Diverse Message Passing

Liang Yang

About Me

2000 | 2004 | 2007

2007 | 2009

2009 | 2010

2010 | 2018

2018 | today

Homepage: http://yangliang.github.io/

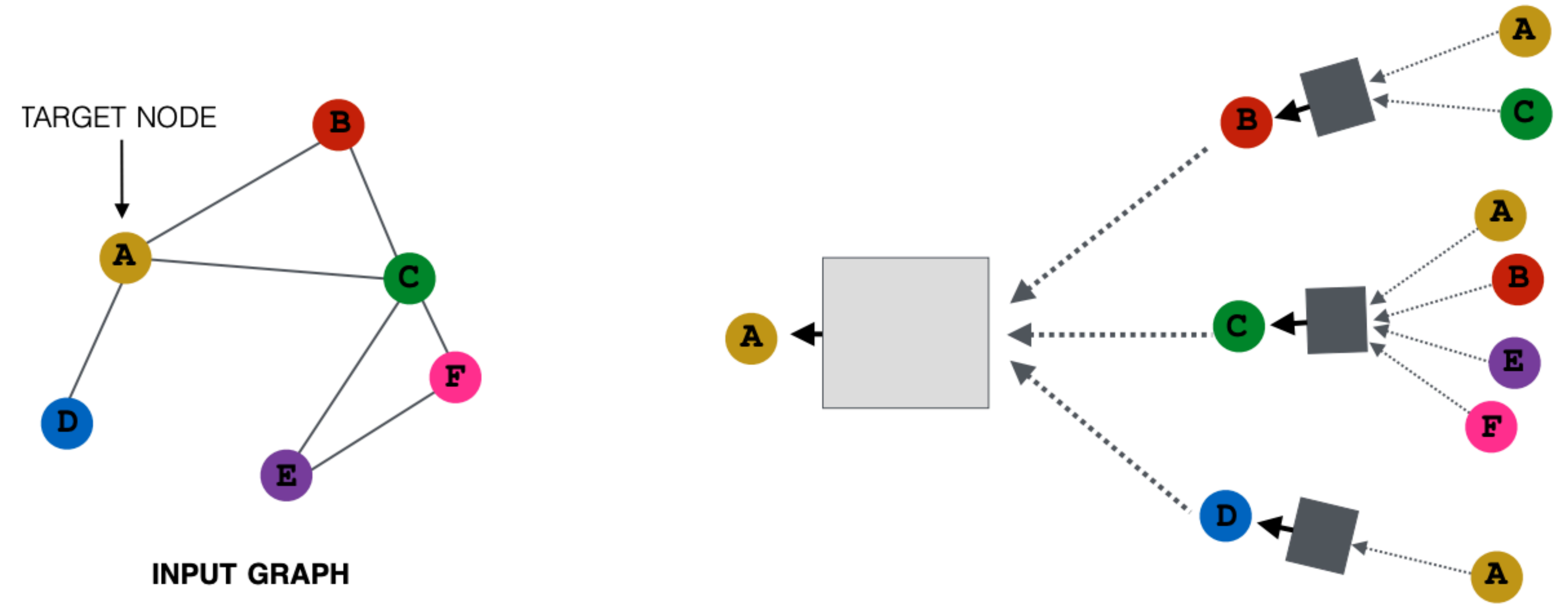
2013 | 2016

Outline

- Existing Message Passing Framework
- Diverse Message Passing
  - Motivations
  - Semi-supervised Task
  - Self-supervised Task
- Theoretical Analysis
- Conclusions

Existing Message Passing

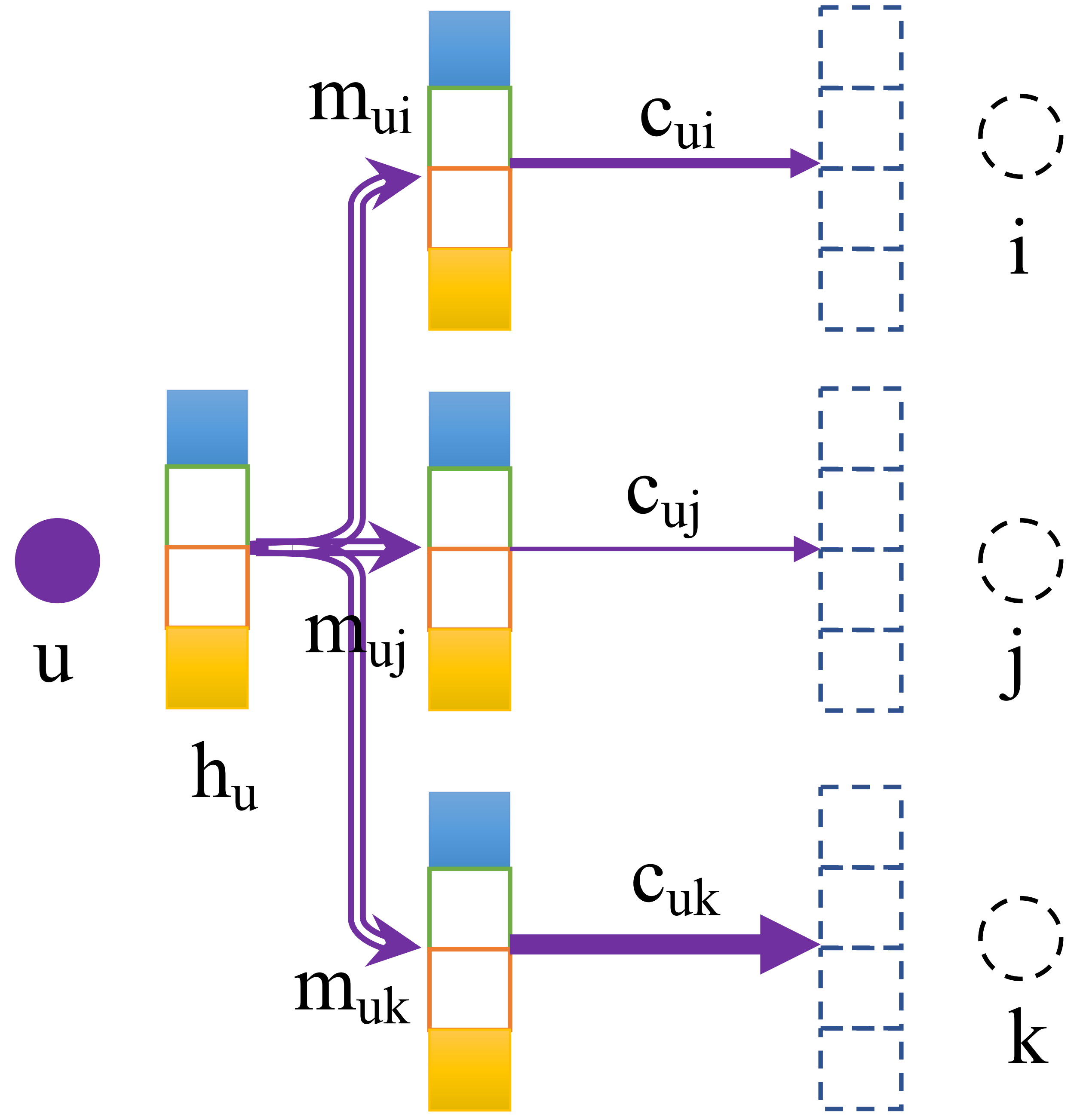
Aggregation-Combination



INPUT GRAPH

$$\bar{\mathbf{h}}_v^k = \mathrm{AGGREGATE}^k\left(\{\mathbf{h}_u^{k-1} | u \in \mathcal{N}(v)\}\right),$$

$$\mathbf{h}_v^k = \mathrm{COMBINATE}^k\left(\mathbf{h}_v^{k-1}, \bar{\mathbf{h}}_v^k\right),$$

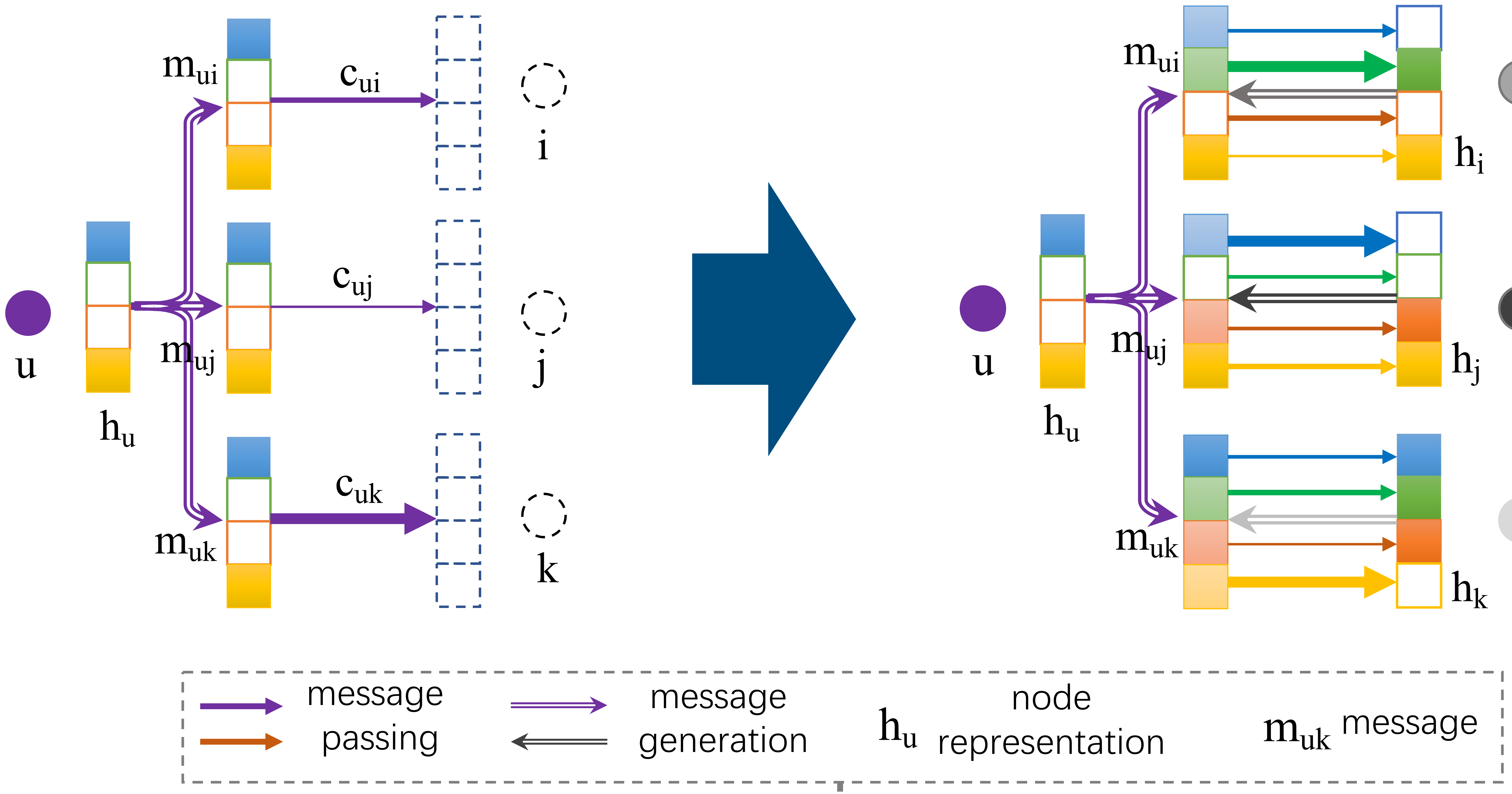


$$\mathbf{h}_v^k = \sigma\left(\left(c_{vv}^k \mathbf{h}_v^{k-1} + \sum_{u \in \mathcal{N}(v)} c_{uv}^k \mathbf{h}_u^{k-1}\right)\mathbf{W}^k\right),$$

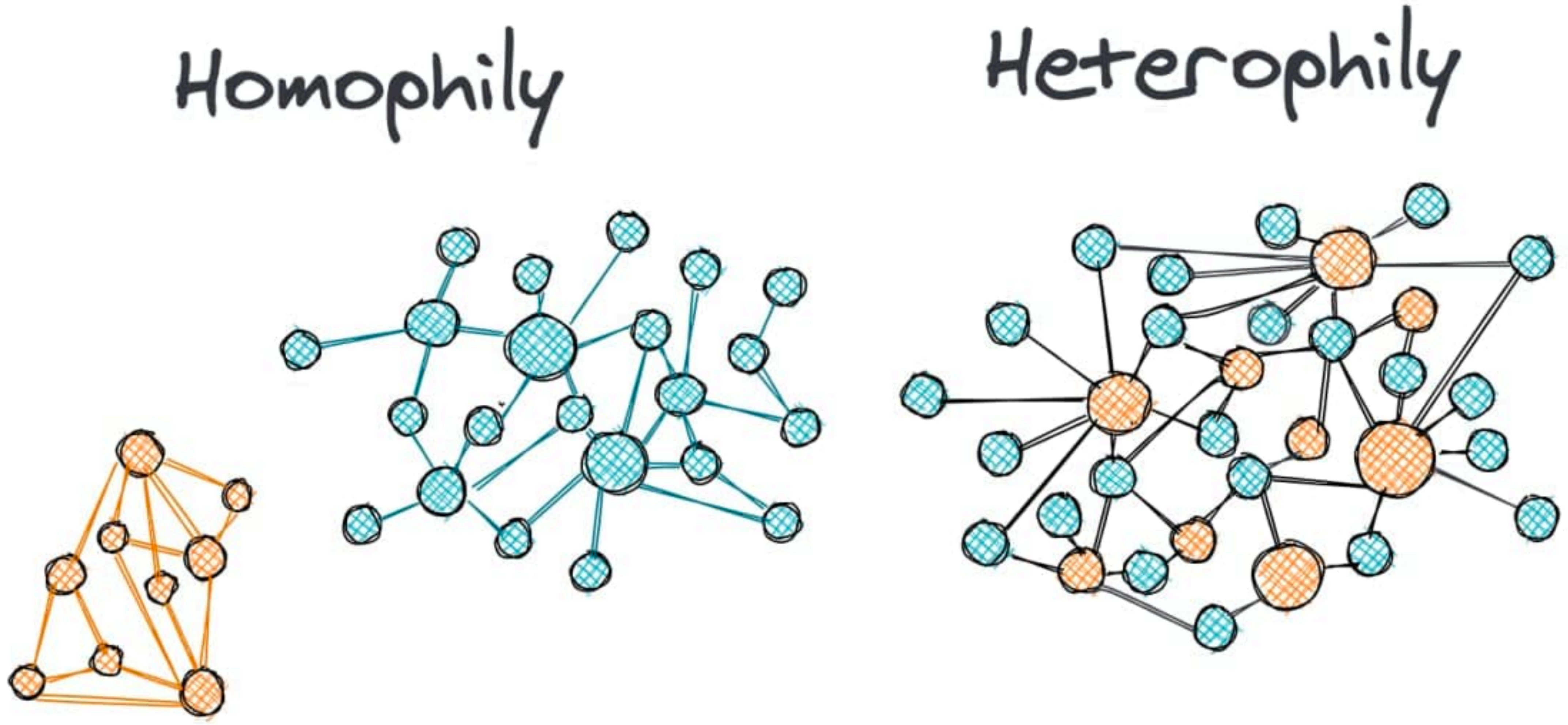
Uniform

In different attribute channels

Diverse Message Passing



Motivations - Attribute Diversity



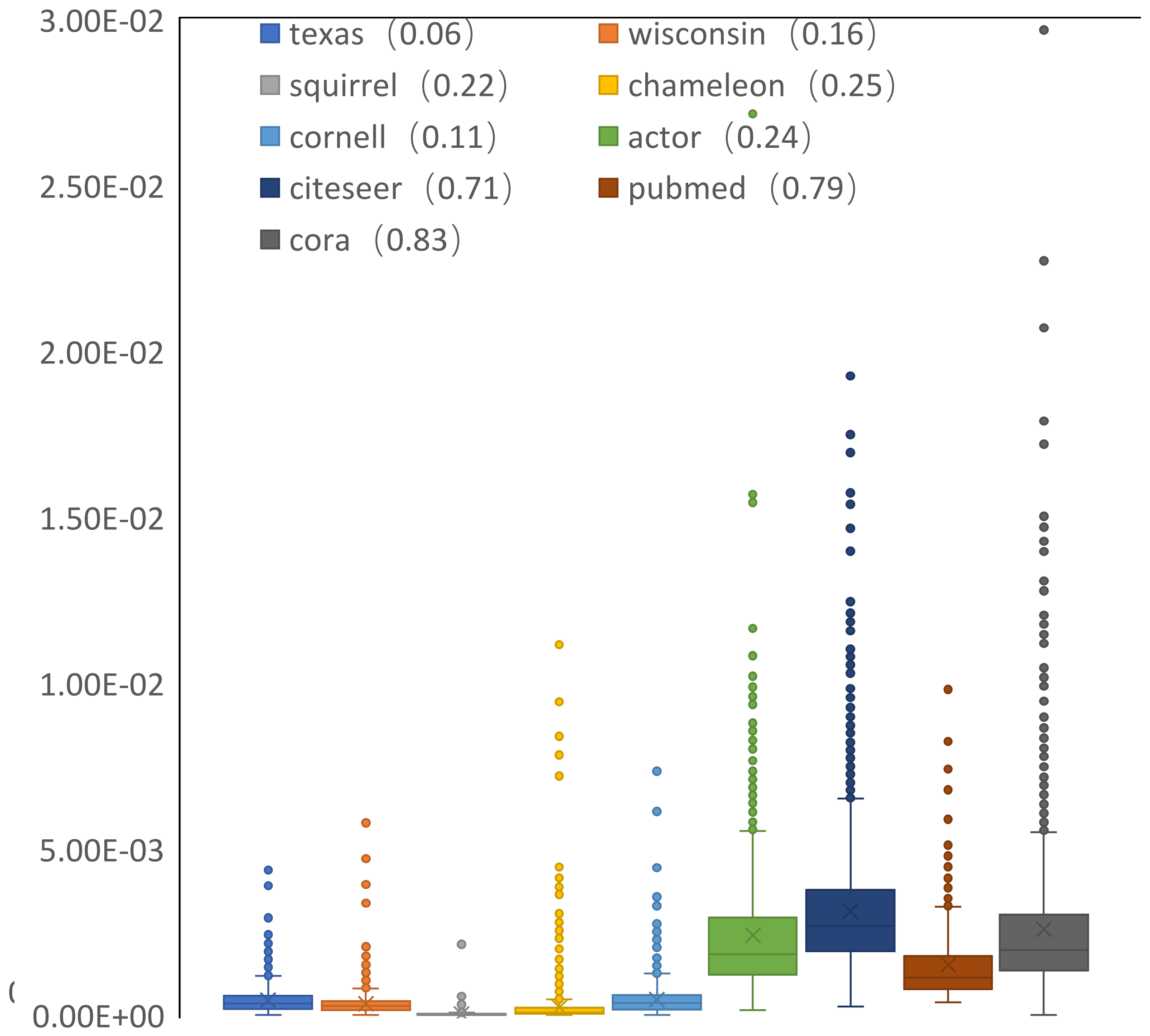
Homophily Rate of Graph

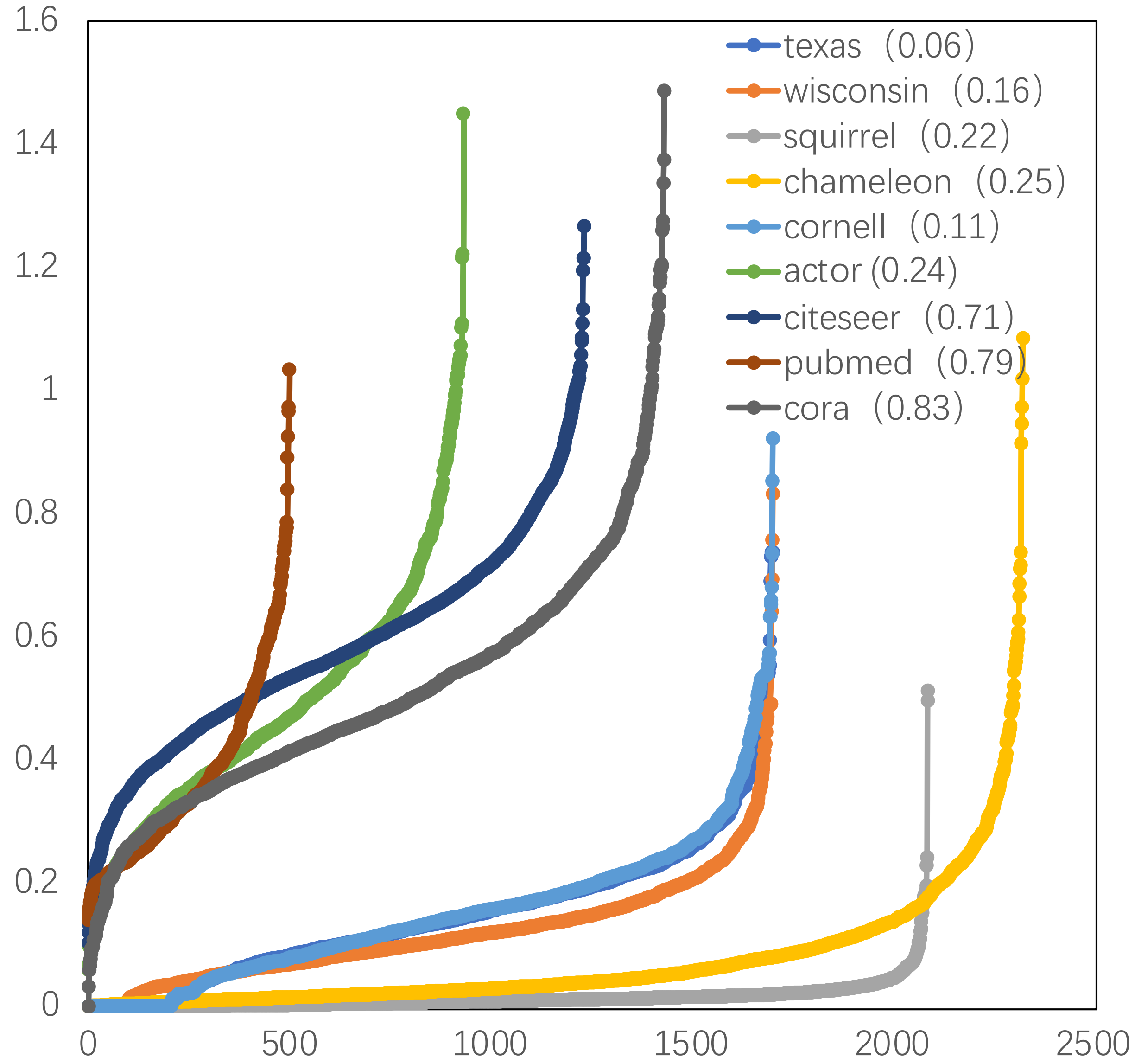
$$\beta = \frac{1}{N}\sum_{v\in V}\frac{\text{Number of } v\text{'s neighbors who have the same label as } v}{\text{Number of } v\text{'s neighbors}}.$$

Topology + Node Label

Graph-level

Attribute As Weak Label





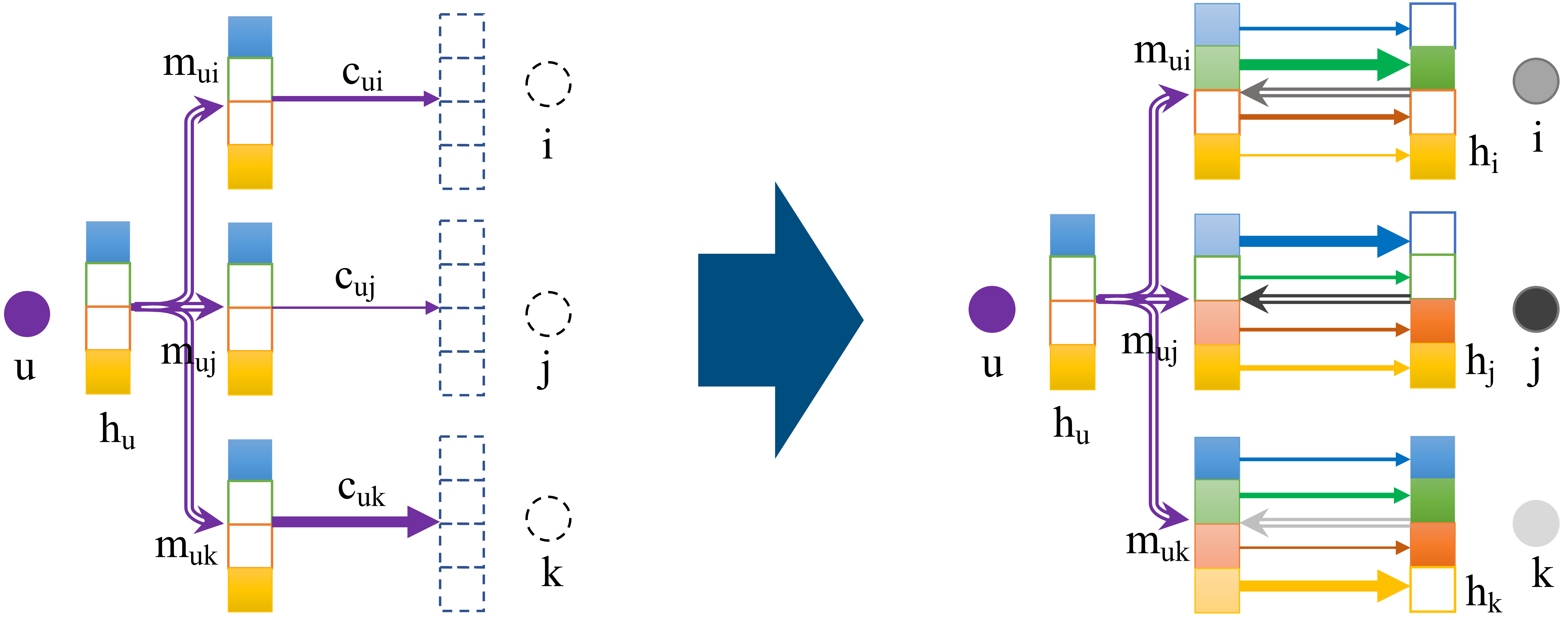
Homophily Rates of Attributes

$$\beta_f = \frac{1}{\sum_{v\in\mathcal{V}} x_{vf}}\sum_{v\in\mathcal{V}}\beta_{vf} = \frac{1}{\sum_{v\in\mathcal{V}} x_{vf}}\sum_{v\in\mathcal{V}}\left(x_{vf}\frac{\sum_{u\in\mathcal{N}(v)} x_{uf}}{d_v}\right),$$

Topology + Node Attribute

Attribute-level

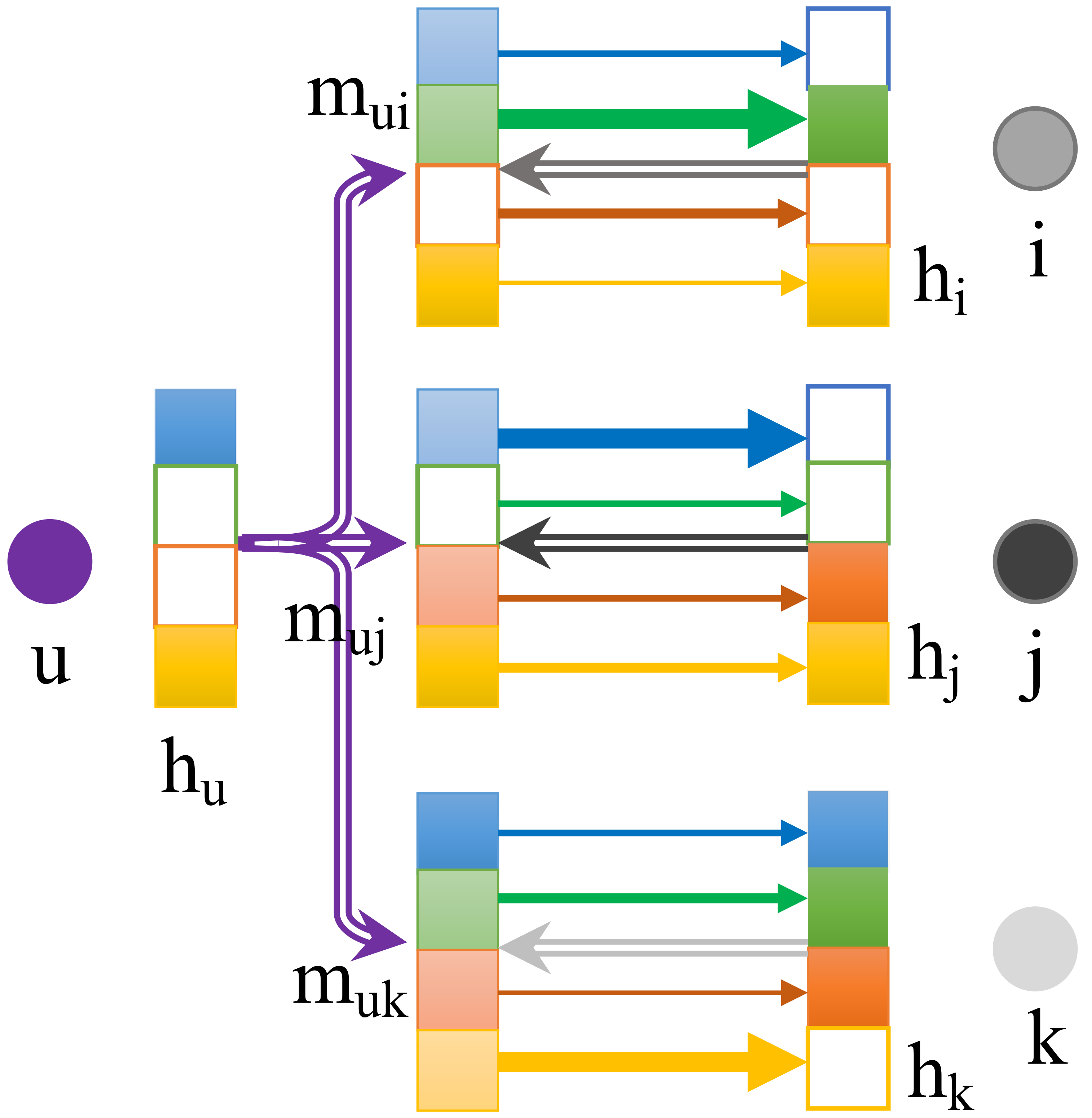
Semi-supervised Diverse Message Passing



$$\mathbf{h}_v^k = \sigma\left(\left(\boxed{c_{vv}^k}\,\mathbf{h}_v^{k-1} + \sum_{u\in\mathcal{N}(v)} \boxed{c_{uv}^k}\,\mathbf{h}_u^{k-1}\right)\mathbf{W}^k\right), \Rightarrow \mathbf{h}_v^k = \sigma\left(\left(\boxed{\mathbf{c}_{vv}^k \odot}\,\mathbf{h}_v^{k-1} + \sum_{u\in\mathcal{N}(v)} \boxed{\mathbf{c}_{uv}^k \odot}\,\mathbf{h}_u^{k-1}\right)\mathbf{W}^k\right)$$

Scalar $\longrightarrow$ **Vector**

Semi-supervised Learning



$$\mathbf{h}_v^k = \sigma\left(\left(\boxed{\mathbf{c}_{vv}^k \odot} \mathbf{h}_v^{k-1} + \sum_{u \in \mathcal{N}(v)} \boxed{\mathbf{c}_{uv}^k \odot} \mathbf{h}_u^{k-1}\right)\mathbf{W}^k\right),$$

Model Complexity O(|E|F)

The first strategy

$$\mathbf{c}_{uv}^k = tanh\left([\mathbf{h}_v^{k-1} || \mathbf{h}_u^{k-1}]\mathbf{W}_c^k\right),$$

Model Complexity O(FxF)

The second strategy

$$\mathbf{c}_{v}^k = tanh\left([\mathbf{h}_v^{k-1} || \bar{\mathbf{h}}_v^{k-1}]\mathbf{W}_c^k\right)$$

$$\boxed{\bar{\mathbf{h}}_v^k = \text{AGGREGATE}^k\left(\{\mathbf{h}_u^{k-1} | u \in \mathcal{N}(v)\}\right),}$$

Model Complexity O(FxF)

Table 2: Mean Classification Accuracy (Bold indicates the best, italics indicates the second best).

Methods	Texas	Wisconsin	Actor	Squirrel	Cham.	Cornell	Citeseer	Pubmed	Cora
GraphSAGE	82.43	81.18	34.23	41.61	58.73	75.95	76.04	88.45	86.90
GCN	64.86	56.86	31.12	32.28	53.51	54.05	75.53	84.71	85.51
GAT	58.38	55.29	26.28	30.62	54.69	58.92	75.46	84.68	82.68
SAGE+JK	83.78	81.96	34.28	40.85	58.11	75.68	76.05	88.34	85.96
Cheby+JK	78.38	82.55	35.14	*45.03*	**63.79**	74.59	74.98	89.07	85.49
GCN+JK	66.49	74.31	34.18	40.45	*63.42*	64.59	74.51	88.41	85.79
GCN-Cheby	77.30	79.41	34.11	43.86	55.24	74.32	75.82	88.72	86.76
MixHop	77.84	75.88	32.22	43.80	60.50	73.51	76.26	85.31	**87.61**
GEOM-GCN	67.57	64.12	31.63	38.14	60.90	60.81	**77.99**	**90.05**	85.27
H2GCN	84.86	*86.67*	**35.86**	36.42	57.11	82.16	*77.04*	*89.40*	*86.92*
DMP-Deg	78.38	80.39	33.09	32.46	54.38	*83.78*	76.87	88.10	86.31
DMP-2-Sum	78.37	84.31	34.93	32.18	55.92	*83.78*	76.27	88.15	85.31
DMP-2-Con	83.78	84.31	34.67	44.28	60.53	*83.78*	75.97	85.31	85.31
DMP-1-Posi	*86.48*	84.31	*35.72*	34.96	51.53	70.27	75.67	88.10	86.11
DMP-1-Sum	*86.48*	86.27	34.21	43.42	50.21	70.27	76.13	88.13	82.28
DMP-1-Con	**89.19**	**92.16**	35.06	**47.26**	62.28	**89.19**	76.43	89.27	86.52

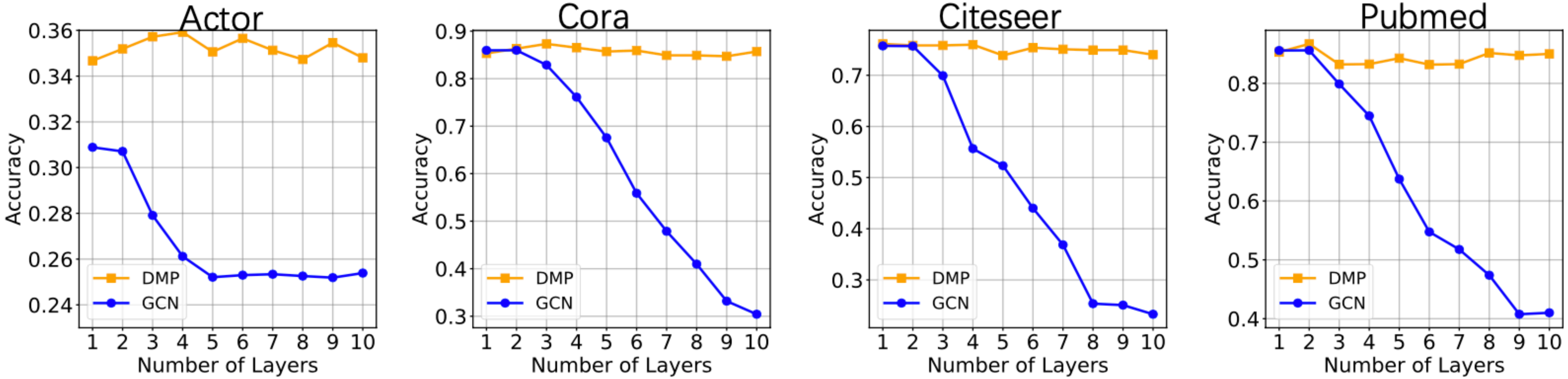


Figure 2: Classification accuracy results with various depths.

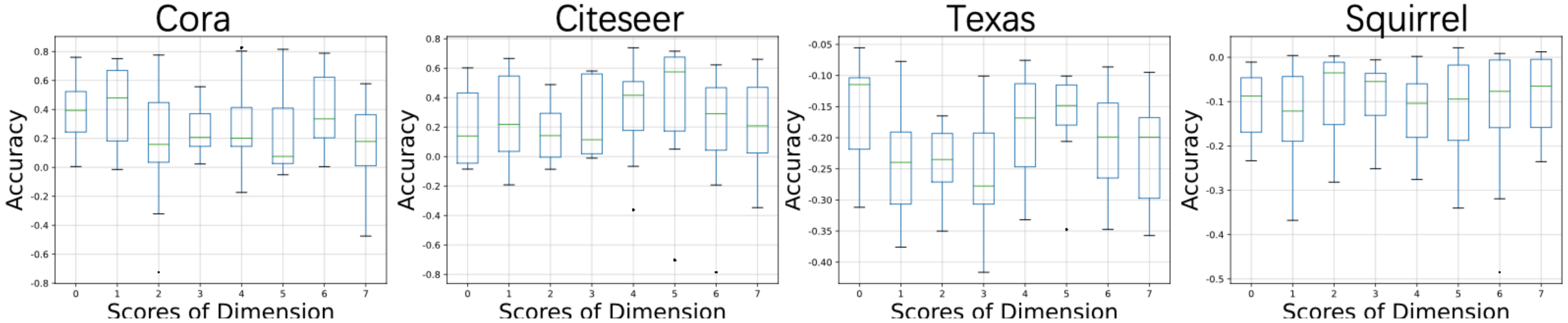
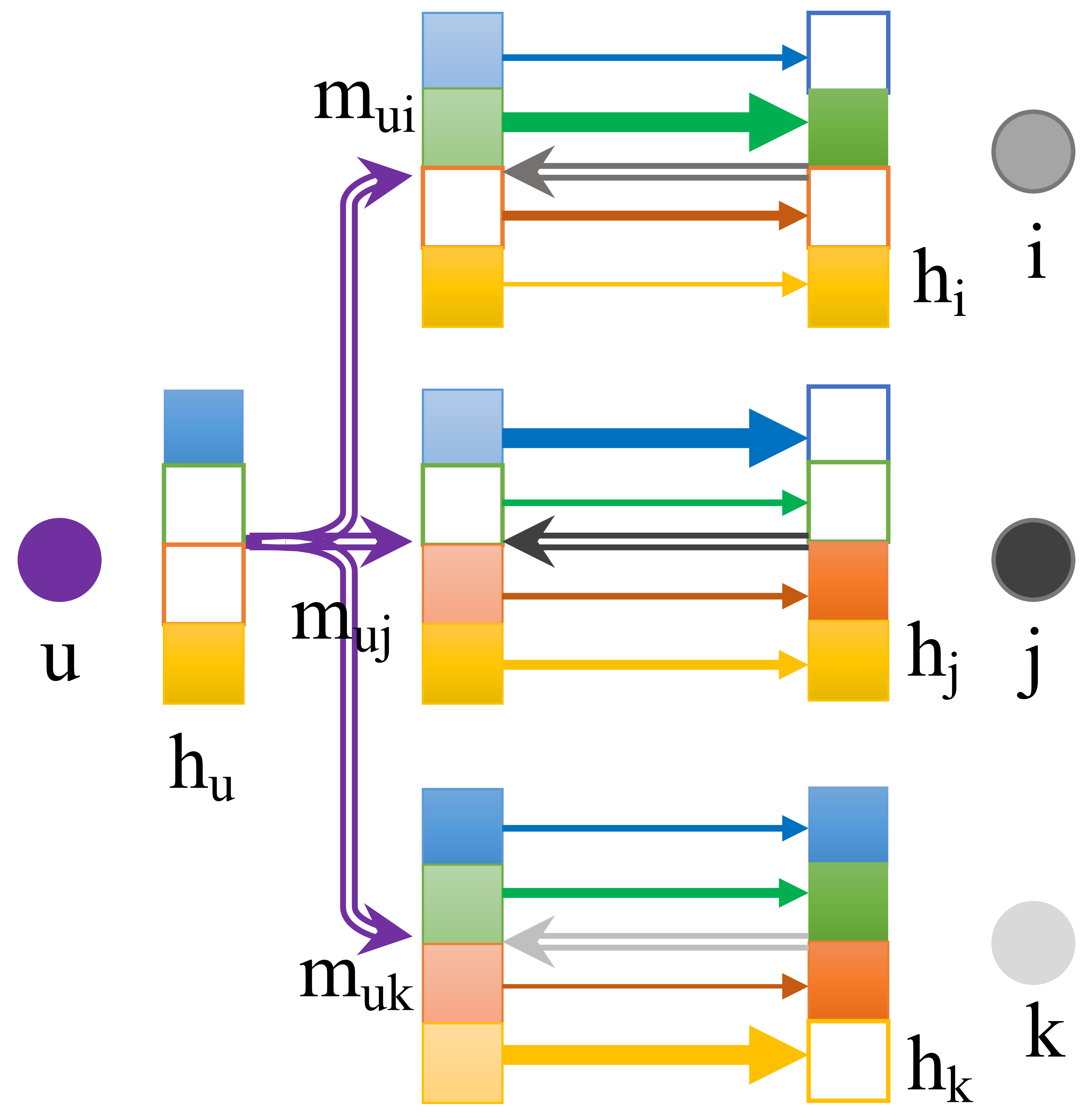


Figure 3: Distributions of learned weights of sampled attribute dimensions.

Self-supervised Diverse Message Passing

Reduce Model Complexity & Preserve Expressive Power



$$\mathbf{h}_v^k = \sigma\left(\left(\boxed{\mathbf{c}_{vv}^k \odot \mathbf{h}_v^{k-1}} + \sum_{u \in \mathcal{N}(v)} \boxed{\mathbf{c}_{uv}^k \odot \mathbf{h}_u^{k-1}}\right)\mathbf{W}^k\right);$$

$$\mathbf{h}_v^k = \sigma\left(\left(\mathbf{m}_{vv}^k + \sum_{u \in \mathcal{N}(v)} \mathbf{m}_{uv}^k\right)\mathbf{W}^k\right)$$

Message

$$\mathbf{m}_{uv}^k = \frac{\mathbf{h}_v^{k-1} \odot \mathbf{h}_u^{k-1}}{< \mathbf{h}_v^{k-1}, \mathbf{h}_u^{k-1} >},$$

Diverse

Interactive

message passing | message generation | h_u node representation | m_{uk} message

Table 3: Node Classification Results in Terms of Accuracy.

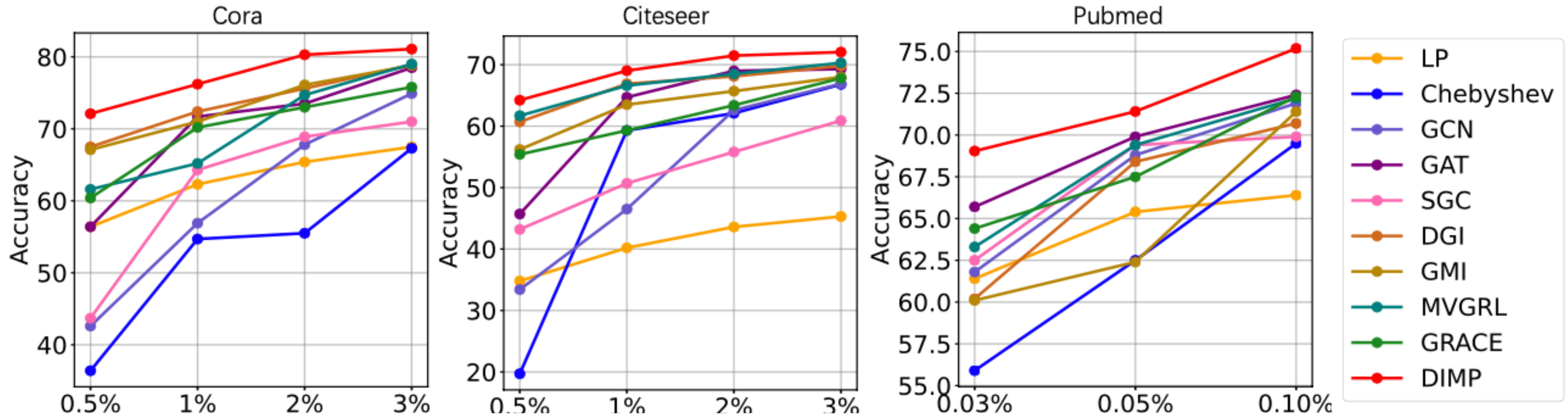
Method	Cora	CiteSeer	PubMed	Amazon-C	Amazon-P	Coauthor-CS	Coauthor-P
MLP	58.2±2.1	59.1±2.3	70.0±2.1	44.9±5.8	69.6±3.8	88.3±0.7	88.9±1.1
LogReg	57.1±2.3	61.0±2.2	64.1±3.1	64.1±5.7	73.0±6.5	86.4±0.9	86.7±1.5
LP	68.0±0.2	45.3±0.2	63.0±0.5	70.8±0.0	67.8±0.0	74.3±0.0	90.2±0.5
Chebyshev	81.2±0.5	69.8±0.5	74.4±0.3	62.6±0.0	74.3±0.0	91.5±0.0	92.1±0.3
GCN	81.5±0.2	70.3±0.3	**79.0±0.4**	76.3±0.5	87.3±1.0	**91.8±0.1**	**92.6±0.7**
GAT	**83.0±0.7**	**72.5±0.7**	**79.0±0.3**	79.3±1.1	86.2±1.5	90.5±0.7	91.3±0.6
MoNet	81.3±1.3	71.2±2.0	78.6±2.3	**83.5±2.2**	**91.2±1.3**	90.8±0.6	92.5±0.9
DGI	81.7±0.6	71.5±0.7	77.3±0.6	75.9±0.6	83.1±0.5	90.0±0.3	91.3±0.4
GMI	80.9±0.7	71.1±0.3	77.9±0.2	76.8±0.1	85.1±0.1	91.0±0.0	OOM
MVGRL	82.9±0.7	72.6±0.7	79.4±0.3	79.0±0.6	87.3±0.3	88.4±0.3	92.6±0.4
GRACE	80.0±0.4	71.7±0.6	79.5±1.1	71.8±0.4	81.8±1.0	90.1±0.8	92.3±0.6
GCA	81.0±0.4	71.9±0.5	80.5±1.1	80.8±0.4	87.1±1.0	91.3±0.4	93.1±0.3
SubG-Con	82.5±0.3	70.8±0.3	73.1±0.5	OOM	OOM	OOM	OOM
DIMP	**83.3±0.5**	**73.3±0.5**	**81.4±0.5**	**83.3±0.4**	**88.7±0.2**	**92.1±0.5**	**94.2±0.4**

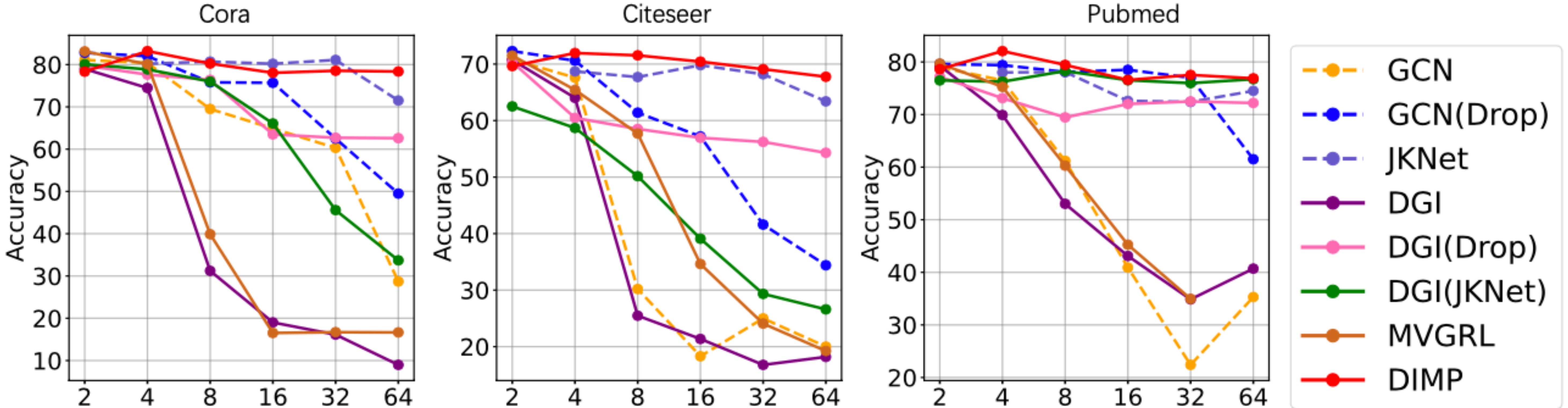
Table 4: Node Clustering Results in Terms of NMI and ARI.

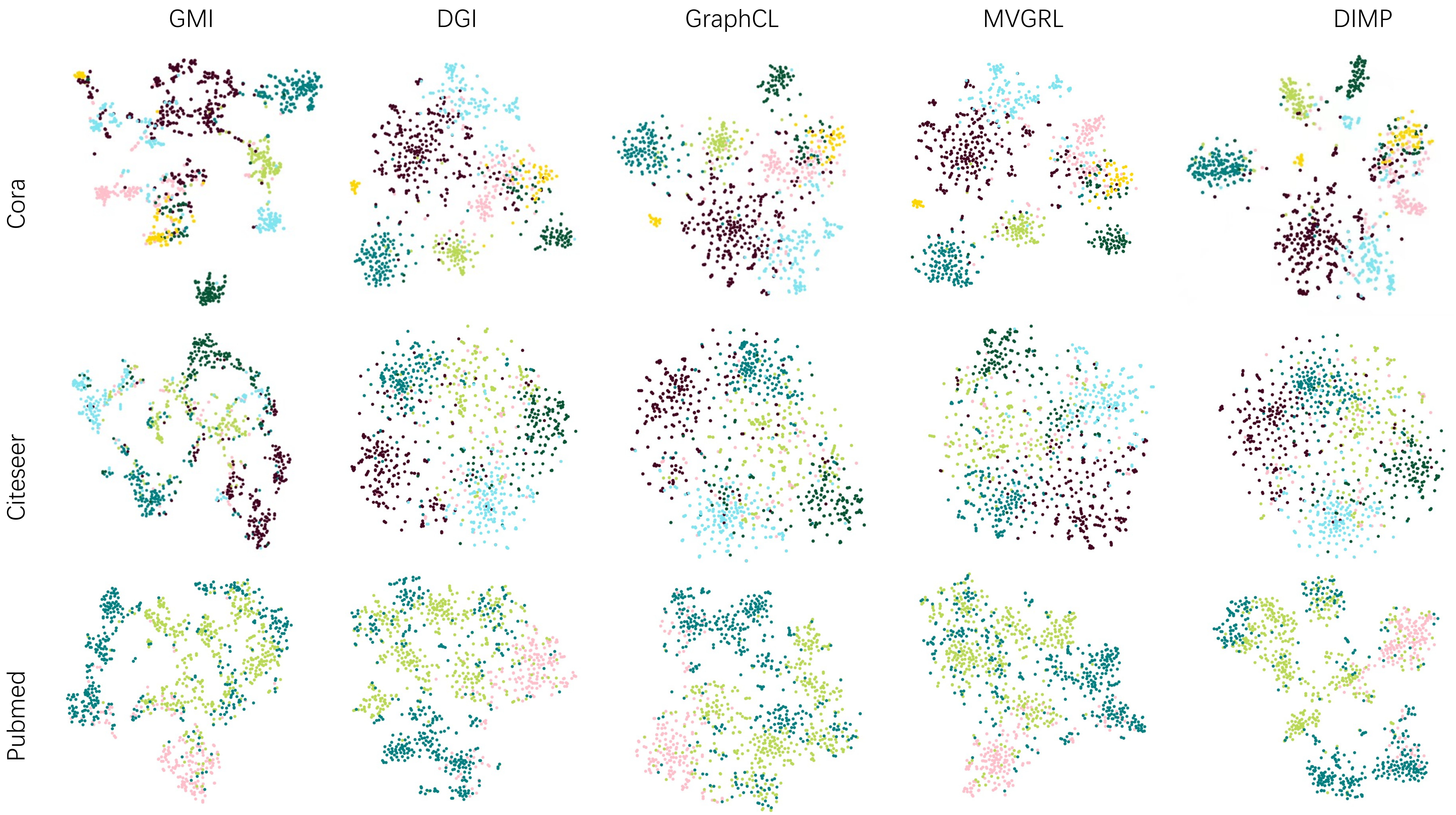
Methods	Cora		Citeseer		Pubmed	
	NMI	ARI	NMI	ARI	NMI	ARI
K-means	0.321	0.230	0.305	0.279	0.001	0.002
Spectral	0.127	0.031	0.056	0.010	0.042	0.002
BigClam	0.007	0.001	0.036	0.007	0.006	0.003
GraphEnc	0.109	0.006	0.033	0.010	0.209	0.184
DeepWalk	0.327	0.243	0.088	0.092	0.279	0.299
GAE	0.429	0.347	0.176	0.124	0.277	0.279
VGAE	0.436	0.346	0.156	0.093	0.229	0.213
MGAE	0.511	0.445	0.412	0.414	0.282	0.248
ARGA	0.449	0.352	0.350	0.341	0.276	0.291
ARVGA	0.450	0.374	0.261	0.245	0.117	0.078
GALA	0.577	0.511	0.441	0.446	0.327	0.321
MVGRL	0.572	0.495	0.469	0.449	0.322	0.296
DIMP	**0.581**	**0.522**	**0.471**	**0.471**	**0.346**	**0.328**

Table 5: Graph Classification Results in Terms of Accuracy.

METHOD	MUTAG	PTC-MR	IMDB-B	IMDB-M	RDT-B
GraphSage	85.1±7.6	63.9±7.7	72.3±5.3	50.9±2.2	-
GCN	85.6±5.8	64.2±4.3	74.0±3.4	51.9±3.8	50.0±0.0
GIN	**89.4±5.6**	64.6±7.0	**75.1±5.1**	**52.3±2.8**	**92.4±2.5**
GAT	89.4±6.1	**66.7±5.1**	70.5±2.3	47.8±3.1	85.2±3.3
DeepWalk	83.7±1.5	57.9±1.3	50.7±0.3	34.7±0.2	-
node2vec	72.6±10.	58.6±8.0	-	-	-
sub2vec	61.1±15.	60.0±6.4	55.3±1.5	36.7±0.8	71.5±0.4
graph2vec	83.2±9.6	60.2±6.9	71.1±0.5	50.4±0.9	75.8±1.0
Infograph	89.0±1.1	61.7±1.4	73.0±0.9	49.7±0.5	82.5±1.4
MVGRL	89.7±1.1	62.5±1.7	74.2±0.7	51.2±0.5	84.5±0.6
GraphCL	86.8±1.3	-	71.1±0.4	-	89.5±0.8
DIMP	**91.5±1.1**	**64.2±1.2**	**74.8±0.8**	**52.0±0.6**	**91.9±0.6**






GMI
DGI
GraphCL
MVGRL
DIMP
Cora
Citeseer
Pubmed

Theoretical Analysis

Diverse Message Passing can prevent over-smoothing issue

Semi-supervised Task

$$\mathbf{h}_v^k = \sigma\left(\left(\boxed{\mathbf{c}_{vv}^k \odot} \mathbf{h}_v^{k-1} + \sum_{u \in \mathcal{N}(v)} \boxed{\mathbf{c}_{uv}^k \odot} \mathbf{h}_u^{k-1}\right)\mathbf{W}^k\right),$$

The connection between learned propagation weights and graph partition

Self-supervised Task

$$\mathbf{h}_v^k = \sigma\left(\left(\boxed{\mathbf{m}_{vv}^k} + \sum_{u \in \mathcal{N}(v)} \boxed{\mathbf{m}_{uv}^k}\right)\mathbf{W}^k\right) \quad \mathbf{m}_{uv}^k = \frac{\mathbf{h}_v^{k-1} \odot \mathbf{h}_u^{k-1}}{< \mathbf{h}_v^{k-1}, \mathbf{h}_u^{k-1} >},$$

The connection between inner-product message and community detection

Theoretical Analysis

Diverse Message Passing can prevent over-smoothing issue

Semi-supervised Task

$$\mathbf{h}_v^k = \sigma\left(\left(\boxed{\mathbf{c}_{vv}^k} \odot \mathbf{h}_v^{k-1} + \sum_{u \in \mathcal{N}(v)} \boxed{\mathbf{c}_{uv}^k} \odot \mathbf{h}_u^{k-1}\right)\mathbf{W}^k\right),$$

The connection between learned propagation weights and graph partition

Self-supervised Task

$$\mathbf{h}_v^k = \sigma\left(\left(\boxed{\mathbf{m}_{vv}^k} + \sum_{u \in \mathcal{N}(v)} \boxed{\mathbf{m}_{uv}^k}\right)\mathbf{W}^k\right) \quad \mathbf{m}_{uv}^k = \frac{\mathbf{h}_v^{k-1} \odot \mathbf{h}_u^{k-1}}{< \mathbf{h}_v^{k-1}, \mathbf{h}_u^{k-1} >},$$

The connection between inner-product message and community detection

Learned propagation weights vs. Graph partition

Theorem 1. *The Uniform Message Passing in Eq. (2) with learnable weights c_{uv} is the gradient descent algorithm of the following objective function with node attribute $\mathbf{X}$ being the initialization of $\mathbf{H}$.*

$$\min_{\mathbf{C},\mathbf{H}} \sum_{u,v} \left(b_{uv}c_{uv} + \gamma c_{uv}^2\right) + 2tr(\mathbf{H}^T \mathbf{L}_C \mathbf{H}), \tag{9}$$

$$s.t.\ \forall u \sum c_{uv} = 1,\ 0 \le c_{uv} \le 1,\ \mathbf{H} \in \mathbf{R}^{N\times F}, \tag{10}$$

where $b_{uv} = g\left(a_{uv}, dis(\mathbf{x}_i, \mathbf{x}_j)\right)$ denotes the similarity between nodes u and v, according to both the topology a_{uv} and the distance between attributes $dis(\mathbf{x}_i, \mathbf{x}_j)$. $\mathbf{A} = [a_{uv}]$ is the adjacency matrix of $\mathcal{G}$. $\mathbf{C}$ represents the collection of c_{uv}, i.e., the adjacency matrix of the learned graph. $\mathbf{L}_C$ stands for the Laplacian matrix of the adjacency matrix $\mathbf{C}$.

Theorem 2. *[Ky Fan's Theorem [30]] There exists*

$$\min_{\mathbf{H}\in\mathbf{R}^{N\times F}, \mathbf{H}^T\mathbf{H}=\mathbf{I}} tr(\mathbf{H}^T \mathbf{L}_C \mathbf{H}) = \sum_{f=1}^{F} \sigma_f(\mathbf{L}_C), \tag{11}$$

where $\sigma_f(\mathbf{L}_C)$ denotes the f^{th} smallest eigenvalue of the Laplacian matrix $\mathbf{L}_C$.

Theorem 3. *[[31, 32]] The multiplicity F of the eigenvalue 0 of the Laplacian matrix $\mathbf{L}_C$ equals to the number of connected components in the graph, whose similarity matrix is $\mathbf{C}$.*

Learned propagation weights vs. Graph partition

$$\mathbf{h}_v^k = \sigma\left(\left(c_{vv}^k \mathbf{h}_v^{k-1} + \sum_{u \in \mathcal{N}(v)} c_{uv}^k \mathbf{h}_u^{k-1}\right)\mathbf{W}^k\right),$$

Theorem 1. *The Uniform Message Passing in Eq. (2) with learnable weights c_{uv} is the gradient descent algorithm of the following objective function with node attribute $\mathbf{X}$ being the initialization of $\mathbf{H}$.*

$$\min_{\mathbf{C},\mathbf{H}} \sum_{u,v} \left(b_{uv} c_{uv} + \gamma c_{uv}^2\right) + 2tr(\mathbf{H}^T \mathbf{L}_C \mathbf{H}), \tag{9}$$

$$s.t.\ \forall u \sum c_{uv} = 1,\ 0 \leq c_{uv} \leq 1,\ \mathbf{H} \in \mathbf{R}^{N \times F}, \tag{10}$$

where $b_{uv} = g\left(a_{uv}, dis(\mathbf{x}_i, \mathbf{x}_j)\right)$ denotes the similarity between nodes u and v, according to both the topology a_{uv} and the distance between attributes $dis(\mathbf{x}_i, \mathbf{x}_j)$. $\mathbf{A} = [a_{uv}]$ is the adjacency matrix of $\mathcal{G}$. $\mathbf{C}$ represents the collection of c_{uv}, i.e., the adjacency matrix of the learned graph. $\mathbf{L}_C$ stands for the Laplacian matrix of the adjacency matrix $\mathbf{C}$.

Theorem 2. *[Ky Fan's Theorem [30]] There exists*

$$\min_{\mathbf{H} \in \mathbf{R}^{N \times F}, \mathbf{H}^T\mathbf{H} = \mathbf{I}} tr(\mathbf{H}^T \mathbf{L}_C \mathbf{H}) = \sum_{f=1}^{F} \sigma_f(\mathbf{L}_C), \tag{11}$$

where $\sigma_f(\mathbf{L}_C)$ denotes the f^{th} smallest eigenvalue of the Laplacian matrix $\mathbf{L}_C$.

Theorem 3. *[[31, 32]] The multiplicity F of the eigenvalue 0 of the Laplacian matrix $\mathbf{L}_C$ equals to the number of connected components in the graph, whose similarity matrix is $\mathbf{C}$.*

Learned propagation weights vs. Graph partition

$$\mathbf{h}_v^k = \sigma\left(\left(c_{vv}^k \mathbf{h}_v^{k-1} + \sum_{u \in \mathcal{N}(v)} c_{uv}^k \mathbf{h}_u^{k-1}\right)\mathbf{W}^k\right),$$

Theorem 1. *The Uniform Message Passing in Eq. (2) with learnable weights c_{uv} is the gradient descent algorithm of the following objective function with node attribute $\mathbf{X}$ being the initialization of $\mathbf{H}$.*

$$\min_{\mathbf{C},\mathbf{H}} \sum_{u,v} \left(b_{uv}c_{uv} + \gamma c_{uv}^2\right) + 2tr(\mathbf{H}^T\mathbf{L}_C\mathbf{H}), \tag{9}$$

$$s.t.\ \forall u \sum c_{uv} = 1,\ 0 \le c_{uv} \le 1,\ \mathbf{H} \in \mathbf{R}^{N \times F}, \tag{10}$$

where $b_{uv} = g\left(a_{uv}, dis(\mathbf{x}_i, \mathbf{x}_j)\right)$ denotes the similarity between nodes u and v, according to both the topology a_{uv} and the distance between attributes $dis(\mathbf{x}_i, \mathbf{x}_j)$. $\mathbf{A} = [a_{uv}]$ is the adjacency matrix of $\mathcal{G}$. $\mathbf{C}$ represents the collection of c_{uv}, i.e., the adjacency matrix of the learned graph. $\mathbf{L}_C$ stands for the Laplacian matrix of the adjacency matrix $\mathbf{C}$.

Theorem 2. *[Ky Fan's Theorem [30]] There exists*

$$\min_{\mathbf{H}\in\mathbf{R}^{N\times F},\mathbf{H}^T\mathbf{H}=\mathbf{I}} tr(\mathbf{H}^T\mathbf{L}_C\mathbf{H}) = \sum_{f=1}^{F} \sigma_f(\mathbf{L}_C), \tag{11}$$

where $\sigma_f(\mathbf{L}_C)$ denotes the f^{th} smallest eigenvalue of the Laplacian matrix $\mathbf{L}_C$.

Theorem 3. *[[31, 32]] The multiplicity F of the eigenvalue 0 of the Laplacian matrix $\mathbf{L}_C$ equals to the number of connected components in the graph, whose similarity matrix is $\mathbf{C}$.*

Learned propagation weights vs. Graph partition

Theorem 1. *The Uniform Message Passing in Eq. (2) with learnable weights c_{uv} is the gradient descent algorithm of the following objective function with node attribute $\mathbf{X}$ being the initialization of* **H**.

$$\mathbf{h}_v^k = \sigma\left(\left(c_{vv}^k \mathbf{h}_v^{k-1} + \sum_{u \in \mathcal{N}(v)} c_{uv}^k \mathbf{h}_u^{k-1}\right)\mathbf{W}^k\right),$$

$$\min_{\mathbf{C},\mathbf{H}} \sum_{u,v} \left(b_{uv}c_{uv} + \gamma c_{uv}^2\right) + 2tr(\mathbf{H}^T\mathbf{L}_C\mathbf{H}), \tag{9}$$

$$s.t.\ \forall u \sum c_{uv} = 1,\ 0 \le c_{uv} \le 1,\ \mathbf{H} \in \mathbf{R}^{N\times F}, \tag{10}$$

Learned propagation weights

where $b_{uv} = g\left(a_{uv}, dis(\mathbf{x}_i, \mathbf{x}_j)\right)$ denotes the similarity between nodes u and v, according to both the topology a_{uv} and the distance between attributes $dis(\mathbf{x}_i, \mathbf{x}_j)$. $\mathbf{A} = [a_{uv}]$ is the adjacency matrix of $\mathcal{G}$. $\mathbf{C}$ represents the collection of c_{uv}, i.e., the adjacency matrix of the learned graph. $\mathbf{L}_C$ stands for the Laplacian matrix of the adjacency matrix $\mathbf{C}$.

Theorem 2. *[Ky Fan's Theorem [30]] There exists*

$$\min_{\mathbf{H}\in\mathbf{R}^{N\times F}, \mathbf{H}^T\mathbf{H}=\mathbf{I}} tr(\mathbf{H}^T\mathbf{L}_C\mathbf{H}) = \sum_{f=1}^{F} \sigma_f(\mathbf{L}_C), \tag{11}$$

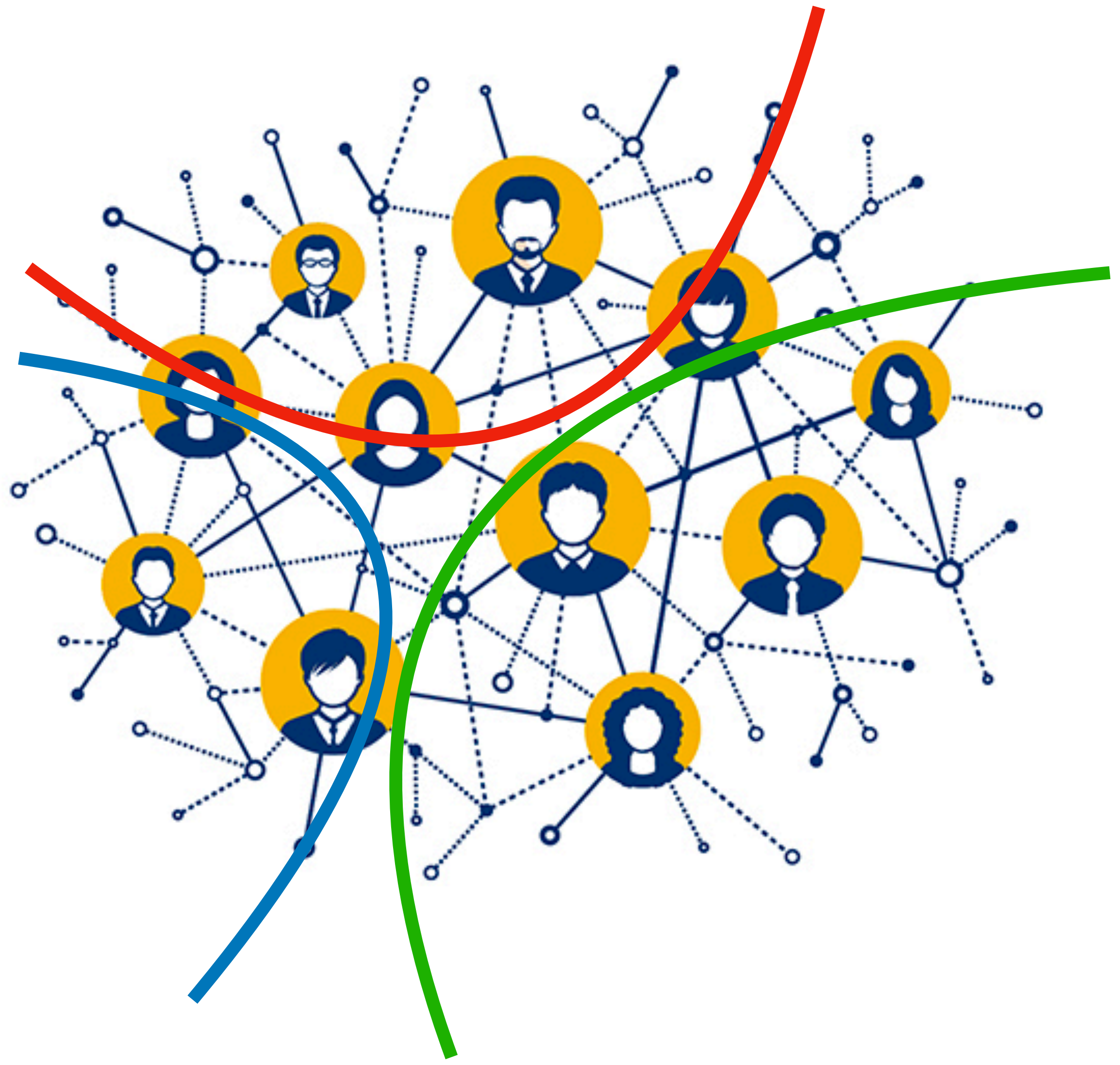
VS.

where $\sigma_f(\mathbf{L}_C)$ denotes the f^{th} smallest eigenvalue of the Laplacian matrix $\mathbf{L}_C$.

Theorem 3. *[[31, 32]] The multiplicity F of the eigenvalue 0 of the Laplacian matrix $\mathbf{L}_C$ equals to the number of connected components in the graph, whose similarity matrix is $\mathbf{C}$.*

Graph partition

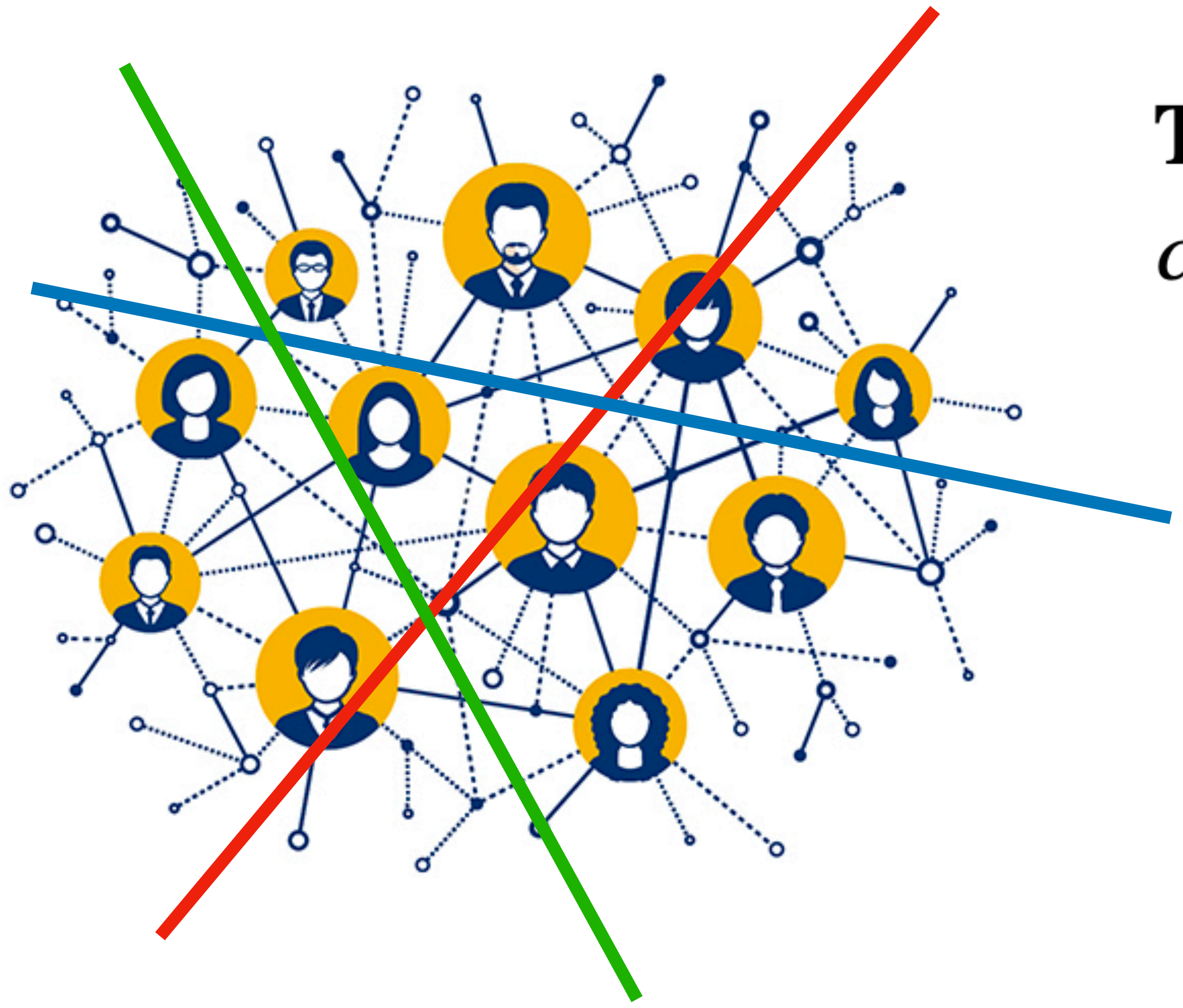
Learned propagation weights vs. Graph partition



Theorem 4. *The Uniform Message Passing in Eq. (2) actually partitions graph into F connected components based on the similarity* $b_{uv} = g\left(a_{uv}, dis(\mathbf{x}_i, \mathbf{x}_j)\right)$ *via*

$$\min_{\mathbf{C}} \sum_{u,v} \left(b_{uv}c_{uv} + \gamma c_{uv}^2\right) \tag{12}$$

$$s.t.\ \forall u \sum c_{uv} = 1,\ 0 \leq c_{uv} \leq 1,\ rank(\mathbf{L}_C) = N - F. \tag{13}$$



Theorem 5. *The Diverse Message Passing in Eq. (4) actually partitions graph into 2 connected components (F groups) based on each similarity* $b_{uv}^{(f)} = g\left(a_{uv}, dis(x_{if}, x_{jf})\right)$ *via*

$$\min_{\mathbf{C}^{(f)}} \sum_{u,v} \left(b_{uv}^{(f)}c_{uv}^{(f)} + \gamma(c_{uv}^{(f)})^2\right), f = 1, ..., F. \tag{14}$$

$$s.t.\ \forall u \sum c_{uv}^{(f)} = 1,\ 0 \leq c_{uv}^{(f)} \leq 1,\ rank(\mathbf{L}_C^{(f)}) = N - 2. \tag{15}$$

Theoretical Analysis

Diverse Message Passing can prevent over-smoothing issue

Semi-supervised Task

$$\mathbf{h}_v^k = \sigma\left(\left(\boxed{\mathbf{c}_{vv}^k} \odot \mathbf{h}_v^{k-1} + \sum_{u \in \mathcal{N}(v)} \boxed{\mathbf{c}_{uv}^k} \odot \mathbf{h}_u^{k-1}\right)\mathbf{W}^k\right),$$

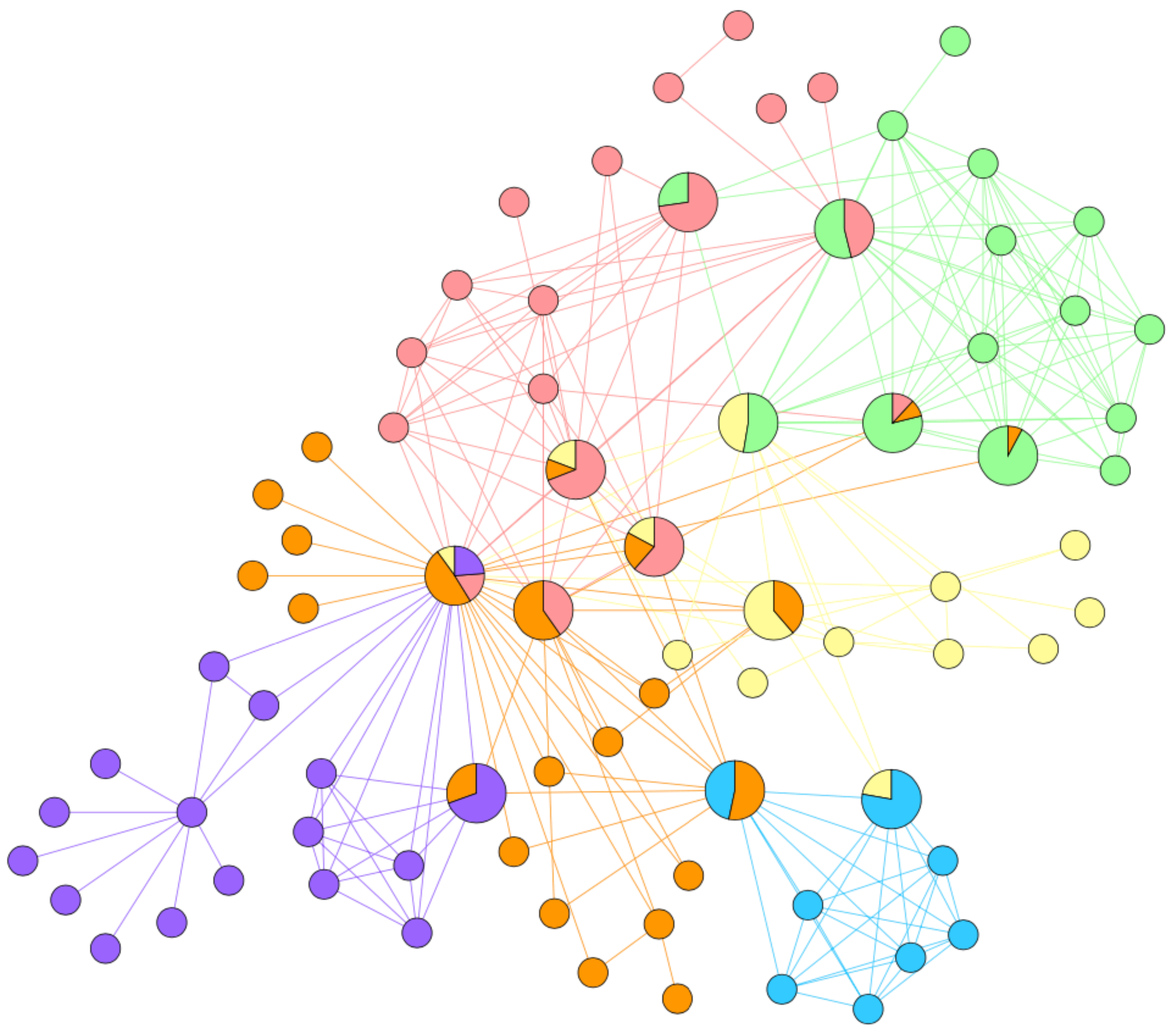
The connection between learned propagation weights and graph partition

Self-supervised Task

$$\mathbf{h}_v^k = \sigma\left(\left(\boxed{\mathbf{m}_{vv}^k} + \sum_{u \in \mathcal{N}(v)} \boxed{\mathbf{m}_{uv}^k}\right)\mathbf{W}^k\right) \quad \mathbf{m}_{uv}^k = \frac{\mathbf{h}_v^{k-1} \odot \mathbf{h}_u^{k-1}}{< \mathbf{h}_v^{k-1}, \mathbf{h}_u^{k-1} >},$$

The connection between inner-product message and community detection

Inner-product message vs Community detection



Theorem 1. *The diverse and interactive message passing in Eqs. (7) and (8) is equivalent to the expectation-maximization (EM) algorithm to maximize the likelihood of generating graph from community structure* $\{\mathbf{h}_u\}_{u=1}^N$ *via Poisson distribution in (Ball, Karrer, and Newman 2011) as follows*

$$P\left(\mathcal{G}\middle|\{\mathbf{h}_u\}_{u=1}^N\right) = \prod_{u<v} \frac{\left(\mathbf{h}_u^T\mathbf{h}_v\right)^{a_{uv}}}{a_{uv}!} \exp\left(-\mathbf{h}_u^T\mathbf{h}_v\right) \times \prod_{u} \frac{\left(\frac{1}{2}\mathbf{h}_u^T\mathbf{h}_u\right)^{a_{uu}/2}}{(a_{uu}/2)!} \exp\left(-\frac{1}{2}\mathbf{h}_u^T\mathbf{h}_u\right). \tag{13}$$

$$\log P\left(\mathcal{G}|\{\mathbf{h}_u\}\right) \geq \sum_{uvz}\left[a_{uv}q_{uv}(z)\log\frac{h_{uz}h_{vz}}{q_{uv}(z)} - h_{uz}h_{vz}\right];$$

Message

$$\mathbf{m}_{uv}^k = \frac{\mathbf{h}_v^{k-1} \odot \mathbf{h}_u^{k-1}}{< \mathbf{h}_v^{k-1}, \mathbf{h}_u^{k-1} >},$$

$$q_{uv}(z) = \frac{h_{uz}h_{vz}}{\sum_z h_{uz}h_{vz}} = \frac{h_{uz}h_{vz}}{< \mathbf{h}_u, \mathbf{h}_v >}.$$

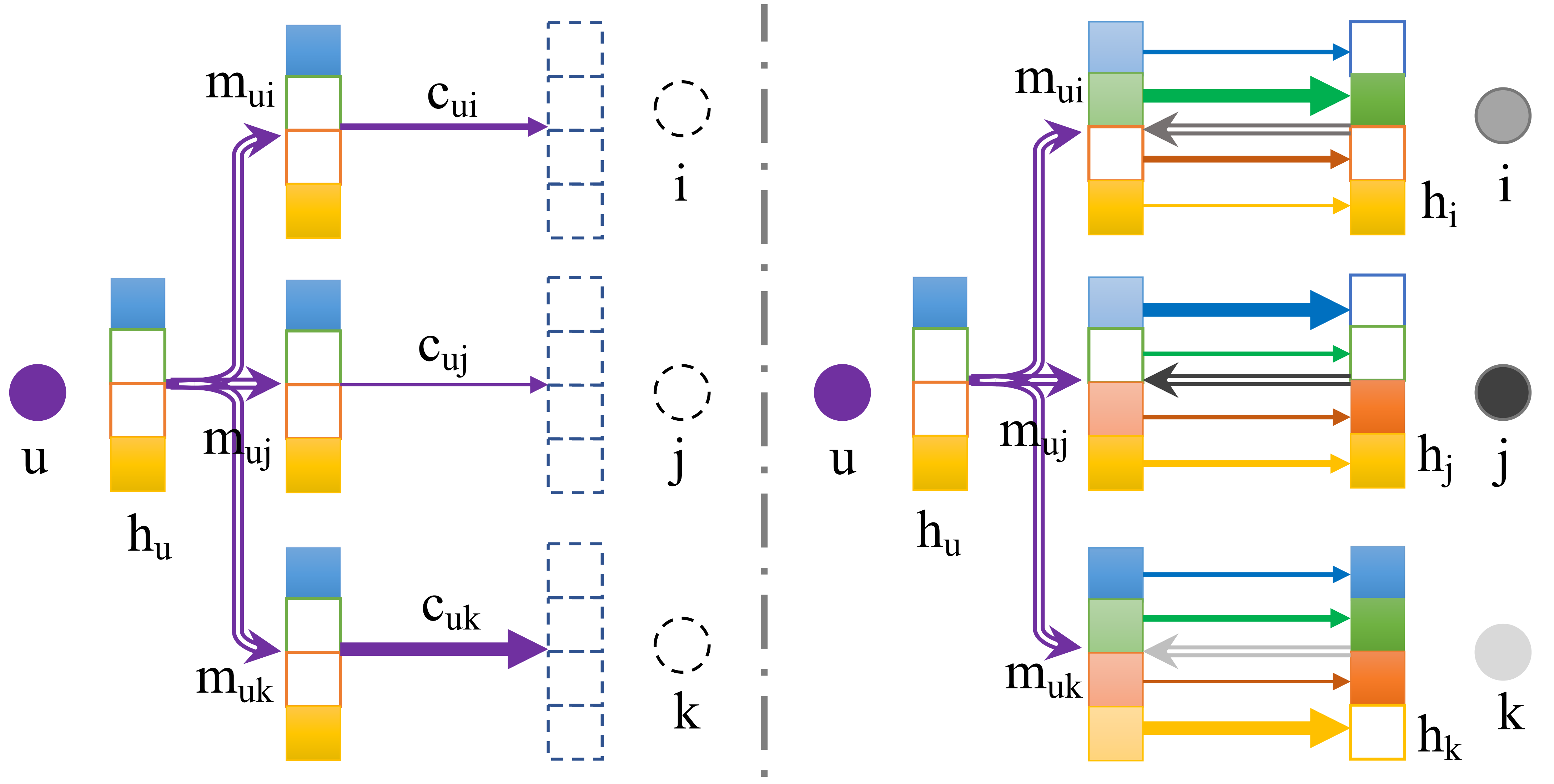
EM

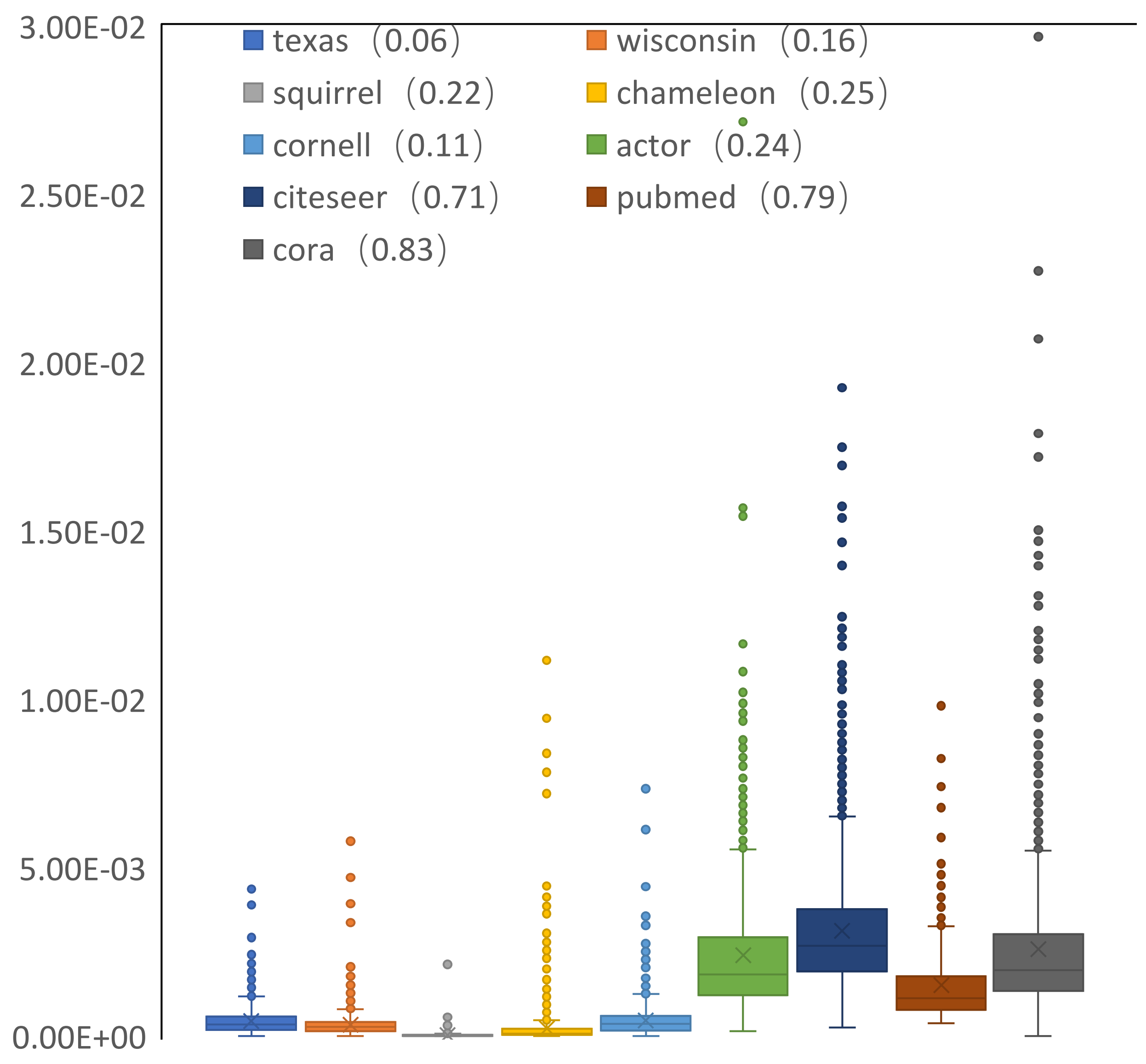
Passing

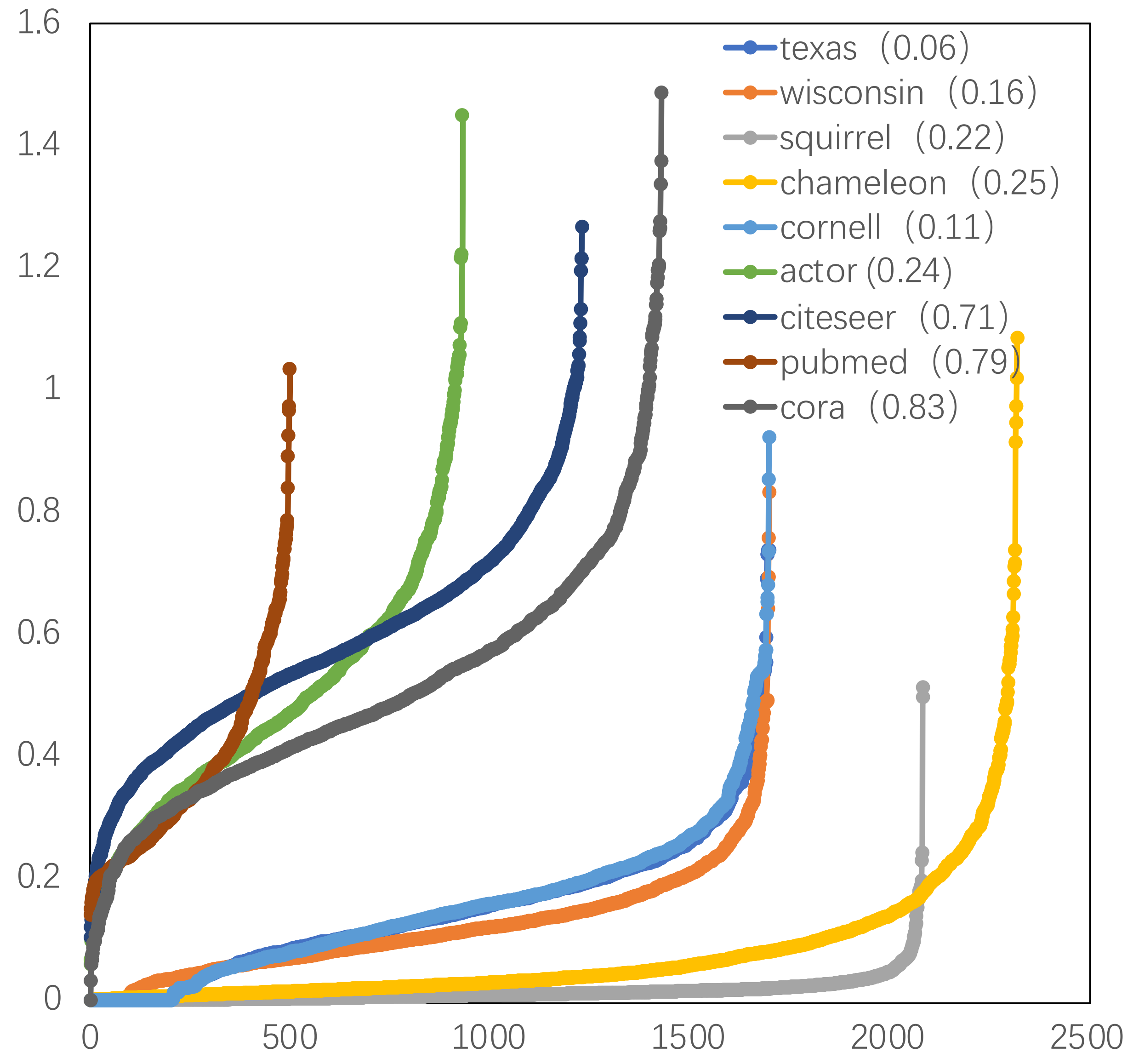
$$\mathbf{h}_v^k = \sigma\Big(\big(\mathbf{m}_{vv}^k + \sum_{u\in\mathcal{N}(v)} \mathbf{m}_{uv}^k\big)\mathbf{W}^k\Big)$$

$$h_{uz} = \frac{\sum_v a_{uv}q_{uv}(z)}{\sum_v h_{vz}}.$$

Conclusions







Semi-supervised Task

$$\mathbf{h}_v^k = \sigma\left(\left(\mathbf{c}_{vv}^k \odot \mathbf{h}_v^{k-1} + \sum_{u\in\mathcal{N}(v)} \mathbf{c}_{uv}^k \odot \mathbf{h}_u^{k-1}\right)\mathbf{W}^k\right);$$

Learned propagation weights vs. Graph partition

Self-supervised Task

$$\mathbf{h}_v^k = \sigma\left(\left(\mathbf{m}_{vv}^k + \sum_{u\in\mathcal{N}(v)} \mathbf{m}_{uv}^k\right)\mathbf{W}^k\right) \quad \mathbf{m}_{uv}^k = \frac{\mathbf{h}_v^{k-1} \odot \mathbf{h}_u^{k-1}}{< \mathbf{h}_v^{k-1}, \mathbf{h}_u^{k-1} >},$$

Inner-product message vs. Community detection

Diverse Message Passing can prevent over-smoothing issue

thank you

Diverse Message Passing

NEURAL INFORMATION PROCESSING SYSTEMS

AAAI-22

河北工業大学

HEBEI UNIVERSITY OF TECHNOLOGY

人工智能与数据科学学院

SCHOOL OF ARTIFICIAL INTELLIGENCE