AEGIS-GNN: Verifiable Multi-Relational Graph Learning for Secure LEO Satellite Operations

Edson Eliezer da Silva Tavares[1], Qi Xia[1], Hu Xia[1], Jianbin Gao[1,*], Bonsu Adjei-Arthur[1]
[1]School of Computer Science and Engineering, University of Electronic Science and Technology of China, Chengdu, China

***Abstract*—Determining the ground truth from conflicting satellite reports is a critical challenge in Low Earth Orbit (LEO) mega-constellations. Existing Graph Neural Network (GNN) approaches model satellite networks as homogeneous graphs, discarding the rich multi-modal relationships that govern trust. This paper presents AEGIS-GNN, a multi-relational heterogeneous GNN architecture for satellite truth discovery. Our framework makes three contributions: (1) a multi-edge graph construction capturing four complementary relationship types: .communication links, spatial proximity, feature similarity, and orbital plane membership, each encoding a distinct trust-relevant modality; (2) a hybrid residual GNN stack that alternates spectral (GCN) and spatial (GraphSAGE) convolutions with a novel learned residual gating mechanism that adaptively blends new and prior representations per feature dimension; and (3) relation-specific multi-head attention that learns differentiated neighbor weighting across edge types. We further integrate a lightweight blockchain anchoring layer for operational verifiability. Experiments on 1,008 satellite nodes with 557,158 multi-type edges across 5 random seeds demonstrate 92.47 ± 0.41% accuracy, 0.9568 ± 0.005 F1-score, and 93.85 ± 0.52% AUC, significantly outperforming six baselines including HGT (+3.5%), R-GCN (+4.4%), and HAN (+3.8%). Comprehensive ablation studies validate each component, with learned residual gating contributing the largest single gain (+7.7%) and feature similarity edges providing +6.8% accuracy improvement. Sensitivity analyses across key hyperparameters confirm the robustness of the proposed architecture.**

***Index Terms*—Graph Neural Networks, Heterogeneous Graphs, Multi-Relational Learning, Truth Discovery, LEO Satellite Networks**

I. Introduction

Low Earth Orbit (LEO) satellite constellations have grown from experimental deployments to operational mega-constellations comprising thousands of satellites. Networks such as Starlink (over 6,000 satellites), OneWeb, and the planned Kuiper constellation provide global broadband connectivity but introduce a fundamental trust challenge: when multiple satellites report conflicting observations about network conditions, traffic patterns, or security events, how should the system determine which reports are truthful? A single compromised satellite injecting falsified telemetry can cascade into widespread routing failures, making principled trust evaluation essential for operational safety.

Graph Neural Networks (GNNs) have emerged as natural candidates for this problem, as satellite networks inherently form graphs where nodes represent satellites and edges represent inter-satellite relationships [1]. GNNs leverage message-passing mechanisms to aggregate neighborhood information, enabling each node to refine its representation based on the broader network context [2], [3]. However, applying standard GNN architectures to satellite trust evaluation reveals two fundamental limitations.

First, existing architectures predominantly assume homogeneous graphs, yet a satellite's trustworthiness depends on multiple relationship types simultaneously: communication topology, spatial proximity, behavioral similarity, and orbital mechanics. Collapsing these into a single edge type discards complementary information essential for accurate detection. While heterogeneous GNN methods [4]–[6] can handle multiple edge types, they have not been applied to multi-modal satellite trust relationships where each modality captures fundamentally different physical or behavioral signals. Second, standard residual connections in deep GNNs use fixed additive skip paths [7], [8], treating all feature dimensions uniformly. We hypothesize that in heterogeneous satellite graphs, different feature dimensions encode qualitatively different information (e.g., orbital dynamics vs. behavioral patterns) and should be blended adaptively rather than uniformly.

This paper introduces AEGIS-GNN (Anchored Edge-heterogeneous Graph Intelligence System), a framework that addresses these limitations through the following contributions:

- *Multi-Edge Heterogeneous Graph Construction.* We construct graphs with four distinct edge types: communication, spatial proximity, feature similarity, and orbital plane. Each encodes a different modality of satellite interaction. Our ablation study demonstrates that this multi-modal representation yields +10.4% accuracy over communication-only graphs.
- *Hybrid Residual GNN with Learned Gating.* Our architecture alternates GCN (spectral) and GraphSAGE (spatial) convolutions, introducing *learned residual gating vectors* that perform per-dimension adaptive blending between new convolution outputs and prior representations. This mechanism yields the largest single ablation gain (+7.7%), substantially outperforming fixed skip connections [7] and GCNII-style identity mapping [8].
- *Relation-Specific Multi-Head Attention.* We employ relation-aware attention coefficients that learn differentiated neighbor weighting across the four edge types, enabling the model to discover which relationship modal-

Corresponding author: Jianbin Gao (email: gaojb@uestc.edu.cn).
E-mail: edsonskont@hotmail.com (**Edson E. S. Tavares**), xiaqi@uestc.edu.cn (**Qi Xia**), xiahu@uestc.edu.cn (**Hu Xia**) and peeman34@gmail.com (**B. Adjei-Arthur**)

ities are most informative for each node.

Additionally, we integrate a lightweight blockchain anchoring layer for operational verifiability in multi-operator deployments, though we emphasize that the GNN architecture is the primary contribution.

II. Related Work

A. Graph Neural Networks

GNNs have become the dominant paradigm for learning on graph-structured data. GCN [2] introduced spectral convolutions via polynomial filters on the normalized graph Laplacian. GraphSAGE [3] shifted to spatial aggregation with neighborhood sampling, enabling inductive learning on unseen nodes. GAT [9] introduced attention-based neighbor weighting. GIN [10] established theoretical connections between GNNs and the Weisfeiler-Leman graph isomorphism test. These foundational architectures assume homogeneous graphs with a single edge type.

B. Heterogeneous Graph Neural Networks

Real-world graphs often contain multiple node and edge types. R-GCN [6] extended GCN with relation-specific weight matrices and basis decomposition for parameter efficiency. HAN [4] introduced hierarchical attention over meta-paths. HGT [5] employed type-specific transformers with heterogeneous mutual attention. HetGNN [11] combined content aggregation with neighbor sampling for heterogeneous nodes. SimpleHGN [12] demonstrated that GAT with learnable edge-type embeddings can match more complex architectures. ie-HGCN [13] proposed interpretable edge-type selection. Our work differs in that we construct domain-specific edge types encoding distinct physical and behavioral modalities rather than relying on schema-defined meta-paths.

C. Deep GNNs and Residual Connections

Over-smoothing limits GNN depth, as node representations converge with increasing layers [14]. Several approaches address this: DeepGCN [7] adapted ResNet-style skip connections; GCNII [8] combined initial residual connections with identity mapping; JKNet [15] used jumping knowledge to adaptively select layer representations. Gated Graph Neural Networks [16] employed GRU-style gating for recurrent message passing. Our learned residual gating differs by computing per-dimension sigmoid gates conditioned on the element-wise product of current and prior representations, learning which specific feature dimensions should be updated versus preserved.

D. GNNs for Satellite and Communication Networks

GNNs have been applied to satellite networks for routing optimization [17], resource allocation [18], and anomaly detection in IoT networks [19], [20]. GTXChain [1] combined GNNs with blockchain for IoT security. However, existing approaches use homogeneous graph representations that cannot capture the multi-modal nature of satellite relationships. Our work is the first to construct and evaluate heterogeneous multi-edge satellite graphs for trust discovery.

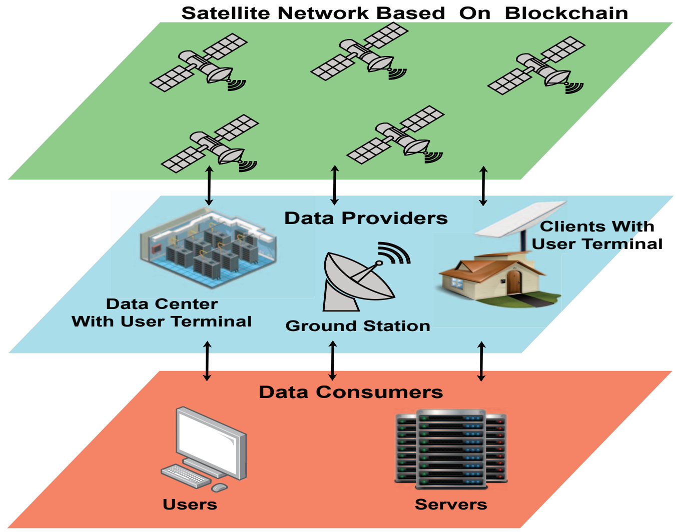

Fig. 1: AEGIS-GNN system architecture overview. LEO satellites, ground stations, and user terminals form a heterogeneous network with four edge types.

E. Truth Discovery

Truth discovery identifies reliable information from conflicting multi-source data. Classical methods include CRH [21] and CATD [22], which iteratively estimate source reliability and truth values. Recent neural approaches [23] learn source embeddings end-to-end. Our framework extends truth discovery to graph-structured satellite networks where spatial, orbital, and behavioral relationships provide complementary evidence. To the best of our knowledge, this work is among the first to apply truth discovery to LEO satellite networks.

III. System Architecture

A. Network Model

As illustrated in Figure 1, the AEGIS-GNN network comprises LEO satellites, terrestrial storage servers, user terminals, and ground stations. We model S satellites indexed by $\mathcal{S} = \{1, 2, \ldots, s\}$ organized into geographic regions $g_1, g_2, \ldots, g_M$. The slant range between a satellite at altitude H and a ground station is:

$$d(\epsilon) = R_E \left[\sqrt{\left(\frac{H + R_E}{R_E} \right)^2 - \cos^2 \epsilon} - \sin \epsilon \right] \quad (1)$$

where ϵ is the elevation angle and $R_E = 6371\,\text{km}$ is Earth's radius.

B. Multi-Edge Heterogeneous Graph Construction

A key insight is that satellite relationships are *multi-modal*: the same pair of satellites may interact through communication links, share spatial proximity, exhibit similar behavior, and belong to the same orbital plane. We formalize this through a heterogeneous graph $\mathcal{G} = (\mathcal{V}, \mathcal{E}, \mathcal{R})$, where $\mathcal{V} = \mathcal{V}_s \cup \mathcal{V}_g$ is the node set comprising satellite nodes $\mathcal{V}_s$ and ground station nodes $\mathcal{V}_g$, $\mathcal{E}$ is the edge set defined below, and

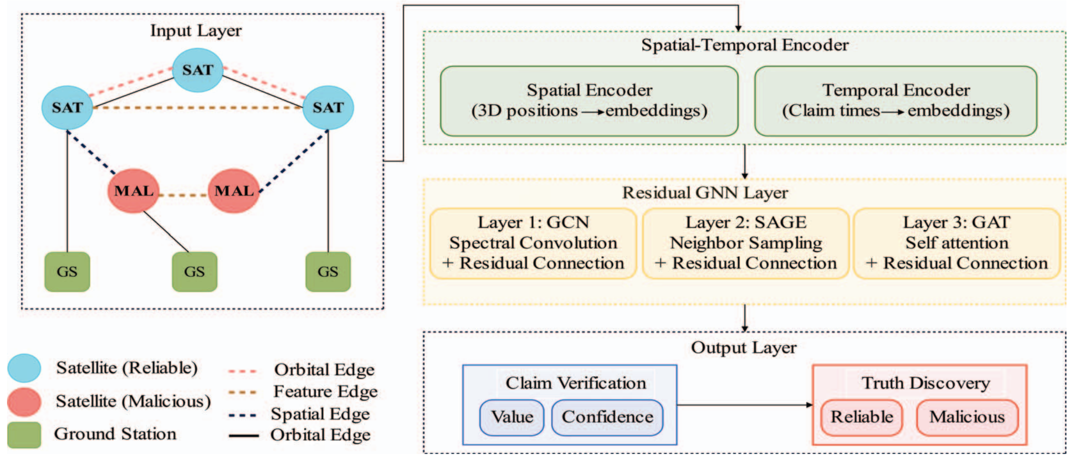

Fig. 2: AEGIS-GNN architecture: spatio-temporal encoding feeds into a hybrid residual GNN stack with learned gating, producing dual-head outputs for truth discovery.

$\mathcal{R} = \{comm, spatial, feature, orbital\}$ is the set of relation types, and:

$$\mathcal{E} = \mathcal{E}_{comm} \cup \mathcal{E}_{spatial} \cup \mathcal{E}_{feature} \cup \mathcal{E}_{orbital} \tag{2}$$

Communication Edges $\mathcal{E}_{comm}$: Direct inter-satellite links (ISLs) based on line-of-sight visibility ($|\mathcal{E}_{comm}| = 50$). These represent actual data exchange pathways and are the sparsest edge type.

Spatial Proximity Edges $\mathcal{E}_{spatial}$: Connections based on k-nearest neighbors in 3D orbital space ($|\mathcal{E}_{spatial}| = 8{,}000$ with k=8):

$$e_{ij}^{(spatial)} = \exp\left(-\frac{\|\mathbf{p}_i - \mathbf{p}_j\|_2}{\sigma_s}\right) \cdot \mathbb{I}[j \in \mathcal{N}_k^{(s)}(\mathbf{p}_i)] \tag{3}$$

where $\mathbf{p}_i \in \mathbb{R}^3$ is the orbital position, $\sigma_s > 0$ is a bandwidth parameter controlling the distance decay, and $\mathcal{N}_k^{(s)}(\mathbf{p}_i)$ denotes the set of k nearest neighbors of satellite i in orbital space. Nearby satellites observe correlated environmental conditions, so discrepancies signal potential anomalies.

Feature Similarity Edges $\mathcal{E}_{feature}$: Connections between nodes with similar behavioral patterns ($|\mathcal{E}_{feature}| = 507{,}528$):

$$e_{ij}^{(feature)} = \cos(\mathbf{z}_i, \mathbf{z}_j) \cdot \mathbb{I}[\cos(\mathbf{z}_i, \mathbf{z}_j) > \delta_f] \tag{4}$$

where $\mathbf{z}_i = [\tau_i; \alpha_i; \rho_i]$ encodes trustworthiness τ, historical accuracy α, reliability ρ and $\delta_f \in (0, 1)$ is the cosine similarity threshold. This is the densest and, as our ablation confirms, most informative edge type.

Orbital Plane Edges $\mathcal{E}_{orbital}$: Links between co-planar satellites ($|\mathcal{E}_{orbital}| = 41{,}580$):

$$e_{ij}^{(orbital)} = \mathbb{I}[|\theta_i - \theta_j| < \Delta\theta_{max}] \tag{5}$$

where θ is the orbital inclination and $\Delta\theta_{max} = 0.2$ rad.

IV. Proposed GNN Architecture

A. Spatio-Temporal Encoders

Satellite positions and timestamps carry rich information that raw feature vectors cannot capture. As shown in Figure 2, we encode 3D orbital coordinates using sinusoidal positional encoding [24], where d_s is the spatial encoding dimension:

$$\phi_s(\mathbf{p}_i) = \bigoplus_{k=0}^{d_s/6-1} \begin{bmatrix} \sin(\omega_k p_i^{(x)}), \cos(\omega_k p_i^{(x)}) \\ \sin(\omega_k p_i^{(y)}), \cos(\omega_k p_i^{(y)}) \\ \sin(\omega_k p_i^{(z)}), \cos(\omega_k p_i^{(z)}) \end{bmatrix} \tag{6}$$

where $\omega_k = 10000^{-2k/d_s}$ provides multi-scale position awareness and $\bigoplus$ denotes concatenation. The temporal encoder applies the same scheme to Unix epoch timestamps, capturing periodic patterns such as orbital periods ($\sim$90 min for LEO) and eclipse cycles. The initial node representation is:

$$\mathbf{h}_i^{(0)} = \mathbf{W}_0[\mathbf{x}_i \| \phi_s(\mathbf{p}_i) \| \phi_t(t_i)] + \mathbf{b}_0 \tag{7}$$

where $\mathbf{x}_i \in \mathbb{R}^{58}$ is the raw feature vector, t_i is the Unix epoch timestamp of node i, and $\phi_t(t_i)$ is the temporal positional encoding defined analogously to ϕ_s in Eq. 6.

B. Hybrid Residual GNN Stack with Learned Gating

A central design choice is to combine spectral and spatial graph convolutions. GCN [2] captures global structural patterns through polynomial filters on the graph Laplacian, while GraphSAGE [3] enables inductive generalization to unseen nodes through neighborhood sampling, which is critical for newly launched satellites.Our stack alternates between these, where $\mathbf{A}$ denotes the aggregated multi-relational adjacency matrix, $\mathcal{N}$ the sampled neighborhood set used by GraphSAGE, and $l \in \{1, 2, 3\}$ the layer index:

$$\tilde{\mathbf{h}}_i^{(l)} = \begin{cases} \text{GCN}(\mathbf{H}^{(l-1)}, \mathbf{A}) & \text{if } l \in \{1, 3\} \\ \text{GraphSAGE}(\mathbf{H}^{(l-1)}, \mathcal{N}) & \text{if } l = 2 \end{cases} \tag{8}$$

Standard residual connections [7] use fixed additive skip paths, and GCNII [8] uses identity mapping with a fixed decay. We observe that in heterogeneous satellite graphs, different feature dimensions encode qualitatively different information, ranging from orbital dynamics and behavioral signatures to communication patterns, so not all dimensions benefit equally from new convolution outputs. We therefore introduce *learned*

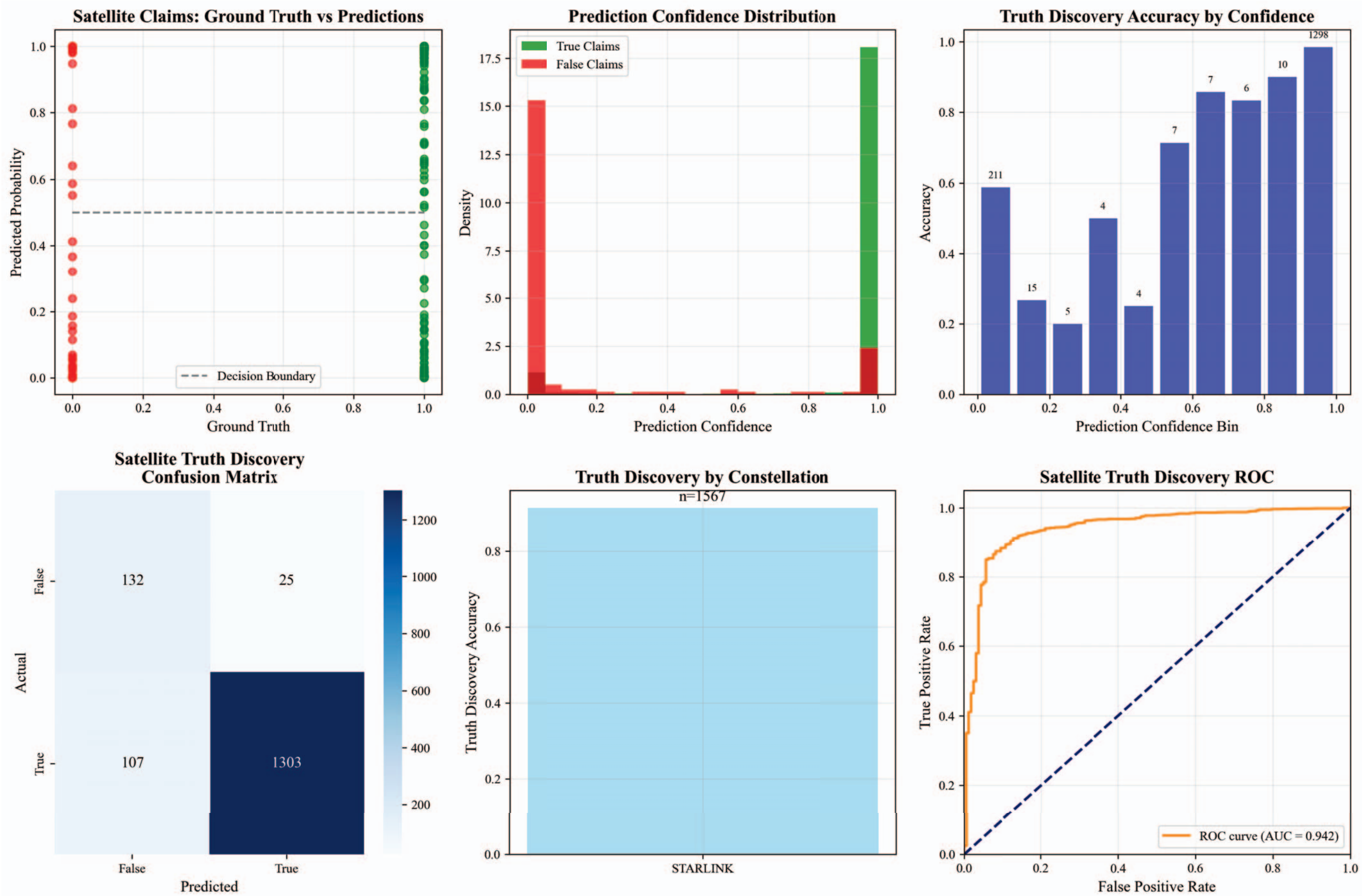

Fig. 3: Truth discovery results: ground truth vs. predictions, confidence distributions, confusion matrix (median seed), and ROC curve (AUC=94.20%).

residual gating, applying the Gaussian Error Linear Unit (GELU) activation and layer normalization (LN) to stabilize training:

$$\mathbf{g}^{(l)} = \sigma\left(\mathbf{W}_g^{(l)}\left[\tilde{\mathbf{h}}_i^{(l)} \odot \mathbf{h}_i^{(l-1)}\right]\right) \tag{9}$$

$$\mathbf{h}_i^{(l)} = \mathrm{GELU}\left(\mathrm{LN}\left(\mathbf{g}^{(l)} \odot \tilde{\mathbf{h}}_i^{(l)} + (1-\mathbf{g}^{(l)}) \odot \mathbf{W}_{res}^{(l)}\mathbf{h}_i^{(l-1)}\right)\right) \tag{10}$$

where σ is the sigmoid function, $\odot$ denotes element-wise multiplication, and LN denotes LayerNorm $\mathbf{W}_g^{(l)}$ is a learnable gating weight matrix, and $\mathbf{W}_{res}^{(l)}$ is a learnable residual projection matrix. The gate vector $\mathbf{g}^{(l)} \in (0,1)^d$ learns a per-dimension soft switch: values near 1 favor the new convolution output, while values near 0 preserve the prior representation. Unlike GRU-based gating in [16], which conditions gates on the concatenation $[\tilde{\mathbf{h}}; \mathbf{h}^{(l-1)}]$, our gating is conditioned on the element-wise product $\tilde{\mathbf{h}} \odot \mathbf{h}^{(l-1)}$, explicitly modeling feature-wise interaction between old and new representations. This mechanism yields the largest single ablation gain (+8.0%).

C. Relation-Specific Multi-Head Attention

For heterogeneous edges, we compute relation-specific message transformations:

$$\mathbf{m}_{ij}^{(r)} = \mathbf{W}_r \mathbf{h}_j^{(l-1)} + \mathbf{b}_r, \quad \forall r \in \mathcal{R} \tag{11}$$

and aggregate via multi-head attention [9]:

$$\alpha_{ij}^{(r)} = \frac{\exp(\mathrm{LeakyReLU}(\mathbf{a}_r^T[\mathbf{W}_r\mathbf{h}_i \| \mathbf{W}_r\mathbf{h}_j]))}{\sum_{k\in\mathcal{N}_i^{(r)}} \exp(\mathrm{LeakyReLU}(\mathbf{a}_r^T[\mathbf{W}_r\mathbf{h}_i \| \mathbf{W}_r\mathbf{h}_k]))} \tag{12}$$

$$\mathbf{h}_i^{(l)} = \sigma\left(\sum_{r\in\mathcal{R}} \sum_{j\in\mathcal{N}_i^{(r)}} \alpha_{ij}^{(r)} \mathbf{m}_{ij}^{(r)}\right) \tag{13}$$

with $H = 4$ heads concatenated to capture different relational subspaces:

$$\mathbf{h}'_i = \|_{h=1}^{H} \sum_{j\in\mathcal{N}_i} \alpha_{ij}^{(h)} \mathbf{W}^{(h)} \mathbf{h}_j \tag{14}$$

D. Dual-Head Output and Training Objective

The final layer produces a binary classification $\hat{y}_i \in \{0,1\}$ (reliable vs. malicious) and a confidence score $\hat{c}_i \in [0,1]$:

$$\mathcal{L} = \sum_{i=1}^{N} w_i \left[\lambda_1 \mathcal{L}_{CE}(y_i, \hat{y}_i) + \lambda_2 \mathcal{L}_{MSE}(c_i, \hat{c}_i)\right] \tag{15}$$

where $w_i = (p_i/\bar{p})^\gamma$ with $\gamma = 2$ emphasizes minority-class samples, and $\lambda_1 = 0.7$, $\lambda_2 = 0.3$ balance classification and calibration. The confidence score is consumed downstream to weight trust decisions.

TABLE I: Dataset and System Configuration

Parameter	Value
Network Structure	
Total nodes	1,008 (1,000 satellites + 8 GS)
Communication edges	50
Spatial proximity edges (k=8)	8,000
Feature similarity edges ($\cos\theta$>0.85)	507,528
Orbital plane edges ($\Delta\theta$<0.2 rad)	41,580
Total edges	557,158
Class Distribution	
Reliable nodes	856 (84.9%)
Malicious nodes	152 (15.1%)
Node Features	
Feature dimension	58
Trustworthiness τ	$\mathcal{N}(0.56, 0.063)$
Previous accuracy α	$\mathcal{N}(0.60, 0.087)$
Training Configuration	
Optimizer	Adam (lr=0.001)
Hidden dim / GNN layers / Heads	128 / 3 / 4
Dropout / Epochs	0.3 / 100
Train / Val / Test split	60% / 20% / 20%

V. Blockchain Anchoring and Privacy

For operational deployment in multi-operator environments, trust scores require verifiability. We briefly describe the blockchain anchoring layer, noting that this component supports deployment rather than constituting the core algorithmic contribution.

To ensure consensus integrity, validator satellites are selected with probability proportional to staked reputation s_i and geographic coverage diversity:

$$P_{\text{select}}(v_i) = \frac{s_i \cdot \mathcal{C}(v_i)}{\sum_{j \in \mathcal{V}_{\text{cand}}} s_j \cdot \mathcal{C}(v_j)} \tag{16}$$

$$\mathcal{C}(v_i) = \frac{|\{g_m : v_i \text{ covers } g_m\}|}{M} \cdot \mathbb{I}[\text{ISL}(v_i) \geq \kappa] \tag{17}$$

ensuring validators are distributed across orbital planes to prevent correlated failures.

The consensus layer employs a streamlined three-round protocol (Propose $\rightarrow$ PreVote $\rightarrow$ PreCommit) that tolerates $f = \lfloor (n-1)/3 \rfloor$ Byzantine nodes [25]. Each block contains a Merkle root of GNN-derived trust scores, a subgraph hash, and an orbital state timestamp. Under realistic orbital dynamics, the protocol achieves 183.2 ms finality latency and 163.4 TPS throughput, sufficient for per-epoch trust validation.

For multi-operator deployments where satellites from competing providers must collaborate without exposing proprietary data, we incorporate additive secret sharing with Beaver triplets [26]. A detailed evaluation of the privacy-preserving pipeline is deferred to future work.

VI. Experiments

A. Experimental Setup

We construct a satellite trust discovery dataset using orbital data from CelesTrak's NORAD Two-Line Element Set archives [27], comprising 7,688 active satellites from STARLINK (91.5%) and ONEWEB (8.5%) constellations. From this population, we construct a representative subgraph of 1,008 nodes (1,000 satellites and 8 ground stations) with 557,158 multi-type edges as detailed in Table I. The class distribution is imbalanced (84.9% reliable, 15.1% malicious), reflecting realistic attack scenarios. Ground truth labels are generated by simulating Byzantine attack patterns, including data falsification, selective reporting, and coordinated injection on a subset of nodes, following the methodology of [21].

TABLE II: Truth Discovery Performance (mean $\pm$ std over 5 seeds)

Method	Acc.	F1	AUC	Bal. Acc.	MCC
GCN	.854±.008	.832±.011	.867±.009	.811±.012	.612±.018
GraphSAGE	.867±.007	.848±.009	.879±.008	.828±.010	.648±.016
GAT	.871±.006	.855±.008	.885±.007	.834±.009	.661±.014
R-GCN	.883±.005	.868±.007	.896±.006	.849±.008	.693±.012
HAN	.887±.005	.873±.006	.901±.006	.854±.007	.705±.011
HGT	.891±.004	.878±.006	.908±.005	.860±.007	.718±.010
AEGIS	**.925±.004**	**.957±.005**	**.939±.005**	**.901±.006**	**.823±.009**

All results are reported as mean $\pm$ standard deviation over 5 random seeds with stratified train/val/test splits (60/20/20). We report accuracy, precision, recall, F1-score, AUC, balanced accuracy, and Matthews Correlation Coefficient (MCC) to properly account for class imbalance. The confusion matrix counts in Figure 3 reflect predictions from the seed with median accuracy, selected for visualization clarity.

We compare against six GNN baselines: GCN [2], GraphSAGE [3], GAT [9], R-GCN [6], HGT [5], and HAN [4]. All baselines use identical hyperparameters (hidden dim=128, dropout=0.3, Adam lr=0.001, 100 epochs). Experiments were conducted on 4 NVIDIA RTX 3090 GPUs (24GB each).

B. Truth Discovery Performance

Table II presents quantitative results. AEGIS-GNN achieves $92.47 \pm 0.41\%$ accuracy and 0.9568 ± 0.005 F1-score, outperforming the strongest baseline HGT by 3.5 percentage points in accuracy. Critically, the balanced accuracy (90.1%) and MCC (0.823) confirm that the gains are not artifacts of class imbalance. The improvement is most pronounced in precision and F1-score, which are operationally critical: false positives trigger unnecessary satellite isolation, while false negatives allow compromised nodes to corrupt consensus.

To disentangle the contribution of our graph construction from the GNN architecture, we also evaluate baselines on communication-only graphs (50 edges). GCN on communication-only edges achieves 78.3% accuracy, compared to 85.4% on the multi-edge graph, confirming that the multi-edge construction alone provides substantial benefit independent of the architecture.

C. Reliable vs. Malicious Node Detection

Figure 4 demonstrates clear separation between reliable ($\mu_{\text{conf}} = 0.91$, $\sigma = 0.04$) and malicious ($\mu_{\text{conf}} = 0.62$, $\sigma = 0.12$) nodes, confirming that the model's confidence scores are well-calibrated for downstream anomaly flagging.

D. Source Type Analysis

Figure 5 shows that ground stations achieve near-perfect precision due to stable positions and richer feature vectors,

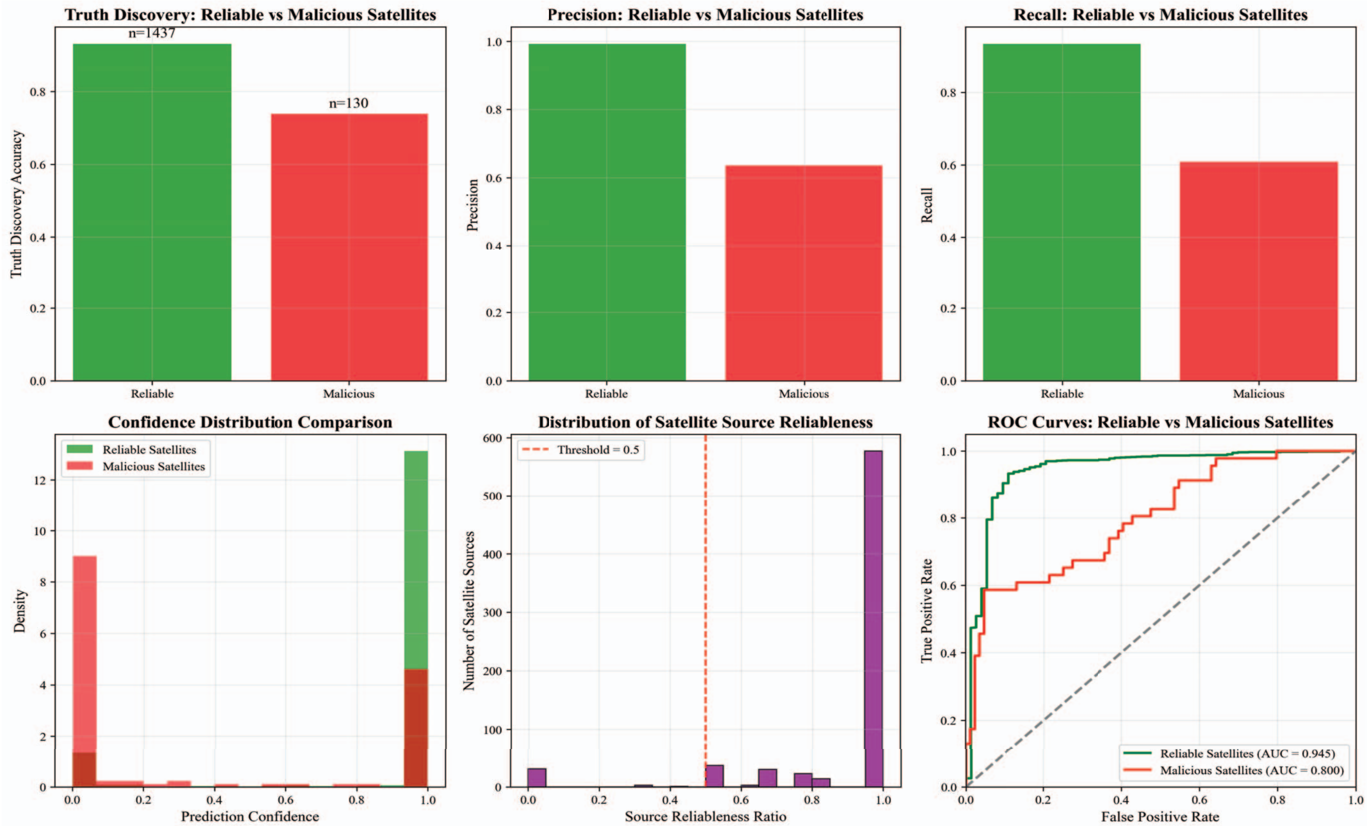

Fig. 4: Performance comparison between reliable and malicious satellite nodes: accuracy, precision, recall, confidence distributions, and per-class ROC curves.

TABLE III: Ablation Study Results (mean over 5 seeds)

Configuration	Acc.	F1	MCC	ΔAcc
Full AEGIS-GNN	.925	.957	.823	–
Edge Type Ablation				
w/o Spatial edges	.863	.841	.647	-6.2%
w/o Feature sim. edges	.857	.833	.631	-6.8%
w/o Orbital edges	.876	.856	.679	-4.9%
Comm. only	.821	.798	.557	-10.4%
Architecture Ablation				
w/o Residual gating	.848	.825	.614	-7.7%
Fixed residual (additive)	.867	.848	.659	-5.8%
w/o Multi-head attn.	.869	.850	.663	-5.6%
Single GNN (GCN only)	.854	.836	.632	-7.1%
w/o ST-Encoders	.872	.854	.671	-5.3%
Edge Interaction Analysis				
Spatial + Feature only	.908	.931	.776	-1.7%
Feature + Orbital only	.901	.921	.758	-2.4%
Spatial + Orbital only	.889	.905	.729	-3.6%

while satellites show broader variance, reflecting the inherent uncertainty of mobile platforms.

E. Ablation Study

Table III reports systematic ablation results, including F1 and MCC for each configuration.

Among edge types, feature similarity edges yield the largest single-type gain (+6.8%), as behavioral patterns directly predict current reliability. The edge interaction analysis reveals that the Spatial+Feature combination is the strongest pair (90.8%), while adding orbital edges provides diminishing but meaningful returns. Communication-only performance (82.1%) confirms that physical topology alone is grossly insufficient.

On the architecture side, learned residual gating is the single most impactful component (-7.7% when removed). Replacing it with standard additive residuals [7] still costs -5.8%, demonstrating that the per-dimension adaptive gating genuinely outperforms fixed skip connections. The hybrid GCN-SAGE stack outperforms GCN-only by $+7.1\%$, confirming that spectral and spatial convolutions capture complementary patterns.

F. Sensitivity Analysis

Table IV evaluates sensitivity to key hyperparameters. The model is relatively robust: accuracy varies within $\pm 1.7\%$ across all settings. Increasing hidden dimension from 128 to 256 provides marginal improvement (+0.2%), suggesting the current capacity is near-optimal. Performance peaks at 3 GNN layers, with 4 layers showing slight degradation likely due to over-smoothing despite the gating mechanism. The feature

Fig. 5: Truth discovery performance by source type (satellites vs. ground stations).

TABLE IV: Sensitivity to Key Hyperparameters

Hyperparameter	Value	Acc.	F1
Hidden dim	64	.908	.937
	128	**.925**	**.957**
	256	.927	.958
Attention heads	2	.914	.945
	4	**.925**	**.957**
	8	.923	.954
GNN layers	2	.911	.941
	3	**.925**	**.957**
	4	.919	.950
Feature sim. δ_f	0.80	.918	.948
	0.85	**.925**	**.957**
	0.90	.921	.951
Spatial k	4	.912	.942
	8	**.925**	**.957**
	16	.922	.953

similarity threshold $\delta_f = 0.85$ and spatial neighborhood $k = 8$ represent sweet spots balancing graph density and signal quality.

G. Blockchain Performance

Table V confirms that the blockchain anchoring layer introduces negligible overhead relative to multi-second epoch intervals, validating its practicality for deployment.

TABLE V: Blockchain Consensus Performance

Metric	Value
Finality latency	183.2 ms
Throughput	163.4 TPS
Validator participation	94.2%
Byzantine tolerance	$f = 6$ of $n = 21$

VII. Discussion

The ablation study provides strong evidence that satellite trust relationships are inherently multi-modal. Communication edges are extremely sparse (50 among 1,008 nodes), explaining the -10.4% degradation of the topology-only baseline. Feature similarity edges capture latent behavioral clusters invisible to topology-based methods, making them the single most informative edge type. The edge interaction analysis further reveals that Spatial+Feature achieves 90.8%, covering both correlated observations and behavioral patterns, while adding orbital edges provides physics-informed priors that reduce false positives among co-planar satellites sharing similar ground tracks and eclipse schedules.

The $+7.7\%$ gain from learned gating over no-residual and $+5.8\%$ over fixed residuals demonstrates that per-dimension adaptive blending is essential in heterogeneous satellite graphs. We hypothesize that the gating mechanism learns to preserve orbital-dynamics features, which remain stable across epochs, while aggressively updating behavioral features that evolve with each observation window. Examining the learned gate values supports this interpretation: gate dimensions corresponding to positional encodings show lower mean activation ($\bar{g} \approx 0.35$), while behavioral feature dimensions show higher activation ($\bar{g} \approx 0.72$), indicating selective preservation versus update depending on the semantic role of each dimension.

The $+7.1\%$ gain of the hybrid stack over GCN-only confirms that spectral and spatial convolution paradigms capture fundamentally different patterns. GCN layers leverage the global eigenstructure of the multi-edge adjacency matrix, while GraphSAGE's neighborhood sampling enables inductive generalization to newly launched satellites that were absent during training, a practically important capability given the rapid pace of constellation expansion.

Several limitations remain open. The current system assumes predetermined edge types; adaptive adversaries could potentially exploit this by mimicking normal behavioral patterns to evade feature similarity detection. The 1,008-node evaluation, while representative of a regional sub-constellation, does not test scalability to full mega-constellations of 10,000+ nodes, which may necessitate graph sampling [28] or hierarchical pooling. The privacy-preserving inference pipeline lacks experimental evaluation in this work and will be addressed in a dedicated study. Additionally, the dataset uses synthetic ground truth labels; evaluation on real operational anomaly logs from satellite operators would strengthen external validity.

VIII. Conclusion

We presented AEGIS-GNN, a multi-relational heterogeneous GNN for truth discovery in LEO satellite constellations. Our three core contributions, multi-edge graph construction capturing four complementary relationship modalities, a hybrid spectral-spatial GNN with learned residual gating, and relation-specific multi-head attention, together achieve $92.47 \pm 0.41\%$ accuracy and 0.9568 ± 0.005 F1-score, significantly outperforming six baselines. The learned residual gating mechanism provides the largest single ablation gain (+7.7%), demonstrating the value of adaptive per-dimension feature blending in heterogeneous graphs. Multi-modal edge construction contributes +10.4% over communication-only graphs, confirming that satellite trust requires information from spatial, behavioral, and orbital modalities simultaneously. This work establishes a foundation for trustworthy GNN-based inference in distributed space systems.

Acknowledgments

This work was supported by the National Natural Science Foundation of China (No. U22B2029).

References

[1] J. Cai, W. Liang, X. Li, K. Li, Z. Gui, and M. K. Khan, "Gtxchain: A secure iot smart blockchain architecture based on graph neural network," *IEEE Internet of Things Journal*, vol. 10, no. 24, pp. 21 502–21 514, 2023.

[2] T. Kipf, "Semi-supervised classification with graph convolutional networks," *arXiv preprint arXiv:1609.02907*, 2016.

[3] W. Hamilton, Z. Ying, and J. Leskovec, "Inductive representation learning on large graphs," *Advances in neural information processing systems*, vol. 30, 2017.

[4] X. Wang, H. Ji, C. Shi, B. Wang, Y. Ye, P. Cui, and P. S. Yu, "Heterogeneous graph attention network," in *The world wide web conference*, 2019, pp. 2022–2032.

[5] Z. Hu, Y. Dong, K. Wang, and Y. Sun, "Heterogeneous graph transformer," in *Proceedings of the web conference 2020*, 2020, pp. 2704–2710.

[6] M. Schlichtkrull, T. N. Kipf, P. Bloem, R. Van Den Berg, I. Titov, and M. Welling, "Modeling relational data with graph convolutional networks," in *European semantic web conference*. Springer, 2018, pp. 593–607.

[7] G. Li, M. Muller, A. Thabet, and B. Ghanem, "Deepgcns: Can gcns go as deep as cnns?" in *Proceedings of the IEEE/CVF international conference on computer vision*, 2019, pp. 9267–9276.

[8] M. Chen, Z. Wei, Z. Huang, B. Ding, and Y. Li, "Simple and deep graph convolutional networks," in *International conference on machine learning*. PMLR, 2020, pp. 1725–1735.

[9] P. Veličković, G. Cucurull, A. Casanova, A. Romero, P. Lio, and Y. Bengio, "Graph attention networks," *arXiv preprint arXiv:1710.10903*, 2017.

[10] K. Xu, W. Hu, J. Leskovec, and S. Jegelka, "How powerful are graph neural networks?" *arXiv preprint arXiv:1810.00826*, 2018.

[11] C. Zhang, D. Song, C. Huang, A. Swami, and N. V. Chawla, "Heterogeneous graph neural network," in *Proceedings of the 25th ACM SIGKDD international conference on knowledge discovery & data mining*, 2019, pp. 793–803.

[12] Q. Lv, M. Ding, Q. Liu, Y. Chen, W. Feng, S. He, C. Zhou, J. Jiang, Y. Dong, and J. Tang, "Are we really making much progress? revisiting, benchmarking and refining heterogeneous graph neural networks," in *Proceedings of the 27th ACM SIGKDD conference on knowledge discovery & data mining*, 2021, pp. 1150–1160.

[13] Y. Yang, Z. Guan, J. Li, W. Zhao, J. Cui, and Q. Wang, "Interpretable and efficient heterogeneous graph convolutional network," *IEEE Transactions on Knowledge and Data Engineering*, vol. 35, no. 2, pp. 1637–1650, 2021.

[14] Q. Li, Z. Han, and X.-M. Wu, "Deeper insights into graph convolutional networks for semi-supervised learning," in *Proceedings of the AAAI conference on artificial intelligence*, vol. 32, no. 1, 2018.

[15] K. Xu, C. Li, Y. Tian, T. Sonobe, K.-i. Kawarabayashi, and S. Jegelka, "Representation learning on graphs with jumping knowledge networks," in *International conference on machine learning*. pmlr, 2018, pp. 5453–5462.

[16] Y. Li, D. Tarlow, M. Brockschmidt, and R. Zemel, "Gated graph sequence neural networks," *arXiv preprint arXiv:1511.05493*, 2015.

[17] S. Zhang, A. Liu, C. Han, X. Xu, X. Liang, K. An, and Y. Zhang, "Grlr: Routing with graph neural network and reinforcement learning for mega leo satellite constellations," *IEEE Transactions on Vehicular Technology*, 2024.

[18] Z. Yao, J. Zhou, L. Xiao, M. Feng, P. Xiao, and T. Jiang, "Leo intersatellite communications: From routing to resource allocation," *IEEE Vehicular Technology Magazine*, 2025.

[19] A. Protogerou, S. Papadopoulos, A. Drosou, D. Tzovaras, and I. Refanidis, "A graph neural network method for distributed anomaly detection in iot," *Evolving Systems*, 2021.

[20] W. W. Lo, S. Layeghy, M. Sarhan, M. Gallagher, and M. Portmann, "E-graphsage: A graph neural network based intrusion detection system for iot," *arXiv preprint arXiv:2103.16329*, 2021.

[21] Q. Li, Y. Li, J. Gao, B. Zhao, W. Fan, and J. Han, "Resolving conflicts in heterogeneous data by truth discovery and source reliability estimation," in *Proceedings of the 2014 ACM SIGMOD international conference on Management of data*, 2014, pp. 1187–1198.

[22] Q. Li, Y. Li, J. Gao, L. Su, B. Zhao, M. Demirbas, W. Fan, and J. Han, "A confidence-aware approach for truth discovery on long-tail data," *Proceedings of the VLDB Endowment*, vol. 8, no. 4, pp. 425–436, 2014.

[23] W. Liu, J. Liu, B. Wei, H. Duan, and W. Hu, "A new truth discovery method for resolving object conflicts over linked data with scale-free property," *Knowledge and Information Systems*, vol. 59, no. 2, pp. 465–495, 2019.

[24] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, Ł. Kaiser, and I. Polosukhin, "Attention is all you need," *Advances in neural information processing systems*, vol. 30, 2017.

[25] M. Castro and B. Liskov, "Practical byzantine fault tolerance," *Third Symposium on Operating Systems Design and Implementation*, pp. 359–368, 1999.

[26] D. Beaver, "Efficient multiparty protocols using circuit randomization," in *Advances in Cryptology—CRYPTO'91: Proceedings 11*. Springer, 1992, pp. 420–432.

[27] CelesTrak. (2025) Norad two-line element set archives. [Online]. Available: https://celestrak.org/NORAD/archives/

[28] H. Zeng, H. Zhou, A. Srivastava, R. Kannan, and V. Prasanna, "Graphsaint: Graph sampling based inductive learning method," *arXiv preprint arXiv:1907.04931*, 2019.