A Meta-Learning Approach for Graph Representation Learning in Multi-Task Settings

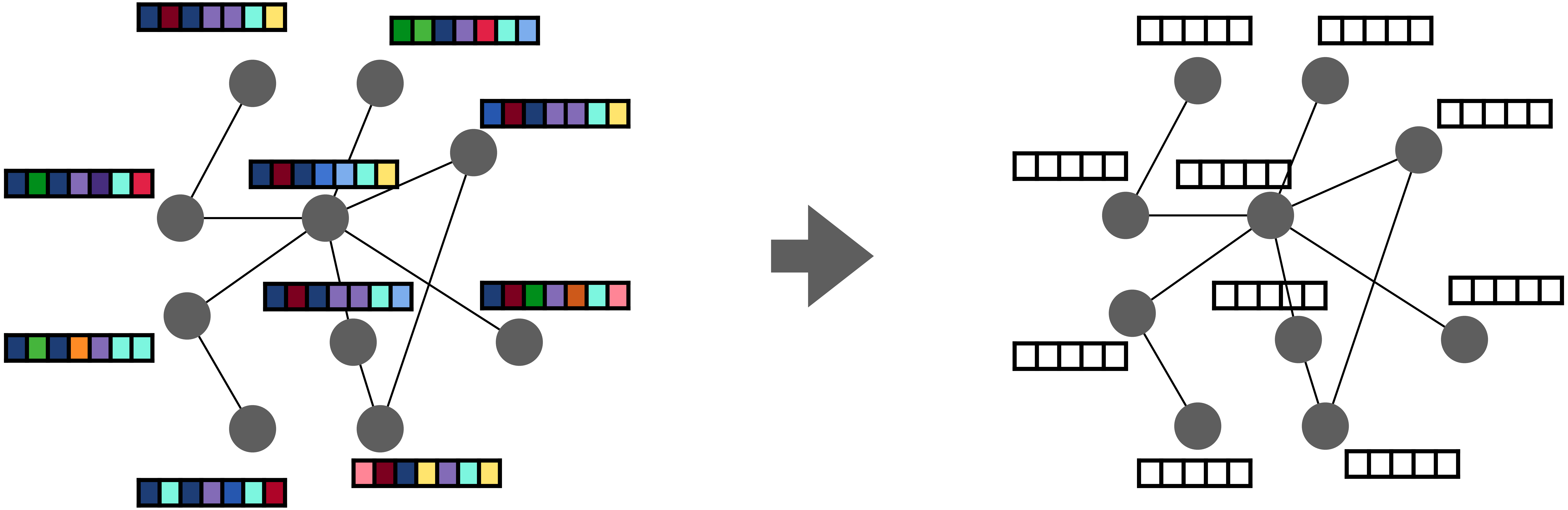
Davide Buffelli, Fabio Vandin
University of Padova
{davide.buffelli, fabio.vandin}@unipd.it

NeurIPS Workshop on Meta-Learning (MetaLearn) 2020

Graph Representation Learning

Node Features $\in \mathbb{R}^D$

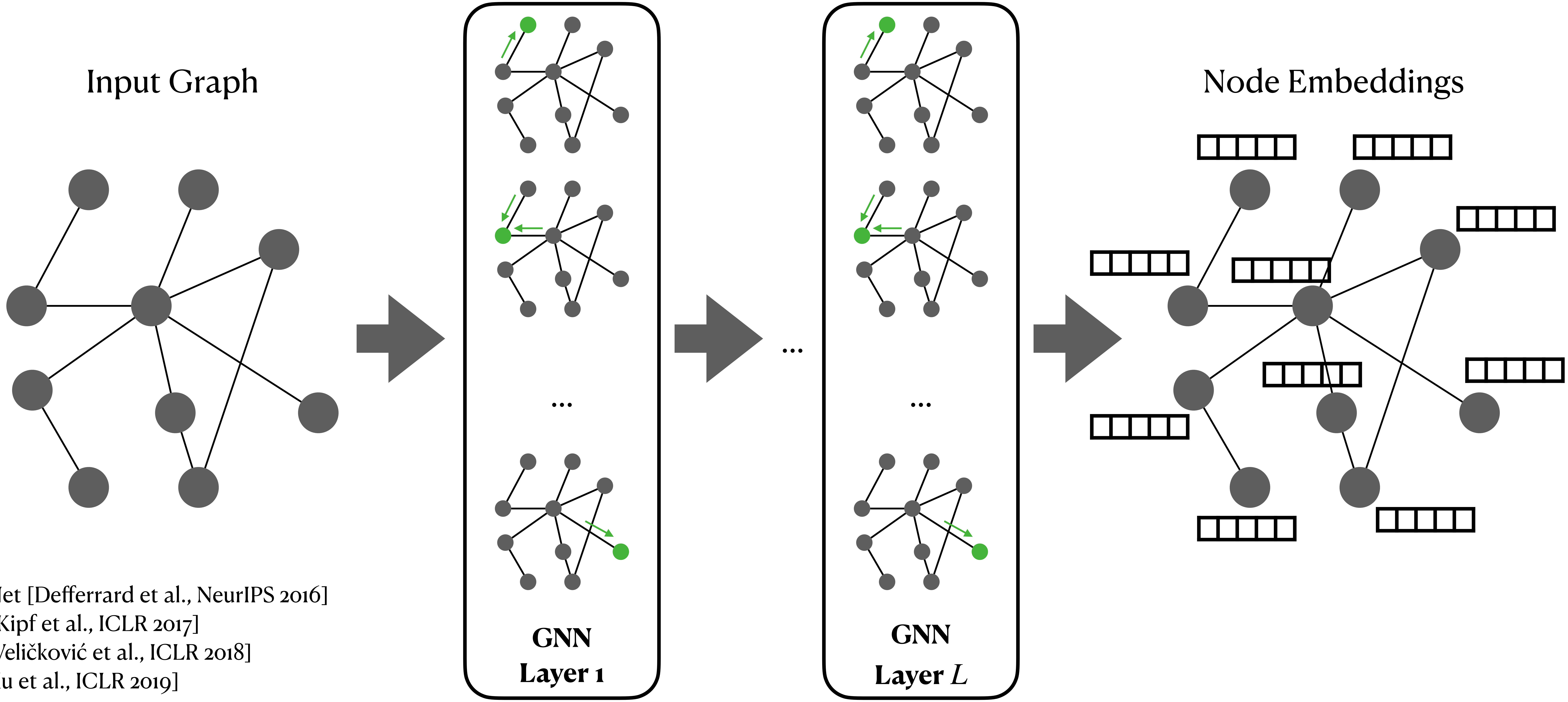
Node Embedding $\in \mathbb{R}^d, \quad d \ll D$



Input Graph

Node Embeddings

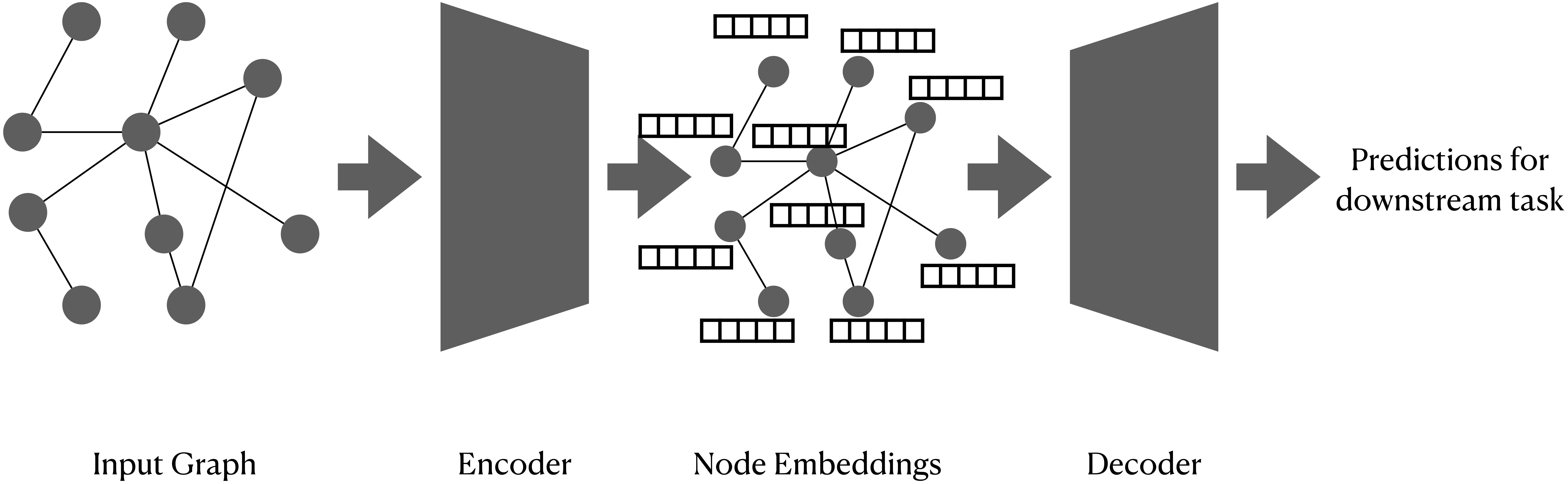
Graph Neural Networks



E.g.
ChebNet [Defferrard et al., NeurIPS 2016]
GCN [Kipf et al., ICLR 2017]
GAT [Veličković et al., ICLR 2018]
GIN [Xu et al., ICLR 2019]

Graph Representation Learning

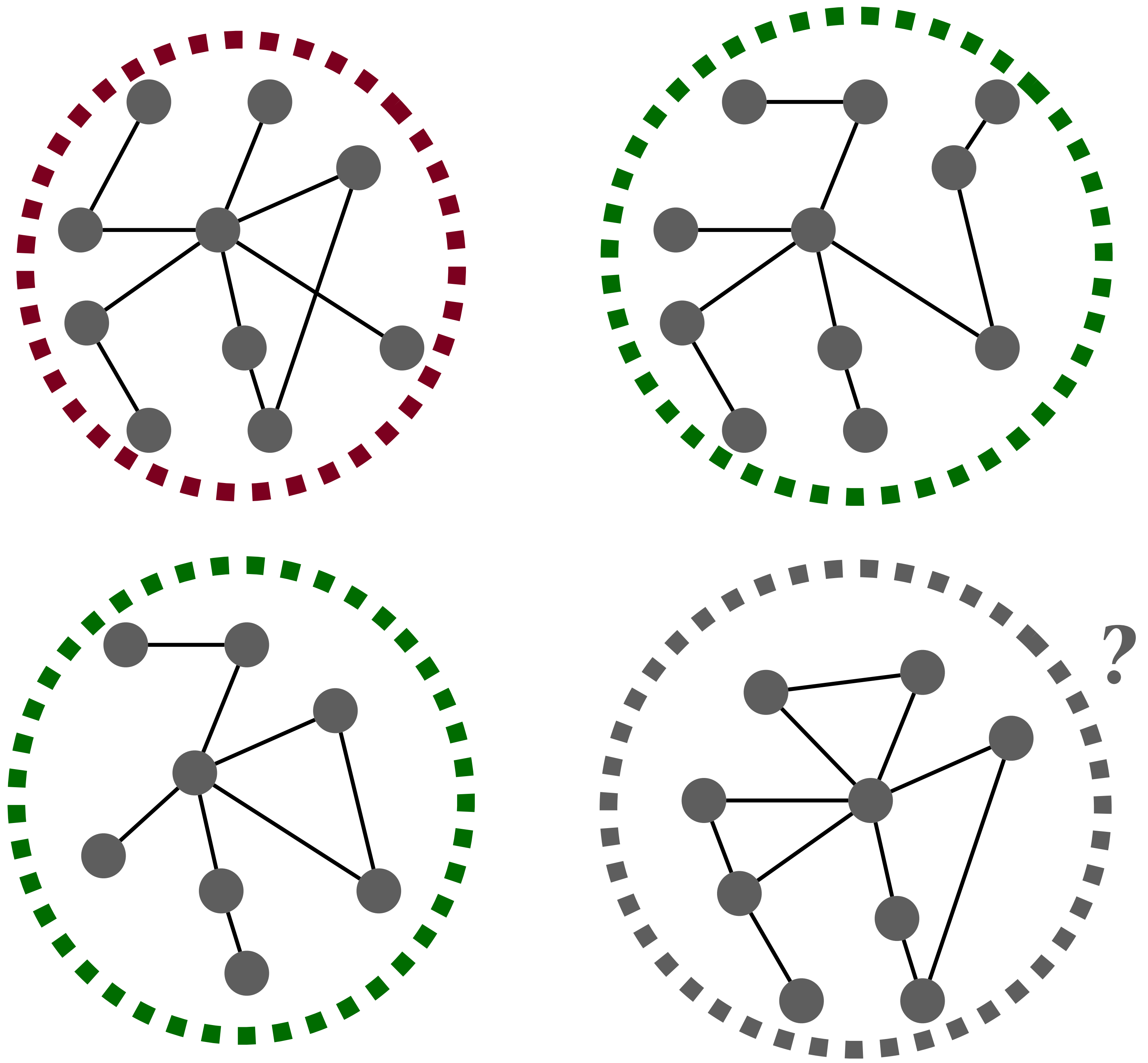
Typical Framework



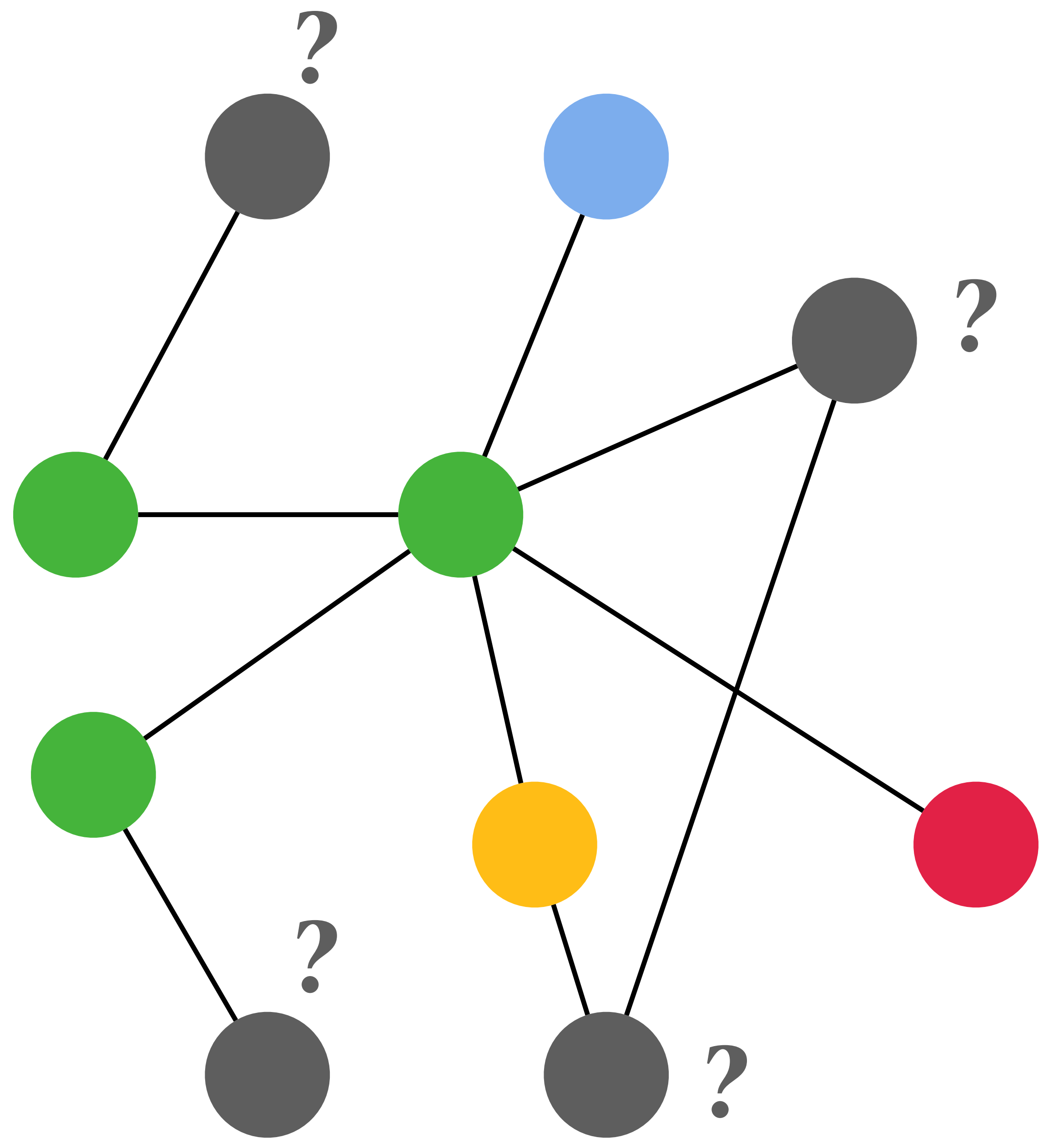
Graph Representation Learning

Applications

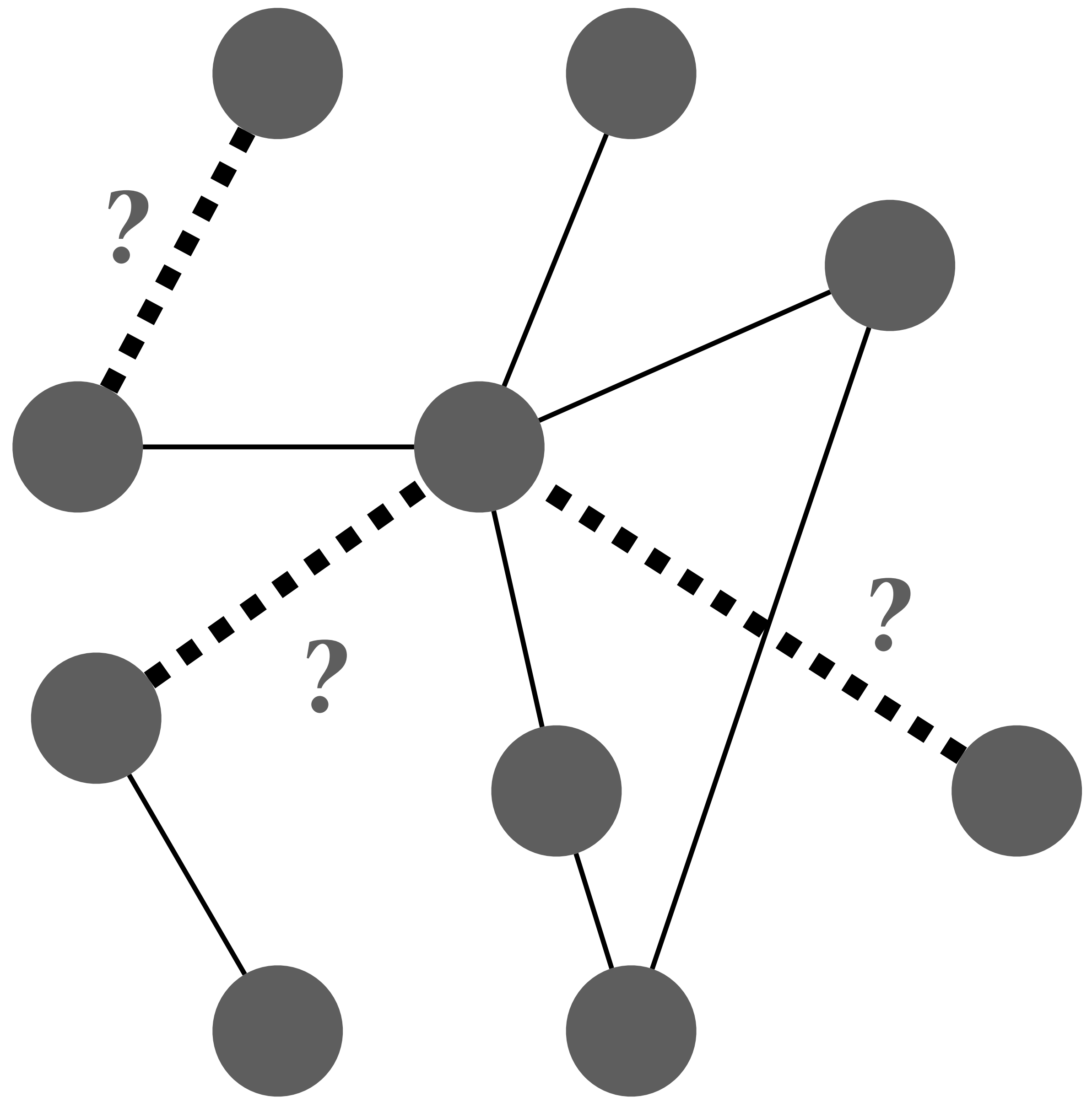
Graph Classification



Node Classification

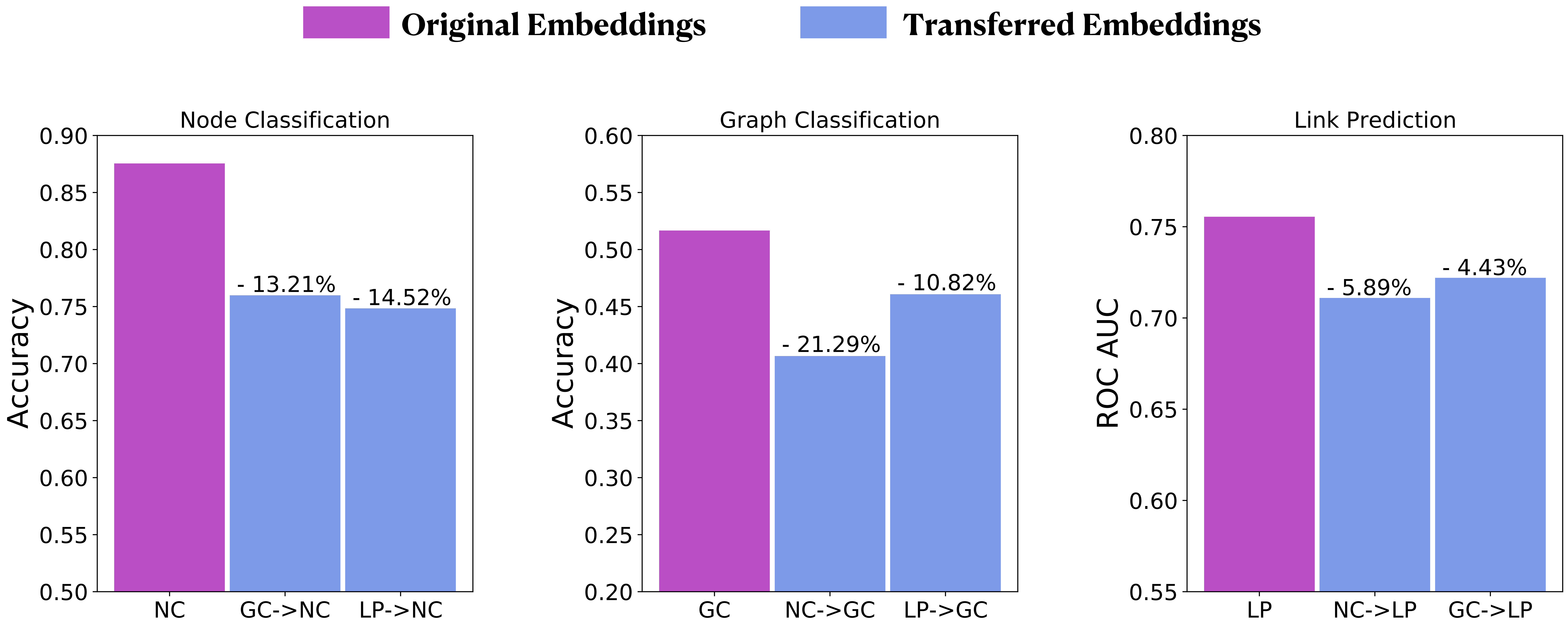


Link Prediction



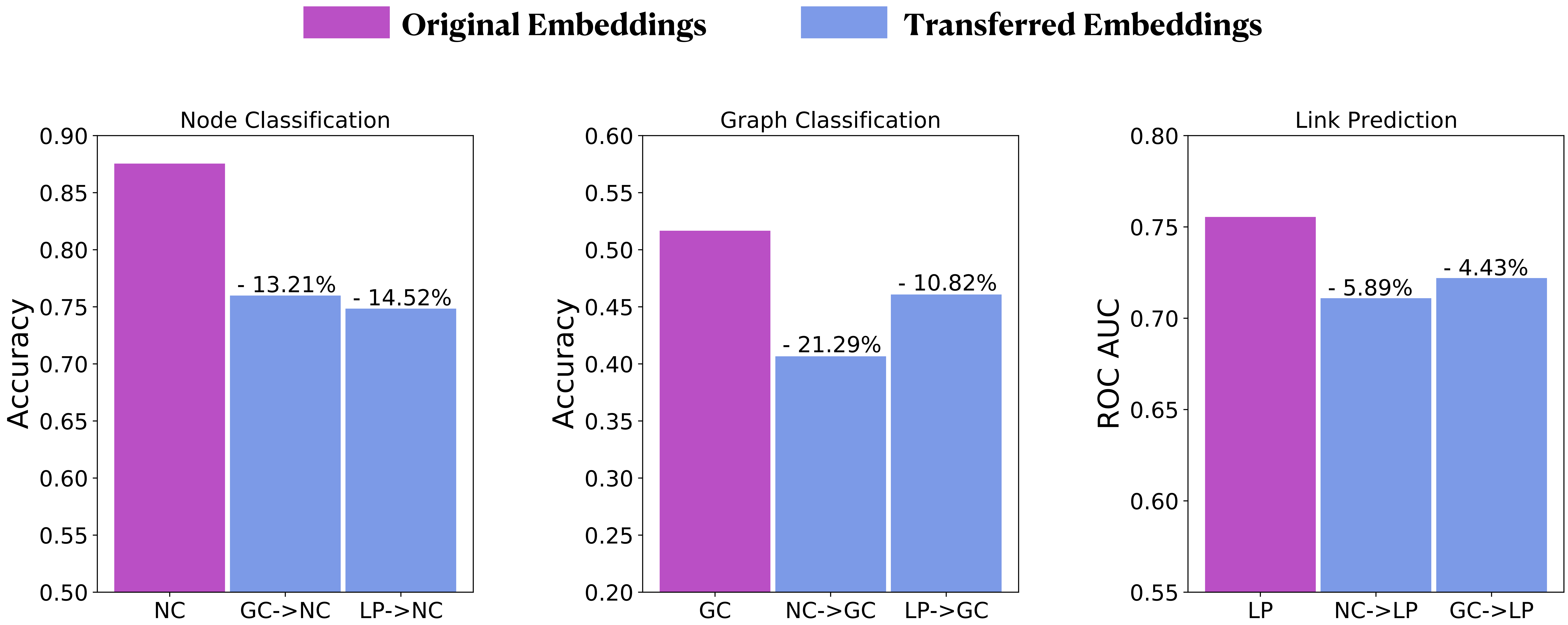
Graph Representation Learning

Transferability of Embeddings



Graph Representation Learning

Transferability of Embeddings



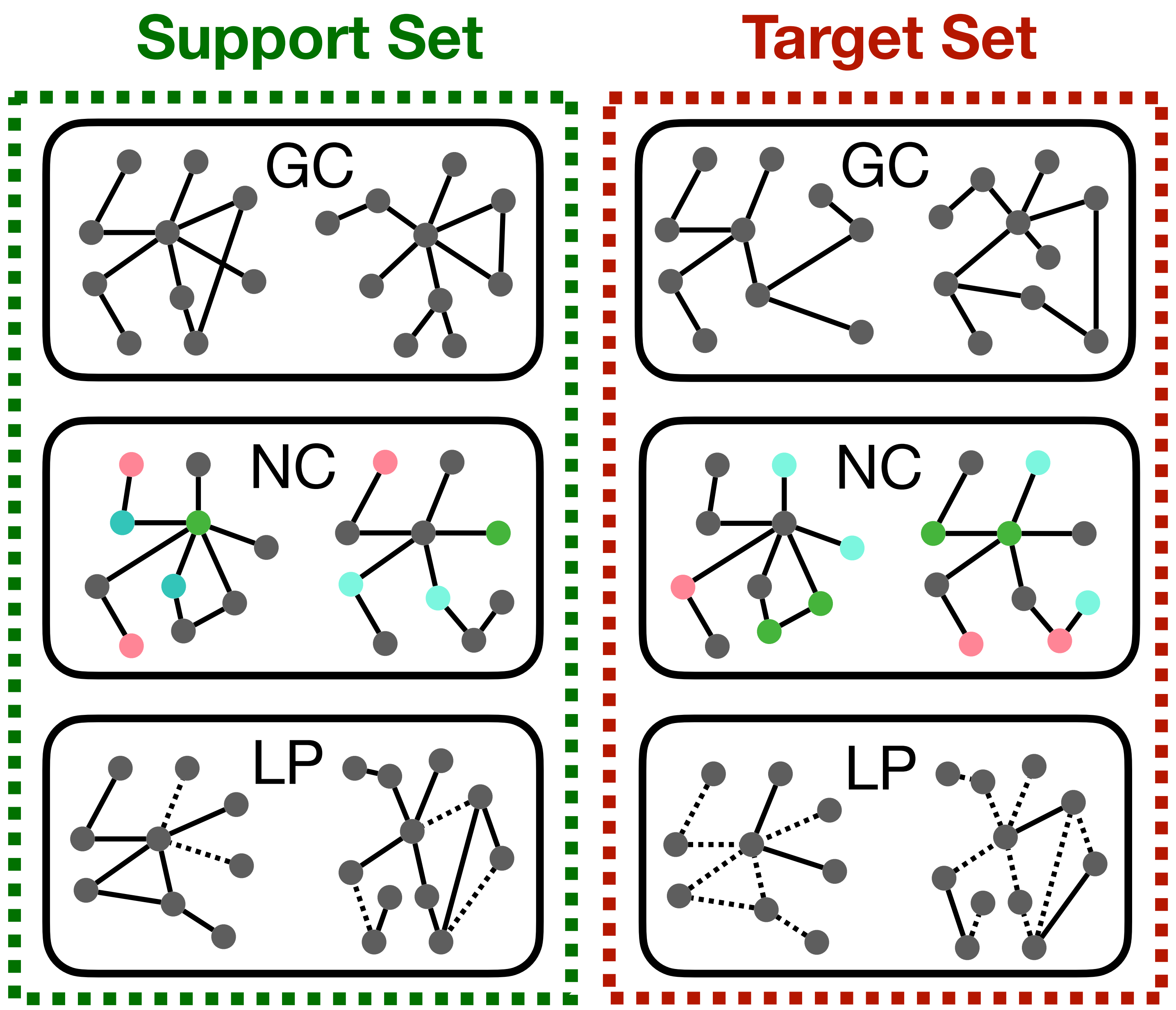
Can we generate node embeddings that generalize across tasks?

SAME: Single-Task Adaptation for Multi-Task Embeddings

Multi-Task Episodes

- 1 task-specific support set per task
- 1 task-specific target set per task

Support Set ↔ Training Set
Target Set ↔ Validation Set



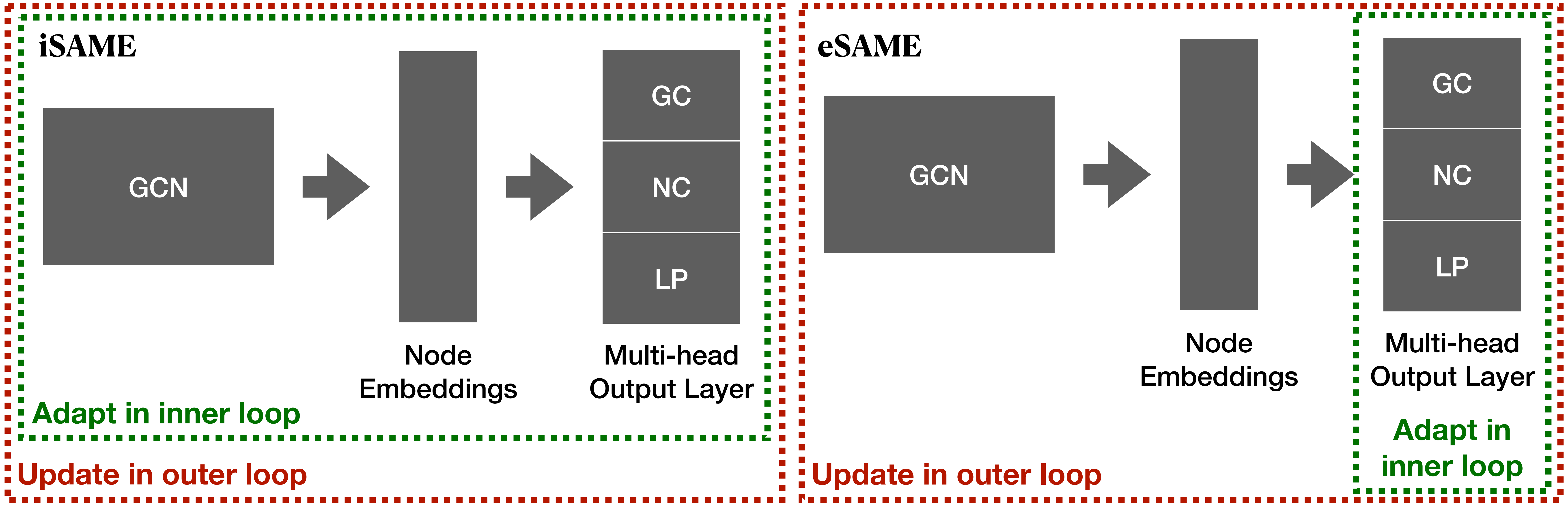
Multi-Learning Procedure

- Separate adaptation for each task
- Unique outer loop update

Two variants:
- eSAME
- iSAME

Algorithm 1: Proposed Meta-Learning Procedure

Input : Model f_θ; Episodes $\mathcal{E} = \{\mathcal{E}_1, .., \mathcal{E}_n\}$
$\texttt{init}(\theta)$
for $\mathcal{E}_i$ *in* $\mathcal{E}$ **do**
  $\texttt{o\_loss} \leftarrow 0$
  for t *in (GC, NC, LP)* **do**
    $\theta'^{(\texttt{t})} \leftarrow \theta$
    $\theta'^{(\texttt{t})} \leftarrow \texttt{ADAPT}(f_\theta, \mathcal{S}^{(\texttt{t})}_{\mathcal{E}_i}, \mathcal{L}^{(\texttt{t})}_{\mathcal{E}_i})$
    $\texttt{o\_loss} \leftarrow \texttt{o\_loss} + \texttt{TEST}(f_{\theta'^{(\texttt{t})}}, \mathcal{T}^{(\texttt{t})}_{\mathcal{E}_i}, \mathcal{L}^{(\texttt{t})}_{\mathcal{E}_i})$
  end
  $\theta \leftarrow \texttt{UPDATE}(\theta, \texttt{o\_loss}, \theta'^{(GC)}, \theta'^{(NC)}, \theta'^{(LP)})$
end



Experiments

1. **Do iSAME and eSAME lead to high quality node embeddings in the single-task setting?**
2. **Do iSAME and eSAME lead to high quality node embeddings in the multi-task setting?**
3. **Do iSAME and eSAME extract information that is not captured by classically trained multi-task models?**
4. **Can the node embeddings learned by iSAME and eSAME be used to perform multiple tasks with comparable or better performance than classical multi-task models?**

Experiments

Do iSAME and eSAME lead to high quality node embeddings in the single-task setting?

Table 1: Results for a single-task model trained in a classical supervised manner (Cl), and a **linear** classifier trained on the embeddings produced by our meta-learning strategies (iSAME, eSAME).

Task	**Model**	**Dataset**			
		ENZYMES	PROTEINS	DHFR	COX2
NC	Cl	87.5 ± 1.9	72.3 ± 4.4	97.3 ± 0.2	96.4 ± 0.3
	iSAME	87.3 ± 0.8	81.8 ± 1.6	96.6 ± 0.3	96.1 ± 0.4
	eSAME	87.8 ± 0.7	82.4 ± 1.6	96.8 ± 0.2	96.5 ± 0.6
GC	Cl	51.6 ± 4.2	73.3 ± 3.6	71.5 ± 2.3	76.7 ± 4.7
	iSAME	50.8 ± 2.9	73.5 ± 1.2	73.2 ± 3.2	76.3 ± 4.6
	eSAME	52.1 ± 5.0	72.6 ± 1.6	71.6 ± 2.4	75.6 ± 4.1
LP	Cl	75.5 ± 3.0	85.6 ± 0.8	98.8 ± 0.7	98.3 ± 0.8
	iSAME	81.7 ± 1.7	84.0 ± 1.1	99.2 ± 0.4	99.1 ± 0.5
	eSAME	80.1 ± 3.4	84.1 ± 0.9	99.2 ± 0.3	99.2 ± 0.7

Experiments

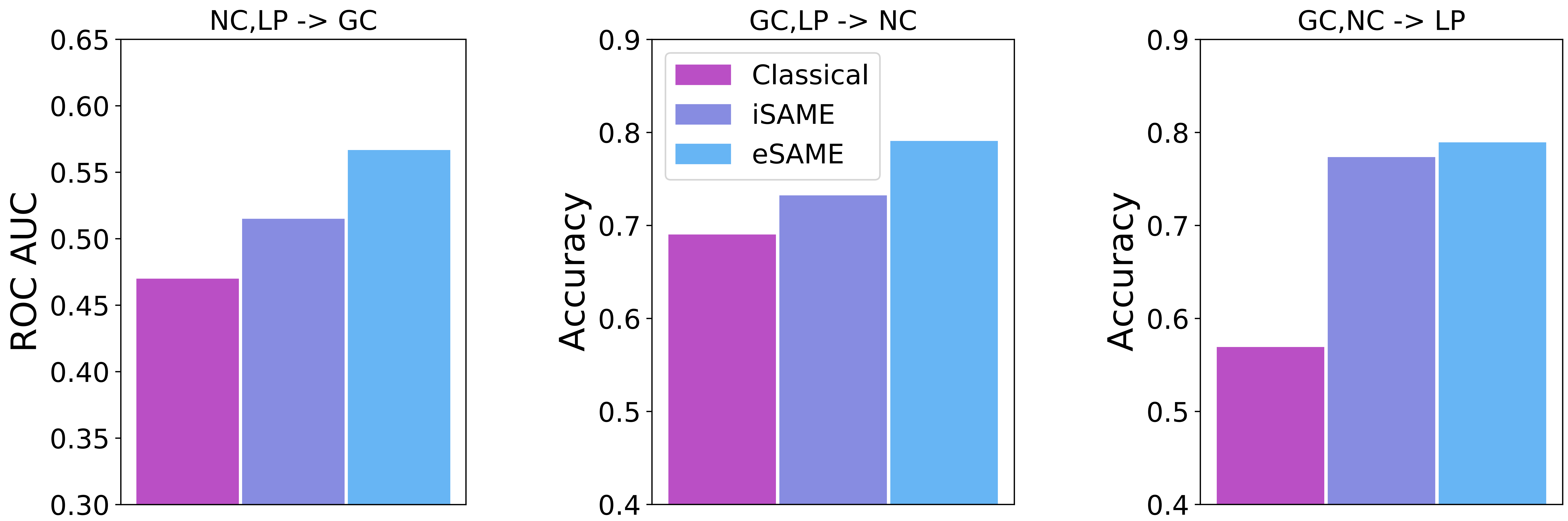
Do iSAME and eSAME lead to high quality node embeddings in the multi-task setting?

Table 2: Results for a single-task model trained in a classical supervised manner, a fine-tuned model (trained on all three tasks, and fine-tuned on the two shown tasks), and a **linear** classifier trained on node embeddings learned with our proposed strategies (iSAME, eSAME) in a multi-task setting.

Task			**Dataset**											
GC	NC	LP	ENZYMES			PROTEINS			DHFR			COX2		
			GC	NC	LP	GC	NC	LP	GC	NC	LP	GC	NC	LP
Classical End-to-End Training														
✓			51.6			73.3			71.5			76.7		
	✓			87.5			72.3			97.3			96.4	
		✓			75.5			85.6			98.8			98.3
Fine-Tuning														
✓	✓		48.3	85.3		73.6	72.0		66.4	92.4		80.0	92.3	
✓		✓	49.3		71.6	69.6		80.7	65.3		58.9	80.2		50.9
	✓	✓		87.7	73.9		80.4	81.5		80.7	56.6		87.4	52.3
iSAME (ours)														
✓	✓		50.1	86.1		73.1	76.6		71.6	94.8		75.2	95.4	
✓		✓	50.7		83.1	73.4		85.2	71.6		99.2	77.5		98.9
	✓	✓		86.3	83.4		79.4	87.7		96.5	99.3		95.5	99.0
✓	✓	✓	50.0	86.5	82.3	71.4	76.6	87.3	71.2	95.5	99.5	75.4	95.2	99.2
eSAME (ours)														
✓	✓		51.7	86.1		71.5	79.2		70.1	95.7		75.6	95.5	
✓		✓	51.9		80.1	71.7		85.4	70.1		99.1	77.5		98.8
	✓	✓		86.7	82.2		80.7	86.3		96.6	99.4		95.6	99.1
✓	✓	✓	51.5	86.3	81.1	71.3	79.6	86.8	70.2	95.3	99.5	77.7	95.7	98.8

Experiments

Do iSAME and eSAME extract information that is not captured by classically trained multi-task models?



Experiments

Can the node embeddings learned by iSAME and eSAME be used to perform multiple tasks with comparable or better performance than classical multi-task models?

Table 3: Δ_m (%) results for a classical multi-task model (Cl), a fine-tuned model (FT; trained on all three tasks and fine-tuned on two) and a **linear** classifier trained on the node embeddings learned using our meta-learning strategies (iSAME, eSAME) in a multi-task setting.

Task GC	Task NC	Task LP	Model	ENZYMES	PROTEINS	DHFR	COX2
✓	✓		Cl	-0.1 ± 0.5	4.0 ± 1.0	-0.3 ± 0.2	0.5 ± 0.1
			FT	-4.5 ± 1.2	0.1 ± 0.5	-7.4 ± 1.4	0.1 ± 0.4
			iSAME	-2.3 ± 0.9	2.7 ± 1.5	-1.2 ± 0.4	-1.6 ± 0.2
			eSAME	-0.8 ± 0.8	3.2 ± 1.4	-1.8 ± 0.3	-1.2 ± 0.3
✓		✓	Cl	-25.3 ± 3.2	-5.3 ± 1.2	-28.3 ± 4.3	-21.4 ± 3.4
			FT	-5.1 ± 1.9	-5.4 ± 1.5	-24.5 ± 3.7	-22.6 ± 3.8
			iSAME	4.1 ± 0.5	-0.2 ± 0.9	0.2 ± 3.2	0.2 ± 0.5
			eSAME	3.2 ± 0.4	-1.2 ± 1.1	-0.7 ± 3.4	-0.8 ± 0.7
	✓	✓	Cl	7.2 ± 2.7	6.8 ± 0.9	-29.1 ± 7.7	-28.2 ± 4.5
			FT	-1.0 ± 0.3	3.1 ± 1.2	-28.9 ± 6.4	-28.3 ± 4.2
			iSAME	4.4 ± 1.1	6.1 ± 1.0	-0.1 ± 6.2	-0.6 ± 2.5
			eSAME	3.9 ± 1.3	6.1 ± 1.1	0.1 ± 6.4	-0.6 ± 2.6
✓	✓	✓	Cl	1.6 ± 1.3	2.9 ± 0.3	-18.9 ± 2.3	-16.9 ± 3.1
			iSAME	1.5 ± 1.0	2.2 ± 0.2	-0.5 ± 1.4	-0.9 ± 1.3
			eSAME	1.8 ± 0.9	2.8 ± 0.2	-1.0 ± 1.7	-0.4 ± 1.2

Thank you for watching!

Don’t hesitate to come to our virtual booth and have a chat.

You can also contact us at:
{davide.buffelli, fabio.vandin}@unipd.it