and the $c_{\ell,i}$ are the corresponding coefficients.

²Throughout the proofs in Part B, we demonstrate the derivation with the scalar x , and the same structure holds for the vector \mathbf{x} .

45 *Proof.* Suppose $\phi_\ell(x) = \sum_{i=1}^\infty c_{\ell,i} \psi_i(x)$, where $c_{\ell,i} = \langle \phi_\ell, \psi_i \rangle$, we have

$$\begin{aligned}
u_\ell(t, x) &= \int_{\mathcal{M}} H(t - t_\ell, x, y) \phi_\ell(y) \text{vol}_{\mathcal{M}}(y) \\
&= \int_{\mathcal{M}} \sum_{i,j} e^{-\lambda_i(t-t_\ell)} \psi_i(x) \psi_i(y) c_{\ell,j} \phi_j(y) \text{vol}_{\mathcal{M}}(y) \\
&= \sum_{i,j} e^{-\lambda_i(t-t_\ell)} \psi_i(x) c_{\ell,j} \int_{\mathcal{M}} \psi_i(y) \psi_j(y) \text{vol}_{\mathcal{M}}(y) \\
&= \sum_{i=1}^\infty e^{-\lambda_i(t-t_\ell)} c_{\ell,i} \psi_i(x)
\end{aligned} \tag{3}$$

46 The Riemannian gradient of $u_\ell(t, \cdot)$ is

$$\text{grad}_x u_\ell(t, x) = \sum_{i=1}^\infty e^{-\lambda_i(t-t_\ell)} c_{\ell,i} \text{grad}_x \psi_i \tag{4}$$

47 Then, the Dirichlet energy of $u_\ell(t, \cdot)$ will be

$$\begin{aligned}
\text{Dir}[u_\ell(t)] &= \frac{1}{2} \int_{\mathcal{M}} |\nabla u_\ell(t, x)|^2 \text{vol}_{\mathcal{M}} \\
&= \frac{1}{2} \int_{\mathcal{M}} \langle \nabla u_\ell, \nabla u_\ell \rangle_g \text{vol}_{\mathcal{M}} \\
&= \frac{1}{2} \sum_{i,j} e^{-(\lambda_i + \lambda_j)(t-t_\ell)} c_{\ell,i} c_{\ell,j} \int_{\mathcal{M}} \langle \nabla_x \psi_i, \nabla_x \psi_j \rangle \text{vol}_{\mathcal{M}} \\
&= -\frac{1}{2} \sum_{i,j} e^{-(\lambda_i + \lambda_j)(t-t_\ell)} c_{\ell,i} c_{\ell,j} \int_{\mathcal{M}} \psi_i \Delta_0 \psi_j \text{vol}_{\mathcal{M}} \\
&= \frac{1}{2} \sum_{i,j} \lambda_j c_{\ell,i} c_{\ell,j} e^{-(\lambda_i + \lambda_j)(t-t_\ell)} \int_{\mathcal{M}} \psi_i \psi_j \text{vol}_{\mathcal{M}} \\
&= \frac{1}{2} \sum_i c_{\ell,i}^2 \lambda_i e^{-2\lambda_i(t-t_\ell)}
\end{aligned} \tag{5}$$

48 since $\phi_\ell = G_{\theta_\ell}[u_{\ell-1}(t_\ell, x)]$, the projection coefficient on ψ_i is

$$\begin{aligned}
|c_{\ell,i}| &= \left| \int_{\mathcal{M}} G_{\theta_\ell}[u_{\ell-1}(t_\ell, x)] \psi_i \text{vol}_{\mathcal{M}} \right| \\
&\leq \|G_{\theta_\ell}\| \|u_{\ell-1}(t_\ell)\| \|\psi_i\| \\
&= \|G_{\theta_\ell}\| \|u_{\ell-1}(t_\ell)\|
\end{aligned}$$

49 From Eq. (3), we have

$$\begin{aligned}
\|u_{\ell-1}(t_\ell)\|^2 &= \int_{\mathcal{M}} \sum_{i,j} c_{\ell-1,i} c_{\ell-1,j} e^{-(\lambda_i + \lambda_j)(t_\ell - t_{\ell-1})} \psi_i(x) \psi_j(x) \text{vol}_{\mathcal{M}} \\
&= \sum_j c_{\ell-1,j}^2 \cdot e^{-2\lambda_j(t_\ell - t_{\ell-1})}
\end{aligned}$$

50

$$\Rightarrow c_{\ell,i}^2 \leq \|G_{\theta_i}\|^2 \sum_j c_{\ell-1,j}^2 e^{-2\lambda_j(t_\ell - t_{\ell-1})} \tag{6}$$

51 Substituting into Eq. (5), we have

$$\text{Dir}[u_\ell(t)] = \frac{1}{2} \sum_{i=1}^\infty c_{\ell,i}^2 \lambda_i e^{-2\lambda_i(t-t_\ell)} \tag{7}$$

$$\leq \frac{1}{2} \|G_{\theta_\ell}\|^2 \sum_{i,j} c_{\ell-1,j}^2 \lambda_i e^{-2\lambda_i(t-t_\ell) - 2\lambda_j(t_\ell - t_{\ell-1})} \tag{8}$$

By induction, we have

$$\text{Dir}[u_\ell(t)] \leq \frac{1}{2} \left[\prod_{k=1}^{\ell} \|G_{\theta_k}\|^2 \right] \sum_{i,j_2 \dots j_\ell} c_{j_2}^2 \lambda_i e^{-2[\lambda_i(t-t_\ell) + \sum_{k=2}^{\ell} \lambda_{j_k}(t_k - t_{k-1})]} \quad (9)$$

$$= \frac{1}{2} w_{\text{model}} \left[\sum_i c_i^2 \lambda_i e^{-2\lambda_i t} + \sum_{\substack{i,j_2 \neq i \\ \dots \\ j_\ell \neq i}} c_{j_2}^2 \lambda_i e^{-2\lambda_i(t-t_\ell) + \sum_{k=2}^{\ell} \lambda_{j_k}(t_k - t_{k-1})} \right] \quad (10)$$

$$(11)$$

where $c_{j_2} = \langle \phi, \psi_{j_2} \rangle$. \square

B.3 Proofs of Theorem 5.1

Theorem 5.1 (Oversquashing under Boundary Conditions). *Given the boundary of a submanifold \mathcal{S} written as $\partial\mathcal{S} = \{0, L\} \times \partial\mathcal{B}_\varepsilon(s)$, we have the following claims hold:*

- For $\varepsilon(s) = \varepsilon_0$, if the Dirichlet condition is applied, the decay of heat kernel is determined by $e^{-(\varepsilon_0^{-2}t)}$, while it is determined by $e^{-(L^{-2}t)}$ under Neumann condition.
- For $\varepsilon(s) = \varepsilon_0 \cosh a(s - \frac{L}{2})$, the decay of heat kernel under both the Dirichlet and Neumann conditions is related to $e^{-(\varepsilon_0 \cosh a \frac{L}{2})^{-2}}$.

For both cases, the decay under Robin condition is upper bounded by the Dirichlet condition and lower bounded by the Neumann condition.

Proof. Consider $\mathcal{S} = \bigcup_{s \in [0, L]} \{s\} \times \mathcal{B}_\varepsilon(s)$, where $\varepsilon(r) : [0, L] \rightarrow (0, \delta)$ with $\delta \in (0, 1)$. Denote the Laplace-Beltrami operator on $[0, L]$ and $\mathcal{B}_\varepsilon(s)$ are $\Delta_{\mathcal{R}}$ and $\Delta_{\mathcal{B}_\varepsilon, s}$. Since $\Delta_{\mathcal{S}} = \Delta_{\mathcal{R}} \otimes I_{\mathcal{B}_\varepsilon, s} + I_{[0, L]} \otimes \Delta_{\mathcal{B}_\varepsilon, s}$, for eigenfunction $\psi^{\mathcal{R}}$ of $\Delta_{\mathcal{R}}$ and $\psi^{\mathcal{B}_\varepsilon, s}$ for $\Delta_{\mathcal{B}_\varepsilon, s}$,

$$\begin{aligned} \Delta_{\mathcal{S}}(\psi_m^{\mathcal{R}} \otimes \psi_\ell^{\mathcal{B}_\varepsilon, s}) &= \Delta_{\mathcal{R}} \psi_m^{\mathcal{R}} \otimes \psi_\ell^{\mathcal{B}_\varepsilon, s} + \psi_m^{\mathcal{R}} \otimes \Delta_{\mathcal{B}_\varepsilon, s} \psi_\ell^{\mathcal{B}_\varepsilon, s} \\ &= (\mu_m^{\mathcal{R}} + \mu_\ell^{\mathcal{B}_\varepsilon, s}) \psi_m^{\mathcal{R}} \otimes \psi_\ell^{\mathcal{B}_\varepsilon, s} \end{aligned} \quad (12)$$

The first case Now, we consider that $\varepsilon(s) = \varepsilon_0$, where $\varepsilon_0 \in (0, \delta)$ is a constant. First, we solve the eigenvalue problem on $\mathcal{R} = [0, L]$ with boundary condition:

On the Dirichlet boundary condition, the eigenvalue problem equation is

$$\begin{cases} \frac{d^2}{dx^2} \psi^{\mathcal{R}} = -\mu^{\mathcal{R}} \psi^{\mathcal{R}}, & x \in (0, L) \\ \psi^{\mathcal{R}}(0) = \psi^{\mathcal{R}}(L) = 0 & x \in \{0, L\} \end{cases} \quad (13)$$

$$\Rightarrow \mu_m^{\mathcal{R}} = \left(\frac{m\pi}{L}\right)^2, \psi_m^{\mathcal{R}}(x) = \sin\left(\frac{m\pi}{L}x\right) \quad (14)$$

On the Neumann boundary condition, the eigenvalue problem equation is

$$\begin{cases} \frac{d^2}{dx^2} \psi^{\mathcal{R}} = -\mu^{\mathcal{R}} \psi^{\mathcal{R}}, & x \in (0, L) \\ \frac{d}{dx} \psi^{\mathcal{R}}|_{x=0} = -\frac{d}{dx} \psi^{\mathcal{R}}|_{x=L} = 0 & x \in \{0, L\} \end{cases} \quad (15)$$

$$\Rightarrow \mu_m^{\mathcal{R}} = \left(\frac{m\pi}{L}\right)^2, \psi_m^{\mathcal{R}}(x) = \cos\left(\frac{m\pi}{L}x\right) \quad (16)$$

For the eigenvalue problem in $\mathcal{B}_\varepsilon(s)$, the sectional surface is a closed ball with radius ε_0 . The Laplace-Beltrami operator $\Delta_{\mathcal{B}, \varepsilon_0} = \frac{1}{\varepsilon_0^2} \Delta_{\mathcal{B}, 1}$, where $\Delta_{\mathcal{B}, 1}$ is the Laplace-Beltrami operator on the unit ball, under the polar coordinate (r, θ) , $r \in [0, 1]$, $\theta \in \mathbb{S}^{n-1}$,

$$\Delta_{\mathcal{B}, 1} = \frac{1}{r^{n-1}} \frac{\partial}{\partial r} \left(r^{n-1} \frac{\partial}{\partial r} \right) + \frac{1}{r^2} \Delta_{\mathbb{S}^{n-1}} \quad (17)$$

75 Using separate variable method, let $\psi^{B_1} = R(r)\Theta(\theta)$, the eigenvalue problem equation is:

$$\frac{1}{r^{n-1}} \frac{d}{dr} \left(r^{n-1} \frac{dR}{dr} \right) \theta + \frac{R}{r^2} \Delta_{\mathbb{S}^{n-1}} \Theta = -\mu^{B,1} R \Theta \quad (18)$$

76 Dividing Θ and let $\Theta(\theta) = Y_k(\theta)$, where $Y_k(\theta)$ is the spherical harmonics such that

$$\Delta_{\mathbb{S}^{n-1}} Y_k = -k(k+n-2)Y_k \quad (19)$$

77 Denote $\lambda_k^\theta = k(k+n-2)$, we have

$$r^2 \frac{dR^2}{dr^2} + (n-1)r \frac{dR}{dr} + (\mu^{B,1} r^2 - \lambda^\theta) R = 0 \quad (20)$$

78 Let $R(r) = r^{-\frac{n-2}{2}} w(r)$, then

$$r^2 \frac{d^2 w}{dr^2} + r \frac{dw}{dr} + (\mu^{B,1} r^2 - (k + \frac{n-2}{2})^2) w = 0 \quad (21)$$

79 Set $z = \sqrt{\mu^{B,\infty}} r$, we have the Bessel's different equation:

$$z^2 \frac{d^2 w}{dz^2} + z \frac{dw}{dz} + (z^2 - (k + \frac{n-2}{2})^2) w = 0 \quad (22)$$

80 Thus, $R(r) = r^{-\frac{n-2}{2}} J_\nu \sqrt{\mu^{B,1}} r$, where $\nu = k + \frac{n-2}{2}$ is the order of the Bessel function J_ν .

81 For the Dirichlet boundary condition, it satisfies

$$R(1) = 0 \Rightarrow \mu_\ell = j_{\nu,\ell}^2, \quad (23)$$

82 where $j_{\nu,\ell}$ is the ℓ -th zeros of J_ν . The eigenfunction and eigenvalues of $\mathcal{B}_{\varepsilon_0}$ is:

$$\mu_\ell^{B_{\varepsilon_0}} = \left(\frac{j_{\nu,\ell}}{\varepsilon_0} \right)^2, \quad \psi_{\ell,k}^{B_{\varepsilon_0}} = r^{-\frac{n-2}{2}} J_\nu \left(\frac{j_{\nu,\ell}}{\varepsilon_0} r \right) Y_k(\theta) \quad (24)$$

83 For the Neumann boundary condition, it satisfies

$$R'(1) = 0 \Rightarrow \mu_\ell^{B,1} = j_{\nu,\ell}'^2, \quad (25)$$

84 where $j_{\nu,\ell}'$ is the ℓ -th zeros of the derivative of J_ν . The eigenfunction and eigenvalues of $\mathcal{B}_{\varepsilon_0}$ is:

$$\mu_\ell^{B_{\varepsilon_0}} = \left(\frac{j_{\nu,\ell}'}{\varepsilon_0} \right)^2, \quad \psi_{\ell,k}^{B_{\varepsilon_0}} = r^{-\frac{n-2}{2}} J_\nu \left(\frac{j_{\nu,\ell}'}{\varepsilon_0} r \right) Y_k(\theta) \quad (26)$$

85 Above all, the eigenvalues of $\mathcal{S} = [0, L] \times \mathcal{B}_{\varepsilon_0}$ are

$$\lambda_{m,\ell}^D = \left(\frac{m\pi}{L} \right)^2 + \left(\frac{j_{\nu,\ell}}{\varepsilon_0} \right)^2 \quad (27)$$

$$\lambda_{m,\ell}^N = \left(\frac{m\pi}{L} \right)^2 + \left(\frac{j_{\nu,\ell}'}{\varepsilon_0} \right)^2 \quad (28)$$

86 Since for the Dirichlet boundary condition, the spectral gap is $\lambda^D = \left(\frac{\pi}{L} \right)^2 + \left(\frac{j_{\nu,1}}{\varepsilon_0} \right)^2$, the decay of

87 $e^{-\lambda^D} = e^{-\left(\frac{\pi}{L} \right)^2 - \left(\frac{j_{\nu,1}}{\varepsilon_0} \right)^2}$ is determined by $e^{-\varepsilon_0^{-2}}$, for the Neumann boundary condition, the spectral

88 gap is $\lambda^N = \left(\frac{\pi}{L} \right)^2 + \left(\frac{j_{0,0}'}{\varepsilon_0} \right)^2 = \left(\frac{\pi}{L} \right)^2$, the decay of $e^{-\lambda^N}$ is determined by $e^{-L^{-2}}$.

89 **The second case** Consider the radius function $\varepsilon(s) = \varepsilon_0 \frac{(e^{a(s-\frac{L}{2})} + e^{-a(s-\frac{L}{2})})}{2} = \varepsilon_0 \cosh a(s - \frac{L}{2})$
 90 for $a > 0, s \in [0, L]$. Since the radius of \mathcal{B} depends on the variable in \mathcal{R} , we can not overlay the
 91 solution on each component to get the eigenvalues. But the separate variable is still working. Let
 92 $l = s - \frac{L}{2}$, the Laplacian-Beltrami operator on \mathcal{S} is

$$\Delta_{\mathcal{S}} = \frac{\partial^2}{\partial l^2} + \frac{1}{\varepsilon_0 \cosh^2 al} \Delta_{\mathcal{B},1} \quad (29)$$

Thus, let $\psi^S = X(l)Y(r, \theta)$, the eigenvalue equation is:

$$\frac{d^2 X}{dl^2} Y + \frac{1}{\varepsilon_0^2 \cosh^2(al)} X \Delta_{B,1} Y = -\lambda^S XY \quad (30)$$

Dividing XY , we have

$$\frac{1}{X} \frac{d^2 X}{dl^2} + \frac{1}{\varepsilon_0^2 \cosh^2(al)} \frac{\Delta_{B,1} Y}{Y} = -\lambda^S \quad (31)$$

Set Y to be the eigenfunction of $\Delta_{B,1}$ as in the first case, we have

$$\frac{d^2 X}{dl^2} + \left(\lambda^S - \frac{\mu^{B,1}}{\varepsilon_0^2 \cosh^2(al)} \right) X = 0 \quad (32)$$

Take $z = \tanh al$,

$$\begin{aligned} \frac{d}{dl} &= \frac{a}{\cosh^2(al)} \frac{d}{dz} = a(1-z^2) \frac{d}{dz} \\ \frac{d^2}{dl^2} &= a^2(1-z^2)^2 \frac{d^2}{dz^2} - 2a^2 z(1-z^2) \frac{d}{dz} \\ &\Rightarrow a^2(1-z^2)^2 \frac{d^2 x}{dz^2} - 2a^2 z(1-z^2) \frac{dx}{dz} + \left(\lambda^S - \frac{\mu^{B,1}(1-z^2)}{\varepsilon_0^2} \right) X = 0 \\ &\Rightarrow (1-z^2) \frac{d^2 X}{dz^2} - 2z \frac{dX}{dz} + \left(\frac{\lambda^S}{a^2(1-z^2)} - \frac{\mu^{B,1}}{a^2 \varepsilon_0^2} \right) X = 0, \end{aligned} \quad (33)$$

which is the associated Legendre differential equation:

$$(1-z^2) \frac{d^2 X}{dz^2} - 2z \frac{dX}{dz} + \left[\ell(\ell+1) - \frac{m^2}{1-z^2} \right] X = 0 \quad (34)$$

with $\ell(\ell+1) = -\frac{\mu^{B,1}}{a^2 \varepsilon_0^2}$, $m^2 = -\frac{\lambda^S}{a^2}$, where ℓ, m are the degree and order of the associated Legendre polynomial, respectively.

We obtain the solution

$$\begin{aligned} X(z) &= C_1 P_\ell^m(z) + C_2 Q_\ell^m(z) \\ &\Rightarrow X(l) = C_1 P_\ell^m(\tanh(al)) + C_2 Q_\ell^m(\tanh(al)) \end{aligned} \quad (35)$$

where P_ℓ^m is the associated Legendre polynomials, Q_ℓ^m is Legendre function of the second kind.

Therefore, for the Dirichlet boundary condition,

$$X(0)Y(\varepsilon_0 \cosh a \frac{L}{2}, \theta) = X(L)Y(\varepsilon_0 \cosh a \frac{L}{2}, \theta) = 0 \quad (36)$$

We find that the spectral gap λ^D mainly relates to $\left(\frac{j_{\nu,1}}{\varepsilon_0 \cosh a \frac{L}{2}} \right)^2$, since $\varepsilon_0 \ll L$, we have

$$\lambda^D \sim \left(\frac{j_{\nu,1}}{\varepsilon_0 \cosh a \frac{L}{2}} \right)^2 \quad (37)$$

for the Neumann boundary condition,

$$\begin{aligned} X'(0)Y(\varepsilon_0 \cosh a \frac{L}{2}, \theta) + X(0)Y'(\varepsilon_0 \cosh a \frac{L}{2}, \theta) &= 0 \\ X'(L)Y(\varepsilon_0 \cosh a \frac{L}{2}, \theta) + X(L)Y'(\varepsilon_0 \cosh a \frac{L}{2}, \theta) &= 0, \end{aligned} \quad (38)$$

the spectral gap λ^N mainly relates to $\left(\frac{j_{\nu,0}}{\varepsilon_0 \cosh a \frac{L}{2}} \right)^2 + \left(\frac{j'_{0,0}}{\varepsilon_0 \cosh a \frac{L}{2}} \right)^2$, meaning that

$$\Rightarrow \lambda^N \sim \left(\frac{j_{\nu,0}}{\varepsilon_0 \cosh a \frac{L}{2}} \right)^2 \quad (39)$$

□

108 B.4 Proofs of Corollary 5.2

109 **Corollary 5.2 (Adaptivity of Robin Condition).** For $\varepsilon(s) : [0, L] \rightarrow (0, \delta)$ and $L \gg \delta$ for a small
 110 $\delta > 0$, the spectral gaps $\lambda^D, \lambda^N, \lambda$ of the Dirichlet condition, Neumann condition, and no boundary
 111 have the following order:

- 112 • For $\varepsilon(s) = \varepsilon_0$, $\lambda^D \geq \lambda \geq \lambda^N$.
- 113 • For $\varepsilon(s) = \varepsilon_0 \cosh a(s - \frac{L}{2})$, if L is fixed, then $\lambda^D \geq \lambda^N \geq \lambda$; if ε_0 is fixed, then
 114 $\lambda \geq \lambda^D \geq \lambda^N$

115 *Proof.* From Cheeger and Buser's inequality (Theorem ??), if the manifold is closed, the spectral
 116 gap of its Laplace-Beltrami operator $\lambda \sim \inf \varepsilon^{n-1}$ when taking S as a submanifold embedded in a
 117 closed manifold.

118 For $\varepsilon(s) = \varepsilon_0$, since ε_0 is small, ε_0^{-2} will be large, and ε_0^{n-1} will be small, thus we have the order

$$\lambda^D \geq \lambda \geq \lambda^N$$

119 For $\varepsilon(s) = \varepsilon_0 \cosh a(s - \frac{L}{2})$, from the proofs of Theorem ??, since $j_{\nu,1} > j_{\nu,0}$, we have $\lambda^D > \lambda^N$.
 120 If L is fixed, since ε_0^{-2} will be large, $\lambda^D \geq \lambda^N \geq \lambda$; if ε_0 is fixed, since $L \gg \varepsilon_0$, $\cosh^{-2} a \frac{L}{2}$ will be
 121 much more smaller than ε_0^{n-1} , then $\lambda \geq \lambda^D \geq \lambda^N$ \square

122 B.5 Proofs of Lemma 5.3

123 **Lemma 5.3 (Oversmoothing under Source Terms).** For the nonhomogeneous heat equation
 124 $\partial_t u(t, \mathbf{x}) = \Delta_g u(t, \mathbf{x}) + f(t, \mathbf{x})$ with a source term $f(t, \mathbf{x}) = \sum_i B_i(t) \psi_i(\mathbf{x})$ and initial condition
 125 $u(0, \mathbf{x}) = \phi(\mathbf{x})$ at $t = 0$, regardless of the existence of boundary condition, there exist some
 126 parameters $\{\alpha_i > 2, \beta_i > 0\}$ satisfying the convergence condition such that

$$\frac{dB_i}{dt} = (1 - \alpha_i + \lambda_i(\beta_i + t))(\beta_i + t)^{-\alpha_i}, \quad B_i(0) = 0.$$

127 Then, the Dirichlet energy will decay polynomially w.r.t time t as

$$Dir[u_t] = \sum_{i=1} \frac{1}{2 - \alpha_i} (\beta_i + t)^{2 - \alpha_i} - \sum_{i=1} (1 + \frac{\lambda_i \beta_i}{2 - \alpha_i}) \beta_i^{1 - \alpha_i}.$$

128

129 *Proof.* Consider the nonhomogeneous heat equation:

$$\begin{cases} \partial_t u(t, x) = \Delta u(t, x) + f(t, x), & x \in \text{int}\mathcal{M}, t > 0 \\ h(x) = 0, & x \in \partial\mathcal{M}, t > 0 \\ u(0, x) = \phi(x), & x \in \mathcal{M}, t = 0 \end{cases} \quad (40)$$

130 The solution is

$$u(t, x) = \int_0^t \int_{\mathcal{M}} H(t - s, x, y) f(s, y) \text{vol}_{\mathcal{M}}(y) ds + \int_{\mathcal{M}} H(t, x, y) \phi(y) \text{vol}_{\mathcal{M}}(y) \quad (41)$$

131 Let $f(t, x) = \sum_i \beta_i(t) \psi_i(x)$, $\beta_i(t) = \langle f_t, \psi_i \rangle$, $\phi(x) = \sum_i c_i \psi_i(x)$, and $c_i = \langle \phi, \psi_i \rangle$. Define

$$u^1(t, x) := \int_0^t \int_{\mathcal{M}} H(t - s, x, y) f(s, y) \text{vol}_{\mathcal{M}}(y) ds \quad (42)$$

$$u^2(t, x) := \int_{\mathcal{M}} H(t, x, y) \phi(y) \text{vol}_{\mathcal{M}}(y) \quad (43)$$

132 Then

$$\begin{aligned}
u^1(t, x) &= \int_0^t \int_{\mathcal{M}} \left[\sum_{i=0}^{\infty} e^{-\lambda_i(t-s)} \psi_i(x) \psi_i(y) \right] \left[\sum_j \beta_j(s) \psi_j(y) \right] \text{vol}_{\mathcal{M}}(y) ds \\
&= \int_0^t \sum_{i=0}^{\infty} e^{-\lambda_i(t-s)} \beta_i(s) \psi_i(x) ds \\
&= \sum_{i=0}^{\infty} e^{-\lambda_i t} \int_0^t e^{\lambda_i s} \beta_i(s) ds \psi_i(x)
\end{aligned} \tag{44}$$

133

$$\begin{aligned}
u^2(t, x) &= \sum_{i=0}^{\infty} e^{-\lambda_i t} c_i \psi_i(x) \int_{\mathcal{M}} \psi_i(y) \psi_i(y) \text{vol}_{\mathcal{M}}(y) \\
&= \sum_{i=0}^{\infty} c_i e^{-\lambda_i t} \psi_i(x)
\end{aligned} \tag{45}$$

$$\tag{46}$$

134 For Dirichlet energy, we have

$$\text{Dir}[u_t] = \frac{1}{2} \int_{\mathcal{M}} \|\text{grad } u_t\|^2 \text{vol}_{\mathcal{M}} = \text{Dir}[u_t^1] + \text{Dir}[u_t^2] + \int_{\mathcal{M}} \langle \text{grad } u_t^1, \text{grad } u_t^2 \rangle_g \text{vol}_{\mathcal{M}} \tag{47}$$

135 For each item, apply the divergence theorem or Green's first identity,

$$\text{Dir}[u_t^1] = -\frac{1}{2} \int_{\mathcal{M}} u_t^1 \Delta_g u_t^1 \text{vol}_{\mathcal{M}} = \frac{1}{2} \sum_{i=1}^{\infty} \frac{1}{\lambda_i} \left[B_i(t) - B_i(0) e^{-\lambda_i t} - e^{-\lambda_i t} \int_0^t e^{\lambda_i s} \frac{dB_i}{ds} ds \right] \tag{48}$$

136

$$\text{Dir}[u_t^2] = -\frac{1}{2} \int_{\mathcal{M}} u_t^2 \Delta u_t^2 \text{vol}_{\mathcal{M}} = \sum_{i=1}^{\infty} c_i^2 \lambda_i e^{-\lambda_i t} \tag{49}$$

137

$$\begin{aligned}
\int_{\mathcal{M}} \langle \text{grad } u_t^1, \text{grad } u_t^2 \rangle d \text{vol}_{\mathcal{M}} &= - \int_{\mathcal{M}} u_t^2 \Delta_g u_t^1 \text{vol}_{\mathcal{M}} \\
&= \sum_{i=1}^{\infty} c_i e^{-\lambda_i t} \left[\beta_i(t) - \beta_i(0) e^{-\lambda_i t} - e^{-\lambda_i t} \int_0^t e^{\lambda_i s} \frac{d\beta_i}{ds} ds \right]
\end{aligned} \tag{50}$$

138 If $B_i(t)$ satisfies

$$\begin{cases} B_i(0) = 0 \\ \frac{dB_i}{dt} = (1 - a_i + \lambda_i(b_i + t))(b_i + t)^{-a_i} \end{cases} \tag{51}$$

139 for $0 < b_i < 1, a_i > 2$, i.e.,

$$B_i(t) = \frac{\lambda_i(b_i + t)^{2-a_i}}{2 - a_i} + (b_i + t)^{1-a_i} - \left(1 + \frac{\lambda_i b_i}{2 - a_i} \right) b_i^{1-a_i} \tag{52}$$

140 Then,

$$\begin{aligned}
E[u_t^1] &= \sum_{i=1}^{\infty} \frac{1}{\lambda_i} \left[B_i(t) - e^{-\lambda_i t} \int_0^t e^{\lambda_i s} \frac{dB_i}{ds} ds \right] \\
&= \sum_{i=1}^{\infty} \frac{1}{2 - a_i} (b_i + t)^{2-a_i} - \left(1 + \frac{\lambda_i b_i}{2 - a_i} \right) b_i^{1-a_i}
\end{aligned} \tag{53}$$

141 So the decay of the Dirichlet energy is mainly determined by

$$\sum_{i=1}^{\infty} \frac{1}{2 - a_i} (b_i + t)^{2-a_i}$$

142 which is polynomially decreasing. \square

143 B.6 Proofs of Theorem 5.4

144 **Theorem 5.4 (Equivalence).** For the homogeneous heat equation $\partial_t u(t, \mathbf{x}) = \Delta_g u(t, \mathbf{x})$, $\mathbf{x} \in$
 145 $\text{int}\mathcal{M}$ with a nonhomogeneous boundary condition of $h(\mathbf{x}) = \gamma(t, \mathbf{x})$, $\mathbf{x} \in \partial\mathcal{M}$ and initial condition
 146 $u(0, \mathbf{x}) = \phi(\mathbf{x})$, $t = 0$, there exists a function $w(t, \mathbf{x}) : \mathcal{M} \rightarrow \mathbb{R}$ such that $h \circ w = \gamma$ in $\partial\mathcal{M}$. By
 147 $v = u - w$, Eq. (??) is equivalent to the nonhomogeneous heat equation satisfying Lemma ??:

$$\begin{cases} \partial_t v(t, \mathbf{x}) = \Delta_g v(t, \mathbf{x}) + \Gamma(w), & \mathbf{x} \in \text{int}\mathcal{M}, t > 0 \\ h(\mathbf{x}) = 0, & \mathbf{x} \in \partial\mathcal{M}, t > 0, \\ v(0, \mathbf{x}) = \phi(\mathbf{x}) - w(0, \mathbf{x}), & \mathbf{x} \in \mathcal{M}, t = 0 \text{ (The initial condition)}, \end{cases}$$

148 where $\Gamma(w) = \Delta_g w(t, \mathbf{x}) - \partial_t w(t, \mathbf{x})$.

149 *Proof.* Introducing an auxiliary function $v = u - w$ such that $h \circ w = 0$ in $\partial\mathcal{M}$. Then,

$$\begin{cases} \Delta_g v = \Delta_g u - \Delta_g w \\ \partial_t v = \partial_t u - \partial_t w \end{cases}$$

150

$$\Rightarrow \partial_t v(t, x) = \Delta_g v(t, x) + \Delta_g w(t, x) - \partial_t w(t, x) \quad (54)$$

151 Given the linearity of all three types of boundary condition, we define

$$\Gamma(w) = \Delta_g w(t, x) - \partial_t w(t, x) \quad (55)$$

152 which satisfies the condition in Lemma ?? on $\text{int}\mathcal{M}$.

153 The equation can be converted into the form in Lemma ??,

$$\begin{cases} \partial_t v(t, x) = \Delta_g v(t, x) + \Gamma(w), & x \in \text{int}\mathcal{M}, t > 0 \\ v(0, x) = \phi(x) - w(0, x), & x \in \mathcal{M}, t = 0 \\ h(x) = 0, & x \in \partial\mathcal{M}, t > 0 \end{cases} \quad (56)$$

154

□

155 B.7 Proofs of Lemma 6.2

156 **Lemma 6.2 (Solution to Heat Equation System).** Let \mathbf{D} , \mathbf{U} and \mathbf{V} denote the diagonal, upper
 157 and lower matrix of \mathbf{L} , respectively. Accordingly to the Jacobi method, the matrix form for the k -th
 158 iteration is derived as follows,

$$\text{diag}(\alpha_i) = \alpha(\mathbf{X}^{(k)}; \theta_\alpha), \text{diag}(\beta_i) = \beta(\mathbf{X}^{(k)}; \theta_\beta), \mathbf{\Gamma} = \gamma(\mathbf{X}^{(k)}; \theta_\gamma), \quad (57)$$

$$\mathbf{X}^{(k+1)} = \mathbf{D}^{-1}(\mathbf{V} + \mathbf{U})\mathbf{X}^{(k)} + \mathbf{D}^{-1}\mathbf{\Gamma}. \quad (58)$$

159 Let $I_i := \mathbb{I}(i \in S)$ be an indicator, $\hat{d}_i = d_i(1 - I_i) + (2I_i - 1) \sum_{j \sim i} I_j$, and $p_i = \frac{\beta_i}{\alpha_i}$.

$$[(\mathbf{D}^{-1}\mathbf{U})\mathbf{X}^{(k)}]_i = \frac{I_i}{\sqrt{\hat{d}_i}} \sum_{j \sim i} \frac{1}{\sqrt{\hat{d}_j}} \mathbf{x}_j^{(k)}, \quad [(\mathbf{D}^{-1}\mathbf{V})\mathbf{X}^{(k)}]_i = \frac{p_i + (1-p_i)I_i}{\sqrt{\hat{d}_i}} \sum_{j \sim i} \frac{I_j}{\sqrt{\hat{d}_j}} \mathbf{x}_j^{(k)}. \quad (59)$$

160

161 *Proof.* Merging the boundary condition and the steady solution, we have

$$\begin{bmatrix} \mathbf{L}_{S,S} & \mathbf{L}_{S,\partial S} \\ \text{diag}(\beta_i)\mathbf{L}_{\partial S,S} & \text{diag}(\alpha_i)\mathbf{I}_{\partial S,\partial S} \end{bmatrix} \begin{bmatrix} \mathbf{X}_S \\ \mathbf{X}_{\partial S} \end{bmatrix} = \begin{bmatrix} \mathbf{F}_S \\ \mathbf{\Gamma}_{\partial S} \end{bmatrix} = \mathbf{\Gamma} \quad (60)$$

162 Now, we use the Jacobi method to solve this nonlinear equation. The diagonal matrix is

$$\mathbf{D} = \begin{bmatrix} \mathbf{I}_{S,S} & \mathbf{0}_{S,\partial S} \\ \mathbf{0}_{\partial S,S} & \text{diag}(\alpha_i)\mathbf{I}_{\partial S,\partial S} \end{bmatrix} \quad (61)$$

163 The lower diagonal and upper diagonal matrices are

$$\mathbf{V} = \begin{bmatrix} \hat{\mathbf{A}}_{S,S}^{low} & \mathbf{0}_{S,\partial S} \\ -\text{diag}(\beta_i)\mathbf{L}_{\partial S,S} & \mathbf{0}_{\partial S,\partial S} \end{bmatrix}, \mathbf{U} = \begin{bmatrix} \hat{\mathbf{A}}_{S,S}^{up} & -\mathbf{L}_{S,\partial S} \\ \mathbf{0}_{\partial S,S} & \mathbf{0}_{\partial S,\partial S} \end{bmatrix} \quad (62)$$

164 For the k -th iteration,

$$\text{diag}(\alpha_i^k) = \alpha(\mathbf{X}^k; \theta_k), \text{diag}(\beta_i^k) = \beta(\mathbf{X}^k; \omega_k), \mathbf{\Gamma} = \gamma(\mathbf{X}^k; \eta_k) \quad (63)$$

165

$$\mathbf{X}^{k+1} = \mathbf{D}^{-1}(\mathbf{V} + \mathbf{U})\mathbf{X}^k + \mathbf{D}^{-1}\mathbf{\Gamma} \quad (64)$$

166 The (i, j) -entry of $\mathbf{D}^{-1}\mathbf{V}$ is

$$(\mathbf{D}^{-1}\mathbf{V})_{ij} = \begin{cases} \frac{1}{\sqrt{d_i^S d_j^S}} & , i \sim j, i \in S, j \in S \\ 0 & , i \sim j, i \in \partial S, j \in \partial S \\ \frac{\beta_i}{\alpha_i \sqrt{d_i^{\partial S} d_j^S}} & , i \sim j, i \in \partial S, j \in S \\ 0 & , else \end{cases} \quad (65)$$

167 The (i, j) -entry of $\mathbf{D}^{-1}\mathbf{U}$ is

$$(\mathbf{D}^{-1}\mathbf{U})_{ij} = \begin{cases} \frac{1}{\sqrt{d_i^S d_j^S}} & , i \sim j, i \in S, j \in S \\ 0 & , i \sim j, i \in \partial S, j \in \partial S \\ \frac{1}{\sqrt{d_i^S d_j^{\partial S}}} & , i \sim j, i \in S, j \in \partial S \\ 0 & , else \end{cases} \quad (66)$$

168

□

169 B.8 Proofs of Theorem 6.3

170 Before proof, we first begin with the following lemma.

171 **Lemma B.1** ([10]). *For the matrix function $[\mathbf{T}(\mathbf{X})]_{ij} := T_{ij}(\mathbf{x}_i)$. If the spectral radius $\rho[\mathbf{T}(\mathbf{X})] < 1$,*
 172 *then*

$$\sum_{k=0}^{\infty} \mathbf{T}(\mathbf{X})^k = (\mathbf{I} - \mathbf{T}(\mathbf{X}))^{-1} \quad (67)$$

173 **Theorem 6.3 (Distance Independence).** *Given two nodes v_i and v_j with distance K , the measure*
 174 *of information transport*

$$\left\| \frac{\partial \mathbf{x}_i^{(K)}}{\partial \mathbf{x}_j} \right\| \leq C_K \left(\sum_{k=0}^K [\mathbf{D}^{-1}(\mathbf{V} + \mathbf{U})]^k \right)_{ij}$$

175 where C_K is a constant related to the GBN layer. As $K \rightarrow \infty$, $\left| \frac{\partial \mathbf{x}_i^K}{\partial \mathbf{x}_j} \right|$ is independent of the distance
 176 K .

177 *Proof.* We give the proof under the scalar-valued case, the vector-valued case can be proved similarly.
 178 Denote $T_{ij} = [\mathbf{D}^{-1}(\mathbf{V} + \mathbf{U})]_{ij}$ and $b_i = (\mathbf{D}^{-1}\mathbf{\Gamma})_i$, then Eq. (??) can be represented element-wise
 179 as

$$x_i^{(k+1)} = \sigma \left(\sum_{j \sim i} T_{ij} \varphi^{(k+1)}(x_j^{(k)}) \right) + b_i(x_i^{(k)}) \quad (68)$$

180 for $k \geq 0$. Now we prove the conclusion

$$\left| \frac{\partial x_i^{(k)}}{\partial x_j} \right| \leq C_k (\delta_{ij} + T_{ij} + \sum_{m=2}^k \sum_{j_m, \dots, j_2} T_{ij_m} T_{j_m j_{m-1}} \dots T_{j_2 j}) \quad (69)$$

181 by induction. For $k = 1$, we have

$$x_i^{(1)} = \sigma \left(\sum_{j \sim i} T_{ij} \varphi^{(1)}(x_j) \right) + b_i(x_i) \quad (70)$$

182 The Jacobian is

$$\begin{aligned}
\left| \frac{\partial x_i^{(1)}}{\partial x_j} \right| &= \left| \sigma' \cdot \left(\frac{\partial T_{ij}}{\partial x_i} \frac{\partial x_i}{\partial x_j} \varphi^{(1)}(x_j) + T_{ij} \varphi'^{(1)}(x_j) \right) + \frac{\partial b_i}{\partial x_i} \frac{\partial x_i}{\partial x_j} \right| \\
&= \left| \sigma' \cdot \left(\frac{\partial T_{ij}}{\partial x_i} \varphi^{(1)}(x_j) \delta_{ij} + T_{ij} \varphi'^{(1)}(x_j) \right) + \frac{\partial b_i}{\partial x_i} \delta_{ij} \right| \\
&\leq \left| \sigma' \cdot \frac{\partial T_{ij}}{\partial x_i} \varphi^{(1)}(x_j) + \frac{\partial b_i}{\partial x_i} \right| \delta_{ij} + \left| \sigma' \cdot \varphi'^{(1)}(x_j) \right| T_{ij} \\
&\leq C_1(\delta_{ij} + T_{ij}),
\end{aligned} \tag{71}$$

183 where $C_1 = \max \left(\sup \left| \sigma' \cdot \frac{\partial T_{ij}}{\partial x_i} \varphi^{(1)}(x_j) + \frac{\partial b_i}{\partial x_i} \right|, \sup \left| \sigma' \cdot \varphi'^{(1)}(x_j) \right| \right)$.

184 Suppose Eq. (69) holds for $k = K$, i.e.,

$$\left| \frac{\partial x_i^{(K)}}{\partial x_j} \right| \leq C_K(\delta_{ij} + T_{ij} + \sum_{m=2}^K \sum_{j_m, \dots, j_2} T_{ij_m} T_{j_m j_{m-1}} \dots T_{j_2 j}) \tag{72}$$

185 Then, for $k = K + 1$, by the chain rule, we have

$$\left| \frac{\partial x_i^{(K+1)}}{\partial x_j} \right| = \left| \sum_{j_{K+1}} \frac{\partial x_i^{(K+1)}}{\partial x_{j_{K+1}}^{(K)}} \frac{\partial x_{j_{K+1}}^{(K)}}{\partial x_j} \right| \leq \sum_{j_{K+1}} \left| \frac{\partial x_i^{(K+1)}}{\partial x_{j_{K+1}}^{(K)}} \right| \left| \frac{\partial x_{j_{K+1}}^{(K)}}{\partial x_j} \right| \tag{73}$$

186

$$\begin{aligned}
\left| \frac{\partial x_i^{(K+1)}}{\partial x_{j_{K+1}}^{(K)}} \right| &= \left| \sigma' \cdot \left(\frac{\partial T_{ij_{K+1}}}{\partial x_i^{(K)}} \frac{\partial x_i^{(K)}}{\partial x_{j_{K+1}}^{(K)}} \varphi^{(K+1)}(x_{j_{K+1}}^{(K)}) + T_{ij_{K+1}} \varphi'^{(K+1)}(x_{j_{K+1}}^{(K)}) \right) + \frac{\partial b_i}{\partial x_i^{(K)}} \frac{\partial x_i^{(K)}}{\partial x_{j_{K+1}}^{(K)}} \right| \\
&= \left| \sigma' \cdot \left(\frac{\partial T_{ij_{K+1}}}{\partial x_i^{(K)}} \varphi^{(K+1)}(x_{j_{K+1}}^{(K)}) \delta_{ij_{K+1}} + T_{ij_{K+1}} \varphi'^{(K+1)}(x_{j_{K+1}}^{(K)}) \right) + \frac{\partial b_i}{\partial x_i^{(K)}} \delta_{ij_{K+1}} \right| \\
&\leq \left| \sigma' \cdot \frac{\partial T_{ij_{K+1}}}{\partial x_i^{(K)}} \varphi^{(K+1)}(x_{j_{K+1}}^{(K)}) + \frac{\partial b_i}{\partial x_i^{(K)}} \right| \delta_{ij_{K+1}} + \left| \sigma' \cdot \varphi'^{(K+1)}(x_{j_{K+1}}^{(K)}) \right| T_{ij_{K+1}} \\
&\leq C_{K+1,K}(\delta_{ij_{K+1}} + T_{ij_{K+1}}),
\end{aligned} \tag{74}$$

187 where $C_{K+1,K} = \max \left(\sup \left| \sigma' \cdot \frac{\partial T_{ij_{K+1}}}{\partial x_i^{(K)}} \varphi^{(K+1)}(x_{j_{K+1}}^{(K)}) + \frac{\partial b_i}{\partial x_i^{(K)}} \right|, \sup \left| \sigma' \cdot \varphi'^{(K+1)}(x_{j_{K+1}}^{(K)}) \right| \right)$.

$$\left| \frac{\partial x_{j_{K+1}}^{(K)}}{\partial x_j} \right| \leq C_K(\delta_{j_{K+1}j} + T_{j_{K+1}j} + \sum_{m=2}^K \sum_{j_m, \dots, j_2} T_{j_{K+1}j_m} T_{j_m j_{m-1}} \dots T_{j_2 j}) \tag{75}$$

188 Substituting into Eq. (73),

$$\begin{aligned}
\left| \frac{\partial x_i^{(K+1)}}{\partial x_j} \right| &\leq C_K C_{K+1,K} \left[\delta_{ij_{K+1}} \left(\delta_{j_{K+1}j} + T_{j_{K+1}j} + \sum_{m=2}^K \sum_{j_m, \dots, j_2} T_{j_{K+1}j_m} T_{j_m j_{m-1}} \dots T_{j_2 j} \right) \right] \\
&\quad + C_K C_{K+1,K} \left[T_{ij_{K+1}} \left(\delta_{j_{K+1}j} + T_{j_{K+1}j} + \sum_{m=2}^K \sum_{j_m, \dots, j_2} T_{j_{K+1}j_m} T_{j_m j_{m-1}} \dots T_{j_2 j} \right) \right] \\
&= C_K C_{K+1,K} (\delta_{ij} + 2T_{ij} + \sum_{m=2}^K \sum_{j_m, \dots, j_2} T_{ij_m} T_{j_m j_{m-1}} \dots T_{j_2 j} \\
&\quad + \sum_{j_{K+1}} T_{ij_{K+1}} T_{j_{K+1}j} + \sum_{m=2}^{K+1} \sum_{j_m, \dots, j_2} T_{ij_m} T_{j_m j_{m-1}} \dots T_{j_2 j}) \\
&= C_K C_{K+1,K} (\delta_{ij} + 2T_{ij} + 3 \sum_{j_2} T_{ij_2} T_{j_2 j} \\
&\quad + 2 \sum_{m=3}^K \sum_{j_m, \dots, j_2} T_{ij_m} T_{j_m j_{m-1}} \dots T_{j_2 j} + \sum_{j_{K+1}, \dots, j_2} T_{ij_m} T_{j_m j_{m-1}} \dots T_{j_2 j}) \\
&= 9C_K C_{K+1,K} (\delta_{ij} + T_{ij} + \sum_{m=2}^{K+1} \sum_{j_m, \dots, j_2} T_{ij_m} T_{j_m j_{m-1}} \dots T_{j_2 j}). \tag{76}
\end{aligned}$$

189 Let $C_{K+1} = 9C_K C_{K+1,K}$, we finish the induction.

190 Since $\left| \frac{\alpha(x_i)}{\beta(x_i)} - \frac{\alpha(x_j)}{\beta(x_j)} \right| \leq |x_i - x_j|$ is a contraction, and the Laplacian \mathbf{L} is normalized, $\rho(\mathbf{T}) < 1$.

191 From Lemma B.1, as $K \rightarrow \infty$,

$$\lim_{K \rightarrow \infty} \left| \frac{\partial x_i^{(K)}}{\partial x_j} \right| = [(\mathbf{I} - \mathbf{T})^{-1}]_{ij}, \tag{77}$$

192 where $\mathbf{T} = \mathbf{D}^{-1}(\mathbf{V} + \mathbf{U})$. □

193 B.9 Derivation of GBN formalism

194 Set Robin condition coefficients $\alpha_i = \alpha(\mathbf{x}_i)$, $\beta_i = \beta(\mathbf{x}_i)$, $\gamma_i = \gamma(\mathbf{x}_i)$, then Robin condition becomes

$$\alpha_i \mathbf{x}_i + \beta_i \sum_{j \sim i, j \in S} (\mathbf{x}_i - \mathbf{x}_j) = \gamma_i \tag{78}$$

195 for each $v_i \in \partial S$.

196 Let the normalized graph Laplacian be reordered to a block matrix

$$\mathbf{L} = \begin{bmatrix} \mathbf{L}_{S,S} & \mathbf{L}_{S,\partial S} \\ \mathbf{L}_{\partial S,S} & \mathbf{L}_{\partial S,\partial S} \end{bmatrix}. \tag{79}$$

197 Then the Robin condition can be converted into matrix form:

$$\text{diag}(\alpha_i) \mathbf{X}_{\partial S} + \text{diag}(\beta_i) \mathbf{L}_{\partial S,S} \mathbf{X}_S = \mathbf{\Gamma}_{\partial S} \tag{80}$$

198 Consider the steady solution of the nonhomogeneous heat equation. We have

$$\mathbf{L}_{S,S} \mathbf{X}_S + \mathbf{L}_{S,\partial S} \mathbf{X}_{\partial S} = \mathbf{F}_S, \tag{81}$$

199 which is the form of Poisson's equation. Merging the above equations, we have

$$\begin{bmatrix} \mathbf{L}_{S,S} & \mathbf{L}_{S,\partial S} \\ \text{diag}(\beta_i) \mathbf{L}_{\partial S,S} & \text{diag}(\alpha_i) \mathbf{I}_{\partial S,S} \end{bmatrix} \begin{bmatrix} \mathbf{X}_S \\ \mathbf{X}_{\partial S} \end{bmatrix} = \begin{bmatrix} \mathbf{F}_S \\ \mathbf{\Gamma}_{\partial S} \end{bmatrix} = \mathbf{\Gamma} \tag{82}$$

200 Now we use the Jacobi method to solve this nonlinear equation. The diagonal matrix is

$$\mathbf{D} = \begin{bmatrix} \mathbf{I}_{S,S} & \mathbf{0}_{S,\partial S} \\ \mathbf{0}_{\partial S,S} & \text{diag}(\alpha_i) \mathbf{I}_{\partial S,S} \end{bmatrix} \quad (83)$$

201 The lower diagonal and upper diagonal matrices are

$$\mathbf{V} = \begin{bmatrix} \hat{\mathbf{A}}_{S,S}^{low} & \mathbf{0}_{S,\partial S} \\ -\text{diag}(\beta_i) \mathbf{L}_{\partial S,S} & \mathbf{0}_{\partial S,S} \end{bmatrix} \quad (84)$$

$$\mathbf{U} = \begin{bmatrix} \hat{\mathbf{A}}_{S,S}^{up} & -\mathbf{L}_{S,\partial S} \\ \mathbf{0}_{\partial S,S} & \mathbf{0}_{\partial S,S} \end{bmatrix} \quad (85)$$

202 for the k -th iteration,

$$\text{diag}(\alpha_i) = \alpha(\mathbf{X}^{(k)}; \theta_k) \quad (86)$$

$$\text{diag}(\beta_i) = \beta(\mathbf{X}^{(k)}; \omega_k) \quad (87)$$

$$\mathbf{\Gamma} = \gamma(\mathbf{X}^{(k)}; \eta_k) \quad (88)$$

$$\mathbf{X}^{(k+1)} = \mathbf{D}^{-1}(\mathbf{V} + \mathbf{U})\mathbf{X}^{(k)} + \mathbf{D}^{-1}\mathbf{\Gamma} \quad (89)$$

203 The (i, j) -entry of $\mathbf{D}^{-1}\mathbf{V}$ is

$$(\mathbf{D}^{-1}\mathbf{V})_{ij} = \begin{cases} \frac{1}{\sqrt{d_i^S d_j^S}} & , i \sim j, i \in S, j \in S \\ 0 & , i \sim j, i \in \partial S, j \in \partial S \\ \frac{\beta_i}{\alpha_i \sqrt{d_i^{\partial S} d_j^S}} & , i \sim j, i \in \partial S, j \in S \\ 0 & , else \end{cases} \quad (90)$$

204 The (i, j) -entry of $\mathbf{D}^{-1}\mathbf{U}$ is

$$(\mathbf{D}^{-1}\mathbf{U})_{ij} = \begin{cases} \frac{1}{\sqrt{d_i^S d_j^S}} & , i \sim j, i \in S, j \in S \\ 0 & , i \sim j, i \in \partial S, j \in \partial S \\ \frac{1}{\sqrt{d_i^S d_j^{\partial S}}} & , i \sim j, i \in S, j \in \partial S \\ 0 & , else \end{cases} \quad (91)$$

205 To give a unified formula, we define the indicator $I_i := \mathbb{I}(i \in S)$, then we define

$$\hat{d}_i = \sum_{j \sim i} I_i I_j + (1 - I_i)(1 - I_j) = d_i(1 - I_i) + (2I_i - 1) \sum_{j \sim i} I_j \quad (92)$$

206 Then,

$$[(\mathbf{D}^{-1}\mathbf{U})\mathbf{X}^{(k)}]_i = \frac{I_i}{\sqrt{\hat{d}_i}} \sum_{j \sim i} \frac{1}{\sqrt{\hat{d}_j}} \mathbf{x}_j^{(k)} \quad (93)$$

$$[(\mathbf{D}^{-1}\mathbf{V})\mathbf{X}^{(k)}]_i = \frac{p_i + (1 - p_i)I_i}{\sqrt{\hat{d}_i}} \sum_{j \sim i} \frac{I_j}{\sqrt{\hat{d}_j}} \mathbf{x}_j^{(k)}, \quad (94)$$

207 where $p_i = \frac{\beta_i}{\alpha_i}$. Since $[\mathbf{D}^{-1}\mathbf{\Gamma}]_i = \mathbf{\Gamma}_i/\alpha_i$, we can replace it with the new variable $\mathbf{\Gamma}_i/\alpha_i = \gamma_i$.

208 Finally we get

$$x_i^{(k+1)} = \frac{I_i}{\sqrt{\hat{d}_i}} \sum_{j \sim i} \frac{1}{\sqrt{\hat{d}_j}} \mathbf{x}_j^{(k)} + \frac{p_i + (1 - p_i)I_i}{\sqrt{\hat{d}_i}} \sum_{j \sim i} \frac{I_j}{\sqrt{\hat{d}_j}} \mathbf{x}_j^{(k)} + \gamma_i \quad (95)$$

Algorithm 1 Graph Boundary conditioned message passing Neural network (GBN)

Input: Graph $\mathcal{G}(\mathcal{V}, \mathcal{E}, \mathbf{A})$, Objective function \mathcal{J} , Number of layers K , The activation function σ .
MLPs with learnable parameters $\alpha(\cdot; \theta_\alpha), \beta(\cdot; \theta_\beta), \gamma(\cdot; \theta_\gamma), \varsigma(\cdot, \eta), \varphi(\cdot, \omega_k)$.

Output: Node representations $\mathbf{X}^{(K)}$, Model parameters.

```

1: while not converge do
2:   Input initial value  $\mathbf{X}^{(0)} = \varphi(\mathbf{X}, \omega_0)$ ;
3:   Compute soft weights for being boundary node by  $I_i = \varsigma(\mathbf{x}_i, \eta)$ ;
4:   for each layer  $k = 1$  to  $K$  do
5:     Compute coefficients  $\alpha_i = \alpha(\mathbf{x}_i^{(k-1)}; \theta_\alpha), \beta_i = \beta(\mathbf{x}_i^{(k-1)}; \theta_\beta), \gamma_i = \gamma(\mathbf{x}_i^{(k-1)}; \theta_\gamma)$ ;
6:     Update node representations  $\mathbf{X}^{(k)}$  by Eq. (??);
7:   end for
8:   Compute the objective function  $\mathcal{J}(\mathbf{X}^{(K)}, \mathbf{Y})$ ;
9:   Train model parameters by Adam optimizer;
10: end while

```

Computational Complexity Computing boundary condition coefficients costs $\mathcal{O}(|\mathcal{V}|d)$, where d is the hidden layer dimension. The matrix multiplication is implemented as neighborhood aggregation of complexity $\mathcal{O}(|\mathcal{E}|)$. The overall complexity of a K -layer GBN is yielded as $\mathcal{O}(K(|\mathcal{V}| + |\mathcal{E}|))$.

D Datasets and Baselines

D.1 Datasets

We evaluate our method on a range of benchmark datasets spanning diverse domains and structural characteristic. Summary statistics are shown in Table 2.

WikiCS A citation graph of Wikipedia articles on Computer Science. Nodes are articles with averaged word embeddings as features; edges represent hyperlinks. The goal is to classify articles into 10 subfields.

Amazon-Computers A product co-purchase graph focused on the ‘‘Computers’’ category from Amazon. Nodes represent products, edges indicate co-purchases, and features are bag-of-words from reviews. The task is product category classification.

Coauthor-CS A co-authorship network from Microsoft Academic Graph. Nodes are authors, edges indicate collaborations, and features are derived from publication texts. The objective is to classify authors into research fields.

Texas A WebKB dataset based on webpages from the University of Texas. Nodes are webpages, edges are hyperlinks, and features come from bag-of-words. The task is to classify webpages into categories like faculty, course, or student.

Wisconsin Similar to Texas, but based on webpages from the University of Wisconsin. Nodes are webpages, edges are hyperlinks, and features use bag-of-words. It is commonly used for evaluating models under small-scale settings.

Roman-Empire A word-level dependency graph from the ‘‘Roman Empire’’ Wikipedia article. Nodes are words; edges connect sequential or syntactically related words. The task is to predict the syntactic role of each word in context. The graph exhibits strong heterophily.

Amazon-Ratings A bipartite graph of users and products from Amazon reviews. Edges represent ratings, with weights as scores. Node features are extracted from review texts. The task is to predict product categories or user preferences.

D.2 Baselines

- GCN [8] employs neighborhood aggregation within the spectral domain.

Table 2: Dataset statistics. The reported number of edges refers to directed edges; for undirected graphs, this count is twice the actual number of edges.

Dataset	Nodes	Edges	Average Degree	Node Features	Classes	Metric
WikiCS	11,701	431,726	36.90	300	10	Accuracy
Amazon-Computers	13,381	491,722	35.76	767	10	Accuracy
Coauthor CS	18,333	163,788	8.93	6,805	15	Accuracy
Texas	183	650	3.55	1,703	5	Accuracy
Wisconsin	251	1030	4.10	1,703	5	Accuracy
Roman-Empire	22,662	65,854	2.91	300	18	Accuracy
Amazon-Ratings	24,492	186100	7.60	300	5	Accuracy

- **GAT** [15] incorporates the attention mechanism into graph-based learning to dynamically weigh neighbor nodes.
- **GIN** [18] enhances graph learning by leveraging MLPs to achieve maximum discriminative power, mimicking the Weisfeiler-Lehman graph isomorphism test process.
- **DR** [5] applies Delaunay triangulation to the node features to rewire the graph, creating a new topology that alters the structural properties to mitigate over-squashing and over-smoothing in graph neural networks.
- **GIN+graphv2** [12] involves processing graph pairs using Graph Isomorphism Networks (GIN) without normalization, focusing on node and graph classification tasks. This approach applies GIN in a batch setting to evaluate performance on various graph datasets.
- **UniFilter** [6] combines low-pass and high-pass filters within each layer and adaptively integrates their embeddings. It also trains a coefficient matrix to measure node correlations for global aggregation.
- **ProxyGap** [7] uses spectral graph pruning to eliminate edges causing over-squashing and over-smoothing, adjusting the graph structure for enhanced neural network performance.
- **BorF** [9] applies Ollivier-Ricci curvature to identify and mitigate over-smoothing and over-squashing in graph neural networks by analyzing and rewriting the graph edges based on curvature values.
- **G²-GCN** [11] introduces gradient gating to dynamically adjust the contribution of each node’s gradient during training, allowing different parts of the graph to learn at varying speeds.
- **GREAD** [2] models graph neural networks using a reaction-diffusion process, where node features evolve through diffusion and reaction steps to capture complex graph structures and dynamics.
- **SWAN** [4] examines oversquashing in graph neural networks through the lens of dynamical systems, analyzing how information flow and node interactions affect the network’s ability to propagate information effectively.
- **CoBFormer** [17] addresses the over-globalizing problem in graph transformers by selectively focusing on local structures and reducing the emphasis on global information, thereby improving the model’s ability to capture fine-grained details.
- **VCR-Graphormer** [3] introduces virtual connections to enable mini-batch processing in graph transformers, allowing for efficient training on large graphs by simulating connections that facilitate information exchange between nodes.
- **Spexphormer** [13] proposes a technique to make graph transformers sparser by selectively retaining important connections and pruning less significant ones, aiming to improve computational efficiency without losing essential information.
- **HGCN** [1] generalizes GAT in the Lorentz model of hyperbolic space in which the graph convolution is conducted in the tangent space.

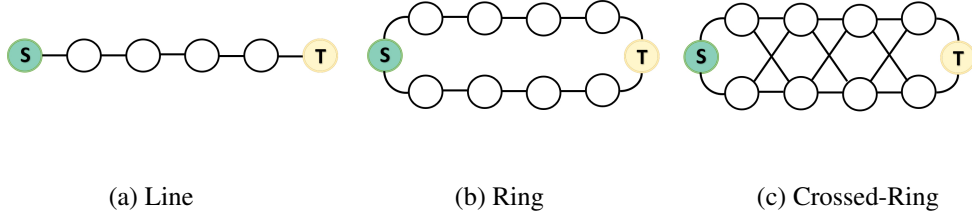


Figure 1: In all three graphs—Line, Ring, and Crossed-Ring—the source node “S” and target node “T” are placed five hops apart.

- **Graph-mamba** [16] focuses on long-range graph sequence modeling by employing selective state spaces to capture distant dependencies and interactions within graph sequences, enhancing the model’s ability to handle complex temporal dynamics.
- **MPNN+VN** [14] investigates the impact of virtual nodes on oversquashing and node heterogeneity in graph neural networks, analyzing how virtual nodes can alter information flow and representation learning.

E Implementation Notes

E.1 Graph transfer

We construct graph transfer datasets inspired by (Gravina et al. 2025), focusing on three types of graph structures: **Line**, **Ring**, and **Crossed-Ring**. Each graph within a task shares the same topology but differs in node features. Specifically, we initialize node features by sampling from a uniform distribution in the interval $[0, 1)$. The source node is initialized with a value of 1, while the target nodes are assigned a value of 0. The objective is to *transfer information from the source node to the target node*, without considering the intermediate nodes. Each graph topology introduces distinct structural properties:

- **Line:** A simple path graph of length n , where the source and target nodes are placed at opposite ends, resulting in a shortest path of length n .
- **Ring:** Cycles of size n , where the source and target nodes are placed at a distance of $\lfloor n/2 \rfloor$ from each other.
- **Crossed-Ring:** Cycles with additional "cross" edges between intermediate nodes. These edges do not reduce the shortest path between source and target, which remains $\lfloor n/2 \rfloor$.

Figure 1 illustrates each graph structure with a source–target distance of $n = 5$. In our experiments, we consider distances $n \in \{3, 5, 10, 50\}$, and use message passing neural networks (MPNNs) with depth equal to n . Unless stated otherwise, input dimension is set to 1 (regression tasks), hidden dimension to 64, and training is run for 2000 epochs. We apply node masking during training and testing to compute loss solely based on the source and the target node. We generate 1000 graphs for training, 100 for validation, and 100 for testing. Final performance is reported as the mean squared error over the test set. The architectural details and hyperparameters of our model are summarized in Table 3.

Table 3: Hyperparameter settings for graph transfer task.

Dataset	hid_dim	Activation	dropout	norm	lr	w_decay
Line	64	Tanh	0	BatchNorm	1e-3	0
Ring	64	Tanh	0	BatchNorm	1e-3	0
Crossed-Ring	64	Tanh	0	BatchNorm	1e-3	0

E.2 Hyperparameter settings

We conduct node classification experiments using a 2-layer GBN network, all the datasets are available as PyTorch Geometric datasets. All datasets are split into 50% training, 25% validation, and 25% testing. We report the final test accuracy as the average over 10 random data splits. The same experimental setup is applied across baseline methods. For experiments on heterogeneous graphs, we utilize a 5-layer model. The architectural details and hyperparameters are summarized in Table 4. Learning rates are set to 3e-5, with a dropout rate to 0.3, 0.2, 0.1 and a hidden dimension size of 512. LayerNorm is used to facilitate deeper models. All experiments are implemented using PyTorch Geometric library. The hardware is NVIDIA GeForce RTX 4090 GPU 24GB memory, and Intel Xeon Platinum 8352V CPU with 120GB RAM. Our code is publicly available at <https://anonymous.4open.science/r/GBN-E854>:

Table 4: Hyperparameter settings for node classification task.

Dataset	n_layers	hid_dim	Activation	dropout	norm	lr	w_decay
Texas	5	512	GELU	0.3	LayerNorm	1e-4	0
Wisconsin	5	512	GELU	0.7	LayerNorm	1e-3	0
Amazon-Ratings	5	512	GELU	0.15	BatchNorm	3e-4	0
Roman-Empire	5	512	GELU	0.15	LayerNorm	3e-4	0
Coauthor-CS	3	512	GELU	0.2	LayerNorm	3e-5	0
AmazonComputers	3	512	GELU	0.2	LayerNorm	3e-5	0
WikiCS	3	512	GELU	0.2	LayerNorm	3e-5	0

References

- [1] Chami, I., Ying, Z., Ré, C., and Leskovec, J. Hyperbolic graph convolutional neural networks. In *Advances in the 32nd NeurIPS*, pp. 4869–4880, 2019.
- [2] Choi, J., Hong, S., Park, N., and Cho, S.-B. GREED: Graph neural reaction-diffusion networks. In *Proceedings of the 40th ICML*, volume 202, pp. 5722–5747. PMLR, 2023.
- [3] Fu, D., Hua, Z., Xie, Y., Fang, J., Zhang, S., Sancak, K., Wu, H., Malevich, A., He, J., and Long, B. Vcr-graphormer: A mini-batch graph transformer via virtual connections. In *Proceedings of the 12th ICLR*. OpenReview.net, 2024.
- [4] Gravina, A., Eliasof, M., Gallicchio, C., Bacciu, D., and Schönlieb, C. On oversquashing in graph neural networks through the lens of dynamical systems. In *Proceedings of the 39th AAAI*, pp. 16906–16914. AAAI Press, 2025.
- [5] Gutteridge, B., Dong, X., Bronstein, M. M., and Giovanni, F. D. Drew: Dynamically rewired message passing with delay. In *Proceedings of the 40th ICML*, pp. 12252–12267. PMLR, 2023.
- [6] Huang, K., Wang, Y. G., Li, M., and Lio, P. How universal polynomial bases enhance spectral graph neural networks: Heterophily, over-smoothing, and over-squashing. In *Proceedings of the 41st ICML*, pp. 20310–20330. PMLR, 2024.
- [7] Jamadandi, A., Rubio-Madrigal, C., and Burkholz, R. Spectral graph pruning against over-squashing and over-smoothing. In *Advances in the 38th NeurIPS*, 2024.
- [8] Kipf, T. N. and Welling, M. Semi-supervised classification with graph convolutional networks. In *Proceedings of the 5th ICLR*. OpenReview.net, 2017.
- [9] Nguyen, K., Hieu, N. M., Nguyen, V. D., Ho, N., Osher, S. J., and Nguyen, T. M. Revisiting over-smoothing and over-squashing using ollivier-ricci curvature. In *Proceedings of the 40th ICML*, volume 202, pp. 25956–25979. PMLR, 2023.
- [10] Ortega, J. M. and Rheinboldt, W. C. *Iterative Solution of Nonlinear Equations in Several Variables*. Society for Industrial and Applied Mathematics, 2000. doi: 10.1137/1.9780898719468.

- 343 [11] Rusch, T. K., Chamberlain, B. P., Mahoney, M. W., Bronstein, M. M., and Mishra, S. Gradient
344 gating for deep multi-rate learning on graphs. In *Proceedings of the 11th ICLR*. OpenReview.net,
345 2023.
- 346 [12] Scholkemper, M., Wu, X., Jadbabaie, A., and Schaub, M. T. Residual connections and nor-
347 malization can provably prevent oversmoothing in gnns. In *Proceedings of the 13rd ICLR*.
348 OpenReview.net, 2025.
- 349 [13] Shirzad, H., Lin, H., Venkatachalam, B., Velingker, A., Woodruff, D. P., and Sutherland, D. J.
350 Even sparser graph transformers. In *Advances in the 38th NeurIPS*, 2024.
- 351 [14] Southern, J., Giovanni, F. D., Bronstein, M. M., and Lutzeyer, J. F. Understanding virtual nodes:
352 Oversmoothing and node heterogeneity. In *Proceedings of the 13rd ICLR*. OpenReview.net,
353 2025.
- 354 [15] Velickovic, P., Cucurull, G., Casanova, A., Romero, A., Liò, P., and Bengio, Y. Graph attention
355 networks. In *Proceedings of the 6th ICLR*. OpenReview.net, 2018.
- 356 [16] Wang, C., Tsepa, O., Ma, J., and Wang, B. Graph-mamba: Towards long-range graph sequence
357 modeling with selective state spaces. *CoRR*, abs/2402.00789, 2024.
- 358 [17] Xing, Y., Wang, X., Li, Y., Huang, H., and Shi, C. Less is more: on the over-globalizing problem
359 in graph transformers. In *Proceedings of the 41st ICML*, volume 235, pp. 54656–54672. PMLR,
360 2024.
- 361 [18] Xu, K., Hu, W., Leskovec, J., and Jegelka, S. How powerful are graph neural networks? In
362 *Proceedings of the 7th ICLR*. OpenReview.net, 2019.