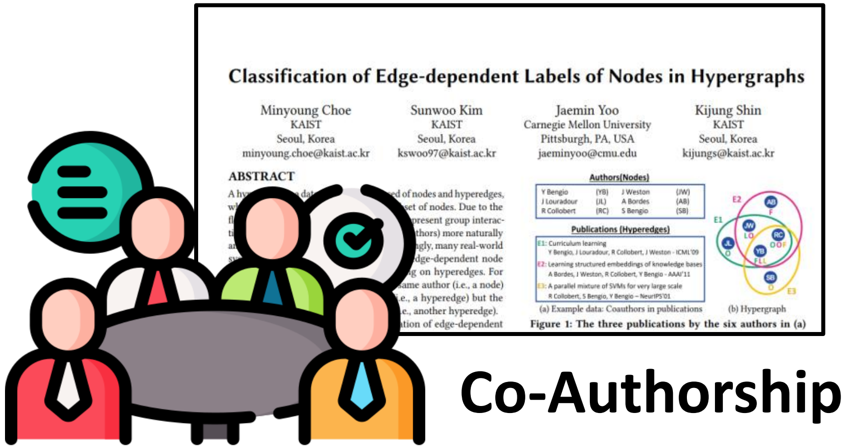
Classification of Edge-dependent Labels of Nodes in Hypergraphs



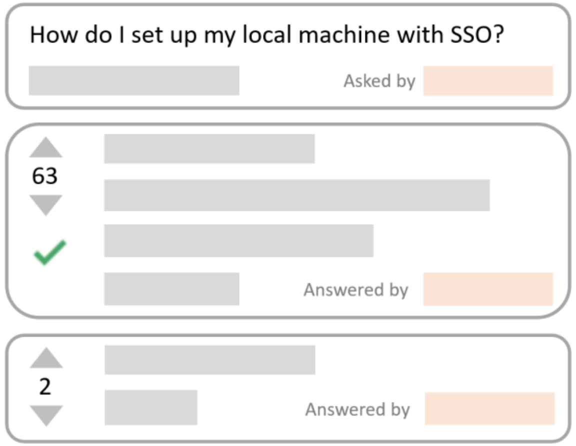

Minyoung Choe Sunwoo Kim Jaemin Yoo Kijung Shin

Group Interactions are EVERYWHERE



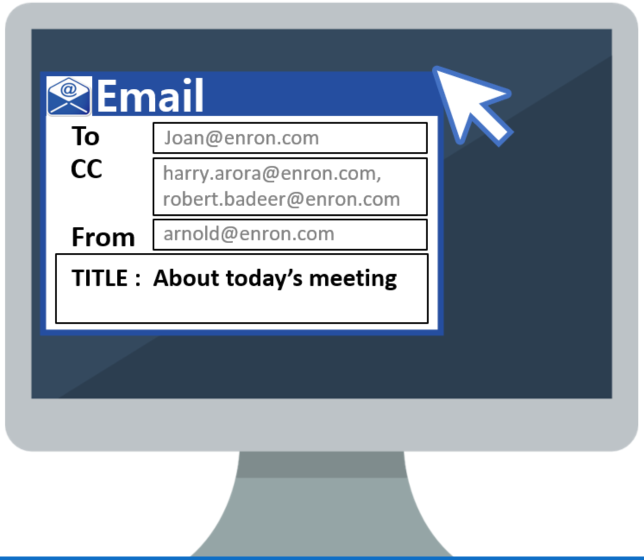
Co-Authorship

Q&A Platform



Multi-player Game

Email



Group Interactions → Hypergraph

Authors(Nodes)

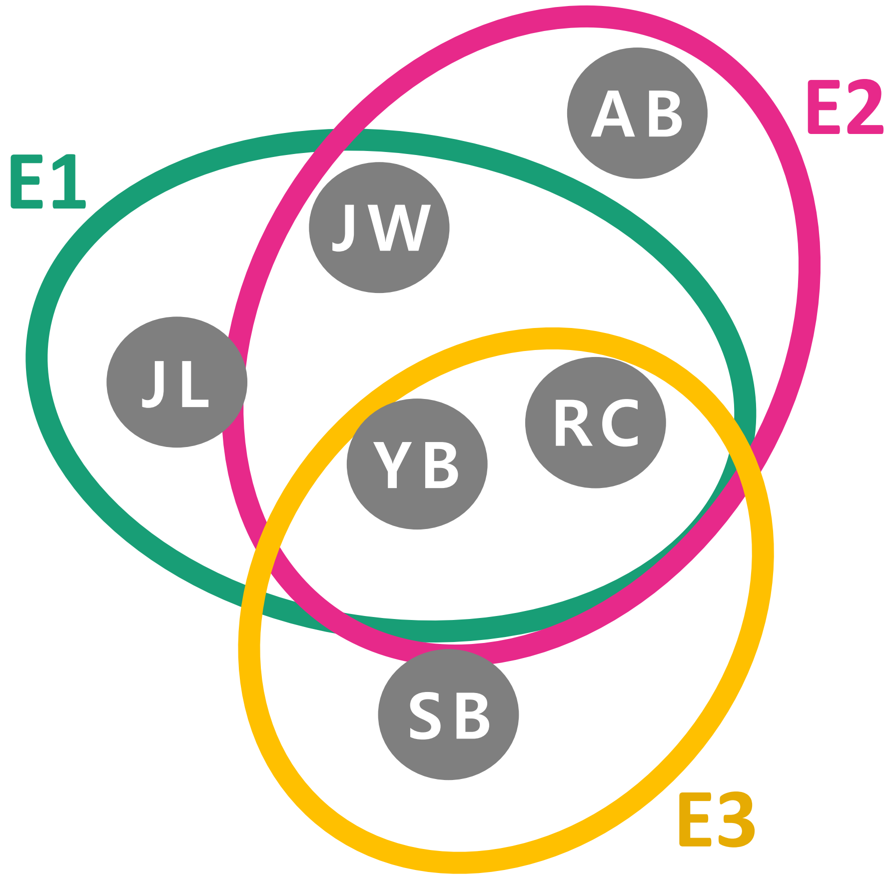
Y Bengio	(YB)	J Weston	(JW)
J Louradour	(JL)	A Bordes	(AB)
R Collobert	(RC)	S Bengio	(SB)

Publications (Hyperedges)

E1: Curriculum learning
Y Bengio, J Louradour, R Collobert, J Weston - ICML'09

E2: Learning structured embeddings of knowledge bases
A Bordes, J Weston, R Collobert, Y Bengio - AAAI'11

E3: A parallel mixture of SVMs for very large scale
R Collobert, S Bengio, Y Bengio – NIPS'01



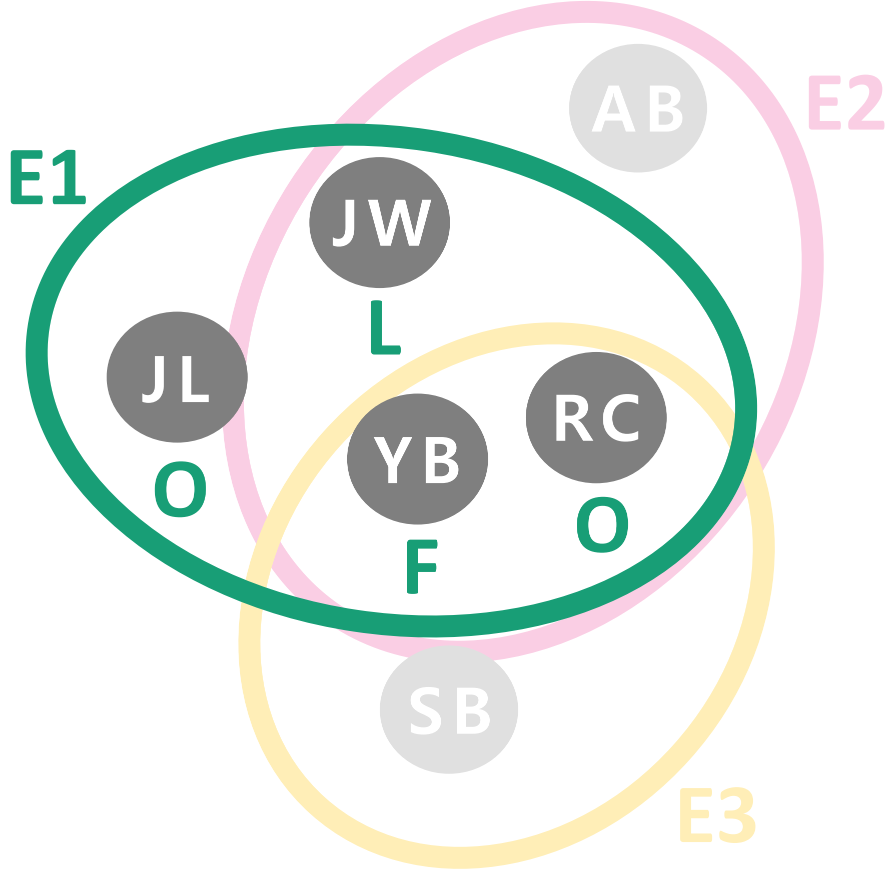
Node Labels are Edge-Dependent

<u>Publications (Hyperedges)</u>

E1: Curriculum learning
Y Bengio, J Louradour, R Collobert, J Weston - ICML'09

First (F) | Others (O) | Last (L)

The importance or role of a node can vary depending on hyperedges!



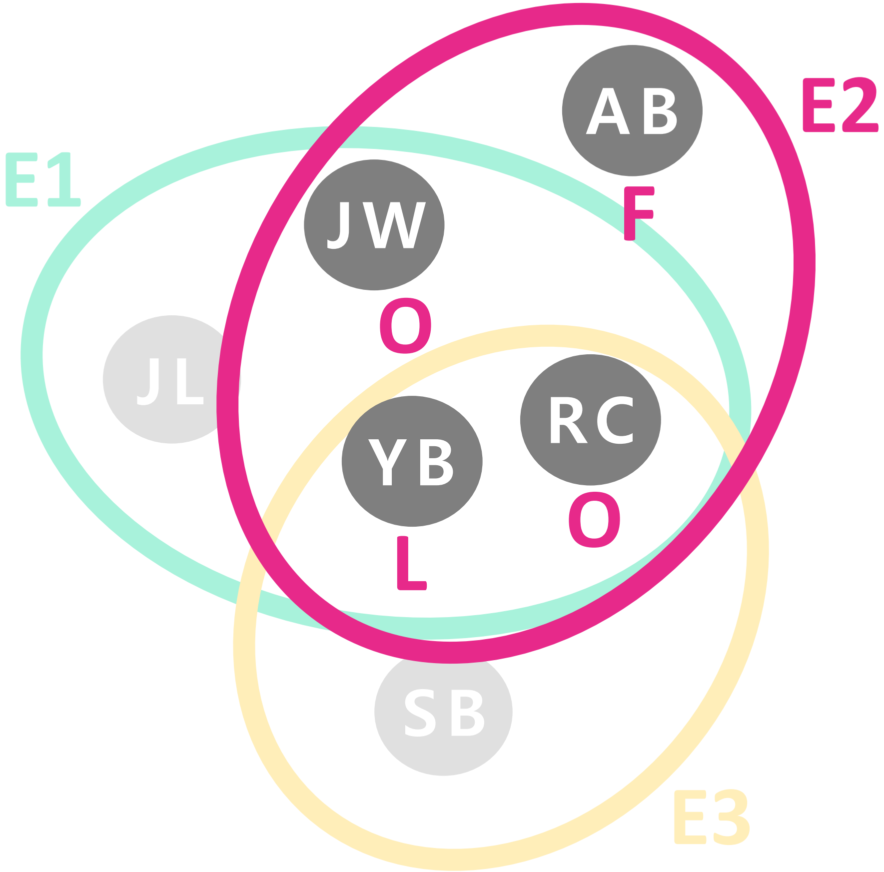
Node Labels are Edge-Dependent

<u>Publications (Hyperedges)</u>

E2: Learning structured embeddings of knowledge bases
A Bordes, J Weston, R Collobert, Y Bengio - AAAI'11

First (F) | Others (O) | Last (L)

The importance or role of a node can vary depending on hyperedges!

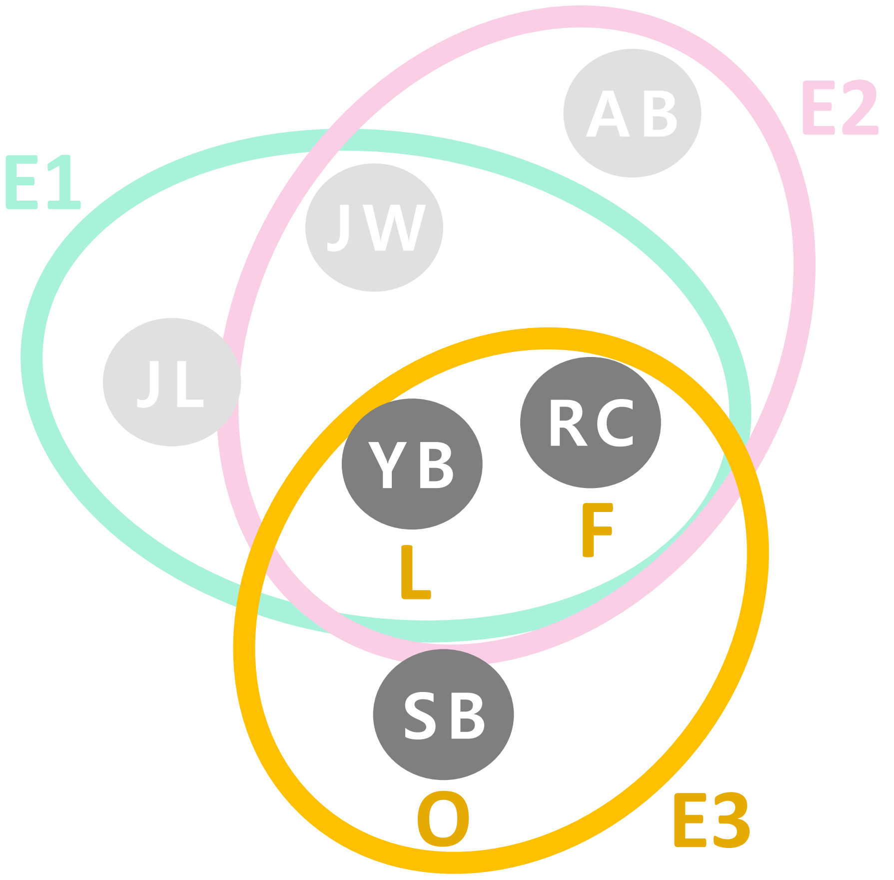


Node Labels are Edge-Dependent

Publications (Hyperedges)

E3: A parallel mixture of SVMs for very large scale
R Collobert, S Bengio, Y Bengio – NIPS’01

First (F) Others (O) Last (L)



The importance or role of a node can vary depending on hyperedges!

Roadmap

1. Introduction
2. **Problem Formulation <<**
3. WHATsNET: Proposed Approach
4. Evaluation
5. Conclusions

Problem Formulation

Edge-Dependent Node Classification

Given

- A hypergraph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$: a node set $\mathcal{V}$ and a hyperedge set $\mathcal{E}$
- A set of edge-dependent node labels in $\mathcal{E}' \subset \mathcal{E}$:

$$y_{v,e}, \forall v \in e \,, \forall e \in \mathcal{E}'$$

- (Optionally) a node feature matrix X

Aim to *accurately predict* the unknown edge-dependent node labels in $\mathcal{E} \backslash \mathcal{E}'$:

$$y_{v,e}, \forall v \in e, \forall e \in \mathcal{E} \backslash \mathcal{E}'$$

Edge-Dependent Node Classification

Co-Authorship

Publication (Hyperedge E1)

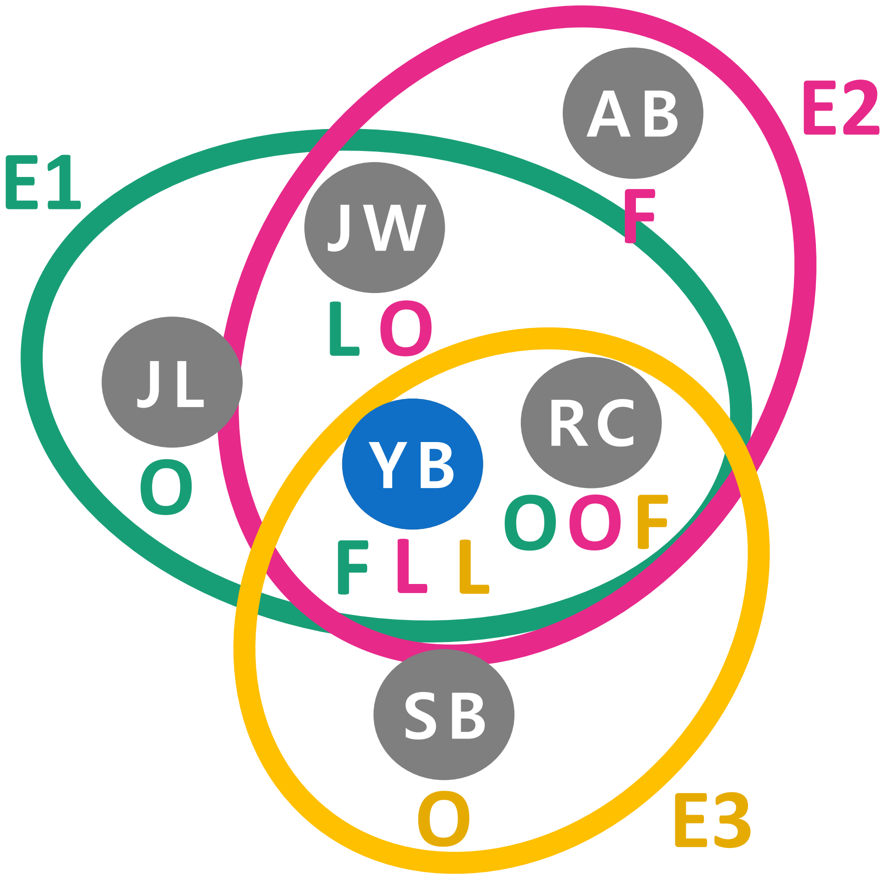
E1: Curriculum learning
Y Bengio, J Louradour, R Collobert, J Weston - ICML'09

Publication (Hyperedge E2)

E2: Learning structured embeddings of knowledge bases
A Bordes, J Weston, R Collobert, Y Bengio - AAAI'11

Publication (Hyperedge E3)

E3: A parallel mixture of SVMs for very large scale
R Collobert, S Bengio, Y Bengio – NIPS'01

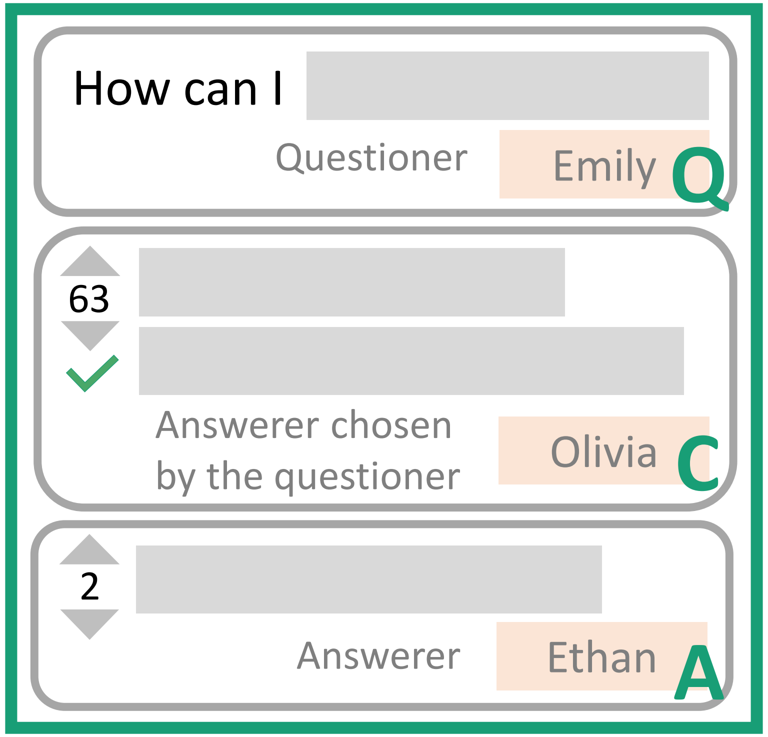


Labels: First, Last, or Others

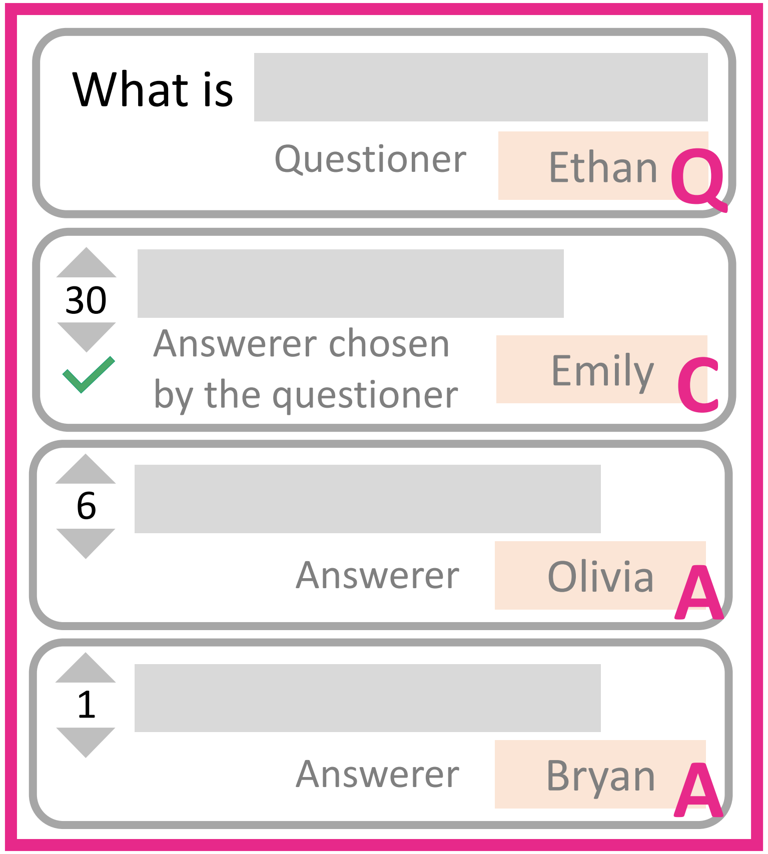
Edge-Dependent Node Classification

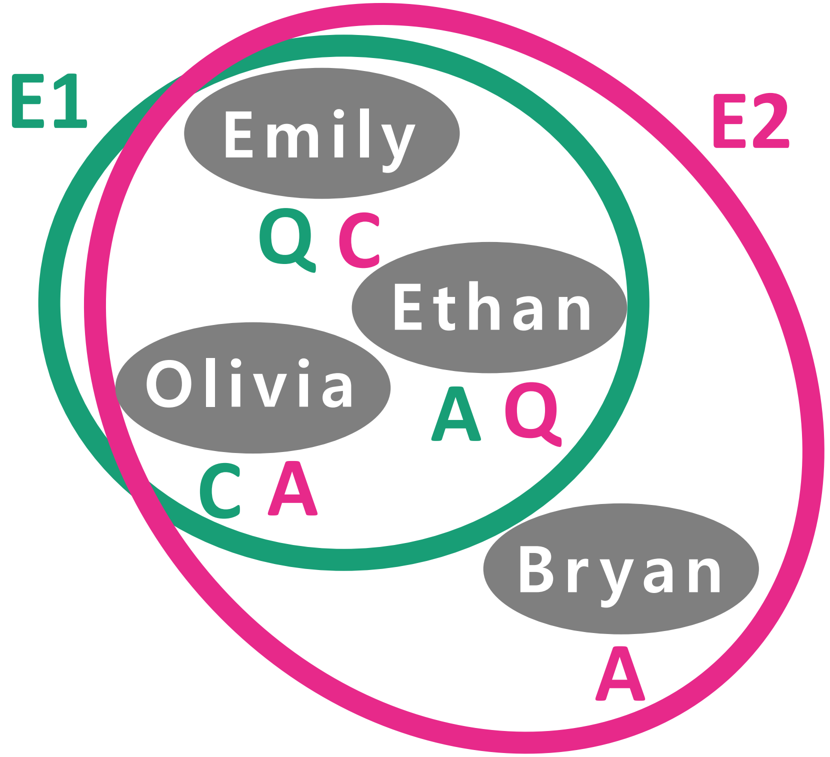
Online Q&A Platforms

Post (Hyperedge E1)



Post (Hyperedge E2)

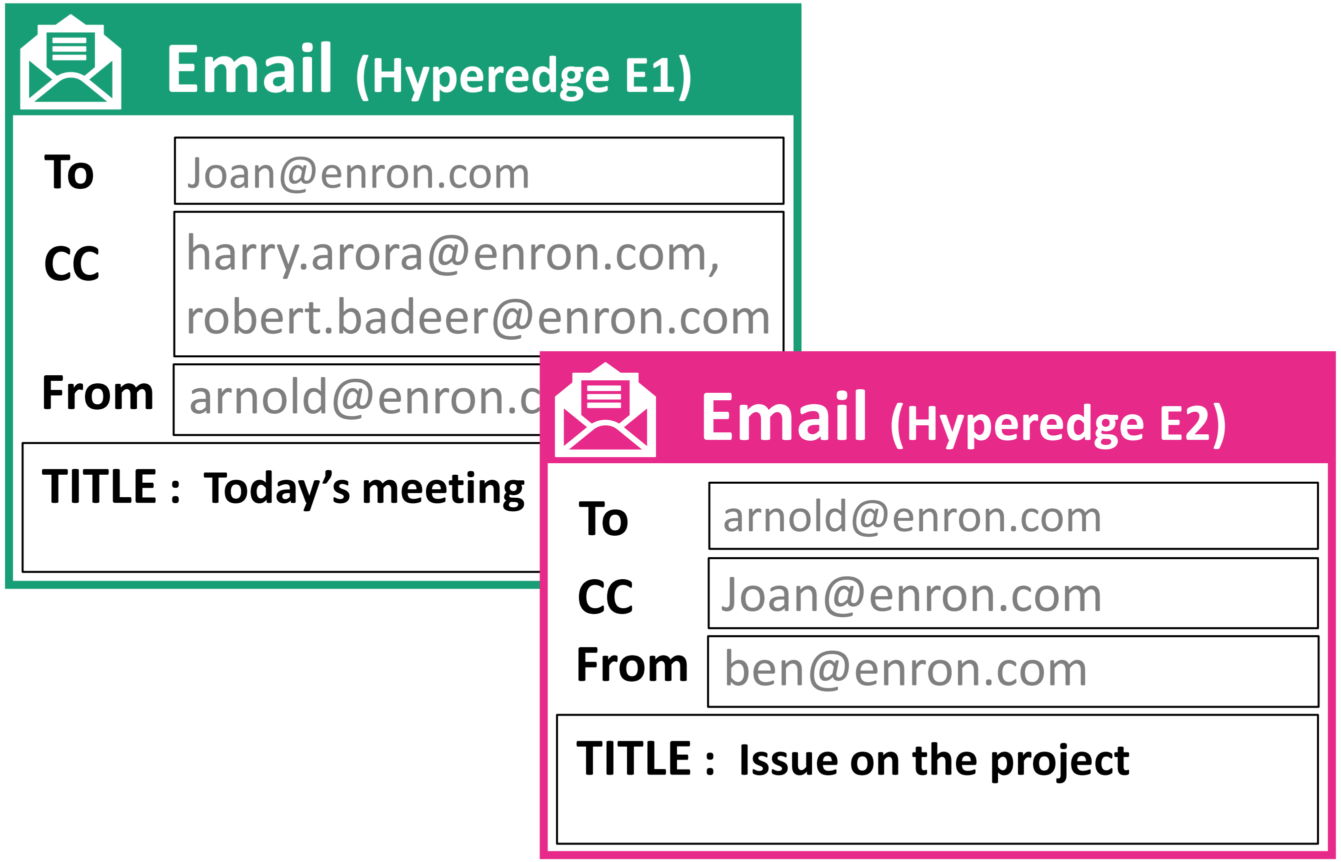


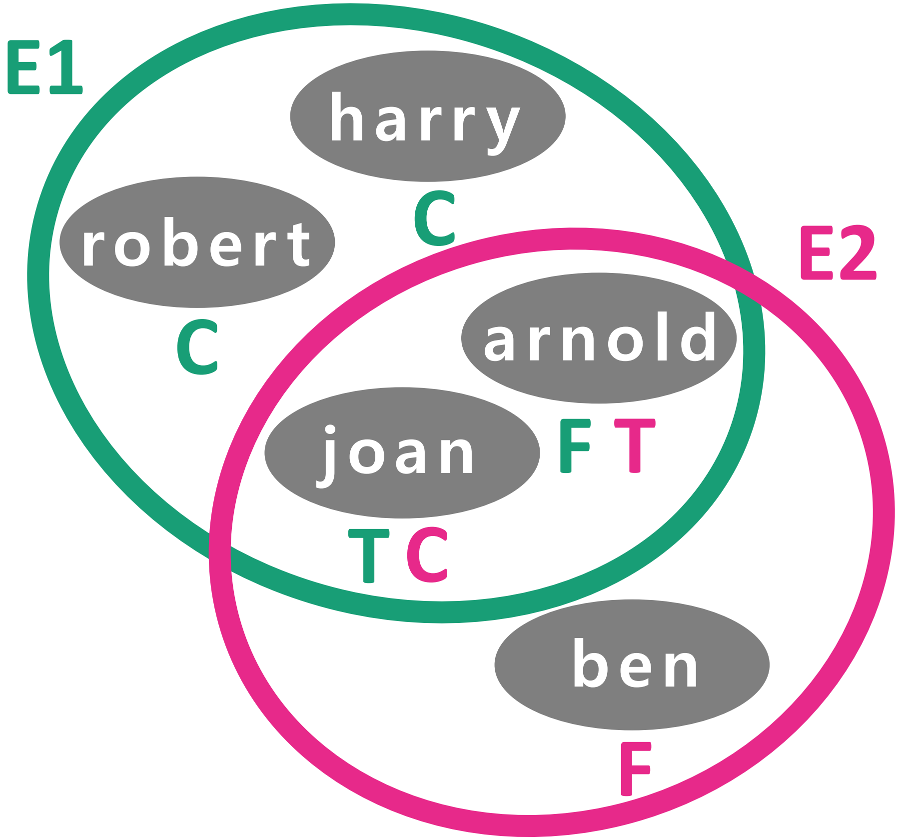


Labels: Questioner, Chosen Answerer, or Answerers

Edge-Dependent Node Classification

Emails



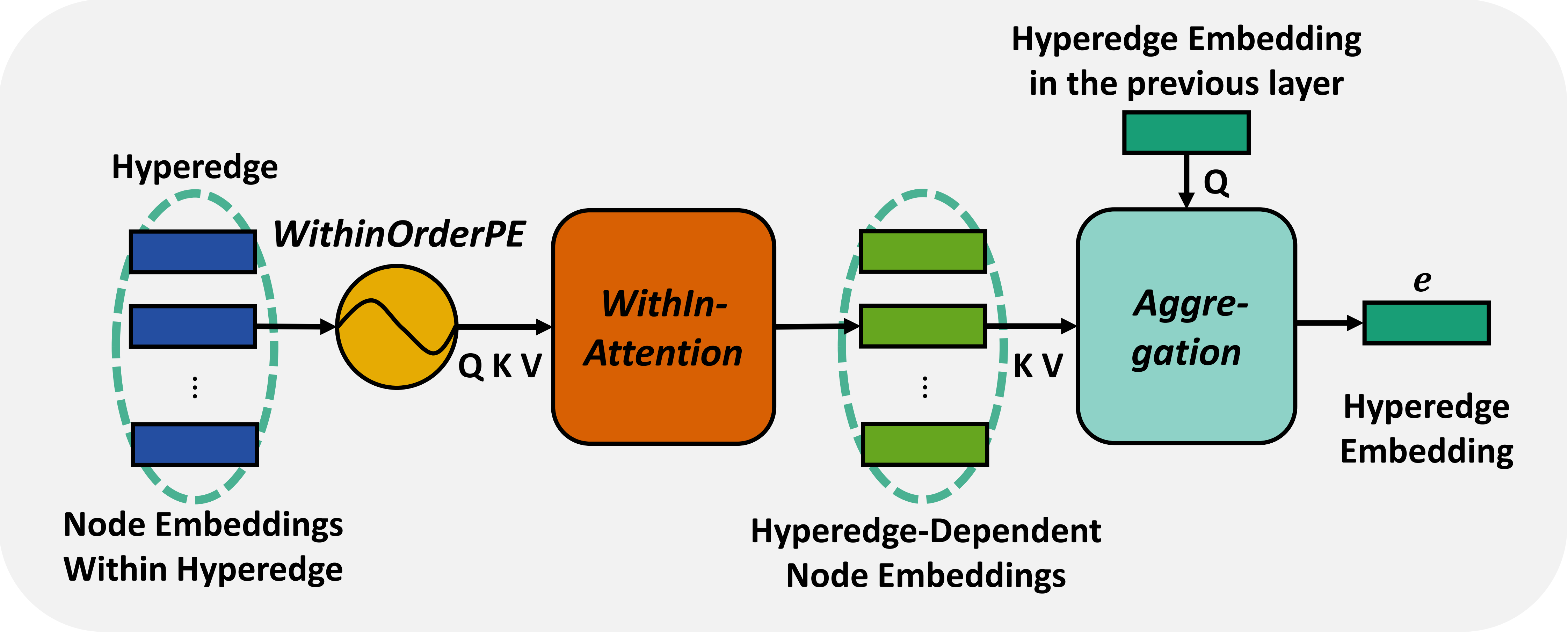


Labels: To, From, or Ccs

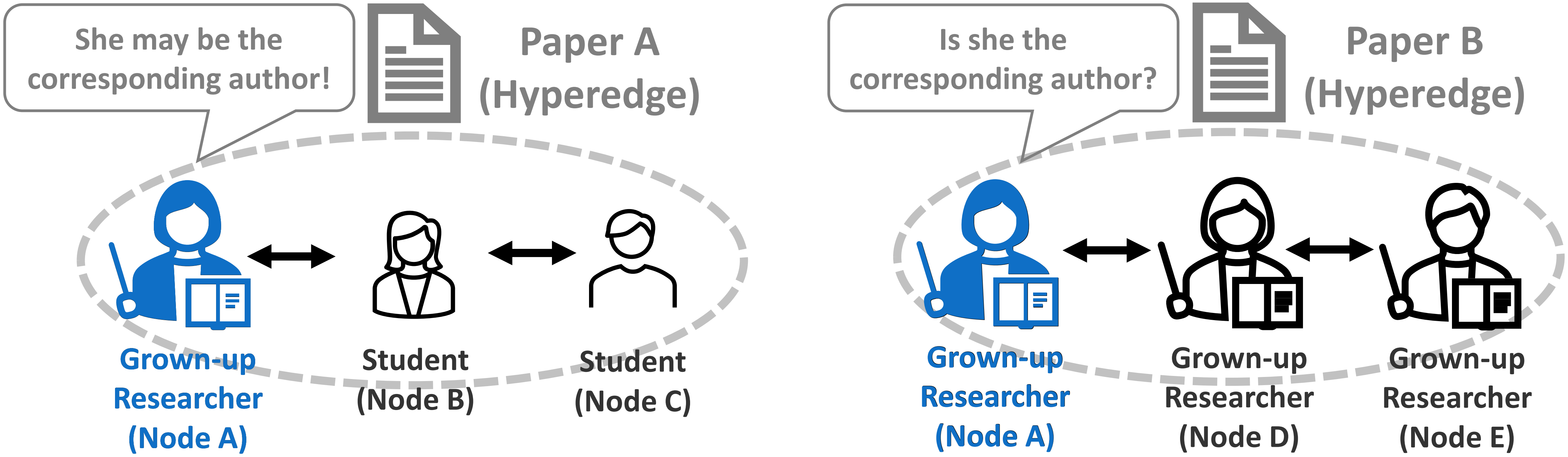
Roadmap

1. Introduction
2. Problem Formulation
3. **WHATsNET: Proposed Approach <<**
4. Evaluation
5. Conclusions

Proposed Model: WHATsNET

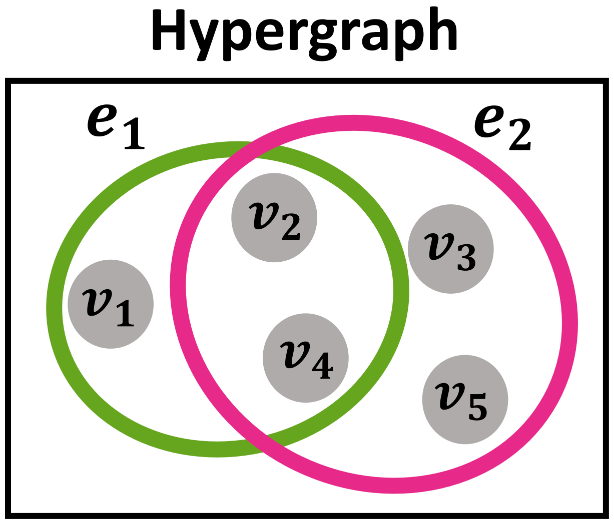


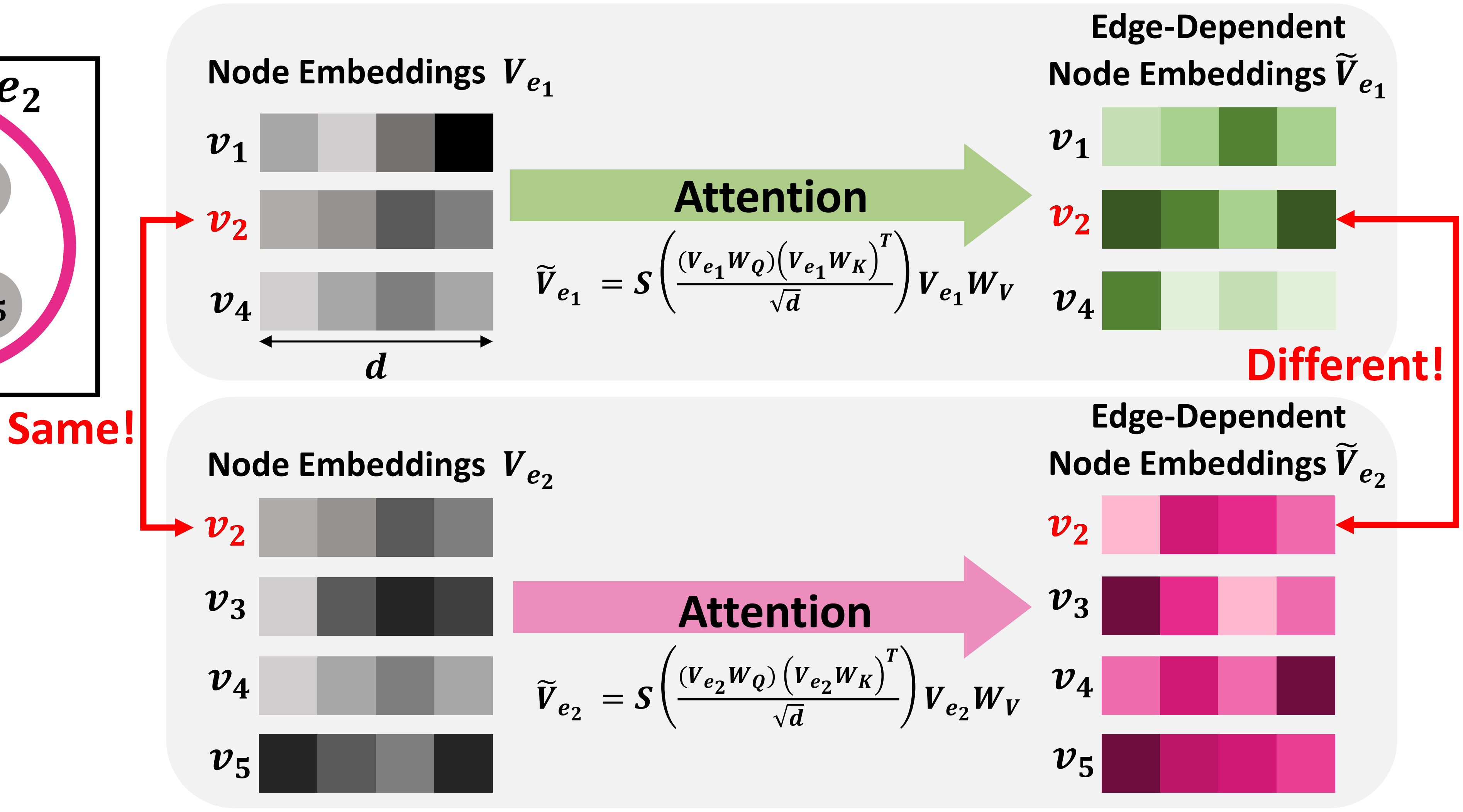
WithInATT: Attention to Other Nodes within Hyperedges



The node label is shaped by the other nodes within the same hyperedge

WithInATT: Attention to Other Nodes within Hyperedges

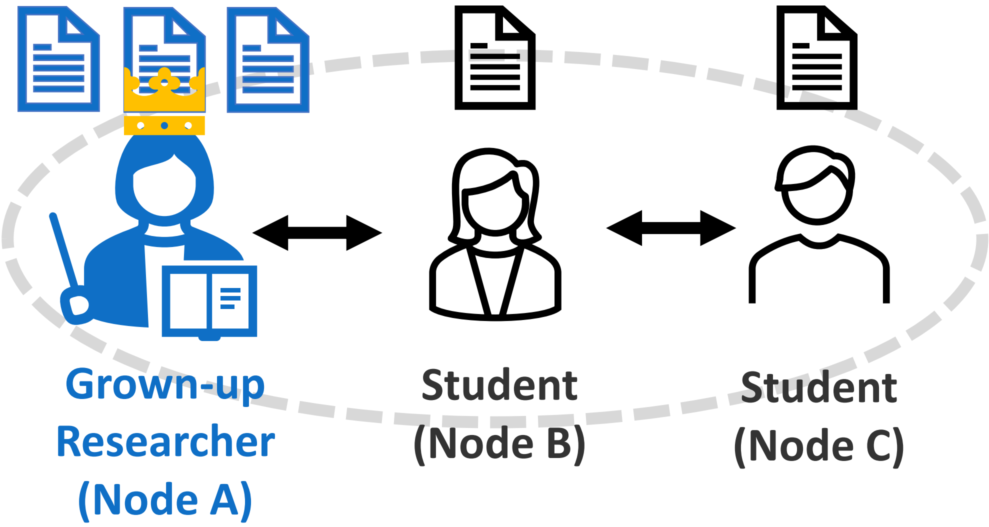




S: Softmax Function
W_Q: Query weight matrix
W_K: Key weight matrix
W_V: Value weight matrix
d: Hidden dimension

WithinOrderPE: Using Centrality for Positional Encoding

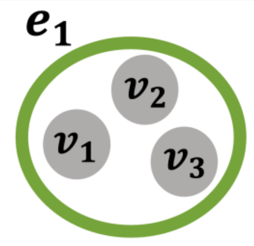
Papers = Degree Centrality



The author with the highest centrality is more likely to be the corresponding author!

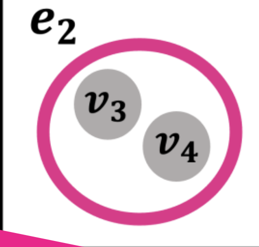
WithinOrderPE: Using Centrality for Positional Encoding

	Centrality Measure 1	Centrality Measure 2	Centrality Measure 3
Node v_1	3	0.7 3th smallest	0.15
Node v_2	1	0.3 1st smallest in e_1	0.15
Node v_3	3	0.5 2ndsmallest	0.15
Node v_4	2	0.5 same in e_2	0.2
⋮	⋮	⋮	⋮



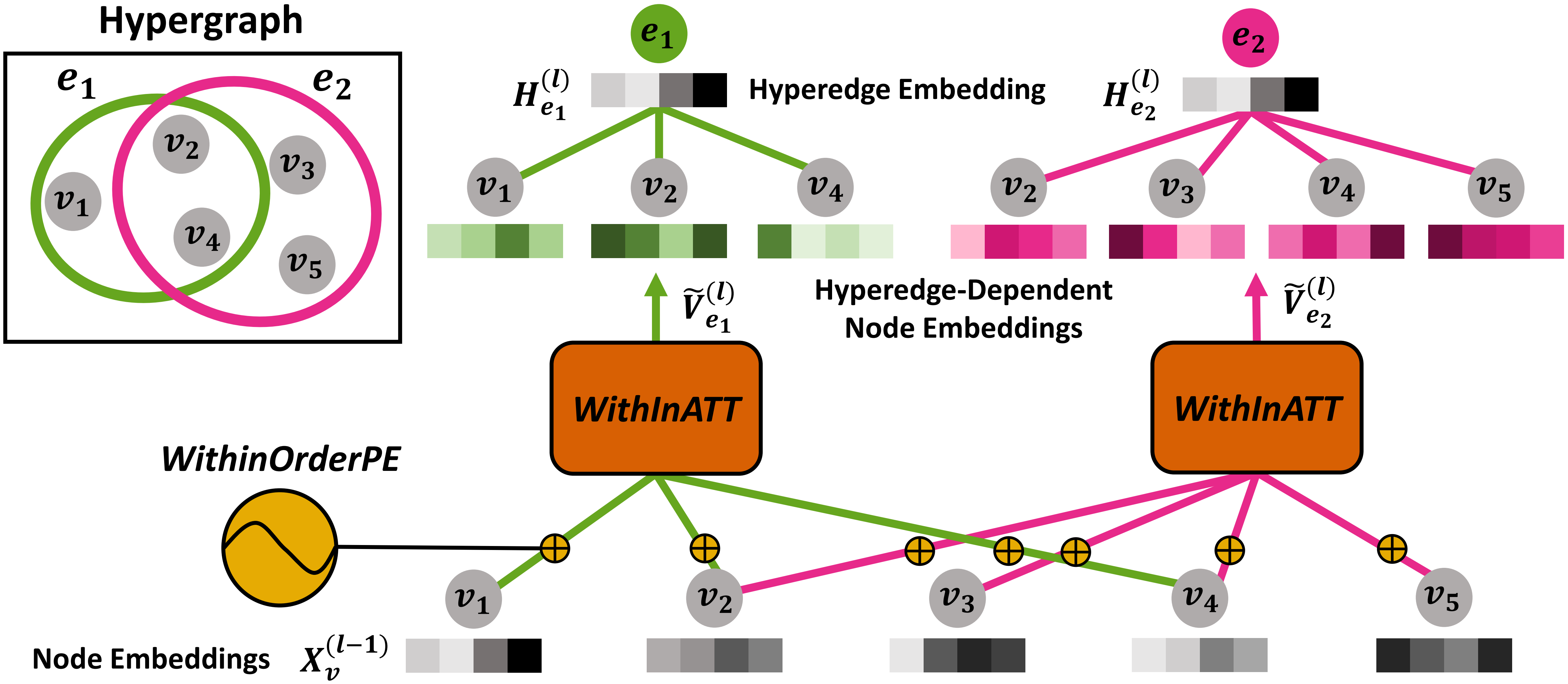
	Within-OrderPE 1	Within-OrderPE 2	Within-OrderPE 3
Node v_1	$2/3$	$\mathbf{3}/(3 = \lvert e_1 \rvert)$	$1/3$
Node v_2	[illegible]	$\mathbf{1}/3$	$1/3$
Node v_3	$2/3$	$\mathbf{2}/3$	$1/3$

Different!



	Within-OrderPE 1	Within-OrderPE 2	Within-OrderPE 3
Node v_3	$2/(2 = \lvert e_2 \rvert)$	$\mathbf{1}/2$	$1/2$
Node v_4	$1/2$	$\mathbf{1}/2$	$2/2$

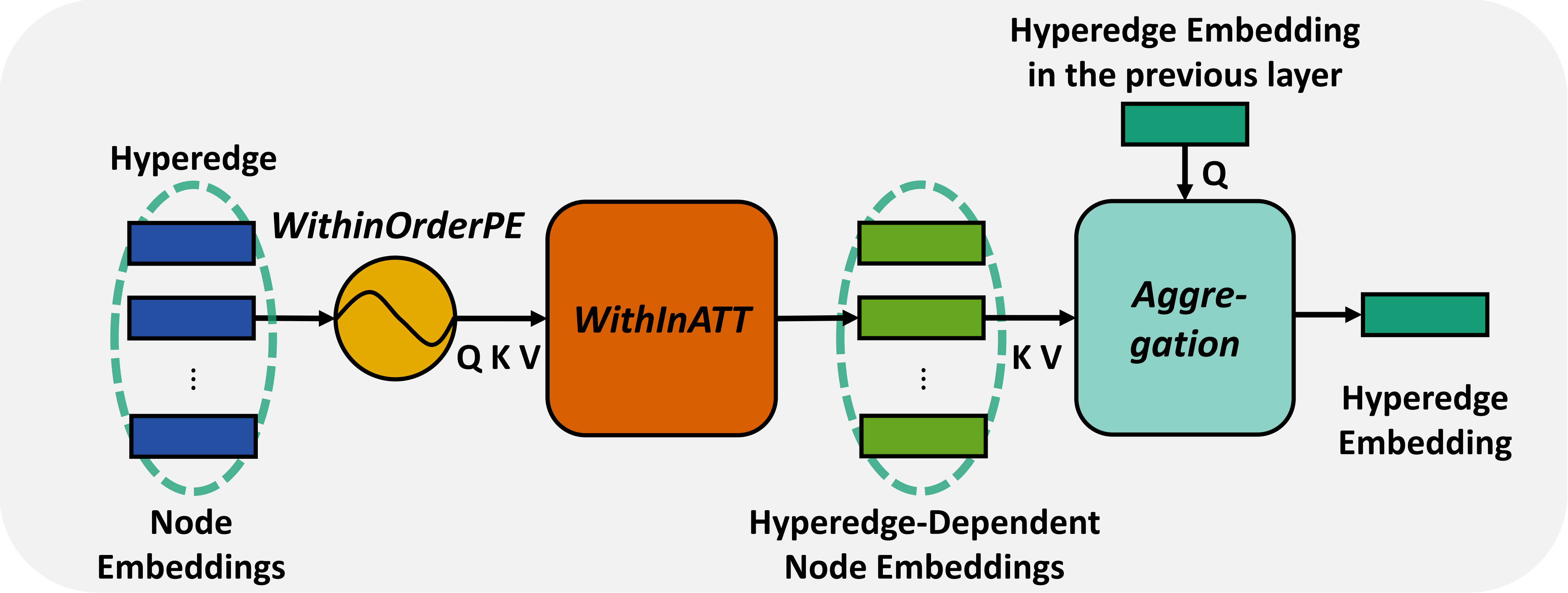
WHATsNET: Our Final Model



WHATsNET: Our Final Model

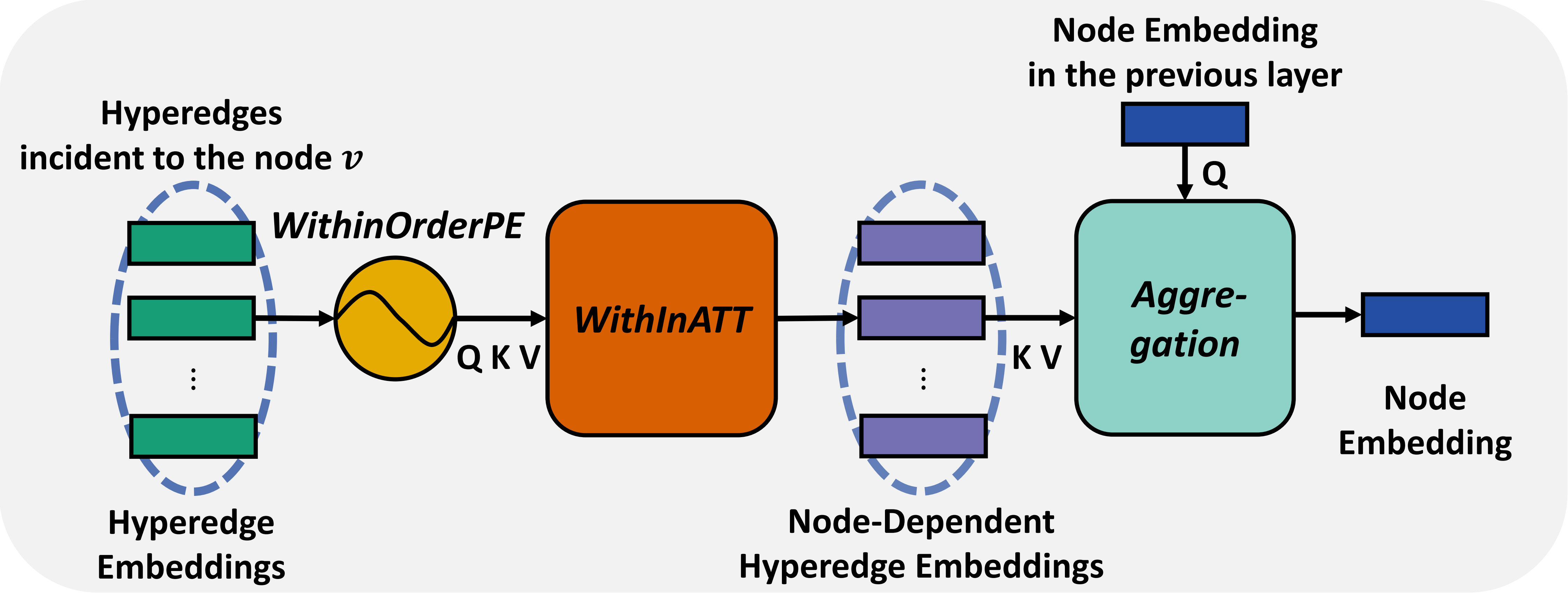
First Update Step) Node→Hyperedge:

Update each hyperedge embedding using node embeddings within it



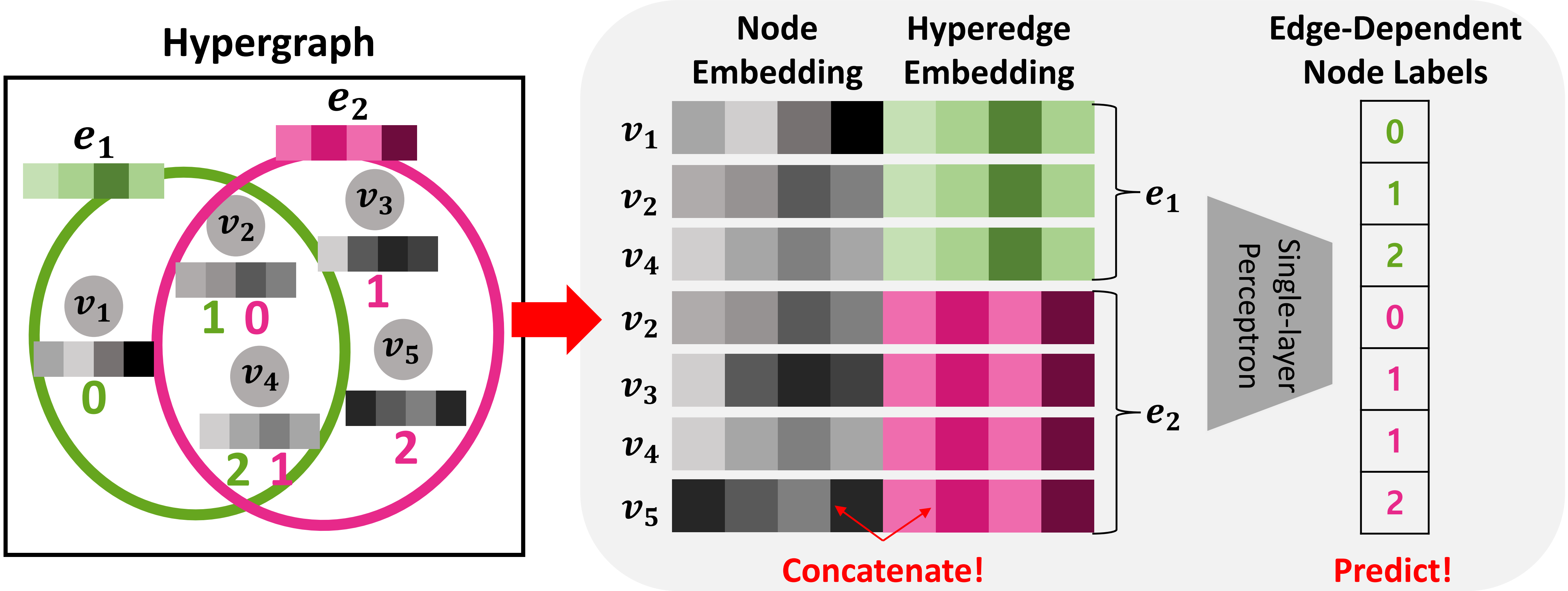
WHATsNET: Our Final Model

Second Update Step) Hyperedge→Node:
Update each node embedding using hyperedge embeddings including it



WHATsNET: Our Final Model

Classifier) Edge-dependent node labels are predicted using the concatenation of node and hyperedge embeddings from the last layer of WHATsNET

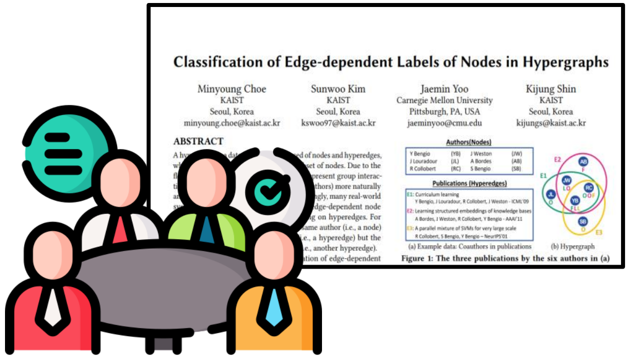


Roadmap

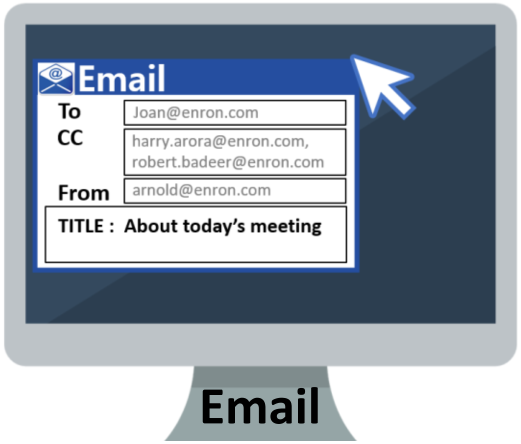
1. Introduction
2. Problem Formulation
3. WHATsNET: Proposed Approach
4. **Evaluation <<**
5. Conclusions

Experiment Setting

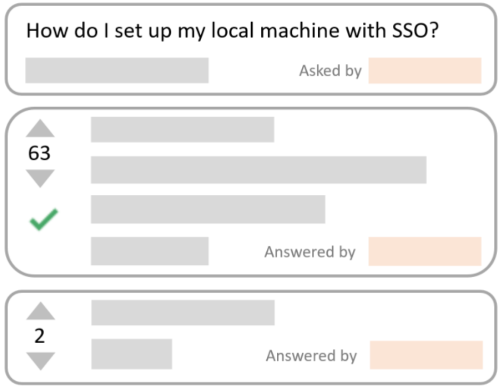
- **Six real-world datasets** from three domains
  - Co-Authorship, Email, and Online Q&A Platform
- **Eight baseline approaches** of hypergraph neural networks
  - HNHN, HGNN, HCHA, HAT, UniGCNII, HNN, HST, and AST
- **Three applications** where edge-dependent node labels are useful
  - Ranking Aggregation, Clustering, and Product Return Prediction



Co-Authorship



Email



Q&A Platform

Q1. Edge-Dependent Node Classification

Q1. Does *WHATsNet* accurately predict the edge-dependent labels of nodes?

A1. WHATsNET *consistently* outperforms 10 competitors in classifying edge-dependent labels of nodes across all 6 datasets!

Results of Edge-Dependent Node Classification

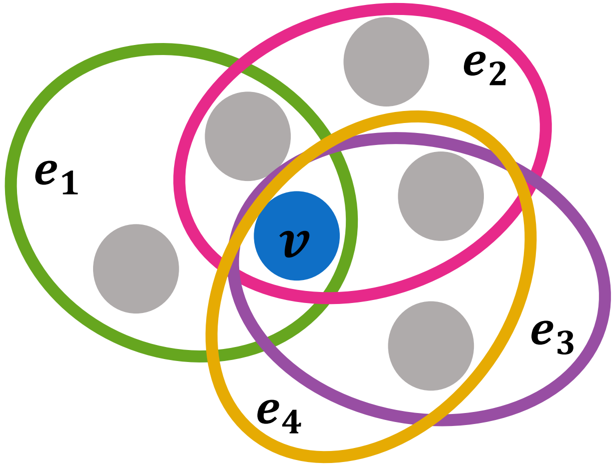
Dataset	Metric	BaselineU	BaselineP	HNHN	HGNN	HCHA	HAT	UniGCNII	HNN	HST	AST	WHATsNET
Coauth-DBLP	MicroF1	0.333 ± 0.001	0.346 ± 0.001	0.486 ± 0.004	0.540 ± 0.004	0.451 ± 0.007	0.503 ± 0.004	0.497 ± 0.003	0.488 ± 0.006	0.564 ± 0.004	0.495 ± 0.038	**0.605** ± 0.002
	MacroF1	0.330 ± 0.001	0.332 ± 0.001	0.478 ± 0.008	0.519 ± 0.002	0.334 ± 0.048	0.483 ± 0.006	0.476 ± 0.002	0.482 ± 0.006	0.549 ± 0.003	0.487 ± 0.040	**0.595** ± 0.002
Coauth-AMiner	MicroF1	0.334 ± 0.000	0.339 ± 0.000	0.520 ± 0.002	0.566 ± 0.002	0.468 ± 0.020	0.543 ± 0.002	0.520 ± 0.001	0.543 ± 0.002	0.596 ± 0.007	0.577 ± 0.005	**0.630** ± 0.005
	MacroF1	0.332 ± 0.000	0.333 ± 0.000	0.514 ± 0.002	0.551 ± 0.004	0.447 ± 0.040	0.533 ± 0.003	0.507 ± 0.001	0.533 ± 0.002	0.583 ± 0.008	0.570 ± 0.002	**0.623** ± 0.007
Email-Enron	MicroF1	0.334 ± 0.001	0.439 ± 0.001	0.738 ± 0.028	0.725 ± 0.004	0.666 ± 0.010	0.817 ± 0.001	0.734 ± 0.010	0.763 ± 0.003	0.779 ± 0.067	0.796 ± 0.014	**0.826** ± 0.001
	MacroF1	0.300 ± 0.001	0.333 ± 0.001	0.637 ± 0.023	0.674 ± 0.003	0.464 ± 0.002	0.753 ± 0.004	0.656 ± 0.010	0.679 ± 0.007	0.681 ± 0.123	0.719 ± 0.020	**0.760** ± 0.004
Email-Eu	MicroF1	0.500 ± 0.001	0.525 ± 0.001	0.643 ± 0.004	0.633 ± 0.001	0.620 ± 0.000	0.669 ± 0.001	0.630 ± 0.005	OutOfMemory	**0.671** ± 0.001	0.666 ± 0.005	**0.671** ± 0.000
	MacroF1	0.493 ± 0.001	0.499 ± 0.001	0.552 ± 0.014	0.533 ± 0.008	0.497 ± 0.001	0.638 ± 0.002	0.565 ± 0.013	OutOfMemory	0.640 ± 0.002	0.624 ± 0.021	**0.646** ± 0.003
Stack-Biology	MicroF1	0.335 ± 0.000	0.368 ± 0.001	0.640 ± 0.005	0.689 ± 0.002	0.589 ± 0.007	0.661 ± 0.005	0.610 ± 0.004	0.618 ± 0.015	0.694 ± 0.002	0.571 ± 0.054	**0.742** ± 0.003
	MacroF1	0.326 ± 0.000	0.334 ± 0.003	0.592 ± 0.006	0.624 ± 0.007	0.465 ± 0.060	0.606 ± 0.005	0.433 ± 0.007	0.568 ± 0.013	0.631 ± 0.006	0.446 ± 0.081	**0.686** ± 0.004
Stack-Physics	MicroF1	0.333 ± 0.001	0.370 ± 0.000	0.506 ± 0.053	0.686 ± 0.004	0.622 ± 0.003	0.708 ± 0.005	0.671 ± 0.022	0.683 ± 0.005	0.755 ± 0.010	0.728 ± 0.039	**0.770** ± 0.003
	MacroF1	0.322 ± 0.001	0.332 ± 0.000	0.422 ± 0.043	0.630 ± 0.002	0.481 ± 0.007	0.643 ± 0.009	0.492 ± 0.016	0.617 ± 0.005	0.666 ± 0.013	0.646 ± 0.046	**0.707** ± 0.004

Q2. Node Label Distribution Preservation

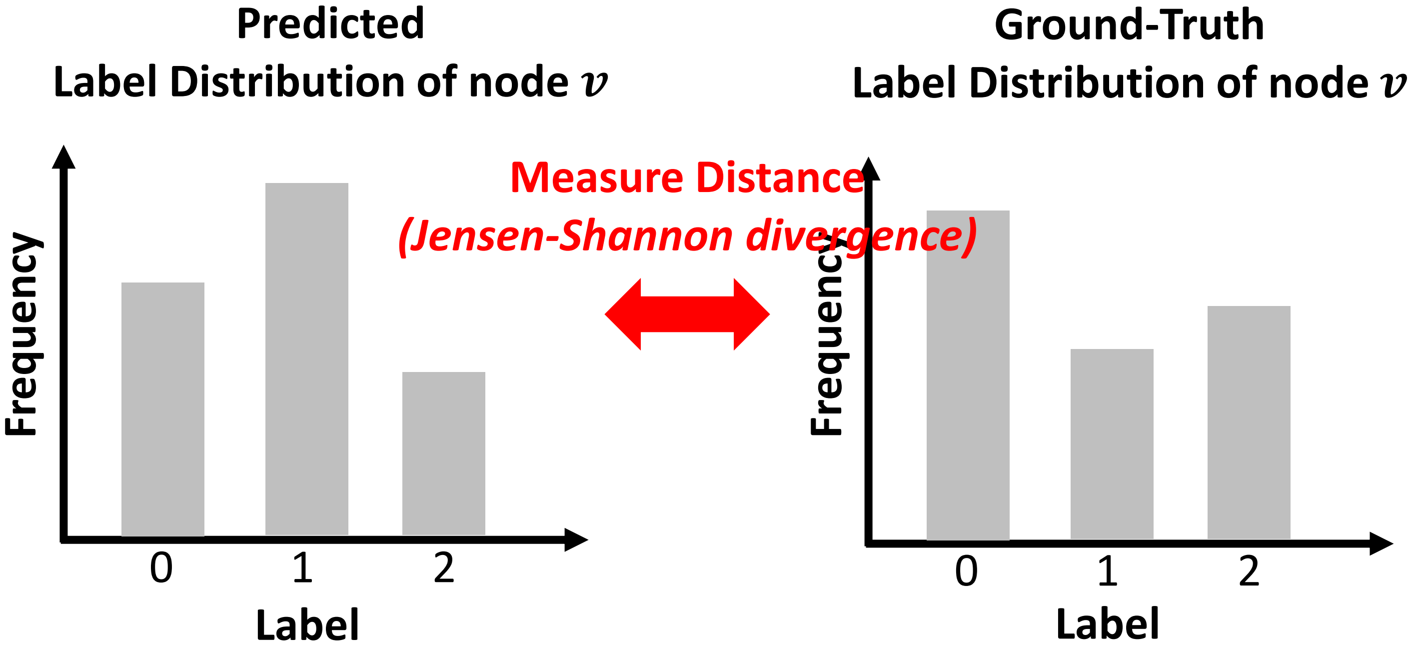
Q2. Does *WHATsNet* classify the same node differently depending on hyperedges?

To assess this, we examine *Node-Level Label Distribution*

= How many times each node has each label



	e_1	e_2	e_3	e_4
Label of v	**0**	**1**	**2**	**1**



Q2. Node Label Distribution Preservation

Q2. Does *WHATsNet* classify the same node differently depending on hyperedges?

A2. WHATsNet preserves well the ground-truth node-level label distributions because it shows the smallest distance between ground-truth and predicted node l abel distribution

Jensen-Shannon divergence (JSD) between Ground-trugh and Predicted Node Label Distributions

Dataset	BaselineU	BaselineP	HNHN	HGNN	HCHA	HAT	UniGCNII	HNN	HST	AST	WHATsNet
Coauth-DBLP	0.532 ± 0.002	0.518 ± 0.002	0.450 ± 0.002	0.394 ± 0.005	0.450 ± 0.002	0.429 ± 0.004	0.449 ± 0.006	0.450 ± 0.003	0.388 ± 0.006	0.453 ± 0.038	**0.350** ± 0.002
Coauth-AMiner	0.529 ± 0.000	0.523 ± 0.000	0.440 ± 0.003	0.372 ± 0.002	0.462 ± 0.002	0.414 ± 0.003	0.424 ± 0.002	0.411 ± 0.003	0.356 ± 0.005	0.382 ± 0.009	**0.328** ± 0.004
Email-Enron	0.486 ± 0.002	0.395 ± 0.001	0.162 ± 0.003	0.212 ± 0.007	0.291 ± 0.007	0.157 ± 0.004	0.187 ± 0.002	0.205 ± 0.002	0.302 ± 0.233	0.178 ± 0.019	**0.136** ± 0.001
Email-Eu	0.199 ± 0.006	0.164 ± 0.004	0.232 ± 0.009	0.291 ± 0.001	0.268 ± 0.002	0.154 ± 0.004	0.300 ± 0.004	OutOfMemory	0.158 ± 0.005	0.168 ± 0.015	**0.151** ± 0.009
Stack-Biology	0.536 ± 0.001	0.467 ± 0.003	0.266 ± 0.007	0.202 ± 0.003	0.237 ± 0.016	0.235 ± 0.006	0.263 ± 0.004	0.311 ± 0.026	0.200 ± 0.002	0.259 ± 0.022	**0.152** ± 0.002
Stack-Physics	0.532 ± 0.001	0.482 ± 0.001	0.286 ± 0.036	0.219 ± 0.008	0.289 ± 0.004	0.227 ± 0.006	0.285 ± 0.050	0.292 ± 0.005	0.162 ± 0.008	0.185 ± 0.021	**0.141** ± 0.003

Q3. Ablation Study of WHATsNet

Q3. Does each component make a meaningful contribution to the performance?

A3-(a). Importance of WithinATT

WHATsNet consistently outperforms the variant without WithinATT.

A3-(b). Improvement by WithinOrderPE

WHATsNet also performs better than the variant without WithinOrderPE

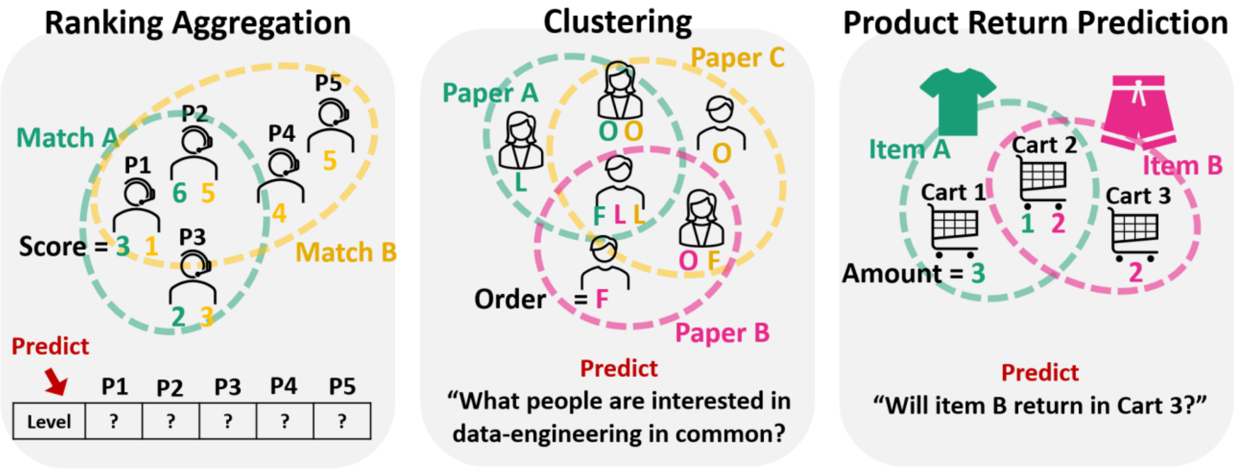
Dataset	Metric	w/o WithinATT	w/o WithinOrderPE	WHATsNet
Coauth-DBLP	MicroF1	0.581 ± 0.004	0.591 ± 0.003	**0.605** ± 0.002
	MacroF1	0.577 ± 0.003	0.584 ± 0.003	**0.595** ± 0.002
Coauth-AMiner	MicroF1	0.604 ± 0.010	0.583 ± 0.095	**0.630** ± 0.005
	MacroF1	0.592 ± 0.013	0.536 ± 0.174	**0.623** ± 0.007
Email-Enron	MicroF1	0.812 ± 0.008	0.825 ± 0.001	**0.826** ± 0.001
	MacroF1	0.747 ± 0.014	**0.762** ± 0.004	0.760 ± 0.004
Email-Eu	MicroF1	0.651 ± 0.019	0.670 ± 0.000	**0.671** ± 0.000
	MacroF1	0.630 ± 0.018	0.638 ± 0.002	**0.646** ± 0.003
Stack-Biology	MicroF1	0.723 ± 0.002	0.732 ± 0.002	**0.742** ± 0.003
	MacroF1	0.656 ± 0.005	0.672 ± 0.004	**0.686** ± 0.004
Stack-Physics	MicroF1	0.752 ± 0.005	0.765 ± 0.002	**0.770** ± 0.003
	MacroF1	0.675 ± 0.010	0.688 ± 0.008	**0.707** ± 0.004
AVG. Ranking	MicroF1	2.83	2.17	**1.00**
	MacroF1	2.83	2.00	**1.17**

Q4. Usefulness in Downstream Tasks

Q4. How can we apply *WHATsNet* to real-world downstream tasks?

Does WHATsNet show usefulness in these tasks?

A4. In *three* downstream tasks, we compare the performance of the tasks using **(a)** ground-truth edge-dependent node labels, **(b)** labels predicted by WHATs-NET, **(c)** No labels

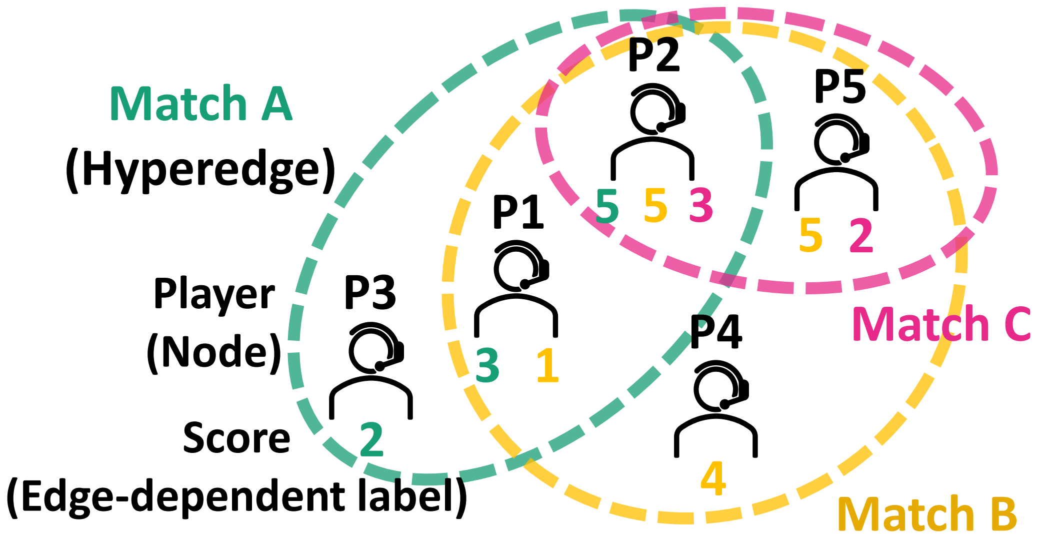


Q4. Usefulness in Downstream Tasks

Q4. How can we apply *WHATsNet* to real-world downstream tasks?

Does WHATsNet show usefulness in these tasks?

A4. Ranking Aggregation Task:



	P1	P2	P3	P4	P5
Match A	3	2	5		
Match B	1	5		4	5
Match C		3			2

Given edge-dependent labels by (a), (b), or (c)

TASK: Predict!

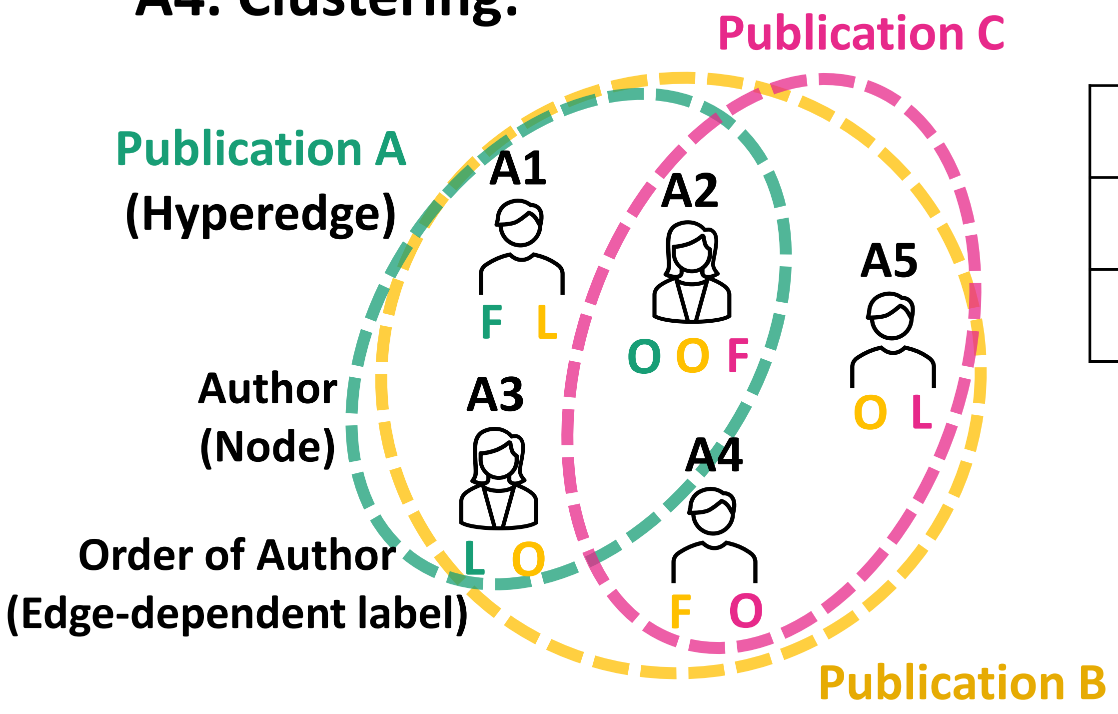
	P1	P2	P3	P4	P5
Level	?	?	?	?	?

Q4. Usefulness in Downstream Tasks

Q4. How can we apply *WHATsNet* to real-world downstream tasks?

Does WHATsNet show usefulness in these tasks?

A4. Clustering:



	A1	A2	A3	A4	A5
Publication A	F	O	L		
Publication B	L	O	O	F	O
Publication C		F		O	L

Given edge-dependent labels by (a), (b), or (c)

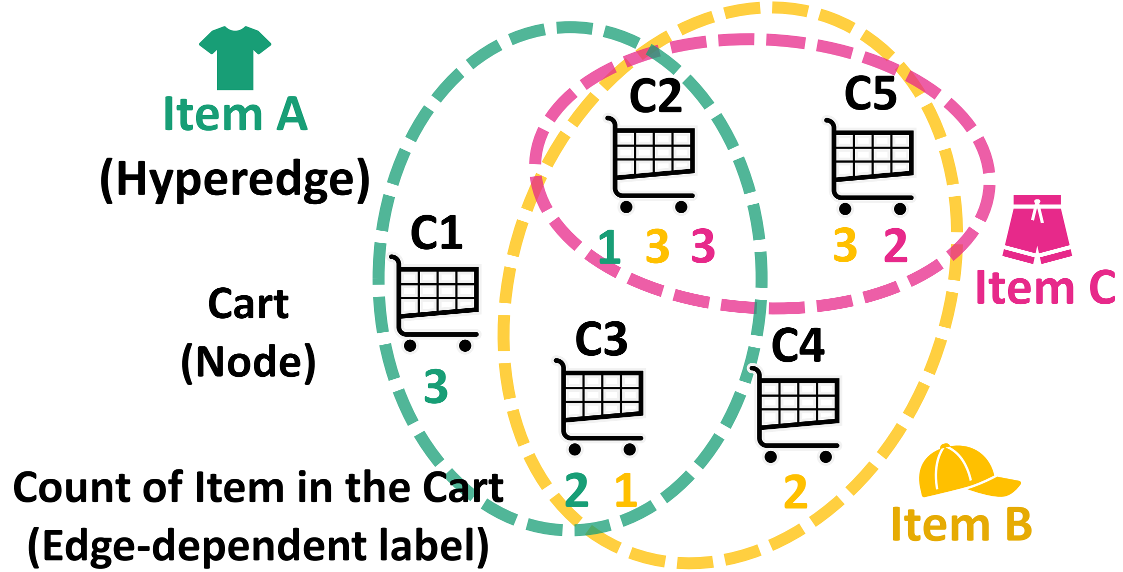
TASK: Cluster publications with the same venue!

Q4. Usefulness in Downstream Tasks

Q4. How can we apply *WHATsNet* to real-world downstream tasks?

Does WHATsNet show usefulness in these tasks?

A4. Product Return Prediction:



	C1	C2	C3	C4	C5
Item A	3	1	2		
Item B	1	3	1	2	3
Item C		3			2

Given edge-dependent labels by (a), (b), or (c)

TASK: Predict the likelihood of products being returned for each cart!

Q4. Usefulness in Downstream Tasks

Q4. How can we apply WHATsNet to real-world downstream tasks?

Does WHATsNet show usefulness in these tasks?

A4. Utilization of edge-dependent node labels predicted by WHATsNet *consistently* leads to performance improvements compared to the results obtained *without* labels.

(a) Ranking Aggregation (Accuracy)

Method	Halo	H-Index
RW [13] w/ Ground Truth	0.711	0.675
RW [13] w/ WHATsNet	**0.714**	0.693
RW [13] w/ HST	0.707	0.695
RW [13] w/ AST	0.706	**0.696**
RW [13] w/o Labels	0.532	0.654

(b) Clustering (NMI: The higher, the better)

Method	DBLP	AMiner
RDC-Spec [23] w/ GroundTruth	**0.221**	**0.359**
RDC-Spec [23] w/ WHATsNet	0.184	0.352
RDC-Spec [23] w/ HST	0.166	0.339
RDC-Spec [23] w/ AST	0.168	0.332
RDC-Spec [23] w/o Labels	0.163	0.338

(c) Product Return Prediction (F1)

Method	Synthetic E-tail
HyperGO [38] w/ GroundTruth	**0.738**
HyperGO [38] w/ WHATsNet	0.723
HyperGO [38] w/ HST	0.724
HyperGO [38] w/ AST	0.721
HyperGO [38] w/o Labels	0.718

Roadmap

1. Introduction
2. Problem Formulation
3. WHATsNET: Proposed Approach
4. Evaluation
5. **Conclusions <<**

Conclusion

- Our contributions are summarized as follows:

✓ **New Problem:** Formulate the edge-dependent node classification problem

✓ **Effective Model:** Propose WHATsNET with WithInATT and WithinOrderPE

✓ **Extensive Experiment:** Show the superiority of WHATsNET over 10 competitors

Code & Dataset: https://github.com/young917/EdgeDependentNodeLabel

Minyoung Choe: minyoung.choe@kaist.ac.kr

Sunwoo Kim: kswoo97@kaist.ac.kr

Jaemin Yoo: jaeminyoo@cmu.edu

Kijung Shin: kijungs@kaist.ac.kr

Minyoung Choe

Sunwoo Kim

Jaemin Yoo

Kijung Shin

Appendix

Table 2: Summary of real-world hypergraphs with edge-dependent labels.

Dataset	Num. of Nodes	Num. of Hyperedges	Max. of Node Deg.	Max. of Hyperedge Size	Sum of Hyperedge Size	Num. of Class 0	Num. of Class 1	Num. of Class 2	Corr. w/ Centrality	Avg. Entropy
Coauth-DBLP	108,484	91,266	236	36	321,011	91,266	138,479	91,266	0.19	0.13
Coauth-AMiner	1,712,433	2,037,605	752	115	5,129,998	2,037,605	1,652,332	1,503,061	0.24	0.13
Email-Enron	21,251	101,124	18,168	948	1,186,521	635,268	450,129	101,124	0.10	0.28
Email-Eu	986	209,508	8,659	59	541,842	209,508	332,334	-	0.24	0.48
Stack-Biology	15,490	26,823	1,318	12	56,257	26,290	18,444	11,523	0.29	0.10
Stack-Physics	80,936	200,811	6,332	48	479,809	194,575	201,121	84,113	0.30	0.12

Appendix

Table 8: Comparison of positional encoding schemes. The symbol $\mathbf{I}_w$ denotes inducing points.

Dataset		w/o PE	Positional Encodings: GraphIT/DK	Positional Encodings: GraphIT/PRWK	Positional Encodings: Shaw/DK	Positional Encodings: Shaw/PRWK	Positional Encodings: LSPE	Positional Encodings: WholeOrderPE	WithinOrderPE: w/o $\mathbf{I}_w$	WithinOrderPE: w/ $\mathbf{I}_w$
Stack Biology	MicroF1	0.732 ± 0.002	0.719 ± 0.006	0.710 ± 0.004	0.731 ± 0.003	0.456 ± 0.014	0.727 ± 0.003	0.732 ± 0.003	**0.737** ± 0.006	**0.737** ± 0.003
	MacroF1	0.672 ± 0.004	0.645 ± 0.012	0.638 ± 0.012	0.667 ± 0.005	0.338 ± 0.038	0.658 ± 0.010	0.669 ± 0.011	**0.680** ± 0.005	0.679 ± 0.007

Table 9: Comparison of node-centrality measures that can be used for WithinOrderPE.

Dataset		Degree	Coreness	Eigenvector	PageRank	H-Coreness	H-Eigenvector	Vector Centrality	WHATsNet	WHATsNet-all
Stack Biology	MicroF1	0.735 ± 0.001	0.739 ± 0.005	0.738 ± 0.003	0.741 ± 0.003	0.738 ± 0.005	0.738 ± 0.004	**0.745** ± 0.002	*0.742* ± 0.003	**0.745** ± 0.002
	MacroF1	0.678 ± 0.004	0.685 ± 0.010	0.681 ± 0.004	0.681 ± 0.004	0.683 ± 0.008	0.681 ± 0.006	*0.687* ± 0.003	0.686 ± 0.004	**0.690** ± 0.005

Table 10: Effect of the number of inducing points and comparison of architectures for hyperedge-to-node message passing.

Dataset		Number of Inducing Points: 2	Number of Inducing Points: 4	Number of Inducing Points: 8	MSG. Passing from Hyperedge to Node: HNHN	MSG. Passing from Hyperedge to Node: HAT	MSG. Passing from Hyperedge to Node: WHATsNet + HNHN	MSG. Passing from Hyperedge to Node: WHATsNet + HAT	MSG. Passing from Hyperedge to Node: WHATsNet
Stack Biology	MicroF1	0.740 ± 0.002	**0.742** ± 0.003	**0.742** ± 0.003	0.640 ± 0.005	0.661 ± 0.005	0.645 ± 0.023	0.670 ± 0.015	**0.742** ± 0.003
	MacroF1	0.682 ± 0.006	0.686 ± 0.004	**0.688** ± 0.002	0.592 ± 0.006	0.606 ± 0.005	0.583 ± 0.029	0.618 ± 0.010	**0.686** ± 0.004

Appendix

Table 11: Effect of inputs to the final classifier

Dataset	Metric	Best Competitor	WHATsNet-IM	WHATsNet
Coauth-DBLP	MicroF1	0.564 ± 0.004	0.602 ± 0.002	**0.605** ± 0.002
	MacroF1	0.549 ± 0.003	0.592 ± 0.002	**0.595** ± 0.002
Coauth-AMiner	MicroF1	0.596 ± 0.007	**0.637** ± 0.003	0.630 ± 0.005
	MacroF1	0.583 ± 0.008	**0.631** ± 0.003	0.623 ± 0.007
Email-Enron	MicroF1	0.779 ± 0.067	**0.858** ± 0.001	0.826 ± 0.001
	MacroF1	0.681 ± 0.123	**0.796** ± 0.004	0.760 ± 0.004
Email-Eu	MicroF1	0.671 ± 0.001	**0.687** ± 0.004	0.671 ± 0.000
	MacroF1	0.640 ± 0.002	**0.660** ± 0.009	0.646 ± 0.003
Stack-Biology	MicroF1	0.694 ± 0.002	0.736 ± 0.003	**0.742** ± 0.003
	MacroF1	0.631 ± 0.006	0.679 ± 0.007	**0.686** ± 0.004
Stack-Physics	MicroF1	0.755 ± 0.010	0.769 ± 0.001	**0.770** ± 0.003
	MacroF1	0.666 ± 0.013	0.692 ± 0.011	**0.707** ± 0.004