Journal of Information Systems Engineering and Management
2024, 9(3)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

Detecting Fraud Across Banks Without Sharing Customer Data: A Federated Graph Learning Approach Under Zero-Trust

Saiprakash Kodela

Independent Researcher, USA

ARTICLE INFO

Received: 01 July 2024

Revised: 20 Aug 2024

Accepted: 27 Aug 2024

ABSTRACT

Fraudulent transactions often involve networks of accounts, customers, devices, merchants, beneficiaries, and financial institutions. An individual bank may therefore observe only a limited portion of a coordinated fraud scheme. Although centralizing transaction data from multiple banks could improve fraud detection, such centralization may conflict with privacy requirements, confidentiality obligations, cybersecurity controls, data-localization rules, and restrictions on cross-border processing. This paper proposes a conceptual federated graph intelligence architecture for collaborative cross-bank fraud detection. Each participating bank constructs and retains a local heterogeneous temporal graph containing customers, accounts, transactions, devices, merchants, and beneficiaries. Compatible graph neural networks are trained locally, while protected model updates are communicated through secure aggregation. Where permitted by law, an independent privacy-preserving entity-matching service provides narrowly defined pseudonymous overlap signals without transferring raw identifiers or complete transaction graphs. The framework separates payment-fraud detection, mule-account identification, fraud-ring discovery, and anti-money-laundering analysis as distinct but related tasks. It also distinguishes bank-level differential privacy from transaction-, edge-, node-, customer-, and subject-level protection. Secure aggregation, contribution clipping, differential privacy, temporal controls, robust validation, explanation mechanisms, human review, and governance controls are treated as complementary safeguards. Because the paper proposes a conceptual architecture rather than an implemented system, it does not report experimental performance. Instead, it presents an evaluation framework covering predictive utility, privacy leakage, calibration, fairness, adversarial robustness, latency, communication cost, and auditability. The analysis concludes that federated graph intelligence may improve cross-bank fraud detection, but it cannot independently establish regulatory compliance. Deployment requires an explicit legal basis, clearly defined institutional responsibilities, privacy-impact assessment, common data standards, independent model validation, security testing, incident management, and meaningful human oversight.

Keywords: federated learning, graph neural networks, cross-bank fraud, privacy-preserving analytics, secure aggregation, differential privacy, regulatory governance

1. INTRODUCTION

Fraud detection is increasingly a network-analysis problem rather than a transaction-classification problem. A fraudulent transaction may appear legitimate when it is considered independently, yet become suspicious when connected to a larger structure involving shared devices, common beneficiaries, rapid transfers, circular fund movements, coordinated account creation, or repeated interactions among apparently unrelated customers. These relationships can be represented as graphs in which customers, accounts, merchants, devices, and transactions are modeled as nodes connected through different types of edges.

The cross-bank dimension makes this problem more difficult. A fraud ring may open accounts at several banks, transfer funds across institutions, use shared devices, and distribute transaction values to avoid institution-specific thresholds. Each bank sees only its local customers, accounts, and transactions. Consequently, an institution may identify isolated suspicious events without seeing the wider network that connects them. FATF has recognized that

collaborative analytics and information sharing may help financial institutions develop a more complete understanding of criminal networks, while also emphasizing that increased processing and sharing of personal data creates privacy and governance risks (Financial Action Task Force [FATF], 2021b, 2022).

Graph neural networks provide a technical approach for learning from relational data. Rather than treating every transaction as independent, a graph neural network aggregates information from connected entities and events. Cheng et al. (2022), for example, demonstrated the relevance of spatial and temporal attention for card-fraud detection. Rao et al. (2022) presented xFraud, which uses a heterogeneous graph to represent multiple entity types and includes an explanation component for fraud analysis. These studies demonstrate the value of relational and temporal representations, although they do not establish the feasibility of collaboration between independent banks under regulatory constraints.

Federated learning provides a possible alternative to centralized data pooling. In a federated architecture, participating institutions retain their local datasets and contribute model updates to a shared training process. Federated graph learning extends this principle to relational data. Liu et al. (2022) showed that graph-based representations, relational attention, and personalized components can be incorporated into a federated system. However, their work addressed social recommendation rather than regulated financial-fraud detection. The banking environment involves different legal duties, attack incentives, data structures, label quality, and consequences for affected individuals.

Federated learning is sometimes described as privacy-preserving because raw records remain within each participant's environment. This description is incomplete. Model updates, gradients, confidence scores, embeddings, and explanation outputs may reveal information about the training data. Gu et al. (2022) demonstrated membership-inference risks in federated learning, while Wainakh et al. (2022) showed that shared gradients may disclose information about users' labels. These findings establish that local data retention must be supplemented by formal privacy mechanisms, secure protocols, controlled outputs, and institutional governance.

The central research question of this paper is:

How can financial institutions collaboratively learn from distributed transaction graphs while limiting information disclosure and complying with applicable data-protection, banking-governance, cybersecurity, and human-oversight requirements?

The paper contributes a conceptual architecture that combines local heterogeneous temporal graphs, federated graph learning, protected entity matching, secure aggregation, differential privacy, robust validation, local explanation, and regulatory governance. It also presents a mathematical formulation and an experimental framework through which the architecture could be evaluated. The paper uses references published in 2022 or earlier and does not claim that the proposed system has already been implemented or empirically validated.

2. CONCEPTUAL AND LITERATURE BACKGROUND

Traditional fraud-detection models frequently evaluate individual transactions using features such as amount, transaction time, location, merchant category, authentication method, and previous customer behavior. Such models can be effective for known patterns but may fail to represent relationships connecting multiple accounts, devices, customers, or institutions. Graph representations address this limitation by making relationships part of the model input.

A heterogeneous graph includes more than one type of node or relationship. In a banking context, node types may include customers, accounts, transactions, merchants, beneficiaries, devices, addresses, telephone numbers, and Internet Protocol identifiers. Edge types may represent account ownership, fund transfers, shared devices, shared addresses, common beneficiaries, merchant interactions, or authentication events. The meaning of a transfer edge differs from the meaning of a shared-device edge, and a heterogeneous graph model can learn relation-specific representations.

The temporal dimension is equally important. Fraudulent activity often involves transaction velocity, bursts, repeated transfers, rapid cash-out, or sequences of events. A model should use only information that was available

when the prediction was made. When a transaction at time (t) is evaluated, later chargebacks, fraud confirmations, or transactions should not influence its features. Failure to preserve this temporal ordering creates future-information leakage and produces unrealistic evaluation results. Cheng et al. (2022) showed the relevance of spatial-temporal graph attention in fraud detection, supporting the inclusion of both relational and temporal structure.Rao et al. (2022) proposed xFraud for explainable fraud-transaction detection in an online retail setting. Its detector uses a heterogeneous graph neural network, while its explanation component produces information intended to assist human users. xFraud supports two elements of the proposed framework: the use of multiple entity types and the need for human-understandable outputs. Nevertheless, the study should not be treated as direct evidence for cross-bank deployment. An online retail platform generally controls its transaction graph within one organizational environment, whereas a banking consortium involves separate legal entities, distinct customer relationships, and restrictions on information disclosure.

Federated graph learning attempts to combine graph modeling with decentralized institutional control. The participants may each hold separate graphs or different partitions of a larger graph. In cross-bank fraud detection, each institution possesses a local subgraph of the payment ecosystem. Some entities may overlap across banks, but those overlaps may be unknown, incomplete, or legally restricted. Parameter aggregation can share general fraud patterns; however, it does not automatically reconstruct the transactions or paths that cross institutional boundaries.

The architecture must also address statistical heterogeneity. Banks differ in size, customer population, transaction channels, products, geographic coverage, fraud prevalence, investigation procedures, feature availability, and label maturity. A model trained at one institution may therefore perform poorly at another. A federated design should combine shared representations with institution-specific parameters and thresholds rather than assuming that one uniform decision function is optimal for every bank.

Privacy risks arise at several stages. Model updates may reveal whether a record participated in training, disclose labels, or expose characteristics of local distributions. Overlap tokens may reveal that a customer, beneficiary, device, or merchant is present at more than one bank. Graph embeddings may contain information about local neighborhoods. Explanations may reveal another institution's customer relationship even when the underlying graph is not directly disclosed. Secure aggregation and differential privacy can reduce some of these risks, but they do not address every threat.

Secure aggregation enables a coordinating service to calculate an aggregate of participant updates without directly observing each individual update. The foundational protocol proposed by Bonawitz et al. (2017) illustrates how participant inputs may be masked while still allowing aggregate computation. Secure aggregation protects the confidentiality of individual contributions from the coordinator, but it does not determine whether those contributions are accurate or non-malicious.

Differential privacy provides a formal way to limit the influence of a defined protected unit on a released output. The protected unit may be a transaction, edge, graph node, customer, person, or participating bank. These definitions are not interchangeable. Noise applied to one complete update from each bank may provide institution-level or client-level protection, subject to the training and accounting mechanism, but it does not automatically provide customer-level protection. Customer-level protection requires contribution bounding and accounting that treat all records associated with one customer as the protected unit.

Table 1 summarizes the principal sources and their role in the conceptual architecture.

Table 1 - Selected literature and relevance to the proposed framework

Source	Principal contribution	Relevance	Limitation
Cheng et al. (2022)	Spatial-temporal graph attention for fraud detection	Supports relational and temporal modeling	Centralized card-fraud setting

Rao et al. (2022)	Heterogeneous graph fraud detection and explanation	Supports multiple entity types and investigator explanations	Online retail rather than cross-bank banking
Liu et al. (2022)	Federated graph learning with personalization	Supports shared and local graph-model components	Social recommendation rather than fraud detection
Gu et al. (2022)	Federated membership-inference attack	Demonstrates that FL updates may leak participation information	Not banking- or graph-specific
Wainakh et al. (2022)	Label leakage from federated gradients	Demonstrates that gradients may reveal labels	Not banking-specific
Bonawitz et al. (2017)	Secure aggregation	Supports confidential aggregation of participant updates	Does not prevent poisoning or endpoint compromise
Dwork and Roth (2014)	Differential-privacy foundations	Supports formal privacy definitions and accounting	Graph and banking application require additional design
FATF (2021a, 2021b, 2022)	Collaborative analytics and privacy-enhancing technologies	Supports governance-oriented analysis	Does not provide universal legal authorization
European Union (2016)	Data-protection principles and safeguards	Supports privacy, security, and accountability controls	Application depends on jurisdiction and circumstances
BCBS (2013)	Risk-data governance and reporting principles	Supports data quality, lineage, and accountability	Not specifically designed for federated fraud models

3. RESEARCH METHOD

This paper uses a conceptual design and structured literature-synthesis method. It does not claim to be a completed systematic review or an empirical evaluation. The publication cutoff is December 31, 2022. Sources were selected based on their relevance to graph-based fraud detection, federated graph learning, federated privacy, secure aggregation, differential privacy, financial-crime collaboration, data protection, and banking data governance.

Priority was given to peer-reviewed journal articles and official regulatory or policy publications. Conference papers and preprints were avoided where suitable journal alternatives were available. Bonawitz et al. (2017) is retained as a foundational exception because its secure-aggregation protocol is directly relevant to the architecture. Literature relating primarily to securities markets, investment trading, stock prediction, or financial-market forecasting was excluded.

The architecture was developed by identifying the functions required for cross-bank fraud detection and mapping those functions to technical, privacy, security, and governance controls. The principal functions include local graph creation, local model training, protected update exchange, optional entity matching, shared-model distribution, local inference, explanation, investigation, validation, and monitoring.

Journal of Information Systems Engineering and Management
2024, 9(3)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

The paper distinguishes conceptual propositions from empirical findings. Claims about the performance of the proposed architecture are expressed as hypotheses for future testing rather than established results. The proposed system should be evaluated only through realistic institution-based graph partitions, chronological testing, privacy attacks, adversarial scenarios, and operational validation.

A complete systematic review would require database-specific search strings, Scopus and Web of Science exports, duplicate removal, title and abstract screening, full-text eligibility assessment, and a documented exclusion log. Those activities are outside the scope of the present conceptual manuscript and should be completed before a submission that claims systematic-review coverage.

4. PROBLEM FORMULATION

Assume a consortium of (K) participating banks. Each Bank (k) maintains a private heterogeneous temporal graph:

$$G_k = (V_k, E_k, X_k, R_k, T_k, Y_k).$$

where:

- (V_k) is the local node set;
- (E_k) is the local edge set;
- (X_k) contains approved node and event features;
- (R_k) identifies relationship types;
- (T_k) contains timestamps;
- (Y_k) contains task-specific labels.

The principal predictive tasks should remain distinct. Payment-fraud detection estimates whether a payment event is fraudulent. Mule-account detection estimates whether an account is being used to receive, transfer, or withdraw fraud-related funds. Fraud-ring discovery identifies connected groups of accounts or entities whose combined behavior indicates coordination. Anti-money-laundering analysis identifies behavior requiring additional investigation but should not be treated as equivalent to confirmed fraud or a legal determination.

The shared graph-encoder parameters are represented by θ, while ϕk represents institution-specific parameters. A multi-task objective may be expressed as:

$$\min_{\theta,\{\phi_k\}} \sum_{k=1}^{K} \alpha_k \left[\begin{array}{l} \mathcal{L}^{(k)}_{\text{payment}} + \lambda_1 \mathcal{L}^{(k)}_{\text{mule}} + \lambda_2 \mathcal{L}^{(k)}_{\text{ring}} \\ + \lambda_3 \mathcal{L}^{(k)}_{\text{calibration}} + \lambda_4 \mathcal{L}^{(k)}_{\text{fairness}} + \lambda_5 \mathcal{R}^{(k)} \end{array} \right],$$

where:

- (θ) denotes shared graph-encoder parameters;
- (ϕ_k) denotes institution-specific parameters;
- (α_k) is a capped aggregation weight;
- $(\mathcal{R}^{(k)})$ is a regularization term;
- each predictive loss corresponds to a separately governed task.

Fraud datasets are usually highly imbalanced because legitimate transactions substantially outnumber confirmed fraud cases. A class-weighted focal-loss function may therefore be used:

Journal of Information Systems Engineering and Management
2024, 9(3)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

$$\mathcal{L}_{\text{fraud}}^{(k)} = -\sum_{i} \alpha_{y_i} \left[\begin{array}{l} y_i(1-p_i)^{\gamma}\log(p_i) \\ + (1-y_i)p_i^{\gamma}\log(1-p_i) \end{array} \right].$$

Here, (p_i) is the predicted fraud probability for observation (i), (y_i) is the corresponding binary label, (α_{yi} }) is the class weight, and (γ) reduces the influence of easily classified observations. The positive-class modulation term is $(1-p_i)^{\gamma}$, while the negative-class modulation term is (p_i^{γ}).

5. FEDERATED GRAPH INTELLIGENCE ARCHITECTURE

The proposed architecture contains three local bank environments, a regulatory and governance control plane, an independent overlap-processing service, a secure aggregation service, a shared-model registry, and local investigation systems.

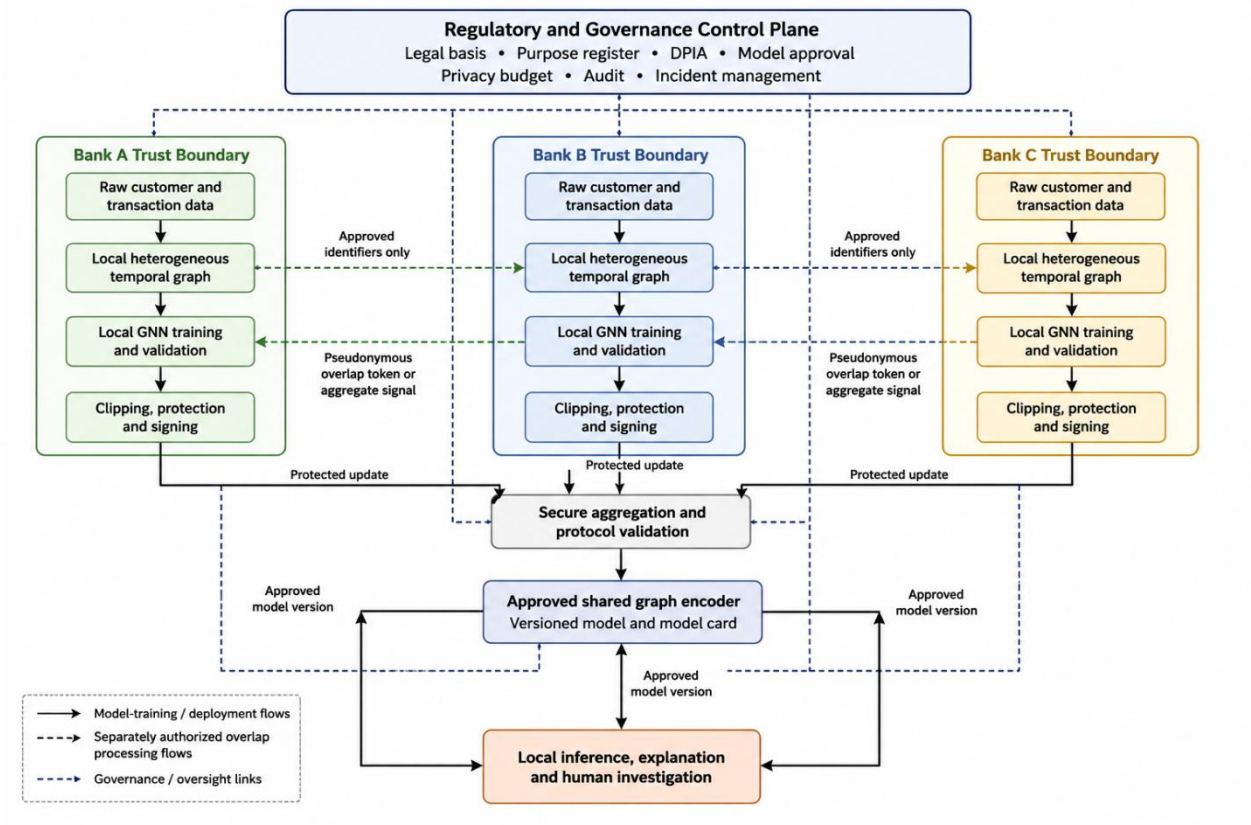

Figure 1- Federated graph intelligence architecture and trust boundaries

Note. Solid arrows represent model-training and deployment flows. Dashed arrows represent separately authorized overlap-processing flows. Raw customer and transaction data remain within the originating bank's trust boundary.

Each bank converts authorized local information into a heterogeneous temporal graph using a common consortium schema. The schema defines node types, edge types, feature definitions, timestamps, permitted identifiers, label maturity, data-quality thresholds, and retention rules. Standardization is necessary because a feature may have different meanings at different institutions. For example, a confirmed-fraud label may represent customer reimbursement at one bank and completed investigation at another.

Journal of Information Systems Engineering and Management
2024, 9(3)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

The local graph model should include relation-specific processing because different edges carry different meanings. A transfer relationship is not equivalent to shared-device use, common ownership, or merchant interaction. Time-aware message passing is also required so that information occurring after a transaction does not influence its historical prediction.

Each bank trains a compatible graph encoder locally. The architecture divides the model into shared and institution-specific components. The shared encoder learns general relational patterns, while local normalization layers, prediction heads, thresholds, or calibration components allow adaptation to institutional differences. This design reduces the risk that a global model optimized for the largest participant will perform poorly at smaller or structurally different banks.

After local training, Bank (k) generates a model update (Δ_k). Its contribution may be clipped as follows:

$$\bar{\Delta}_k = \Delta_k \min\left(1, \frac{C_B}{\|\Delta_k\|_2}\right)$$

where (C_B) is the maximum approved bank-update norm. Clipping limits the influence of one participant and provides a basis for privacy and poisoning controls.

If noise is applied to the complete bank update,

the resulting protection should ordinarily be described as bank-level or client-level differential privacy, assuming the full mechanism and privacy accountant support that claim. It should not automatically be described as customer-level differential privacy.

Secure aggregation computes:

$$\bar{\Delta}_k = \Delta_k \min\left(1, \frac{C_B}{\|\Delta_k\|_2}\right)$$

$$\left[\tilde{\Delta}_k\right] = \bar{\Delta}_k + \mathcal{N}\left(0, \sigma_B^2 C_B^2 I\right),$$

without revealing each unmasked participant update to the aggregation service. Secure aggregation reduces coordinator visibility, but it does not guarantee that every update is honest. A malicious bank may submit manipulated gradients intended to weaken the model or conceal a targeted fraud pattern. The architecture therefore requires authenticated participants, signed model code, bounded contributions, validation against local holdout data, aggregate anomaly testing, rollback capability, and controlled incident procedures.

Cross-bank fraud detection may require information about overlapping entities. Parameter aggregation alone cannot reveal that a device, merchant, beneficiary, or legal entity is present in more than one bank. The architecture therefore includes an optional independent overlap service. This service may use private set intersection, trustee-managed pseudonymization, secure multiparty computation, or a narrowly defined trusted execution environment.

The overlap service should return only the minimum authorized information, such as a pseudonymous overlap token, an institution-count category, a coarse risk indicator, or an aggregate community signal. It should not distribute clear-text customer identifiers, complete graph neighborhoods, another bank's customer list, or unrestricted investigative information.

Entity resolution introduces the possibility of false matches. An inaccurate match may incorrectly connect unrelated customers and increase their risk score. The matching process must therefore define acceptable

identifiers, matching thresholds, false-match rates, correction procedures, access rules, retention periods, re-identification testing, and prohibited secondary uses.

The global model is versioned and distributed through an approved model registry. Each version should be accompanied by a model card describing the purpose, training population, approved tasks, privacy mechanism, performance limitations, known risks, validation status, and prohibited uses.

Inference and explanation occur locally. An investigator may receive a transaction-risk score, account-risk score, suspicious path, velocity indicator, shared-device signal, or relation-contribution summary. Cross-bank explanations should not reveal the identity of another participating institution or expose its raw records unless a separate legally authorized information-sharing process permits such disclosure.

Table 2 summarizes the architecture components.

Component	Function	Information processed	Principal control
Local graph environment	Constructs bank-specific temporal graph	Raw local customer and transaction records	Institutional trust boundary
Local GNN trainer	Learns local graph representations	Local graph, features, and labels	Controlled code and validation
Update-protection layer	Clips, signs, and protects updates	Model update	Contribution bounds and privacy controls
Secure aggregation service	Computes aggregate update	Masked participant updates	Threshold secure aggregation
Overlap service	Identifies approved cross-bank overlaps	Restricted identifiers or cryptographic representations	Independent authorization and minimization
Shared-model registry	Stores approved model versions	Aggregated parameters and documentation	Version control and approval
Local inference service	Produces fraud-risk scores	Local graph and approved shared model	Local access control
Case-management system	Supports investigator decisions	Alerts, explanations, and case outcomes	Human review and audit logging
Governance control plane	Controls purpose, privacy, security, and lifecycle	Policies, approvals, and logs	Independent oversight

6. PRIVACY, SECURITY, AND THREAT MODEL

The architecture assumes that technical and institutional participants may fail or behave adversarially. The coordinator may be honest but curious, a participating bank may submit malicious updates, participants may collude, an external attacker may compromise infrastructure, and an authorized insider may misuse access.

The coordinator should not be able to inspect individual updates. Secure aggregation addresses this threat by limiting the coordinator to an aggregate output. However, the final model and aggregate may still reveal information. Differential privacy should therefore be considered where a formal privacy guarantee is required.

Privacy granularity must be explicitly defined. Transaction-level privacy treats datasets as neighboring when one transaction is added or removed. Edge-level privacy protects one relationship. Node-level privacy protects one node

and its incident relationships. Customer-level privacy protects the full contribution of one customer within a bank. Subject-level privacy attempts to protect one person whose records may occur at several institutions. Bank-level privacy protects one participating bank's complete contribution.

These guarantees require different clipping and accounting mechanisms. Customer-level privacy cannot be obtained merely by clipping the final update of each bank. The bank must first limit the combined influence of all nodes, edges, transactions, and features associated with each customer. Subject-level privacy is even more difficult because the same person may appear at multiple banks under different identifiers.

Privacy reporting should include the adjacency definition, clipping unit, clipping threshold, noise mechanism, values of ($\in$) and (Δ), number of training rounds, participant-sampling assumptions, and composition method. Differential privacy is a property of a specified randomized mechanism; it is not a general label for decentralization.

Membership-inference and label-leakage attacks should be included in security testing. Gu et al. (2022) showed that prediction-confidence sequences may support membership inference in federated learning. Wainakh et al. (2022) demonstrated that gradients may disclose label information. These attacks should be tested under realistic cross-silo conditions rather than assuming that raw-data localization is sufficient protection.

Secure aggregation introduces a privacy–robustness tension. Hiding individual updates protects participants from coordinator inspection, but it can make conventional update-level poisoning detection more difficult. Possible controls include authenticated training software, hardware-backed keys, verifiable contribution bounds, aggregate-level tests, minimum cohort sizes, validation rounds, and rapid rollback.

External attackers may target the aggregation service, model registry, overlap service, communication network, encryption keys, or local training systems. Controls should include encryption in transit and at rest, mutual authentication, network segmentation, least-privilege access, hardware-backed key storage, vulnerability management, security monitoring, and incident-response procedures.

Insider threats are especially important because explanations and overlap information may be valuable for legitimate investigations while also being susceptible to misuse. Access should be role-based, purpose-specific, logged, and periodically reviewed. Sensitive investigative data should remain separate from the federated training environment.

Table 3 presents the principal threats and controls.

Threat	**Potential consequence**	**Proposed safeguards**	**Remaining limitation**
Curious coordinator	Infers local information from updates	Secure aggregation and minimum cohort size	Aggregate or final model may still leak
Malicious bank	Poisons the shared model	Bounded updates, attestation, validation, rollback	Hidden updates complicate attribution
Participant collusion	Infers another bank's contribution	Threshold aggregation and access separation	Small consortiums remain vulnerable
Compromised overlap service	Reveals entity relationships	Independent operation, encryption, minimization	Matching itself creates sensitive information
External attacker	Steals models, tokens, or keys	Segmentation, authentication, monitoring	Endpoint compromise remains possible
Insider misuse	Uses alerts for	Least privilege, logging,	Legitimate access

Journal of Information Systems Engineering and Management
2024, 9(3)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

	unauthorized purposes	dual control	cannot be risk-free
False entity match	Incorrectly associates unrelated customers	Conservative thresholds and correction process	Matching errors cannot be eliminated
Temporal leakage	Inflates historical performance	Event-time processing and chronological testing	Data arrival delays may remain
Explanation leakage	Reveals another bank's relationship	Local explanations and redaction	Some overlap information may remain sensitive

7. REGULATORY AND GOVERNANCE CONSTRAINTS

Regulatory compliance cannot be inferred solely from the fact that raw data remain local. Federated processing may still involve personal data, profiling, derived information, pseudonymous identifiers, model updates, or cross-border transfers. Whether a particular architecture is lawful depends on jurisdiction, purpose, legal basis, institutional roles, data categories, contractual arrangements, and effects on individuals.

The General Data Protection Regulation provides an illustrative framework for purpose limitation, data minimization, accuracy, storage limitation, security, accountability, data protection by design, and impact assessment. It also regulates certain decisions based solely on automated processing that produce legal or similarly significant effects. Article 22 should not be described as prohibiting every automated fraud score. Its application depends on whether the decision is solely automated and whether it creates the required level of effect (European Parliament & Council of the European Union, 2016).

The architecture therefore separates risk scoring from final adverse action. A model may prioritize an alert for investigation, but customer blocking, account closure, reporting, or other significant actions should be governed by applicable law, institutional policy, and meaningful human review. Human involvement must be substantive rather than a formal approval of an automated recommendation.A data-protection impact assessment or equivalent risk assessment should be completed before production deployment. The assessment should examine the purpose, necessity, proportionality, data flows, privacy unit, entity-matching process, attack surface, explanation outputs, cross-border processing, and effects of false positives.

FATF publications on data pooling, collaborative analytics, and private-sector information sharing recognize that privacy-enhancing technologies may facilitate collaboration. They also emphasize the importance of legal frameworks, public-authority involvement, privacy rules, governance, and responsible implementation. FATF guidance should not be interpreted as a universal authorization for unrestricted data sharing or joint profiling (FATF, 2021a, 2021b, 2022).

BCBS 239 was designed for risk-data aggregation and reporting rather than federated fraud detection. Its principles are nevertheless relevant by analogy because they emphasize governance, data architecture, accuracy, completeness, timeliness, adaptability, lineage, and supervisory review. The consortium should maintain shared definitions, versioned transformations, data-quality thresholds, model lineage, validation records, and senior-management accountability.

The governance lifecycle begins with a clearly defined fraud task and prohibited uses. The consortium must then establish legal authority, contractual roles, data-controller or processor responsibilities where applicable, liability, intellectual-property rights, withdrawal procedures, and dispute resolution. The graph schema, features, labels, privacy mechanism, matching protocol, and retention rules should be approved before training.

Journal of Information Systems Engineering and Management
2024, 9(3)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

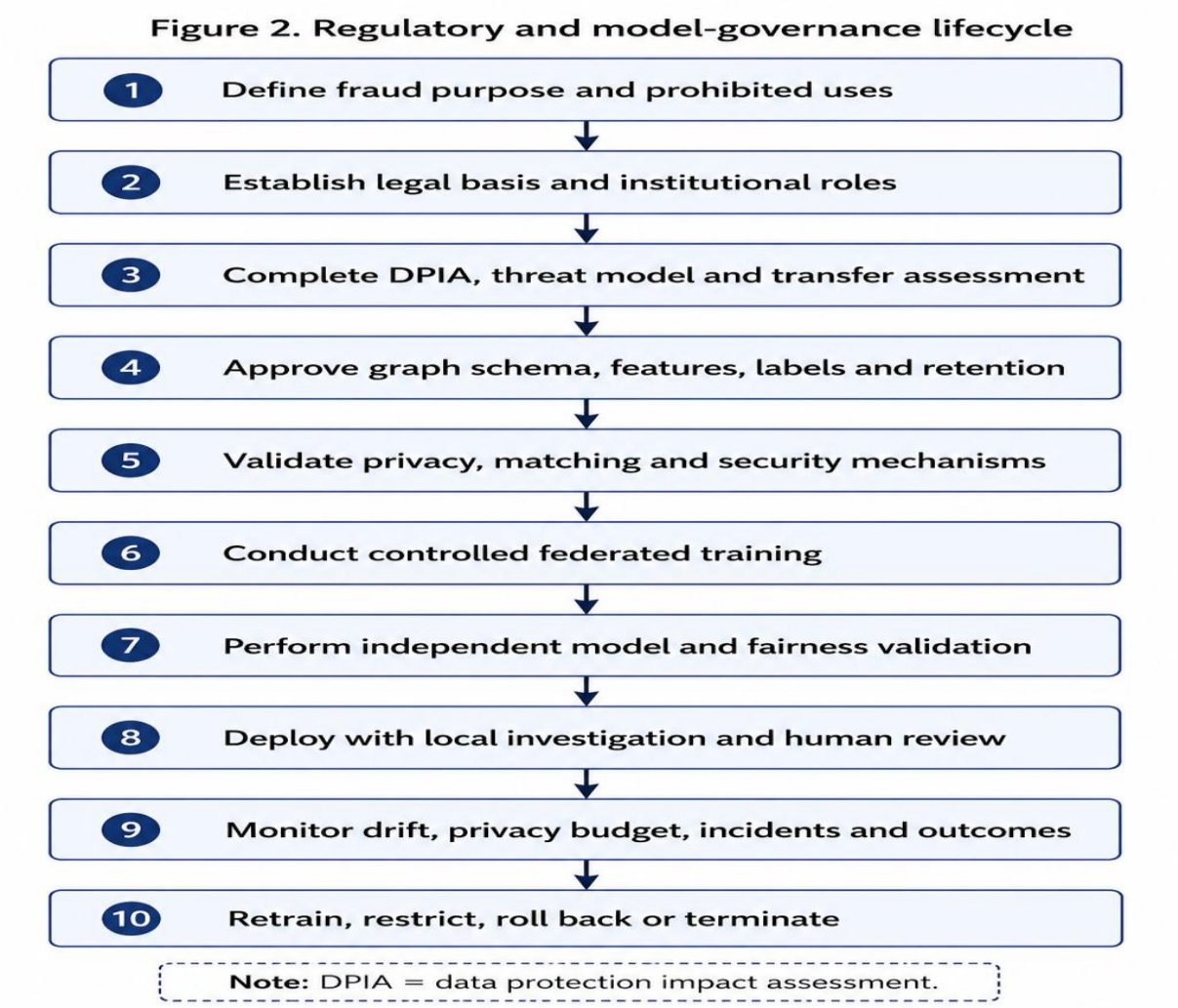

Figure 2- Regulatory and model-governance lifecycle

Note. DPIA means data-protection impact assessment. A material change in purpose, data, model behavior, privacy mechanism, or jurisdiction should initiate a new review.The consortium agreement should specify that shared models and overlap outputs may be used only for approved fraud-related purposes. Reuse for marketing, credit scoring, employee monitoring, or unrelated customer profiling should be prohibited unless independently authorized.

Every model version should undergo independent validation. Validation should assess predictive performance, privacy, fairness, robustness, data quality, temporal leakage, entity-resolution errors, explanation behavior, and operational controls. Approval should be time-limited and subject to continuing monitoring.

8. EVALUATION FRAMEWORK

The proposed architecture should be evaluated against local and collaborative baselines. Because the paper is conceptual, the following measures are recommendations rather than reported findings.

The data should be partitioned by real or realistically simulated institutions. Randomly dividing transactions from one dataset into artificial clients would not reproduce the cross-bank problem. Training, validation, and testing should be separated chronologically. Fraud labels should reflect realistic confirmation delays and should not be treated as immediately available.The primary baselines should include a local tabular classifier, a local graph neural network, a centralized graph model in a legally approved research environment, a federated non-graph model, a federated graph model without overlap information, and the complete protected architecture.

Journal of Information Systems Engineering and Management
2024, 9(3)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

The centralized model should be used only as an experimental upper-bound comparator. It should not be presented as a recommended production design. Its performance may indicate the value lost because cross-bank topology is incomplete.

Evaluation should examine unequal bank sizes, differences in fraud prevalence, missing features, different graph densities, delayed labels, participant dropout, entity-matching errors, malicious updates, and changing fraud patterns. A single random train-test split would not be sufficient.

The following hypotheses may be tested:

H_1: Federated graph learning improves multi-bank fraud detection relative to independent local models.
H_2: Protected cross-bank overlap signals improve detection beyond parameter aggregation alone.
H_3:Institution-specific model components improve performance under non-identical bank distributions.
H_4:Customer-level privacy controls reduce privacy leakage but impose a greater utility cost than bank-level controls.

H_5:Secure aggregation limits coordinator visibility while increasing the difficulty of identifying malicious updates.

Fraud detection should be evaluated using precision-recall area under the curve, precision at investigation capacity, recall at a fixed false-positive rate, cost-weighted recall, expected prevented loss, and investigator workload. Ordinary accuracy is likely to be misleading because legitimate transactions dominate the dataset.

Calibration should be measured because banks may interpret risk scores as probabilities. The Brier score, expected calibration error, and reliability diagrams should be reported for each institution. A model that ranks cases effectively but produces poorly calibrated probabilities may still create governance and threshold-setting problems.

Privacy evaluation should combine formal accounting with empirical attacks. Formal reporting should state the privacy unit, $(\in)$, (Δ), clipping threshold, training rounds, and assumptions. Empirical testing should include membership inference, label inference, reconstruction where relevant, overlap-token re-identification, repeated-round leakage, collusion, and explanation disclosure.

Robustness testing should include random update corruption, targeted poisoning, model backdoors, participant dropout, label errors, and compromised local clients. Testing should measure both aggregate performance and targeted harm to specific fraud patterns or institutions.

Fairness testing should be conducted only for attributes that may lawfully and ethically be used for audit. The analysis should examine false-positive and false-negative rates, alert rates, calibration, and investigation outcomes. A fairness constraint should not be assumed to eliminate discrimination; it is one element of a broader impact assessment.

Operational evaluation should measure communication volume, training rounds, local computation, secure-aggregation overhead, matching-service latency, inference latency, model-registry availability, incident-recovery time, and rollback time.

Table- 4 Recommended evaluation measures

Evaluation dimension	**Measures**
Predictive utility	PR-AUC, precision at capacity, recall at fixed false-positive rate
Financial utility	Prevented loss, investigation cost, net expected benefit
Calibration	Brier score, expected calibration error, reliability curve
Temporal validity	Detection delay, monthly performance, drift

Journal of Information Systems Engineering and Management
2024, 9(3)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

	degradation
Cross-bank value	Improvement over local models and performance on multi-bank patterns
Privacy	(\varepsilon), (\delta), membership inference, label inference, re-identification
Robustness	Poisoning resistance, dropout tolerance, corrupted-label sensitivity
Entity resolution	Match precision, recall, false-match rate, correction rate
Fairness	False-positive and false-negative differences and calibration by audit group
Explanation	Fidelity, stability, usefulness, and disclosure risk
Efficiency	Bytes per round, convergence rounds, training time, P95/P99 latency
Governance	Audit completeness, traceability, approval time, and rollback time

Statistical reporting should include multiple random seeds, multiple chronological windows, confidence intervals, institution-level results, the worst-performing institution, failed experiments, threshold-selection procedures, and ablation studies. Ablation studies should remove overlap signals, temporal edges, personalization, secure aggregation, privacy noise, and fairness constraints one at a time.

9. DISCUSSION

Federated graph intelligence offers an attractive middle position between isolated institutional models and centralized transaction pooling. It can allow banks to learn shared patterns while retaining control over raw records. Graph models can represent relationships and temporal structures that transaction-level classifiers may miss. Nevertheless, the architecture creates new challenges that should not be underestimated.

The first challenge is incomplete topology. Federated model averaging can transfer parameters but cannot automatically reveal cross-bank transaction chains. Optional overlap information may improve detection, yet it also creates additional privacy and governance risks. The value of each overlap signal must therefore be tested against its disclosure risk.

The second challenge is institutional heterogeneity. A fraud pattern common at one bank may be absent at another. Features and labels may have different meanings, and global averaging may reduce the performance of specialized local models. Shared encoders and personalized prediction heads offer one response, but they require careful validation.

The third challenge is the tension between privacy and robustness. Secure aggregation hides individual contributions, while poisoning defenses often require inspection of those contributions. A secure design must combine privacy and integrity mechanisms rather than assuming that either can be added independently.

The fourth challenge is graph privacy. A person may be represented through multiple accounts, devices, addresses, and transactions. Removing one customer may alter a large graph neighborhood. Customer-level privacy is therefore more difficult than transaction-level privacy. Subject-level privacy is more difficult again because one person may appear at several institutions.

The fifth challenge is explanation disclosure. Investigators need meaningful reasons for alerts, but graph paths and shared-entity information may reveal another bank's customer relationships. Explanation interfaces must be treated as regulated information-sharing channels rather than neutral visualization tools.

The sixth challenge is legal diversity. The same architecture may be permissible in one jurisdiction and prohibited or restricted in another. Model updates, embeddings, pseudonymous tokens, or risk scores may be treated differently under different legal systems. A global technical architecture therefore requires regional and jurisdiction-specific governance.

The architecture should not replace existing fraud investigators or statutory processes. It should prioritize cases, identify network patterns, and support analysis. Final decisions must consider model uncertainty, evidence quality, customer circumstances, legal duties, and investigator judgment.

10. LIMITATIONS AND FUTURE RESEARCH

The paper presents a conceptual framework and does not provide empirical proof that federated graph intelligence improves cross-bank fraud detection. Future research must implement the architecture using realistic institution-based graph partitions, chronological evaluation, delayed labels, and operational constraints.

The literature synthesis is structured but is not a completed Scopus or Web of Science systematic review. A future review should document the exact database search strings, search date, number of retrieved records, duplicate removal, screening criteria, publication-type verification, and reasons for exclusion.

Protected entity matching remains an unresolved design area. Future research should compare private set intersection, trustee-managed pseudonymization, secure multiparty computation, and trusted execution environments. These comparisons should include matching accuracy, computation cost, collusion resistance, correction procedures, and re-identification risk.

Future work should also compare transaction-, edge-, node-, customer-, subject-, and bank-level differential privacy. The studies should report utility, privacy loss, and attack success for each definition rather than referring generally to privacy-preserving federated learning.

Robust aggregation under secure aggregation requires additional investigation. Future systems may use verifiable update bounds, code attestation, secure robust statistics, and controlled validation cohorts. These mechanisms must be evaluated against malicious participants rather than only random noise.

Fraud labels are delayed, incomplete, and institution-dependent. Future research should examine positive-unlabeled learning, weak supervision, label maturity, investigator disagreement, and delayed feedback. Unreported fraud should not automatically be treated as a confirmed legitimate transaction.

Future research should evaluate whether explanations improve investigator effectiveness and whether they disclose sensitive information. Suitable measures include explanation fidelity, consistency, investigation time, case-confirmation rate, and re-identification risk.

Finally, the regulatory analysis should be adapted to specific jurisdictions through collaboration with financial regulators, data-protection authorities, participating banks, security specialists, consumer representatives, and legal experts.

11. CONCLUSION

This paper proposed a conceptual federated graph intelligence framework for cross-bank fraud detection under regulatory constraints. The framework represents local banking data as heterogeneous temporal graphs and combines local graph-model training with protected update aggregation. Where legally authorized, an independent overlap service provides minimized cross-bank signals without distributing complete customer or transaction graphs.The proposed framework distinguishes payment fraud, mule-account detection, fraud-ring discovery, and AML investigation. It also distinguishes bank-level privacy from customer-, transaction-, node-, edge-, and subject-level privacy. These distinctions are necessary because different tasks and privacy units require different labels, controls, evaluation procedures, and legal justifications.

Federated learning reduces direct raw-data movement but does not eliminate information leakage. Secure aggregation protects individual updates from coordinator inspection but does not prevent poisoning. Differential privacy requires an explicit protected unit and privacy accountant. Cross-bank entity matching and model explanations may create disclosure risks even when raw records remain local.

The principal conclusion is that federated graph intelligence should be treated as one component of a wider institutional system. A defensible deployment requires explicit legal authority, purpose limitation, common data standards, secure aggregation, appropriate differential privacy, matching controls, adversarial testing, human review, independent validation, auditability, incident response, and continuing regulatory oversight.

Federated graph intelligence may help banks identify fraud networks that no institution can observe independently. Its value, however, depends not only on predictive accuracy but also on whether the collaboration is lawful, secure, explainable, proportionate, and accountable.

REFERENCES

[1] Basel Committee on Banking Supervision. (2013). *Principles for effective risk data aggregation and risk reporting*. Bank for International Settlements.

[2] Bonawitz, K., Ivanov, V., Kreuter, B., Marcedone, A., McMahan, H. B., Patel, S., Ramage, D., Segal, A., & Seth, K. (2017). Practical secure aggregation for privacy-preserving machine learning. In *Proceedings of the 2017 ACM SIGSAC Conference on Computer and Communications Security* (pp. 1175–1191). Association for Computing Machinery. https://doi.org/10.1145/3133956.3133982

[3] Cheng, D., Wang, X., Zhang, Y., & Zhang, L. (2022). Graph neural network for fraud detection via spatial-temporal attention. *IEEE Transactions on Knowledge and Data Engineering, 34*(8), 3800–3813. https://doi.org/10.1109/TKDE.2020.3025588

[4] Dwork, C., & Roth, A. (2014). The algorithmic foundations of differential privacy. *Foundations and Trends in Theoretical Computer Science, 9*(3–4), 211–407. https://doi.org/10.1561/0400000042

[5] European Data Protection Board. (2018). *Guidelines on automated individual decision-making and profiling for the purposes of Regulation 2016/679* (WP251rev.01).

[6] European Parliament & Council of the European Union. (2016). Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data. *Official Journal of the European Union, L 119*, 1–88.

[7] Financial Action Task Force. (2021a). *Opportunities and challenges of new technologies for AML/CFT*. FATF/OECD.

[8] Financial Action Task Force. (2021b). *Stocktake on data pooling, collaborative analytics and data protection*. FATF/OECD.

[9] Financial Action Task Force. (2022). *Partnering in the fight against financial crime: Data protection, technology and private sector information sharing*. FATF/OECD.

[10] Gu, Y., Bai, Y., & Xu, S. (2022). CS-MIA: Membership inference attack based on prediction confidence series in federated learning. *Journal of Information Security and Applications, 67*, Article 103201. https://doi.org/10.1016/j.jisa.2022.103201

[11] Liu, Z., Yang, L., Fan, Z., Peng, H., & Yu, P. S. (2022). Federated social recommendation with graph neural network. *ACM Transactions on Intelligent Systems and Technology, 13*(4), Article 55, 1–24. https://doi.org/10.1145/3501815

[12] Rao, S. X., Zhang, S., Han, Z., Zhang, Z., Min, W., Chen, Z., Shan, Y., Zhao, Y., & Zhang, C. (2022). xFraud: Explainable fraud transaction detection. *Proceedings of the VLDB Endowment, 15*(3), 427–436. https://doi.org/10.14778/3494124.3494128

[13] Wainakh, A., Ventola, F., Müßig, T., Keim, J., Garcia Cordero, C., Zimmer, E., Grube, T., Kersting, K., & Mühlhäuser, M. (2022). User-level label leakage from gradients in federated learning. *Proceedings on Privacy Enhancing Technologies, 2022*(2), 227–244. https://doi.org/10.2478/popets-2022-0043