A Multi-Feature Semantic Fusion and Bipartite Graph-Based Risk Identification Approach for Project Participation

Yue Wang[1,2], Yujie Hu[1,2], Wenjing Chang[1,2] and Jianjun Yu[1,*]
[1]*Computer Network Information Center, CAS, CAS Informatization Plaza No.2 Dong Sheng Nan Lu, Haidian District, Beijing 100083, China*
[2]*University of Chinese Academy of Sciences, No.19A Yuquan Road, Shijingshan District, Beijing 100049, China*

Keywords: Risk Identification, Graph Neural Networks, Bipartite Graphs, Multi-Semantic Feature Fusion.

Abstract: In the complex landscape of project management, ensuring the authenticity of participant involvement is paramount to achieving fairness, enforceability, and desired outcomes. Addressing the challenges posed by the heterogeneous nature of graphs, the underutilization of rich attribute information, and the scarcity of anomaly labels, we propose a Project Participation Authenticity Risk Identification Graph Neural Network (PARI-GNN), a novel architecture leveraging graph-based anomaly detection techniques to assess authenticity risks in project participation. PARI-GNN include a novel framework for risk identification using heterogeneous graphs. This method transforms heterogeneous graphs into bipartite graphs and combines multi-feature semantic fusion techniques with bipartite graph structures, providing a robust solution for identifying inauthentic participation. We evaluate our proposed model using real-world data. The experimental outcomes affirm the superior performance of PARI-GNN in accurately discerning authenticity risks, demonstrating the efficacy and competitive advantage of the proposed framework over a variety of state-of-the-art methodologies.

1 INTRODUCTION

The current landscape of project management involves a multitude of participants, each with varying degrees of involvement and contribution. Ensuring the authenticity of participant engagement is crucial for the fairness of the project application, the enforceability of the project and the achievement of desired outcomes. Unfortunately, similar to financial fraud, projects also face fraudulent practices such as false participation, exaggerated participation levels, or falsified contributions. These deceptive acts can lead to misallocation of resources, project delays, or even failures, thereby inflicting significant costs on organizations and stakeholders involved.

The challenge lies in the accurate identification and assessment of these contributions, a task that is increasingly difficult with the growing scale and complexity of projects. To combat these challenges, it is essential to get robust methods for identifying and assessing the risks of inauthentic participation in projects. But traditional methods of verification, relying on surface-level indicators or simple data statistics, are often manual, time-consuming, and prone to human error, which may not effectively detect or adapt to sophisticated or novel schemes. Hence, there is a growing need for advanced techniques that can learn from complex patterns and interactions inherent in project data.

Graph-based anomaly detection presents a promising solution to address these issues. Graphs naturally encapsulate the relational information among participants and project elements, providing a rich framework for identifying irregularities in participant behaviors and interactions.

However, several obstacles need addressing. Firstly, heterogeneous graphs, comprising diverse node and edge types, pose a challenge due to their complexity, demanding sophisticated mechanisms for effective risk identification. Secondly, traditional graph learning methods often overlook rich attribute information within nodes or edges, requiring advanced representation learning techniques to integrate this valuable context. Lastly, the scarcity of

[*] Corresponding author

Wang, Y., Hu, Y., Chang, W. and Yu, J.
A Multi-Feature Semantic Fusion and Bipartite Graph-Based Risk Identification Approach for Project Participation.
DOI: 10.5220/0012742500003690
In *Proceedings of the 26th International Conference on Enterprise Information Systems (ICEIS 2024) - Volume 1*, pages 907-914
ISBN: 978-989-758-692-7; ISSN: 2184-4992

anomaly labels in graph data complicates the training of supervised models, necessitating alternative approaches for effective anomaly detection.

In response to these challenges, we propose a novel architecture that leverages graph-based anomaly detection techniques to assess authenticity risks in project participation. Our model, which we refer to as the Project participation Authenticity Risk Identification Graph Neural Network (PARI-GNN), is designed to integrate multi-feature data and heterogeneous graph structures, effectively bridge the gap in anomaly label scarcity, and provide interpretable results for decision-makers.

The main contributions of this paper are as follow:

- We present a novel framework for risk identification that employs heterogeneous graphs, seamlessly integrating knowledge graph construction with subgraph extraction and transformation techniques. Subsequently, it transforms these heterogeneous graphs into bipartite graphs, leveraging bipartite graph-based methodologies to efficiently identify risks.
- We propose a multi-feature semantic fusion bipartite graph risk identification model. Leveraging multi-feature semantic fusion techniques, it integrates multi-feature data by encoding bipartite graphs containing rich node and edge features. It conducts structural and feature decoding independently, applying reconstruction loss and scoring functions to identify risks.
- We evaluate our proposed model using real-world data. The experimental results demonstrate the effectiveness of the proposed framework and its superiority compared to a variety of state-of-the-art methods.

The remainder of this paper is organized as follows. Section 2 formally introduces the problem definition. Section 3 reviews related work in the field. In Section 4, we detail the proposed model for identifying the authenticity risks associated with personnel involvement in projects. Experimental evaluation on real-world datasets is presented in Section 5. The paper is concluded in Section 6 with a summary of our findings.

2 RELATED WORK

Graph Anomaly Detection. Risk identification is fundamentally a form of anomaly detection. As networks increasingly model various complex systems, research on graph-based anomaly detection has garnered widespread attention. Classical graph anomaly detection models are constrained by their shallow learning mechanisms, which limit their capacity to discern complex interaction patterns within graphs (Li et al., 2017; Ding et al., 2019). Models for graph anomaly detection based on deep neural network architectures have seen rapid development due to their exceptional performance. Graph Convolutional Networks (GCNs) (Kipf et al., 2016) provide a scalable semi-supervised learning method for graph-structured data, while GraphSAGE (Hamilton et al., 2017) extends the GCN architecture to efficiently generate node embeddings using node feature information, such as textual attributes. Graph Attention Networks (GATs) (Veličković et al., 2017) address the shortcomings of previous methods based on graph convolution or its approximations using masked self-attention layers. DOMINANT (Ding et al., 2019) offers an architecture akin to autoencoders for graph anomaly detection, and AnomalyDAE (Fan et al., 2020) expands this architecture using dual autoencoders. GAD-NR (Roy et al., 2023) is a novel variant of GAE (M. Tang et al., 2022), integrating neighborhood reconstruction for graph anomaly detection. AdONE (Bandyopadhyay et al., 2020) learns by differentiating between network's structural and attribute-based embeddings, minimizing the impact of outliers in a coupled manner. FIW-GNN (Yan et al., 2023) proposes a feature-importance weighted graph neural network as a solution for credit card fraud detection. However, most of these models are tailored for anomaly detection in homogenous graphs, what we need to deal with is risk identification for heterogeneous graphs.

Heterogeneous Graph Anomaly Detection. Paper (Wang et al., 2022) provides a comprehensive review of the latest developments in heterogeneous graph embedding methods and techniques, demonstrating the success of heterogeneous graph embedding technologies in addressing real-world application issues with broader impacts. Paper (Bing et al., 2023) systematically summarizes and analyzes the existing Heterogeneous Graph Neural Networks (HGNNs), categorizing them based on their neural network architectures. GEM (Liu et al., 2018) introduces a heterogeneous graph neural network approach for detecting malicious accounts, proposing an attention mechanism to learn the significance of different types of nodes and employing a sum operator to model the aggregation patterns for each node type. HAN (Wang et al., 2019) presents a novel heterogeneous graph neural network based on hierarchical attention, capable of capturing the complex structures and rich semantics underlying heterogeneous graphs. Paper (Zhang et al., 2019) exploits the attribute heterogeneous information network to identify key players in underground forums. However, most of these models are tailored for anomaly detection in nodes, with relatively limited research on heterogeneous graph edge anomaly detection (Ma et al., 2021).

3 PROBLEM DEFINITION

In this work, we address the challenge of assessing authenticity risks in project participation using graph-based anomaly detection technology. Our objective is to leverage the structure of bipartite graphs to represent complex relationships and interactions within projects, facilitating the identification of anomalies that signify such risks.

Definition 1: *Personnel-Project Bipartite Graph.* We define *B* to represent the participation relationships between personnel and projects and *A* represent the adjacency matrix of *B*.

$$B = (U, V, E_B),\ A_{ij}=w_{ij} \tag{1}$$

Where *U* denotes the set of personnel nodes and *V* represents the set of project nodes, *U* and *V* are disjoint sets of nodes such that each edge $e \in E_B$ connects a node from set *U* to a node from set *V*, and w_{ij} denotes the specific amount of person-years input by individual u_i to project v_j.

Definition 2: Project Participation Risk Identification Task. Building upon the definition of *B*, the task of project participation risk identification can be further specified as identifying anomalous patterns or structures within the graph. These anomalies may indicate authenticity risks associated with individual participation in projects.

The spectrum of authenticity risks in project participation is diverse, encompassing scenarios such as an individual's engagement in an inordinately large number of projects simultaneously, typified by exceeding the threshold of ten projects. A substantial variance in the content and disciplinary trajectories of the projects that an individual contributes to may also signal potential risks. Equally problematic are situations where individuals are recorded as active participants despite having departed from the team. In addition, atypical patterns in the interrelations among project collaboration team members could be symptomatic of deeper issues. These factors collectively contribute to the complexity of assessing authenticity risks. Therefore, we have defined these risk rules using Equations 2.1 to 2.4.

- Personnel Participation Function.

$$f_{participation}(u) = |E(u)| \tag{2.1}$$

Where *E(u)* is the set of edges associated with individual $u \in U$, indicating the number of projects a person participates in.

- Project Diversity Metric.

$$D(u) = \sum_{p_i, p_j \in P(u), i \neq j} d(p_i, p_j) \tag{2.2}$$

Where *d(p* —

Where $d(p_i, p_j)$ is the dissimilarity function between projects p_i and p_j, which can be defined based on project content or disciplinary direction.

- Departure Indicator Function.

$$I_{departed}(u) = \begin{cases} 1, & if\ u\ has\ departed \\ 0, & otherwise \end{cases} \tag{2.3}$$

Where *u* is a personnel node.

- Cooperation Anomaly Metric.

$$A(u_i, u_j) = f_{anomaly\_score}(u_i, u_j) \tag{2.4}$$

Where *u* is a personnel node.

Following, we will introduce our proposed risk identification framework, which extracts subgraphs from knowledge graphs and transforms them into bipartite graphs, synthesizing structural complexity and multidimensional feature representations to assess the authenticity risk in project participation.

4 METHODOLOGY

In this section, we present the proposed framework of PARI-GNN in detail, which is designed to identify authenticity risks in project participation. Our approach is grounded in the construction and analysis of a bipartite graph, derived from the intricate relationships between individuals and projects. This bipartite graph serves as the foundation for applying our advanced deep learning framework, which seamlessly integrates multi-feature semantic fusion and graph-based analysis to uncover and quantify potential risks. The architecture of the deep model is illustrated in Figure 1.

4.1 Generating Bipartite Graphs

A knowledge graph can illustrate a panoramic view of complex relationships among entities such as researchers, institutions, funding sources, and research fields. It holds high value for analysing and discovering potential project risks. The construction

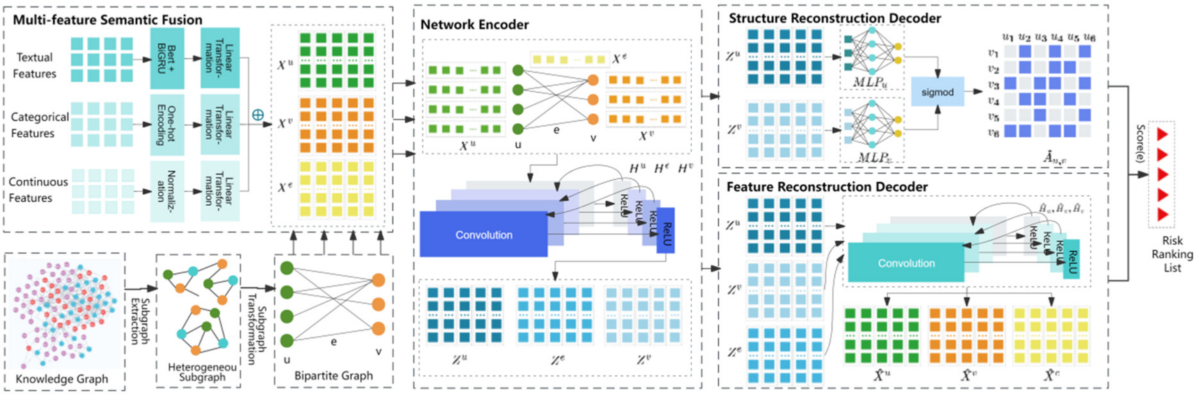

Figure 1: The overall framework of our proposed PARI-GNN which is designed to identify authenticity risks in project participation. It encompasses the process of constructing a knowledge graph from real-world project data, extracting heterogeneous subgraphs, and transforming them into bipartite graphs. Subsequently, employing multi-feature semantic fusion techniques facilitates semantic enhancement of attributes. Further, based on the bipartite graph, an encoder-decoder framework is constructed, utilizing reconstruction error and scoring functions to output a list of project participation risk assessments.

of the project knowledge graph in this paper aims to integrate scattered and disparate project data. We first identify project-related entities from unstructured project abstracts and task documents, integrate entities and relationships from structured and unstructured project datasets to construct triples, and obtain an ontology model pre-built for the project dataset. Entity linking based on triple knowledge is then performed to complete the construction of the project knowledge graph.

Subsequently, we extract heterogeneous subgraphs from the constructed knowledge graph. These subgraphs preserve the complex relationships and attributes from the original knowledge graph and focused on the task of identifying authenticity risks in project participation. The extraction process utilizes algorithms to traverse the knowledge graph, collecting nodes and edges that meet our criteria.

The next step is to transform the heterogeneous subgraphs into bipartite graphs. The edges in this bipartite graph only connect nodes from these two different sets, representing personnel participation in projects. To accomplish this transformation, we employ a mapping strategy, where the associations of other entities in the subgraph with individuals or projects are respectively transformed into attributes of individual or project entities and allocated to attribute sets.

Our method simplifies project participation into a bipartite graph, enabling deep learning and graph-based anomaly detection to effectively identify authenticity risks by analysing interactions and detecting risky patterns. This approach utilizes knowledge graphs to extract heterogeneous subgraphs, offering a clear framework for detailed risk analysis in project participation networks.

4.2 Multi-Feature Semantic Fusion

The multi-feature semantic fusion method employed in this study aims to enhance the semantic understanding of node and edge attributes in the model. Through this method, the model integrates text, discrete, and continuous type attribute data into a unified semantic space, thereby enhancing the model's understanding and predictive performance.

We utilize BERT to embed text data and extract textual embeddings. Then, we employ BiGRU networks to capture bidirectional information in the text data, enhancing semantic understanding and generating deep semantic features, denoted as F_t. Discrete attributes are encoded using one-hot encoding and transformed into low-dimensional dense vectors, producing categorical features represented as F_c. Continuous attributes undergo standardization to ensure consistency, resulting in numerical features denoted as F_n.

The obtained feature vectors are concatenated and then fed into three fully connected layers for feature fusion and dimensionality reduction, resulting in the fused feature vector F_{fusion}.

The formula for this feature fusion process is as follows.

$$F_{fusion} = ReLU(MLP(F_t \oplus F_c \oplus F_n)) \tag{3}$$

For the bipartite graph consisting of personnel and project nodes along with their edges, after undergoing multi-feature semantic fusion operations, feature vectors X^u, X^v, and X^e are respectively generated for nodes and edges.

4.3 Network Encoder

In the model architecture proposed for evaluating the authenticity risk of project participation, the network encoder plays a crucial role. This component is designed to handle the intricate interactions between personnel and project represented by a bipartite graph, capturing the structural and relational information embedded in the graph through convolutional operations on the graph structure.

The encoder takes the node and edge attribute graphs of personnel participating in projects as input, generating latent representations of nodes and edges. The latent representations of nodes in the network require encoding of features from all neighboring (including themselves) nodes, node structures, and edge features connecting to K-hop nodes. The encoder consists of m layers of convolutional layers, and the encoding process is as follows:

- Initial embedding: The input of the first layer is initialized with the original input features of nodes and edges.

$$\begin{gathered}\{H^u\}^{(0)} = X^u \\ \{H^v\}^{(0)} = X^v \\ \{H^e\}^{(0)} = X^e\end{gathered} \tag{4}$$

- Neighborhood aggregation: For each node, the encoder aggregates features from its direct neighbors, with subsequent layers taking the output of the previous layer as input. This step is crucial for capturing the dynamic relationships between individuals and projects.
- Feature transformation: The aggregated features are transformed through neural network layers to enable the model to learn complex patterns and relationships in the data.
- Latent representation output: The output of the encoder is the latent representation of each node and edge, encapsulating its attributes and its relationship context within the project participation network.

$$\begin{gathered}Z^u = \{H^u\}^{(m)} \\ Z^v = \{H^v\}^{(m)} \\ Z^e = \{H^e\}^{(m)}\end{gathered} \tag{5}$$

Where m denotes the number of convolutional layers.

4.4 Structure Reconstruction Decoder

The input to the structural reconstruction decoder comprises Z_u and Z_v, which are the latent representations of personnel nodes and project nodes, generated by the network encoder from the bipartite graph of personnel participating in projects. Subsequently, separate multilayer perceptrons MLP_u and MLP_v are designed for Z_u and Z_v, respectively. They are employed to process the latent features, capture the nonlinear relationships between features, and transform them into a space suitable for reconstructing the adjacency matrix.

The outputs of MLP_u and MLP_v are combined in pairs to form a matrix, representing the logical potential edges between the two sets of nodes. The element values of the matrix are compressed into the range of (0, 1) by the sigmoid function, generating the reconstructed adjacency matrix.

$$\begin{aligned}&P(\hat{A}_{i,j}) \\ &= Sigmod(MLP_u(Z_i^u) \\ &\cdot MLP_v(Z_j^v)^T) \quad u\epsilon U, v\epsilon V\end{aligned} \tag{6}$$

4.5 Feature Reconstruction Decoder

The Feature Reconstruction Decoder receives the latent representations of nodes and edges as input, denoted as Z^u, Z^v, and Z^e, which are generated by the network encoder. The objective of the Feature Reconstruction Decoder is to reconstruct the features of nodes and edges from these latent representations. The latent feature matrices undergo convolutional operations to capture the local connectivity patterns among the features.

The output of the convolution is a set of feature maps, which are subsequently passed through the nonlinear activation function *ReLU*. This transforms the feature maps into activated features, namely $\hat{H}^u$, $\hat{H}^v$ *and* $\hat{H}^e$, which encapsulate both the original latent features and the local structures learned through the convolutional filters.

Finally, these activated features are mapped back to the reconstructed feature space, thereby completing the feature decoding process.

$$\begin{aligned}&\hat{X}^e \\ &= f_{ReLU}\big(Z^e \\ &\parallel Agg(Z^u, Z^v \mid \forall\,(u,v) \in E_B)\big)\end{aligned} \tag{7}$$

4.6 Reconstruction Loss and Risk Score

In our proposed framework designed to predict authenticity risks in project participation, we have integrated a reconstruction loss function with a scoring function. The reconstruction loss function measures the accuracy of the graph's overall reconstruction, with a focus on the edges that represent project participation. The scoring function derives anomaly scores from the magnitude of reconstruction loss for each edge, identifying anomalous participation patterns that may harbor authenticity risks.

$$L_{recon} = (1-\alpha)\,\cdot\,MSE(X^e,\hat{X}^e) + \alpha\,\cdot\,BCE\big(A,\hat{A}\big) \quad (8)$$

where α is a controlling parameter.

The construction of the reconstruction loss function L_{recon} aims to evaluate the quality of the reconstructed graph by assessing the fidelity of the reconstructed features $\hat{X}^e$ and adjacency matrix $\hat{A}$ against their original counterparts X^e and A. This function employs a weighted combination of mean squared error and binary cross-entropy loss. The mean squared error assesses the average squared difference between the estimated and actual values for continuous data, while the binary cross-entropy loss quantifies the distance between probability distributions for binary classification tasks.

$$Score\big(e_{i,j}\big) = (1-\alpha)\,\cdot\,MSE(x^e_{i,j},\hat{x}^e_{i,j}) + \alpha\,\cdot\,BCE\big(e_{i,j}\big) \quad (9)$$

The anomaly scoring function $Score\big(e_{i,j}\big)$ assigns a numerical value to each edge within the graph, reflecting the level of anomaly based on the reconstruction loss. It uses the same loss components to quantify the degree of deviation of each edge's reconstructed features $\hat{x}^e_{i,j}$ and presence $e_{i,j}$ from their expected patterns. High scores indicate significant anomalies and potential authenticity risks within the context of project participation.

5 EXPERIMENTS

To evaluate our approach, we applied PARI-GNN to perform authenticity risk identification of personnel participation in projects using data from a large-scale project management system sourced from the real world.

5.1 Datasets

In our study, we initially utilized knowledge graph technology to construct a comprehensive knowledge graph with 57,162 nodes and 1,220,522 edges, featuring entities like organizations, personnel, projects, disciplines, and sources of tasks. This graph was refined into a bipartite graph focusing on personnel and projects, where other entities were redefined as node attributes. To address the lack of real-world risk data, we enriched the dataset by injecting anomalies into the bipartite graph using techniques from the literature, notably referencing GraphBEAN (Fathony et al., 2023) for injecting a mix of structural and feature anomalies. Following the methodology for anomaly injection from Paper (J. Tang, et al., 2022), we assigned degrees of anomaly and labeled all affected edges and nodes as anomalous. The resulting dataset for our experiments contains 1,003 personnel nodes, 2,304 project nodes, and 14,819 edges, with details on feature dimensionality, number, and fraction of anomalous nodes outlined in Table 1.

Table 1: Details of the dataset with injected anomalies.

nodes/edges	number	feat	anomalies	fraction
personnel	1003	14	47	4.69%
project	2304	843	89	3.86%
edge	14819	4	1076	7.26%

5.2 Experimental Settings

In this section, we present detailed experimental settings, including baseline methods for comparison, evaluation metrics and parameter setup.

Comparison Methods. We compare the proposed PARI-GNN with the following popular anomaly detection methods: AnomalyDAE (Fan et al., 2020), DOMINANT (Ding et al., 2019), Adone (Bandyopadhyay et al., 2020).

Evaluation Metrics. In the experiments, two evaluation metrics are employed to measure the performance of different models:

ROC-AUC (Davis & Goadrich, 2006). Receiver Operating Characteristic (ROC) is a widely used evaluation metric in anomaly detection methods (Peng et al., 2018). The closer the value is to 1, the higher the quality of the method.

PR-AUC (Davis & Goadrich, 2006). We also utilize the area under the Precision-Recall (PR) curve as an evaluation metric. Compared to the ROC curve, it provides richer information about algorithm performance in the case of imbalanced datasets.

Parameter Setup. In the implementation of our PARI-GNN model, we meticulously configured the architecture and training parameters to optimize performance for the task of personnel involvement authenticity verification. Some key experimental parameter settings are shown in Table 2.

Table 2: Parameter setup of PARI-GNN.

Parameter	Value
Epochs	100
Learning Rate	0.001
Hidden Channels	32
Edge Prediction Latent	32

5.3 Experimental Results

In this study, we present a comprehensive evaluation of PARI-GNN against three baseline models, namely AnomalyDAE, DOMINANT, and Adone, using ROC-AUC and PR-AUC as performance metrics. We present the experimental results in Table 3.

Table 3: Performance of different methods w.r.t. PR AUC and ROC AUC.

Models	PR-AUC	ROC-AUC
Adone	0.4847	0.8506
DOMINANT	0.7746	0.9249
AnomalyDAE	0.8648	0.9511
PARI-GNN	0.9400	0.9777

The analysis of ROC curve outcomes reveals that our model, PARI-GNN, surpasses baseline models with an AUC of 0.9777, highlighting its exceptional proficiency in differentiating between positive and negative classes at various thresholds. This performance indicates a remarkable equilibrium between True Positive Rate (TPR) and False Positive Rate (FPR), underscoring PARI-GNN's effectiveness in minimizing false positives while maximizing true positives, as illustrated in Figure 2. Furthermore, PARI-GNN's performance on the PR curve, with an AUC of 0.9400 as depicted in Figure 3, confirms its capability to maintain high precision across diverse recall levels, which is crucial in contexts where avoiding false positives is paramount.

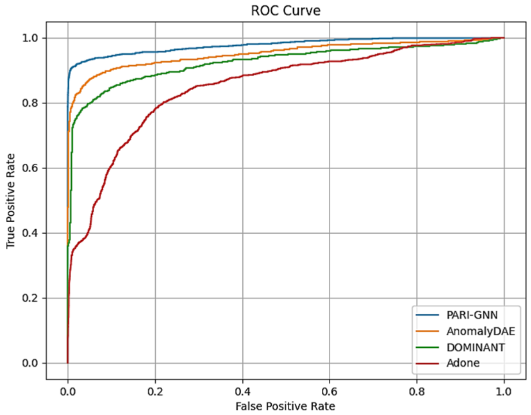

Figure 2: The ROC curve of the proposed model compared to the baseline models.

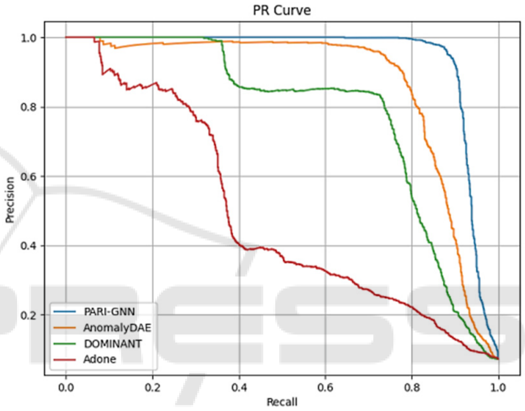

Figure 3: The PR curve of the proposed model compared to the baseline models.

In our comparative analysis, the AnomalyDAE model ranked second with a PR AUC of 0.8648 and a ROC AUC of 0.9511, showing strong discriminative capability but falling short of PARI-GNN, particularly due to a higher false positive rate affecting precision at elevated recall levels.

The DOMINANT model, achieving a PR AUC of 0.7746 and a ROC AUC of 0.9249, demonstrated moderate efficacy. Its performance is hampered by a drop in precision at higher recall levels, suggesting challenges in maintaining prediction confidence amidst class overlap or noise.

The Adone model records the lowest AUC scores of 0.4847 for PR and 0.8506 for ROC, reflecting substantial room for improvement. The marked drop in precision after a modest recall level in the PR curve indicates a tendency to misclassify negative samples as positive, which can be detrimental in precision-critical tasks.

6 CONCLUSIONS

This study has presented PARI-GNN, an innovative model designed to tackle the challenges of identifying authenticity risks in project participation. Through a comprehensive evaluation against baseline models AnomalyDAE, DOMINANT, and Adone, PARI-GNN demonstrated superior performance, particularly in distinguishing between positive and negative classes with high accuracy and precision. The model's success, as evidenced by its outstanding AUC scores on both ROC and PR curves, confirms the effectiveness of employing a graph-based anomaly detection approach integrated with multi-feature semantic fusion techniques. PARI-GNN not only advances the state of the art in anomaly detection within project management contexts but also provides a scalable and interpretable framework for decision-makers to assess and mitigate risks of inauthentic participation. Future work will focus on further refining the model's capabilities, exploring additional data sources, and extending its applicability to other domains requiring robust authenticity verification.

REFERENCES

Li, J., Dani, H., Hu, X., & Liu, H. (2017, August). Radar: Residual analysis for anomaly detection in attributed networks. In IJCAI (Vol. 17, pp. 2152-2158).

Kipf, T. N., & Welling, M. (2016). Semi-supervised classification with graph convolutional networks. arXiv preprint arXiv:1609.02907.

Hamilton, W., Ying, Z., & Leskovec, J. (2017). Inductive representation learning on large graphs. Advances in neural information processing systems, 30.

Veličković, P., Cucurull, G., Casanova, A., Romero, A., Lio, P., & Bengio, Y. (2017). Graph attention networks. arXiv preprint arXiv:1710.10903.

Ma, X., Wu, J., Xue, S., Yang, J., Zhou, C., Sheng, Q. Z., ... & Akoglu, L. (2021). A comprehensive survey on graph anomaly detection with deep learning. IEEE Transactions on Knowledge and Data Engineering.

Tang, M., Yang, C., & Li, P. (2022). Graph auto-encoder via neighborhood wasserstein reconstruction. arXiv preprint arXiv:2202.09025.

Roy, A., Shu, J., Li, J., Yang, C., Elshocht, O., Smeets, J., & Li, P. (2023). GAD-NR: Graph Anomaly Detection via Neighborhood Reconstruction. arXiv preprint arXiv:2306.01951.

Yan, K., Gao, J., & Matsypura, D. (2023, October). FIW-GNN: A Heterogeneous Graph-Based Learning Model for Credit Card Fraud Detection. In 2023 IEEE 10th International Conference on Data Science and Advanced Analytics (DSAA) (pp. 1-10). IEEE.

Peng, Z., Luo, M., Li, J., Liu, H., & Zheng, Q. (2018, July). ANOMALOUS: A Joint Modeling Approach for Anomaly Detection on Attributed Networks. In IJCAI (pp. 3513-3519).

Davis, J., & Goadrich, M. (2006, June). The relationship between Precision-Recall and ROC curves. In Proceedings of the 23rd international conference on Machine learning (pp. 233-240).

Ding, K., Li, J., Bhanushali, R., & Liu, H. (2019, May). Deep anomaly detection on attributed networks. In Proceedings of the 2019 SIAM International Conference on Data Mining (pp. 594-602). Society for Industrial and Applied Mathematics.

Fan, H., Zhang, F., & Li, Z. (2020, May). Anomalydae: Dual autoencoder for anomaly detection on attributed networks. In ICASSP 2020-2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP) (pp. 5685-5689). IEEE.

Bandyopadhyay, S., N, L., Vivek, S. V., & Murty, M. N. (2020, January). Outlier resistant unsupervised deep architectures for attributed network embedding. In Proceedings of the 13th international conference on web search and data mining (pp. 25-33).

Fathony, R., Ng, J., & Chen, J. (2023, June). Interaction-focused anomaly detection on bipartite node-and-edge-attributed graphs. In 2023 International Joint Conference on Neural Networks (IJCNN) (pp. 1-10). IEEE.

Tang, J., Li, J., Gao, Z., & Li, J. (2022, June). Rethinking graph neural networks for anomaly detection. In International Conference on Machine Learning (pp. 21076-21089). PMLR.

Wang, X., Bo, D., Shi, C., Fan, S., Ye, Y., & Philip, S. Y. (2022). A survey on heterogeneous graph embedding: methods, techniques, applications and sources. IEEE Transactions on Big Data, 9(2), 415-436.

Bing, R., Yuan, G., Zhu, M., Meng, F., Ma, H., & Qiao, S. (2023). Heterogeneous graph neural networks analysis: a survey of techniques, evaluations and applications. Artificial Intelligence Review, 56(8), 8003-8042.

Liu, Z., Chen, C., Yang, X., Zhou, J., Li, X., & Song, L. (2018, October). Heterogeneous graph neural networks for malicious account detection. In Proceedings of the 27th ACM international conference on information and knowledge management (pp. 2077-2085).

Wang, X., Ji, H., Shi, C., Wang, B., Ye, Y., Cui, P., & Yu, P. S. (2019, May). Heterogeneous graph attention network. In The world wide web conference (pp. 2022-2032).

Zhang, Y., Fan, Y., Ye, Y., Zhao, L., & Shi, C. (2019, November). Key player identification in underground forums over attributed heterogeneous information network embedding framework. In Proceedings of the 28th ACM international conference on information and knowledge management (pp. 549-558).