Chungnam
National
University

1st CNU-UNIST Workshop:

Hyperbolic Heterogeneous Graph Representation Learning

Jongmin Park

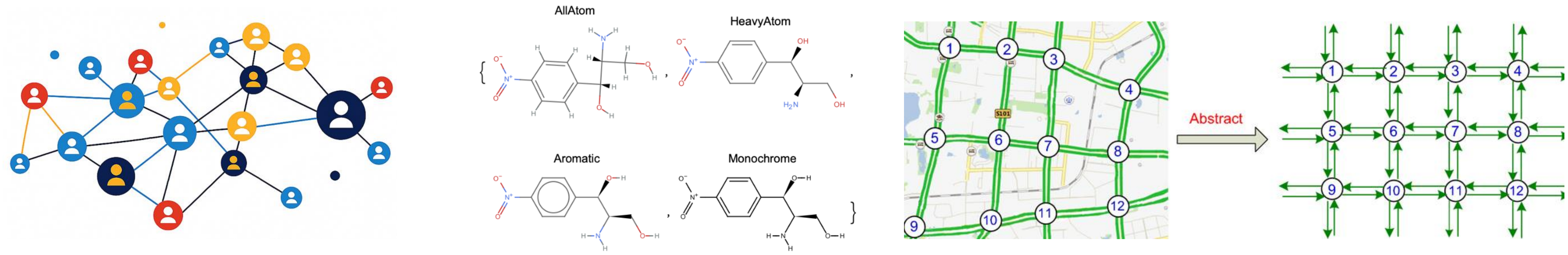
Department of Computer Science and Engineering
Chungnam National University

2026.02.12

Introduction

Introduction

- **Graphs**
  - ✓ A graph consists of **nodes** (entities) and **edges** (relationships).
  - ✓ Represents relationship between entities effectively.
  - ✓ Can be utilized in **various domain** for **modeling relationships between entities**.



(a) Social network **(b) Molecular network** **(c) Traffic network**

Fig1. Various domain of graph modeling.

Introduction

- **Heterogeneous graphs**
  - ✓ A heterogeneous graph consists of **various types of nodes and links.**
  - ✓ Can effectively **represent heterogeneous information** observed in the real world.

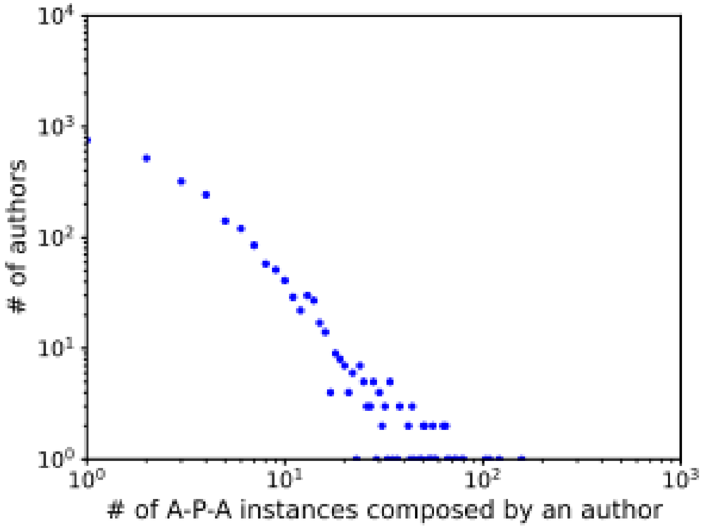
- **Metapath instances within a heterogeneous graph**
  - ✓ **Metapath** : Sequence of nodes/link types.
  - ✓ **Metapath instance** : Node sequence within a given metapath.

A-P-A Metapath : Co-authorship information.
A-P-A Metapath instance : Paper 1 is written by Author 1 and Author 2.

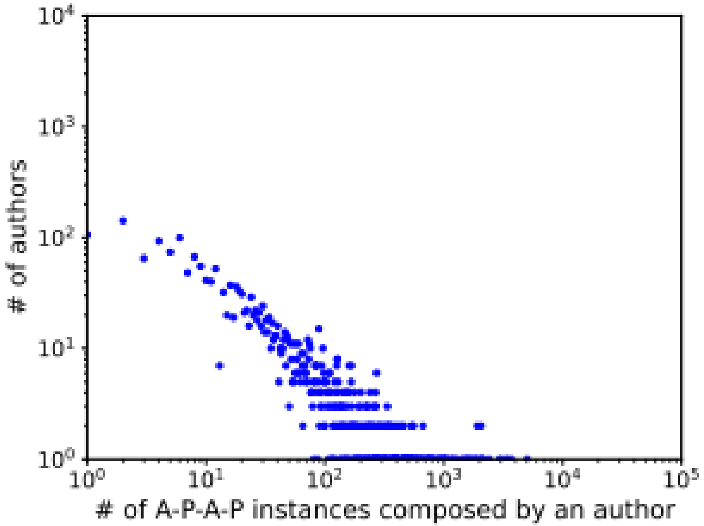
Fig 2. Academic network with two types of nodes.

Introduction

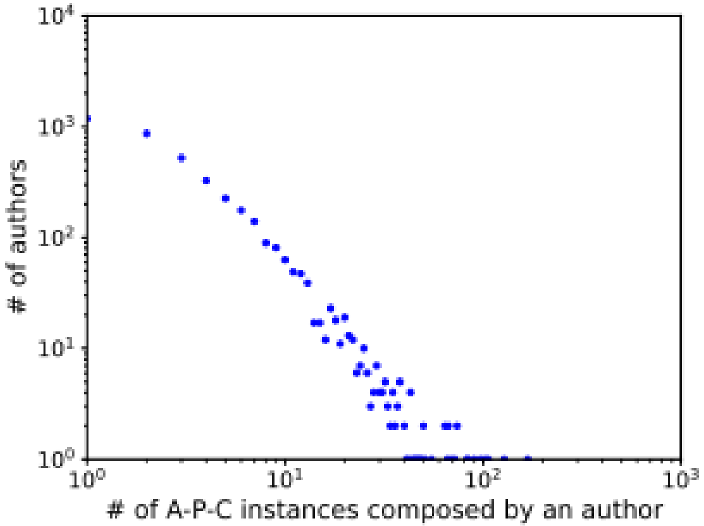
- **Hierarchical structures of a heterogeneous graph**
  - ✓ From metapath instances, we can observe **multiple hierarchical structures**.
  - ✓ However, it is difficult to preserve such structures in Euclidean space.



(a) A-P-A in DBLP



(b) A-P-A-P in DBLP



(c) A-P-C in DBLP

Fig 3. Metapath instance distributions of some metapaths in DBLP dataset.

Introduction

- **Hyperbolic space**
  - ✓ Hyperbolic space has a **negative curvature** and **expands exponentially**.
  - ✓ Hyperbolic space is **effective to represent hierarchical structures**.

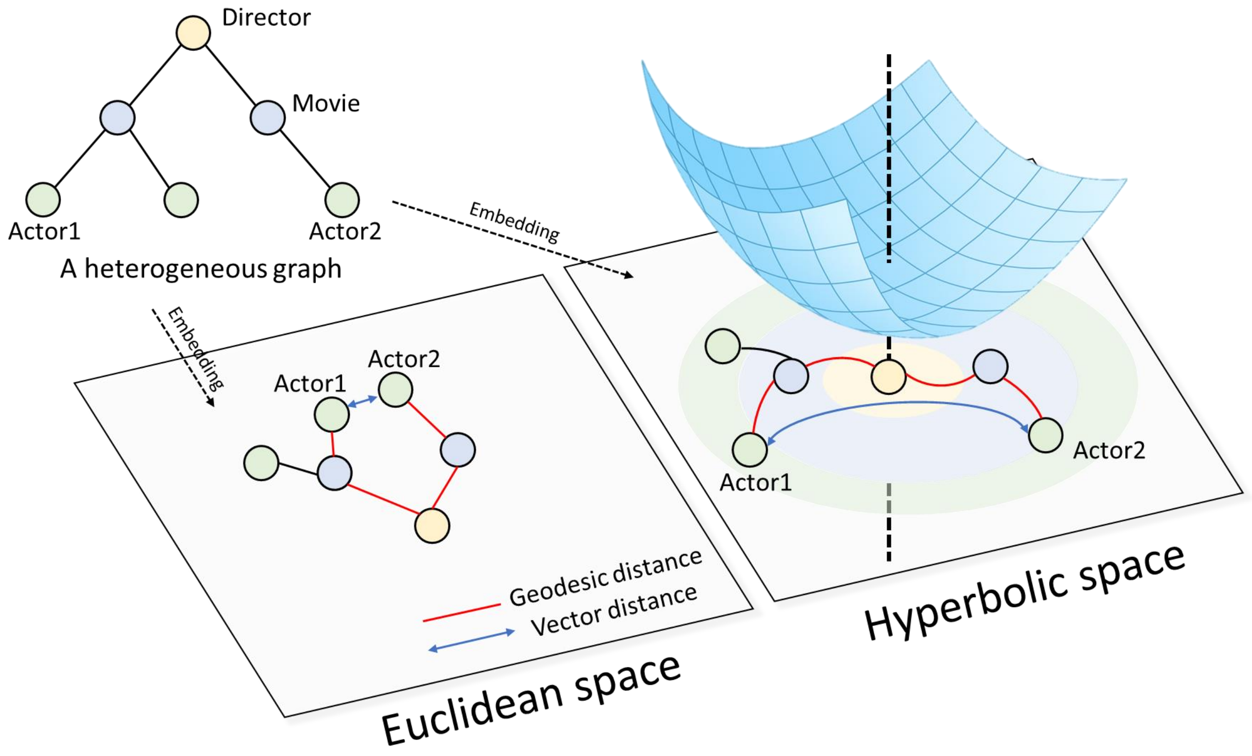


Fig 4. Comparison between Euclidean space and Hyperbolic space.

Part 1

Hyperbolic Heterogeneous Graph Neural Networks

Background

- **Hyperbolic Heterogeneous Graph Embedding**
  - ✓ Euclidean operations are reformulated under hyperbolic space to enable representation learning in curved space.

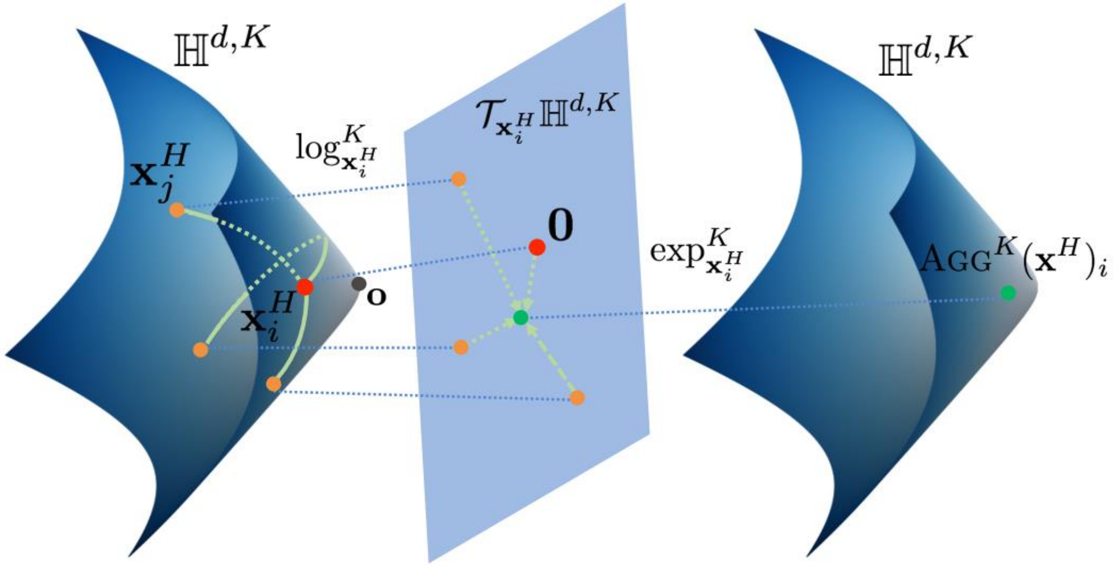
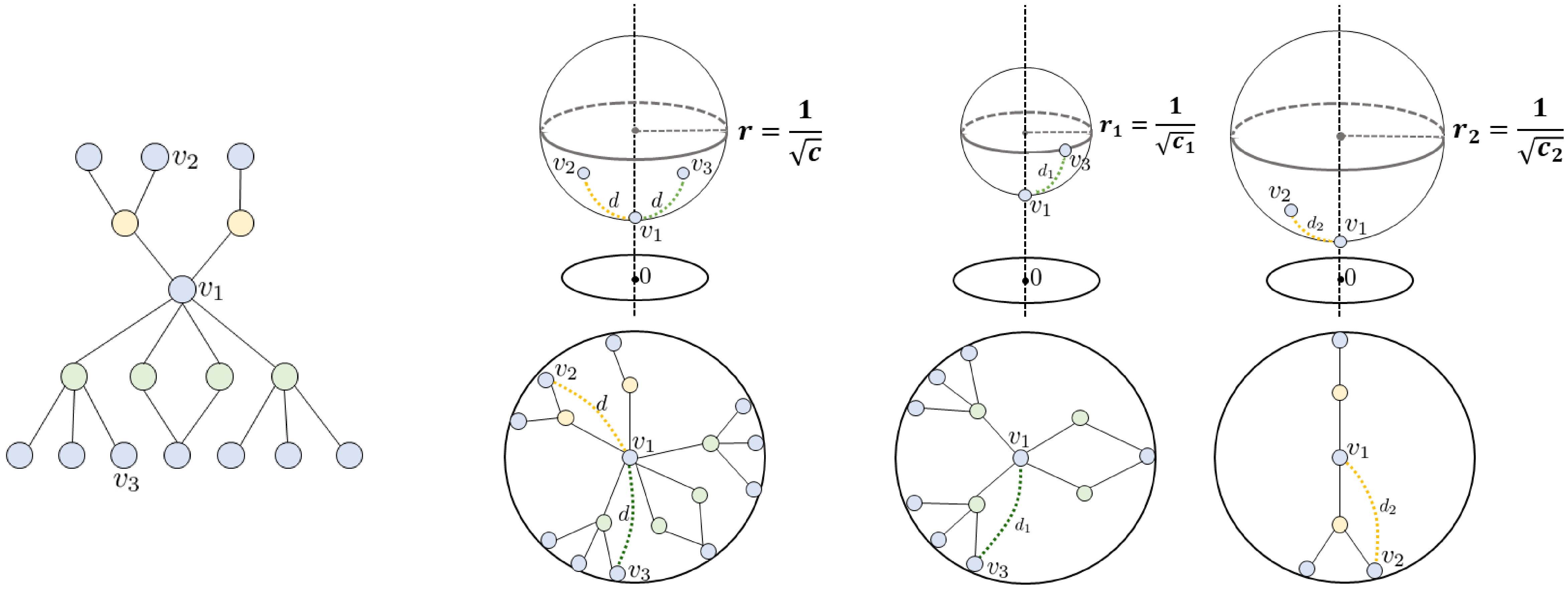


Fig 5. An example of graph convolution operation in the hyperbolic space.

Background

- **Comparison between Single and Multiple Hyperbolic Spaces**



(a) A heterogeneous graph **(b) Single Hyperbolic space** **(c) Multiple Hyperbolic spaces**

Fig 6. Comparison between single and multiple hyperbolic spaces.

Methodologies

- **Multi-Hyperbolic Space-based Heterogeneous Graph Attention Network (MSGAT)**
  - ✓ To effectively capture multiple hierarchical structures, we propose MSGAT.
  - ✓ Graph attention mechanism in **multiple metapath-specific Hyperbolic spaces**.
  - ✓ Each metapath-specific Hyperbolic space has a negative curvature that **reflects its underlying distribution**.

Methodologies

- **Overall Framework of MSGAT**

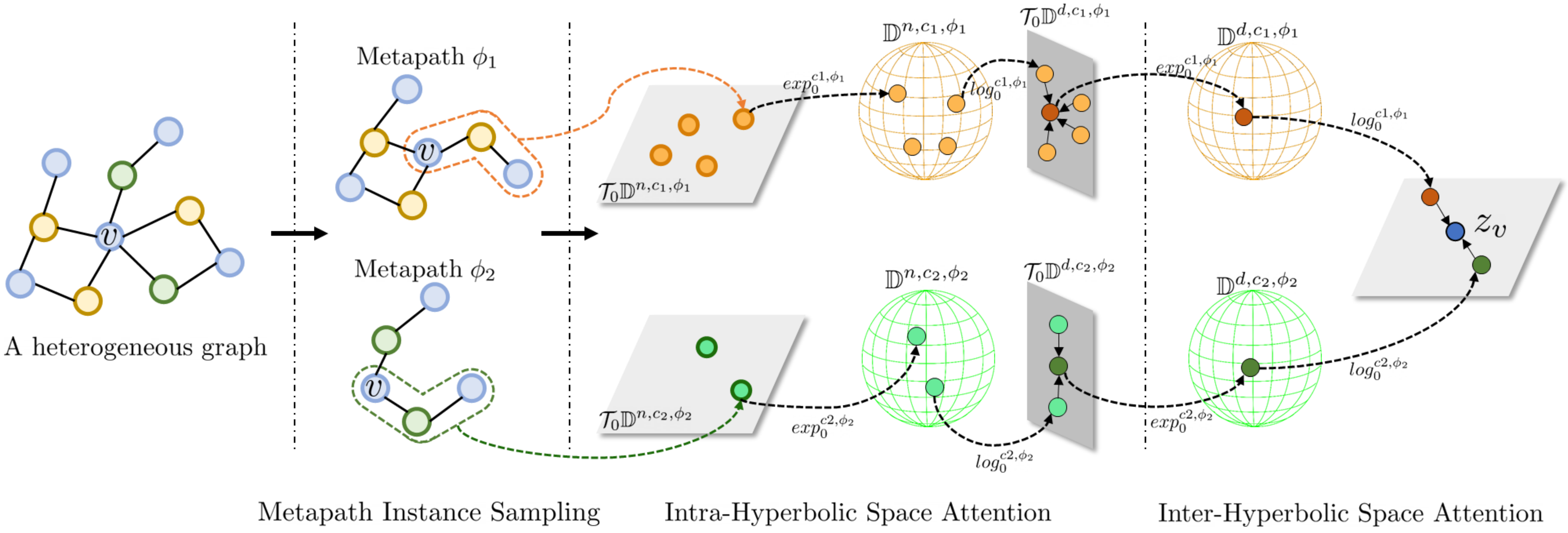


Fig 7. An overview of MSGAT.

Experiments

- **Datasets**
  - ✓ Four real-world heterogeneous graph datasets were used.

IMDb dblp computer science bibliography acm Association for Computing Machinery last.fm

Datasets	IMDB	DBLP	ACM	LastFM
Domain	TV program database	Citation Network	Citation Network	Music database
# Nodes	12,772	18,405	8,989	20,612
# Links	18,644	33,973	12,961	107,242
# Node types	3	3	3	3
# Link types	2	2	2	2
Tasks	Node classification & clustering			Link prediction

Table 1. Statistics of datasets.

Experiments

- **Baselines**
  - ✓ The baselines for the four categories.

Categories	Models
ⅰ. **Euclidean homogeneous** GNNs	GCN, GAT
ⅱ. **Hyperbolic homogeneous** GNNs	HGCN
ⅲ. **Euclidean heterogeneous** GNNs	HAN, MAGNN, GTN, HGT, GraphMSE, Simple-HGN
ⅳ. **Hyperbolic heterogeneous** GNNs	McH-HGCN, SHAN, HHGAT

Table 2. The baselines.

Experiments

- ## Task 1 : Node Classification

  - ✓ In most cases, proposed models outperform existing baselines.
  - ✓ A comparison of category iii and **iv** shows that leveraging **hyperbolic space to learn hierarchical structures is effective** in heterogeneous graph representation learning.

Datasets	IMDB				DBLP				ACM			
Metric	Macro-F1				Macro-F1				Macro-F1			
Train %	20%	40%	60%	80%	20%	40%	60%	80%	20%	40%	60%	80%
GCN	52.17	53.20	54.35	54.19	87.51	88.55	89.44	89.45	83.08	87.34	88.80	88.43
HGCN	54.38	57.05	57.86	57.92	91.69	91.93	92.60	92.58	87.29	89.19	90.01	90.03
HAN	56.19	56.84	58.95	58.61	92.63	92.35	92.86	92.73	87.88	90.54	91.22	91.35
HGT	56.14	57.12	61.52	63.69	90.36	91.57	92.32	93.46	89.12	89.15	90.57	93.45
Simple-HGN	59.97	61.94	66.73	67.56	93.48	93.98	94.01	94.25	92.25	92.64	93.06	93.55
SHAN	62.23	63.98	66.68	68.49	94.27	94.33	94.50	94.67	92.56	92.88	94.10	**94.94**
HHGAT	63.16	65.07	65.72	67.42	94.19	94.27	94.90	94.77	91.34	92.92	94.28	93.91
MSGAT	**65.75**	**68.07**	**71.42**	**70.03**	**95.44**	**95.54**	**95.67**	**95.29**	**92.73**	**93.95**	**94.83**	94.01

Table 3. Experimental results for the node classification task (Macro-F1).

Experiments

- **Task 1 : Node Classification (Cont.)**
  - ✓ In most cases, proposed models outperform existing baselines.
  - ✓ A comparison of category **iii** and **iv** shows that leveraging **hyperbolic space to learn hierarchical structures is effective** in heterogeneous graph representation learning.

Datasets	IMDB				DBLP				ACM			
Metric	Micro-F1				Micro-F1				Micro-F1			
Train %	20%	40%	60%	80%	20%	40%	60%	80%	20%	40%	60%	80%
GCN	52.13	53.34	54.61	54.37	88.21	88.68	90.01	90.14	87.75	87.86	88.40	88.56
HGCN	54.46	57.02	58.01	58.54	92.06	92.31	93.16	93.21	88.09	90.06	90.51	91.10
HAN	56.71	56.68	58.26	59.35	92.35	92.87	93.42	93.54	91.20	91.78	92.39	92.03
HGT	57.97	58.80	62.63	67.01	91.46	92.05	92.72	92.57	89.59	90.70	91.18	91.77
Simple-HGN	63.76	65.60	69.29	69.35	94.17	93.87	94.71	94.68	91.91	92.86	93.33	93.53
SHAN	64.31	<u>66.56</u>	69.57	69.42	94.53	94.60	94.92	<u>95.36</u>	<u>92.38</u>	93.37	<u>94.46</u>	**94.56**
HHGAT	<u>65.76</u>	66.34	<u>70.40</u>	<u>69.61</u>	<u>94.66</u>	<u>94.72</u>	<u>95.15</u>	95.34	92.36	<u>93.46</u>	94.34	93.72
MSGAT	**69.09**	**70.95**	**73.60**	**73.37**	**95.79**	**95.90**	**95.98**	**95.85**	**92.96**	**93.91**	**94.88**	<u>94.05</u>

Table 4. Experimental results for the node classification task (Micro-F1).

Experiments

- ## Task 2 : Node Clustering

Datasets	IMDB		DBLP		ACM	
Metric	**NMI**	**ARI**	**NMI**	**ARI**	**NMI**	**ARI**
GCN	7.84	8.12	75.37	77.14	51.73	53.42
HGCN	10.29	11.10	76.48	79.36	60.19	62.06
HAN	11.21	11.49	77.03	82.53	61.24	64.11
HGT	14.55	16.59	79.02	80.28	67.88	72.56
Simple-HGN	17.58	19.51	82.38	85.71	69.91	72.07
SHAN	20.60	22.56	82.39	86.13	72.90	77.73
HHGAT	20.75	22.80	83.14	85.91	72.49	77.92
MSGAT	**24.06**	**26.33**	**84.38**	**88.27**	**73.33**	**78.28**

Table 5. Experimental results for the node clustering task.

- ## Task 3 : Link Prediction

Dataset	Metric	GCN	HGCN	HAN	HGT	Simple-HGN	HHGAT	MSGAT
LastFM	**ROC-AUC**	43.68	46.71	48.35	47.78	53.85	54.37	**55.77**
	F1-Score	56.15	57.23	57.11	61.16	63.02	62.85	**63.39**

Table 6. Experimental results for the link prediction task.

Experiments

- **Ablation Study**
  - ✓ The results of ablation study demonstrate the effectiveness of using multiple hyperbolic spaces.

Datasets	IMDB				DBLP				ACM			
Metric	Macro-F1	Micro-F1	NMI	ARI	Macro-F1	Micro-F1	NMI	ARI	Macro-F1	Micro-F1	NMI	ARI
MSGAT	**71.42**	**73.60**	**24.06**	**26.33**	**95.67**	**95.98**	**84.38**	**88.27**	**94.83**	**94.88**	**73.33**	**78.28**
CONCAT	68.03	70.75	22.24	24.91	94.23	94.54	82.06	84.65	93.33	93.91	72.73	75.82
EUCLID	64.72	67.17	16.72	14.65	93.09	93.51	79.32	83.84	90.02	90.09	70.24	93.18
SINGLE	66.06	68.77	21.65	25.91	93.76	93.94	80.95	86.06	92.39	92.67	73.16	77.93

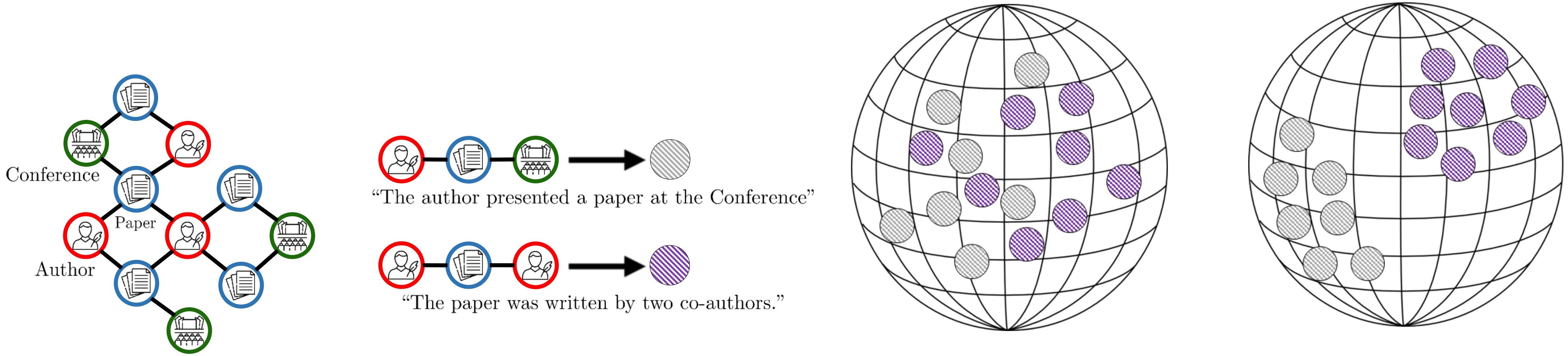
Table 7. Experimental results for the ablation study.

Part 2

Metapath-based Hyperbolic Contrastive Learning

Background

- **Indistinguishability between Metapath Embeddings**
  - ✓ Since **different metapaths embeddings** are represented **in similar positions**, the advantages of hyperbolic space **cannot be effectively utilized**.



(a) An academic network. **(b) Metapaths.** **(c) Worst case of metapath embeddings in hyperbolic space.** **(d) Optimal case of metapath embeddings in hyperbolic space.**

Fig 8. An example of indistinguishability between metapath embeddings limits the effectiveness of hyperbolic space.

Background

- **Contrastive Learning**
  - ✓ Contrastive learning can address the aforementioned limitations.
  - ✓ Objective : **maximize** the similarity between **positive sample pairs** and **minimize** the similarity between **negative pairs.**
  - ✓ Sample selection strategy : For given anchor (or query), select **similar ones** as **positive samples** and **dissimilar ones** as **negative samples**.

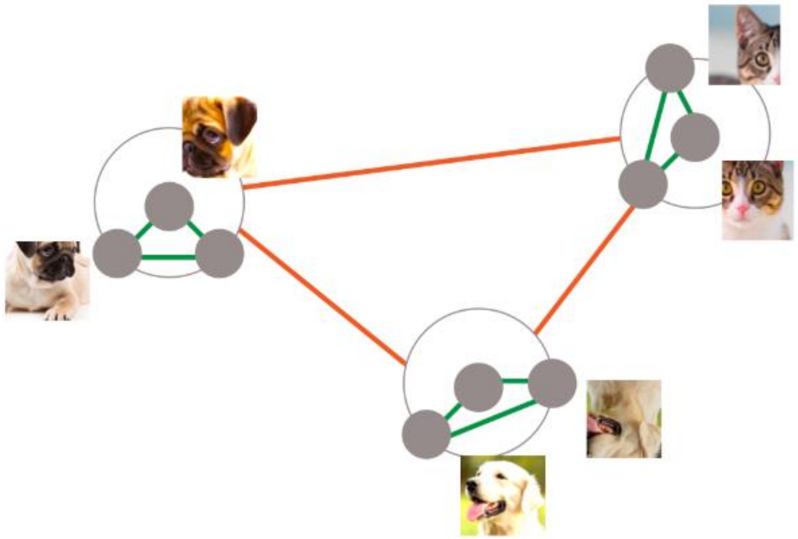


Fig 9. An example of contrastive learning.

Methodology

- **Metapath-based Hyperbolic Contrastive Learning (MHCL)**
  - ✓ **Positive** samples : **same** metapath embedding.
  - ✓ **Negative** samples : **different** metapath embeddings.
  - ✓ Objective : **Minimize** the distance between **same metapaths** and **maximize** the similarity between **different metapaths.**

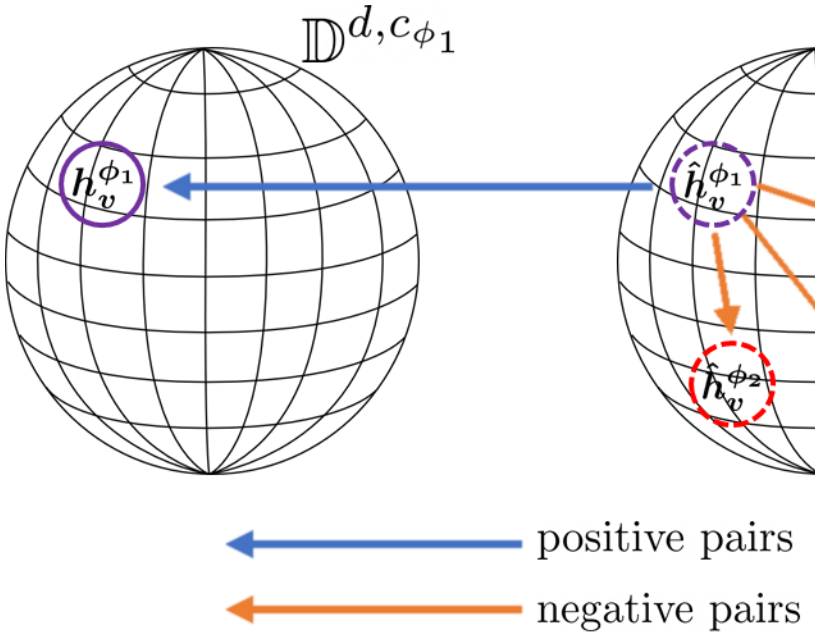


Fig 10. A basic concept of MHCL.

Methodology

- **Overall framework of MHCL**

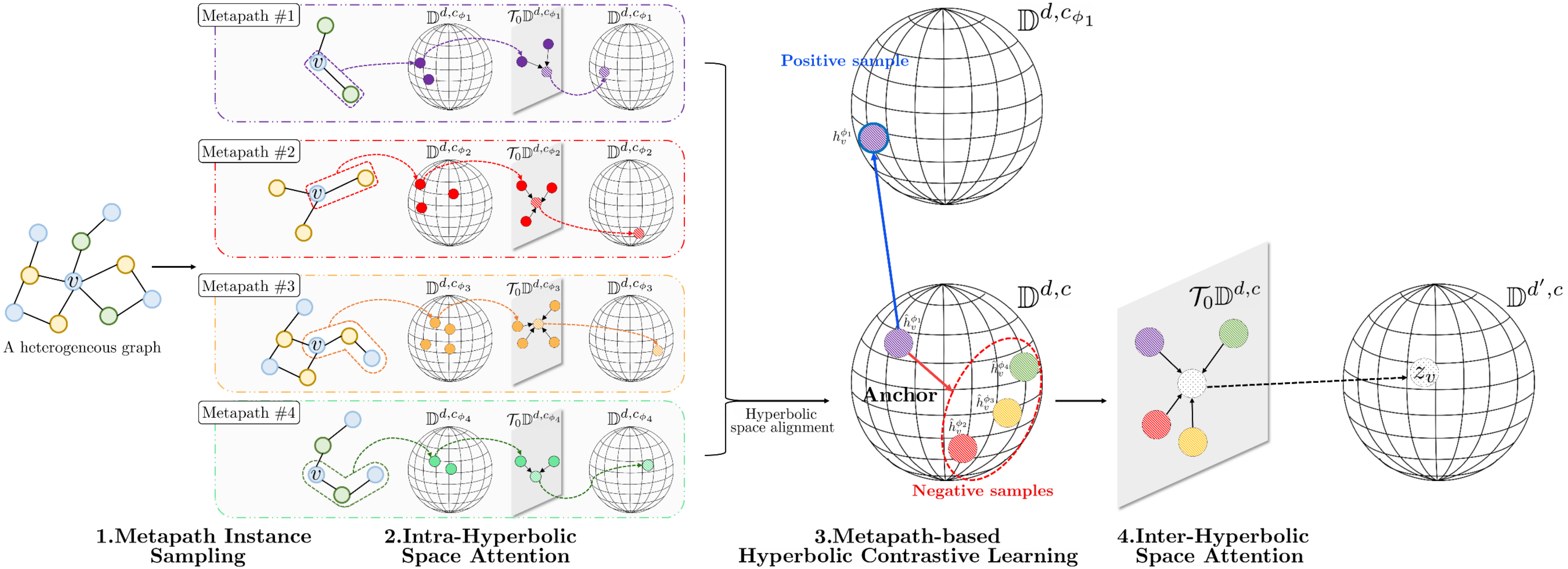


Fig 11. An overview of MHCL.

Experiments

- **Datasets**
  - ✓ Four-real world heterogeneous graph datasets.

Datasets	IMDB	DBLP	ACM	LastFM
Domain	TV program database	Citation Network	Citation Network	Music database
# Nodes	12,772	18,405	8,989	20,612
# Links	18,644	33,973	12,961	107,242
# Node types	3	3	3	3
# Link types	2	2	2	2
Tasks	Node classification & clustering			Link prediction

Table 8. Statistics of datasets.

Experiments

- **Baselines**
  - ✓ The baselines for the four categories.

Categories	Models
ⅰ. **Euclidean homogeneous** GNNs	GCN, GAT
ⅱ. **Hyperbolic homogeneous** GNNs	HGCN, HGCL
ⅲ. **Euclidean heterogeneous** GNNs	HAN, MAGNN, GTN, HGT, Simple-HGN
ⅳ. **Hyperbolic heterogeneous** GNNs	McH-HGCN, SHAN, HHGAT, MSGAT

Table 9. The baselines.

Experiments

- **Task 1 : Node Classification**
  - ✓ In most cases, MHCL outperformed existing hyperbolic GNNs.
  - ✓ A comparison of category **ii** and **iii** shows that **learning heterogeneous information is more crucial** than structural information in heterogeneous graphs.

Datasets	IMDB				DBLP				ACM			
Metric	Macro-F1				Macro-F1				Macro-F1			
Train %	20%	40%	60%	80%	20%	40%	60%	80%	20%	40%	60%	80%
HGCN	54.38	57.08	57.86	57.92	91.69	91.93	92.60	92.58	87.29	89.19	90.01	90.03
HGCL	56.86	58.44	58.90	59.65	92.85	93.12	93.57	93.65	88.11	89.87	91.36	91.57
HAN	56.19	56.84	58.95	58.61	92.63	92.35	92.86	92.73	87.88	90.54	91.22	91.35
Simple-HGN	59.97	61.94	66.73	67.56	93.48	93.98	94.10	94.25	92.25	92.64	93.06	93.55
SHAN	62.23	63.98	66.68	68.49	94.27	94.33	94.50	94.67	92.56	92.88	94.10	**94.94**
HHGAT	63.16	65.07	65.72	67.42	94.19	94.27	94.90	94.77	91.34	92.92	94.28	93.91
MSGAT	<u>65.75</u>	<u>68.07</u>	<u>71.42</u>	<u>70.03</u>	<u>95.44</u>	<u>95.54</u>	**95.67**	<u>95.29</u>	<u>92.73</u>	<u>93.95</u>	<u>94.83</u>	94.01
MHCL	**66.11**	**69.08**	**72.29**	**71.27**	**95.63**	**95.71**	<u>95.38</u>	**95.70**	**93.86**	**94.85**	**95.16**	<u>94.89</u>

Table 10. Experimental results for the node classification task. (Macro-F1)

Experiments

- **Task 1 : Node Classification (Cont.)**
  - ✓ In most cases, MHCL outperformed existing hyperbolic GNNs.
  - ✓ A comparison of category **ii** and **iii** shows that **learning heterogeneous information is more crucial** than structural information in heterogeneous graphs.

Datasets	IMDB				DBLP				ACM			
Metric	Micro-F1				Micro-F1				Micro-F1			
Train %	20%	40%	60%	80%	20%	40%	60%	80%	20%	40%	60%	80%
HGCN	54.46	57.08	58.01	58.54	92.06	92.31	93.16	93.21	88.09	90.06	90.51	91.10
HGCL	55.52	59.50	59.94	60.08	93.32	93.74	93.98	94.30	89.36	89.91	90.66	92.41
HAN	56.71	56.68	58.26	59.35	92.35	92.87	93.42	93.54	91.20	91.78	92.39	92.03
Simple-HGN	63.76	65.60	69.29	69.35	94.17	93.87	94.71	94.68	91.91	92.86	93.33	93.53
SHAN	64.31	66.56	69.57	69.42	94.53	94.60	94.92	95.36	92.38	93.37	94.46	<u>94.56</u>
HHGAT	65.76	66.34	70.40	69.61	94.66	94.72	95.15	95.34	92.36	93.46	94.34	93.72
MSGAT	<u>69.09</u>	<u>70.95</u>	<u>73.60</u>	<u>73.37</u>	<u>95.79</u>	<u>95.90</u>	**95.98**	<u>95.85</u>	<u>92.96</u>	<u>93.91</u>	<u>94.88</u>	94.05
MHCL	**69.54**	**71.71**	**74.63**	**74.24**	**95.82**	**96.07**	<u>95.70</u>	**96.05**	**93.83**	**94.82**	**95.47**	**94.96**

Table 11. Experimental results for the node classification task. (Micro-F1)

Experiments

- ## Task 2 : Node Clustering

Datasets	IMDB		DBLP		ACM	
Metric	NMI	ARI	NMI	ARI	NMI	ARI
HGCN	10.29	11.10	76.48	79.36	60.19	62.06
HGCL	12.89	14.78	78.06	80.49	62.84	64.50
HAN	11.21	11.49	77.03	82.53	61.24	64.11
Simple-HGN	17.58	19.51	82.38	85.71	69.91	72.07
SHAN	20.60	22.56	82.39	86.13	72.90	77.73
HHGAT	20.75	22.80	83.14	85.91	72.49	77.92
MSGAT	24.06	26.33	84.38	88.27	73.33	78.28
MHCL	**26.18**	**30.87**	**84.52**	**88.89**	**76.62**	**81.73**

Table 12. Experimental results for the node clustering task.

- ## Task 3 : Link Prediction

Dataset	Metric	HGCN	HGCL	HAN	HGT	Simple-HGN	HHGAT	MSGAT	MHCL
LastFM	ROC-AUC	46.71	46.99	48.32	47.78	53.85	54.37	55.77	**56.39**
	F1-Score	57.23	58.03	57.11	61.16	63.02	62.85	63.39	**63.48**

Table 13. Experimental results for the link prediction task.

Experiments

- **Ablation Study**
  - ✓ The ablation study demonstrates the effectiveness of metapath-based contrastive learning across various embedding spaces.

Datasets	IMDB				DBLP				ACM			
Metric	Macro-F1	Micro-F1	NMI	ARI	Macro-F1	Micro-F1	NMI	ARI	Macro-F1	Micro-F1	NMI	ARI
MHCL	**72.29**	**74.63**	**26.18**	**30.87**	**95.38**	**95.70**	**84.52**	**88.89**	**95.16**	**95.47**	**76.62**	**81.73**
$\text{MHCL}_{\text{w/o Cont}}$	70.92	73.18	23.90	27.70	94.90	95.15	83.69	87.84	94.60	94.53	73.85	78.01
$\text{MHCL}_{\text{Single+Cont}}$	67.88	69.49	23.73	28.63	94.42	94.85	80.26	86.38	93.23	93.20	74.16	78.85
$\text{MHCL}_{\text{Single}}$	66.19	68.35	21.33	25.40	93.74	93.86	80.12	86.05	92.40	92.46	73.53	78.10
$\text{MHCL}_{\text{Euclid+Cont}}$	66.59	68.37	21.79	26.17	93.60	93.73	79.75	83.98	91.39	91.43	71.04	74.82
$\text{MHCL}_{\text{Euclid}}$	64.56	67.32	16.93	16.16	93.18	93.53	78.94	83.68	90.32	90.51	70.16	73.31

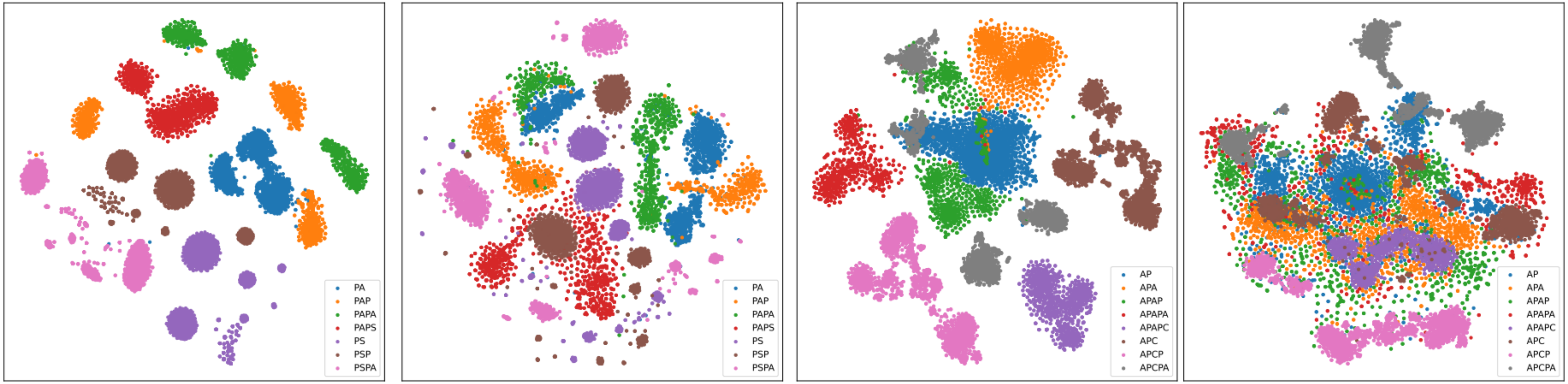
Table 14. Results of the ablation study.

Models	$\text{MHCL}_{\text{w/o Cont}}$	$\text{MHCL}_{\text{Single+Cont}}$	$\text{MHCL}_{\text{Single}}$	$\text{MHCL}_{\text{Euclid+Cont}}$	$\text{MHCL}_{\text{Euclid}}$
Embedding Space	Multi-Hyperbolic	Singe-Hyperbolic		Euclidean	
Contrastive Learning	✗	✓	✗	✓	✗

Table 15. Variant models for ablation study.

Experiments

- **Visualization**
  - ✓ The visualization of metapath embeddings shows that using MHCL **enhances the separability** among **different metapaths**.



(a) ACM w/ contrastive learning (b) ACM w/o contrastive learning (c) DBLP w/ contrastive learning (d) DBLP w/o contrastive learning

Figure 12. Visualization of metapath embeddings.

Part 3

Hyperbolic Heterogeneous Graph Transformer

Background

- **Previous Hyperbolic Heterogeneous Graph Learning**
  - ✓ **Frequent mappings** between the hyperbolic space and the tangent space.
  - ✓ **Prior knowledge** is required to leverage metapaths effectively.
  - ✓ Message-passing makes it **difficult to learn global hierarchical structures**.

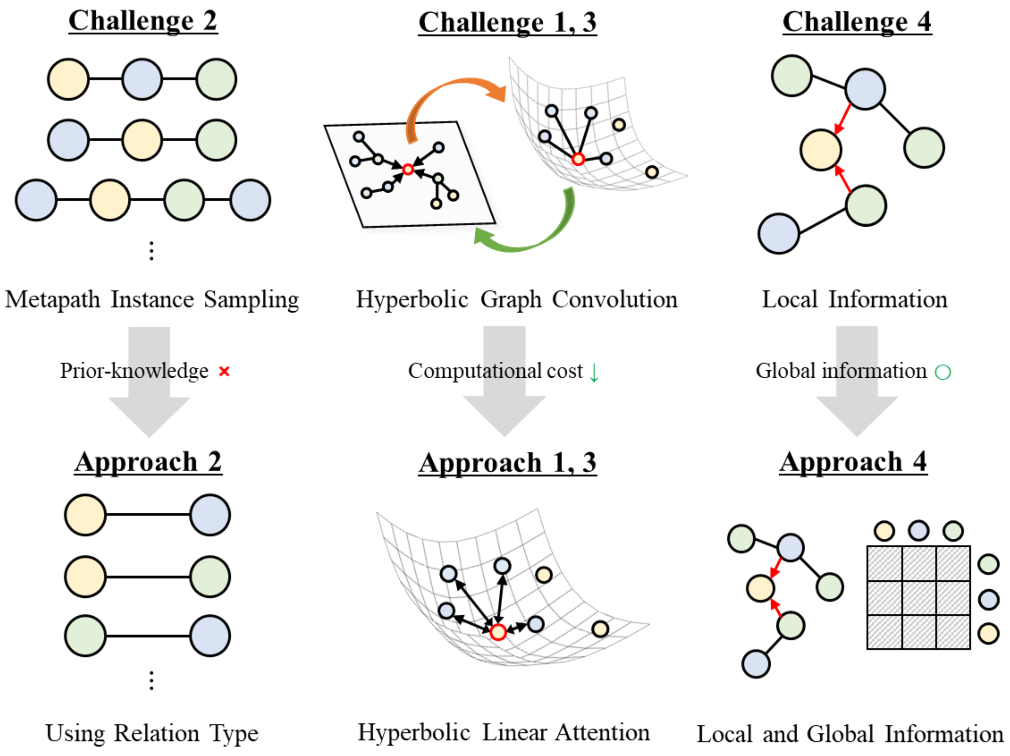


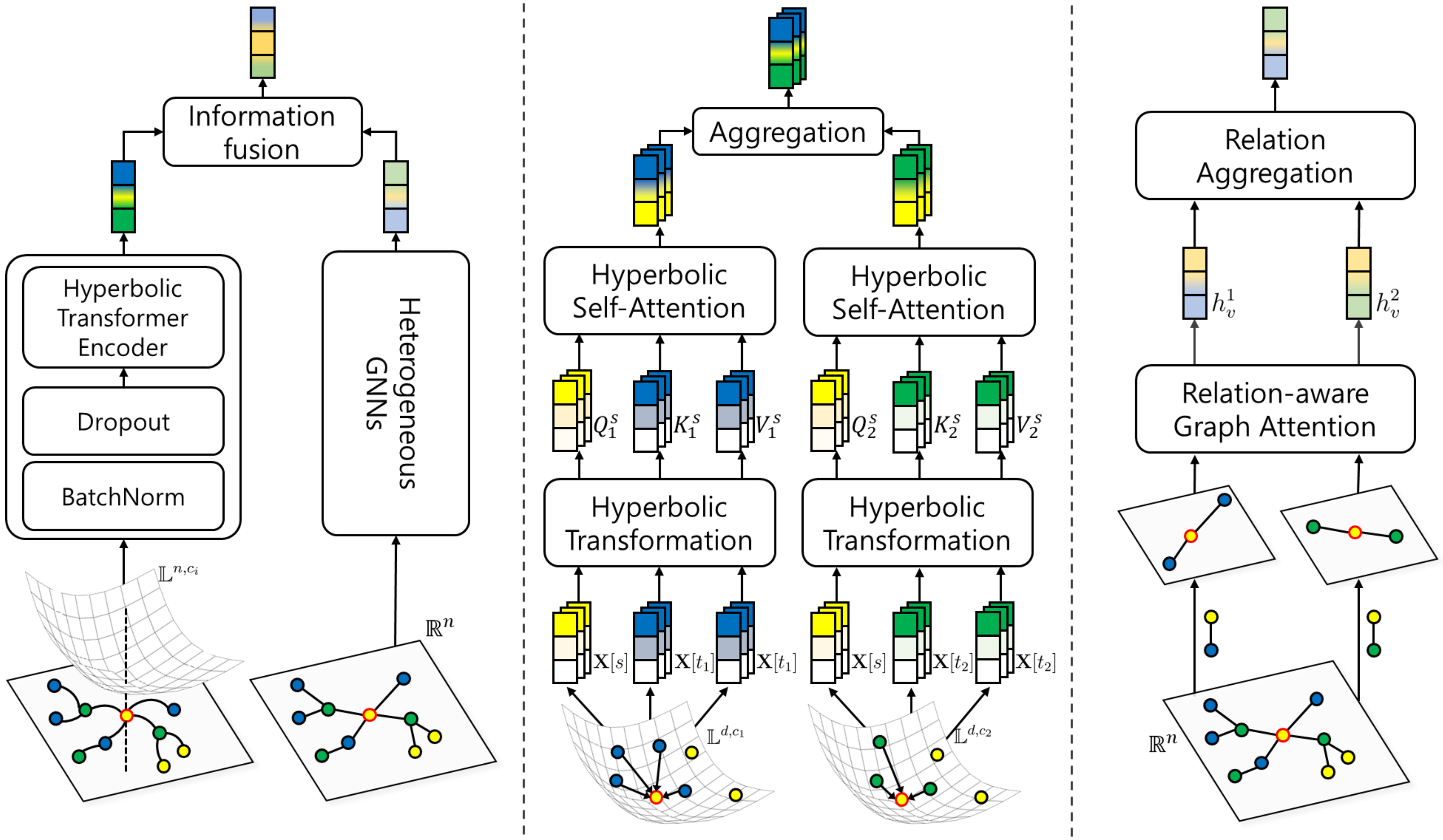
Figure 13. Previous limitations and proposed solutions.

Background

- **Graph Transformer**
  - ✓ Learn relationships between distant nodes using **self-attention**.
  - ✓ Encodes **structural information** to represent **diverse graph structures**.
  - ✓ Time complexity : $O(n^2)$
- **Linear Transformer**
  - ✓ Reduces computation via **approximate attention methods**.
  - ✓ Linear time complexity **suited for large-scale data**.
  - ✓ Time complexity : $O(n)$
- **Hyperbolic Transformer**
  - ✓ Representation learning on hyperbolic space via Transformer models.
  - ✓ **Learn hierarchical structures** effectively.
  - ✓ Time complexity : $O(n)$ ~ $O(n^2)$

Methodology

- **Overall framework of HypHGT**



(a) Overall Framework **(b) Hyperbolic Transformer Encoder** **(c) Heterogeneous GNNs**

Figure 14. Framework

Experiments

- **Datasets**
  - ✓ Three-real world heterogeneous graph datasets.

Datasets	IMDB	DBLP	ACM
Domain	TV program database	Citation Network	Citation Network
# Nodes	12,772	18,405	8,989
# Links	18,644	33,973	12,961
# Node types	3	3	3
# Link types	2	2	2
Tasks	Node classification		

Table 16. Statistics of datasets.

Experiments

- **Baselines**
  - ✓ The baselines for the four categories.

Categories	Models
ⅰ. **Euclidean homogeneous** models	GCN, GAT
ⅱ. **Hyperbolic homogeneous** models	HGCN, Hypformer
ⅲ. **Euclidean heterogeneous** models	HAN, MAGNN, GTN, HGT, Simple-HGN
ⅳ. **Hyperbolic heterogeneous** models	SHAN, HHGAT, MSGAT

Table 17. The baselines.

Experiments

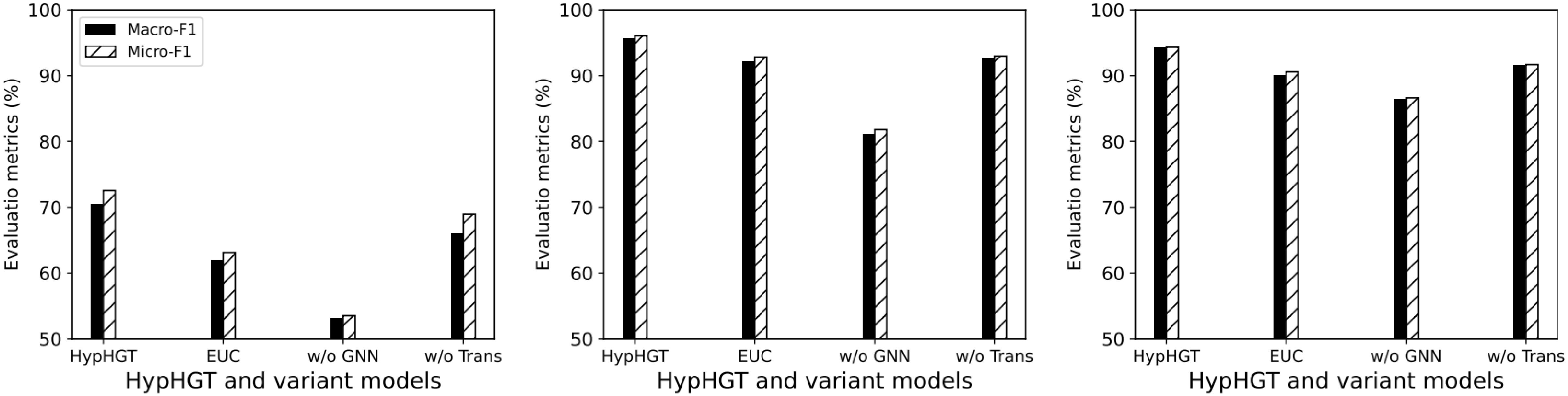
- **Node Classification**
  - ✓ In all cases, HypHGT outperformed existing baselines.
  - ✓ Compared to Hypformer, HypHGT can more effectively learn relation-based heterogeneity in heterogeneous graphs.

Datasets	IMDB		DBLP		ACM	
Metric	Macro-F1	Micro-F1	Macro-F1	Micro-F1	Macro-F1	Micro-F1
GAT	57.06	57.53	91.29	92.15	88.95	89.06
HGCN	57.98	58.68	92.70	93.39	90.26	90.71
Hypformer	65.12	68.15	94.18	94.36	93.81	93.90
GTN	62.73	64.26	93.83	94.18	92.54	92.56
HGT	62.87	63.29	93.96	94.02	91.79	92.07
SHAN	66.75	69.99	94.46	94.98	93.71	<u>94.32</u>
HHGAT	66.38	70.28	93.76	94.56	93.62	93.14
MSGAT	<u>68.91</u>	<u>70.45</u>	<u>94.51</u>	<u>95.28</u>	<u>93.84</u>	93.95
HypHGT	**70.47**	**72.56**	**95.68**	**96.04**	**94.24**	**94.35**

Table 18. Experimental results for the node classification task.

Experiments

- **Ablation Study**
  - ✓ The ablation study demonstrates the effectiveness of each proposed module.

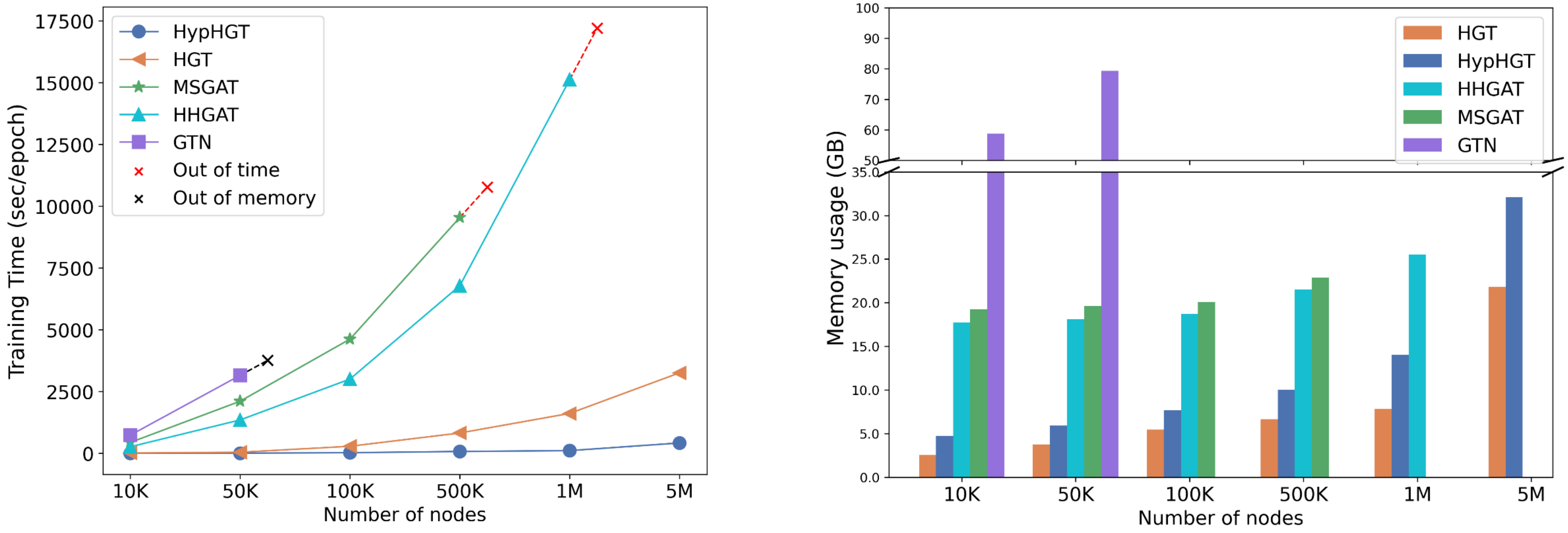


(a) Results on IMDB dataset. **(b) Results on DBLP dataset.** **(c) Results on ACM dataset.**

Figure 15. Results of the ablation study.

Experiments

- **Scalability Analysis**
  - ✓ HypHGT efficiently scales to large synthetic heterogeneous graphs compared to other baselines.



(a) Time consumption for varying number of nodes. **(b) Memory consumption for varying number of nodes.**

Figure 16. Scalability analysis of HypHGT on synthetic heterogeneous graphs.

Experiments

- **Efficiency Analysis**
  - ✓ Efficiency analysis demonstrates that the hyperbolic linear-attention reduces computational costs.

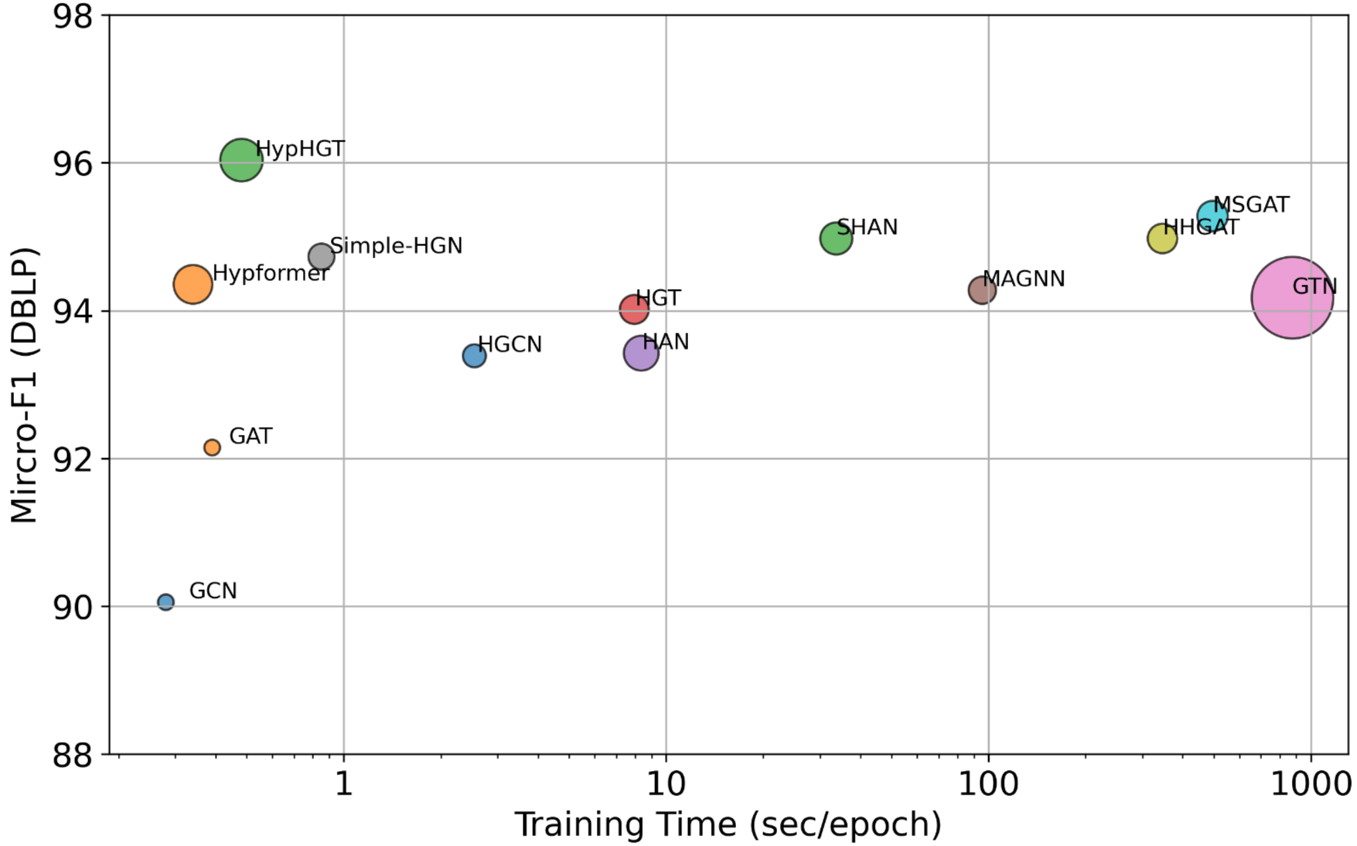


Figure 17. Time and memory comparison for baselines on DBLP dataset.

Q&A