Spatio-Temporal Graphs Beyond Grids: Benchmark for Maritime Anomaly Detection

Anonymous Author(s)
Affiliation
Address
email

Abstract

Spatio-temporal graph neural networks (ST-GNNs) have achieved notable success in structured domains such as road traffic and public transportation, where spatial entities can be naturally represented as fixed nodes. In contrast, many real-world systems including maritime traffic lack such fixed anchors, making the construction of spatio-temporal graphs a fundamental challenge. Anomaly detection in these non-grid environments is particularly difficult due to the absence of canonical reference points, the sparsity and irregularity of trajectories, and the fact that anomalies may manifest at multiple granularities. In this work, we introduce a novel benchmark dataset for anomaly detection in the maritime domain, extending the Open Maritime Traffic Analysis Dataset (OMTAD) into a benchmark tailored for graph-based anomaly detection. Our dataset enables systematic evaluation across three different granularities: node-level, edge-level, and graph-level anomalies. We plan to employ two specialized LLM-based agents: *Trajectory Synthesizer* and *Anomaly Injector* to construct richer interaction contexts and generate semantically meaningful anomalies. We expect this benchmark to promote reproducibility and to foster methodological advances in anomaly detection for non-grid spatio-temporal systems.

1 Introduction

Spatio-temporal graph neural networks (ST-GNNs) have been extensively studied in domains such as road traffic forecasting and public transportation systems [19, 3, 2]. A common characteristic of these applications is that the underlying spatial entities like road intersections, bus stops, or subway stations can be naturally defined as fixed nodes. This inherent grid-like structure makes the construction of spatio-temporal graphs straightforward and facilitates the modeling of both spatial dependencies and temporal dynamics. Consequently, anomaly detection in such structured environments has received significant attention and demonstrated promising results [1].

However, there are many cases both in real-world and scientific domains where situations do not conform to these assumptions. In particular, there exist domains where fixed spatial anchors are absent or physically ambiguous. The maritime environment represents one of the most prominent examples: unlike road traffic systems, the open sea does not provide natural fixed nodes such as intersections or road segments. Although artificial proxies such as waypoints, port coordinates, or grid discretizations can be imposed, these methods are often ad hoc and fail to capture the continuous and dynamic nature of vessel trajectories. This fundamental challenge renders the construction of a meaningful spatio-temporal graph a non-trivial task. We expect that such *non-grid spatio-temporal systems* will become increasingly common, not only in maritime monitoring but also in emerging domains such as drone swarms and aerial traffic management.

Submitted to AI for Science workshop (NeurIPS 2025). Do not distribute.

Performing anomaly detection in these settings is even more challenging. First, the lack of fixed spatial anchors complicates the definition of normal versus abnormal interactions among moving entities. Second, the inherent sparsity and irregularity of the trajectories make it difficult to design robust models. Third, anomalous patterns may manifest at multiple levels: individual entities (node-level anomalies), unusual pairwise interactions (edge-level anomalies), or entire subgroups behaving abnormally (graph-level anomalies). These challenges highlight the need for systematic benchmarks that enable rigorous evaluation and foster methodological innovations [7]. There are several Marine datasets

To address this gap, in this paper we introduce a novel benchmark dataset for anomaly detection in the maritime domain. Our dataset is designed to support anomaly detection tasks at three granularities: *(i) node-level anomalies*, capturing abnormal single-entity behaviors, *(ii) edge-level anomalies*, reflecting irregular inter-entity interactions, and *(iii) graph-level anomalies*, identifying collective abnormal events. Inspired by recent advances in graph anomaly detection across node-, edge-, and graph-level settings [5], we aim to provide a unified testbed that allows the community to explore and compare methods across multiple anomaly detection settings. To construct the dataset in a principled manner, we use two large language model(LLM)-based agents: *Trajectory Synthesizer*, which augments inter-vessel contexts by enriching sparse neighborhoods, and an *Anomaly Injector*, which introduces diverse anomalies guided by high-level prompts. We believe this contribution will not only facilitate research on maritime anomaly detection, but also establish a foundation for studying anomaly detection in broader non-grid spatio-temporal systems.

2 Dataset

We build our benchmark upon the **Open Maritime Traffic Analysis Dataset (OMTAD)** [9], a publicly available and openly licensed collection of vessel trajectories derived from AIS signals. OMTAD covers the West Australian offshore region (105–116°E, 36–15°S) from 2018 to 2020, and provides **19,124** trajectories across four vessel categories: Cargo (14,384), Tanker (4,020), Fishing (466), and Passenger (254). Each AIS record includes vessel identifiers, geolocation, kinematic information such as course over ground (COG) and speed over ground (SOG), and UTC timestamps.

2.1 Limitations and Our Extensions

While OMTAD provides a well-organized and open collection of vessel tracks, it has two key limitations that prevent direct use for graph-based anomaly detection. First, although some trajectories are physically close enough to form meaningful spatio-temporal graphs, many occur in isolation without nearby neighbors, making graph construction difficult. Second, it only contains *normal* trajectories and thus provides no anomaly labels. These issues hinder systematic benchmarking of graph-based anomaly detection.

To address these limitations, we extend OMTAD in two complementary ways.

- **Trajectory synthesis in sparse regions.** For vessels without nearby neighbors, we generate synthetic but physically plausible companion trajectories. These synthetic neighbors are created by perturbing SOG, COG, and geolocation values within bounded ranges, ensuring that even isolated vessels can be embedded into meaningful spatio-temporal graphs.
- **Anomaly injection.** Since no anomalies are provided, we introduce anomalies through a controlled injection process. Instead of rigid rules, we rely on prompt-driven generation to produce diverse anomalies across node, edge, and graph levels, aligning them with semantically meaningful maritime scenarios.

This extension is not only practical but also justified: preliminary experiments in Appendix A.2 show that even under relatively naive anomaly settings, graph-based models consistently outperform purely temporal baselines. This validates that repurposing OMTAD into a graph-based anomaly detection benchmark captures meaningful structural signals and provides a solid foundation for further extensions. Through these steps, OMTAD is repurposed into a unified benchmark dataset that supports systematic evaluation of anomaly detection in non-grid spatio-temporal graphs. Detailed configurations of the trajectory synthesis and anomaly injection agents are provided in the Appendix C.

References

[1] A. Deng and B. Hooi. Graph neural network-based anomaly detection in multivariate time series. In *Proceedings of the AAAI conference on artificial intelligence*, volume 35, pages 4027–4035, 2021.

[2] S. Guo, Y. Lin, N. Feng, C. Song, and H. Wan. Attention based spatial-temporal graph convolutional networks for traffic flow forecasting. In *Proceedings of the AAAI conference on artificial intelligence*, volume 33, pages 922–929, 2019.

[3] Y. Li, R. Yu, C. Shahabi, and Y. Liu. Diffusion convolutional recurrent neural network: Data-driven traffic forecasting. In *International Conference on Learning Representations*, 2018.

[4] N. Lim, B. Hooi, S.-K. Ng, X. Wang, Y. L. Goh, R. Weng, and J. Varadarajan. Stp-udgat: Spatial-temporal-preference user dimensional graph attention network for next poi recommendation. In *Proceedings of the 29th ACM international conference on information & knowledge management*, pages 845–854, 2020.

[5] Y. Lin, J. Tang, C. Zi, H. V. Zhao, Y. Yao, and J. Li. Unigad: Unifying multi-level graph anomaly detection. *Advances in neural information processing systems*, 37:136120–136148, 2024.

[6] J. Liu, J. Li, and C. Liu. Ais-based kinematic anomaly classification for maritime surveillance. *Ocean Engineering*, 305:118026, 2024.

[7] X. Ma, J. Wu, S. Xue, J. Yang, C. Zhou, Q. Z. Sheng, H. Xiong, and L. Akoglu. A comprehensive survey on graph anomaly detection with deep learning. *IEEE transactions on knowledge and data engineering*, 35(12):12012–12038, 2021.

[8] S. Mao, E. Tu, G. Zhang, L. Rachmawati, E. Rajabally, and G.-B. Huang. An automatic identification system (ais) database for maritime trajectory prediction and data mining. In *Proceedings of ELM-2016*, pages 241–257. Springer, 2017.

[9] M. Masek, C. P. Lam, T. Rybicki, J. Snell, D. Wheat, L. Kelly, D. Glassborow, and C. Smith-Gander. The open maritime traffic analysis dataset. In *24th International Congress on Modelling and Simulation (MODSIM)*, pages 948–954, 2021.

[10] D. Nguyen and R. Fablet. A transformer network with sparse augmented data representation and cross entropy loss for ais-based vessel trajectory prediction. *IEEE Access*, 12:21596–21609, 2024.

[11] D. Nguyen, R. Vadaine, G. Hajduch, R. Garello, and R. Fablet. Geotracknet—a maritime anomaly detector using probabilistic neural network representation of ais tracks and a contrario detection. *IEEE Transactions on Intelligent Transportation Systems*, 23(6):5655–5667, 2021.

[12] I. Obradović, M. Miličević, and K. Žubrinić. Machine learning approaches to maritime anomaly detection. *Naše more: znanstveni časopis za more i pomorstvo*, 61(5-6):96–101, 2014.

[13] B. Qiao, K. Li, W. Zhou, S. Li, Q. Lu, and S. Hu. Botsim: Llm-powered malicious social botnet simulation. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 39, pages 14377–14385, 2025.

[14] M. Riveiro, G. Pallotta, and M. Vespe. Maritime anomaly detection: A review. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, 8(5):e1266, 2018.

[15] S. K. Singh and F. Heymann. Machine learning-assisted anomaly detection in maritime navigation using ais data. In *2020 IEEE/ION Position, Location and Navigation Symposium (PLANS)*, pages 832–838. IEEE, 2020.

[16] S. Wang, Y. Li, H. Xing, and Z. Zhang. Vessel trajectory prediction based on spatio-temporal graph convolutional network for complex and crowded sea areas. *Ocean Engineering*, 298:117232, 2024.

[17] Z. Wu, S. Pan, G. Long, J. Jiang, and C. Zhang. Graph wavenet for deep spatial-temporal graph modeling. *arXiv preprint arXiv:1906.00121*, 2019.

[18] T. Yang, Y. Nian, S. Li, R. Xu, Y. Li, J. Li, Z. Xiao, X. Hu, R. Rossi, K. Ding, et al. Ad-llm: Benchmarking large language models for anomaly detection. *arXiv preprint arXiv:2412.11142*, 2024.

[19] B. Yu, H. Yin, and Z. Zhu. Spatio-temporal graph convolutional networks: A deep learning framework for traffic forecasting. *arXiv preprint arXiv:1709.04875*, 2017.

[20] Y. Zang, L. Deng, S. Sun, Y. Ai, D. Geng, and X. Kang. Abnormal behavior detection of vessels based on deep learning algorithm: case study. In *2021 6th International Conference on Transportation Information and Safety (ICTIS)*, pages 205–212. IEEE, 2021.

[21] Y. Zhang, Q. Jin, M. Liang, R. Ma, and R. W. Liu. Vessel behavior anomaly detection using graph attention network. In *International Conference on Neural Information Processing*, pages 291–304. Springer, 2023.

A Motivation

To verify the feasibility of our proposed benchmark, we conducted a preliminary experiment by injecting synthetic anomalies into the OMTAD dataset. This was necessary because maritime data lacks ground-truth anomaly labels and defining anomalies is highly context-dependent, with no clear consensus even within the maritime community. By introducing controlled perturbations, we created a testbed to examine whether graph-level anomaly detection tasks can be meaningfully supported in this setting.

A.1 Anomaly Injection

We synthesized anomalies by perturbing vessel trajectories at the node level. For each trajectory of length w, a contiguous anomaly block of size $m = r_{\text{node}} w$ was chosen, where $r_{\text{node}} \in \{r_1, r_2, r_3\}$ is the node anomaly ratio. The block was placed by sampling a start index $s \sim \mathcal{U}(0, w - m)$, which defined a binary anomaly mask z_t indicating anomalous segments. Nodes within the anomaly block were perturbed in their Speed Over Ground (SOG) and Course Over Ground (COG) features [6]. Formally, we modeled the rates of change of SOG (a) and COG (ω) as normally distributed, $a \sim \mathcal{N}(\mu_a, \sigma_a^2)$ and $\omega \sim \mathcal{N}(\mu_\omega, \sigma_\omega^2)$, where $a_i = (\text{SOG}_i - \text{SOG}_{i-1})/\Delta t$ and $\omega_i = (\text{COG}_i - \text{COG}_{i-1})/\Delta t$. To create significant deviations, we replaced them with $a_i^* = \mu_a + k \cdot \sigma_a$ and $\omega_i^* = \mu_\omega + k \cdot \sigma_\omega$ with $k > 3$, which ensures perturbed values lie outside the 99.7% confidence interval of normal behavior. The updated SOG and COG values were then iteratively applied over the anomaly block. A trajectory was labeled anomalous ($y_{\text{traj}} = 1$) if at least one node was perturbed, and normal ($y_{\text{traj}} = 0$) otherwise. As mentioned earlier, defining anomalies in the maritime context is inherently difficult, and even within this domain there is no established consensus. Therefore, we restrict our anomaly definition to *kinematic movement anomalies* based on SOG and COG deviations.

Node- and graph-level anomaly ratios. In our design, anomaly prevalence is controlled at two complementary levels. First, the *node anomaly ratio* $r_{\text{node}} \in (0, 1]$ specifies the fraction of anomalous nodes within a trajectory. Given a trajectory of length w, the anomalous span is set to $m = r_{\text{node}}\, w$ nodes, realized as a consecutive block of length m. This choice reflects the temporal persistence of real-world incidents, since anomalies are more likely to appear as sustained abnormal behaviors (e.g., equipment malfunction, evasive maneuvers, adverse weather conditions, or loitering) rather than isolated spikes.

Second, the *trajectory anomaly ratio* $r_{\text{traj}} \in (0, 1]$ denotes the fraction of trajectories labeled anomalous at the graph level, formally defined as

$$r_{\text{traj}} \;=\; \frac{1}{N} \sum_{i=1}^{N} \mathbb{1}\Big(y_{\text{traj}}^{(i)} = 1\Big)\,,$$

where N is the total number of trajectories.

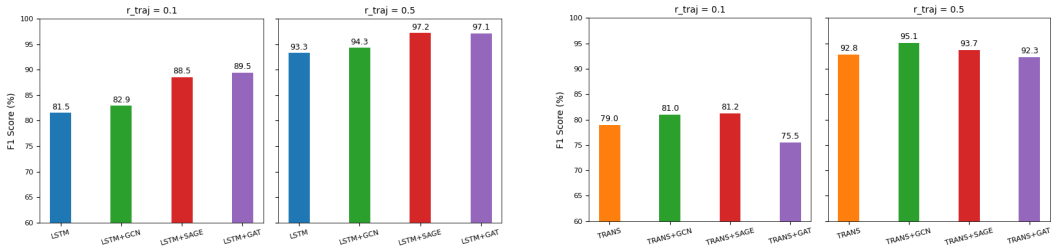
Thus, r_{node} controls the intra-trajectory anomaly density, while r_{traj} governs the dataset-level class balance for graph-level detection.

A.2 Preliminary Experiment

We conducted a preliminary study under the setting of *graph-level anomaly detection*, where each vessel trajectory is classified as either normal or anomalous. In this setup, we varied the trajectory anomaly ratio $r_{\text{traj}} \in \{0.1, 0.5\}$ while fixing the node anomaly ratio at $r_{\text{node}} = 0.5$. We compared standard time-series models (LSTM, Transformer) with their *Time-series + GNN* counterparts, which incorporate GNN modules, as summarized in Fig. 1a and Fig. 1b.

To construct graph inputs for the GNNs, we applied the OPTICS clustering algorithm to spatial snapshots at each timestamp t, grouping nearby vessels into dynamic clusters without imposing predefined constraints on the number or shape of clusters. From each cluster, a fixed number k of vessel trajectories was sampled to ensure that the adjacency matrix remained consistent across all graphs. For each sampled set, we then built a directed temporal graph over a w-hour observation window, where each node corresponds to a time-stamped vessel state, resulting in exactly $k \times w$ nodes per graph.

Although this preliminary experiment primarily focuses on *graph-level* anomaly detection with varying r_{traj} and fixed r_{node}, the same framework can naturally be extended to support *node-level* anomaly detection.



(a) (b)

Figure 1: Preliminary results of time-series models and their GNN-integrated variants under different r_{traj} settings. (a) LSTM-based models. (b) Transformer-based models. "TRANS" denotes Transformer.

A.3 Findings

The results show that GNN-integrated models consistently outperform purely temporal baselines across all anomaly ratios. This demonstrates that graph-based modeling provides a more natural fit for capturing maritime dynamics, where vessel states and inter-vessel interactions must be considered jointly. At the same time, it is important to emphasize that our injection strategy perturbed only the simplest navigational features. Real-world maritime anomalies are far more diverse, including illegal rendezvous between vessels, loitering behaviors, spoofed AIS signals, or sudden deviations due to environmental conditions. Thus, while this experiment confirmed the viability of our framework, it represents only a simplified case. Our ultimate goal is to generalize this process by employing LLM-based agents to automatically generate and annotate richer, semantically meaningful anomalies, thereby creating a more realistic and versatile benchmark environment.

B Related Works

Spatio-temporal GNNs in structured domains. Spatio-temporal graph neural networks (ST-GNNs) have achieved notable success in domains where the underlying graph structure is fixed and well defined, such as road traffic [19, 3, 17], public transportation [2], and mobility systems [4, 1]. In these settings, nodes typically correspond to pre-defined spatial anchors (e.g., intersections, stations, or sensors), which makes the construction of spatio-temporal graphs straightforward and effective. However, the assumptions of stable topologies and fixed node identities do not generalize to non-grid environments such as the open sea, where spatial anchors are absent and trajectories are highly irregular.

Maritime Anomaly Detection and Datasets. Maritime anomaly detection has emerged as a challenging task in the maritime domain due to the dynamic and unstructured nature of vessel movements. A comprehensive survey [14] highlights the difficulty of defining anomalous behaviors and reviews a wide range of approaches. Classical machine learning techniques have long been applied to AIS data, including supervised and unsupervised methods for identifying irregular navigation patterns [12, 15]. More recently, deep learning approaches have demonstrated stronger capacity for modeling complex temporal dependencies, such as CNN- and RNN-based models for abnormal behavior detection [20] and probabilistic neural representations like GeoTrackNet [11]. Transformer-based methods have also been introduced, with TrAISformer [10] achieving state-of-the-art results in AIS-based trajectory prediction. In parallel, graph-based methods have gained momentum for their ability to explicitly capture vessel-to-vessel interactions, including graph attention networks for anomaly detection [21] and spatio-temporal graph convolutional networks for trajectory prediction in crowded sea areas [16].

A key bottleneck in advancing this line of research lies in the lack of standardized open datasets. While several AIS-based datasets exist [8], they are often incomplete, commercial, or unavailable for public use. The Open Maritime Traffic Analysis Dataset (OMTAD) [9] represents an important step toward openness by providing cleaned and processed AIS tracks for multiple vessel types. Nevertheless, OMTAD has not been designed as an anomaly detection benchmark, and in particular, it lacks systematic definitions and annotations for multi-level anomalies. Our work addresses this gap by extending OMTAD into a benchmark dataset tailored for graph-based anomaly detection across node, edge, and graph levels.

LLM-based Anomaly Injection and Benchmark Augmentation. Recent studies have begun to explore the potential of LLMs in supporting anomaly detection tasks. For instance, AD-LLM [18] presents the first comprehensive benchmark that systematically examines how LLMs can be leveraged for anomaly detection across multiple dimensions, including zero-shot detection, data augmentation, and model selection. This line of work demonstrates the broad applicability of LLMs in enhancing anomaly detection pipelines. However, these efforts primarily remain at an abstract level and provide limited insights into fine-grained dataset augmentation grounded in real-world domain data.

In parallel, BotSim [13] introduces an LLM-powered end-to-end simulation toolkit for malicious social botnet generation, enabling downstream evaluation of bot detection methods. This framework illustrates how LLM agents can be utilized to construct diverse and semantically meaningful anomaly scenarios in a simulation setting. Nevertheless, the maritime domain remains underexplored: despite the availability of AIS-based datasets such as OMTAD, there has been little research on using LLMs to perform precise, domain knowledge–driven anomaly injection.

To bridge this gap, we extend OMTAD into a benchmark dataset specifically designed for maritime anomaly detection. To the best of our knowledge, this is the first attempt to systematically augment a real-world maritime dataset with LLM-based anomaly injection, providing a platform for training and evaluating anomaly detection methods in non-grid spatio-temporal systems.

C Dataset Construction Method

Overview We adopt a two–agent architecture specialized for dataset construction: (1) *Trajectory Synthesizer*, which enriches inter-vessel connectivity through augmentation of local contexts, and (2) *Anomaly Injector*, which introduces anomalies guided by high-level text prompts. Both agents operate under a common *Coordinator* that manages data flow, prepares structured perception inputs, enforces constraints, and validates outputs. This design separates augmentation (ensuring sufficient structural density) from anomaly generation (ensuring semantic variety), providing a flexible and reproducible pipeline for benchmark creation.

C.1 Coordinator Workflow

For each focal vessel v over a given window $[t_0, t_1]$, the Coordinator executes a simple loop: (i) construct a standardized *perception bundle* from AIS and environmental metadata, (ii) dispatch it to the Trajectory Synthesizer to obtain an augmented multi-vessel graph $\mathcal{G}$, (iii) pass the synthesized graph and perception context to the Anomaly Injector to apply prompt-driven modifications and produce labels, and (iv) collect provenance, validation logs, and final artifacts for dataset assembly.

Table 1: Perception schema consumed by both agents.

Category	Fields
AIS	MMSI, t, latitude, longitude, SOG, COG
Derived	ΔSOG/Δt, ΔCOG/Δt
Env	wind/wave/current bins, visibility proxy
Provenance	source trajectory IDs

In this way, augmentation and anomaly injection are decoupled but remain interoperable under a single orchestrator.

C.2 Shared Environment Perception Schema

Both agents consume a common schema that represents vessel states and their context in a slot-filled format. The specific categories and fields are mentioned in Table 1.

This schema ensures that both augmentation and injection modules operate on consistent, validated inputs. All fields follow fixed units and identifiers, and missing values are explicitly marked to maintain determinism.

C.2.1 Agent 1: Trajectory Synthesizer (Augmentation)

Goal. Increase the density and diversity of meaningful inter-vessel interactions so that GNN-based methods can better exploit spatial context while preserving physical plausibility.

Main Idea. The Trajectory Synthesizer enriches local graph structures by adding trajectories around each vessel to ensure sufficient connectivity and realistic interaction density.

Components.

- **Neighbor-based augmentation:** If physically close vessels are present, their trajectories are directly included to form proximity-based edges and enrich inter-vessel connectivity.
- **Synthetic augmentation:** In sparse regions where nearby vessels are absent, the agent generates additional "virtual neighbors" by sampling trajectories similar to the focal vessel. Their SOG, COG, latitude, and longitude values are perturbed within realistic variation ranges to preserve plausibility while increasing graph density.

Outputs. An augmented spatio-temporal graph that combines original vessel tracks with either actual or synthesized neighbors, including provenance information indicating which trajectories were real and which were generated.

C.2.2 Agent 2: Anomaly Injector (Prompt-Driven)

Goal. Introduce diverse and semantically meaningful anomalies into trajectories in order to support node-, edge-, and graph-level anomaly detection tasks.

Main Idea. The Anomaly Injector operates from high-level *text prompts* rather than fixed perturbation rules, allowing flexible and context-aware anomaly creation.

Components.

- **Prompt Interpretation:** Parsing natural language descriptions of anomalies (e.g., unusual speed changes, risky encounters, or group loitering) into structured intent.
- **Scenario Realization:** Mapping the interpreted intent into corresponding edits of the spatio-temporal graph, such as modifying single-node kinematics, vessel-to-vessel interactions, or group-level patterns.
- **Label Generation:** Attaching anomaly labels (node, edge, or graph level) along with rationale text that traces back to the original prompt.

Outputs. A set of modified trajectories and anomaly labels, where each label is tied to a prompt, anomaly type, and severity level, accompanied by rationale text for interpretability.

D Future Works

While our current work lays the foundation for a benchmark on non-grid spatio-temporal anomaly detection, several important directions remain for future development. First, we plan to consolidate the proposed pipeline into a reproducible framework that can automatically synthesize augmented trajectories and inject anomalies through prompt-driven agents. The next step is to curate a finalized version of the dataset. We will release the dataset under an open license to encourage broad adoption and reproducibility, accompanied by scripts that enable researchers to regenerate augmented or injected variants deterministically.

Second, to establish a reference point for the community, we will benchmark a variety of baseline methods on the dataset. This includes purely temporal sequence models such as LSTM and Transformer, hybrid spatio-temporal GNN models, and recent graph anomaly detection architectures designed for node-, edge-, and graph-level tasks. Comprehensive evaluation across different anomaly ratios and scenarios will provide insights into the strengths and limitations of each model class.

Finally, we aim to extend the anomaly definitions beyond the initial kinematic-focused injections. In particular, we plan to incorporate more semantically complex anomalies, such as illegal encounters, coordinated group behaviors, or procedural violations near ports and restricted areas. Leveraging LLM-based agents in conjunction with domain rules will allow us to gradually expand the scope of the benchmark, bridging the gap between controlled synthetic anomalies and realistic, context-dependent maritime events. In parallel, we recognize that the task-specific labeling strategy itself requires careful refinement. Defining consistent and interpretable labels across node-, edge-, and graph-level tasks is non-trivial, and we plan to investigate principled ways of assigning task-aware labels that capture both local anomalies and their broader contextual implications.