MAGNN: Metapath Aggregated GNN for Heterogenous Graph Embedding [1]

Fu, Zhang, Meng, and King

Presented by: Rozhan Akhound-Sadegh

UNIVERSITY OF WATERLOO | FACULTY OF MATHEMATICS

1
Introduction

Heterogeneous Graphs

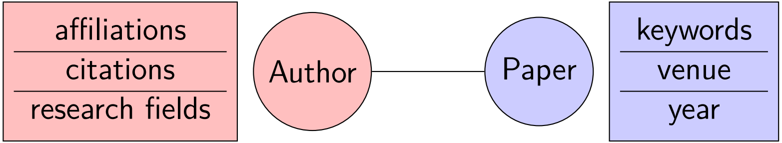
- GNNs assume homogeneous graphs (1 node type and 1 edge type)

Heterogeneous Graphs

- GNNs assume homogeneous graphs (1 node type and 1 edge type)
- Heterogenous graphs: node/edge attributes in different feature spaces

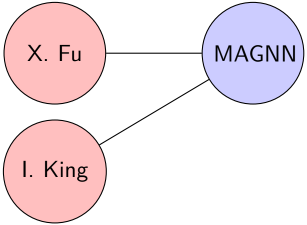
Heterogeneous Graphs

- GNNs assume homogeneous graphs (1 node type and 1 edge type)
- Heterogenous graphs: node/edge attributes in different feature spaces
- Co-authorship network



Metapaths

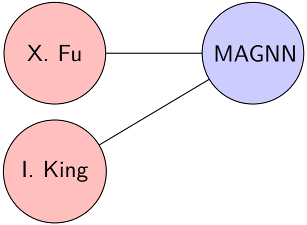
- Author-Paper-Author



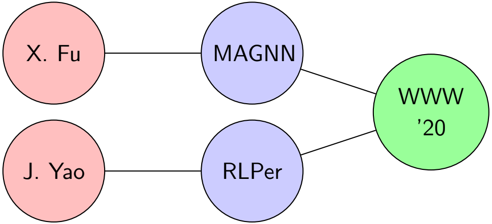
UNIVERSITY OF WATERLOO | FACULTY OF MATHEMATICS

Metapaths

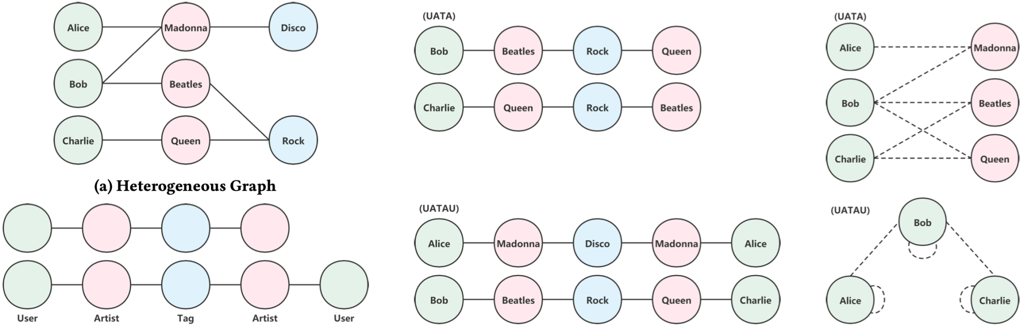
- Author-Paper-Author



- Author-Paper-Venue-Paper-Author



Metapaths and Metapath-based Graphs



(a) Heterogeneous Graph (b) Metapaths (c) Metapath Instances (d) Metapath-based Graphs

Previous Work Limitations

- Do not leverage node content features
- Discard all intermediate nodes along metapath
- Rely on a single metapath for graph embedding
    - Require manual selection
    - Loss of information (from other metapaths)

2
Metapath Aggregated Graph Neural Network

MAGNN Architecture

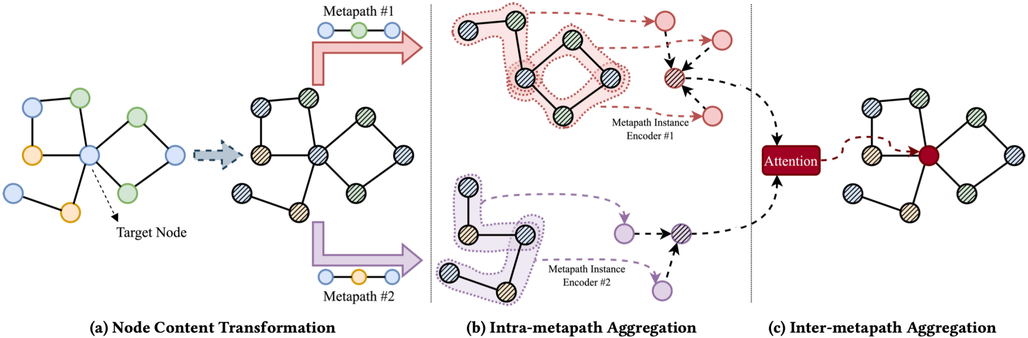


Figure 2: The overall architecture of MAGNN (path instances that start and end with the target node are omitted for clarity).

MAGNN Architecture Overview

- Node content transformation

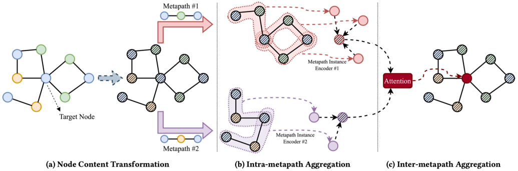


Figure 2: The overall architecture of MAGNN (path instances that start and end with the target node are omitted for clarity).

MAGNN Architecture Overview

- Node content transformation
  - Type-specific linear transformations to transform heterogenous node attributes to the same latent space

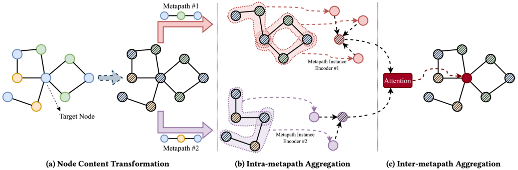
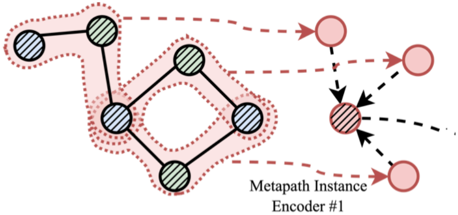


Figure 2: The overall architecture of MAGNN (path instances that start and end with the target node are omitted for clarity).

MAGNN Architecture Overview

- Node content transformation
- Intra-metapath aggregation



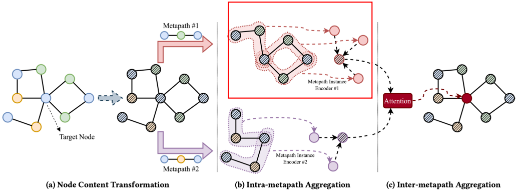


Figure 2: The overall architecture of MAGNN (path instances that start and end with the target node are omitted for clarity).

MAGNN Architecture Overview

- Node content transformation
- Intra-metapath aggregation
- Inter-metapath aggregation

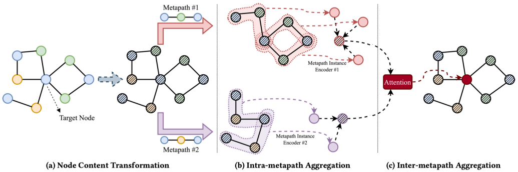


Figure 2: The overall architecture of MAGNN (path instances that start and end with the target node are omitted for clarity).

Node Transformation

- Project into latent space

$$\mathbf{h}'_v = \mathbf{W}_A \cdot \mathbf{x}_v^A$$

Node Transformation

- Project into latent space

$$\mathbf{h}'_v = \mathbf{W}_A \cdot \mathbf{x}^A_v$$

- node type $A \in \mathcal{A}$

Node Transformation

- Project into latent space

$$\mathbf{h}'_v = \mathbf{W}_A \cdot \mathbf{x}_v^A$$

- node type $A \in \mathcal{A}$
- node $v \in \mathcal{V}_A$

Node Transformation

- Project into latent space

$$\mathbf{h}'_v = \mathbf{W}_A \cdot \mathbf{x}_v^A$$

- feature vector $\mathbf{x}_v^A \in \mathbb{R}^{d_A}$

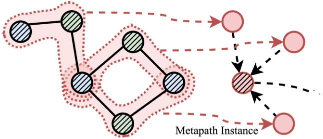
Node Transformation

- Project into latent space

$$\mathbf{h}'_v = \boxed{\mathbf{W}_A} \cdot \mathbf{x}_v^A$$

- feature vector $\mathbf{x}_v^A \in \mathbb{R}^{d_A}$
- parametric weight matrix for type A nodes $\mathbf{W}_A \in \mathbb{R}^{d' \times d_A}$

Intra-metapath Aggregation I: Instance Encoding

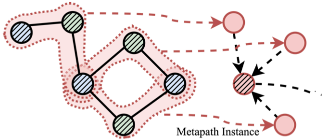


- metapath instance encoder:

$$\mathbf{h}_{P(v,u)} = f_\theta(P(v,u)) = f_\theta\left(\mathbf{h}'_v, \mathbf{h}'_u, \left\{\mathbf{h}'_t, \forall t \in \{m^{P(v,u)}\}\right\}\right)$$

UNIVERSITY OF WATERLOO | FACULTY OF MATHEMATICS

Intra-metapath Aggregation I: Instance Encoding

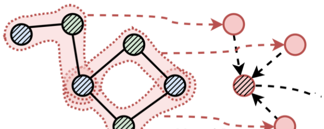


- metapath instance encoder:

$$\mathbf{h}_{P(v,u)} = f_\theta(P(v,u)) = f_\theta\left(\mathbf{h}'_v, \mathbf{h}'_u, \left\{\mathbf{h}'_t, \forall t \in \{m^{P(v,u)}\}\right\}\right)$$

- $P(v, u)$: metapath connecting target node v with metapath-based neighbor u

Intra-metapath Aggregation I: Instance Encoding

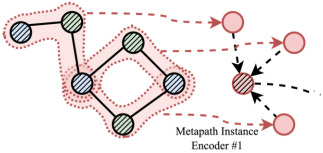


- metapath instance encoder:

$$\mathbf{h}_{P(v,u)} = f_\theta(P(v,u)) = f_\theta\left(\mathbf{h}'_v, \mathbf{h}'_u, \left\{\mathbf{h}'_t, \forall t \in \{m^{P(v,u)}\}\right\}\right)$$

- Intermediate nodes $\{m^{P(v,u)}\} = P(v,u) \setminus \{u, v\}$

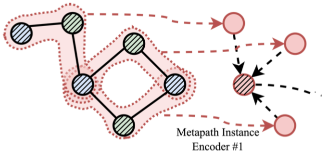
Intra-metapath Aggregation II: Aggregation with Attention



- Graph attention: sum metapath instances of P related to target node v

UNIVERSITY OF WATERLOO | FACULTY OF MATHEMATICS

Intra-metapath Aggregation II: Aggregation with Attention



- Graph attention: sum metapath instances of P related to target node v
- $$e_{vu}^{P} = \text{LeakyReLU}\left(\mathbf{a}_P^\top \cdot \left[\mathbf{h}_v' \,||\, \mathbf{h}_{P(v,u)}\right]\right)$$

$$\mathbf{h}_v^P = \sigma\left(\sum_{u \in \mathcal{N}_v^P} \alpha_{vu}^P \cdot \mathbf{h}_{P(v,u)}\right), \quad \alpha_{vu}^P = \frac{\exp\left(e_{vu}^P\right)}{\sum_{s \in \mathcal{N}_v^P} \exp(e_{vs}^P)}$$

Intra-metapath Aggregation II: Aggregation with Attention

- Graph attention: sum metapath instances of P related to target node v
- $$e_{vu}^{P} = \text{LeakyReLU}\left(\mathbf{a}_P^{\top} \cdot \left[\mathbf{h}'_v \,||\, \mathbf{h}_{P(v,u)}\right]\right)$$
$$\mathbf{h}_v^{P} = \sigma\left(\sum_{u \in \mathcal{N}_v^P} \alpha_{vu}^{P} \cdot \mathbf{h}_{P(v,u)}\right), \quad \alpha_{vu}^{P} = \frac{\exp\left(e_{vu}^{P}\right)}{\sum_{s \in \mathcal{N}_v^P} \exp(e_{vs}^{P})}$$
- parameterized attention vector $\mathbf{a}_P^{\top} \in \mathbb{R}^{2d'}$

Intra-metapath Aggregation II: Aggregation with Attention

- Graph attention: sum metapath instances of P related to target node v
- $$e^P_{vu} = \text{LeakyReLU}\left(\mathbf{a}_P^\top \cdot \left[\mathbf{h}'_v \mid\mid \mathbf{h}_{P(v,u)}\right]\right)$$
$$\mathbf{h}^P_v = \sigma\left(\sum_{u \in \mathcal{N}^P_v} \alpha^P_{vu} \cdot \mathbf{h}_{P(v,u)}\right), \quad \alpha^P_{vu} = \frac{\exp\left(e^P_{vu}\right)}{\sum_{s \in \mathcal{N}^P_v} \exp(e^P_{vs})}$$
- Normalized importance weight α^P_{vu}

Inter-metapath Aggregation

- For a node type A, we have latent vectors $\{\mathbf{h}_v^{P_1}, \mathbf{h}_v^{P_2}, \cdots, \mathbf{h}_v^{P_M}\}$ for each $v \in \mathcal{V}_A$

Inter-metapath Aggregation

- For a node type A, we have latent vectors $\{\mathbf{h}_v^{P_1}, \mathbf{h}_v^{P_2}, \cdots, \mathbf{h}_v^{P_M}\}$ for each $v \in \mathcal{V}_A$
- Use attention mechanism to learn weights for metapaths

$$s_{P_i} = \frac{1}{|\mathcal{V}_A|} \sum_{v \in \mathcal{V}_A} \tanh\left(\mathbf{M}_A \cdot \mathbf{h}_v^{P_i} + \mathbf{b}_A\right) \qquad e_{P_i} = \mathbf{q}_A^\top \cdot s_{p_i}$$

$$\mathbf{h}_v^{\mathcal{P}_A} = \sum_{P \in \mathcal{P}_A} \beta_P \cdot \mathbf{h}^P, \quad \beta_{P_i} = \frac{\exp\left(e_{P_i}\right)}{\sum_{P \in \mathcal{P}_A} \exp\left(e_{P_P}\right)}$$

Inter-metapath Aggregation

- For a node type A, we have latent vectors $\{\mathbf{h}_v^{P_1}, \mathbf{h}_v^{P_2}, \cdots, \mathbf{h}_v^{P_M}\}$ for each $v \in \mathcal{V}_A$
- Use attention mechanism to learn weights for metapaths

$$s_{P_i} = \frac{1}{|\mathcal{V}_A|} \sum_{v \in \mathcal{V}_A} \tanh\left(\mathbf{M}_A \cdot \mathbf{h}_v^{P_i} + \mathbf{b}_A\right) \qquad e_{P_i} = \mathbf{q}_A^\top \cdot s_{p_i}$$

$$\mathbf{h}_v^{\mathcal{P}_A} = \sum_{P \in \mathcal{P}_A} \beta_P \cdot \mathbf{h}^P, \qquad \beta_{P_i} = \frac{\exp\left(e_{P_i}\right)}{\sum_{P \in \mathcal{P}_A} \exp\left(e_{P_P}\right)}$$

Inter-metapath Aggregation

- For a node type A, we have latent vectors $\{\mathbf{h}_v^{P_1}, \mathbf{h}_v^{P_2}, \cdots, \mathbf{h}_v^{P_M}\}$ for each $v \in \mathcal{V}_A$
- Use attention mechanism to learn weights for metapaths

$$s_{P_i} = \frac{1}{|\mathcal{V}_A|} \sum_{v \in \mathcal{V}_A} \tanh\left(\mathbf{M}_A \cdot \mathbf{h}_v^{P_i} + \mathbf{b}_A\right) \qquad e_{P_i} = \mathbf{q}_A^\top \cdot s_{p_i}$$

$$\mathbf{h}_v^{\mathcal{P}_A} = \sum_{P \in \mathcal{P}_A} \beta_P \cdot \mathbf{h}^P, \qquad \beta_{P_i} = \frac{\exp\left(e_{P_i}\right)}{\sum_{P \in \mathcal{P}_A} \exp\left(e_{P_P}\right)}$$

Inter-metapath Aggregation

- For a node type A, we have latent vectors $\{\mathbf{h}_v^{P_1}, \mathbf{h}_v^{P_2}, \cdots, \mathbf{h}_v^{P_M}\}$ for each $v \in \mathcal{V}_A$
- Use attention mechanism to learn weights for metapaths

$$s_{P_i} = \frac{1}{|\mathcal{V}_A|} \sum_{v \in \mathcal{V}_A} \tanh\left(\mathbf{M}_A \cdot \mathbf{h}_v^{P_i} + \mathbf{b}_A\right) \qquad e_{P_i} = \mathbf{q}_A^\top \cdot s_{p_i}$$

$$\mathbf{h}_v^{\mathcal{P}_A} = \sum_{P \in \mathcal{P}_A} \beta_P \cdot \mathbf{h}^P, \qquad \beta_{P_i} = \frac{\exp\left(e_{P_i}\right)}{\sum_{P \in \mathcal{P}_A} \exp\left(e_{P_P}\right)}$$

- Apply final linear layer

$$\mathbf{h}_v = \sigma\left(\mathbf{W}_o \cdot \mathbf{h}_v^{\mathcal{P}_A}\right)$$

Metapath Instance Encoders f_θ

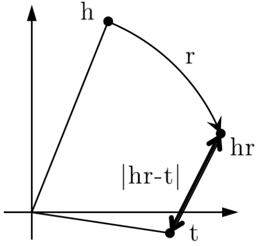
- Mean $f_\theta(P(v,u) := \mathtt{MEAN}\left(\{\mathbf{h}'_t, \forall t \in P(v,u)\}\right)$

Metapath Instance Encoders f_θ

- Mean $f_\theta(P(v,u) := \texttt{MEAN}\left(\{\mathbf{h}'_t, \forall t \in P(v,u)\}\right)$
- Linear $f_\theta(P(v,u) := \mathbf{W}_P \cdot \texttt{MEAN}\left(\{\mathbf{h}'_t, \forall t \in P(v,u)\}\right)$

Metapath Instance Encoders f_θ

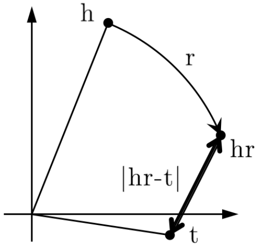
- Mean $f_\theta(P(v,u) := \texttt{MEAN}\left(\{\mathbf{h}'_t, \forall t \in P(v,u)\}\right)$
- Linear $f_\theta(P(v,u) := \mathbf{W}_P \cdot \texttt{MEAN}\left(\{\mathbf{h}'_t, \forall t \in P(v,u)\}\right)$
- Relational Rotation



(b) RotatE models $\mathbf{r}$ as rotation in complex plane.

Metapath Instance Encoders f_θ

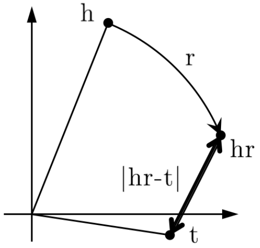
- Mean $f_\theta(P(v,u) := \texttt{MEAN}\left(\{\mathbf{h}'_t, \forall t \in P(v,u)\}\right)$
- Linear $f_\theta(P(v,u) := \mathbf{W}_P \cdot \texttt{MEAN}\left(\{\mathbf{h}'_t, \forall t \in P(v,u)\}\right)$
- Relational Rotation
  - Mean and Linear encoders ignore node ordering in metapath



(b) RotatE models $\mathbf{r}$ as rotation in complex plane.

Metapath Instance Encoders f_θ

- Mean $f_\theta(P(v,u) := \texttt{MEAN}\left(\{\mathbf{h}'_t, \forall t \in P(v,u)\}\right)$
- Linear $f_\theta(P(v,u) := \mathbf{W}_P \cdot \texttt{MEAN}\left(\{\mathbf{h}'_t, \forall t \in P(v,u)\}\right)$
- Relational Rotation
  - Mean and Linear encoders ignore node ordering in metapath
  - $\mathbf{o}_0 = \mathbf{h}'_{t_0} = \mathbf{h}'_u$
    $\mathbf{o}_i = \mathbf{h}'_{t_i} + \mathbf{o}_{i-1} \odot \mathbf{r}_i$
    $\mathbf{h}_{P(v,u)} = \frac{\mathbf{o}_n}{n+1}$



(b) RotatE models $\mathbf{r}$ as rotation in complex plane.

3
Experiments

Training

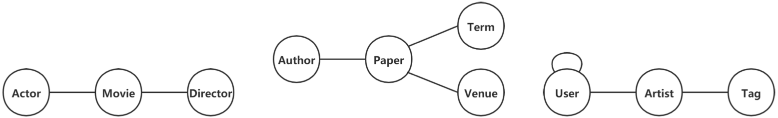
- Semi-supervised

$$\mathcal{L} = -\sum_{v \in \mathcal{V}_L} \sum_{c=1}^{C} \mathbf{y}_v[c] \cdot \log \mathbf{h}_v[c],$$

- Unsupervised (w/ negative sampling)

$$\mathcal{L} = -\sum_{(u,v) \in \Omega} \log \sigma \left(\mathbf{h}_u^\top \cdot \mathbf{h}_v \right) - \sum_{(u',v') \in \Omega^-} \log \sigma \left(-\mathbf{h}_{u'}^\top \cdot \mathbf{h}_{v'} \right),$$

Datasets



(a) IMDb (b) DBLP (c) Last.fm

Dataset	Node	Edge	Metapath
IMDb	# movie (M): 4,278 # director (D): 2,081 # actor (A): 5,257	# M-D: 4,278 # M-A: 12,828	MDM MAM DMD DMAMD AMA AMDMA
DBLP	# author (A): 4,057 # paper (P): 14,328 # term (T): 7,723 # venue (V): 20	# A-P: 19,645 # P-T: 85,810 # P-V: 14,328	APA APTPA APVPA
Last.fm	# user (U): 1,892 # artist (A): 17,632 # tag (T): 1,088	# U-U: 12,717 # U-A: 92,834 # A-T: 23,253	UU UAU UATAU AUA AUUA ATA

Baselines

- Homogeneous: Dropping all node features

Baselines

- Homogeneous: Dropping all node features
  - LINE

Baselines

- Homogeneous: Dropping all node features
  - LINE
  - node2vec

Baselines

- Homogeneous: Dropping all node features
  - LINE
  - node2vec
- Heterogenous Model

Baselines

- Homogeneous: Dropping all node features
  - LINE
  - node2vec
- Heterogenous Model
  - ESim

Baselines

- Homogeneous: Dropping all node features
  - LINE
  - node2vec
- Heterogenous Model
  - ESim
  - metapath2vec

Baselines

- Homogeneous: Dropping all node features
  - LINE
  - node2vec
- Heterogenous Model
  - ESim
  - metapath2vec
  - HERec

Baselines

- Homogeneous: Dropping all node features
  - LINE
  - node2vec
- Heterogenous Model
  - ESim
  - metapath2vec
  - HERec
  - GATNE

Baselines

- Homogeneous: Dropping all node features
  - LINE
  - node2vec
- Heterogenous Model
  - ESim
  - metapath2vec
  - HERec
  - GATNE
  - HAN

Baselines

- Homogeneous: Dropping all node features
  - LINE
  - node2vec
- Heterogenous Model
  - ESim
  - metapath2vec
  - HERec
  - GATNE
  - HAN
- Run on Best Metapath-based Homogeneous Graph

Baselines

- Homogeneous: Dropping all node features
  - LINE
  - node2vec
- Heterogenous Model
  - ESim
  - metapath2vec
  - HERec
  - GATNE
  - HAN
- Run on Best Metapath-based Homogeneous Graph
  - GCN

Baselines

- Homogeneous: Dropping all node features
  - LINE
  - node2vec
- Heterogenous Model
  - ESim
  - metapath2vec
  - HERec
  - GATNE
  - HAN
- Run on Best Metapath-based Homogeneous Graph
  - GCN
  - GAT

Experiments

- Node Classification

Experiments

- Node Classification

Dataset	Metrics	Train %	Unsupervised					Semi-supervised			
			LINE	node2vec	ESim	metapath2vec	HERec	GCN	GAT	HAN	MAGNN
IMDb	Macro-F1	20%	44.04	49.00	48.37	46.05	45.61	52.73	53.64	56.19	**59.35**
		40%	45.45	50.63	50.09	47.57	46.80	53.67	55.50	56.15	**60.27**
		60%	47.09	51.65	51.45	48.17	46.84	54.24	56.46	57.29	**60.66**
		80%	47.49	51.49	51.37	49.99	47.73	54.77	57.43	58.51	**61.44**
	Micro-F1	20%	45.21	49.94	49.32	47.22	46.23	52.80	53.64	56.32	**59.60**
		40%	46.92	51.77	51.21	48.17	47.89	53.76	55.56	57.32	**60.50**
		60%	48.35	52.79	52.53	49.87	48.19	54.23	56.47	58.42	**60.88**
		80%	48.98	52.72	52.54	50.50	49.11	54.63	57.40	59.24	**61.53**
DBLP	Macro-F1	20%	87.16	86.70	90.68	88.47	90.82	88.00	91.05	91.69	**93.13**
		40%	88.85	88.07	91.61	89.91	91.44	89.00	91.24	91.96	**93.23**
		60%	88.93	88.69	91.84	90.50	92.08	89.43	91.42	92.14	**93.57**
		80%	89.51	88.93	92.27	90.86	92.25	89.98	91.73	92.50	**94.10**
	Micro-F1	20%	87.68	87.21	91.21	89.02	91.49	88.51	91.61	92.33	**93.61**
		40%	89.25	88.51	92.05	90.36	92.05	89.22	91.77	92.57	**93.68**
		60%	89.34	89.09	92.28	90.94	92.66	89.57	91.97	92.72	**93.99**
		80%	89.96	89.37	92.68	91.31	92.78	90.33	92.24	93.23	**94.47**

Experiments

- Node Classification
- Node Clustering

Dataset	Metrics	Unsupervised					Semi-supervised			
		LINE	node2vec	ESim	metapath2vec	HERec	GCN	GAT	HAN	MAGNN
IMDb	NMI	1.13	5.22	1.07	0.89	0.39	7.46	7.84	10.79	**15.58**
	ARI	1.20	6.02	1.01	0.22	0.11	7.69	8.87	11.11	**16.74**
DBLP	NMI	71.02	77.01	68.33	74.18	69.03	73.45	70.73	77.49	**80.81**
	ARI	76.52	81.37	72.22	78.11	72.45	77.50	76.04	82.95	**85.54**

Experiments

- Node Classification
- Node Clustering
- Link Prediction

Dataset	Metrics	LINE	node2vec	ESim	metapath2vec	HERec	GCN	GAT	GATNE	HAN	MAGNN
Last.fm	AUC	85.76	67.14	82.00	92.20	91.52	90.97	92.36	89.21	93.40	**98.91**
	AP	88.07	64.11	82.19	90.11	89.47	91.65	91.55	88.86	92.44	**98.93**

Ablation Study

- Encoder

Ablation Study

- Encoder
  - $\text{MAGNN}_{\text{rot}}$: Relational rotation encoder

Ablation Study

- Encoder
  - $\text{MAGNN}_{\text{rot}}$: Relational rotation encoder
  - $\text{MAGNN}_{\text{linear}}$: Linear encoder

Ablation Study

- Encoder
  - $\text{MAGNN}_{\text{rot}}$: Relational rotation encoder
  - $\text{MAGNN}_{\text{linear}}$: Linear encoder
  - $\text{MAGNN}_{\text{avg}}$: MEAN encoder

Ablation Study

- Encoder
  - $\text{MAGNN}_{\text{rot}}$: Relational rotation encoder
  - $\text{MAGNN}_{\text{linear}}$: Linear encoder
  - $\text{MAGNN}_{\text{avg}}$: MEAN encoder
- MAGNN_{sm}: Single best metapath

Ablation Study

- Encoder
  - $\text{MAGNN}_{\text{rot}}$: Relational rotation encoder
  - $\text{MAGNN}_{\text{linear}}$: Linear encoder
  - $\text{MAGNN}_{\text{avg}}$: MEAN encoder
- MAGNN_{sm}: Single best metapath
- MAGNN_{nb}: Only metapath-based neighbors

Ablation Study

- Encoder
  - $\text{MAGNN}_{\text{rot}}$: Relational rotation encoder
  - $\text{MAGNN}_{\text{linear}}$: Linear encoder
  - $\text{MAGNN}_{\text{avg}}$: MEAN encoder
- MAGNN_{sm}: Single best metapath
- MAGNN_{nb}: Only metapath-based neighbors
- $\text{MAGNN}_{\text{feat}}$: No node features

Ablation Results

Variant	IMDb				DBLP				Last.fm	
	Macro-F1	Micro-F1	NMI	ARI	Macro-F1	Micro-F1	NMI	ARI	AUC	AP
$\text{MAGNN}_{\text{feat}}$	48.87	50.36	5.82	5.30	92.80	93.32	77.17	82.15	N/A	N/A
MAGNN_{nb}	58.45	58.84	12.87	11.98	92.61	93.15	77.64	82.60	93.68	92.95
MAGNN_{sm}	56.77	56.64	11.90	11.84	93.19	93.69	79.48	84.39	92.54	91.52
$\text{MAGNN}_{\text{avg}}$	59.66	59.78	13.64	15.27	93.13	93.44	79.31	84.30	98.63	98.57
$\text{MAGNN}_{\text{linear}}$	57.80	57.96	9.80	8.49	93.21	93.52	78.95	83.89	98.56	98.48
$\text{MAGNN}_{\text{rot}}$	**60.43**	**60.63**	**15.58**	**16.74**	**93.51**	**93.94**	**80.81**	**85.54**	**98.91**	**98.93**

Node Pair Embedding Visualization

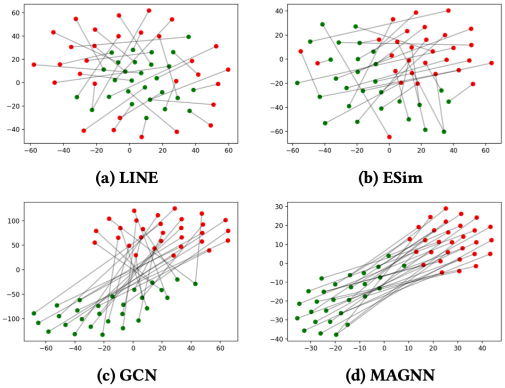


Figure 4: Embedding visualization of node pairs in Last.fm.

References I

[1] Xinyu Fu, Jiani Zhang, Ziqiao Meng, and Irwin King. "MAGNN: Metapath Aggregated Graph Neural Network for Heterogeneous Graph Embedding". In: *Proceedings of The Web Conference 2020*. WWW '20. Taipei, Taiwan: Association for Computing Machinery, 2020, pp. 2331–2341. ISBN: 9781450370233. DOI: 10.1145/3366423.3380297. URL: https://doi.org/10.1145/3366423.3380297.

UNIVERSITY OF WATERLOO | FACULTY OF MATHEMATICS