Discrete Curvature Graph Information Bottleneck

Xingcheng Fu[1,2] [*]**, Jian Wang**[1,2]**, Yisen Gao**[3]**, Qingyun Sun**[4] [*]**, Haonan Yuan**[4]**,
Jianxin Li**[4]**, Xianxian Li**[1,2]

[1]Guangxi Key Lab of Multi-source Information Mining & Security, Guangxi Normal University, Guilin, China
[2]Key Lab of Education Blockchain and Intelligent Technology, Ministry of Education, Guangxi Normal University, China
[3]Institute of Artificial Intelligence, Beihang University, Beijing, China
[4]School of Computer Science and Engineering, Beihang University, Beijing, China
{fuxc, lixx}@gxnu.edu.cn, wj668811@stu.gxnu.edu.cn, {yisengao, sunqy, yuanhn, lijx}@buaa.edu.cn

Abstract

Graph neural networks(GNNs) have been demonstrated to depend on whether the node effective information is sufficiently passing. Discrete curvature (*Ricci curvature*) is used to study graph connectivity and information propagation efficiency with a geometric perspective, and has been raised in recent years to explore the efficient message-passing structure of GNNs. However, most empirical studies are based on directly observed graph structures or heuristic topological assumptions and lack in-depth exploration of underlying optimal information transport structures for downstream tasks. We suggest that graph curvature optimization is more in-depth and essential than directly rewiring or learning for graph structure with richer message-passing characterization and better information transport interpretability. From both graph geometry and information theory perspectives, we propose the novel Discrete **Curv**ature **G**raph **I**nformation **B**ottleneck (**CurvGIB**) framework to optimize the information transport structure and learn better node representations simultaneously. CurvGIB advances the *Variational Information Bottleneck* (*VIB*) principle for Ricci curvature optimization to learn the optimal information transport pattern for specific downstream tasks. The learned *Ricci curvature* is used to refine the optimal transport structure of the graph, and the node representation is fully and efficiently learned. Moreover, to reduce the computational complexity of Ricci curvature, we combine *Ricci flow* and *VIB* to deduce a curvature optimization approximation to form a tractable IB objective function. Extensive experiments on various datasets demonstrate the superior effectiveness and interpretability of CurvGIB.

1 Introduction

Through the unique mechanism of message passing, graph neural networks (GNNs) are capable of effectively capturing complex relationships and structural information between nodes. Extensive empirical studies have proved that whether the node information can be sufficient and effectively propagated is crucial, and the efficiency of message passing directly impacts the learning capability of GNNs. For different downstream tasks and applications, studying the optimal message passing structure in GNNs is of crucial necessity
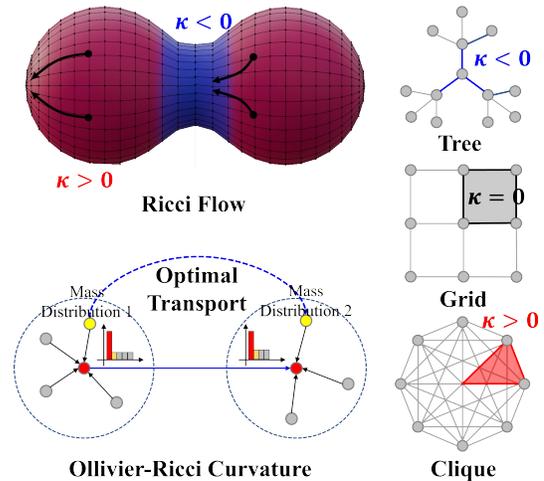
---

[*]Co-corresponding Author.

Figure 1: Ricci flow, Ollivier-Ricci curvature and different curvature-relevant topologies.

and importance for improving their overall performance and adaptability.

In practice, the properties of graphs are often determined by critical structures relevant to downstream tasks, rather than the entire graph structure (Sun et al. 2021). In response to this observation, two main research directions have emerged. Some of the existing work focuses on the exploration and mining of critical subgraph structures (Yu et al. 2020) relevant to tasks by leveraging some information-theoretic techniques, such as maximum mutual information (Wu et al. 2020) and information bottleneck (Tishby, Pereira, and Bialek 2000). On the other hand, other works focus on the messaging perspective, and these works are devoted to addressing messaging issues such as excessive smoothing and excessive compression, which can seriously impair GNNs' performance. Recently, the definition of graph curvature has provided a new perspective and insight for understanding and alleviating technical bottlenecks in the message-passing mechanism (Topping et al. 2021; Nguyen et al. 2022; Sun et al. 2024). However, the first existing works only focuses on finding task-relevant information and lacks a deep understanding of the message-

passing process in GNNs; the second existing works relies primarily on heuristic graph geometric priors based on the original topology to guide structure re-wiring, making it difficult to bridge and correlate with downstream tasks. There is still a lack of a unified perspective to comprehensively understand and mine the underlying task-relevant optimal transport structures in GNNs.

Rethinking the above issues, we suggest that the representation and generalization of GNNs fundamentally depend on whether the underlying transport structure of effective information is optimal in the broad and different downstream task scenarios. To extract task-relevant information more efficiently, the Information Bottleneck (IB) principle provides us with a theoretical framework, guiding us on how to maximize the retention of task-relevant information while compressing irrelevant information. As shown in Figure 1, graph curvature as a geometric definition of optimal transport, offers intuitive and effective metrics for the connectivity patterns of local topological structures. Consequently, a natural question arises: ***Can we combine the perspectives of information theory and graph geometry to develop a novel framework that can learn the optimal transport of underlying task-relevant effective information?*** This unified framework is expected to provide important guidance for our deep understanding of the underlying transmission structure of effective information in GNNs, and further broadening the application of GNNs in various practical scenarios.

In this paper, we present a unified discrete **Curv**ature **G**raph **I**nformation **B**ottleneck (**CurvGIB**) framework through both information theory and graph geometry perspectives to mining the underlying optimal transport structure of task-relevant graph information. The core idea is to leverage IB to learn the downstream task-relevant graph curvature rather than the topological structure directly, as the graph curvature better reflects the local connection patterns of the graph. Furthermore, we utilize graph curvature to distill the underlying optimal transport structure of task-relevant information, serving as a new structure for Graph Neural Networks (GNNs) training. Additionally, to address the issue of high computational complexity in discrete graph curvature differentiation, we improve an approximate optimization algorithm for Ricci curvature within our framework based on Ricci flow and employ a bi-level optimization technique to ensure efficient and stable learning of IB and Ricci curvature simultaneously. Our contributions are summarized as follows:

- CurvGIB advances both the Information Bottleneck principle and discrete graph curvature for graph learning, providing an elegant and universal framework in the perspectives of information theory and graph geometry.

- CurvGIB provide strong geometric intuition and interpretability, and it can mine the underlying optimal transport structure of GNNs to improve the model performance simultaneously, by using the IB principle to learn the optimal graph curvatures.

- Extensive experiment results in node classification demonstrate that the proposed CurvGIB has superior effectiveness compared to other baselines.

2 Related Work

2.1 Graph Structure Learning

Graph Structure Learning (GSL) parameterizes the adjacency matrix using probabilistic models, fully parameterized models, or metric learning models and jointly optimizes the adjacency matrix and GNN parameters for downstream tasks. For example, Yixin Liu et al.(Liu et al. 2022) use contrasting learning to generate optimal graph structures, maximizing the consistency between the learned topology and self-enhancing learning objectives. Zhen Zhang et al.(Zhang et al. 2021) propose a multi-view pooling operator, which enhances the robustness of graph structures and preserves low-level topological information. However, GSL methods often face high time complexity, limiting their scalability.

2.2 Graph Curvature

Discrete curvature methods can be categorized into two approaches: **re-weighting** and **re-wiring**. The **re-weighting** approach influences the information propagation path by adjusting the weights of nodes or edges. BORF(Nguyen et al. 2022) uses the Ollivier-Ricci curvature to reconstruct the neighborhoods of the paired nodes, but its approximation can amplify the impact of noisy edges. SDRF(Topping et al. 2021) employs balanced Forman curvature for neighborhood reconstruction, but it has weaker handling of nodes in graph boundary regions and involves higher computational costs. The **re-wiring** approach optimizes information transmission by directly modifying the graph structure. Some works (Ye et al. 2020; Fu et al. 2021, 2023) has innovatively incorporated graph curvature information into GNNs, significantly enhancing the performance of graphs with complex structures. Despite the potential shown by these methods, most existing research focuses on static graphs, with limited exploration of dynamic or large-scale graphs.

2.3 Information Bottleneck for Graph Learning

As a graph compression theory, information bottleneck has been widely concerned in recent years, and a lot of work attempts to use information bottlenecks to obtain the most sufficient representation for various graph-related downstream tasks. For example,(Wu et al. 2020) proposes GIB which achieves the most efficient node representation by maximizing mutual information between representations and labels while minimizing mutual information between representations and original inputs. With a variational IB principle, it can obtain sufficient information for node representation learning. (Yu et al. 2020) proposes SIB which directly reveals the vital substructure in the subgraph level. It leverages the mutual information estimator from the Deep Variational Information Bottleneck (*VIB*) (Alemi, Fischer, and Dillon 2017) for irregular graph data as the GIB objective. It tries to compress the original structure and help nodes aggregate the most efficient information coming from its neighborhood. (Sun et al. 2021) proposes VIB-GSL which can extract the underlying relations from the essence of representation learning, which is related to downstream tasks. It generates new structures by sampling original features with a specific probability distribution to mask original structures.
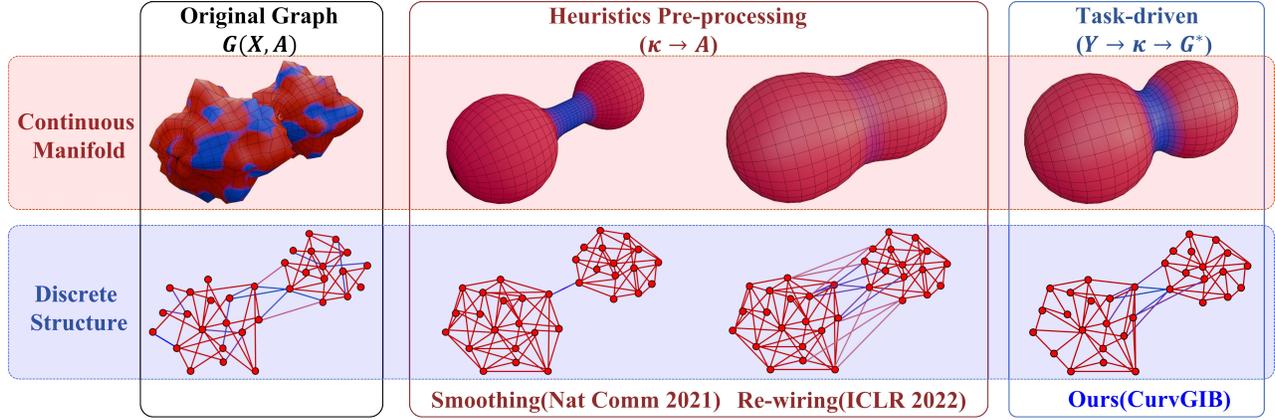
Figure 2: Instances of the geometric perspective, comparison different paradigms for graph structure refinement.

3 Prelimilary

3.1 Information Bottleneck

Following standard practice in the IB literature (Tishby, Pereira, and Bialek 2000), given data X, representation Z of X and target Y, (X, Y, Z) are following the Markov Chain $<Y \to X \to Z>$.

Definition 1 (Information Bottleneck)**.** *For the input data X and its label Y, the **Information Bottleneck** principle aims to learn the minimal sufficient representation Z:*

$$Z = \arg \min_{Z} -I(Z, Y) + \beta I(Z, X), \qquad (1)$$

where β is the Lagrangian multiplier trading off sufficiency and minimality.

Variational Information Bottleneck (VIB) (Alemi, Fischer, and Dillon 2017) proposed a variational approximation of the IB objective for deep neural networks:

$$\mathcal{L} = \frac{1}{N} \sum_{i=1}^{N} \int dZ p(Z|X_i) \log q(Y_i|Z) \\ + \beta \mathcal{D}_{KL} \left(p(Z|X_i), r(Z) \right), \qquad (2)$$

where $q(Y_i|Z)$ is the variational approximation to $p(Y_i|Z)$ and $r(Z)$ is the variational approximation of Z's prior distribution.

3.2 Discrete Graph Curvature

The overlap of two balls is inherently connected to the cost of transportation moving one ball to the other, yielding the Ollivier definition of Ricci curvature which is proposed by (Ollivier 2009).

Definition 2 (Olliver-Ricci Curvature)**.** *In a graph, given the mass distribution $m_u^{\alpha}(v)$, node $\boldsymbol{u} \in V$*

$$m_u^{\alpha}(v) = \begin{cases} \alpha, & v = u, \\ (1 - \alpha) \frac{1}{|N(u)|}, & v \in N(u), \\ 0, & otherwise. \end{cases} \qquad (3)$$

Where $|N(u)|$ is the set of nodes linking to u. The Ollivier's Ricci curvature of a node pair (i, j) is defined as follows:

$$Ric^{\alpha}(i, j) = 1 - \frac{W(m_i^{\alpha}, m_j^{\alpha})}{d(i, j)}, \qquad (4)$$

where $W(m_i^{\alpha}, m_j^{\alpha})$ is the Wasserstein distance between the probability (mass) distributions on nodes i and j, and $d(i, j)$ is the distance defined on the graph.

4 Discrete Curvature Graph Information Bottleneck

We propose a supervised optimization model of graph structure based on *Ricci flow* and *VIB*, named as CurvGIB. Specifically, we use *Ricci flow* to optimize the discrete Ollivier-Ricci curvature guide by *VIB*, learning a mapping function of curvature that measures the importance of all message edges in the neighborhood of the target node. The overall architecture of CurvGIB is shown in Fig. 3.

4.1 CurvGIB Principle Derivation

In this work, we focus on learning the critical information transport patterns during GNNs training, rather than directly reconstructing or learning the graph structures. Intuitively, we want to use the IB principle to find the "important" information in the training of GNNs, and then learn an optimal transport for the critical information, named *IB-Curvature*.

Given a graph $G = (X, A)$, the Ricci curvature $\mathcal{K}$ of each edge can be obtained by Eq. 4. However, according to the Eq. 4, the Ricci curvature is depended on the Wasserstein distance in each epoch, which leads to unacceptable computational complexity. Inspired by DeepRcci (Sun et al. 2023), we give a differentiable IB-guided curvature following their approximate definition. Specifically, the mass distribution is redefined using the Laplacian matrix $\boldsymbol{L}^{\alpha} = \alpha \boldsymbol{I} + (1 - \alpha) \boldsymbol{D}^{-1} \boldsymbol{A}$ of the $\mathbf{A}$. We rewrite the Eq. (3) as:

$$[\boldsymbol{L}^{\alpha}]_{ij} = \begin{cases} \alpha, & i = j \\ (1 - \alpha) \frac{1}{\boldsymbol{D}_{ii}}, & [\boldsymbol{A}]_{ij} = 1 \\ 0, & \text{Otherwise} \end{cases} \qquad (5)$$
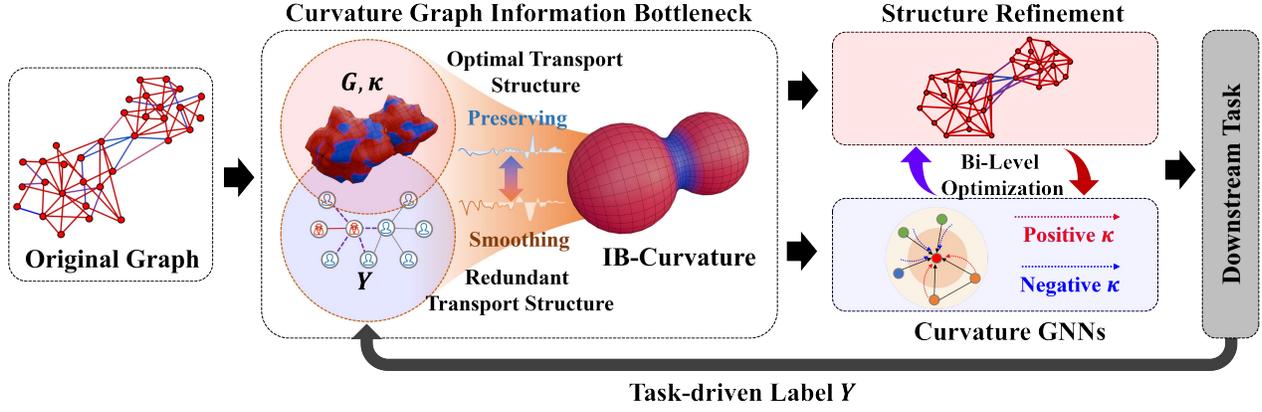
Figure 3: The proposed CurvGIB principle and its overall framework.

where α is the mass weight of each node, and the D is the diagonal degree matrix. Then, the differentiable IB-Curvature based on edge weight matrix $\mathcal{K}$ can be define as:

$$\kappa_{\mathrm{IB}}(i,j) = 1 - \frac{[\boldsymbol{L}^\alpha(\boldsymbol{A})f(\mathbf{Z})]_i - [\boldsymbol{L}^\alpha(\boldsymbol{A})f(\mathbf{Z})]_j}{d(\mathbf{z}_i, \mathbf{z}_j)}, l \quad (6)$$

with an affine transform $f : \mathbb{R}^{N \times H} \rightarrow \mathbb{R}^N$ and the $d(\cdot, \cdot)$ is the geodesic distance in latent space. According to Eq. (6), it can be observed that this approximate curvature definition is differentiable with respect to representation $\mathbf{Z}$.

Then we can give a formal definition of the *Discrete Curvature Graph Information Bottleneck* (CurvGIB) for graphs based on *IB-Curvature*.

Definition 3 (Curvature Graph Information Bottleneck). *For a graph $G = (X, A)$, label Y and its Ricci curvature κ of each edges, the optimal curvature κ_{IB} of node pair (i, j) found by Information Bottleneck is denoted as:*

$$\begin{aligned} \mathbf{Z}_{\mathrm{IB}}, \kappa_{\mathrm{IB}} &= \arg \min_{\mathbf{Z}, \kappa} \mathrm{CurvGIB}(\mathbf{X}, \mathbf{Y}, \mathbf{Z}, \kappa) \\ &\triangleq \arg \min_{\mathbf{Z}, \kappa}[-I(\mathbf{Z} \mid \kappa; \mathbf{Y}) + \beta I(\mathbf{Z} \mid \kappa; \mathbf{X})], \end{aligned} \quad (7)$$

Intuitively, the IB-curvature is updated with the hidden states learned by the IB-guided GNNs model and thus can be formulated as a constrained optimization form. For the IB constrain, the first term $-I(\mathbf{Z}|\mathcal{K}; Y)$ encourages that optimal transport information of label is preserved, and the second term $I(\mathbf{Z}|\mathcal{K}; \mathbf{X}))$ encourages that label-irrelevant information in message-passing is dropped. The learning procedure can be expressed through the Markov Chain: $\mathrm{MC}_{\mathrm{CurvGIB}} :< \mathbf{Y} \rightarrow \mathbf{Z}_\kappa \rightarrow \mathcal{K} >$. IB-Curvature represents the optimal transport pattern of the information most relevant to Y. As shown in Figure 2, negative IB-Curvature represents that this critical information tends to be transported on the "backbone" path of the graph, while positive IB-Curvature represents that critical information tends to propagate in the local highly connective structure.

Variational Bounds of CurvGIB. We respectively introduce the lower bound of $I(\mathbf{Z}|\kappa; \mathbf{Y})$ with respect to (Poole et al. 2019), and the upper bound of $I(\mathbf{Z}|\kappa; \mathbf{X})$, for Eq. (7).

Proposition 1 (**Lower bound of** $I(\mathbf{Z}|\kappa; \mathbf{Y})$). *For node representation $\mathbf{Z}$ with label $Y \in \mathbb{Y}$ and Ricci Curvature κ learned from A, we have*

$$I(\mathbf{Z}|\kappa; \mathbf{Y}) \geq \int p((\mathbf{Z}, \mathbf{Y})|\kappa) \log q(\mathbf{Y}|(\kappa, \mathbf{Z})) d\mathbf{Y} d\kappa d\mathbf{Z}, \quad (8)$$

where q is the variational posterior distribution.

Proposition 2 (**Upper bound of** $I(\mathbf{Z}|\kappa; \mathbf{X})$). *For node representation $\mathbf{Z}$ with label $Y \in \mathbb{Y}$ and Ricci Curvature κ learned from A, we have*

$$I(\mathbf{Z}|\kappa; \mathbf{X}) \leq \int p(\mathbf{X}) p(\mathbf{Y}|(\mathbf{X}, \kappa)) \log \frac{p(\mathbf{Y}|(\mathbf{X}, \kappa))}{r(\mathbf{Z}|\kappa)} d\kappa d\mathbf{X} d\mathbf{Z}, \quad (9)$$

where r is the variational posterior distribution.

Proofs for Proposition 1 and 2 are provided in Appendix Proof section.

4.2 Instantiating the CurvGIB Framework

For simultaneously optimizing the IB curvature and learning the IB representation, we propose a bi-level optimization framework based on Ricci flow and curvature-aware GNNs. The CurvGIB framework instantiating as shown in Figure 3.

IB-Curvature and Representation Learning. Discrete graph curvature measures how easily the label information flows through an edge and can be guided to control the strength of message-passing. We first update the Ricci curvature of the refined structure $\boldsymbol{A}^\star$ according to Eq. 3 for each edge,

$$\kappa_{\mathrm{IB}}(i,j) = 1 - \frac{[\boldsymbol{L}^\alpha(\boldsymbol{A}^\star)f(\mathbf{Z})]_i - [\boldsymbol{L}^\alpha(\boldsymbol{A}^\star)f(\mathbf{Z})]_j}{d(\mathbf{z}_i, \mathbf{z}_j)}. \quad (10)$$

Then, we advance (Ye et al. 2020) by using an affine transform $f : \mathbb{R}^{N \times M} \rightarrow \mathbb{R}^N$, and $f(\mathbf{X}) = \mathbf{W}\mathbf{X} + \mathbf{b}$. The loss and aggregation function of curvature-aware GNNs with curvature κ are formulated as follows:

$$\mathbf{Z}_\kappa = \mathrm{CurvGNN}(\mathbf{X}, \mathbf{Y}, \kappa_{\mathrm{IB}})$$

$$\mathrm{AGG}(\mathbf{Z}^{(l)}, \kappa_{\mathrm{IB}}) = \sigma\left(\mathbf{Z}^{(l)} + f\left(\mathcal{K}^{(l)}(\kappa_{\mathrm{IB}}) \cdot \mathbf{Z}_j^{(l)}\right)\right), \quad (11)$$

where $\mathcal{K}$ is the edge weight matrix obtained by IB-Curvature (we will introduce the definition later), and σ is the activate function.

Then, the IB principle guides GNNs to obtain critical information. The optimal transport information representation $\mathbf{Z}_\kappa$ with fixed curvature κ is:

$$\begin{aligned} \mathbf{Z}_\kappa &= \arg\min_{\mathbf{Z}_\kappa} \text{CurvGIB}\left(\mathbf{X}, \mathbf{Y}, \mathbf{Z}, \kappa\right), \\ &= \arg\min_{\mathbf{Z}_\kappa} \left(-I(\mathbf{Z}_\kappa; Y) + \beta I(\mathbf{Z}_\kappa; \mathbf{X})\right), \end{aligned} \quad (12)$$

where β is Lagrangian multiplier, where smaller β indicates more critical information was retained to $\mathbf{Z}_\kappa$.

Structure Refinement with IB-Curvature. In differential geometry, Ricci flow evolves a smooth manifold, allowing positive and negative curvatures to evolve in their respective curvature directions, ultimately partitioning the entire manifold into multiple weakly connected sub-manifolds. Existing work primarily employs reverse Ricci flow to broaden the topological "bottlenecks" of message passing, thereby alleviating issues such as over-squashing. In our paper, we aim to learn the optimal transport IB-curvature by using Ricci flow adaptively. Specifically, we follow existing work define a discrete Ricci flow as a edge-weighted graph with the edge weight matrix $\mathcal{K}$ for IB-Curvature as:

$$\mathcal{K}^{(l+1)} = \left(1 - \kappa_{\text{IB}}^{(l)}\right) d^{(l)}(\mathbf{Z}_\kappa, \mathbf{Z}_\kappa^T), \quad (13)$$

where l is the epoch number. According to Eq. 6, the updated Ricci curvature is: $1 - \frac{[L^\alpha(A)f(Z_\kappa)]_i - [L^\alpha(A)f(Z_\kappa)]_j}{d(z_i^\kappa, z_j^\kappa)}$

For each pair of nodes, we jointly optimize the edge sampling probability $\pi_{ij} = \text{sigmoid}(\mathcal{K}_{ij})$ by IB-Curvature weights and graph representation learning. We follow (Jang, Gu, and Poole 2016) to give the concrete relaxation of the Bernoulli distribution to update π:

$$\begin{aligned} [\boldsymbol{A}^\star]_{ij} &\sim \text{Ber}(\pi_{ij}) \\ &\approx \text{sigmoid}\left(\frac{1}{\tau}\left(\log\frac{\pi_{u,v}}{1 - \pi_{u,v}} + \log\frac{\epsilon}{1 - \epsilon}\right)\right), \end{aligned} \quad (14)$$

where $\epsilon \sim \text{Uniform}(0, 1)$ and $\tau \in \mathbb{R}^+$ is the temperature.

The refined structure A^* is optimized by maximizing the likelihood as

$$\boldsymbol{A}^\star = \arg\min_{\boldsymbol{A}^\star}\left[-\sum_{i=1}^{N}\sum_{j=1}^{N}\log p\left([\boldsymbol{A}]_{ij} = 1 \mid \boldsymbol{A}^\star\right)\right]. \quad (15)$$

Bi-level Optimization. According to Eq. (7) of the CurvGIB definition, the optimization of IB-Curvature κ_κ and representation $\mathbf{Z}_{\text{IB}}$ depend on each other, we introduce a bi-level optimization mechanism to ensure the effectiveness and stability of training. The overall optimization function is formalized as follows:

$$\begin{aligned} &\min \text{CurvGIB}\left(\mathbf{X}, \mathbf{Y}, \mathbf{Z}_\kappa, \kappa_{\text{IB}}\right), \\ &\text{s.t. } \kappa_{\text{IB}} \in \arg\min_{\kappa_{\text{IB}}} \text{IBCurv}(\mathbf{Z}_\kappa). \\ &\text{IBCurv}(\mathbf{Z}_\kappa) = \sum_{ij}(1 - \kappa_{\text{IB}}(i, j))\dot{d}(\mathbf{z}_i, \mathbf{z}_j) \end{aligned} \quad (16)$$

where IBCurv is obtained by combining Eq. (10), (13) and (6). The overall process of CurvGIB is shown in Algorithm 1.

Algorithm 1: The overall process of CurvGIB

Input: Graph $G = (X, A)$ with label Y; Number of training epochs $E, E1, E2$;
Output: weighted node representation Z, predicted label $\hat{Y}$

1 Parameter initialization;
2 Initialization κ with Eq.(10)
3 **for** $i = 0, 1, 2, \cdots, E$ **do**
    // IB-Curvature and Representation Learning
4   **for** $i = 0, 1, 2, \cdots, E1$ **do**
5   Aggregate node information with Eq.(11)
      Optimize the first phase with Eq.(12)
6   **end**
    // Structure Refinement with IB-Curvature
7   **for** $i = 0, 1, 2, \cdots, E2$ **do**
8   Calculate edge weight matrix $\mathcal{K}$ with Eq.(13)
      Sample structures with Bernoulli distribution by using $\mathcal{K}$ as a probability matrix with Eq.(14)
9   Optimize the second phase with Eq.(15)
10   **end**
11   Optimize the overall framework with Eq.(16)
12 **end**

Dataset	Nodes	Edges	Avg. Degree	Labels
Cora	2,708	5,429	3.90	7
Citeseer	3,312	4,732	2.79	6
PubMed	19,717	44,338	4.50	3
CS	18,333	81,894	8.93	40
Physics	34,493	247,962	14.38	5
Amazon-C	13,752	287,209	41.75	10
Amazon-P	13,381	245,778	36.71	10

Table 1: Statistics of datasets.

5 Experiments

We evaluate CurvGIB [1] on two tasks: node classification and graph denoising, to verify whether CurvGIB can retain critical structures conducive to message passing to improve the efficiency and robustness of graph representation learning. We then perform an extensive and comprehensive validation analysis to further demonstrate that CurvGIB can preserve critical graph structure while compressing unnecessary redundant information.

5.1 Datasets

As shown in Table 1, we conduct experiments on several real-world datasets: (1) Citation network: Cora and Citeseer (Kipf and Welling 2017) are citation networks of machine learning academic papers and PubMed is a citation network of biomedical academic papers. (2) Co-occurrence

---

[1]https://github.com/RingBDStack/CurvGIB

Method	Cora		CiteSeer		PUBMED		CS		PHYSICS		AMAZON-C		AMAZON-P	
	ACC	Macro-F1	ACC	Macro-F1	ACC	Macro-F1	ACC	Macro-F1	ACC	Macro-F1	ACC	Macro-F1	ACC	Macro-F1
GCN	81.5±0.5	81.9±1.0	70.9±0.5	<u>82.4±0.9</u>	79.0±0.3	<u>81.8±0.5</u>	91.1±0.5	82.5±0.6	92.8±1.0	81.8±0.5	82.6±2.4	82.4±0.6	91.2±1.2	81.4±0.4
GAT	83.0±0.7	76.8±17.8	72.5±0.7	77.4±17.9	79.0±0.3	72.6±23.4	90.5±0.6	75.8±18.9	92.5±0.9	74.5±21.1	78.0±19.0	75.9±21.4	85.1±20.3	76.1±19.7
GIN	86.6±0.4	77.6±8.9	63.4±0.6	77.9±8.5	86.6±0.1	76.8±7.7	53.3±0.4	78.5±8.2	86.6±0.1	75.7±7.7	54.3±19.0	77.7±8.9	51.4±17.0	77.3±7.8
GraphSAGE	83.4±3.2	<u>82.2±8.1</u>	73.4±4.9	**84.7±8.5**	<u>86.9±4.7</u>	**86.1±5.5**	92.4±5.7	84.0±8.2	<u>95.0±4.8</u>	83.7±7.0	80.4±13.7	**84.2±7.3**	89.7±7.7	<u>85.1±6.4</u>
SDRF	86.3±0.3	58.8±29.6	72.6±0.3	59.5±13.7	85.2±0.1	66.4±23.1	92.2±0.1	65.4±25.5	OOM	OOM	<u>86.6±0.2</u>	60.0±24.9	91.6±0.3	63.9±25.6
BORF	<u>87.5±0.2</u>	62.7±26.7	<u>73.8±0.2</u>	58.7±14.3	86.1±0.4	68.0±21.5	<u>93.1±0.1</u>	71.8±21.9	OOM	OOM	OOM	OOM	<u>93.6±0.2</u>	66.6±24.6
CurvGN-n	82.7±0.7	78.3±0.5	72.1±0.6	65.4±0.4	79.2±0.5	75.4±0.8	92.9±0.3	86.4±0.8	94.3±0.2	85.2±0.6	86.5±0.7	69.1±2.2	92.5±0.5	80.6±1.7
SUBLIME	84.2±0.5	71.7±0.5	73.5±0.6	66.4±1.4	81.0±0.6	73.7±1.6	92.3±0.7	<u>87.4±1.7</u>	94.2±0.3	<u>91.9±0.5</u>	72.3±0.5	42.0±4.0	81.3±0.3	66.6±2.2
MVPOOL-SL	83.6±0.4	78.1±7.6	72.6±0.5	61.1±3.2	80.0±0.4	60.1±11.2	92.0±0.5	84.9±6.2	94.0±0.6	<u>90.0±1.1</u>	79.5±1.5	68.5±16.9	88.5±1.2	79.0±16.3
CurvGIB	**89.0±0.9**	**84.3±0.9**	**75.5±0.1**	73.5±1.4	**88.0±0.1**	78.9±0.9	**94.2±0.5**	**91.2±0.5**	**96.5±0.3**	87.7±0.5	**91.9±1.4**	<u>83.9±1.7</u>	**93.7±1.4**	**94.8±4.1**

Table 2: Summary of node classification results: "average accuracy ± standard deviation". <u>Underlined</u>: best performance of specific backbones, **bold**: best results of each dataset.

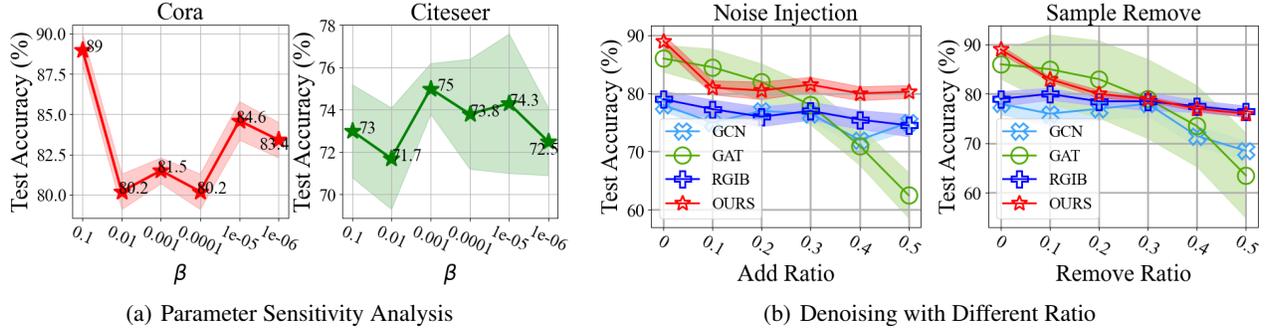(a) Parameter Sensitivity Analysis   (b) Denoising with Different Ratio

Figure 4: Parameter sensitivity and robustness analysis on Cora and Citeseer.

network (Shchur et al. 2019): Coauthor CS and Coauthor Physics are co-authorship graphs based on the Microsoft Academic Graph from the KDD Cup 2016 challenge. Amazon Computers and Amazon Photos (Shchur et al. 2019) which are segments of the Amazon co-purchase graph.

5.2 Experimental Setup

Baselines. We compare the CurvGIB with a number of node-level baselines which can be divided into 4 categories, including traditional graph neural networks (Kipf and Welling 2017; Xu et al. 2018; Hamilton, Ying, and Leskovec 2017; Velickovic et al. 2017), graph structure learning method (Liu et al. 2022; Zhang et al. 2021), information bottleneck method(Wang et al. 2022) and Graph curvature method(Ye et al. 2020; Topping et al. 2021; Nguyen et al. 2022), to demonstrate the effectiveness and robustness of CurvGIB .

Parameter Setting. We set both the information bottleneck size K and the embedding dimension of baseline methods as 64. For CurvGIB,we perform the depth search of $l \in \{2, 4, 6, 8\}$ for each dataset,and perform hyperparameter search of $\beta \in \{10^{-1}, 10^{-2}, 10^{-3}, 10^{-4}, 10^{-5}, 10^{-6}\}$ for each dataset.

5.3 Results And Analysis

Node Classification. We perform 10-fold cross-validation and report the average accuracy, average F1 score, and the standard deviation across the 10 folds in Table 2. As

shown in Table 2, the proposed CurvGIB consistently outperforms all baselines on all datasets. Especially in the dataset Amazon-Computers, our method outperforms all baselines by a large margin. Generally, the graph structure learning method(SUBLIME, $\text{MVPOOL}_{\text{SL}}$) only shows small improvement or even have a negative impact on some datasets (Amazon-Photos and Amazon-Computers), which is because they are constrained by the generated graph structures which miss the critical structure information from the original graph. In addition, the curvature-based methods (BORF, SDRF, CurvGN, CurvGIB) all perform well, and in particular our method gets a significant improvement.

Parameter Sensitivity. We explore the influence of the Lagrangian multiplier β trading off prediction and compression in Eq.(1, 7, 8, 9). Figure 4 (a) shows the changing trend of the node classification accuracy on Cora and Citeseer. Based on the results, we make the following observations: (1) Remarkably, the node classification accuracies of CurvGIB variation across different β show the same changing behavior on both datasets. Citeseer has high $\beta$is due to its difficulty in learning enough subgraph information from many different concatenations of small graphs, and Cora requires a higher β to compress more information. (2) Appropriate value of β can greatly increase the model's performance. CurvGIB achieves the best balance of prediction and compression with $\beta = 10^{-1}$ and $\beta = 10^{-3}$ on Cora and Citeseer, respectively.
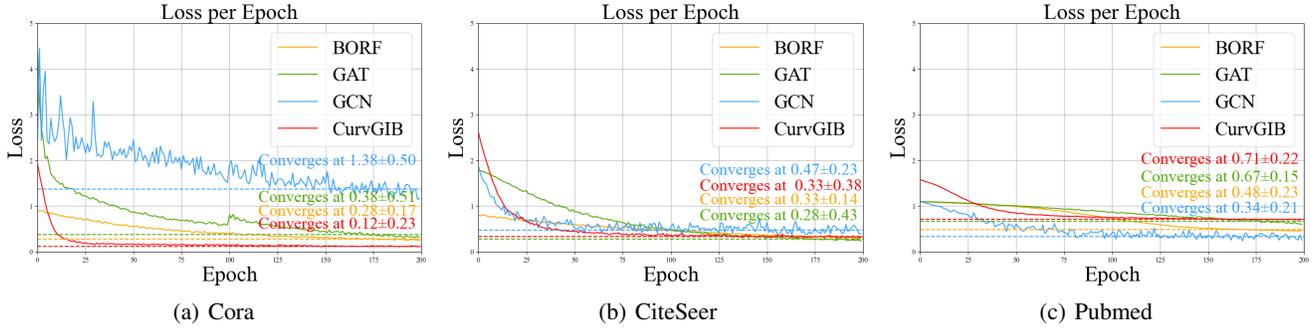
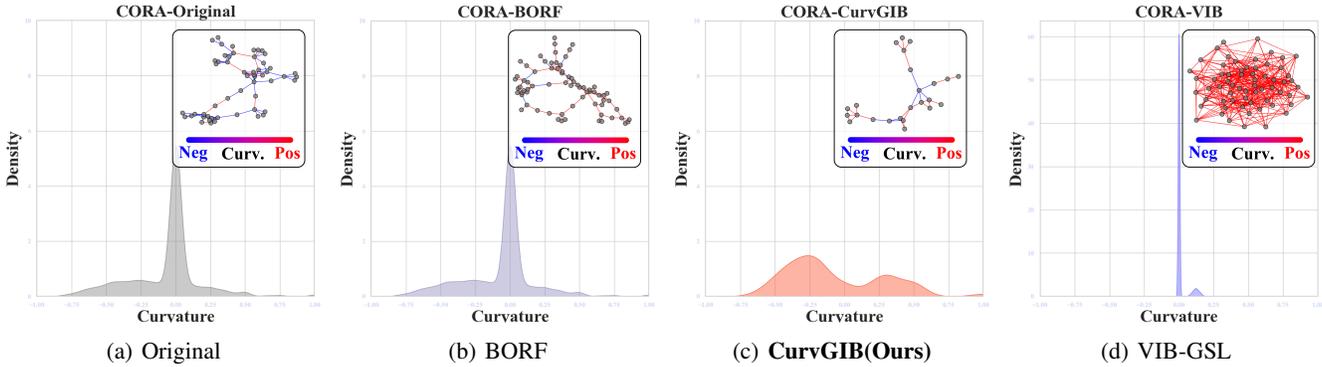Figure 5: Learning Stability Analysis on Different Datasets

Figure 6: Distributions of Ricci curvature and visualizations of sampling sub-structures after learning on Cora.

Graph Denoising. To evaluate the robustness of our framework by adding or deleting edges on Cora. Specifically, for each graph in the dataset, we randomly remove (if edges exist) or add (if no such edges) $10\%, 20\%, 30\%, 40\%, 50\%$ edges. As shown in Figure 4 (b), the classification accuracy of GCN dropped by 10% with 50% missing edges and dropped by 5% with 50% noisy edges. CurvGIB achieves better results with small performance degradation, and with the increase of the ratio, the accuracy of CurvGIB remains stable gradually.

Learned Graph Curvature. We tested different models on the Cora dataset and visualized the Ricci curvature of edges before and after processing. As shown in Figure 6, **BORF** tends to remove edges with negative curvature as much as possible while ensuring that the curvature distribution remains unchanged. **VIB-GSL** (Sun et al. 2021) tends to process the original graph into a fully connected graph. Our model can adjusts to the structure that is most conducive to message passing based on the specific dataset.

Training stability. CurvGIB deduces a tractable variational approximation for the **IB** objective, which facilitates the training stability. In this subsection, we analyze the convergence of CurvGIB and other baselines on Cora, Citeseer and Pubmed with a learning rate of 0.001. As shown in Figure 5, except **GCN**, other models all coverage steadily. However, CurvGIB converges 2.5 times as fast as model **GAT**

and converges 2 times as fast as model **BORF**. It benefits from our CurvGIB principle, which improve the model to retain critical structures for sufficient information transport.

6   Conclusion

In this study, we introduced the Discrete **Curv**ature **G**raph **I**nformation **B**ottleneck (**CurvGIB**) framework, which demonstrates significant advantages in optimizing the information transport structure of Graph Neural Networks (GNNs). By incorporating Ricci curvature, CurvGIB enhances information propagation efficiency in complex graph structures from a geometric perspective, providing a more precise and effective approach compared to direct rewiring or learning graph structures. The main contributions of CurvGIB include optimizing discrete curvature to improve graph structures, extending the Information Bottleneck (IB) principle to learn optimal information transport patterns for specific downstream tasks, reducing computational complexity through a tractable curvature optimization approximation, and demonstrating superior performance and interpretability through extensive empirical validation. However, the current research focuses primarily on node-level tasks and has not yet been extended to graph-level tasks. Future work should explore how CurvGIB can be adapted for graph-level tasks and improve its applicability and effectiveness in different scenarios and applications.

References

Alemi, A. A.; Fischer, I.; and Dillon, J. V. 2017. Deep Variational Information Bottleneck. *ArXiv*, abs/1612.00410.

Fu, X.; Li, J.; Wu, J.; Sun, Q.; Ji, C.; Wang, S.; Tan, J.; Peng, H.; and Philip, S. Y. 2021. ACE-HGNN: Adaptive curvature exploration hyperbolic graph neural network. In *2021 IEEE international conference on data mining (ICDM)*, 111–120. IEEE.

Fu, X.; Wei, Y.; Sun, Q.; Yuan, H.; Wu, J.; Peng, H.; and Li, J. 2023. Hyperbolic geometric graph representation learning for hierarchy-imbalance node classification. In *Proceedings of the ACM Web Conference 2023*, 460–468.

Hamilton, W. L.; Ying, Z.; and Leskovec, J. 2017. Inductive Representation Learning on Large Graphs. In *Neural Information Processing Systems*.

Jang, E.; Gu, S. S.; and Poole, B. 2016. Categorical Reparameterization with Gumbel-Softmax. *ArXiv*, abs/1611.01144.

Kipf, T. N.; and Welling, M. 2017. Semi-Supervised Classification with Graph Convolutional Networks. arXiv:1609.02907.

Liu, Y.; Zheng, Y.; Zhang, D.; Chen, H.; Peng, H.; and Pan, S. 2022. Towards Unsupervised Deep Graph Structure Learning. *Proceedings of the ACM Web Conference 2022*.

Nguyen, K. D.; Nguyen, T. M.; Ho, N.; Nguyen, K. N.; Nong, H.; and Nguyen, V. P. 2022. Revisiting Oversmoothing and Over-squashing using Ollivier's Ricci Curvature. In *International Conference on Machine Learning*.

Ollivier, Y. 2009. Ricci curvature of Markov chains on metric spaces. *Journal of Functional Analysis*, 256(3): 810–864.

Poole, B.; Ozair, S.; Van Den Oord, A.; Alemi, A.; and Tucker, G. 2019. On variational bounds of mutual information. In *International Conference on Machine Learning*, 5171–5180. PMLR.

Shchur, O.; Mumme, M.; Bojchevski, A.; and Günnemann, S. 2019. Pitfalls of Graph Neural Network Evaluation. arXiv:1811.05868.

Sun, L.; Hu, J.; Zhou, S.; Huang, Z.; Ye, J.; Peng, H.; Yu, Z.; and Yu, P. 2024. Riccinet: Deep clustering via a riemannian generative model. In *Proceedings of the ACM on Web Conference 2024*, 4071–4082.

Sun, L.; Huang, Z.; Wu, H.; Ye, J.; Peng, H.; Yu, Z.; and Philip, S. Y. 2023. DeepRicci: Self-supervised Graph Structure-Feature Co-Refinement for Alleviating Oversquashing. In *ICDM*, 558–567. IEEE.

Sun, Q.; Li, J.; Peng, H.; Wu, J.; Fu, X.; Ji, C.; and Yu, P. S. 2021. Graph Structure Learning with Variational Information Bottleneck. In *AAAI*, volume abs/2112.08903.

Tishby, N.; Pereira, F. C.; and Bialek, W. 2000. The information bottleneck method. *arXiv preprint physics/0004057*.

Topping, J.; Di Giovanni, F.; Chamberlain, B. P.; Dong, X.; and Bronstein, M. M. 2021. Understanding over-squashing and bottlenecks on graphs via curvature. In *ICLR*.

Velickovic, P.; Cucurull, G.; Casanova, A.; Romero, A.; Lio', P.; and Bengio, Y. 2017. Graph Attention Networks. *ArXiv*, abs/1710.10903.

Wang, J.; Luo, M.; Li, J.; Liu, Z.; Zhou, J.; and Zheng, Q. 2022. Toward Enhanced Robustness in Unsupervised Graph Representation Learning: A Graph Information Bottleneck Perspective. *IEEE Transactions on Knowledge and Data Engineering*, 36: 4290–4303.

Wu, T.; Ren, H.; Li, P.; and Leskovec, J. 2020. Graph Information Bottleneck. *ArXiv*, abs/2010.12811.

Xu, K.; Hu, W.; Leskovec, J.; and Jegelka, S. 2018. How Powerful are Graph Neural Networks? *ArXiv*, abs/1810.00826.

Ye, Z.; Liu, K. S.; Ma, T.; Gao, J.; and Chen, C. 2020. Curvature Graph Network. In *International Conference on Learning Representations*.

Yu, J.; Xu, T.; Rong, Y.; Bian, Y.; Huang, J.; and He, R. 2020. Graph Information Bottleneck for Subgraph Recognition. *ArXiv*, abs/2010.05563.

Zhang, Z.; Bu, J.; Ester, M.; Zhang, J.; Li, Z.; Yao, C.; Huifen, D.; Yu, Z.; and Wang, C. 2021. Hierarchical Multi-View Graph Pooling With Structure Learning. *IEEE Transactions on Knowledge and Data Engineering*, 35: 545–559.