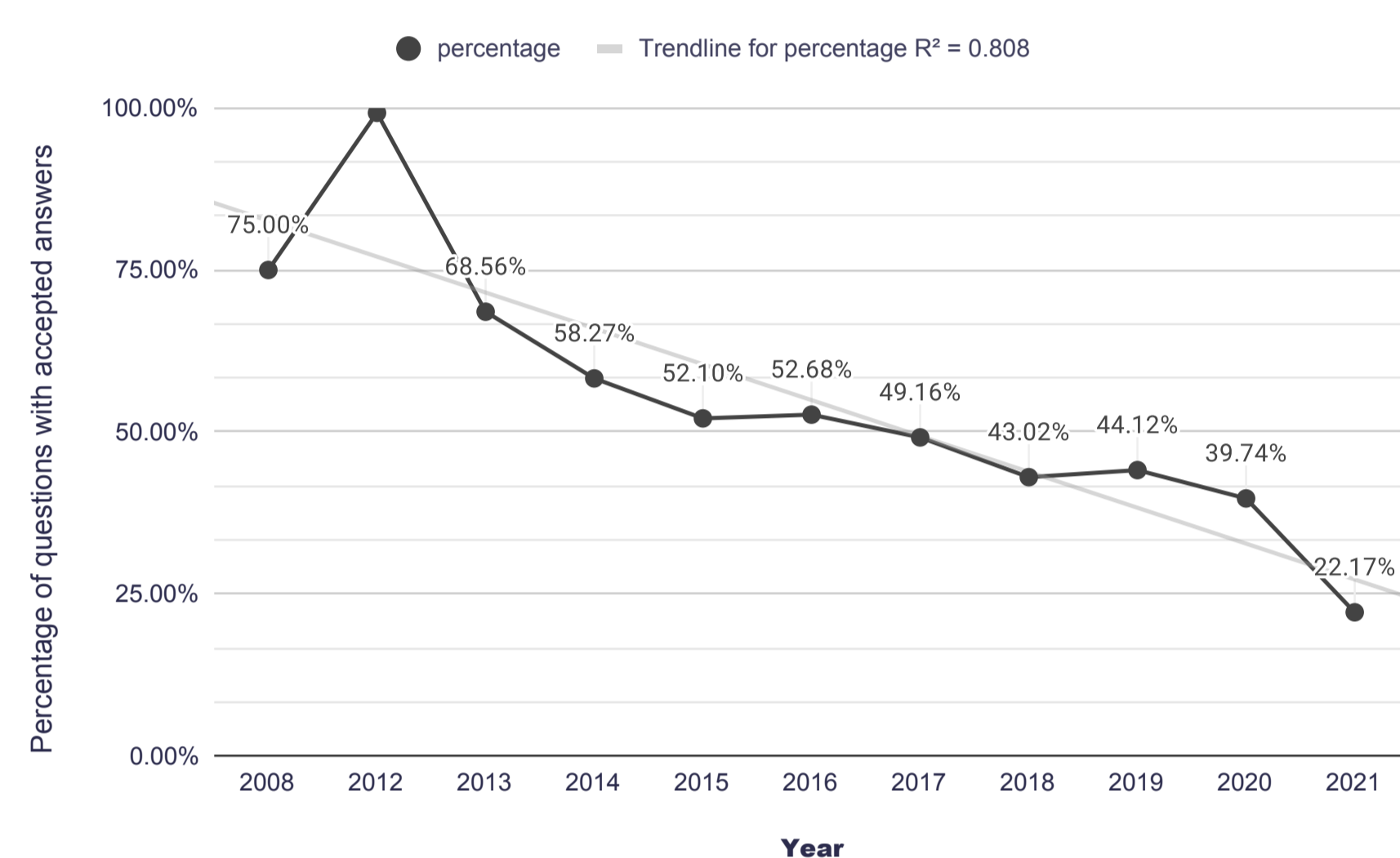


## Work Summary

Many questions on the Stack Exchange (SE) question-answering (Q&A) platform still go unresolved. We have developed a novel approach to identify these unresolved questions by analysing the surrounding communication networks. Using a property graph model (PGM) and graph neural networks (GNNs), our approach captures both communication structure and message content. Tests on historical data from three SE communities show our method outperforms baselines, focusing only on question content. We believe that the introduced approach can help improve the viability of the platform and subsequently improve the knowledge-sharing process facilitated by it. See [AF23] for complete details.

## Motivation

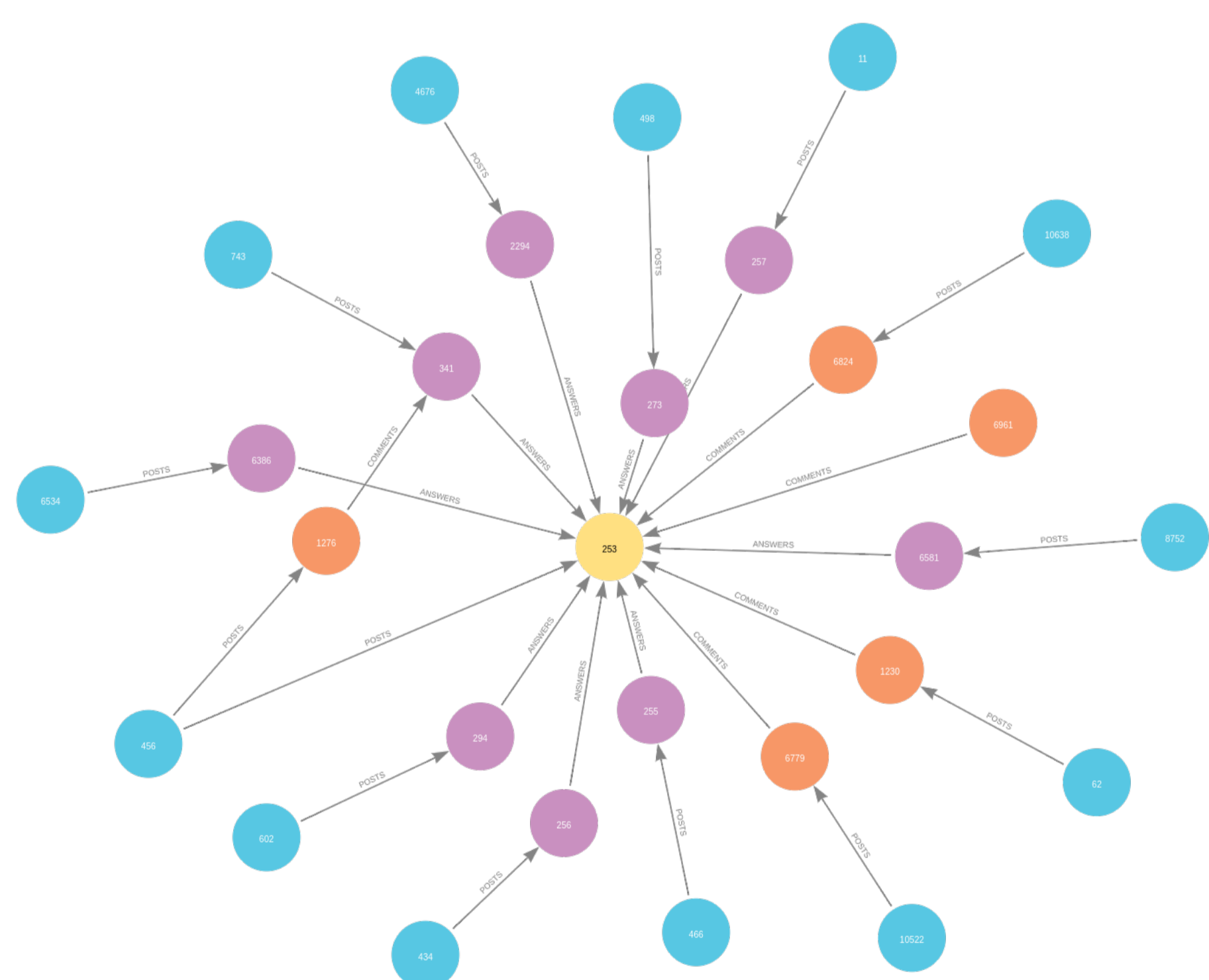


**Figure 1:** For the Computer Science SE, like many other communities on the SE platform, the percentage of resolved questions follows a decreasing trend over the years

## Proposed Approach and Experiments

### The Approach

- PGM [BFVY18] to model the communication network formed around questions
- Two GNN architectures: GCN [KW16] and GGNN [YYL20]
- Three node representations: *text embeddings plus node type*, *only text embeddings*, and *only node type*



**Figure 2:** The communication graph of a question; the blue nodes represent users, the purple nodes represent answers, the orange nodes represent comments, and the yellow node in the middle represents the question

### Experiments

- Used the whole data dump of three SE communities, namely, {Computer, Data, Political} Science SE in the experiments
- Logistic regression (LogReg) and few-shot learners (few-shot(x)) for baselines
- Evaluation metrics: accuracy, precision, recall, and f1-score
- Stratified 5-fold cross-validation
- 5x2cv t-test [Die98] with p-value at 0.005
- Notebooks including the full information and data, are shared on GitHub

## Select Result

**Table 1:** Results on Computer Science SE dataset (with majority class ratio 0.54)

Method	Node feature set	Metric	Mean	SD
GCN	Text embeddings + node type	Accuracy	0.68	0.01
		Recall	0.64	0.04
		Precision	<b>0.65</b>	0.01
		F1-score	0.65	0.02
	Text embeddings	Accuracy	0.64	0.00
		Recall	0.54	0.04
		Precision	0.62	0.01
	Node type	F1-score	0.58	0.02
		Accuracy	0.65	0.00
Recall		0.67	0.01	
GGNN	Text embeddings + node type	Precision	0.61	0.00
		F1-score	0.64	0.01
		Accuracy	<b>0.70</b>	0.01
	Text embeddings	Recall	0.81	0.05
		Precision	0.63	0.02
		F1-score	0.71	0.01
		Accuracy	0.68	0.01
	Node type	Recall	0.73	0.03
		Precision	0.63	0.01
		F1-score	0.68	0.01
		Accuracy	0.68	0.00
		Recall	<b>0.94</b>	0.02
LogReg	Precision	0.59	0.00	
	F1-score	<b>0.73</b>	0.01	
	Accuracy	0.56	0.00	
Few-shot(5)	Recall	0.32	0.01	
	Precision	0.54	0.01	
	F1-score	0.40	0.01	
Few-shot(10)	Accuracy	0.50	0.02	
	Recall	0.44	0.19	
	Precision	0.46	0.01	
Few-shot(20)	F1-score	0.43	0.11	
	Recall	0.47	0.06	
	Accuracy	0.50	0.01	
Few-shot(10)	Precision	0.45	0.01	
	F1-score	0.46	0.03	
	Accuracy	0.49	0.00	
Few-shot(20)	Recall	0.53	0.07	
	Precision	0.45	0.00	
	F1-score	0.49	0.03	

## Funding, Data, and the Code

This work was carried out in the context of the Trondheim Analytica project, supported by the NTNU Digital Transformation programme. Furthermore, the code and the data used in the experiments are available on GitHub.com (<https://github.com/habedi/GNNforUnresolvedQuestions>).

## References

- [AF23] Hassan Abedi Firouzjaei. A deep learning-based approach for identifying unresolved questions on Stack Exchange Q&A communities through graph-based communication modelling. *Int. J. Data Sci. Anal.*, Sep 2023.
- [BFVY18] Angela Bonifati, George Fletcher, Hannes Voigt, and Nikolay Yakovets. *Querying Graphs*. Morgan & Claypool Publishers, 2018.
- [Die98] Thomas G. Dietterich. Approximate Statistical Tests for Comparing Supervised Classification Learning Algorithms. *Neural Comput.*, 10(7):1895–1923, 1998.
- [KW16] Thomas N. Kipf and Max Welling. Semi-Supervised Classification with Graph Convolutional Networks. *arXiv preprint arXiv:1609.02907*, 2016.
- [YYL20] Jiaxuan You, Zhitao Ying, and Jure Leskovec. Design Space for Graph Neural Networks. In *NeurIPS*, pages 17009–17021, 2020.