Airdrop Hunter Detection via PageRank-Augmented Multimodal Graph Neural Networks

Jiajie Shi[1,2,3†], Yuyang Qin[1,2,4†], Hengming Dai[5,6], Xiaoyi Fan[1,2], Haihan Duan[1,2*]
[1]Artificial Intelligence Research Institute, Shenzhen MSU-BIT University, Guangdong, China
[2]Guangdong-Hong Kong-Macao Joint Laboratory for Emotion Intelligence and Pervasive Computing, Guangdong, China
[3]The University of Hong Kong, Hong Kong, China
[4]The Chinese University of Hong Kong, Shenzhen, Guangdong, China
[5]Institute of International Rivers and Eco-Security, Yunnan University, Kunming, China
[6]School of Earth Sciences, Yunnan University, Kunming, China
Email: jiajieshi@connect.hku.hk, yuyangqin1@link.cuhk.edu.cn, daihm@ynu.edu.cn,
xiaoyi.fan@smbu.edu.cn, duanhaihan@smbu.edu.cn

2025 IEEE International Conference on Cloud Computing Technology and Science (CloudCom) | 979-8-3315-6634-0/25/$31.00 ©2025 IEEE | DOI: 10.1109/CLOUDCOM67567.2025.11331465

***Abstract*—Airdrops are a widely used mechanism in Web3 ecosystems to incentivize early users by distributing governance tokens. However, these mechanisms are increasingly targeted by airdrop hunters—malicious actors who exploit token distribution systems through address farming, automated scripts, and behavioral camouflage. While prior work such as ARTEMIS leverages multimodal features and local transaction patterns to detect such behavior, it lacks a global understanding of wallet influence in the transaction graph. In this paper, we propose an enhanced detection framework that augments the ARTEMIS by incorporating PageRank-based global centrality as an additional structural feature. This allows the model to better distinguish superficially active wallets from those with broader influence in the network. We evaluate our method on real-world Non-Fungible Token (NFT) data from the Blur marketplace and achieve state-of-the-art performance. Furthermore, a feature substitution experiment reveals that simple degree-based features alone can achieve near-perfect performance, even outperforming PageRank, suggesting that the labels are strongly coupled with topological properties. These findings highlight both the effectiveness of structural augmentation and the potential risks of shortcut learning in graph-based detection systems.**

***Index Terms*—Blockchain Security, Airdrop Hunter Detection, Graph Neural Networks, PageRank, On-chain Behavior Analysis, Non-Fungible Token**

I. Introduction

The rapid development of blockchain and decentralized finance (DeFi) has fostered a new era of user participation and asset distribution strategies in the Web3 ecosystem [1]. Among these, airdrops—Fungible Token (FT) or Non-Fungible Token (NFT) distributions to eligible addresses—have become a widely adopted mechanism to incentivize early adopters, reward loyal users, and bootstrap new communities [2]. Airdrops not only help projects gain traction but also create early liquidity and user engagement. However, their increasing popularity has inadvertently attracted strategic exploitation by malicious actors known as *airdrop hunters*.

These actors often engage in Sybil attacks by generating or controlling multiple pseudo-anonymous wallets to illegitimately claim disproportionate airdrop allocations. Unlike traditional Sybil threats in peer-to-peer networks, airdrop hunters are economically incentivized actors who actively simulate legitimate user behaviors—such as NFT purchases, staking, or social interactions—to circumvent heuristic-based eligibility filters. In practice, such manipulation leads to unfair token distribution, rapid post-airdrop dumping, and compromised governance processes, all of which undermine the long-term sustainability of Web3 projects.

Existing mitigation strategies—such as IP address tracking, CAPTCHA-based human verification, and basic rule-based filtering—offer limited effectiveness in adversarial blockchain ecosystems. This is primarily due to the pseudonymous nature of blockchain interactions and the ease with which attackers can deploy large-scale automated scripts to bypass these safeguards, rendering traditional defenses insufficient against sophisticated Sybil behaviors and airdrop exploitation. To mitigate these threats, recent research has proposed more sophisticated detection mechanisms based on graph learning. Among them, ARTEMIS [3] stands out as a multimodal graph neural network (GNN) framework that integrates behavioral signals with visual and textual NFT features to detect airdrop hunters as a node classification task. Such multimodal modeling approaches have also proven effective in broader metaverse and identity modeling scenarios [4].

Despite its effectiveness, ARTEMIS and related models rely primarily on local sampling, neighborhood aggregation, and shallow transaction patterns. These methods lack a global understanding of node roles in the broader wallet interaction network, which is crucial for identifying highly coordinated but stealthy hunter networks. For example, airdrop hunters may operate across disjoint subgraphs and avoid forming dense clusters, making them harder to detect with localized GNNs alone.

To address this limitation, we propose a structural enhancement to the ARTEMIS framework by incorporating the global PageRank centrality score into each wallet's node

†Jiajie Shi and Yuyang Qin are visiting students at Artificial Intelligence Research Institute and Guangdong-Hong Kong-Macao Joint Laboratory for Emotion Intelligence and Pervasive Computing, Shenzhen MSU-BIT University, China.
*Haihan Duan is the corresponding author (duanhaihan@smbu.edu.cn).

representation. PageRank captures the relative importance of nodes within the entire transaction network and has been widely used in applications ranging from web search to fraud detection [5], [6]. By integrating this signal, our model can better distinguish truly influential wallet nodes from those that are only superficially active in limited contexts.

In addition, we perform a feature substitution analysis by replacing PageRank with traditional graph metrics—namely, in-degree and out-degree. Interestingly, the model performance not only did not degrade but dramatically improved, suggesting that both global and local structural signals are strongly correlated with hunter labels. This raises important questions about potential feature redundancy, shortcut learning, and representation bias, underscoring the need for more rigorous feature disentanglement in future research.

To validate our approach, we utilize the Blur NFT marketplace dataset used in ARTEMIS [3], which includes verified airdrop hunter addresses, transactional relationships, and associated multimodal NFT features. Through extensive experiments and qualitative case studies, we demonstrate that our proposed framework outperforms existing baselines across precision, recall, and F1-score, and reveals interpretable structural patterns behind collusive airdrop-hunting behavior. The codes are available at https://github.com/kid1412-hku/Airdrop-Hunter-Detection-Cloudcom2025.git.

Our contributions are summarized as follows:

- We propose a PageRank-enhanced GNN-based framework that improves upon the ARTEMIS model by incorporating global structural awareness into wallet representations.
- We conduct a feature substitution analysis using degree-based metrics to assess the relative importance of global versus local structural signals, uncovering potential risks of shortcut learning.
- We demonstrate the effectiveness of the enhanced model on a real-world NFT transaction dataset from the Blur marketplace, achieving strong performance across Precision, Recall, and F1-score metrics, and provide a visual interpretation of hunter communities.

II. Related Work

A. Airdrop Hunters and Sybil Behavior in Web3

Airdrop hunters are a growing group of economically incentivized adversaries in Web3 ecosystems, who systematically exploit token distribution mechanisms (designed to reward early adopters or community contributors) via Sybil attacks, automation, and behavioral camouflage—including creating numerous controlled wallet addresses, timing NFT interactions strategically, and mimicking organic user behavior to maximize token rewards [7], [8].

Their tactics draw on classical Sybil attacks in distributed systems [9], where adversaries generate fake identities to gain disproportionate influence; in blockchain contexts, this manifests as orchestrated transaction patterns, wallet cycling, and exploitation of multi-account reward schemes.

Unchecked airdrop hunting carries notable socio-economic consequences: studies [7], [10] show it distorts token governance outcomes, inflates user engagement metrics, damages decentralized ecosystems' long-term credibility, and links to post-airdrop token dumping (causing price volatility and reduced community cohesion)—undermining fair, merit-based token distribution.

Despite growing awareness, most existing efforts remain descriptive, focusing on post-hoc observations or heuristic labeling of suspicious behaviors [3], rather than robust, generalizable detection frameworks. Challenges like noisy labeling, evolving attacker tactics, and multi-modal behavior simulation further render traditional rule-based methods increasingly ineffective.

B. Detection of Airdrop Hunters: Graph-Based Approaches

In response to these challenges, a growing body of work has proposed graph-based modeling techniques to capture the relational nature of Web3 transactions. One of the most representative frameworks is **ARTEMIS** [3], which treats the detection task as a node classification problem over heterogeneous transaction graphs. By integrating textual and visual metadata of NFTs with wallet-level behavioral features, ARTEMIS offers a multimodal perspective that enhances model robustness.

Several follow-up works have extended this direction. **ARTEMIX** [11], for instance, combines multimodal learning with community detection and time-series anomaly detection, enabling detection of subtle collusive patterns and temporal inconsistencies. Although this hybrid design improves explainability and clustering effectiveness, it introduces dependency on handcrafted rules and post-processing heuristics, which may not generalize well to adversarial settings or across NFT platforms with differing airdrop policies.

Other graph learning models such as GCN [12], GraphSAGE [13], and GAT [14] have been evaluated in similar Sybil or fraud detection tasks [15], but they primarily operate on local node neighborhoods. As a result, their expressiveness is limited in capturing globally dispersed coordination rings—a critical weakness when detecting airdrop hunters who distribute their activity across multiple wallets to avoid localized clustering signals.

Several recent studies in fraud detection [16] emphasize the need for incorporating higher-order structure, temporal dynamics, and heterogeneous modalities. These directions point toward more unified detection architectures that combine local transaction patterns, wallet attributes, and long-range topological signals.

C. PageRank in Blockchain Analysis

To enhance global reasoning capabilities, researchers have begun to integrate global structural signals such as node centrality and clustering coefficients. Among these, **PageRank** [5] has emerged as a particularly effective tool due to its ability to quantify node importance within directed, weighted graphs.

In the blockchain domain, PageRank has been used to detect financial fraud [17], evaluate wallet trustworthiness [18], and assess influence in DAO governance [19]. Its adoption in NFT ecosystems, however, remains sparse. One notable application is in [20], where PageRank helped uncover laundering flows by identifying wallets disproportionately central to transaction loops.

By assigning more weight to nodes with recursive importance—those who transact with other important nodes—PageRank offers a perspective on influence and coordination that goes beyond frequency or degree. When integrated into graph learning pipelines, this signal complements GNNs' local propagation mechanisms, allowing for multiscale representation learning that captures both behavioral and structural anomalies.

Our work builds on this intuition by augmenting the ARTEMIS framework with PageRank-based features. This enhancement allows the model to better differentiate between superficially active but structurally peripheral actors and those exhibiting strategic centrality within collusive networks. To the best of our knowledge, this is the first attempt to integrate global graph signals directly into the airdrop hunter detection pipeline.

III. Methodology

A. Problem Formulation

We formulate airdrop hunter detection as a supervised binary node classification task on a directed and attributed graph $G = (V, E, X)$ constructed from NFT transactions. Specifically:

- V denotes the set of wallet addresses (nodes),
- E represents the set of directed edges where $(u, v) \in E$ implies wallet u transferred an NFT to v,
- $X \in \mathbb{R}^{|V| \times d}$ is the node feature matrix, which fuses multimodal NFT content, behavioral metrics, and structural properties.

Each node $v \in V$ is associated with a binary label $y_v \in \{0, 1\}$, where $y_v = 1$ indicates an airdrop hunter and $y_v = 0$ otherwise. The learning objective is to optimize a classifier f_θ that maps wallet nodes to class labels, using a stratified training-validation split and evaluated via precision, recall, and F1-score.

B. Data Source and Labeling

The transaction data was collected from the Blur NFT marketplace between October 19, 2022 and April 1, 2023 via the Etherscan API, yielding 2,453,280 on-chain NFT transfers across 203,370 wallet addresses. We also scraped metadata for 1,155,947 NFT assets, including image URLs, textual descriptions, and trait annotations using OpenSea APIs [21].

For label generation, we adopted the behavior-centric clustering and expert verification framework of [10], where wallets are embedded using interaction patterns, then clustered via Agglomerative Hierarchical Clustering (AHC). Blockchain experts manually labeled each cluster, resulting in 4,808 airdrop hunter labels (about 4% of the total). To ensure robustness, transaction loop detection and wash-trading diagnostics [20] were also performed.

C. Multimodal and Behavioral Feature Design

Our feature set spans three perspectives: asset content, wallet behavior, and graph topology.

1) Multimodal NFT Features: To capture the intrinsic value of received assets, we apply pretrained BERT [22] and ViT [23] models to extract text and image embeddings from NFT metadata. The resulting vectors are fused via gated attention and aggregated across each wallet's transaction history, forming a content-level feature representation.

2) Behavioral Features: Inspired by [7], [24]–[26], we incorporate wallet activeness (contract call frequency, unique counterparties), airdrop frequency (not used in training), and manipulation indicators such as Benford's Law deviation and NFT flipping frequency. These indicators help differentiate legitimate users from synthetic farming behavior.

D. Graph Structural Features

1) Path-Aware Neighborhood Sampling: We follow [3] to design a neighborhood sampling strategy that filters 1-hop and 2-hop neighbors based on transaction path similarity. This alleviates noise from large hubs and preserves trading semantics in the subgraph structure.

2) PageRank Centrality (Our Extension): As an enhancement to ARTEMIS, we introduce PageRank as a global structural feature to capture each wallet's influence within the transaction graph. Specifically, the directed PageRank score for node v is computed as:

$$\mathrm{PR}(v) = \frac{1-d}{|V|} + d \sum_{u \in \mathrm{In}(v)} \frac{\mathrm{PR}(u)}{\mathrm{OutDegree}(u)}, \tag{1}$$

where $d = 0.85$ is the damping factor, $\mathrm{In}(v)$ denotes the set of predecessor nodes, and $|V|$ is the total node count. The iterative process continues until convergence, and normalized scores are appended to node features.

To implement this, we convert the PyTorch Geometric graph into a NetworkX object and compute the PageRank scores on the undirected version of the transaction graph using CPU resources. These scores are then concatenated as an additional scalar feature dimension for each node:

$$\mathbf{x}'_v = [\mathbf{x}_v \parallel \mathrm{PR}(v)], \tag{2}$$

where $\mathbf{x}_v$ is the original node feature vector. The updated features are transferred back to the GPU for downstream GNN training, ensuring consistent device alignment with the labels.

This feature enriches the GNN's ability to differentiate between locally active hubs and globally significant participants. For example, wallets exhibiting high transaction volume within closed loops (i.e., farming clusters) tend to receive lower PageRank scores than dispersed yet central actors. Thus, PageRank helps suppress false positives and improves recall for stealthy airdrop hunters.

E. Model Architecture

Our model is built upon the ARTEMIS framework [3], which employs a 2-layer Graph Neural Network (GNN) with path-aware neighbor sampling, fused multimodal features, and behavior-derived indicators. As shown in Figure 1, the architecture takes as input the concatenated representation of multimodal content features, behavioral features, and PageRank scores, and outputs a binary classification label per node. The overall structure enables the model to leverage both local patterns and global signals when reasoning over transaction behavior.

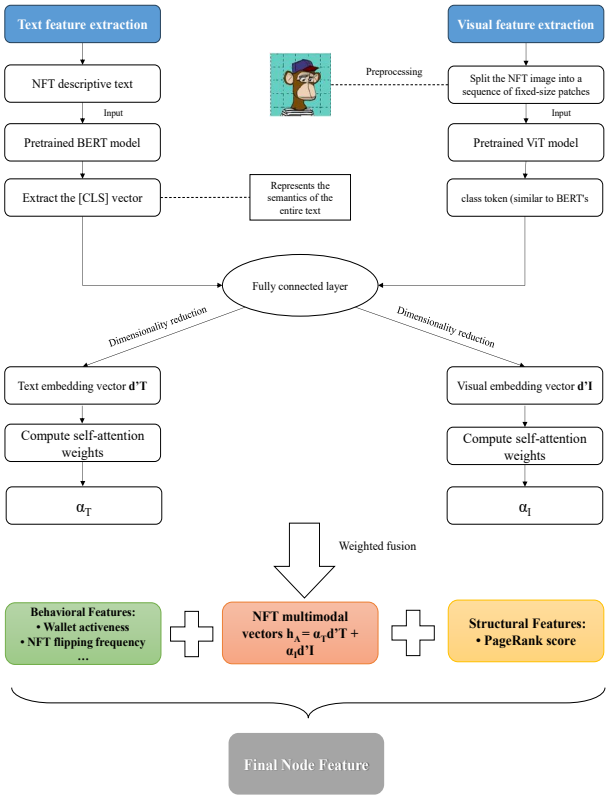

Fig. 1. Node feature construction diagram. The model integrates multimodal NFT content embeddings, behavioral features, and structural features into a unified node representation.

IV. Experiments and Analysis

A. Experimental Setup

To rigorously evaluate the effectiveness of our airdrop hunter detection model, we conducted extensive experiments on the constructed dataset under a unified experimental environment. The experimental setup is summarized as follows.

We split the full dataset using a 9:1 ratio into training and validation sets. The training set is used for model fitting and hyperparameter tuning, while the validation set is used to assess the generalization ability of the model on unseen data.

The hardware and software configurations used during training and evaluation are summarized in TABLE I:

TABLE I
Hardware and Software Configuration

Category	Details
Hardware	
Processor	AMD Ryzen 7 4800H (8 cores, 2.9 GHz)
Memory	16 GB DDR4
GPU	NVIDIA GeForce GTX 1650 (4GB VRAM)
Software	
OS	Windows 11
Development Environment	Jupyter Notebook
Framework	PyTorch 2.5.0, CUDA 12.4

The multimodal feature extraction module utilizes two parallel pipelines: one based on Vision Transformer (ViT-base, patch size 16×16, image size 224×224) and the other based on a pretrained BERT model to encode image and text features of NFTs, respectively.

The graph learning component is a two-layer Graph Neural Network (GNN) used for wallet node classification. Key hyperparameters used in the experiments are listed in TABLE II:

TABLE II
Model Hyperparameter Settings

GNN Module		Transformer Module	
Component	**Setting**	**Component**	**Setting**
Neighbor Sampling Size	8	Hidden Size	768
Dropout Rate	0.5	Attention Heads	12
Batch Size	256	Transformer Layers	12
Aggregation Depth (K)	3		
GNN Layers	2		

For baseline comparison, we include traditional machine learning models such as Support Vector Machine (SVM) and Light Gradient Boosting Machine (LightGBM), random-walk-based graph embedding methods including DeepWalk and node2vec, as well as a variety of Graph Neural Network (GNN) models—namely, Graph Convolutional Network (GCN), Graph Sample and Aggregation (GraphSAGE), Graph Attention Network (GAT), and Graph Isomorphism Network (GIN). We additionally consider two multimodal GNN-based baselines, ARTEMIS and ARTEMIX, which integrate visual, textual, and behavioral features for airdrop hunter detection. For methods based on random walks (DeepWalk and Node2Vec), the number of walks is set to 20, the walk length is 5, and the context size is 10 [3]. All methods are evaluated under the same data split and metric definitions for fair comparison. The selected baseline models are listed in TABLE III:

For Training Configuration, we perform five independent training runs to ensure robustness and stability of the model. In each run, the model is trained for a maximum of 100 epochs using the Adam optimizer with a learning rate of 1×10^{-4} and weight decay of 5×10^{-4}. Early stopping with a patience of 10 epochs is employed to prevent overfitting, based on validation loss improvement.

TABLE III
BASELINE COMPARISON MODELS

Model Category	Methods
Traditional ML	SVM [27], LightGBM [28]
Graph Embedding	DeepWalk [29], Node2Vec [30]
GNN Models	GCN [12], GraphSAGE [13], GAT [14], GIN [31]
Multimodal GNN	ARTEMIS [3], ARTEMIX [11]

The training set is sampled using a class-balanced WeightedRandomSampler to mitigate class imbalance. Each mini-batch contains 256 wallet nodes. During each epoch, the model parameters are updated via backpropagation using the binary cross-entropy loss with a positive class weighting scheme. Model performance is evaluated per epoch using precision, recall, and F1-score.

To maintain fairness and reproducibility, we apply the same sampling strategy, neighbor aggregation sizes ([8, 1, 1]), and evaluation metrics across all experiments. The best-performing model checkpoint (based on validation F1-score) is retained for final test set evaluation.

B. Evaluation Metrics and Methodology

Due to the highly imbalanced nature of the airdrop hunter detection task—where positive samples represent only a small fraction of the dataset—accuracy is not a reliable evaluation metric. A model can achieve high accuracy by simply predicting the majority class while failing to identify any true hunters.

To better assess model performance under such imbalance, we adopt three widely used metrics: **Precision**, **Recall**, and **F1-score**. Precision quantifies the proportion of correctly predicted hunter addresses among all predictions, helping reduce false positives. Recall measures the fraction of actual hunters correctly identified, ensuring comprehensive coverage. F1-score, the harmonic mean of Precision and Recall, provides a balanced summary of both.

These metrics collectively offer a more informative and robust evaluation framework, particularly in blockchain scenarios where both false positives (mislabeling legitimate users) and false negatives (overlooking hunters) can have practical implications.

C. Detection Results and Analysis

Table IV presents the detection performance of various baseline and proposed methods in terms of Precision, Recall, and F1-score. Our model—denoted as **ARTEMIS-PageRank**—extends the original ARTEMIS framework [3] by incorporating a PageRank-based global centrality feature. This enhancement yields near-perfect detection results across all metrics.

Introducing PageRank leads to significant performance gains, especially in Precision (from 0.820 to 0.988) and Recall (from 0.833 to 0.976). PageRank provides a global view of node importance within the transaction graph, which complements ARTEMIS's local behavior and multimodal features.

TABLE IV
AIRDROP HUNTER DETECTION PERFORMANCE COMPARISON

Method	Precision	Recall	F1-score
SVM [27]	0.744	0.544	0.629
LightGBM [28]	0.793	0.597	0.680
DeepWalk [29]	0.567	0.501	0.496
Node2Vec [30]	0.620	0.502	0.500
GCN [12]	0.648	0.896	0.752
GraphSAGE [13]	0.562	0.934	0.701
GAT [14]	0.464	0.873	0.579
GIN [31]	0.680	0.903	0.776
ARTEMIS [3]	0.820	0.833	0.826
ARTEMIX [11]	0.928	0.869	0.898
ARTEMIS-PageRank	**0.988**	**0.976**	**0.982**

This structural signal has proven effective in a range of graph mining tasks for capturing global centrality [32].

1) Complementing Local Sampling and Behavior Features: The original ARTEMIS framework emphasizes local path sampling and behavior-based node features. While effective in capturing micro-level signals like loop trades and abnormal trading patterns, such approaches may struggle to recognize globally significant but structurally dispersed airdrop hunters. This challenge is consistent with known limitations of message-passing GNNs, which are inherently constrained by their local receptive fields [33].

By incorporating PageRank, the model gains awareness of the macro structure of the network. It can better differentiate:

- **Superficially active users**: high-frequency behavior in local clusters but negligible global presence.
- **Truly influential users**: participants with wide-reaching interactions across multiple NFT projects or wallet communities.

2) Reducing False Positives and False Negatives: This structural augmentation helps reduce both false positives and false negatives. Low-PageRank users with suspicious local activity are less likely to be misclassified as legitimate, while globally connected, evasive hunters are more likely to be correctly recalled.

3) Multiscale Representation Learning: With the integration of PageRank, the model simultaneously benefits from:

- **Local path sampling** to detect loops and chains;
- **Multimodal semantics and behavioral features** to flag speculative activity;
- **Global structural positioning** to identify market-influential nodes.

D. Case Study

a) Case 1: Node 4636 — Passive Sink with Repeated Aggregation from Few Sources: Node 4636 was predicted as an airdrop hunter with high confidence (0.9862). It shows a completely passive behavior pattern—receiving 176 NFT transfers from 73 unique addresses while sending none (out-degree = 0). Its PageRank is extremely low (2.27×10^{-5}), indicating minimal influence in the overall transaction graph.

The edge weight analysis reveals significant aggregation from a subset of addresses. While 44 edges represent single

transfers, the remaining 29 edges correspond to repeated interactions, including one address that sent 26 transfers and others with 10, 11, and 22 transfers respectively. This implies that although the wallet interacts with many sources, **a small group of addresses contributes a disproportionately high volume** of transactions, potentially indicating a hub-and-spoke control structure.

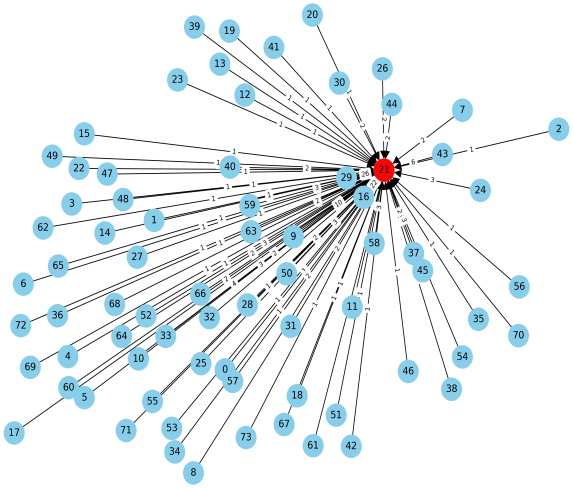

Fig. 2. Incoming transaction graph for Node 4636. Node 4636 is shown in red, edge labels denote NFT transfer counts.

Figure 2 illustrates a clear star-shaped structure, with Node 4636 at the center. The incoming edges are unidirectional, and edge weights highlight the coordination level. **To improve clarity, all node indices in the figure are remapped starting from 0, with Node 4636 represented as the red node labeled 21**. Such a pattern is consistent with "sink behavior," where a central wallet collects assets from disposable addresses that likely belong to the same controller.

The model effectively detects this address by combining degree asymmetry, repeated transaction patterns, and global isolation. The existence of several high-frequency inbound edges contributes strongly to the classification.

b) Case Study 2: Transaction Network among Verified Airdrop Hunters: To further explore behavioral collusion patterns, we construct a transaction graph among verified airdrop hunters by selecting only those nodes with true labels equal to 1. Unlike Case 1, which focuses on the behavior of a single wallet, this case highlights the collective trading structure among multiple confirmed airdrop hunters.

As illustrated in Figure 3, we extract the largest weakly connected component among hunter nodes and select a subset of 15 representative wallets based on their PageRank scores for clarity and interpretability. The edge weights represent the number of transactions between each pair, and the node color intensity encodes the PageRank score—a proxy for the structural importance of each wallet in the subgraph.

Several interesting patterns emerge:

- Dense and reciprocal connections exist among many hunters, indicating possible coordinated behavior. Some wallets (e.g., nodes 0 and 1) have dozens of incoming and outgoing edges, implying they may act as "relay hubs" or central orchestrators in hunter communities.
- All edge weights are 1, suggesting these interactions are likely minimal and intentional—potentially just enough to simulate engagement or meet eligibility thresholds.

This visualization underscores the value of structural analysis beyond individual behavior. Detecting densely connected clusters of known malicious actors can expose coordination rings that traditional address-level heuristics might miss.

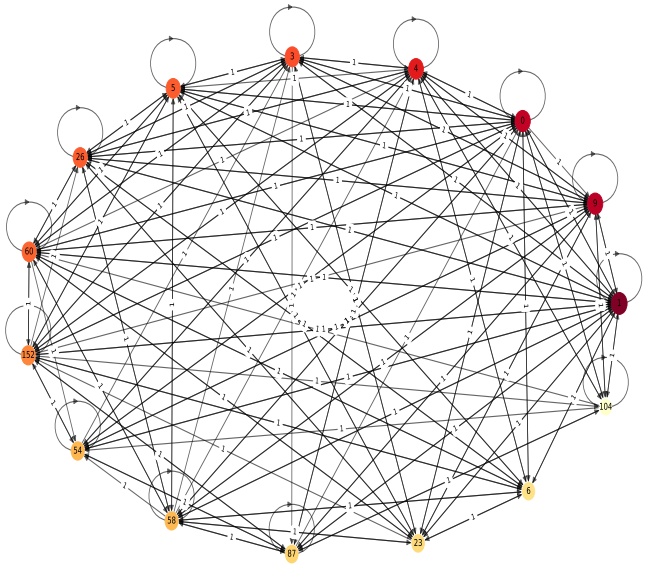

Fig. 3. Representative subgraph of the largest connected transaction network among verified airdrop hunters. The visualization includes 15 selected nodes based on their PageRank rankings. Edge directions indicate fund transfers, edge weights show transaction counts, and node colors reflect PageRank scores (darker = more influential).

E. Summary of Findings

The experimental results consistently demonstrate the effectiveness and interpretability of our proposed detection framework. Compared with traditional machine learning models, random-walk-based embeddings, and standard GNNs, our enhanced multimodal GNN—ARTEMIS-PageRank—achieves superior performance across all evaluation metrics. Notably, it surpasses the original ARTEMIS and ARTEMIX models by integrating global structural signals through PageRank centrality.

This improvement validates our hypothesis that incorporating global graph topology can complement local behavior patterns and multimodal semantics, enabling the model to better distinguish between legitimate users and coordinated airdrop hunters.

Beyond quantitative performance, our case studies further showcase the interpretability of the model. The first case reveals how node-level asymmetry and repeated aggregation

behaviors flag sink addresses, while the second exposes dense substructures among labeled hunters, suggesting collective manipulation.

Overall, the proposed approach not only achieves state-of-the-art performance but also provides actionable insights into the behavioral and structural signatures of airdrop hunting—laying a foundation for real-world deployment and deeper adversarial understanding.

V. Limitations and Future Directions

Despite the near-optimal performance of our improved ARTEMIS model on the Blur dataset, several limitations and future improvements remain:

- **Dataset Bias and Label Incompleteness:** The dataset only covers Blur marketplace transactions from October 2022 to April 2023. It may not generalize to other platforms (e.g., OpenSea, Rarible) or future conditions due to differences in trading volume, user behavior, and NFT types. Additionally, the reliance on agglomerative clustering followed by expert labeling introduces possible subjectivity and label noise [34], especially for ambiguous or hybrid wallets that do not conform clearly to the hunter or non-hunter categories.
- **Overfitting Risk and Class Imbalance:** The integration of multimodal, behavioral, and structural features increases the dimensionality of the input space, which may lead to overfitting, particularly in the presence of label imbalance. With only 4% of positive samples, the model might memorize patterns specific to training data, as evidenced by the high precision observed on the validation set. Additionally, our evaluation was based solely on a 90/10 random split, lacking an external test set, which limits our understanding of true generalization performance [35].
- **Adversarial Evolution:** Once the detection methods are publicly known, adversaries may modify their strategies to evade classification. Examples include the use of proxy wallets, splitting behavior across addresses, or injecting random low-frequency transactions to mimic benign patterns. Our current model does not explicitly address robustness under such adversarial settings, and future work should explore defense mechanisms, such as adversarial training or perturbation-based evaluations [36].
- **Real-Time Deployment Challenge:** PageRank, while effective in capturing global importance and flow characteristics, requires iterative computation over the entire graph structure, which is computationally expensive and unsuitable for real-time detection. In streaming or low-latency environments, such as live airdrop monitoring systems, more scalable approximations (e.g., personalized PageRank, Monte Carlo sampling) or incrementally updatable variants [37] are necessary to maintain responsiveness without compromising accuracy.
- **Structural Feature Redundancy:** In a controlled ablation experiment, we replaced PageRank with normalized in-degree and out-degree features and still achieved near-perfect results: **Precision = 0.999**, **Recall = 1.000**, and **F1-score = 0.999**. This underscores the strong predictive power of degree-based cues and suggests that labels may be closely tied to simple topological statistics. While effective, such reliance poses risks of shortcut learning, where the model depends on shallow graph patterns rather than deeper behavioral semantics. This aligns with known structural biases in GNNs, where high-degree nodes or homophilic neighborhoods are overly favored. As airdrop hunters can manipulate their degree distribution—e.g., through sparse or synthetic transactions—to evade detection, such features may become unreliable. To mitigate this, future work could explore contrastive learning to differentiate nodes by structural and semantic traits, or adopt structure-aware regularization to reduce overfitting to degree-based patterns. Additionally, disentangled GNNs that separate structural and semantic signals show promise for robust, interpretable learning. A systematic evaluation under adversarial perturbations can further assess model resilience. Recent advances [38] highlight the need for sustainable and decentralized learning architectures under adversarial and resource-constrained settings, offering valuable directions for strengthening Web3 security systems.

In summary, while our approach achieves strong performance, addressing the above limitations is crucial for generalizability, robustness, and practical deployment.

VI. Conclusion

In this paper, we addressed the growing challenge of detecting airdrop hunters in NFT marketplaces by enhancing the ARTEMIS framework with global structural signals. Leveraging a real-world dataset from the Blur NFT platform, we constructed a multimodal graph that incorporates textual, visual, behavioral, and topological features. Our proposed model introduces PageRank-based centrality as an additional input feature, enabling more precise identification of strategically dispersed airdrop hunter behaviors. Extensive experiments demonstrated that the improved model significantly outperforms both traditional machine learning baselines and existing GNN variants, achieving near-perfect performance across precision, recall, and F1-score. In particular, feature substitution experiments—replacing PageRank with in-degree and out-degree—further revealed the strong predictive power of structural connectivity. While this supports the importance of graph features, it also highlights the risk of shortcut learning and potential structural overfitting, pointing to the need for robust feature disentanglement and validation strategies. While our approach offers strong detection capability, it also highlights open challenges regarding label quality, cross-platform generalizability, and real-time scalability. Future work will focus on evaluating the method across multiple NFT ecosystems (e.g., OpenSea, Rarible), incorporating adversarial learning techniques to improve robustness, and exploring lightweight structural alternatives to PageRank for online deployment.

Overall, our study provides both a practical solution and theoretical insight into the detection of airdrop hunters, contributing to the broader effort of maintaining fairness and trust within Web3 ecosystems. In future applications, this system can be adapted to decentralized compliance monitoring, Sybil-resistant NFT auctions, and token distribution fairness auditing in real-world Web3 deployments.

ACKNOWLEDGMENT

This work was supported in part by the Key Field Projects of Ordinary Universities in Guangdotng Province (No. 2025ZDZX3050); and in part by the National Natural Science Foundation of China (No. 62576213); and in part by Engineering Technology Research Center for Ordinary Universities in Guangdong Province (No. 2024GCZX005); in part by Guangdong-Hong Kong-Macao Joint Laboratory for Emotional Intelligence and Pervasive Computing, Artificial Intelligence Research Institute, Shenzhen MSU-BIT University.

REFERENCES

[1] H. Chen, H. Duan, M. Abdallah, Y. Zhu, Y. Wen, A. E. Saddik, and W. Cai, "Web3 metaverse: State-of-the-art and vision," *ACM Transactions on Multimedia Computing, Communications and Applications*, vol. 20, no. 4, pp. 1–42, 2023.

[2] Y. Peng and H. Duan, "Blockchain-based incentive mechanism in internet of things: Survey and vision," in *2024 IEEE International Conference on Smart Internet of Things (SmartIoT)*. IEEE, 2024, pp. 555–562.

[3] C. Zhou, H. Chen, H. Wu, J. Zhang, and W. Cai, "Artemis: Detecting airdrop hunters in nft markets with a graph learning system," in *Proceedings of the ACM Web Conference 2024 (WWW '24)*. ACM, 2024.

[4] H. Duan, J. Li, S. Fan, Z. Lin, X. Wu, and W. Cai, "Metaverse for social good: A university campus prototype," in *Proceedings of the 29th ACM international conference on multimedia*, 2021, pp. 153–161.

[5] L. Page, S. Brin, R. Motwani, and T. Winograd, "The pagerank citation ranking: Bringing order to the web," Stanford InfoLab, Tech. Rep. 1999-66, 1999. [Online]. Available: http://ilpubs.stanford.edu:8090/422/

[6] H. Wang, X. Liu, and L. Wu, "Structure-aware fraud detection in transaction networks," *ACM Transactions on Knowledge Discovery from Data (TKDD)*, 2022.

[7] F. Zhang, H. He, and Y. Tang, "Token airdrops in web3: Motivation, risks, and defense," *arXiv preprint arXiv:2212.08955*, 2022.

[8] J. Liu, Z. Wang, and S. Jiang, "Sybil attacks in blockchain: A survey," *IEEE Access*, vol. 10, pp. 120 518–120 536, 2022.

[9] J. R. Douceur, "The sybil attack," in *International workshop on peer-to-peer systems*. Springer, 2002, pp. 251–260.

[10] S. Fan, T. Min, X. Wu, and W. Cai, "Altruistic and profit oriented: Making sense of roles in web3 community from airdrop perspective," in *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems (CHI '23)*. Hamburg, Germany: ACM, 2023, pp. 1–16.

[11] Y. Qin, T. Ma, H. Chen, and H. Duan, "Artemix: A community-boosting-based framework for airdrop hunter detection in the web3 community," *Blockchain*, vol. 2, p. 0010, 2024.

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

[13] W. L. Hamilton, R. Ying, and J. Leskovec, "Inductive representation learning on large graphs," in *Advances in Neural Information Processing Systems*, vol. 30, 2017.

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

[15] Z. Feng, L. Zhao, X. Chen, and L. Wu, "Graph neural networks for blockchain analysis: Methods and applications," *IEEE Transactions on Computational Social Systems*, vol. 9, no. 6, pp. 1952–1965, 2022.

[16] Z. Liu, C. Zhang, Y. Wang, X. Zhang, Z. Liu, and J. Li, "A survey on graph neural networks for fraud detection," *ACM Computing Surveys*, vol. 55, no. 8, pp. 1–38, 2023.

[17] W. Chen, L. Zhang, and J. Wu, "Detecting ponzi schemes on ethereum: Towards accurate financial crime analytics," *IEEE Transactions on Information Forensics and Security*, vol. 15, pp. 1226–1240, 2020.

[18] Z. Liu, Y. Yuan, Q. Liu, and Y. Wang, "Walletrank: Reputation management for ethereum accounts using transaction graphs," in *Proceedings of the 30th ACM International Conference on Information & Knowledge Management*, 2021.

[19] K. Xu, Z. Liu, Y. Chen, and K. Ren, "Decentralized identity meets sybil resistance: A survey and future directions," in *Proceedings of the IEEE International Conference on Blockchain and Cryptocurrency (ICBC)*. IEEE, 2023, pp. 1–10.

[20] T. Li, Y. Zhang, and C. Wu, "Detecting wash trading loops in nft markets using path aggregation," *arXiv preprint arXiv:2303.07856*, 2023.

[21] OpenSea Developer Docs, "NFT Metadata Standards," https://docs.opensea.io/docs/metadata-standards, accessed: 2025-06-28.

[22] J. Devlin, M.-W. Chang, K. Lee, and K. Toutanova, "Bert: Pre-training of deep bidirectional transformers for language understanding," in *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics (NAACL)*. ACL, 2019.

[23] A. Dosovitskiy, L. Beyer, A. Kolesnikov, D. Weissenborn, X. Zhai, T. Unterthiner, M. Dehghani, M. Minderer, G. Heigold, S. Gelly, J. Uszkoreit, and N. Houlsby, "An image is worth 16x16 words: Transformers for image recognition at scale," in *Proceedings of the International Conference on Learning Representations (ICLR)*. OpenReview, 2021.

[24] S. J. Das and A. Kumar, "Detecting anomalies in blockchain transactions using benford's law," *IEEE Transactions on Blockchain*, vol. 1, no. 2, pp. 104–115, 2021.

[25] X. Zhao and Y. Liu, "Quantifying wallet activeness in ethereum: Contract calls and gas patterns," *IEEE Access*, vol. 10, pp. 12 345–12 358, 2022.

[26] M. Yang, E. C. Ngai, X. Hu, B. Hu, J. Liu, E. Gelenbe, and V. C. Leung, "Digital phenotyping and feature extraction on smartphone data for depression detection," *Proceedings of the IEEE*, 2025.

[27] C. Cortes and V. Vapnik, "Support-vector networks," *Machine learning*, vol. 20, no. 3, pp. 273–297, 1995.

[28] G. Ke, Q. Meng, T. Finley, T. Wang, W. Chen, W. Ma, Q. Ye, and T.-Y. Liu, "Lightgbm: A highly efficient gradient boosting decision tree," in *Advances in neural information processing systems (NeurIPS)*, 2017, pp. 3146–3154.

[29] B. Perozzi, R. Al-Rfou, and S. Skiena, "Deepwalk: Online learning of social representations," in *Proceedings of the 20th ACM SIGKDD*. ACM, 2014, pp. 701–710.

[30] A. Grover and J. Leskovec, "node2vec: Scalable feature learning for networks," in *Proceedings of the 22nd ACM SIGKDD*. ACM, 2016, pp. 855–864.

[31] K. Xu, W. Hu, J. Leskovec, and S. Jegelka, "How powerful are graph neural networks?" in *International Conference on Learning Representations (ICLR)*, 2019.

[32] R. A. Rossi and N. K. Ahmed, "Proximity measures for graph mining: A survey," *ACM Computing Surveys*, vol. 52, no. 4, pp. 1–35, 2020.

[33] U. Alon and E. Yahav, "On the bottleneck of graph neural networks and its practical implications," in *International Conference on Learning Representations (ICLR)*, 2021.

[34] X. Fan, X. Li, Y. Zhang, H. Zhao, and J. Wang, "Altruistic or opportunistic? detecting airdrop hunters in web3 markets," *Proceedings of the ACM Web Conference 2023*, 2023.

[35] Y. Zhang, Y. Xu, Z. He, K. Chen, and K. Ren, "Tokenscope: Detecting fraudulent smart contracts on ethereum via behavioral graph learning," *IEEE Transactions on Dependable and Secure Computing*, 2022.

[36] Q. He, T. Wu, M. Liu, W. Chen, and J. Li, "Adversarial attacks on node embeddings via graph poisoning," in *Proceedings of the 30th ACM International Conference on Information & Knowledge Management*, 2021.

[37] B. Bahmani, A. Chowdhury, and A. Goel, "Fast incremental and personalized pagerank," in *Proceedings of the VLDB Endowment*, vol. 4, no. 3, 2010, pp. 173–184.

[38] H. Duan, T. Ma, Y. Qin, R. Zeng, W. Cai, V. C. Leung, and X. Hu, "Derelayl: Sustainable decentralized relay learning," *IEEE Transactions on Mobile Computing*, 2025.