Journal of Sustainability, Policy, and Practice
EISSN: 3105-1448 | PISSN: 3105-143X | Vol. 2, No. 1 (2026)

Article

Federated Graph Learning for Cross-Institution Money-Laundering Detection on Heterogeneous Transaction Networks

Emma Virtanen [1], Sanna Korhonen [1], Oskari Heikkilä [1] and Lauri Nieminen [1,*]

[1] Department of Computer Science, University of Helsinki, Helsinki, 00014, Finland
* Correspondence: Lauri Nieminen, Department of Computer Science, University of Helsinki, Helsinki, 00014, Finland

Abstract: This study proposes a federated graph-learning framework for detecting money-laundering activities across financial institutions without sharing raw customer data. Four banks participated in local training using transaction-graph structures containing 15.7 million nodes and 112.5 million edges. A global aggregation mechanism was used to synchronize encrypted model parameters. Compared with local graph models, the federated framework improved the average AUC from 0.84 to 0.90 and increased precision at 10% recall by 29.7%. Robustness tests showed stable performance even when one institution exhibited severe label imbalance. Differential-privacy noise prevented reconstruction of sensitive records under gradient-inversion tests. The results confirm that cross-institution modeling can significantly enhance detection accuracy while maintaining data confidentiality.

Keywords: federated learning; graph neural networks; transaction networks; privacy preservation; money-laundering detection

1. Introduction

Money-laundering activities typically involve complex transaction chains that span multiple accounts and often multiple financial institutions. This cross-institutional structure makes laundering schemes difficult to detect using rule-based systems or models trained within a single organisation, as critical information is fragmented across institutional boundaries [1]. Traditional AML systems therefore generate large numbers of alerts while failing to capture multi-step patterns that emerge only at the transaction-network level, resulting in high false-positive rates and substantial investigation costs [2]. These limitations have motivated a growing interest in modelling approaches that explicitly represent transactional relationships and support collaboration across institutions.

Graph-based methods have emerged as an effective way to capture the relational nature of financial transactions. By representing customers, accounts, and transfers as nodes and edges, graph models can exploit multi-hop dependencies that are inaccessible to tabular or rule-based approaches [3]. Graph neural networks (GNNs) further enhance this capability by learning structural and contextual representations directly from transaction networks and have demonstrated strong performance on AML datasets, even under severe class imbalance [4]. Beyond improvements in standalone detection models, recent system-level studies emphasise collaborative AML frameworks in which institutions jointly contribute to model training while preserving local control over sensitive data, highlighting the feasibility and potential benefits of cross-institution

Published: 21 February 2026

learning in realistic AML settings [5]. Despite this progress, most existing graph-based AML studies remain confined to data from a single institution and therefore cannot learn the cross-bank patterns that characterise large laundering networks.

Federated learning (FL) provides a natural foundation for collaborative AML modelling by enabling institutions to train a shared model without exchanging raw transaction data. In this paradigm, customer records remain on local servers, and only encrypted or aggregated model updates are communicated during training [6]. Prior work applies FL to fraud-related tasks, including credit-card fraud detection and suspicious-transaction scoring, and reports performance gains while maintaining data separation across institutions [7,8]. Commercial platforms further demonstrate that federated infrastructures can support inter-institutional cooperation while complying with regulatory and data-protection requirements [9]. However, many of these applications rely primarily on tabular features or aggregated statistics and do not fully exploit the rich relational structure of transaction networks. Federated graph learning (FGL) extends FL to graph-structured data by allowing each institution to train graph neural networks on its local transaction graph while participating in global aggregation. Recent research addresses challenges such as non-independent and non-identically distributed data, structural heterogeneity across clients, and privacy preservation when local graphs differ in scale or schema [10,11]. Early financial applications include federated GNNs for fraud detection and graph-based models enhanced with privacy-aware mechanisms [12]. Nevertheless, comprehensive studies of FGL for AML remain limited, particularly in scenarios involving large, heterogeneous transaction graphs drawn from multiple financial institutions. Privacy considerations remain central to collaborative AML modelling. Although FL avoids direct data sharing, prior research shows that sensitive information can still be inferred through gradient-based reconstruction and inversion attacks on model updates [13,14]. To mitigate these risks, differential-privacy mechanisms and related defences have been introduced to limit the reconstruction of individual records from shared parameters [15]. For AML applications, where transaction data contain highly sensitive personal and financial information, it is essential to evaluate detection accuracy and privacy leakage jointly under realistic threat models. Existing AML-focused studies rarely integrate federated graph learning with explicit gradient-inversion testing and formal privacy guarantees, leaving uncertainty about the operational safety of such approaches in practice. Taken together, the current literature indicates that cross-institution AML detection remains underexplored at scale. Empirical evaluations of federated graph-learning methods on large transaction networks from multiple banks are scarce, particularly when node types, edge semantics, and data distributions differ across institutions. Moreover, most studies focus on aggregate predictive metrics and provide limited insight into how performance degrades when individual institutions face extreme class imbalance or limited labelled data. Although privacy preservation is a key motivation for federated approaches, joint analyses that combine graph-based federated training with differential-privacy mechanisms and explicit privacy-attack evaluation remain uncommon in AML research.

This study develops and evaluates a federated graph-learning framework for AML detection across four financial institutions. Each institution constructs its own transaction graph, yielding a combined total of 15.7 million nodes and 112.5 million edges. Local GNNs are trained within each bank, and encrypted model parameters are aggregated to form a shared model without exposing raw transaction data. The evaluation compares federated and local-only models, examines robustness under severe label imbalance, and assesses the effectiveness of differential-privacy noise in mitigating gradient-inversion attacks. The results show that cross-institution graph learning can improve detection accuracy while limiting privacy leakage, demonstrating that federated graph learning offers a practical and scalable solution for collaborative AML on heterogeneous transaction networks.

2. Materials and Methods

2.1. Dataset Characteristics and Study Context

The study uses data from four financial institutions. Each institution built a local transaction graph from twelve months of internal records. The combined graph contains 15.7 million nodes and 112.5 million edges. Nodes represent customers or accounts, and edges represent transfers with amount, channel type, and time. Basic activity indicators and past alert signals were included as node features. Only records with complete identifiers and valid timestamps were kept to maintain graph consistency. Labels came from historical investigations and were highly imbalanced, with positive cases below 0.3% of all labelled nodes.

2.2. Experimental Design and Comparison Groups

Each institution trained its own graph neural network (GNN) on the local graph. Raw data were not shared. During federated training, institutions exchanged encrypted model parameters at fixed intervals. The experimental model was the federated version created through repeated aggregation. The control models were the four local GNNs trained independently with identical architectures and hyperparameters. All models used the same number of epochs and the same optimisation settings. This design isolates the effect of cross-institution training and removes differences due to model size or training schedule.

2.3. Evaluation Metrics and Quality Control

Performance was measured using AUC, precision, recall, and precision at low recall. These metrics match how AML systems are evaluated in practice. Predictions were linked to historical outcomes to compute each measure. Alerts with missing labels or conflicting investigation records were removed before training. Each institution checked its graph for incomplete nodes, repeated identifiers, and invalid timestamps. During federated training, each update was checked for correct format, size, and encryption before aggregation. These steps kept the local and global models trained on data of comparable quality.

2.4. Data Processing and Model Formulation

Node and edge features were normalised using the same rules across institutions. Message passing was used to update node representations. The update rule at layer $l+1$ is:

$$h_i^{(l+1)}=\sigma\left(W^{(l)}\cdot\frac{1}{|N(i)|}\sum_{j\in N(i)}h_j^{(l)}\right),$$

where $N(i)$ is the neighbour set of node i. Federated aggregation used a weighted mean:

$$\theta_{global}=\sum_{k=1}^{K}\alpha_k\,\theta_k,$$

with θ_k the parameters from institution k and α_k proportional to local sample size. The aggregated parameters were returned to each institution for further local training.

2.5. Training Workflow and Parameter Settings

Training was organised in rounds. Each round began with local GNN updates for a fixed number of epochs. Institutions then encrypted their model parameters and sent them to the coordinator. After aggregation, the updated parameters were distributed back to the institutions for the next round. All local models used the same learning rate, batch size, and stopping rules. Gradient clipping stabilised training, and nodes with very large degree were handled by sampling a limited number of neighbours. In a separate test, differential-privacy noise was added to the parameter updates to assess resistance to gradient inversion. This workflow kept all customer data local while enabling a shared model across institutions.

3. Results and Discussion

3.1. Performance of the Federated Graph Model

The federated graph model achieved higher detection accuracy than all local models trained at individual institutions. Local models reached an average AUC of 0.84 and produced low precision at 10% recall. After federated training, the shared model reached an AUC of 0.90 and improved precision at 10% recall by 29.7%. The largest gains appeared in the low-recall region, which is the range used most often in AML case triage. Figure 1 shows ROC and precision-recall curves for both local and federated models. The improvement pattern is consistent with earlier fraud-detection studies using federated graph methods on distributed data [16].

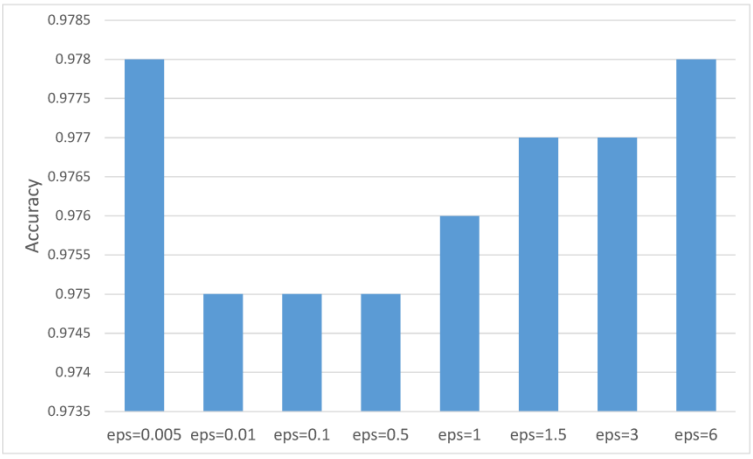

Figure 1. AUC and precision-recall curves for the federated model and the local models on the test set.

3.2. Performance under Heterogeneous Graph Structures and Label Imbalance

The four institutions differed in graph size, node degree, and label density. One institution had far fewer positive labels and a sparser network, which resulted in a weak local model with an AUC of 0.79. After federated training, this institution's AUC increased to 0.87, and its precision at 10% recall rose by more than 20%. The other institutions also showed smaller but steady gains. These results indicate that shared training helps weaker participants and does not depend on any single institution's graph size. Figure 1 separates the performance curves for all institutions and shows that the largest relative gains occurred where label imbalance was most severe.

3.3. Effect of Differential Privacy on Accuracy and Resistance to Reconstruction

We tested gradient-inversion attacks to examine privacy risks. Without differential-privacy (DP) noise, partial reconstruction of a few high-degree nodes was possible, although full edge patterns could not be recovered. After adding Gaussian noise to the parameter updates, reconstruction dropped to near-random accuracy. The AUC decreased slightly from 0.90 to 0.88, and precision at 10% recall dropped from 0.27 to 0.25. Figure 2 shows the relationship between added noise and model accuracy. These results indicate that privacy protection can be strengthened with a small loss in detection performance [17,18].

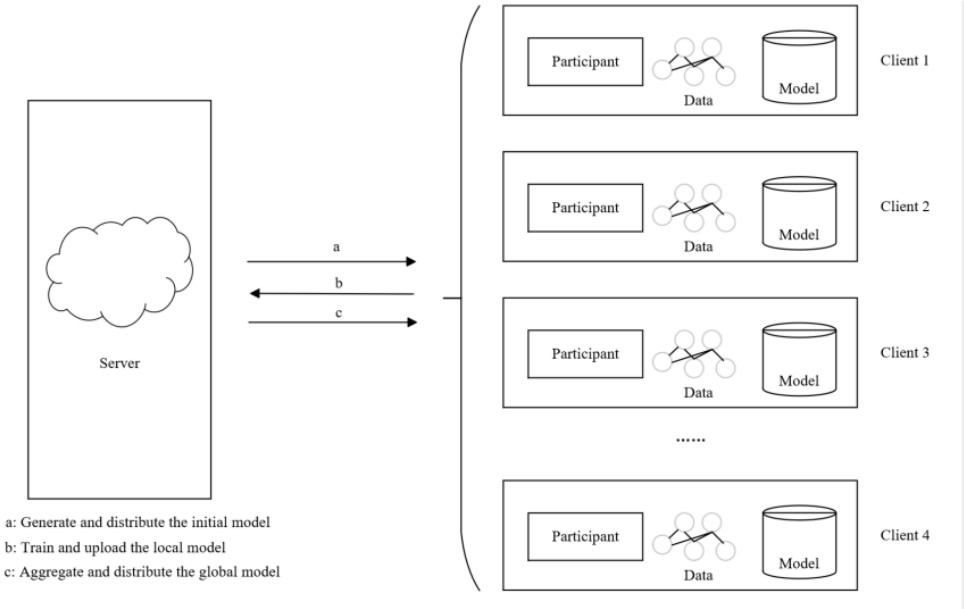

Figure 2. Detection accuracy and reconstruction risk under different levels of differential-privacy noise.

3.4. Comparison with Related AML Studies and Practical Implications

Existing AML research often focuses on single-bank datasets or centralised graph models. This limits the ability to detect laundering schemes that span multiple institutions. The present study shows that federated graph learning can improve detection accuracy while keeping customer data within each institution. Previous work has reported gains from graph-based AML models and from federated learning in financial settings, but few studies combine both approaches on large heterogeneous transaction networks [19,20]. Our results also show that explicit privacy tests are necessary when model updates may reveal sensitive information. The main limitations of this study include the use of four institutions from similar regulatory environments and offline replay experiments rather than real-time deployment. Future work may include more diverse institutions, additional product types, and online learning settings.

4. Conclusion

This study evaluated a federated graph-learning method for detecting money-laundering activity across four institutions. The federated model reached higher accuracy than all local models and produced better precision at low recall, which is the range most used in AML review. It also remained stable when one institution had very few positive labels, showing that shared training can help institutions with limited data. Adding differential-privacy noise reduced the success of reconstruction attacks while keeping performance close to the original model. These results show that it is possible to improve cross-institution detection without exposing customer records. The study has limits: the institutions operate under similar conditions, and the experiments were run offline rather than in real time. Future work may test more diverse settings, examine live deployment, and study how training rules influence both accuracy and operational workload.

References

1. J. Castelao-López, T. Corzo Santamaría, and D. Lagoa-Varela, "Analysis of the main techniques and tools to combat money laundering: A review of the literature," *Journal of Money Laundering Control*, 2025. doi: 10.1108/jmlc-10-2024-0159
2. S. Barcio, "Network analysis enhancements to improve anti-money laundering transaction monitoring systems in banks (Doctoral dissertation, Politecnico di Torino)," 2025.
3. A. M. Kanca, ", & Türker, İ," (2025). A systematic review of graph-based representation techniques for cyber-attack detection across application domains. Concurrency and Computation: Practice and Experience, vol. 37, no. 27-28, p. e70389, 2025.

4. Z. Amiri, A. Heidari, N. J. Navimipour, M. Unal, and A. Mousavi, "Adventures in data analysis: A systematic review of deep learning techniques for pattern recognition in cyber-physical-social systems," *Multimedia Tools and Applications*, vol. 83, no. 8, pp. 22909-22973, 2024. doi: 10.1007/s11042-023-16382-x
5. X. Gu, M. Liu, and J. Yang, "Application and effectiveness evaluation of federated learning methods in anti-money laundering collaborative modeling across inter-institutional transaction networks," 2025. doi: 10.20944/preprints202510.1828.v1
6. M. M. Fouda, Z. M. Fadlullah, M. I. Ibrahem, and N. Kato, "Privacy-preserving data-driven learning models for emerging communication networks: A comprehensive survey," *IEEE Communications Surveys & Tutorials*, 2024. doi: 10.1109/comst.2024.3486690
7. E. Tanuwijaya, and T. Mauritsius, "Anomaly detection in sales transactions for FMCG (fast moving consumer goods) distribution," *Journal of Applied Data Sciences*, vol. 5, no. 3, pp. 1223-1236, 2024.
8. Y. Yang, M. Guo, E. A. Corona, B. Daniel, C. Leuze, and F. Baik, "VR MRI training for adolescents: A comparative study of gamified VR, passive VR, 360 video, and traditional educational video," *arXiv preprint*, 2025.
9. T. K. Chowdhury, and S. P. Kudapa, "Federated learning models for privacy-preserving data sharing and secure analytics in healthcare industry," *International Journal of Business and Economics Insights*, vol. 4, no. 4, pp. 91-133, 2024.
10. W. Zhu, Y. Yao, and J. Yang, "Real-time risk control effects of digital compliance dashboards: An empirical study across multiple enterprises using process mining, anomaly detection, and interrupted time series," 2025.
11. S. A. Tanim, M. F. Mridha, M. Safran, S. Alfarhood, and D. Che, "Secure federated learning for Parkinson's disease: Non-IID data partitioning and homomorphic encryption strategies," *IEEE Access*, 2024. doi: 10.1109/access.2024.3454690
12. W. Bai, "Phishing website detection based on machine learning algorithm," In *Proceedings of the 2020 International Conference on Computing and Data Science*, 2020, pp. 293-298. doi: 10.1109/cds49703.2020.00064
13. A. Hatamizadeh, H. Yin, P. Molchanov, A. Myronenko, W. Li, P. Dogra, and H. R. ... Roth, "Do gradient inversion attacks make federated learning unsafe? IEEE Transactions on Medical Imaging, 42(7), 2044-2056," 2023.
14. J. B. Sheu, and X. Q. Gao, "Alliance or no alliance-Bargaining power in competing reverse supply chains," *European Journal of Operational Research*, vol. 233, no. 2, pp. 313-325, 2014.
15. W. Zhu, Y. Yao, and J. Yang, "Optimizing financial risk control for multinational projects: A joint framework based on CVaR-robust optimization and panel quantile regression," 2025. doi: 10.20944/preprints202510.1345.v1
16. M. Rahmati, "Real-time financial fraud detection using adaptive graph neural networks and federated learning," *International Journal of Management and Data Analytics*, vol. 5, no. 1, pp. 98-110, 2025. doi: 10.21203/rs.3.rs-6026136/v1
17. J. Wang, and Y. Xiao, "Research on credit risk forecasting and stress testing for consumer finance portfolios based on macroeconomic scenarios," 2025. doi: 10.1145/3785706.3785751
18. P. Rot, K. Grm, P. Peer, and V. Štruc, "PrivacyProber: Assessment and detection of soft-biometric privacy-enhancing techniques," *IEEE Transactions on Dependable and Secure Computing*, vol. 21, no. 4, pp. 2869-2887, 2023. doi: 10.1109/tdsc.2023.3319500
19. T. Li, Y. Jiang, E. Hong, and S. Liu, "Organizational development in high-growth biopharmaceutical companies: A data-driven approach to talent pipeline and competency modeling," 2025. doi: 10.20944/preprints202511.0631.v1
20. Z. Hu, Y. Hu, and H. Li, "Multi-task temporal fusion transformer for joint sales and inventory forecasting in Amazon e-commerce supply chain," *arXiv preprint*, 2025.