HRD-Net: Hierarchical Relational Diffusion Network for Anomaly Detection in Multivariate Time-series Data

1st Hongbo Zhao
School of Computer Science and Technology
East China Normal University
Shanghai, China
51275901107@stu.ecnu.edu.cn

2nd Zhe Wang
School of Computer Science and Technology
East China Normal University
Shanghai, China
52275901012@stu.ecnu.edu.cn

3rd Yinjie Teng
School of Computer Science and Technology
East China Normal University
Shanghai, China
51275901082@stu.ecnu.edu.cn

4th Yihong Huang
School of Computer Science and Technology
East China Normal University
Shanghai, China
51275901062@stu.ecnu.edu.cn

5th Kai Zhang
School of Computer Science and Technology
East China Normal University
Shanghai, China
kzhang980@gmail.com

***Abstract*—Multivariate time series anomaly detection is critical for ensuring the safety of complex systems with wide applications. Recently, Graph neural networks have emerged as a powerful tool for anomaly detection due to its ability to model intricate dependencies among different system components. However, the multi-layer information aggregation in GNNs often results in a receptive field that is insensitive to local topological contexts or "edge densities". This may introduce irrelevant nodes across community boundaries to contaminate message passing, degrade the prediction, and finally lead to false positive alerts. In this paper, we propose Hierarchical Relational Diffusion Network (HRD-Net) to build contextualized relational fields in GNNs for anomaly detection. HRD-Net uses random-walk based diffusion to recover the relational context for each node, with adaptive restart probabilities to balance the trade-off between local communities and global interactions under different scales. The resultant probability maps serve as a hierarchical "topological firewall" to suppress spurious correlations while preserving truly relevant neighbors to guarantee the quality of message passing. Furthermore, HRD-Net enables the construction of a more reliable multi-scale anomaly scoring system by integrating reconstruction errors from both individual scales and their aggregated representation. Experiments on six challenging benchmarks demonstrate the potential of HRD-Net over altogether 12 methods, with case studies to interpret and validate its capability to disentangle equipment faults from sensor drift artifacts in real-world maintenance logs.**

***Index Terms*—Anomaly detection, Multivariate time series, Graph neural networks, Multi-scale learning.**

I. Introduction

Multivariate time series anomaly detection aims to identify rare observations that deviate from normal patterns, which is crucial for ensuring the reliability of complex systems in domains such as industrial production, healthcare, and finance [1]–[5]. This task is challenging due to the scarcity and diversity of anomalies, complex inter-channel dependencies, and temporal dynamics.

Graph neural networks (GNNs) have recently shown strong potential by modeling variables as nodes and their relationships as edges. Representative works include MTAD-GAT [6], GDN [7], and GReLeN [8], along with more advanced methods such as GTA [9], GANF [10], and MTGFlow [11]. However, existing GNNs rely on multi-hop message passing based on shortest-path distance, which overlooks richer local topological context (e.g., path density and connectivity patterns) and may introduce "noise leakage" by aggregating irrelevant information.

To address this limitation, we propose Hierarchical Relational Diffusion Network (HRD-Net) for multivariate time series anomaly detection. HRD-Net leverages random walk with restart (RWR) to capture topology-aware relational contexts, which are sensitive to local structural properties. By introducing multiple restart probabilities, HRD-Net models multi-scale dependencies and forms a hierarchical "topological firewall" that filters spurious correlations while preserving meaningful

interactions. This multi-scale structure further enables robust anomaly scoring by integrating representations across different levels.

Contributions:

- We propose HRD-Net, a relational diffusion framework based on RWR that captures topology-aware local structures beyond conventional multi-hop neighborhoods.
- We design a multi-scale mechanism with adaptive restart probabilities to suppress noise propagation and enhance message passing quality.
- Extensive experiments on six benchmark datasets demonstrate the effectiveness of HRD-Net with strong performance and interpretable case studies.

II. Related Work

Early time-series anomaly detection methods rely on density-, distance-, and one-class classification models [12], [13], while recent approaches are dominated by deep learning, particularly reconstruction-based models such as Autoencoders and their variants [14]–[18]. However, these methods often treat multivariate time series as a whole and overlook explicit inter-variable dependencies. To address this, graph-based methods leverage GNNs [19], [20] to model variables as nodes and relationships as edges, typically constructing graphs from pairwise correlations and detecting anomalies via prediction or reconstruction errors [7], [11]. Nevertheless, such approaches are limited by fixed-depth message passing. Recent works explore higher-order structures via diffusion and random-walk mechanisms [21], [22], where Random Walk with Restart (RWR) enables topology-aware dependency modeling. Yet, existing RWR-based methods are mainly designed for static graphs and rely on a single restart probability, limiting their ability to capture multi-scale structures. In contrast, our HRD-Net employs multiple restart probabilities to construct a hierarchical, multi-scale relational representation, enabling more expressive structural modeling and robust anomaly detection.

III. Methodology

The overall architecture of HRD-Net is shown in Figure 2, which comprises three key modules: (1) Base Graph Construction, (2) Hierarchical Relational Diffusion and Prediction, and (3) Multi-scale Loss and Anomaly Scoring.

Problem Formulation. Given a multivariate time series $\mathbf{X} \in \mathbb{R}^{N\times T}$ with N variables over T time steps, we detect anomalous time steps via next-step forecasting on a sliding window. At time t, the model takes a look-back window $\mathbf{X}_t = \{\mathbf{x}^{(t-L+1)}, \ldots, \mathbf{x}^{(t)}\} \in \mathbb{R}^{N\times L}$ as input and predicts $\mathbf{y} = \mathbf{x}^{(t+1)} \in \mathbb{R}^N$ via $\mathbf{g} = f(\mathbf{X}_t)$. The anomaly score at time t is defined by the prediction error, with larger errors indicating higher anomaly likelihood.

A. Base Graph Construction

We first construct a base graph (or adjacency matrix $\mathbf{A}$) to model the the intrinsic, pairwise relationships between the N different channels of a time series, which is the first step needed for all GNN-based methods. Unlike methods that learn embeddings, we compute this graph directly from the raw signals to preserve maximal information. The graph is represented by an adjacency matrix $\mathbf{A} \in \mathbb{R}^{N\times N}$, where each entry $\mathbf{A}_{ij}$ quantifies the correlation or similarity between channel i and channel j. The most well-known measure is the Pearson correlation coefficients,

$$\mathbf{A}_{ij} = \frac{\sum_{t=1}^{L}(x_i^{(t)} - \mu_i)(x_j^{(t)} - \mu_j)}{\sqrt{\sum_{t=1}^{L}(x_i^{(t)} - \mu_i)^2}\sqrt{\sum_{t=1}^{L}(x_j^{(t)} - \mu_j)^2}}. \quad (1)$$

Other measures (e.g., Spearman, MIC, Chatterjee) are also applicable; in our experiments we adopt Chatterjee correlation for stability.

The matrix $\mathbf{A}$ captures pairwise dependencies between all channels of the time series in the current time window, serving as the key input for subsequent relational diffusion.

B. Hierarchical Relational Diffusion and Prediction

Hierarchical Relational Diffusion and Prediction is the core module to build a contextualized relational fields. It uses RWR with varying jump-back probabilities to build multi-scale receptive fields, which are fine-tuned with a lightweight CNN layer before feeding to Graph Attention network (GAT) for channel wise predictions.

1) RWR-based Relational Diffusion: Traditional GNNs capture high-order dependencies via stacked message passing, but often ignore local topology (e.g., link density), introducing irrelevant neighbors and causing noisy updates and false positives.

To address this, we adopt *Random Walk with Restart (RWR)* [23], [24] to explore topology-aware, functionally relevant neighbors. Starting from node i, RWR iterates as

$$\mathbf{p}_i^{(l+1)} = (1-\alpha)\widehat{\mathbf{A}}\mathbf{p}_i^{(l)} + \alpha\mathbf{e}_i, \quad (2)$$

with closed-form solution

$$\mathbf{P} = (1-\alpha)(\mathbf{I} - \alpha\widehat{\mathbf{A}})^{-1}\mathbf{e}. \quad (3)$$

$\mathbf{P}$ serves as an augmented adjacency matrix encoding topology-aware proximity over all paths. The restart probability α controls diffusion: large α emphasizes local communities, while small α captures long-range dependencies.

We select $\alpha \in 0.1, \ldots, 0.9$ by minimizing $L_{\text{uni-scale}}$, and keep the top-K values to construct hierarchical graph views $\mathbf{P}^k$ (with $K = 3$).

2) CNN-based Pattern Learning: The RWR-generated graphs are further refined by a lightweight CNN, enabling end-to-end learning of relational structures and capturing recurring local motifs across time.

Each scale k is processed by a dedicated CNN block CNN_k (one 3×3 convolution with 16 kernels):

$$\widetilde{\mathbf{P}}^k = \text{CNN}_k(\mathbf{P}^k), \quad (4)$$

where k corresponds to a restart probability α^k.

The refined $\widetilde{\mathbf{P}}^k$ serves as a semantically enhanced adjacency matrix for the downstream GAT forecasting module.

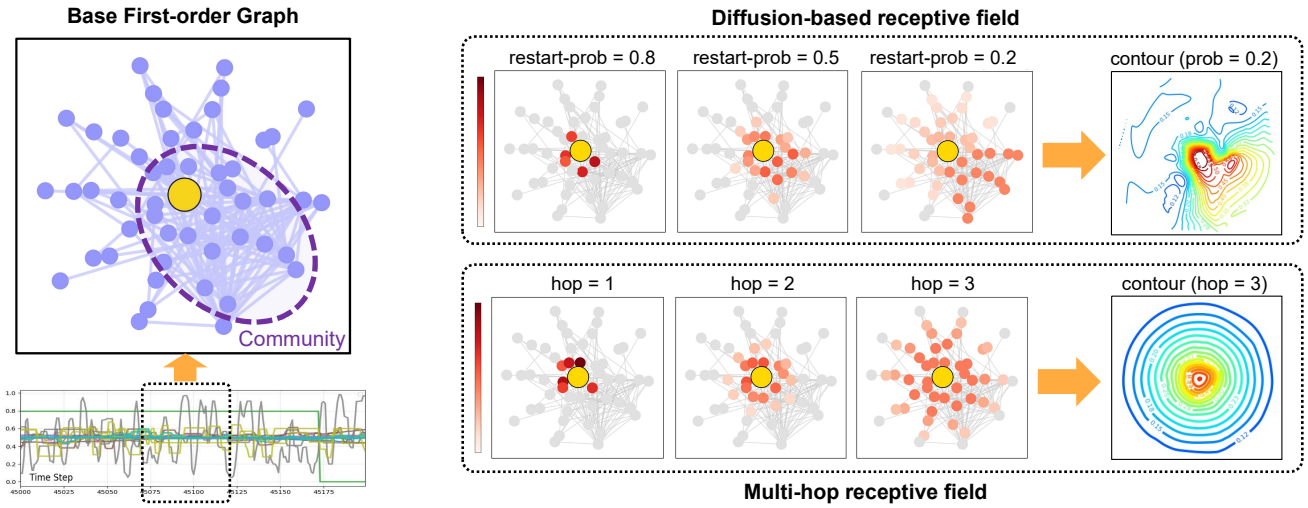

Fig. 1. Left: WADI time series and its correlation graph for a data segment. Right: receptive fields of random-walk diffusion (top - sensitive to local topological contexts) and GNN (bottom - based purely on hop distance) with different depths.

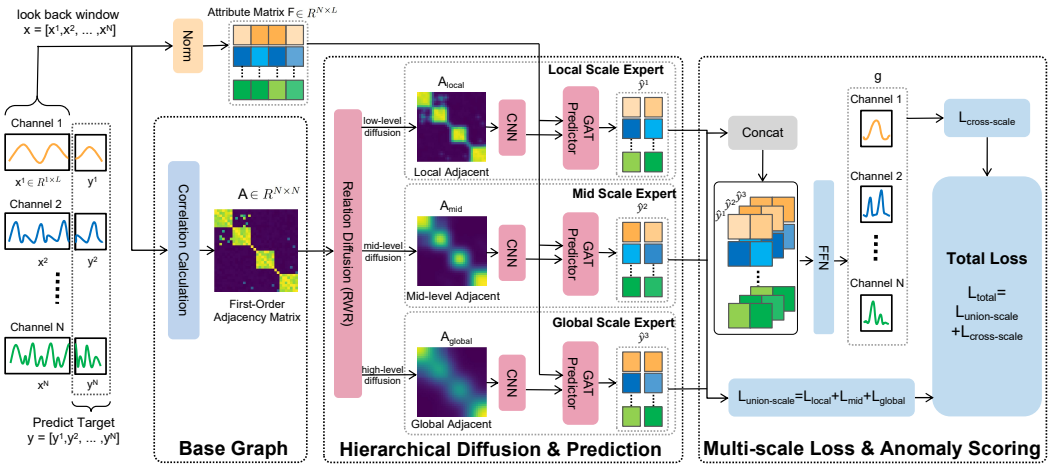

Fig. 2. The structure of HRD-Net, including (1) Base Graph Construction, (2) Hierarchical Diffusion and Prediction, and (3) Multi-scale Loss and Anomaly Scoring. See detailed introductions in the Method Section.

3) GAT-based Predictor: The GAT in each branch k takes the original time-series data $\mathbf{X}_t$ and the graph view $\widetilde{\mathbf{P}}^k$ corresponding to the k-th structural scale as inputs. We sparsify each matrix $\widetilde{\mathbf{P}}^k$ by preserving only the top-10% largest entries column-wise, yielding the final sparse graph $\widetilde{\mathbf{P}}_s^k$.

The GAT then update the representation for each node by attending over its neighbors. For each node i, its attention coefficients γ_{ij}^k are computed for its neighbors $j \in \mathcal{N}_i$ with $\mathcal{N}_i$ defined by the ith column of $\widetilde{\mathbf{P}}_s^k$ (including node i). An un-normalized attention score e_{ij}^k is calculated as:

$$e_{ij}^k = \text{LeakyReLU}(\mathbf{a}^T[\mathbf{W}\mathbf{x}_i || \mathbf{W}\mathbf{x}_j]), \tag{5}$$

where $\mathbf{W}$ is the shared linear transformation weight matrix, $||$ denotes concatenation, and $\mathbf{a}$ is a learnable attention vector. These scores are then normalized via the softmax function over the node's extended neighborhood $\mathcal{N}_i \cup \{i\}$:

$$\gamma_{ij}^k = \text{softmax}_j(e_{ij}^k) = \frac{\exp(e_{ij}^k)}{\sum_{l \in \mathcal{N}_i \cup \{i\}} \exp(e_{il}^k)}. \tag{6}$$

Finally, the updated representation for node i is computed as a weighted aggregation of transformed features from its neighbors and itself, followed by a non-linear activation σ:

$$\mathbf{z}_i^k = \sigma\left(\gamma_{ii}^k \mathbf{W}\mathbf{x}_i + \sum_{j \in \mathcal{N}_i} \gamma_{ij}^k \mathbf{W}\mathbf{x}_j\right). \tag{7}$$

We aggregate the node representations for each scale-k as $\mathbf{Z}^k = [\mathbf{z}_1^k, \ldots, \mathbf{z}_N^k]$, which is then turned to the prediction values for the scale-k with a feed-forward network, i.e., $\hat{\mathbf{y}}^k = \text{FFN}(\mathbf{Z}^k)$.

C. Multi-scale Loss Aggregation and Anomaly Scoring

HRD-Net leverages multi-scale predictions to build a robust anomaly scoring mechanism.

We first compute the average per-scale error:

$$\mathcal{L}_{\text{uni-scale}} = \frac{1}{K}\sum_{k=1}^{K}\frac{1}{N}\sum_{i=1}^{N}(y_i - \hat{y}_i^k)^2. \tag{8}$$

To fuse multi-scale information, each $\hat{\mathbf{y}}^k$ is transformed by $\mathbf{W}^k$ and concatenated, followed by a feedforward layer:

$$\mathbf{g} = f_\theta\left([\mathbf{W}^1\hat{\mathbf{y}}^1 \;||\; \cdots \;||\; \mathbf{W}^K\hat{\mathbf{y}}^K]\right). \tag{9}$$

The fused prediction yields a cross-scale loss:

$$\mathcal{L}\text{cross-scale} = \frac{1}{N}\sum i = 1^N(y_i - g_i)^2. \tag{10}$$

The final objective combines both:

$$\mathcal{L}\text{total} = \mathcal{L}\text{uni-scale} + \mathcal{L}_{\text{cross-scale}}. \tag{11}$$

At inference, anomaly scores are computed from normalized prediction errors $|y_i - g_i|$, and the system-level score is the maximum across sensors; values above a threshold indicate anomalies.

IV. Experiment

1) Datasets: We used six popular benchmarks in recent SOTA studies: SWaT [25], with 51 channels from a water treatment facility with attack instances; WADI [26], a water distribution system with 123 sensors in 2017,2019; PSM [27], with 25 expert-labeled metrics from eBay's server infrastructure; MSL [28], with 55 sensors from the Mars rover; and SMD [29], a 38-dimensional dataset from an internet company.

2) Evaluation Metrics: Following recent work on rigorous evaluation [30], we report point-wise F1-Score for fair comparison. We do not use Point-Adjustment (PA), as it can overestimate performance by marking an entire anomaly segment as detected if only one point is identified. All metrics are computed at the time-step level.

3) Baselines: We selected 12 competitive methods (classical, reconstruction-based, GNN-based and Transformer-based), including: LSTM-AE [16], MAD-GAN [18], MTAD-GAT [6], USAD [31], GDN [7], GTA [9], Anomaly Transformer [32],GANF [10], NPSR [33], MTGFlow [11], M2N2 [34] and SimAD [35].

4) Experimental Setting: Our implementation is based on PyTorch and runs on an NVIDIA RTX 3090 GPU. The input window size is set to $L = 5$ for fast-paced industrial datasets (SWaT, WADI-17/19) and $L = 15$ for server datasets (PSM, MSL, SMD). We use $K = 3$ adaptive scales, with restart probabilities selected from $0.1, \ldots, 0.9$ based on the smallest $L_{\text{uni-scale}}$.Each scale adopts a CNN module with stacked 3×3 convolutions (16 channels, padding=1). The GAT backbone has two layers with 64/128 hidden dimensions depending on dataset complexity.We train using Adam (lr=10^{-3}, $\beta_1 = 0.9$, $\beta_2 = 0.99$), batch size 64, for up to 60 epochs with early stopping (patience 15). Baselines follow their official implementations.

A. Results

Results in Table I show that HRD-Net outperforms both Transformer-based and GNN-based methods with an average F1 improvement of 3.2%. Despite this, it remains highly parameter-efficient (20k parameters), significantly smaller than recent Transformer baselines. The lower performance of Anomaly Transformer is due to our use of non-adjusted F1-scores, consistent with prior studies [33], [34].

B. Ablation Studies

Impact of HRD-Net Components. We evaluate each component in Table II.

Removing relational diffusion ("w/o Relation Diffusion") causes significant drops (3.7% on SWaT, 3.5% on WADI-19), highlighting its critical role.

Using only a single scale also degrades performance: global-only drops 2.8%, and local-only drops 2.6% on WADI-19, showing that hierarchical multi-scale information is necessary.

Removing CNN ("w/o CNN") leads to smaller but consistent declines (1.6% on SWaT, 1.4% on WADI-19), indicating it further refines relational patterns beyond RWR.

Loss Function Design. We also investigated the design of the loss function (anomaly scores) by comparing several alternatives: (1) using only the uni-scale loss $L_{\text{uni-scale}}$, (2) using only the cross-scale loss $L_{\text{cross-scale}}$, and (3) different aggregation strategies such as imbalanced summation, max, or min aggregation of the two losses. Table III reports the corresponding results.

The comparisons show that the balanced summation of cross-scale and uni-scale losses consistently yields the best performance, while relying solely on either loss degrades results. Moreover, max/min aggregation proves unstable and leads to significantly poorer performance, and summing all individual branch losses without averaging also underperforms. These findings indicate that both types of prediction errors are important and should be properly balanced to reflect the system anomaly state.

Case Study on "Topological Firewall". We use the SWaT dataset with 51 sensors for a case study. Here, we select a short segment of 500 time steps during a **normal** operational period, focusing on sensor AIT202 as the center node of interest. As in Figure 3 (left), the base correlational graph reveals several functional communities. AIT202 is embedded within cluster A, surrounded by topologically and functionally related sensors, while also maintaining a first-order connection to P502, which belongs to a distinct cluster B. Figure 3 (right) further highlights the signal characteristics: AIT202 exhibits a relatively flat pattern during normal operation, while P502 displays routine periodic fluctuations.

Given this topological and signal context, we compare the behavior of HRD-Net and a conventional GNN:

- Our HRD-Net assigns **a very low weight (0.063)** to AIT202-P502 connection. This low weight stems from the local view (α=0.9), which identifies that the nodes belong to distinct communities and are unrelated. This weak linkage acts as a "topological firewall" to suppress

TABLE I

F1-SCORE (WITHOUT POINT-ADJUSTMENT) FOR 12 METHODS IN TIME SERIES ANOMALY DETECTION ON SIX BENCHMARK DATASETS.

Method	Year	Category	Datasets						Avg.
			SWaT	WADI-17	WADI-19	PSM	MSL	SMD	
LSTM-VAE	2018	VAE-based	77.6	22.7	20.6	45.5	21.2	43.5	38.5
MAD-GAN	2019	GAN-based	77.0	37.0	33.1	47.1	26.7	22.0	40.5
MTAD-GAT	2020	GNN-based	78.4	43.7	41.0	52.0	27.5	41.0	47.3
USAD	2020	AE-based	79.2	23.3	20.6	47.9	21.1	42.6	39.1
GDN	2021	GNN-based	81.2	57.0	47.1	50.3	21.7	44.1	50.2
GTA	2021	GNN-based	76.1	53.1	51.5	51.3	21.8	43.5	49.6
GANF	2022	GNN-based	77.8	44.5	43.2	52.3	25.6	44.3	48.0
Anomaly Transformer	2022	Transformer	22.1	10.0	11.2	43.4	19.1	11.1	19.5
NPSR	2023	Transformer	82.0	55.0	51.6	51.0	26.1	47.6	52.2
MTGFlow	2023	GNN-based	79.7	46.0	44.3	53.2	21.7	35.1	46.7
M2N2	2024	AE-based	78.4	14.8	15.6	43.3	30.2	36.1	36.4
SimAD	2025	Transformer	82.0	**63.0**	53.2	52.1	30.0	33.7	52.3
Ours (HRD-Net)	2026	GNN-based	**84.8**	57.1	**55.1**	**53.5**	**33.3**	**49.3**	**55.5**

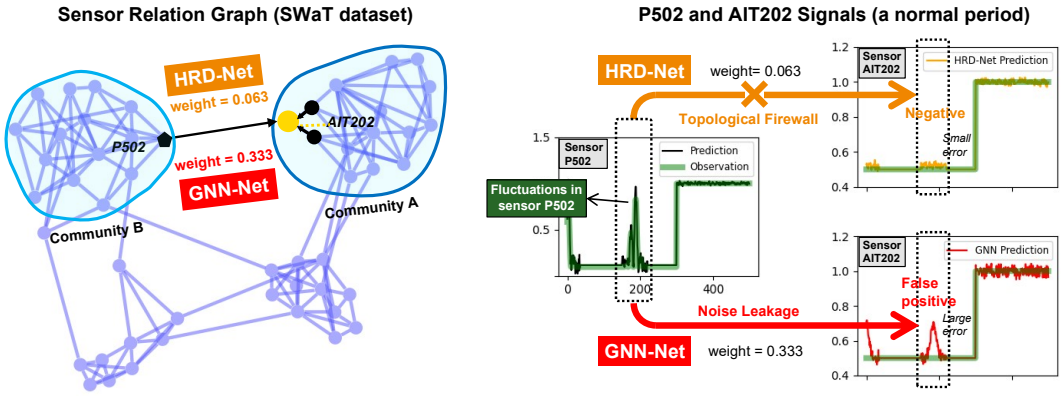

Fig. 3. **Left**: base correlation graph for SWaT, with AIT202 embedded in functional cluster A and its 1st-order neighbor P502 in cluster B; GNN assigns a high weight when aggregating info. from P502 to AIT202, while HRD-Net recognizes P502 as irrelevant and assigns a low weight. **Right**: HRD-Net blocks propagation of P502 signal fluctuations to AIT202; GNN allows this noise leakage to contaminate AIT202 features, resulting in inflated prediction errors and false positive alerts

TABLE II

ABLATION STUDIES ON HRD-NET MODULES

Components	SWaT	WADI-19
HRD-Net (Full Model)	**84.8**	**55.0**
w/o Relation Diffusion	81.1	51.5
w/o Hierarchy (Global Only)	82.7	52.2
w/o Hierarchy (Local Only)	82.5	52.4
w/o CNN	83.2	53.6

TABLE III

ABLATION STUDY OF TOTAL LOSS DEFINITION

Loss Function Strategy	SWaT	WADI-19
Cross-Scale + Uni-Scale(Ours)	**84.8**	**55.0**
Cross-Scale Only	82.9	52.1
Uni-Scale Only	82.3	53.4
Sum of All Losses	83.5	51.1
Max Aggregation	82.1	51.2
Min Aggregation	81.3	48.3

irrelevant signals from P502 - a low prediction error for AIT202 thus follows and correctly identified as normal.

- In contrast, GNN assigns **a much higher weight (0.333)** to the same edge merely due to the first-order connection between AIT202 and P505. This naive aggregation disregards functional context and introduces P502's fluctuating signal into AIT202's feature update - a "noise leakage" to corrupt the prediction for AIT202, inflate its prediction error, and finally triggers false positive detection.

This case study reveals how HRD-Net leverages topological context to selectively filter neighbors in the message-passing process, thereby reducing prediction error and avoiding false positives. As shown in Figure 4, HRD-Net achieves a higher F1-score primarily by minimizing such false positive alarms (the overlap of the blue and orange curves) compared to GNN-based model (WADI-19 dataset).

V. CONCLUSION

In this paper, we proposed the Hierarchical Relational Diffusion Network (HRD-Net), a novel framework that uses

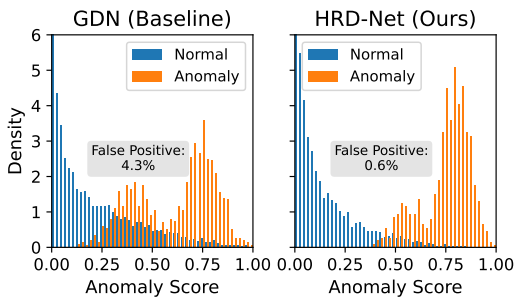

Fig. 4. Distribution of anomaly scores for normal (blue) and anomalous time-steps (orange) in GDN and HRD-Net.

a random-walk based diffusion mechanism to build a contextualized relational field. By using adaptive restart probabilities, HRD-Net forms a hierarchical "topological firewall" that balances local and global contexts to ensure robust message passing, and hence reliable profiling of the diverse anomalous state of complex systems. Experiments on six benchmarks validated the effectiveness of our method.

REFERENCES

[1] A. Blázquez-García, A. Conde, U. Mori, and J. A. Lozano, "A review on outlier/anomaly detection in time series data," *ACM computing surveys (CSUR)*, vol. 54, no. 3, pp. 1–33, 2021.

[2] S. Crépey, N. Lehdili, N. Madhar, and M. Thomas, "Anomaly detection in financial time series by principal component analysis and neural networks," *Algorithms*, vol. 15, no. 10, p. 385, 2022.

[3] J. Pereira and M. Silveira, "Learning representations from healthcare time series data for unsupervised anomaly detection," in *2019 IEEE international conference on big data and smart computing (BigComp)*. IEEE, 2019, pp. 1–7.

[4] T. Chen *et al.*, "Anomaly detection in semiconductor manufacturing through time series forecasting using neural networks," Ph.D. dissertation, Massachusetts Institute of Technology, 2018.

[5] M. Bawaneh and V. Simon, "Anomaly detection in smart city traffic based on time series analysis," in *2019 International Conference on Software, Telecommunications and Computer Networks (SoftCOM)*. IEEE, 2019, pp. 1–6.

[6] Y. Zhao, Z. Nasrullah, and Z. Li, "Multivariate time-series anomaly detection via graph attention network," in *IEEE International Conference on Big Data (Big Data)*. IEEE, 2020, pp. 2563–2573.

[7] A. Deng and B. Hooi, "Graph neural network-based anomaly detection in multivariate time series," in *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 35, no. 5, 2021, pp. 4027–4035.

[8] W. Zhang, C. Zhang, and F. Tsung, "Grelen: Multivariate time series anomaly detection from the perspective of graph relational learning." in *IJCAI*, 2022, pp. 2390–2397.

[9] J. Chen, X. Zhang, Z. He, and Y. Wang, "Learning graph structure with transformer for multivariate time series anomaly detection," in *Proceedings of the 30th ACM International Conference on Information and Knowledge Management (CIKM)*, 2021, pp. 2913–2917.

[10] E. Dai and J. Chen, "Graph-augmented normalizing flows for anomaly detection of multiple time series," in *International Conference on Learning Representations*, 2022.

[11] Q. Zhou, J. Chen, H. Liu, S. He, and W. Meng, "Detecting multivariate time series anomalies with zero known label," in *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 37, no. 4, 2023, pp. 4963–4971.

[12] F. Angiulli and F. Fassetti, "Detecting distance-based outliers in streams of data," in *Proceedings of the sixteenth ACM conference on Conference on information and knowledge management*, 2007, pp. 811–820.

[13] B. Schölkopf and A. J. Smola, *Learning with kernels: support vector machines, regularization, optimization, and beyond*. MIT press, 2002.

[14] C. C. Aggarwal *et al.*, *Data mining: the textbook*. Springer, 2015, vol. 1, no. 3.

[15] D. P. Kingma, M. Welling *et al.*, "Auto-encoding variational bayes," 2013.

[16] D. Park, Y. Hoshi, and C. C. Kemp, "A multimodal anomaly detector for robot-assisted feeding using an lstm-based variational autoencoder," *IEEE Robotics and Automation Letters*, vol. 3, no. 3, pp. 1544–1551, 2018.

[17] C. Zhang, D. Song, Y. Chen, X. Feng, C. Lumezanu, W. Cheng, J. Ni, B. Zong, H. Chen, and N. V. Chawla, "A deep neural network for unsupervised anomaly detection and diagnosis in multivariate time series data," in *Proceedings of the AAAI conference on artificial intelligence*, vol. 33, no. 01, 2019, pp. 1409–1416.

[18] D. Li, D. Chen, B. Jin, L. Shi, J. Goh, and S.-K. Ng, "Mad-gan: Multivariate anomaly detection for time series data with generative adversarial networks," in *International conference on artificial neural networks*. Springer, 2019, pp. 703–716.

[19] T. N. Kipf and M. Welling, "Semi-supervised classification with graph convolutional networks," in *International Conference on Learning Representations (ICLR)*, 2017.

[20] P. Velickovic, G. Cucurull, A. Casanova, A. Romero, P. Lio, and Y. Bengio, "Graph attention networks," in *International Conference on Learning Representations (ICLR)*, 2018.

[21] J. Atwood and D. Towsley, "Diffusion-convolutional neural networks," *Advances in neural information processing systems*, vol. 29, 2016.

[22] J. Gasteiger, S. Weißenberger, and S. Günnemann, "Diffusion improves graph learning," *Advances in neural information processing systems*, vol. 32, 2019.

[23] L. Lovász and P. Winkler, "A note on the last new vertex visited by a random walk," *Journal of graph theory*, vol. 17, no. 5, pp. 593–596, 1993.

[24] H. Tong, C. Faloutsos, and J.-Y. Pan, "Fast random walk with restart and its applications," in *Sixth international conference on data mining (ICDM'06)*. IEEE, 2006, pp. 613–622.

[25] J. Goh, S. Adepu, K. N. Junejo, and A. Mathur, "A dataset to support research in the design of secure water treatment systems," in *International conference on critical information infrastructures security*. Springer, 2016, pp. 88–99.

[26] C. M. Ahmed, V. R. Palleti, and A. P. Mathur, "Wadi: a water distribution testbed for research in the design of secure cyber physical systems," in *Proceedings of the 3rd international workshop on cyber-physical systems for smart water networks*, 2017, pp. 25–28.

[27] A. Abdulaal, Z. Liu, and T. Lancewicki, "Practical approach to asynchronous multivariate time series anomaly detection and localization," in *Proceedings of the 27th ACM SIGKDD conference on knowledge discovery & data mining*, 2021, pp. 2485–2494.

[28] K. Hundman, V. Constantinou, C. Laporte, I. Colwell, and T. Soderstrom, "Detecting spacecraft anomalies using lstms and nonparametric dynamic thresholding," in *Proceedings of the 24th ACM SIGKDD international conference on knowledge discovery & data mining*, 2018, pp. 387–395.

[29] Y. Su, Y. Zhao, C. Niu, R. Liu, W. Sun, and D. Pei, "Robust anomaly detection for multivariate time series through stochastic recurrent neural network," in *Proceedings of the 25th ACM SIGKDD international conference on knowledge discovery & data mining*, 2019, pp. 2828–2837.

[30] D. Wagner, T. Michels, F. C. Schulz, M. Rudolph, and M. Kloft, "Timesead: Benchmarking deep time-series anomaly detection," 2022.

[31] J. Audibert, P. Michiardi, F. Guyard, S. Marti, and M. A. Zuluaga, "Usad: Unsupervised anomaly detection on multivariate time series," in *Proceedings of the 26th ACM SIGKDD international conference on knowledge discovery & data mining*, 2020, pp. 3395–3404.

[32] J. Xu, H. Wu, J. Wang, and M. Long, "Anomaly transformer: Time series anomaly detection with association discrepancy," in *International Conference on Learning Representations*, 2022.

[33] C.-Y. A. Lai, F.-K. Sun, Z. Gao, J. H. Lang, and D. Boning, "Nominality score conditioned time series anomaly detection by point/sequential reconstruction," *Advances in Neural Information Processing Systems*, vol. 36, pp. 76 637–76 655, 2023.

[34] D. Kim, S. Park, and J. Choo, "When model meets new normals: Test-time adaptation for unsupervised time-series anomaly detection," in *Proceedings of the AAAI conference on artificial intelligence*, vol. 38, no. 12, 2024, pp. 13 113–13 121.

[35] Z. Zhong, Z. Yu, X. Xi, Y. Xu, W. Cao, Y. Yang, K. Yang, and J. You, "Simad: A simple dissimilarity-based approach for time series anomaly detection," 2025.