2025 IEEE International Conference on Web Services (ICWS) | 979-8-3315-5563-4/25/$31.00 ©2025 IEEE | DOI: 10.1109/ICWS67624.2025.00096

MGG-AD: Multi-Granularity Graph-Based Anomaly Detection in IoT Systems

Yi Li*, Zhangbing Zhou*†, Boris Sedlak‡, Schahram Dustdar‡
* *School of Information Engineering, China University of Geosciences (Beijing), Beijing 100083, China*
† *Computer Science Department, TELECOM SudParis, Evry 91011, France*
‡ *Distributed Systems Group, TU Wien, Vienna 1040, Austria*

***Abstract*—Internet of Things (IoT) systems gained significant attention for monitoring and optimizing processes. To ensure real-time detections with low latency, IoT applications often monitor individual components through a microservice network, deployed close to IoT devices. Existing methods for multivariate time series anomaly detection typically construct one global graph for identifying deviations in predicted or reconstructed attribute features. However, consider an active node that suddenly experiences a sharp drop in connections or established unexpected links; these structural anomalies would be totally overlooked. To address these limitations, this paper proposes Multi-Granularity Graph Anomaly Detection (MGG-AD), a novel approach that captures both attribute and topological dependencies within IoT systems. Specifically, we construct a multi-granularity dependency graph from a global graph and multiple local subgraphs that define geographical correlations among IoT devices. First, at the attribute-level, we detect contextual deviations by reconstructing feature representations and contrasting attributes across local subgraphs. Second, at the topological-level, we identify abnormal structural variations by comparing local subgraphs with the global graph—called contrastive learning. We evaluated MGG-AD on two publicly available datasets and against state-of-the-art methods—we found that our solution provides higher detection accuracy and robustness, underlining its suitability for dynamic IoT systems.**

I. Introduction

ANOMALY detection is an integral part of microservice workflows for ensuring the reliability of individual services [1]. To optimize the stability and efficiency of processes, Internet of Things (IoT) applications commonly monitor multivariate time series throughout highly distributed IoT devices [2], [3]. To detect abnormal behavior in such IoT systems, the monitoring process can be split into multiple web services that are executed on geographically distributed edge devices [4]. However, anomalies occur infrequently, and some anomaly types may not manifest within a given monitoring period [5], leading to highly imbalanced datasets [6]. Furthermore, the absence of representative samples for certain anomaly types hinders comprehensive modeling of task dependencies in web services, making anomaly detection particularly challenging.

Existing multivariate time series anomaly detection methods can be broadly categorized into generative [7] and predictive approaches [8]. These methods primarily focus on capturing feature representations of normal system attributes and identifying anomalies based on deviations in reconstruction or prediction [9]. Generative models, such as Generative Adversarial Networks (GANs), and predictive models, like Graph Neural Networks (GNNs), extract attribute features from time series data, where each attribute is typically associated with a sensor device deployed in a network. However, in real-world IoT-based web services, attributes often result from complex interactions among multiple devices, making the one-to-one correspondence assumption rarely valid [10]. This limitation not only increases deployment costs and energy consumption but also fails to capture complex dependencies across spatially distributed devices.

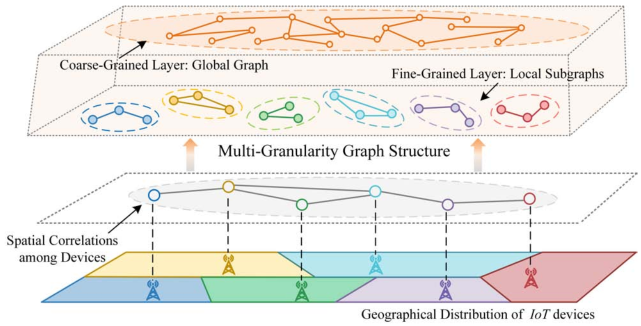
Coarse-Grained Layer: Global Graph
Fine-Grained Layer: Local Subgraphs
Multi-Granularity Graph Structure
Spatial Correlations among Devices
Geographical Distribution of *IoT* devices

Fig. 1. Our Multi-Granularity Graph Structure (MGG) and its motivation. At the bottom, IoT devices are distributed geographically. The third layer captures spatial correlations, guiding local subgraph construction. The fine-grained layer models local attribute dependencies, while the coarse-grained layer represents a global dependency graph for system-wide interactions.

To capture the intricate dependencies between attributes from multiple interacting sensors, researchers have explored graph-based modeling strategies. Fig. 1 shows an example for such a strategy, where multiple IoT devices—6 in the bottom layer—are deployed across different locations to monitor various system attributes—3 for each, as in the local subgraphs. Existing methods typically only construct a global graph, as shown in the topmost layer, where all attributes are treated as a unified entity, merging both intra-regional and cross-regional attributes. While such a holistic modeling has demonstrated effectiveness in anomaly detection [11], it fails to capture fine-grained spatio-temporal dependencies, inherent in web services. In particular, critical local dependencies between neighboring devices may be overshadowed when aggregating information at a global scale, limiting the ability to detect fine-grained anomalies in complex environments.

Recent advancements in graph anomaly detection have

2836-3868/25/$31.00 ©2025 IEEE
DOI 10.1109/ICWS67624.2025.00096

introduced contextual subgraph modeling to address this limitation. For example, SL-GAD [12] constructs subgraphs centered on target nodes and employs contrastive learning to align local subgraphs with global representations, enhancing discriminative ability in anomaly detection. As shown by the white layer in Fig. 1, the geographical deployment of IoT devices creates a localized graph encoding spatio-temporal dependencies. Inspired by this, we introduce a Multi-Granularity Graph (MGG) representation, where (1) local subgraphs preserve spatially correlated attribute relationships and (2) a global graph captures system-wide attribute interactions. Instead of treating all attributes as unified, we partition the system into multiple subgraphs to capture multi-granular dependencies of web services at different levels. Beyond local dependencies, spatial relationships among devices form a hierarchical structure, revealing correlations between local subgraphs and the global graph. Capturing these topological dependencies is essential for improving predictive or reconstructive performance—enhancing detection accuracy.

Despite recent advances in graph-based modeling, existing methods [13] primarily focus on attribute-level dependencies across monitored metrics, overlooking structural anomalies (e.g., a web service suddenly communicating with a new node). Consequently, they struggle to identify anomalies manifesting as topological deviations rather than feature irregularities. To address this, we propose a Multi-Granularity Graph-based Anomaly Detection (MGG-AD) framework, which captures both attribute- and structure-level anomalies in IoT web services. We first construct a multi-granularity dependency graph to model local and global correlations between IoT devices and their monitored attributes. Building on this structure, we introduce a hybrid learning framework combining generative and contrastive approaches. At the attribute level, a graph attention network (GAT)-based encoder-decoder reconstructs node features across local subgraphs, ensuring feature consistency. At the structure level, an attention-based readout module encodes graph-level representations to highlight topological deviations. By leveraging multi-granularity graphs, MGG-AD effectively captures dependencies across different scales, enabling robust detection of diverse anomaly patterns.

To summarize, the main contributions of our work are:

- A multi-granularity dependency graph that models both local and global attribute correlations, capturing hierarchical dependencies for effective anomaly detection.
- A hybrid generative-discriminative framework is introduced, where generative modeling ensures feature consistency for attribute-level anomalies detection, and contrastive learning measures structural deviation across different granularities.
- A comprehensive anomaly scoring mechanism that integrates attribute-level reconstruction errors and structural contrastive deviations, enabling accurate identification of anomalous behaviors.

To evaluate our solution—the MGG-AD framework—we conducted extensive experiments on two publicly available datasets. The results show that our method achieves higher accuracy than established techniques for anomaly detection.

The reminder of this paper is organized as follows. Section II introduces the two fundamental types of anomalies. Section III provides an overview of the proposed framework and its core components. Section IV to Section VI describe the multi-granularity graph construction, the generative-contrastive modeling process, and the anomaly scoring mechanism in detail. Section VII presents experimental setups and evaluation results. Related work is discussed in Section VIII, and Section IX concludes the paper.

II. Preliminaries

This section defines the two fundamental types of anomalies in IoT systems. For the remaining paper, we broadly categorize them into attribute anomalies and structural anomalies.

*1) **Attribute Anomalies**:* An attribute anomaly occurs when the statistical or functional relationships among different system attributes deviate significantly from expected patterns. When monitoring device utilization, possible attributes are CPU and memory usage, or I/O load, which are also present in the Application Server Dataset[1] (ASD) evaluated later.

For example, high CPU utilization is typically associated with increased I/O activity and memory consumption. However, if CPU usage surges unexpectedly while I/O activity remains low, this may indicate an impending system failure or a software malfunction.

*2) **Structural Anomalies**:* A structural anomaly refers to unexpected changes in the topological relationships between system entities, as such, they disrupt usual interaction patterns. This class of anomalies also occur in the later evaluated Server Machine Dataset[2](SMD), indicating a web server malfunction.

For example, servers typically exchange network traffic continuously. If a previously active server suddenly experiences a sharp drop in connections or establishes unexpected links, this may indicate a network anomalies or failure.

III. Methodology Overview

To systematically detect attribute anomalies and structural anomalies in IoT systems, we introduce a coherent methodology that combines multi-granularity graph modeling, generative learning, and contrastive learning. The core idea is to jointly model both attribute dependencies and structural relationships, allowing for a more comprehensive detection of anomalous behaviors. As illustrated in Fig. 2, our proposed framework consists of three main components:

1) Multi-Granularity Graph Learning: We construct a hierarchical graph structure that represent attribute dependencies at multiple topological levels. A fine-grained graph is built to model local spatially correlated attributes within specific system regions, preserving local attribute interactions. Meanwhile, a coarse-grained graph captures system-wide dependencies and provides a global perspective on attribute relationships. This multi-granularity representation effectively

[1] https://github.com/zhhlee/InterFusion/tree/main/data

[2] https://github.com/NetManAIOps/OmniAnomaly/tree/master/ServerMachineDataset

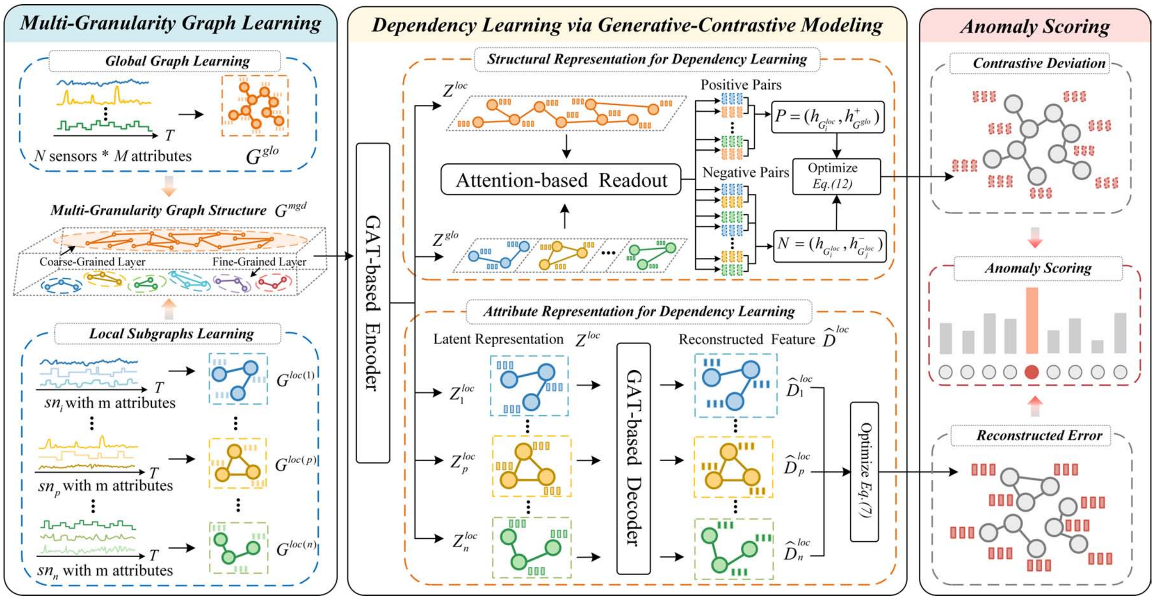

Fig. 2. Framework of MMG-AD for Anomaly Detection. The framework consists of three key modules: (i) *Multi-Granularity Graph Learning*, where local subgraphs and a global graph are constructed to form the multi-granularity dependency graph; (ii) *Dependency Learning via Generative-Contrastive Modeling*, which captures structural representations and attribute representations to model dependencies; and (iii) *Anomaly Scoring*, where anomalies are identified by combining contrastive deviations and reconstruction errors.

encodes both localized variations and high-level structural dependencies within the system.

2) Dependency Learning via Generative-Contrastive Modeling: To reconstruct node attributes and learn the normal patterns of attribute relationships, we employ a GAT-based encoder-decoder architecture. Simultaneously, a contrastive learning module is designed to enforce structural consistency by modeling the relationships between subgraphs and the global graph. By jointly optimizing these two module with a combined loss function, the model learns representations that distinguish normal patterns from anomalies across both attribute and structure levels.

3) Anomaly Scoring: Anomalies are identified based on deviations from learned normal patterns. The anomaly score is computed as a combination of reconstruction error, which reflects deviations in attribute values, and contrastive deviation, which captures structural inconsistencies. By integrating these two perspectives, the framework ensures comprehensive scoring and robust anomaly detection.

IV. Multi-Granularity Graph Learning

This section presents the construction of the multi-granularity dependency graph, which serves as the foundation for anomaly detection. We first introduce the dependency graph structure, outlining how attribute relationships are captured, and then describe the multi-granularity graph construction, where local and global dependency graphs are integrated into a unified representation. This hierarchical design allows the model to effectively capture both fine-grained attribute dependencies and global structural patterns.

A. Dependency Graph Structure Learning

To model the complex dependencies among sensor attributes in IoT systems, we construct a dependency graph $\mathcal{G}$ that encodes spatio-temporal relationships within multivariate time series data. In an IoT system, n devices are deployed across different geographical locations, with each device monitoring m attributes. The collected multivariate time series data can be represented as: $D = \{d_{ij}(t) | i \in \{1, 2, ..., N\}, j \in \{1, 2, ..., M\}, t \in \{1, 2, ..., T\}\}$, where $d_{ij}(t)$ denotes the observation of attribute j from device i at timestamp t.

An embedding vector ev_{ij} is introduced to capture both temporal and spatial dependencies among attributes across different devices. The dependency strength between two attributes is measured by their embedding similarity, where a higher similarity value indicates a stronger dependency.

To construct the graph structure, we first initialize a fully connected weighted graph, where each attribute node considers all others as candidate neighbors. The actual neighbors are then selected based on their similarity scores. Specifically, the normalized dot product $\mathcal{T}$ is computed to quantify the similarity between node sn_{ij} and its candidate neighboring node $sn_{il} \in cdn_{ij}$:

$$\mathcal{T}_{ilij} = \frac{ev_{ij} \ ev_{il}}{\|ev_{ij}\| \cdot \|ev_{il}\|}, \quad sn_{il} \in cdn_{ij}, \tag{1}$$

where $\mathcal{T}_{ilij}$ measures the dependency strength between node sn_{ij} and its candidate neighboring node sn_{il}.

This process allows that the constructed dependency graph $\mathcal{G}$ effectively encodes attribute relationships, serving as the foundation for multi-granularity graph modeling.

B. Multi-Granularity Graph Learning

To effectively model both localized attribute dependencies and system-wide correlations in IoT systems, we construct a multi-granularity dependency graph $\mathcal{G}^{mgd}$. This structure integrates local dependency subgraphs, which capture fine-grained relationships within individual devices, and a global dependency graph, which encodes broader correlations among attributes across different devices. By combining these two perspectives, the multi-granularity graph provides a comprehensive representation of the system, facilitating robust anomaly detection.

1) Local Dependency Graph: To exploit fine-grained attribute dependencies within individual devices, we construct multiple local dependency subgraphs $\mathcal{G}^{loc} = \{\mathcal{G}^{loc(1)}, \mathcal{G}^{loc(2)}, ..., \mathcal{G}^{loc(n)}\}$. Each subgraph $\mathcal{G}^{loc(p)}$ corresponds to a single device and consists of m nodes, each representing one monitored attribute.

We first initialize a fully connected subgraph for each device, where the dependence strength between nodes is calculated using the similarity metric in Equation 1. To construct a sparse yet informative structure, each node retains only its top-k most similar neighbors, forming the final local adjacency matrix:

$$\mathbf{A}^{loc} = [\mathbf{A}^{loc(1)}, ..., \mathbf{A}^{loc(p)}, ..., \mathbf{A}^{loc(n)}], \tag{2a}$$

$$\mathbf{A}^{loc(p)}_{ilij} = 1\{il \in Topk(\{\mathcal{T}^{loc(1)}_{ilij} : ih \in cdn_{ij}\})\}, \tag{2b}$$

where $\mathbf{A}^{loc(p)}_{ilij} \in \{0, 1\}$ denotes whether an edge exists from node sn_{ij} to sn_{il}. The parameter k controls the sparsity of each local subgraph and is fine-tuned through experiments.

2) Global Dependency Graph: To model system-wide attribute correlations, we construct a global dependency graph $\mathcal{G}^{glo}$ by evaluating the similarity among all $n \times m$ nodes. Similar to the local subgraph construction, each node retains only its top-k' most similar neighbors, forming the global adjacency matrix: T

$$\mathbf{A}^{glo}_{ilij} = 1\{il \in Topk'(\{\mathcal{T}_{ihij} : ih \in cdn_{ij}\})\}, \tag{3}$$

where $\mathbf{A}^{glo}_{ilij} \in \{0, 1\}$ indicates whether a directed edge exists from node sn_{ij} to sn_{il}. k' determines the sparsity of the global graph and is adjusted based on empirical evaluation.

3) Integration into a Multi-Granularity Graph: As illustrated in Fig. 1, the multi-granularity dependency graph $\mathcal{G}^{mgd}$ is structured into two hierarchical layers, each capturing dependencies at different scales. Before constructing these layers, as shown in the gray region of Fig. 1, the system is first partitioned based on the geographical distribution of IoT devices, preserving spatial correlations among them. (i) *Fine-Grained Layer (Bottom Layer in Orange Region):* Each local dependency subgraph represents attribute relationships within a single device, capturing fine-grained dependencies among its monitored attributes. (ii) *Coarse-Grained Layer (Top Layer in Orange Region):* This layer models attribute dependencies across all devices, providing a system-wide perspective on relationships and facilitating the detection of global anomalies.

By integrating information from local and global perspectives, this multi-granularity graph provides a structured representations of the IoT system. It effectively captures both attribute-level and structure-level anomalies, enhancing the accuracy and robustness of anomaly detection.

V. Dependency Learning Via Generative-Contrastive Modeling

To capture both attribute-level and structure-level dependencies, we employ a hybrid modeling approach that integrates generative learning for attribute representation and contrastive learning for structural consistency. First, a GAT-based encoder-decoder reconstructs node features across subgraphs to ensure representation consistency (Section V-A). Then, a readout module generates graph-level representations to assess structural similarity and detect anomalies (Section V-B). This framework effectively captures both fine-grained feature dependencies and global structural patterns.

A. Attribute Dependency Learning via Generative Modeling

To model attribute dependencies in a multi-granularity graph, we design a GAT-based encoder-decoder framework, which captures feature correlations among nodes and reconstructs their attribute representations. By learning normal patterns of attribute relationships, the framework facilitates anomaly detection through reconstruction errors.

1) Graph Attention Network for Attribute Representation: The encoder and decoder both leverage GAT to model dependencies between nodes and their neighbors. In the multi-granularity graph structure $\mathcal{G}^{mgd}$, each node represents an IoT device, and edges encode dependencies among devices. The encoder maps input node features $\mathbf{D}$ into a latent representation: $\mathbf{Z} = \mathrm{GAT}_{\mathrm{enc}}(\mathbf{D}, \mathbf{A})$. The attention mechanism assigns dynamic weights to neighbors, capturing fine-grained dependencies:

$$\alpha_{ij} = \frac{\exp(\mathrm{leakyReLU}(\mathbf{a}^{\mathrm{T}}[\mathbf{W}_e\mathbf{H}_i \| \mathbf{W}_e\mathbf{H}_j]))}{\sum_{k \in \mathcal{N}(i)} \exp(\mathrm{leakyReLU}(\mathbf{a}^{\mathrm{T}}[\mathbf{W}_e\mathbf{H}_i \| \mathbf{W}_e\mathbf{H}_j]))}, \tag{4}$$

where $\mathbf{a}$ is a trainable attention vector, and $\|$ denotes vector concatenation. This mechanism allows the encoder to focus on the most relevant neighbors, which is particularly important for learning the dependencies among IoT devices in the network.

The decoder then reconstructs node features from $\mathbf{Z}$ to $\hat{\mathbf{D}} = \mathrm{GAT}_{\mathrm{dec}}(\mathbf{Z}, \mathbf{A})$. Similar to the encoder, the decoder also applies the attention mechanism. This reconstruction ensures that learned representations preserve structural and attribute relationships.

2) Multi-Granularity Learning for Structural Dependencies: To effectively model structural dependencies among IoT devices, we adopt a multi-granularity learning approach that considers both global and local graph structures. The encoder extracts attribute representations at different levels, while the decoder reconstructs node features to ensure that learned representations preserve both local and global dependencies.

Specifically, the encoder processes both the global graph $\mathcal{G}^{glo}$ and multiple local subgraphs $\mathcal{G}_i^{loc}$. The latent representation of each node is obtained as follows:

$$\mathbf{Z}^{glo} = \text{GAT}_{\text{enc}}(\mathbf{D}^{glo}, \mathbf{A}^{glo}), \quad \mathbf{Z}_i^{loc} = \text{GAT}_{\text{enc}}(\mathbf{D}_i^{loc}, \mathbf{A}_i^{loc}) \tag{5}$$

where the latent embeddings from all local subgraphs are concatenate as $\mathbf{Z}^{loc} = \mathbf{Z}_1^{loc} \oplus \mathbf{Z}_2^{loc} \oplus ... \oplus \mathbf{Z}_n^{loc}$ to integrate local-level representations. This hierarchical encoding strategy allows the model to capture dependencies at different scales, providing a comprehensive representation of IoT device relationships.

The decoder reconstructs node features at both global and local levels to ensure that the encoded representations retain essential structural and attribute relationships:

$$\hat{\mathbf{D}}^{glo} = \text{GAT}_{\text{dec}}(\mathbf{Z}^{glo}, \mathbf{A}^{glo}), \quad \hat{\mathbf{D}}_i^{loc} = \text{GAT}_{\text{dec}}(\mathbf{Z}_i^{loc}, \mathbf{A}_i^{loc}) \tag{6}$$

Similarly, the reconstructed features of all local subgraphs are concatenated as $\hat{\mathbf{D}}^{loc} = \hat{\mathbf{D}}_1^{loc} \oplus \hat{\mathbf{D}}_2^{loc} \oplus ... \oplus \hat{\mathbf{D}}_n^{loc}$. This ensures that the learned representations effectively capture the hierarchical structure of IoT networks.

By integrating global and local structural dependencies, the multi-granularity learning framework enhances the ability to detect anomalies. The learned representations provide a more fine-grained understanding of normal behavior patterns, making it easier to identify deviations that indicate potential anomalies.

3) Anomaly Detection via Reconstruction Errors: To assess node abnormality, we compare the reconstructed features $\hat{\mathbf{D}}$ with the original features $\mathbf{D}$ and the reconstruction error is defined serving as an indicator for measuring node anomalousness:

$$\mathcal{L}_{\text{recon}} = \frac{1}{N \cdot T} \sum_{i=1}^{N} \sum_{t=1}^{T} (\hat{\mathbf{D}}_i(t) - \mathbf{D}_i(t))^2, \tag{7}$$

where $\hat{\mathbf{D}}_i(t)$ and $\mathbf{D}_i(t)$ are the reconstructed and original features of node sn_i at time t, respectively. A higher reconstruction error indicates a greater likelihood of anomalies, as abnormal patterns deviate significantly from learned normal dependencies.

B. Structural Dependency Learning via Contrastive Analysis

To detect structural anomalies, we propose a multi-granularity contrastive learning framework that captures topological dependencies at different scales. Unlike attribute-based reconstruction, this module learns structural normal patterns by comparing the relationships between global and local subgraphs. It consists of three key components: (i) a GAT-based encoder to extract node-level structural representations, is introduced in Section V-A, (ii) an attention-based readout module to aggregate graph-level representations, and (iii) a contrastive learning module to distinguish normal and anomalous structural patterns.

1) Attention-based Readout Module: To capture graph dependencies across different granularities, we aggregate node-level features into graph-level representations.

As introduced in Section V-A, the GAT-based encoder extracts node embeddings from both global and local subgraphs as $\mathbf{Z}^{glo}$ and $\mathbf{Z}^{loc} = \mathbf{Z}_1^{loc} \oplus \mathbf{Z}_2^{loc} \oplus ... \oplus \mathbf{Z}_n^{loc}$. An attention-based readout function is applied to obtain the final graph-level features:

$$\mathbf{h}_{\mathcal{G}_i^{loc}} = \text{READOUT}(\mathbf{h}_{sn_{ij}} | sn_{ij} \in \mathbf{V_i}), \tag{8a}$$

$$\mathbf{h}_{\mathcal{G}^{glo}} = \text{READOUT}(\mathbf{h}_{sn_i} | sn_i \in \mathbf{V}), \tag{8b}$$

where $\mathbf{h}_{sn_i}$ and $\mathbf{h}_{sn_{ij}}$ are the node-level features, and READOUT$(\cdot)$ aggregates them into a single graph representation. This produces graph embeddings for both local subgraphs and the global graph, which serve as inputs for contrastive learning.

2) Contrastive Learning for Structural Dependency: The contrastive learning module is designed to capture the structural dependencies between subgraphs and the global graph, modeling normal structural relationships. It optimizes representations by maximizing similarity within the same structure while minimizing similarity across different structures.

Positive Pairs: Each subgraph $\mathcal{G}_i^{loc}$ is naturally part of the global graph $\mathcal{G}^{glo}$. To enforce their structural consistency, we define positive pairs as:

$$\mathcal{P} = (\mathbf{h}_{\mathcal{G}_i^{loc}}, \mathbf{h}^{+}_{\mathcal{G}^{glo}}), \tag{9}$$

where $\mathbf{h}_{\mathcal{G}_i^{loc}}$ and $\mathbf{h}_{\mathcal{G}^{glo}}$ are the graph embeddings of the subgraph $\mathcal{G}_i^{loc}$ and global graph, respectively.

Negative Pairs: Each subgraph has its unique geographical and functional role. To ensure they remain distinct, we define negative pairs as:

$$\mathcal{N} = (\mathbf{h}_{\mathcal{G}_i^{loc}}, \mathbf{h}^{-}_{\mathcal{G}_j^{loc}}), \tag{10}$$

where $\mathbf{h}_{\mathcal{G}_j^{loc}}(j \neq i)$ is the graph embeddings of the subgraph $\mathcal{G}_j^{loc}$.

The cosine similarity is employed to measure the agreement between graph-level features, providing a stable and interpretable metric. The similarity between a subgraph $\mathcal{G}_i^{loc}$ and the global graph $\mathcal{G}^{glo}$ is calculated as:

$$sim = \frac{{\mathbf{h}_{\mathcal{G}_i^{loc}}}^{T} \mathbf{h}_{\mathcal{G}^{glo}}}{\|\mathbf{h}_{\mathcal{G}_i^{loc}}\| \|\mathbf{h}_{\mathcal{G}^{glo}}\|}, \tag{11}$$

The similarity score sim_{+} for positive pairs should be maximized to capture the inherent structural consistency, while the similarity score sim_{-} for negative pairs should be minimized to ensure structural distinction between subgraphs.

3) Anomaly Detection via Structural Deviations: To detect structural anomalies, we leverage contrastive learning to establish normal structural dependencies between subgraphs and the global graph. The contrastive loss $\mathcal{L}_{\text{contrast}}$ is defined as follows:

$$\mathcal{L}_{\text{contrast}} = -\log \frac{\exp(sim(\mathbf{h}_{\mathcal{G}_i^{loc}}, \mathbf{h}_{\mathcal{G}^{glo}})/\tau)}{\sum_{\mathbf{h}_{\mathcal{G}_j^{loc}} \in \mathcal{N}} \exp(sim(\mathbf{h}_{\mathcal{G}_i^{loc}}, \mathbf{h}_{\mathcal{G}_j^{loc}})/\tau)}, \tag{12}$$

where τ is the temperature parameter that controls the sharpness of the similarity distribution. The denominator aggregates similarity scores over all negative pairs.

Given the learned normal subgraph-global dependencies, we detect anomalies by measuring structural deviations: (i) A low similarity sim_+ between subgraph and global graph suggests an unexpected structural inconsistency, indicating that a subgraph deviates significantly from its expected global strcuture. (ii) A high similarity sim_- between different subgraphs suggests an unusual structural overlap, indicating that two subgraphs, which should be distinct, exhibit an abnormally similar pattern. These deviations indicate potential anomalies, making contrastive learning an effective strategy for structural anomaly detection in IoT networks.

VI. Anomaly Scoring

To effectively detect anomalies in IoT networks, we develop a multi-granularity anomaly scoring mechanism based on both attribute reconstruction errors and structural contrastive deviations. For this, we combine fine-grained (subgraph-level) and global (graph-level) information to measure deviations from normal patterns. This section introduces the joint optimization of reconstruction and contrastive learning modules, followed by the anomaly scoring function that integrates multiple perspectives for robust anomaly detection.

A. Joint Optimization

To capture both attribute dependencies and structural consistency, we jointly optimize two loss components: the reconstruction loss $\mathcal{L}_{\text{recon}}$ for accurate attribute reconstruction and the contrastive loss $\mathcal{L}_{\text{contrast}}$ for preserving structural relationships by maximizing agreement in positive pairs and minimizing it in negative pairs. The joint objective function is formulated as:

$$\mathcal{L}_{\text{joint}} = \alpha \mathcal{L}_{\text{recon}} + \beta \mathcal{L}_{\text{contrast}}, \tag{13}$$

where α and β are parameters that control the relative contributions of the two losses. This flexible weighting allows the model to balance attribute and structural learning based on the specific anomaly characteristics of the dataset.

B. Anomaly Scoring Mechanism

Anomalies are identified by measuring deviations from learned normal patterns. The anomaly score for each node is defined as a weighted sum of the reconstruction error and the contrastive error. Specifically, reconstruction error f_i^{recon} is computed to measure attribute deviations by calculating the Mean Squared Error (MSE) between the original and reconstructed values:

$$f_i^{\text{recon}} = \frac{1}{T}\sum_{t=1}^{T}(\hat{\mathbf{X}}_i(t) - \mathbf{X}_i(t))^2 \tag{14}$$

Contrastive error f_i^{contrast} measures structural deviations by computing one minus the cosine similarity between the subgraph embedding $\mathbf{h}_{\mathcal{G}_i^{loc}}$ and the global graph embedding $\mathbf{h}_{\mathcal{G}^{glo}}$:

$$f_i^{\text{contrast}} = 1 - sim(\mathbf{h}_{\mathcal{G}_i^{loc}}, \mathbf{h}_{\mathcal{G}^{glo}}) \tag{15}$$

The final anomaly score is computed as:

$$f_i = \alpha' f_i^{\text{recon}} + \beta' f_i^{\text{contrast}}, \tag{16}$$

where α' and β' control the contributions of attribute and structural deviations, respectively.

A threshold μ is set based on normal data statistics, and a node is flagged as anomalous if $f_i > \mu$. This approach ensures that anomalies, which deviate significantly from both normal attribute distributions and structural patterns, are effectively affected.

The proposed framework integrates multi-granularity learning to enhance anomaly detection in IoT networks. It combines (i) Attribute-level learning for accurate feature reconstruction, (ii) Structure-level learning for preserving subgraph-global relationships, and (iii) joint anomaly scoring to detect both local attribute anomalies and global structural deviations. This hybrid approach leverages attribute and topological perspectives, effectively capturing complex IoT anomalies that cannot be captured by traditional methods.

VII. Experimental Evaluation

This section presents the experimental evaluation conducted to assess the performance of MGG-AD. We first introduce the datasets and evaluation metrics used in our study. Next, we describe the baseline methods and the experimental setup. Finally, we evaluate the anomaly detection performance, conduct an ablation study to assess the contribution of different components, and perform a sensitivity analysis to determine the optimal parameter setting for anomaly detection.

A. Datasets and Evaluation Metrics

1) Datasets: We evaluate our approach on two publicly available datasets: ASD [7] (Application Server Dataset) and SMD [11] (Server Machine Dataset). Both datasets consist of multivariate time series (MTS) collected from industrial server systems, with each entity characterized by multiple system metrics. ASD contains stable servers with labeled anomalies categorized into temporal, inter-metric, and inter-metric-temporal types. SMD includes multiple machines, from which we select stable ones to avoid concept drift.

The dataset statistics, including time span, system metrics name, and anomaly proportions, are summarized in Table I. Notice, how both data sets contain a similar rate of anomalies, precisely 4.61% for ASD, and 5.84% for SMD. The datasets combine a mix of structural anomalies and attribute anomalies for the system metrics, as introduced in Section II.

2) Evaluation Metrics: We evaluate detection performance using standard precision (*Pre*), recall (*Rec*), and F1-score (*F1*), computed as: *F1* $= (2 \times$ *Pre* $\times$ *Rec*$)/($*Pre* $+$ *Rec*$)$ where *Precision = TP/(TP + FP)* and *Recall = TP/(TP + FN)*, with *TP*, *FP*, and *FN* denoting the number of true positives, false positives, and false negatives, respectively.

TABLE I
SUMMARY OF KEY STATISTICS FOR THE TWO PUBLIC DATASETS, ASD AND SMD, INCLUDING THE NUMBER OF ENTITIES AND METRICS, TIME COVERAGE, SYSTEM METRIC NAMES, TRAINING AND TESTING SIZES, AND ANOMALY RATIO.

Datasets	#Entities	#Metrics	Time Coverage	System Metrics Name	Train	Test	Anomalies (%)
ASD [7]	12	19	45 Days	CPU-Related Metrics, Memory Usage	102,331	51,840	4.61
SMD [11]	12	38	5 Weeks	CPU Load, Network Usage	304,168	304,174	5.84

B. Baselines and Experimental Setup

We compare MGG-AD with five state-of-the-art baselines: OmniAnomaly [11] uses a stochastic recurrent model for reconstruction-based detection; InterFusion [7] adopts hierarchical VAEs to capture temporal and inter-metric patterns; GDN [14] introduces a GAT network to model inter-sensor dependencies for anomaly prediction; MTAD-GAT [15] learns temporal and feature dependencies through parallel graph attention layers, combining forecasting and reconstruction objectives; and FuSAGNet [8] combines sparse autoencoders and GNNs to learn latent representations and graph structures for forecasting-driven anomaly detection.

In our method, we set $k = 10$ and $k' = 30$ for ASD, and $k = 15$ and $k' = 40$ for SMD. We set $(\alpha, \beta) = (0.7, 1.3)$ for ASD and $(1.2, 0.8)$ for SMD. The anomaly scoring weights are set as $\alpha' = 0.6$ and $\beta' = 0.4$. The number of training epochs is set to 300 for ASD and 150 for SMD. The discrimination modules were trained using the Adam optimizer with a learning rate of 0.001. Next, in Section VII-C1, we compare the performance of our approach with state-of-the-art models. Afterward, in Section VII-C2, we gain deeper insights into MGG-AD by analyzing each component's effectiveness.

TABLE II
ANOMALY DETECTION PERFORMANCE (PRECISION(%), RECALL(%), AND F1-SCORE(%)) OF DIFFERENT METHODS ON ASD AND SMD DATASETS WITH LABELED GROUND-TRUTH ANOMALIES.

Method	ASD			SMD		
	Pre	*Rec*	*F1*	*Pre*	*Rec*	*F1*
OmniAnomaly	84.72	75.16	79.63	72.25	59.81	65.51
InterFusion	80.35	67.77	73.51	82.69	74.28	78.26
GDN	89.67	85.17	87.36	78.09	69.94	73.78
MTAD-GAT	85.13	77.43	81.09	89.23	82.61	85.80
FuSAGNet	90.08	88.21	89.13	85.37	84.14	84.75
MGG-AD	93.41	90.86	92.11	91.27	87.35	89.26

C. Performance Evaluation

1) Overall Performance: TABLE II presents the performance evaluation results of MGG-AD and the baseline models on two public datasets. Overall, MGG-AD achieves the highest F1-score across both datasets, demonstrating a strong balance between precision and recall. Notably, while FuSAGNet achieves a comparable recall on the SMD dataset, MGG-AD maintains superior consistency across all three metrics. On average, MGG-AD achieves an F1-score of 0.92 on ASD and 0.89 on SMD, surpassing all baselines and demonstrating its effectiveness in multivariate time series anomaly detection.

Compared to classical time series-based anomaly detection methods such as OmniAnomaly and InterFusion, these methods primarily rely on temporal modeling but struggle to capture multivariate dependencies effectively. While they are sensitive to long-term temporal variations, they lack structural modeling capabilities, making them prone to missing complex inter-metric anomalies, as reflected in their relatively low recall. Specifically, OmniAnomaly, which employs a VAE-based generative approach, does not explicitly encode topological dependencies, leading to a higher false-negative rate when anomalies arise from metric correlations rather than individual values. Similarly, InterFusion introduces attention mechanisms to improve upon OmniAnomaly but still lacks a structural representations of spatial-temporal dependencies. Consequently, it fails to capture complex cross-metric anomalies, resulting in suboptimal performance.

To address these limitations, MGG-AD leverages a multi-granularity graph structure to model both local and global relationships. This design enhances the detection of cross-metric anomalies, leading to a 15.5% increase in recall on ASD and a 15.7% improvement in F1-score on SMD compared to these purely temporal modeling methods.

In contrast to time series-based methods, GNN-based approaches, such as GDN and MTAD-GAT, explicitly model metric dependencies, enhancing the detection of inter-metric anomalies and demonstrating stronger performance in capturing multivariate relationships. However, these methods still face specific limitations. GDN focuses on learning spatial dependencies but does not explicitly model temporal dependencies, limiting its recall when detecting anomalies that evolve over time. MTAD-GAT enhances robustness by integrating graph attention mechanisms with dual-stream prediction and reconstruction but lacks a clear distinction between global and local dependencies, compromising its ability to capture hierarchical relationships in multivariate data.

To further refine spatial-temporal dependency modeling, MGG-AD integrates hierarchical graph representations, distinguishing between localized and global patterns. By explicitly capturing both levels of dependencies, MGG-AD achieves a 5.7% increase in recall on ASD compared to GDN and a 2.0% improvement in precision on SMD compared to MTAD-GAT, demonstrating the effectiveness of its structured anomaly detection approach.

Beyond these methods, FuSAGNet, which integrates spatial-temporal attention mechanisms, achieves a high recall in detecting anomalies. However, it tends to overfit certain abnormal patterns, leading to a slightly lower precision due to an increased false-positive rate. In contrast, MGG-AD introduces contrastive learning to enhance the ability of model to distinguish between normal and abnormal patterns. This design improves anomaly discrimination and mitigates overfitting to specific abnormal patterns, a limitation observed in FuSAGNet. As a result, MGG-AD achieves a 3.0% improvement in F1-score on ASD and a 5.9% precision increase on SMD, demonstrating superior robustness.

To summarize, MGG-AD consistently outperforms all baseline methods by a significant margin. By capturing both fine-grained and coarse-grained dependencies, it constructs a structured representation of multivariate relationships, addressing the hierarchical modeling limitations of methods such as OmniAnomaly and InterFusion. Furthermore, MGG-AD simultaneously detects both attribute-level and structure-level anomalies within its multi-granularity graph structure, overcoming the constraints of MTAD-GAT and FuSAGNet, which focus solely on attribute-level deviations. By integrating these advances, MGG-AD provides a more reliable and accurate solution for anomaly detection in IoT environments, ensuring robust and interpretable performance across diverse datasets.

TABLE III
ABLATION STUDY: IMPACT OF DIFFERENR COMPOENTS OF MGG-AD ON ANOMALY DETECTION PERFORMANCE (PRECISION(%), RECALL(%), AND F1-SCORE(%)) ON ASD AND SMD DATASETS.

Method	ASD			SMD		
	Pre	*Rec*	*F1*	*Pre*	*Rec*	*F1*
MGG-AD	93.41	90.86	92.11	91.27	87.35	89.26
w/o $\mathcal{G}^{loc}$	92.40	84.77	88.42	89.11	80.34	84.49
w/o $\mathcal{G}^{glo}$	89.42	86.89	88.13	86.12	87.98	87.04
w/o Contrast	90.57	82.86	86.53	88.17	79.36	83.53

2) Ablation Study: This section examines the impact of key components in MGG-AD, in particular, the Local Subgraphs $\mathcal{G}^{loc}$, the Global Graph $\mathcal{G}^{glo}$, and Contrastive Learning between local and global structures. To evaluate the contributions of each component, we systematically remove one of the component and assess its effect on the performance. The results, summarized in TABLE III, highlight the necessity of each element in preserving high detection accuracy.

Impact of Local Subgraphs $\mathcal{G}^{loc}$. To examine the role of local subgraphs, we exclude $\mathcal{G}^{loc}$ and rely solely on the global graph. As shown in TABLE III, the absence of local subgraphs leads to a decreasing in recall of 6.64% (90.86 → 84.77) in ASD and 7.01% (87.35 → 80.34) in SMD. This performance drop highlights the inability of the global graph alone to capture fine-grained dependencies, leading to missed certain anomalies. This observation underscores the necessary of $\mathcal{G}^{loc}$ in detecting subtle abnormal behaviours that may be overlooked when relying only on global correlations.

Impact of the Global Graph $\mathcal{G}^{glo}$. We further investigate the impact of removing the global graph $\mathcal{G}^{glo}$, where the model learns dependencies exclusively from local subgraphs. Without global structural information, the precision decreases by 3.99% (93.41 → 89.42) on ASD and 5.15% (91.27 → 86.12) on SMD. This decline indicates that the absence of $\mathcal{G}^{glo}$ limits the ability of model to detect anomalies that manifest at a global scale, demonstrating its importance in maintaining overall detection accuracy.

Impact of Contrastive Learning. To assess the role of contrastive learning, we remove this mechanism, leaving only attribute-level reconstruction for anomaly detection. The exclusion of contrastive learning leads to a substantial drop in F1-score by 5.58% (92.11 → 86.53) on ASD and 5.73% (89.26 → 83.53) on SMD. This decline indicates that contrastive learning is essential for capturing structural dependencies across different granularities, enhancing the ability of model to distinguish normal from abnormal patterns.

In summary, the ablation study confirms that each component of MGG-AD plays a crucial role in anomaly detection. By capturing both fine-grained and coarse-grained dependencies, their integration enhances detection accuracy, as evidenced by the experiments results.

3) Sensitivity Analysis: This section investigates how the choice of k (local subgraph neighbors) and k' (global graph neighbors) affects detection accuracy across two datasets. By varying one parameter while keeping the other fixed, we evaluate their impact on capturing local and global dependencies.

The number of neighbors per node determines graph sparsity, which in turn affects anomaly detection. As shown in Fig. 3 and Fig. 4, performance remains relatively stable across different (k, k') settings, without abrupt degradation or significant fluctuations. This indicates that the model effectively preserves structural information within a reasonable range of sparsity, allowing robust detection of anomalies.

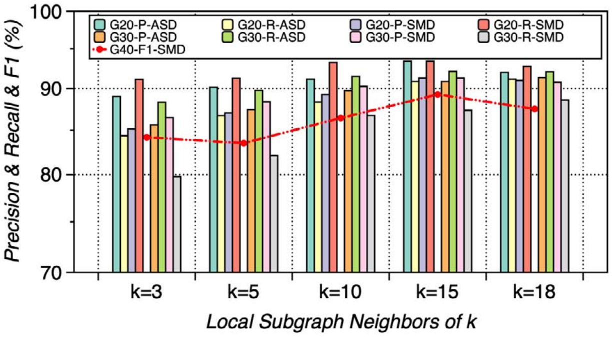

Fig. 3. Impact of varying local connection degree k on precision and recall. Labels such as "L5-P-ASD" indicate precision at $k = 5$ on ASD, while "L15-R-SMD" represents recall at $k = 15$ on SMD.

Effect of Varying k. To examine the impact of k, we fix $k' = 20, 30$ and vary k among $3, 5, 10, 15, 18$. As illustrated in Fig. 3, a smaller k results in sparser local subgraphs, leading to weaker inter-attribute dependencies and reduced effectiveness in detecting attribute-level anomalies. For example, in the ASD dataset with $k' = 30$, increasing k from 5 to 10 improves recall from 88.32 to 90.86, highlighting the benefit of enhanced local connectivity. However, further increasing k to 15 or beyond yields diminishing returns, as excessive local connections may introduce redundant information, which may obscure fine-grained anomaly patterns. A similar trend is observed in the SMD dataset: with $k' = 40$, increasing k from 10 to 15 improves the F1-score from 88.43 to 89.26, demonstrating the advantage of capturing more local dependencies. However, beyond this threshold, detection accuracy stabilizes and is not further improved by additional local connections.

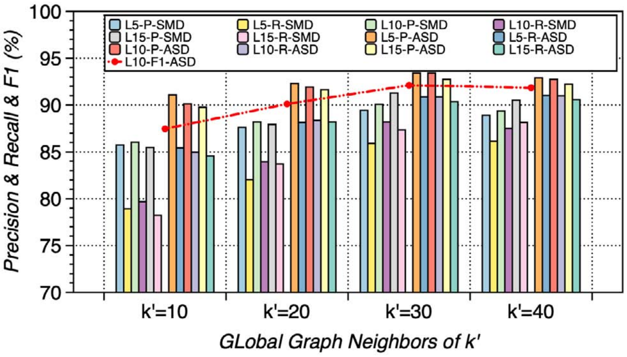

Fig. 4. Impact of varying global connection degree k' on precision and recall. Labels such as "G20-P-ASD" indicate precision at $k' = 20$ on ASD, while "G30-R-SMD" represents recall at $k' = 30$ on SMD.

Effect of Varying k'. To analyze the effect of k', we fix $k = 5, 10, 15$ and vary k' among $10, 20, 30, 40$. As depicted in Fig. 4, when $k' = 10$, the global graph is relatively sparse, resulting in weaker structural dependencies and reduced capability to detect topological anomalies. Conversely, as k' increases, excessive global connections introduce redundant information, leading to a slight precision decline rather than consistent improvement. Moreover, the interplay between k and k' reveals that an imbalance between local and global structures can hinder detection performance. For instance, when $k = 3$ and $k' = 30$, the dominance of the global graph may obscure localized anomaly patterns, weakening the ability to capture fine-grained dependencies. Additionally, an excessively large global adjacency matrix increases computational overhead without significant performance gains, underscoring the importance of maintaining a balanced graph structure.

TABLE IV
ANALYSIS OF THE PERFORMANCE IMPACT FROM VARYING THE RATIO OF LOCAL (k) AND GLOBAL (k') CONNECTIONS. THE RESULT SUGGEST AN OPTIMAL TRADE-OFF LEADS TO IMPROVED DETECTION ACCURACY.

Dataset	k	k'	Best F1-score	Optimal k'/k Ratio
ASD	10	30	92.11	3.00
SMD	15	40	89.26	2.67

As summarized in Table IV, the results suggest an optimal balance between local and global connections. The best performance is achieved when the number of global neighbors is approximately 2.5 to 3 times that of local neighbors, allowing the model to capture fine-grained attribute dependencies while maintaining essential structural information. Specifically, the highest F1-score are observed at (k, k') pairs of $(10, 30)$ for ASD (92.11) and $(15, 40)$ for SMD (89.26), highlighting the importance of this ratio for optimal anomaly detection.

VIII. Related Work and Comparison

Multivariate time series anomaly detection has been extensively developed to capture complex dependencies among sensor attributes in IoT systems. Existing approaches can be broadly categorized into attribute-level dependency modeling and generative reconstruction-based methods. However, these approaches often fail to simultaneously capture both attribute-level and structural-level anomalies, leading to suboptimal performance in complex IoT environments.

A. *Attribute Dependency Modeling*

Recent studies have leveraged GNNs to model dependencies among sensor attributes by constructing graph structures from multivariate time series data. For example, VGCRN [16] employs a variational graph convolutional recurrent network to capture spatial and temporal dependencies. Anomalous behaviours are identified by measuring the deviations between forecasted and observed values. Similarly, a GAN-based model [17] have been developed to learn attribute correlations within an adaptive graph structure. DCdetector [13] focuses on time series anomaly detection via contrastive representation learning, primarily targeting attribute-level anomalies and lacks structural modeling. Structured modeling is crucial for capturing these relationships between microservice variables [18]. However, the respective work focuses on real-time optimization of microservice pipelines, rather than anomaly detection. While these methods effectively capture attribute-level dependencies, they often treat attributes as graph nodes without considering their spatial context [19], limiting their ability to model multi-granularity relationships.

B. *Generative Reconstruction-Based Methods*

An alternative approach focuses on learning the normal behavior of time series via generative models and reconstructing expected patterns to detect anomalies. For instance, PW-GAN-GP [20] utilizes a predictive Wasserstein generative adversarial network to estimate the high-dimensional distribution of time series data, where anomalies are detected based on prediction errors from an LSTM-based generator. Similarly, TSMAE [21] employs a memory-augmented autoencoder to reconstruct latent representations of multivariate time series and identify anomalies via reconstruction loss. However, these methods focus primarily on attribute dependencies and lack an explicit structural representation of the IoT system, leading to potential blind spots in detecting topological anomalies.

Structural modeling has been explored in performance disgnosis, such as CauseInfer [22], which leverages a hierarchical causality graph to infer system bottlenecks. While effective analyzing cloud environments, these approaches are not directly designed for anomaly detection in IoT settings, where both attribute and topological deviations must be considered.

C. Comparison with our Approach

Despite the progress made by existing methods, most approaches either (i) focus solely on attribute-level dependencies or (ii) rely on time-series reconstruction without explicitly modeling structural relationships. Moreover, these methods neglect the multi-granularity nature of dependencies in IoT systems, where sensory attributes exhibit both local and global correlations. To address these limitations, we propose a multi-granularity graph-based framework that simultaneously captures attribute and structural dependencies. By integrating generative learning (reconstruction-based) and contrastive learning (structure-based) methods within a hierarchical graph representation, our approach effectively models both fine-grained local dependencies and global topological structures. This allows more accurate anomaly detection by leveraging deviations in both attribute values and structural patterns.

IX. Conclusion

This paper presents a Multi-Granularity Graph Anomaly Detection (MGG-AD) framework to enhance anomaly detection in IoT systems by modeling both attribute-level and structure-level dependencies. Instead of treating all attributes as a unified entity, our multi-granularity dependency graph captures both local and global correlations among monitoring attributes for web services. By integrating a generative-discriminative mechanism for attribute anomaly detection and contrastive learning for structural anomaly detection, our method effectively identifies abnormal patterns across different representation levels. Extensive experiments on publicly available datasets demonstrate that MGG-AD achieves higher accuracy than state-of-the-art techniques for anomaly detection. In future work, we plan to extend our validation to broader ranger of IoT scenarios beyond current public datasets and conduct detailed case studies in real-world applications to enhance detection interpretability. Furthermore, to support latency-sensitive and resource-constrained deployments, we will explore lightweight model optimization techniques and conduct systematic runtime and energy overhead analysis.

Acknowledgments

This work was supported in part by the National Natural Science Foundation of China (Grants No. 62372420, 62402460), the National Science and Technology Major Project (Grant No. 2024ZD1001900), the China Geological Survey (CGS) work project: “Geoscience literature knowledge services and decision supporting” (Grant No. DD20230139), the Fundamental Research Funds for the Central Universities (Grants No. 2652023001, 2652023063), and the China Scholarship Council Program (Grant No. 202406400050).

References

[1] Y. Li, Z. Zhou, S. Deng, X. Sun, X. Xue, S. Yangui, and W. Gaaloul, “Accurate Anomaly Detection Leveraging Knowledge-Enhanced GAT,” in *IEEE International Conference on Web Services*, pp. 568–577, 2024.

[2] Y. Wu, H.-N. Dai, and H. Tang, “Graph Neural Networks for Anomaly Detection in Industrial Internet of Things,” *IEEE Internet of Things Journal*, vol. 9, no. 12, pp. 9214–9231, 2021.

[3] X. Xue, Y. Guo, S. Chen, and S. Wang, “Analysis and Controlling of Manufacturing Service Ecosystem: A Research Framework based on the Parallel System Theory,” *IEEE Transactions on Services Computing*, vol. 14, no. 6, pp. 1598–1611, 2021.

[4] T. Huang, et al, “A Semi-Supervised VAE based Active Anomaly Detection Framework in Multivariate Time Series for Online Systems,” in *Proceedings of the ACM Web Conference*, pp. 1797-1806, 2022.

[5] Y. Li, Z. Zhou, et al, “Accurate Anomaly Detection with Energy Efficiency in IoT-Edge-Cloud Collaborative Networks,” *IEEE Internet of Things Journal*, vol. 10, no. 19, pp. 16959–16974, 2023.

[6] S. Li, S. Li, M. Xie, K. Gong, J. Zhao, C. H. Liu, and G. Wang, “End-to-End Transferable Anomaly Detection via Multi-Spectral Cross-Domain Representation Alignment,” *IEEE Transactions on Knowledge and Data Engineering*, vol. 35, no. 12, pp. 12194–12207, 2021.

[7] Z. Li, Y. Zhao, J. Han, Y. Su, R. Jiao, X. Wen, and D. Pei, “Multivariate Time Series Anomaly Detection and Interpretation Using Hierarchical Inter-Metric and Temporal Embedding,” in *Proceedings of the 27th ACM SIGKDD Conference on Knowledge Discovery & Data Mining*, pp. 3220–3230, 2021.

[8] S. Han and S. S. Woo, “Learning Sparse Latent Graph Representations for Anomaly Detection in Multivariate Time Series,” in *Proceedings of the 28th ACM SIGKDD Conference on Knowledge Discovery and Data Mining*, pp. 2977–2986, 2022.

[9] S. Kim, K. Choi, H.-S. Choi, B. Lee, and S. Yoon, “Towards A Rigorous Evaluation of Time-Series Anomaly Detection,” in *Proceedings of the AAAI Conference on Artificial Intelligence*, pp. 7194–7201, 2022.

[10] T. Wilson, et al, “DeepGPD: A Deep Learning Approach for Modeling Geospatio-Temporal Extreme Events,” in *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 36, no. 4, pp. 4245–4253, 2022.

[11] Y. Su, et al, “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*, pp. 2828-2837, 2019.

[12] Y. Zheng, M. Jin, Y. Liu, L. Chi, K. T. Phan, and Y.-P. P. Chen, “Generative and Contrastive Self-Supervised Learning for Graph Anomaly Detection,” *IEEE Transactions on Knowledge and Data Engineering*, vol. 35, no. 12, pp. 12220–12233, 2021.

[13] Y. Yang, C. Zhang, T. Zhou, Q. Wen, and L. Sun, “DCdetector: Dual Attention Contrastive Representation Learning for Time Series Anomaly Detection,” in *Proceedings of the 29th ACM SIGKDD Conference on Knowledge Discovery and Data Mining*, pp. 3033–3045, 2023.

[14] 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, pp. 4027–4035, 2021.

[15] H. Zhao, et al, “Multivariate Time-Series Anomaly Detection via Graph Attention Network,” in *2020 IEEE International Conference on Data Mining (ICDM)*, pp. 841–850, 2020.

[16] W. Chen, L. Tian, B. Chen, L. Dai, Z. Duan, and M. Zhou, “Deep Variational Graph Convolutional Recurrent Network for Multivariate Time Series Anomaly Detection,” in *Proceedings of the 39th International Conference on Machine Learning*, vol. 162, pp. 3621–3633, 2022.

[17] H. Kang and P. Kang, “Transformer-based Multivariate Time Series Anomaly Detection Using Inter-Variable Attention Mechanism,” *Knowledge-Based Systems*, vol. 290, pp. 111507, 2024.

[18] B. Sedlak, V. C. Pujol, P. K. Donta, and S. Dustdar, “Markov Blanket Composition of SLOs,” in *IEEE International Conference on Edge Computing and Communications (EDGE)*, pp. 128-138, 2024.

[19] C. Lin, B. Du, L. Sun, and L. Li, “Hierarchical Context Representation and Self-Adaptive Thresholding for Multivariate Anomaly Detection,” *IEEE Transactions on Knowledge and Data Engineering*, vol. 36, no. 7, pp. 3139–3150, 2024.

[20] S. Qi, J. Chen, P. Chen, P. Wen, X. Niu, and L. Xu, “An Efficient GAN-based Predictive Framework for Multivariate Time Series Anomaly Prediction in Cloud Data Centers,” *The Journal of Supercomputing*, vol. 80, pp. 1268–1293, 2024.

[21] H. Gao, et al, “TSMAE: A Novel Anomaly Detection Approach for Internet of Things Time Series Data Using Memory-Augmented Autoencoder,” *IEEE Transactions on Network Science and Engineering*, vol. 10, no. 5, pp. 2978–2990, 2023.

[22] P. Chen, Y. Qi, and D. Hou, “CauseInfer: Automated End-to-End Performance Diagnosis with Hierarchical Causality Graph in Cloud Environment,” *IEEE Transactions on Services Computing*, vol. 12, no. 2, pp. 214–230, 2019.