



Kent Academic Repository

Sun, Zhongtian, Wang, Jingyun, Alamri, Ahmed and Cristea, Alexandra I (2025) *SPAR-GNN: Knowledge Tracing with Behavioural Patterns and Selective LLM Feedback*. In: *Artificial Intelligence in Education. Posters and Late Breaking Results, Workshops and Tutorials, Industry and Innovation Tracks, Practitioners, Doctoral Consortium, Blue Sky, and WideAIED*. . pp. 328-335. Springer ISBN 978-3-031-99263-6. E-ISBN 978-3-031-99264-3.

Downloaded from

<https://kar.kent.ac.uk/112605/> The University of Kent's Academic Repository KAR

The version of record is available from

<https://doi.org/10.1007/9>

This document version

Author's Accepted Manuscript

DOI for this version

Licence for this version

CC BY (Attribution)

Additional information

Versions of research works

Versions of Record

If this version is the version of record, it is the same as the published version available on the publisher's web site. Cite as the published version.

Author Accepted Manuscripts

If this document is identified as the Author Accepted Manuscript it is the version after peer review but before type setting, copy editing or publisher branding. Cite as Surname, Initial. (Year) 'Title of article'. To be published in **Title of Journal** , Volume and issue numbers [peer-reviewed accepted version]. Available at: DOI or URL (Accessed: date).

Enquiries

If you have questions about this document contact ResearchSupport@kent.ac.uk. Please include the URL of the record in KAR. If you believe that your, or a third party's rights have been compromised through this document please see our [Take Down policy](https://www.kent.ac.uk/guides/kar-the-kent-academic-repository#policies) (available from <https://www.kent.ac.uk/guides/kar-the-kent-academic-repository#policies>).

SPAR-GNN: Knowledge Tracing with Behavioural Patterns and Selective LLM Feedback

Zhongtian Sun^{1,2}[0000-0003-0489-5203], Jingyun Wang³[0000-0001-9325-1789],
Ahmed Alamri⁴[0000-0001-9258-3503], and Alexandra
Cristea³[0000-0002-1454-8822]

¹ University of Kent

`z.sun-256@kent.ac.uk`

² University of Cambridge

³ Durham University

`{jingyun.wang,alexandra.i.cristea}@durham.ac.uk`

⁴ University of Jeddah

`asalamri4@uj.edu.sa`

Abstract. Most knowledge tracing systems prioritise performance prediction over the delivery of personalised, actionable feedback. This paper proposes **SPAR-GNN** (Selective Pattern-Aware Reasoning with Graph Neural Networks), a novel framework that combines heterogeneous GNN-based modelling with selectively triggered feedback from a Large Language Model (LLM). SPAR-GNN represents students, problems, skills and virtual pattern nodes that encode latent behavioural traits such as frustration and hint overuse within a unified graph structure. A customised Heterogeneous Graph Layer learns student representations, while the LLM is activated only for at-risk learners identified through behavioural and performance indicators. This reduces computational overhead while enhancing pedagogical relevance. SPAR-GNN consistently outperforms strong baselines across multiple metrics, demonstrating both predictive accuracy and robustness. Observational analysis also reveals interpretable links between behaviours and learning progression, supporting fairness-aware interventions.

Keywords: Knowledge Tracing · Graph Neural Networks · Personalised Feedback · Behavioural Modelling · Large Language Models

1 Introduction

Personalised learning constitutes a foundational component of intelligent tutoring systems (ITS), with accurate *knowledge tracing* serving as a prerequisite for the delivery of tailored instructional interventions.[2, 4]. However, traditional methods such as Bayesian Knowledge Tracing (BKT) and Deep Knowledge Tracing (DKT) focus primarily on correctness prediction, offering limited capacity for interpretability or personalised feedback [7, 3]. More recent approaches like

Self-Attentive Knowledge Tracing (SAKT) improve model transparency but still lack mechanisms for delivering real-time, adaptive guidance [15, 13].

Recent advances in Graph Neural Networks (GNNs) and Large Language Models (LLMs) present an opportunity to bridge this gap. GNNs effectively model complex relationships [20, 18, 19, 10, 24, 22, 8], but typically lack integrated feedback mechanisms. Conversely, while LLMs have exhibited considerable potential in generating educational content[12], their efficacy remains limited in predictive tasks and pedagogical interventions within educational contexts. Current evaluations suggest that LLMs demonstrate stronger capabilities in content generation than in anticipating learner needs or dynamically adapting instructional strategies to optimize pedagogical outcomes.

This paper introduces **SPAR-GNN** (*Selective Pattern-Aware Reasoning with Graph Neural Networks*), a novel framework that combines heterogeneous GNN-based knowledge tracing with behaviour-sensitive LLM feedback. SPAR-GNN incorporates virtual *pattern nodes* to represent latent behavioural archetypes, such as persistent frustration or hint overuse, enabling semantic message passing between students with similar learning challenges. Rather than fine-tuning LLMs across all interactions, SPAR-GNN selectively triggers feedback for learners identified as at risk based on predicted performance and recent behavioural indicators, enhancing both computational efficiency and pedagogical relevance.

To support fairness-aware intervention, we conduct an observational analysis of behaviour–performance correlations, offering an interpretable and lightweight alternative to full causal inference. We evaluate SPAR-GNN on the ASSISTments 2017 dataset—a widely adopted benchmark in educational AI that features rich behavioural signals and structured skill mappings—demonstrating the framework’s applicability to real-world learning contexts. Overall, SPAR-GNN advances the dual aims of knowledge tracing: delivering accurate predictions of learner progression and enabling personalised, equitable support within a scalable architecture.

2 Related Work

Bayesian knowledge tracking and deep knowledge tracking have been widely used in predicting student performance. Bayesian Knowledge Tracing models student learning as a probabilistic transition process [14], while Deep Knowledge Tracing leverages recurrent neural networks to model sequential dependencies [7]. However, these models lack interpretability and fail to generate actionable feedback [1, 21]. Recently, GNNs have been applied in knowledge tracing to model complex relationships between students, concepts and interactions [23, 9]. Although Graph-based knowledge tracking introduces student concept graphs to enhance prediction accuracy, it do not incorporate personalised feedback mechanisms [25].

LLMs, such as GPT and T5, have shown promise in generating instructional content and answering student queries [12]. Prior works have fine-tuned LLMs for knowledge tracking [15, 13], yet they incur high computational costs and lack

adaptability in feedback generation. This approach selectively employs LLMs only in high-impact learning scenarios, optimising efficiency and effectiveness.

3 Methodology

3.1 Graph-Based Knowledge Tracing

This paper proposes SPAR-GNN to model student learning dynamics as a heterogeneous graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ comprising nodes for students, problems, and skills. To capture higher-level learning traits, we introduce virtual **pattern_nodes** representing behavioural archetypes (e.g., persistent frustration, over-reliance on hints). These nodes serve as semantic anchors, linking otherwise unconnected students through shared behavioural signals and enabling message passing across latent affective or cognitive states.

Edges encode key interactions:

- **Attempts:** (`student` $\rightarrow$ `problem`) denotes problem-solving events.
- **Skill mappings:** (`problem` $\rightarrow$ `skill`) specify assessed skills.
- **Hint-informed learning:** (`student` $\rightarrow$ `skill`) captures post-hint success.
- **Behavioural signals:** (`student` $\rightarrow$ `pattern_node`) encode latent affective or cognitive traits.

A single-layer HGTCConv enables relation-specific message passing across this schema. Initial node features $\mathbf{h}_v^{(0)}$ are updated to $\mathbf{h}_v^{(1)}$ via attention-weighted aggregation, and final student embeddings are passed to an MLP for outcome prediction.

3.2 Selective LLM-Driven Feedback

To ensure efficiency and relevance, SPAR-GNN triggers a LLM (LLaMA-3-8B) only for at-risk students—those with predicted scores $\hat{y}_s < 0.4$ and recent behavioural signals indicating high frustration, boredom, or hint dependency. Feedback is generated using a concise prompt incorporating the student ID, dominant skill and key behavioural concerns.:

Student {id} is struggling with skill '{skill}' due to {issue}.

Reply with:

- A simple explanation
- One practical tip
- One real-life example

The resulting responses are cached and logged for transparency and reproducibility. This selective prompting strategy enables targeted, context-aware guidance while substantially reducing token usage and computational overhead, thereby improving scalability in real-world settings.

3.3 Fairness through Behavioural Correlation Analysis

To assess fairness and potential behavioural biases, we analyse Pearson correlations between key behavioural signals (e.g., frustration levels, hint frequency) and model-predicted outcomes. For each behavioural feature x_b , we compute the standardised correlation:

$$\rho_{x_b, \hat{y}} = \frac{\text{Cov}(x_b, \hat{y})}{\sigma_{x_b} \cdot \sigma_{\hat{y}}}.$$

In addition, k -means clustering is applied to the learned student embeddings to uncover latent groupings and assess whether certain behavioural profiles are disproportionately associated with low predicted performance or frequent feedback triggering. These diagnostic tools offer an interpretable lens into the fairness dynamics of the system, laying groundwork for more equitable interventions.

4 Experiments

4.1 Datasets

The ASSISTments 2017 dataset⁵[11] is derived from student interactions within the ASSISTments intelligent tutoring system. It includes a total of 942,816 problem-solving interactions from 686 unique students across 102 different questions. Among these interactions, 38.2% (359,966) were answered correctly, while the remaining 61.8% (582,850) were incorrect. The dataset records fine-grained response patterns, including timestamps, skill mappings, and behavioural indicators such as frustration and hint usage. These features make it particularly well-suited to our study of behaviour-aware knowledge tracing and selective feedback generation [5, 12].

4.2 Evaluation Metrics

We evaluate all models using a comprehensive set of metrics that reflect both predictive accuracy and pedagogical utility:

- **Area Under the Curve (AUC)** – Measures the model’s ability to distinguish between correct and incorrect responses.
- **F1 Score** – The harmonic mean of precision and recall, capturing the trade-off between false positives and false negatives.
- **Root Mean Squared Error (RMSE)** – Quantifies the average prediction error magnitude.
- **Accuracy (Acc)** – The proportion of correctly predicted outcomes.
- **Balanced Accuracy (Bal Acc)** – The average of recall obtained on each class, mitigating class imbalance effects.
- **Precision** – The proportion of predicted positives that are truly correct.
- **Recall** – The proportion of actual positives correctly identified.

⁵ <https://pykt-toolkit.readthedocs.io/en/latest/datasets.html#assistments2017>

4.3 Results

Table 1 presents the performance of baseline models, including classical machine learning, LLM-based and deep learning methods. We report standard metrics: AUC, F1, RMSE, Accuracy, Balanced Accuracy, Precision and Recall. Our model, **SPAR-GNN**, consistently outperforms baselines across most metrics, highlighting its predictive strength and robustness. Results for naïve, LLM-only, and deep learning models are drawn from [12].

Table 1. Results for ASSISTments 2017 dataset. Best values in **bold**.

Family	Model	AUC	F1	RMSE	Acc	Bal Acc	Precision	Recall
Naive Baselines	Mean [16]	0.50	0.00	0.48	0.63	0.50	0.00	0.00
	NaP [16]	0.60	0.26	0.48	0.65	0.55	0.56	0.17
	NaP Skills [12]	0.59	0.35	0.51	0.63	0.56	0.50	0.27
LLM	0-Shot [12]	0.50	0.30	0.51	0.57	0.50	0.37	0.26
	FT Min [12]	0.66	0.38	0.46	0.66	0.58	0.59	0.28
	FT Extended [12]	0.68	0.39	0.46	0.67	0.59	0.63	0.28
DL	BKT [27]	0.63	0.27	0.47	0.65	0.55	0.60	0.17
	Best-LR [6]	0.70	0.48	0.45	0.69	0.63	0.62	0.39
	DKT [14]	0.77	0.58	0.42	0.72	0.68	0.64	0.54
	SAKT [13]	0.70	0.51	0.46	0.69	0.63	0.60	0.43
	SAINT+ [17]	0.54	0.01	0.60	0.63	0.50	0.36	0.01
Ours	SPAR-GNN	0.845	0.67	0.45	0.80	0.77	0.65	0.69

4.4 Qualitative Analysis of Selective LLM Feedback

We qualitatively assess feedback generated by the LLM for students identified as at-risk based on low predicted performance and adverse behavioural signals (e.g., high frustration or frequent hint usage). Prompts follow a consistent structure:

- **Simple Explanation:** A concise, accessible summary of the skill.
- **Practical Tip:** Behaviour-based advice tailored to the student’s difficulty
- **Real-Life Example:** A relatable scenario illustrating the skill in context.

Table 2 presents representative feedback examples across diverse students and skill areas, illustrating the LLM’s ability to generate personalised, context-aware guidance that addresses both cognitive and affective barriers to learning. Cognitive difficulties are supported through clear explanations and strategic scaffolding (e.g., stepwise problem-solving in Skill 27, analogy-based comprehension in Skill 26), while emotional challenges such as frustration or boredom are addressed through self-regulatory strategies like the Pomodoro technique or reduced hint reliance. By embedding suggestions in related scenarios, the feedback promotes

Table 2. Selected LLM-generated feedback for struggling students

ID	Skill	Feedback Summary
1825	Skill 58 (Data Analysis)	Defined data analysis, encouraged real-world connections, and illustrated with a coffee shop example.
2572	Skill 27 (Problem Solving)	Advised breaking complex problems into steps, using a flat tyre to illustrate iterative problem-solving.
1380	Skill 26 (Reading Comprehension)	Suggested overcoming boredom with the Pomodoro technique, and used a step-by-step cake recipe as an analogy for extracting main ideas.
672	Skill 92 (Hint Dependency)	Advised reducing over-reliance on hints through incremental learning, supported by an example involving language learning and vocabulary recall.
1154	Skill 43 (Data Analysis)	Encouraged contextual learning through personal interest, with an example on analysing customer data in a coffee shop for marketing strategy.

both engagement and relevance. Overall, these results demonstrate that *selective* LLM prompting, guided by behavioural and predictive signals can provide targeted, pedagogically meaningful support without the overhead of continual inference.

4.5 Behavioural Correlates of Predicted Performance

To assess how the model aligns with student behaviour, we analyse correlations between behavioural features and predicted performance, by examining:

- **Frustration** (RES_FRUSTRATED): A normalised measure of the student’s emotional struggle during problem-solving.
- **Hint Usage**: The average number of hints requested per student across all problem attempts.

Pearson correlations reveal a weak positive link between frustration and predicted performance ($r = 0.12$), suggesting that some frustrated students remain engaged. In contrast, hint usage shows a moderate negative correlation ($r = -0.41$), indicating that excessive reliance on hints is associated with underperformance. These findings support the model’s behavioral sensitivity and motivate the use of selective interventions tailored to students’ evolving emotional and cognitive states.

5 Conclusion

This paper introduced **SPAR-GNN**, a graph-based knowledge tracing framework that integrates behavioural signals and semantic pattern nodes within a

heterogeneous graph. Using a customised HGT architecture and selectively triggered LLM feedback, SPAR-GNN achieves strong predictive performance ($AUC = 0.845$) while enhancing pedagogical efficiency. Unlike prior methods that fine-tune LLMs across all interactions, SPAR GNN invokes LLMs selectively for at-risk students, improving computational efficiency while maintaining pedagogical relevance. Observational analysis of frustration and hint usage provides a lightweight and interpretable view of fairness without requiring full causal inference. This study is limited to a single dataset (ASSISTments 2017) and one LLM backbone. Future work can explore generalisation across diverse datasets and model architectures, including those that capture behavioural and affective signals (e.g., hint usage, emotional states, engagement logs). Additionally the integration of generative approaches such as adversarial diffusion models [26], offers a promising direction for improving domain adaptability and generating synthetic behavioural traces for low resource setting. Such extensions may also enable more principled causal reasoning and robustness across varied educational contexts.

References

1. Abdelrahman, G., Wang, Q., Nunes, B.: Knowledge tracing: A survey. *ACM Computing Surveys* **55**(11), 1–37 (2023)
2. Alamri, A., Sun, Z., Cristea, A.I., Senthilnathan, G., Shi, L., Stewart, C.: Is mooc learning different for dropouts? a visually-driven, multi-granularity explanatory ml approach. In: *Intelligent Tutoring Systems: 16th International Conference, ITS 2020, Athens, Greece, June 8–12, 2020, Proceedings 16*. pp. 353–363. Springer (2020)
3. Alamri, A., Sun, Z., Cristea, A.I., Stewart, C., Pereira, F.D.: Mooc next week dropout prediction: weekly assessing time and learning patterns. In: *Intelligent Tutoring Systems: 17th International Conference, ITS 2021, Virtual Event, June 7–11, 2021, Proceedings 17*. pp. 119–130. Springer (2021)
4. Carlon, M.K.J., Cross, J.S.: Knowledge tracing for adaptive learning in a metacognitive tutor. *Open Education Studies* **4**(1), 206–224 (2022)
5. Feng, M., Heffernan, N., Koedinger, K.: Addressing the assessment challenge with an online system that tutors as it assesses. *User modeling and user-adapted interaction* **19**, 243–266 (2009)
6. Gervet, T., Koedinger, K., Schneider, J., Mitchell, T., et al.: When is deep learning the best approach to knowledge tracing? *Journal of Educational Data Mining* **12**(3), 31–54 (2020)
7. Ghosh, A., Heffernan, N., Lan, A.S.: Context-aware attentive knowledge tracing. In: *Proceedings of the 26th ACM SIGKDD international conference on knowledge discovery & data mining*. pp. 2330–2339 (2020)
8. Harit, A., Sun, Z.: Causal spherical hypergraph networks for modelling social uncertainty. *arXiv preprint arXiv:2506.17840* (2025)
9. Harit, A., Sun, Z., Yu, J., Al Moubayed, N.: Monitoring behavioral changes using spatiotemporal graphs: A case study on the studentlife dataset. In: *NeurIPS 2024 Workshop on Behavioral Machine Learning* (2024)
10. Harit, A., Sun, Z., Yu, J., Moubayed, N.A.: Breaking down financial news impact: A novel ai approach with geometric hypergraphs. *arXiv preprint arXiv:2409.00438* (2024)

11. Liu, Z., Liu, Q., Chen, J., Huang, S., Tang, J., Luo, W.: pykt: A python library to benchmark deep learning based knowledge tracing models. In: Thirty-sixth Conference on Neural Information Processing Systems Datasets and Benchmarks Track (2022)
12. Neshaei, S.P., Davis, R.L., Hazimeh, A., Lazarevski, B., Dillenbourg, P., Käser, T.: Towards modeling learner performance with large language models. arXiv preprint arXiv:2403.14661 (2024)
13. Pandey, S., Karypis, G.: A self-attentive model for knowledge tracing. arXiv preprint arXiv:1907.06837 (2019)
14. Piech, C., Bassen, J., Huang, J., Ganguli, S., Sahami, M., Guibas, L.J., Sohl-Dickstein, J.: Deep knowledge tracing. *Advances in neural information processing systems* **28** (2015)
15. Pu, S., Becker, L.: Self-attention in knowledge tracing: Why it works. In: International Conference on Artificial Intelligence in Education. pp. 731–736. Springer (2022)
16. Sarsa, S., Leinonen, J., Hellas, A.: Empirical evaluation of deep learning models for knowledge tracing: Of hyperparameters and metrics on performance and replicability. arXiv preprint arXiv:2112.15072 (2021)
17. Shin, D., Shim, Y., Yu, H., Lee, S., Kim, B., Choi, Y.: Saint+: Integrating temporal features for ednet correctness prediction. In: LAK21: 11th International Learning Analytics and Knowledge Conference. pp. 490–496 (2021)
18. Sun, Z., Harit, A., Cristea, A.I., Wang, J., Lio, P.: Money: Ensemble learning for stock price movement prediction via a convolutional network with adversarial hypergraph model. *AI Open* **4**, 165–174 (2023)
19. Sun, Z., Harit, A., Cristea, A.I., Wang, J., Lio, P.: A rewiring contrastive patch performer framework for graph representation learning. In: 2023 IEEE International Conference on Big Data (BigData). pp. 5930–5939. IEEE (2023)
20. Sun, Z., Harit, A., Cristea, A.I., Yu, J., Shi, L., Al Moubayed, N.: Contrastive learning with heterogeneous graph attention networks on short text classification. In: 2022 International Joint Conference on Neural Networks (IJCNN). pp. 1–6. IEEE (2022)
21. Sun, Z., Harit, A., Lio, P.: Actionable interpretability via causal hypergraphs: Unravelling batch size effects in deep learning. arXiv preprint arXiv:2506.17826 (2025)
22. Sun, Z., Harit, A., Yu, J., Wang, J., Liò, P.: Advanced hypergraph mining for web applications using sphere neural networks. In: Companion Proceedings of the ACM on Web Conference 2025. pp. 1316–1320 (2025)
23. Wu, Z., Huang, L., Huang, Q., Huang, C., Tang, Y.: Sgkt: Session graph-based knowledge tracing for student performance prediction. *Expert Systems with Applications* **206**, 117681 (2022)
24. Wynn, A., Wang, J., Sun, Z., Shimada, A.: Analysing learner behaviour in an ontology-based e-learning system: A graph neural network approach (2024)
25. Yang, Y., Shen, J., Qu, Y., Liu, Y., Wang, K., Zhu, Y., Zhang, W., Yu, Y.: Gikt: a graph-based interaction model for knowledge tracing. In: Machine learning and knowledge discovery in databases: European conference, ECML PKDD 2020, Ghent, Belgium, September 14–18, 2020, proceedings, part I. pp. 299–315. Springer (2021)
26. Yu, J., Sun, Z., Luo, S.: Adversarial diffusion model for unsupervised domain-adaptive semantic segmentation. arXiv preprint arXiv:2412.16859 (2024)

27. Yudelson, M.V., Koedinger, K.R., Gordon, G.J.: Individualized bayesian knowledge tracing models. In: Artificial intelligence in education: 16th international conference, AIED 2013, Memphis, TN, USA, July 9-13, 2013. proceedings 16. pp. 171–180. Springer (2013)