← Go to **AAAI 2026 Conference** homepage (/group?id=AAAI.org/2026/Conference)

Scaling Multi-Agent Epistemic Planning through GNN-Derived Heuristics

Giovanni Briglia (/profile?id=~Giovanni_Briglia1),
Francesco Fabiano (/profile?id=~Francesco_Fabiano1),
Stefano Mariani (/profile?id=~Stefano_Mariani2) 👁

📄 (/pdf?id=5XuhfxmrLZ)

📅 24 Jul 2025 (modified: 15 Sept 2025)   📁 Submitted to AAAI-26   👁 Conference, Area Chairs, Senior Program Committee, Program Committee, Authors   📑 Revisions (/revisions?id=5XuhfxmrLZ)   🔖 BibTeX   © CC BY 4.0 (https://creativecommons.org/licenses/by/4.0/)

Serve As Reviewer: 👁 Francesco Fabiano (/profile?id=~Francesco_Fabiano1)
Keywords: 👁 Multi-Agent Epistemic Planning, Graph Neural Networks, Heuristics Search, Learning for Planning
Primary Keyword: PRS: Learning for Planning and Scheduling
TL;DR: We study the use of GNNs to provide heuristics that guide Multi-Agent Epistemic Planning---where states are represented as Kripke structures---to improve scalability by learning from prior solved instances and capturing structural patterns.
Secondary Keywords: MAS: Multiagent Planning, KRR: Reasoning with Beliefs, SO: Heuristic Search, SO: Learning to Search, PRS: Planning/Scheduling and Learning

Abstract:
Multi-agent Epistemic Planning (MEP) is an autonomous planning framework for reasoning about both the physical world and the beliefs of agents, with applications in domains where information flow and awareness among agents are critical. The richness of MEP requires states to be represented as Kripke structures, i.e., directed labeled graphs. This representation limits the applicability of existing heuristics, hindering the scalability of epistemic solvers, which must explore an exponential search space without guidance, resulting often in untractability. To address this, we exploit Graph Neural Networks (GNNs) to learn patterns and relational structures within epistemic states, to guide the planning process. GNNs, which naturally capture the graph-like nature of Kripke models, allow us to derive meaningful estimates of state quality---e.g., the distance from the nearest goal---by generalizing knowledge obtained from previously solved planning instances. We integrate these heuristics into an epistemic planning pipeline and evaluate them against standard baselines, showing significant improvements in the scalability of multi-agent epistemic planning.

Country Of Institutions: 👁 Italy, United Kingdom
Supplementary Material: ⬇ zip (/attachment?id=5XuhfxmrLZ&name=supplementary_material)
Profile Policy Agreement: 👁 I confirm that all authors have up-to-date OpenReview profiles, including their current position, institution-affliated email address, and DBLP URL. I understand that submissions with incomplete author profiles will be subject to desk rejection.
Submission Number: 19203

| Filter by reply type... ▼ | Filter by author... ▼ | Search keywords... | Sort: Newest First |

☰ ☷ ☶   – = ☰   🔗

👁 Everyone | Program Chairs | Submission19203... | Submission19203... | Submission19203...   *5 / 5 replies shown*
Submission19203... | Submission19203... | Submission19203... | Submission19203...
Submission19203... | Submission19203... | ✖

Add: **Withdrawal**   **Ethics Chair Author Comment**

Paper Decision

Decision   by Program Chairs   📅 15 Sept 2025, 14:18 (modified: 15 Sept 2025, 16:47)

👁 Program Chairs, Area Chairs, Senior Program Committee, Program Committee, Authors

📄 Revisions (/revisions?id=WiExBjKZoy)

Decision: Reject

Comment:

This paper did not advance to Phase 2 review for the AAAI-26 conference.

The paper was reviewed by at least two human reviewers in Phase 1, and unfortunately it did not garner sufficient support for promotion to Phase 2. This decision is final. The provided reviews contain more detail, which we hope will be beneficial to the progress of this research.

Add: **Ethics Chair Author Comment**

A review of paper 19203

Official Review   by Program Committee LQZs   📅 29 Aug 2025, 13:42 (modified: 22 Sept 2025, 14:57)

👁 Program Chairs, Area Chairs, Senior Program Committee, Program Committee LQZs, Authors

📄 Revisions (/revisions?id=VszXJMuEcu)

Review:

This paper proposes embedding a regression-based heuristic function learned by GNNs into the MEP framework. By leveraging the graph representation of Kripke structures, the model predicts the distance from a single e-state to the nearest goal. The predicted value is then used as the heuristic function in A*, thereby reducing the number of node expansions in MEP. Experimental results demonstrate that the GNN-based A* algorithm significantly decreases node search volume compared to directly applying BFS.

Strengths

The integration of GNNs with MEP in this work represents a certain degree of novelty.

Weaknesses

1. The optimality of the A* algorithm requires the heuristic function to be a lower bound of the true distance. However, the GNN-based deep learning approach adopted in this paper is a data-driven approximation method, which cannot guarantee this property in theory. As a result, it is difficult to have sufficient confidence in the effectiveness of the proposed method.
2. In search problems, total search time, in addition to the number of expanded nodes, is also an important evaluation criterion. Nevertheless, according to the experimental results presented in the appendix, although the proposed method generally reduces the number of expanded nodes compared to BFS, it does not exhibit a clear advantage in terms of search time.
3. Beyond node expansion and search time, the proposed method also fails to demonstrate significant improvements in problem-solving success rate over BFS. Moreover, it does not achieve a problem-solving rate comparable to that of H-EFP, which raises concerns regarding the overall significance of the work.
4. The experimental section of the paper only includes performance comparisons, but lacks ablation studies to validate the effectiveness of individual design components, such as the state representation method, DFS-based pruning, and others.
5. The methodology section suffers from a lack of clarity. In most cases, it relies heavily on textual descriptions without sufficient mathematical formulations or illustrative figures. Furthermore, the emphasis is not well placed: too much space is devoted to describing the neural network architecture, which is not central to the paper's main contributions.

In summary, while this paper presents a degree novelty by integrating GNNs with MEP, it still suffers from issues regarding theoretical guarantees and the necessity of the proposed approach. Therefore, I recommend rejection.

Rating: 4: Ok but not good enough - rejection

Confidence: 5: The reviewer is absolutely certain that the evaluation is correct and very familiar with the relevant literature

Add: **Ethics Chair Author Comment**

Interesting Paper with Some Issues

Official Review by Program Committee 2Rpn   🗓 27 Aug 2025, 14:58 (modified: 22 Sept 2025, 14:57)

👁 Program Chairs, Area Chairs, Senior Program Committee, Program Committee 2Rpn, Authors

📑 Revisions (/revisions?id=HP3bDsRkrD)

Review:

Summary

This paper is about multi-agent epistemic planning. More precisely, it presents a method for training graph neural networks (GNNs) as goal-distance estimators (i.e., heuristic functions) that can guide heuristic searches. Empirical evaluation shows that these trained GNNs provide effective guidance toward the goal by reducing the number of node expansions compared to an uninformed search algorithm. Furthermore, the evaluation demonstrates that the trained GNN-based heuristic function perform favorably to other "classical heuristics" and provides complementary strengths.

Overall Assessment:

The combination of learning and search in a complex setting such as multi-agent epistemic planning is an interesting and relevant topic for the AAAI community. The paper has some weaknesses in terms of presenting and motivating certain design decisions (see the detailed comments). however, while the approach appears ad hoc in some respects, it may serve as a starting point for future research.

Overall, I think the paper has too many unanswered questions about the design of the pipeline and issues of self-containedness because it makes many references to related work without providing an exact formalization. For this reason, I am leaning towards rejecting the paper in its current form.

Detailed Comments

- It is not clear to me why A* is the search algorithm of choice in this setting. The GNN-based heuristic offers no guarantees, particularly with regard to admissibility, and the state space is finite. Therefore, a more heuristics-driven, aggressive search algorithm, such as greedy best-first search, seems to be a more natural choice. Is there a reason to use A* over a greedy best-first search? In any case, it would be beneficial to discuss this point in the paper or maybe try to run greedy best-first search.
- e-State Representation: From the description, it seems that the predicates that are true in a state are essentially ignored. The explanation for doing so is practical: the number of possible e-states. However, I'm missing a explanation of why this approach makes intuitive sense. As I understand it, the hashing completely loses the relationship of similarity between e-states, which can result in the loss of quite a few (probably a lot) important information. The impact of these design decisions is unfortunately not fully apparent in the write-up (or at least it did not convey this well to me).

References (Minor)

In my opinion, the citations for some concepts do not fully attribute the original works that introduced the ideas. Don't get me wrong; the current citations are fine. However, I believe adding the original work would give credit where credit is due.

- The citation for "classical heuristic planning" is Helmert (2006). This is appropriate because it describes the fundamental planning system that is still used by many today. However, I think it would also be good to cite the paper that essentially started this paradigm: Blai Bonet, Hector Geffner: Planning as heuristic search. Artif. Intell. 129(1-2): 5-33 (2001)
- A* search was original introduced by Peter E. Hart, Nils J. Nilsson, Bertram Raphael: A Formal Basis for the Heuristic Determination of Minimum Cost Paths. IEEE Trans. Syst. Sci. Cybern. 4(2): 100-107 (1968)

Minors

- In the introduction, it would be helpful to spell out what *deep* stand for, as it is done in the experiment section: dynamic epistemic logic-based planner.
- Considering the limited explanation of the mA* language and formalism: I think this is fine given the limited space and the existing pointers. However, I briefly wondered why the state space is finite (as stated in the paper) before realizing that the actions do not allow for announcements and we are not in a dynamic DEL setting. While this may be obvious to some readers, highlighting the actions and their capabilities could help.

Questions

1. Could you comment on the decision to use A* instead of a greedy best-first search? None of the used heuristics are admissible (afaik), so it doesn't seem like the natural choice for an analysis.

Rating: 5: Marginally below acceptance threshold

Confidence: 3: The reviewer is fairly confident that the evaluation is correct

Add: | **Ethics Chair Author Comment** |

AI Review

AI Review   by Program Committee AI   📅 27 Aug 2025, 05:54 (modified: 15 Sept 2025, 20:14)

👁 Program Chairs, Area Chairs, Senior Program Committee, Program Committee AI, Program Committee Submitted, Authors

📑 Revisions (/revisions?id=0ZTTrOxYtO)

Review:

Title: Scaling Multi-Agent Epistemic Planning through GNN-Derived Heuristics

Synopsis of the paper The paper tackles the lack of effective heuristic guidance in multi-agent epistemic planning by learning a distance-to-go estimate directly from epistemic states represented as Kripke structures. It proposes a graph neural network that consumes a joint encoding of the current epistemic state and the goal, trained on state–distance pairs produced via depth-limited search with stochastic sampling and backtracking. The learned regressor is integrated into an A^* planner to prioritize promising states. Experiments across several benchmark domains show large reductions in node expansions versus breadth-first search and evidence of knowledge transfer across domains.

Summary of Review This is an original and relevant contribution for epistemic planning: learning heuristics on full Kripke structures, with a workable data-generation pipeline and integration into a modern solver. The empirical evidence indicates consistent search-space reductions, scalability to longer plans, and cross-domain transfer, though the comparison to the prior heuristic portfolio (H-EFP) is limited to solved counts. The main weaknesses are the under-specified and potentially lossy state/goal graph encoding, limited baselines and ablations, and several technical inconsistencies (most notably the normalization range and percentage reductions). These issues appear correctable with clarifications, additional experiments, and minor revisions, and addressing them would substantially improve the paper's rigor and reproducibility.

Strengths

- Clear motivation and positioning
  - The paper convincingly motivates why heuristic guidance is critical in multi-agent epistemic planning and why classical heuristics do not directly apply to Kripke-structured states. It positions the contribution relative to prior planning-graph-based heuristics in epistemic planners and to solver-level efficiencies.
- Novel technical idea and practical pipeline
  - Learning a distance-to-go heuristic directly over Kripke structures with a graph neural network, while explicitly encoding the goal alongside the state, is original in the epistemic-planning setting. The training data-generation process (depth-limited search with stochastic sampling and backtracking to recover distances) is practical and scalable enough to produce large numbers of state–label pairs from a small set of problems.
- Integration into a working solver with appropriate search
  - The heuristic is integrated into A^* within a re-implementation of a state-of-the-art solver. The system uses bisimulation-based reductions and duplicate checks, ensuring a non-trivial comparison baseline.
- Compelling empirical indications
  - Across multiple setups (per-domain training, cross-domain training, and a scalability test), the learned heuristic reduces node expansions substantially compared to breadth-first search and retains near-optimal plan lengths on the solved subset. In the scalability test, the learned heuristic solves instances that overwhelm uninformed search.
- Sensible choices to improve robustness and reporting
  - The paper uses interquartile mean and IQR-based dispersion to summarize heavy-tailed metrics (Agarwal et al., 2021). The training design includes class balancing over distance bins and target normalization to stabilize training, which are reasonable and effective practices.

Weaknesses

- Under-specified and potentially lossy state/goal encoding
  - The graph encoding uses hashed world identifiers (plus a "repetition number") normalized to $[0, 1]$ and then embedded, but the paper does not precisely define the "repetition number," how fluent valuations are

exposed as features, or how the goal encoding is connected to the e-state graph beyond high-level prose. This risks (a) imposing arbitrary ordinal structure on categorical identifiers, (b) reduced generalization to unseen worlds/goals, and (c) potential collisions conflating distinct worlds. These are central to the approach and should be specified in the main text, not only in the appendix.

- Baselines, ablations, and fairness of comparisons
  - The core baseline is breadth-first search. The comparison to H-EFP is restricted to solved-instance counts and allows H-EFP to run multiple heuristics in parallel, which is not directly comparable to single-heuristic runs. There are no within-solver baselines such as planning-graph heuristics, greedy best-first, or weighted A^*, nor ablations on key design choices (with/without goal encoding, alternative feature sets, number/type of message-passing layers, pooling choices, or class-balancing strategies).
- Limited runtime reporting and inference overhead
  - Results emphasize nodes expanded, while wall-clock time is largely omitted due to the lack of batched inference. Since the GNN introduces non-trivial scoring overhead, the absence of runtime analysis limits practical assessment.
- Technical inconsistencies and notation issues
  - Normalization range: with \alpha=(\text{max_val}-\text{min_val})/D_{\max} and \beta=\text{min_val}, the normalized target is in [\text{min_val},\text{max_val}], not $[0, 1]$, contrary to the text. The later clamping to [\text{min_val},\text{max_val}] is consistent with the intended range.
  - Loss notation: the paper alternates between $\tilde{d}$ and a barred symbol in the MSE, and momentarily refers to the "true normalized distance $d \in [0, 1]$," which conflicts with the earlier definition.
  - GCN equation: the all-ones vector used in $\tilde{D} = \operatorname{diag}(\tilde{A}\mathbf{1})$ is not defined in text.
  - Planning problem tuple: the definition mixes $P = \langle D, I, G \rangle$ with a nested form placing I and G inside D, creating ambiguity; also, Definition 1 vacillates between (M, s) being a pair and s being included inside M.
- Percentage-reduction calculations
  - Several "% Reduction" entries do not match the reported node counts when computed as $(\mathrm{BFS} - \mathrm{GNN})/\mathrm{BFS}$. For example, Setup #1 Test reports GNN 75 vs BFS 343 (implying roughly 78% reduction), but the table lists 48%. This affects headline magnitudes and should be corrected or the aggregation method precisely defined and justified.
- Reproducibility gaps
  - Important procedural details are missing or deferred: exact node and edge feature definitions (including fluent encodings), hashing function and collision handling, the definition and use of the "repetition number," labeling with multiple goals, the per-domain $D_{\max}$ and treatment of distances exceeding $D_{\max}$, sampling probabilities and exploration caps, dataset sizes per domain, random seeds, and A^* configuration details (tie-breaking, re-expansions). The "Train/Test" columns mix instance splits with per-split aggregates, which is sometimes ambiguous.

Suggestions for Improvement

- Precisely specify the state/goal graph encoding in the main text
  - Provide a small running example (e.g., 3 worlds, 2 agents) that maps a Kripke structure and a goal to the exact tensors used by the GNN: node features, edge indices, and edge attributes. Formally define the "repetition number," describe how fluent valuations are encoded (e.g., bitmask for moderate $|\mathcal{F}|$, feature hashing into fixed-size vectors, or explicit learned embeddings per fluent), and explain how the goal is attached to the state graph and aligned across instances. If node IDs are categorical, avoid normalizing IDs to $[0, 1]$ as a scalar; prefer one-hot encodings or learned embeddings to eliminate spurious ordinality.
  - Consider fast, generalizable graph features or kernels as complementary or alternative inputs to the GNN encoder (e.g., Weisfeiler–Leman-style features or graph kernels), which have shown strong efficiency and transfer in planning (Chen & Thiébaux, 2024) and can be combined with classical ML heuristics (Chen, Trevizan, & Thiébaux, 2024).
- Expand baselines, ablations, and fairness
  - Within the same solver, add comparisons to planning-graph heuristics and search variants such as greedy best-first and weighted A^*. For H-EFP, report node expansions and times on the common solved subset under identical single-threaded conditions, alongside the portfolio result. Include ablations on: goal encoding on/off; alternative node/edge features (topology-only vs. fluent-aware); message-passing depth and operator (e.g., GINE vs. GCN vs. GAT); and pooling choice (mean vs. sum vs. attention). A simple non-graph baseline (e.g., MLP over hand-crafted structural features) would help justify the graph backbone.
- Report runtime and prototype batching
  - Include wall-clock time and its variance on representative subsets, and relate it to node expansions. Implement a simple batching strategy (e.g., accumulate a batch from the open list before scoring, or

interleave a cheap heuristic while accumulating) and quantify its effect on time-to-solution. Metareasoning or portfolio scheduling could help decide when to evaluate the GNN versus cheaper guidance (Ma et al., 2020; Budd, Lacerda, & Hawes, 2024).

- Improve the learning target and analysis
  - Explore training using admissible lower bounds as targets or constraints, or ranking-based losses, which can better align with search behavior and improve heuristic quality (Núñez-Molina et al., 2024). Calibrate predictions and analyze error by depth (e.g., scatter plots of predicted vs. true distances), and discuss how saturation is handled when true distances exceed $D_{\max}$ (e.g., censored regression or explicit "beyond-bound" flag).
  - Consider alternative learned guidance signals beyond distances, such as action or successor scoring policies (as in learned global search heuristics for CSPs) trained with policy-gradient or imitation-style objectives (Tönshoff et al., 2023). This could reduce the need for precise distance calibration while still improving open-list ordering.
- Clarify search integration and data labeling
  - Document A^* tie-breaking, duplicate detection, and whether re-expansions occur under heuristic inconsistency. Clarify how labels are computed with multiple goals (minimum distance), how unreachable states are handled (discarded vs. special value), and how the truncation at depth d affects labels and training (e.g., censoring).
- Strengthen positioning and discuss alternatives
  - Discuss when learned heuristics are preferable versus adopting epistemic formalisms designed to remain tractable (Engesser, Herzig, & Perrotin, 2024). Contrast explicit search with learned guidance against recent RL-style integrations with epistemic logic that avoid full Kripke search during planning (Engesser et al., 2025). For transfer across domains and goals, consider a meta-learning or fine-tuning regime to quickly adapt the heuristic to new distributions (Chen et al., 2023).
  - If target applications entail preferences over epistemic objectives, discuss how the approach could be extended to preference-aware state scoring beyond pure distance (Klassen, Muise, & McIlraith, 2023). Recent LLM-modulo planning frameworks could also help synthesize domain priors or features to bootstrap the heuristic in new domains (Kambhampati et al., 2024), though GNNs remain a natural fit for Kripke-structured states.
- Correct technical and reporting issues
  - Fix the normalization statement so the range is [\text{min_val}, \text{max_val}], ensure consistent notation for the normalized target ($\tilde{d}$), define the all-ones vector in the GCN equation, and make Definition 1 and the planning-problem tuple unambiguous. Recompute percentage reductions or clearly explain the aggregation method. Where "Train" and "Test" denote different aggregates (e.g., per-split vs. per-domain), add a caption note. Where feasible, add confidence intervals via bootstrap for IQM (Agarwal et al., 2021).
  - Minor editorial items: fix misspellings and name variants in citations and text to aid reproducibility.

References

- Agarwal, R., Schwarzer, M., Castro, P. S., Courville, A. C., & Bellemare, M. G. (2021). Deep reinforcement learning at the edge of the statistical precipice. In Advances in Neural Information Processing Systems 34.
- Budd, M., Lacerda, B., & Hawes, N. (2024). Stop! Planner time: Metareasoning for probabilistic planning using learned performance profiles. Proceedings of the AAAI Conference on Artificial Intelligence, 38(18), 20053–20060. https://doi.org/10.1609/aaai.v38i18.29983 (https://doi.org/10.1609/aaai.v38i18.29983)
- Chen, D. Z., & Thiébaux, S. (2024). Graph learning for numeric planning. In Advances in Neural Information Processing Systems 37.
- Chen, D. Z., Trevizan, F., & Thiébaux, S. (2024). Return to tradition: Learning reliable heuristics with classical machine learning. Proceedings of the International Conference on Automated Planning and Scheduling, 34, 68–76.
- Chen, J., Wang, J., Zhang, Z., Cao, Z., Ye, T., & Chen, S. (2023). Efficient meta neural heuristic for multi-objective combinatorial optimization. In Advances in Neural Information Processing Systems 36.
- Engesser, T., Herzig, A., & Perrotin, E. (2024). Towards epistemic-doxastic planning with observation and revision. Proceedings of the AAAI Conference on Artificial Intelligence, 38(9), 10501–10508. https://doi.org/10.1609/aaai.v38i9.28919 (https://doi.org/10.1609/aaai.v38i9.28919)
- Engesser, T., Le Marre, T., Lorini, E., Schwarzentruber, F., & Zanuttini, B. (2025). A simple integration of epistemic logic and reinforcement learning. In Proceedings of the 24th International Conference on Autonomous Agents and Multiagent Systems (AAMAS 2025), 686–694.
- Kambhampati, S., Valmeekam, K., Guan, L., Verma, M., Stechly, K., Bhambri, S., Saldyt, L., & Murthy, A. (2024). LLMs can't plan, but can help planning in LLM-modulo frameworks. arXiv:2402.01817.

- Klassen, T. Q., Muise, C., & McIlraith, S. A. (2023). Planning with epistemic preferences. In Proceedings of the 20th International Conference on Principles of Knowledge Representation and Reasoning (pp. 752–756). https://doi.org/10.24963/kr.2023/76 (https://doi.org/10.24963/kr.2023/76)
- Ma, T., Ferber, P., Huo, S., Chen, J., & Katz, M. (2020). Online planner selection with graph neural networks and adaptive scheduling. Proceedings of the AAAI Conference on Artificial Intelligence, 34(4), 5077–5084. https://doi.org/10.1609/aaai.v34i04.5949 (https://doi.org/10.1609/aaai.v34i04.5949)
- Núñez-Molina, C., Asai, M., Mesejo, P., & Fernández-Olivares, J. (2024). On using admissible bounds for learning forward search heuristics. Proceedings of the Thirty-Third International Joint Conference on Artificial Intelligence (pp. 6761–6769). https://doi.org/10.24963/ijcai.2024/747 (https://doi.org/10.24963/ijcai.2024/747)
- Silver, T., Chitnis, R., Curtis, A., Tenenbaum, J. B., Lozano-Pérez, T., & Kaelbling, L. P. (2021). Planning with learned object importance in large problem instances using graph neural networks. Proceedings of the AAAI Conference on Artificial Intelligence, 35(13), 11962–11971. https://doi.org/10.1609/aaai.v35i13.17421 (https://doi.org/10.1609/aaai.v35i13.17421)
- Tönshoff, J., Kisin, B., Lindner, J., & Grohe, M. (2023). One model, any CSP: Graph neural networks as fast global search heuristics for constraint satisfaction. Proceedings of the Thirty-Second International Joint Conference on Artificial Intelligence (pp. 4280–4288). https://doi.org/10.24963/ijcai.2023/476 (https://doi.org/10.24963/ijcai.2023/476)

Add: **Ethics Chair Author Comment**

The motivation and contributions of this paper are relatively clear; however, there are still some unclear descriptions in the Methods section and some ambiguous parts in the Experiments section. Since my understanding of this field is not very deep, I need to further discuss certain aspects with the authors. Therefore, I decide to err on the side of caution and give it a score of 5 for now.

Official Review by Program Committee 1H7B   📅 23 Aug 2025, 16:33 (modified: 22 Sept 2025, 14:57)
👁 Program Chairs, Area Chairs, Senior Program Committee, Program Committee 1H7B, Authors
📄 Revisions (/revisions?id=VI2NeA8Epk)

Review:
Summary

This paper proposes a Graph Neural Network (GNN)-based heuristic approach to enhance the scalability of Multi-Agent Epistemic Planning (MEP). MEP faces significant challenges due to the complexity of state spaces that arise from reasoning about information flow and belief updates among agents, rendering traditional planning methods inefficient. By leveraging GNNs to learn patterns and relationships within states, the authors generate effective heuristic estimates that guide the planning process and reduce the search space. Experimental results demonstrate substantial improvements in planning efficiency, with a significant reduction in the number of expanded nodes across multiple standard benchmarks, as well as good generalization capabilities across different domains. This work offers a novel, data-driven solution to the field of MEP, holding important theoretical and practical implications.

Strength

The motivation and contributions of this paper are relatively clear. By utilizing Graph Neural Networks (GNNs) to learn heuristic methods, it addresses the scalability issues in MEP caused by the complexity of the state space, thereby improving planning efficiency.

Weakness

1. All the tables are completely misformatted, and the formula encoding in the Methods section is also incorrect; these need to be updated.
2. The writing in the Methods section is somewhat vague and hard for me to follow. ① I'm unclear about the "Build the Dataset" part here—is this dataset the same as the one you use for training? If not, how can you ensure that the estimation model trained on the training set can be applied to another problem?
3. The sampled data points in the Experiments section show very clear results. But how can the precision of the experimental results be guaranteed?
4. In the Related work section, ① I don't know in what ways using GNNs is better than the original heuristic methods, such as those that manually design A* cost functions. This is also something I'm not quite clear about in the Experiments section; ② I'd like to know how this work differs from existing Multi-agent RL work.

While I'm not very familiar with the background of this paper, I have read some additional related literature, though not extensively. I'm very open to communicating with the authors.

Rating: 5: Marginally below acceptance threshold

Confidence: 2: The reviewer is willing to defend the evaluation, but it is quite likely that the reviewer did not understand central parts of the paper