ICLR

One for All: Towards Training One Graph Model for All Classification Tasks

Hao Liu[1*], Jiarui Feng[1*], Lecheng Kong[1*], Ningyue Liang[1], Dacheng Tao[2], Yixin Chen[1], Muhan Zhang[3†]

[1] Washington University in St. Louis [2] Nanyang Technological University [3] Peking University

Washington University in St. Louis

NANYANG TECHNOLOGICAL UNIVERSITY SINGAPORE

北京大學 PEKING UNIVERSITY

* Contributed equally.

† Corresponding Author.

Motivation: Graph Foundation Model

- A foundation model is a model trained on broad data that can be adapted to a wide range of downstream tasks.

- Language domain:

OpenAI ChatGPT4.0

Text Generation

Translation

Question Answering

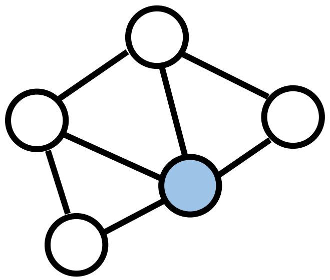
…

- Graph domain:

A single GNN model can only deal with a single dataset and a single task.

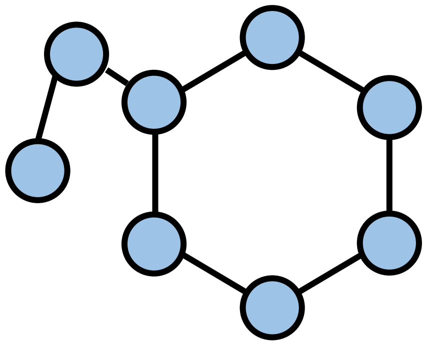
Goal (One for All): Train one graph model for all classification tasks

Node Classification



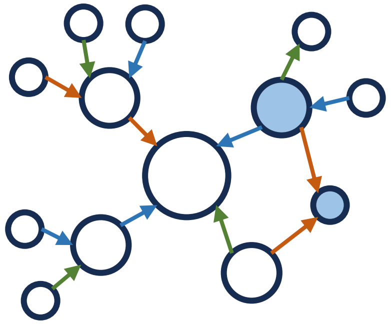
What's the category of this **node**?
1. … 2. … 3. …

Graph Classification



Does this **molecular graph** inhibit HIV?
Yes or No

Link Prediction



What's the relationship between **these two nodes?**
1. … 2. … 3. …

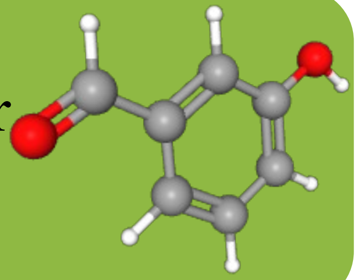
Challenge 1: Input to GNNs are Different

For example,
  Node feature of citation graph are Bag-of-word vector of paper's title and abstract.
  Node feature of molecular graph are indices of nominal features of atoms.

Citation Network

{ 'Attention': 1, 'is': 3, 'all': 2, 'you': 1, 'need': 1, 'The': 1, 'dominant': 1, 'sequence': 2, 'transduction': 1, 'models': 2, 'are': 1, … }

Molecular Graph



{ 'Atom_Type_Index': 1, 'Atomic_Number': 6, 'In_Ring': 0, 'Bond_Type': 2 }

✗ Feature Type ✗ Feature Dimension

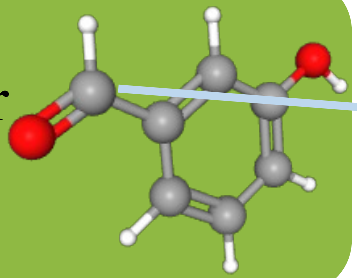
Solution to Challenge 1: Unify Feature with TAGs

- Use readable texts to describe nodes and edges.

Citation Network

Paper title and abstract: Attention is all you need. The dominant sequence transduction models are …

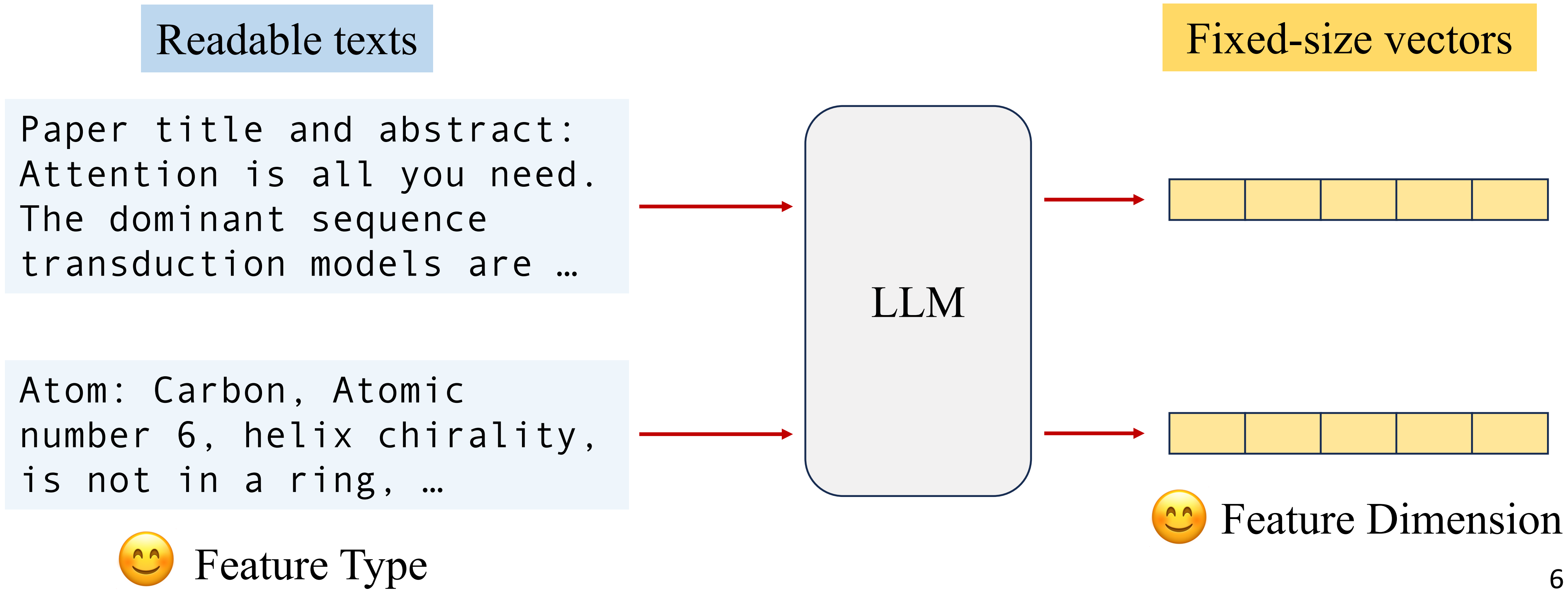
Molecular Graph



Atom: Carbon, Atomic number 6, helix chirality, is not in a ring, …

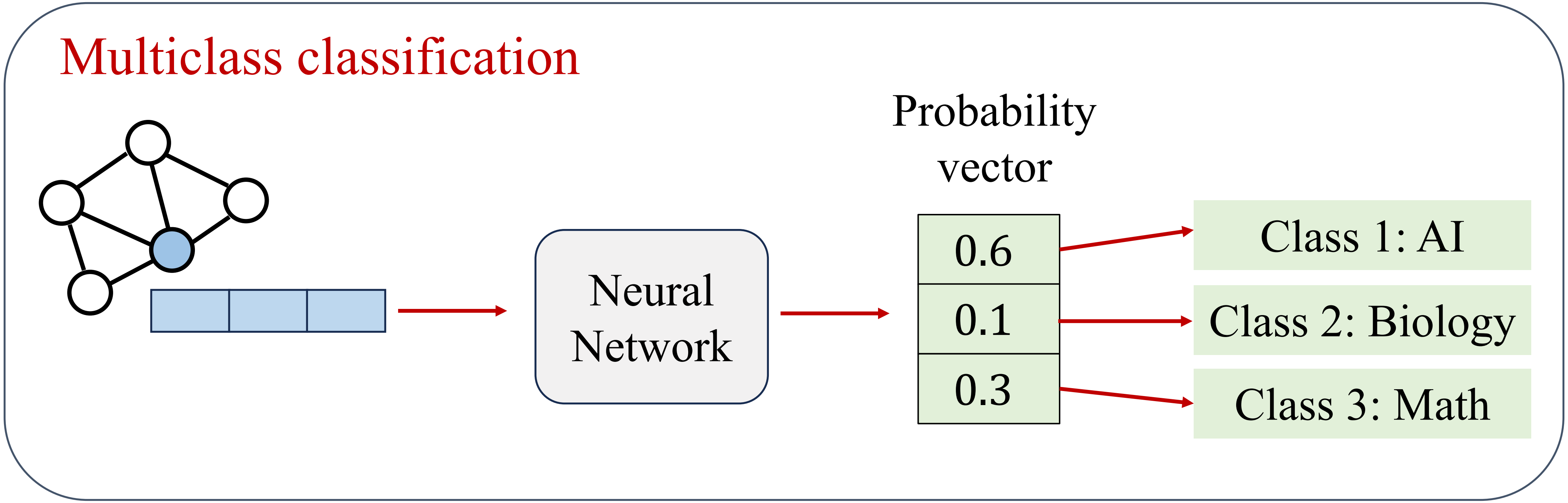
Solution to Challenge 1: Unify the Input to GNNs

- Apply an LLM to encode graph attributes to the same space.



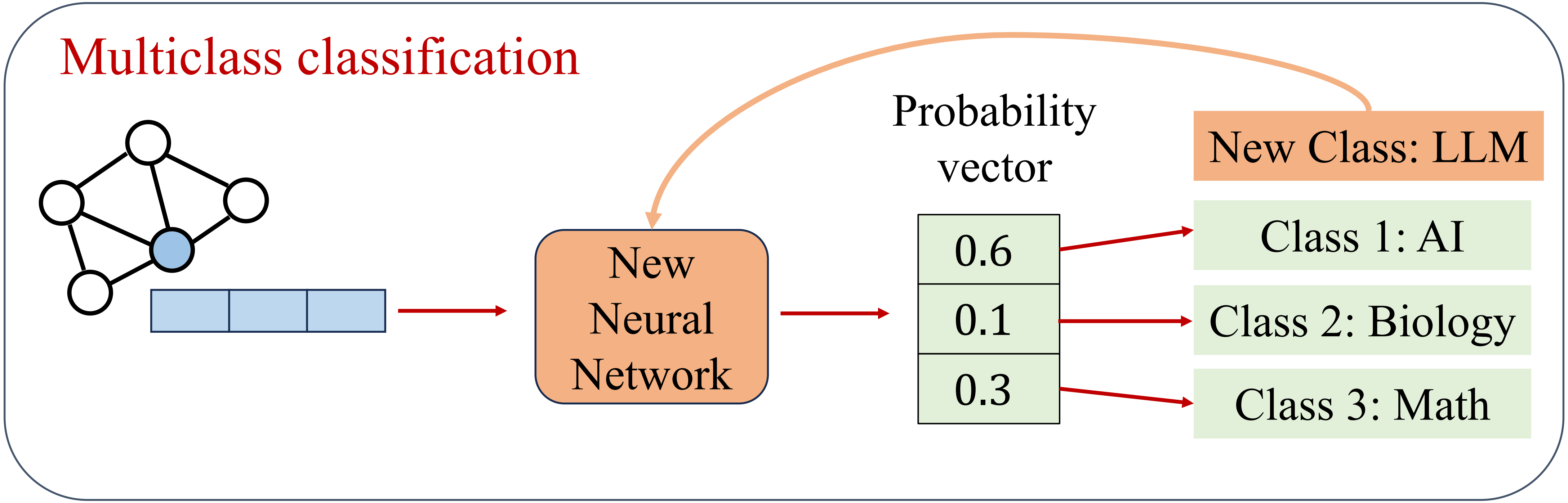
Challenge 2: Output from GNNs are Different

Different graph datasets have different labels,
such that the output from GNNs are different across datasets.



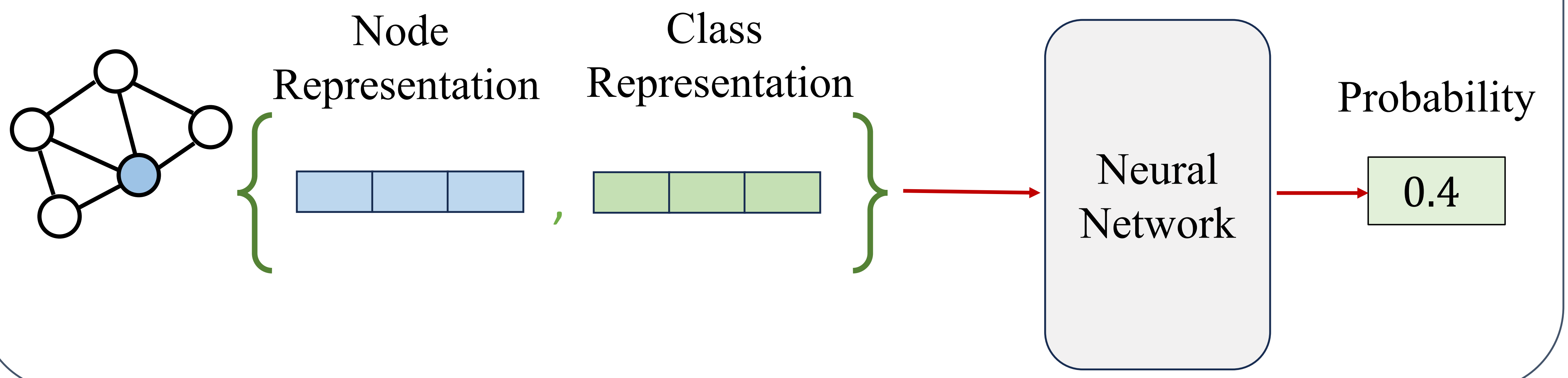
Challenge 2: Output from GNNs are Different

Need to train new model when classes change.

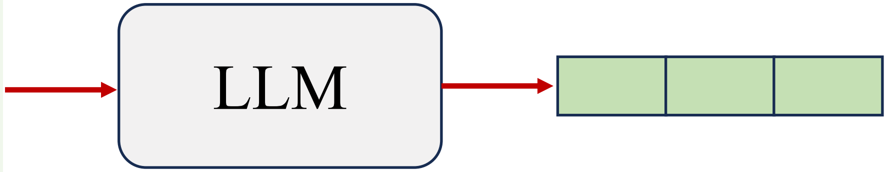


Solution to Challenge 2: Include Class Information in the Input

Multiple binary classification with class information in input

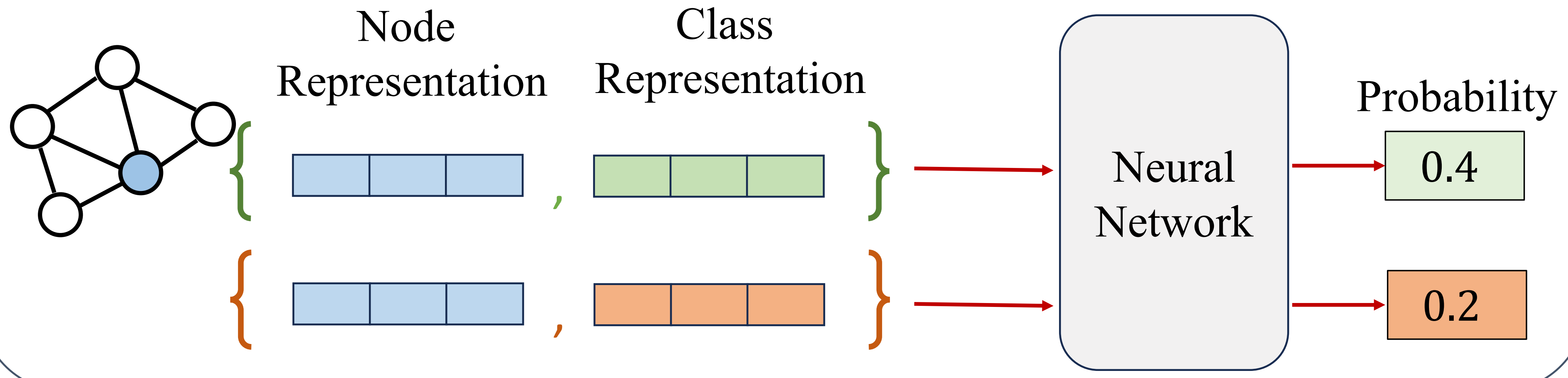



```
Literature Category. cs.AI
(Artificial Intelligence). Covers
all areas of AI except Vision …
```

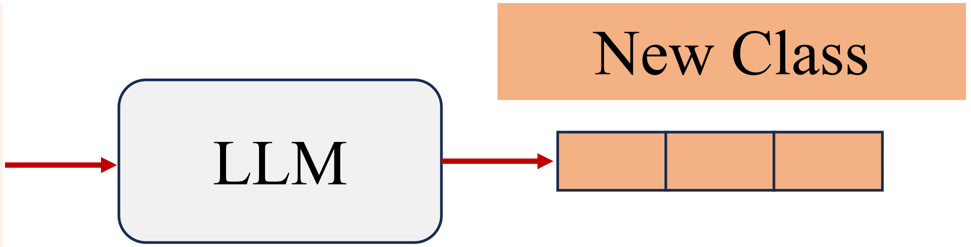


Solution to Challenge 2:
Include Class Information in the Input

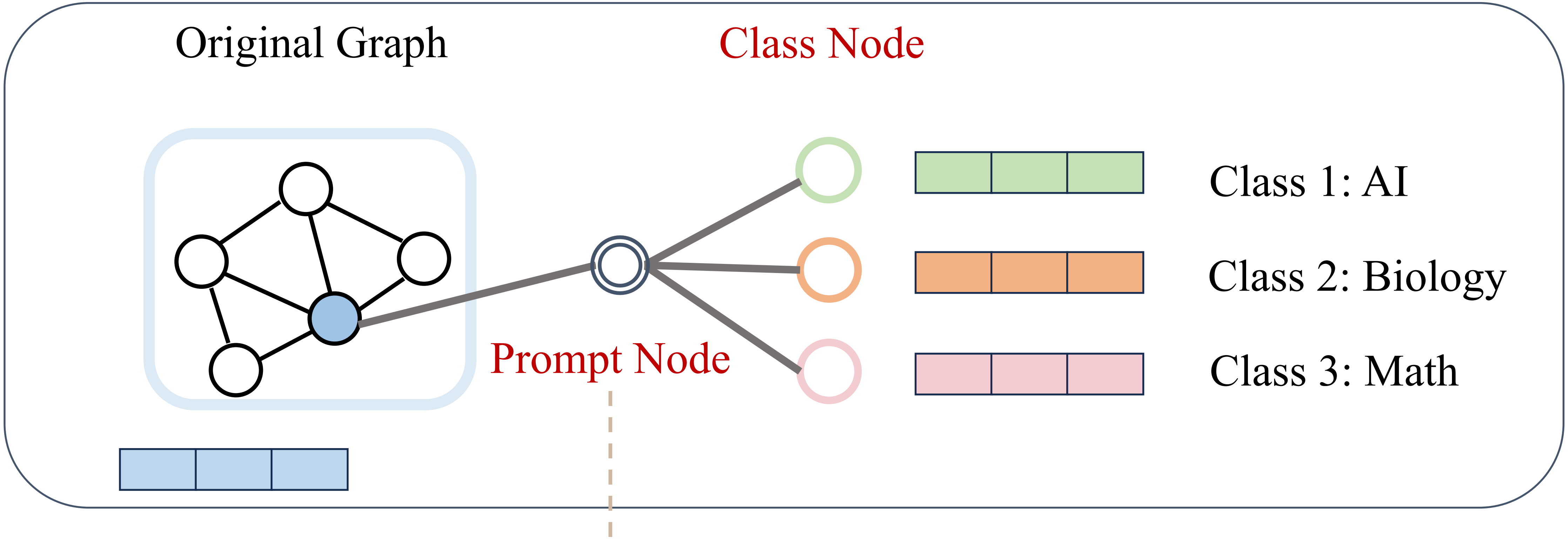
Multiple binary classification with class information in input




```
Literature Category. cs.CV
(Computer Vision and Pattern
Recognition). Covers image
processing, computer vision …
```



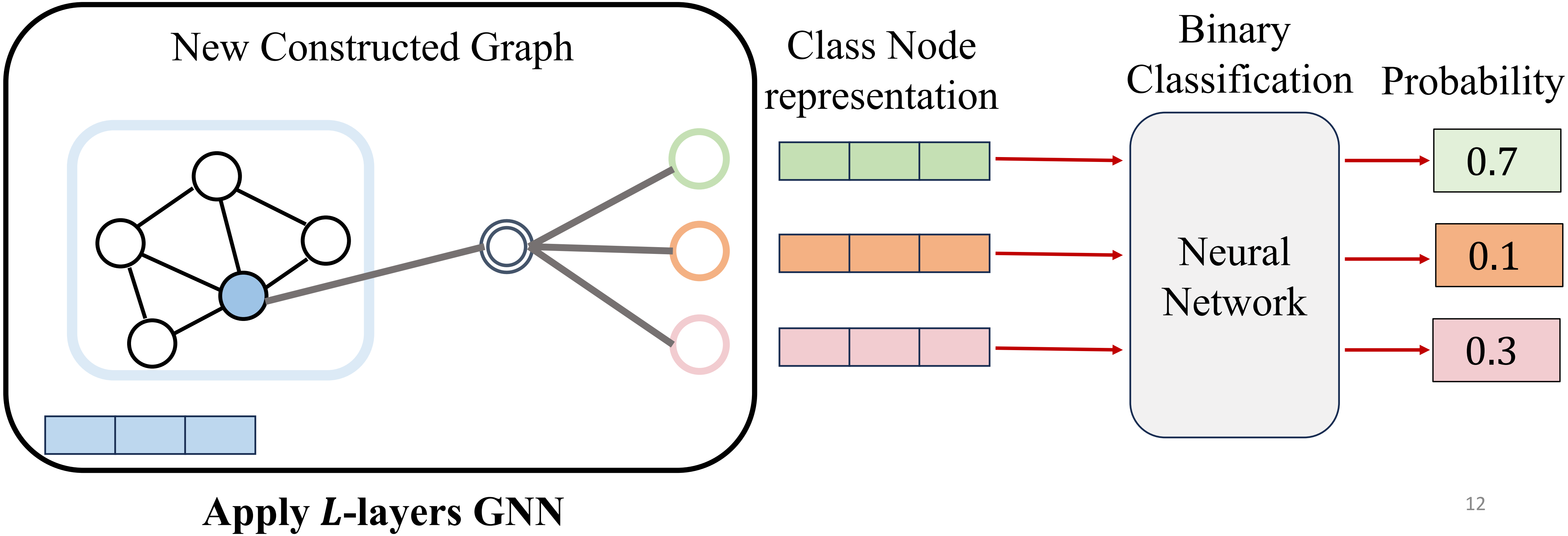
Solution to Challenge 2: Class Node Representation as Input



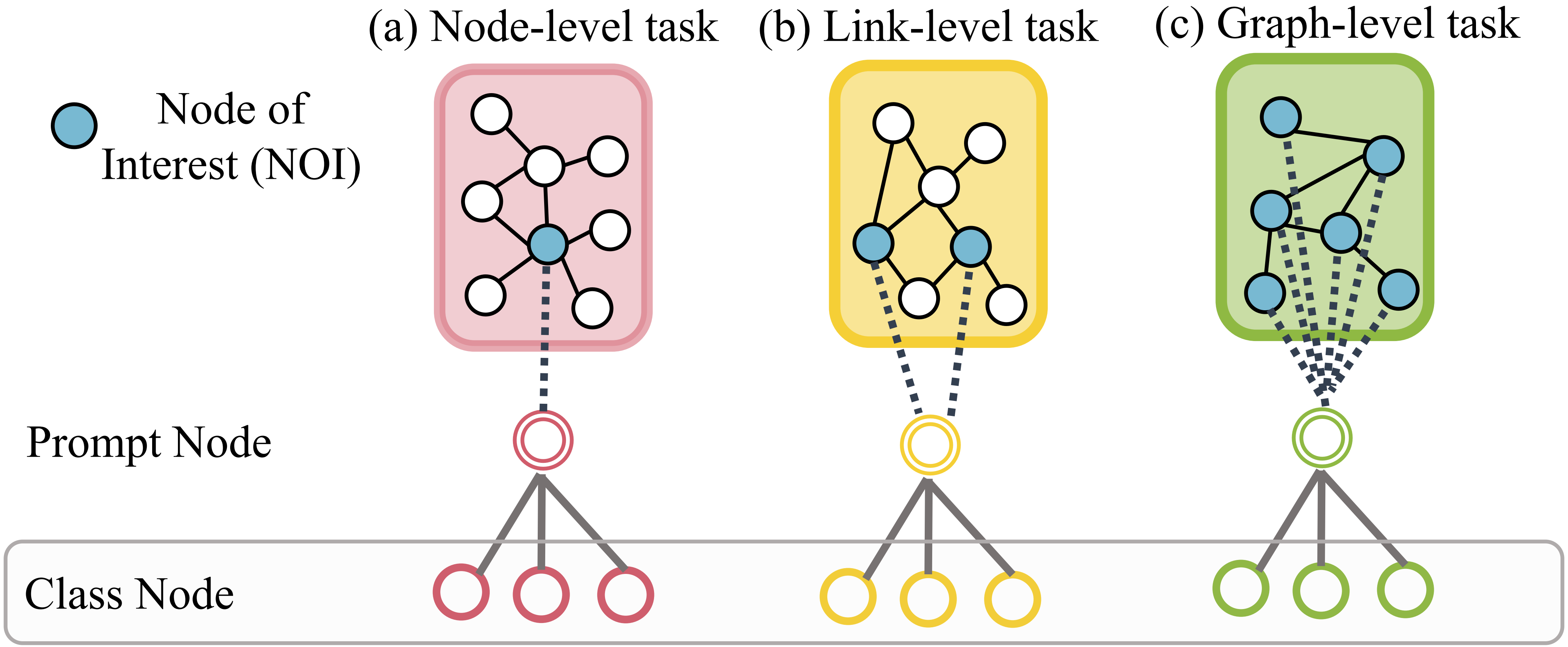

```
Node Classification on the
literature category of the paper.
```

Solution to Challenge 2: Class Node Representation as Input

After applying GNN, representation of class node include information of both class node and original node.

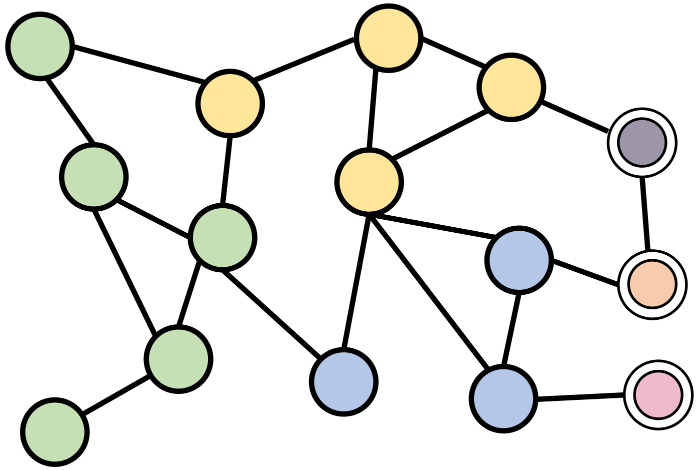


Unify Different-level Tasks



All tasks are transformed to a binary classification task on class node representation.

Limited Data Regime 1: Few-shot Scenario



Nodes of base classes

Nodes of novel classes

Given:

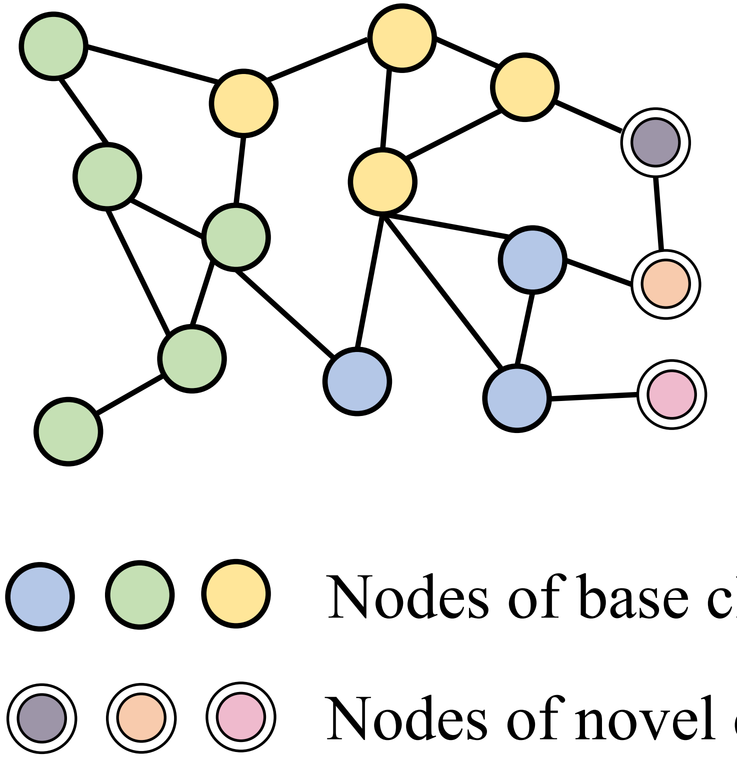
1. Abundant labeled nodes in base classes
2. Limited labeled nodes in novel classes

Target:

Classify unlabeled nodes in novel classes

If # of novel classes is N and # of labeled nodes per class is k, this task is called N-way k-shot few-shot task.

Limited Data Regime 2: Zero-shot Scenario



Nodes of base classes

Nodes of novel classes

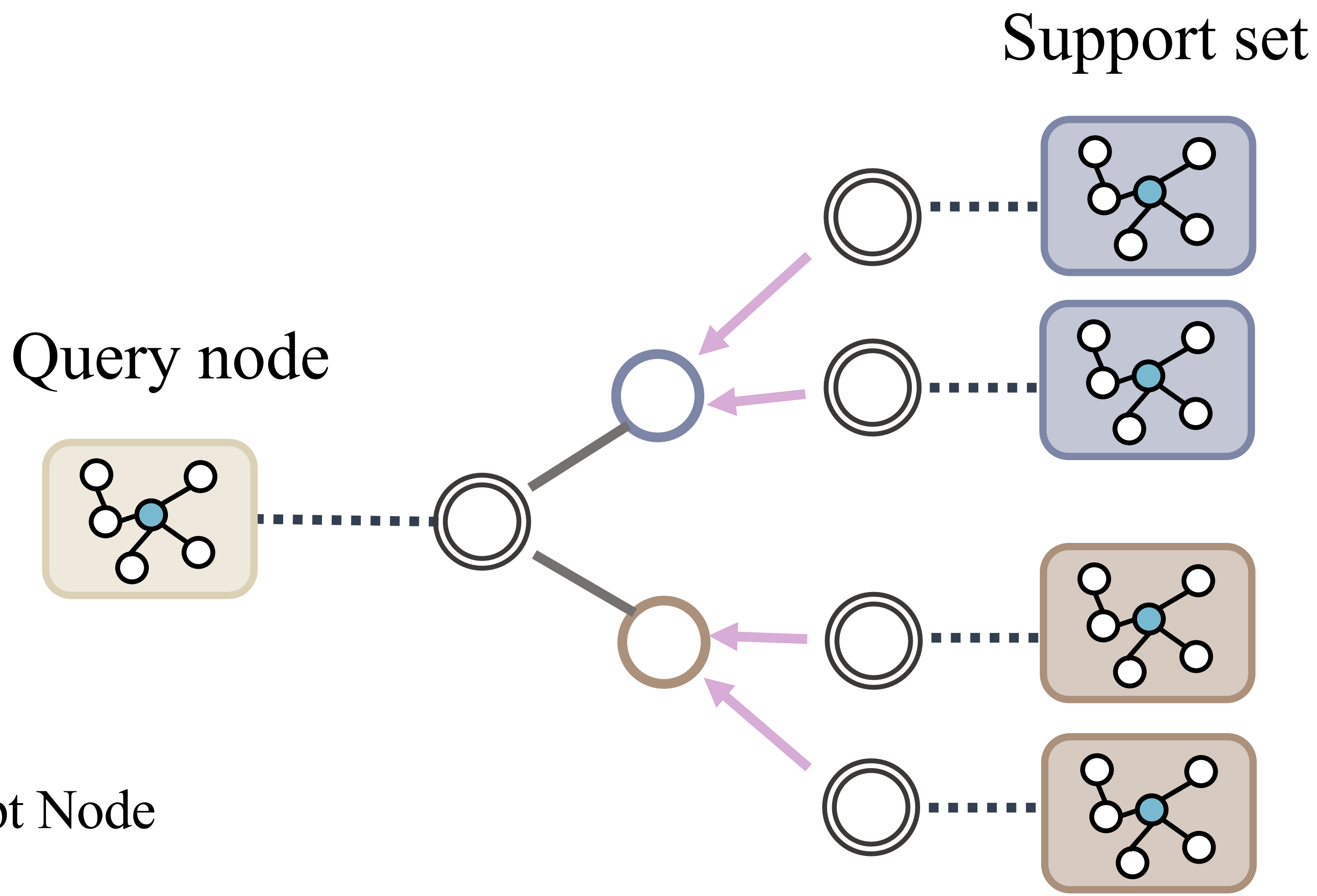
Given:

1. Abundant labeled nodes in base classes
2. ~~Limited labeled nodes in novel classes~~

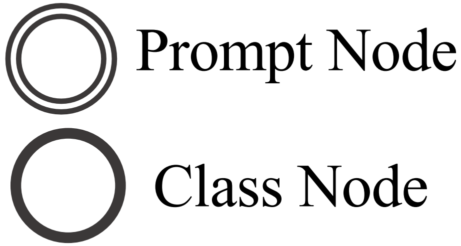
Target:

Classify unlabeled nodes in novel classes

Few-shot Scenario: Graph Construction



Connect support set graph structure to corresponding class node



Supervised Learning Experiment

OFA trains a single graph model on all datasets and tasks simultaneously.

Table 2: Results on supervised learning (first).

Task type Metric	Cora Link AUC ↑	Cora[1] Node Acc ↑	PubMed Link AUC ↑	PubMed[1] Node Acc ↑	ogbn-arxiv[1] Node Acc ↑	Wiki-CS Node Acc ↑	HIV Graph AUC ↑
GCN	90.40±0.20	78.86±1.48	91.10±0.50	74.49±0.99	74.09±0.17	79.07±0.10	75.49±1.63
GAT	93.70±0.10	**82.76±0.79**	91.20±0.10	75.24±0.44	74.07±0.10	**79.63±0.10**	74.45±1.53
OFA-ind-st	91.87±1.03	75.61±0.87	98.50±0.06	73.87±0.88	75.79±0.11	77.72±0.65	73.42±1.14
OFA-st	94.04±0.49	75.90±1.26	98.21±0.02	75.54±0.05	75.54±0.11	78.34±0.35	78.02±0.17
OFA-e5	92.83±0.38	72.20±3.24	98.45±0.05	77.91±1.44	75.88±0.17	73.02±1.06	**78.29±1.48**
OFA-llama2-7b	94.22±0.48	73.21±0.73	**98.69±0.10**	77.80±2.60	77.48±0.17	77.75±0.74	74.45±3.55
OFA-llama2-13b	**94.53±0.51**	74.76±1.22	98.59±0.10	**78.25±0.71**	**77.51±0.17**	77.65±0.22	76.71±1.19

Table 3: Results on supervised learning (second).

Task type Metric	WN18RR Link Acc ↑	FB15K237 Link Acc ↑	PCBA Graph APR ↑
GCN	67.40±2.40	74.20±1.10	20.20±0.24
GIN	57.30±3.40	70.70±1.80	22.66±0.28
OFA-ind-st	97.22±0.18	**95.77±0.01**	22.73±0.32
OFA-st	96.91±0.11	95.54±0.06	24.83±0.10
OFA-e5	97.84±0.35	95.27±0.28	**25.19±0.33**
OFA-llama2-7b	98.08±0.16	95.56±0.05	21.35±0.94
OFA-llama2-13b	**98.14±0.25**	95.69±0.07	21.54±1.25

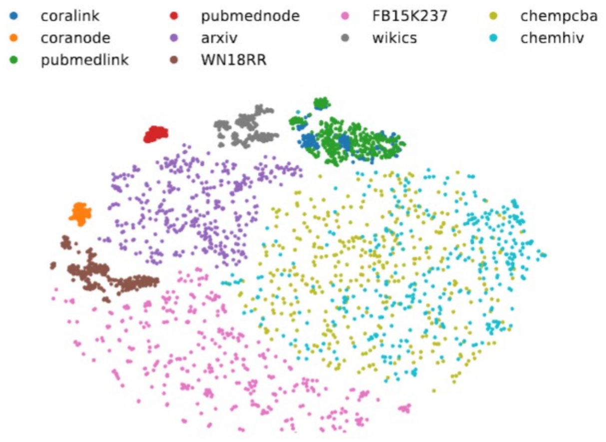


Figure 3: Output embedding space of NOI prompt nodes on all datasets for OFA-joint-st.

Few-shot and Zero-shot Setting

Table 5: Few-shot and Zero-shot results (Acc) on ogbn-arxiv and Cora (Node-level).

# Way	ogbn-arxiv-5-way (Transductive)				Cora-2-way (Transfer)		
Task	5-shot	3-shot	1-shot	0-shot	5-shot	1-shot	0-shot
GPN	50.53±3.07	48.32±3.80	38.58±1.61	-	63.83±2.86	56.09±2.08	-
TENT	60.83±7.45	56.03±8.90	45.62±10.70	-	58.97±2.40	54.33±2.10	-
GLITTER	56.00±4.40	57.44±4.90	47.12±2.73	-	-	-	-
TLP-BGRL	50.13±8.78	46.21±7.92	35.81±8.58	-	81.31±1.89	59.16±2.48	-
TLP-SURGL	77.89±6.46	74.19±7.55	61.75±10.07	-	92.49±1.02	81.52±2.09	-
Prodigy	61.09±5.85	58.64±5.84	48.23±6.18	-	-	-	-
OFA-joint-lr	61.45±2.56	59.78±2.51	50.20±4.27	46.19±3.83	76.10±4.41	67.44±4.47	56.92±3.09

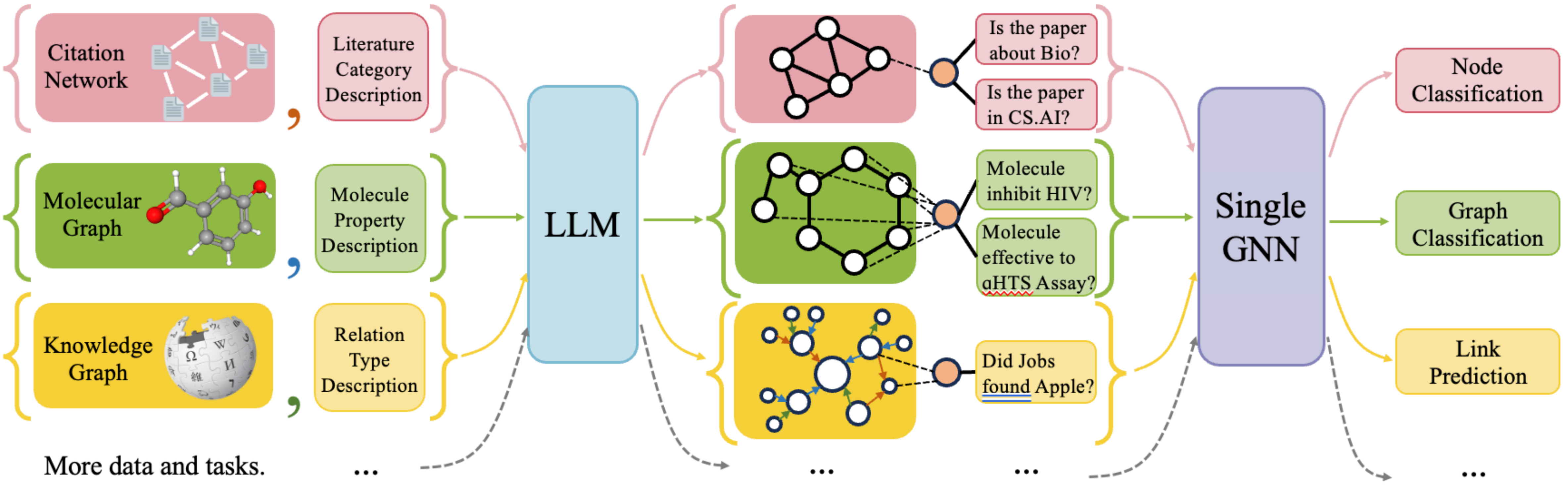
Table 6: Few-shot and Zero-shot results (Acc) on FB15K237 and WN18RR (Link-level).

# Way	FB15K237-20-way (Transductive)		WN18RR-5-way (Transfer)	
Task	5-shot	0-shot	5-shot	0-shot
Prodigy	74.92±6.03	-	-	-
OFA-joint-lr	82.56±1.58	70.20±2.40	46.32 ±4.18	30.96±5.46

Table 7: Few-shot and Zero-shot results (AUC) on HIV and PCBA (Graph-level & Transfer setting).

# Way	HIV-2-way		PCBA-2-way	
Task	5-shot	0-shot	5-shot	0-shot
Galactica-1.3B	-	33.85	-	52.02
GIMLET	-	66.24	-	62.11
OFA-joint-lr	63.58±1.81	35.67±4.46	51.53±9.94	60.62±5.45

Whole Picture of our Model: OFA



Unify Input | Unify Output | One model for all tasks

Thank you!

Arxiv: https://arxiv.org/abs/2310.00149

Code: https://github.com/LechengKong/OneForAll