

الجمهورية الشعبية الديمقراطية الجزائرية
People's Democratic Republic of Algeria
وزارة التعليم العالي و البحث العلمي
Ministry of Higher Education and Scientific Research
المدرسة العليا للإعلام الآلي 8 ماي 5491 - سيدي بلعباس
Higher School of Computer Science
8 Mai 1945 - Sidi Bel Abbès



Final Year Thesis

To obtain the diploma of: **Engineer**

Field of Study: **Computer Science**

Specialization: **Artificial Intelligence and Data Science**

Theme

A Lambda-Based Heterogeneous Temporal Graph Neural Network with Relation-Aware Attention for Real-Time Financial Fraud Detection

Presented by

AMIEUR Zineb Ichrake HAMDI Aya

Defended on: **June, 2026**

In front of the jury composed of

Dr. Khaldi Belkacem
Dr. Malki Abdelhamid
Dr. Laaziz Lahlou
Pr. Kara Nadjia
Dr. Elhannani Souad

President of the Jury
Thesis Supervisor
Co-Supervisor
Co-Supervisor
Examiner

Academic Year: 2025/2026

Abstract

The rapid growth of digital payment systems has significantly increased the volume and sophistication of financial fraud worldwide. Traditional detection approaches treat transactions as independent instances, limiting their ability to capture the complex relational and temporal patterns inherent in financial networks. Graph Neural Networks have emerged as a powerful paradigm to address these limitations, enabling the detection of fraud patterns that are invisible to conventional approaches.

This thesis presents a novel framework for real-time financial fraud detection over heterogeneous temporal graphs. The proposed system follows a Lambda architecture that decouples deep offline graph computation from lightweight real-time inference: a Batch Processing Network builds rich entity embeddings through stacked Relation-Specific Self-Attentive Layers and Gated Recurrent Units, cached in Redis for instant retrieval, while a Real-Time Inference Network combines these embeddings with incoming transaction features to produce a fraud score within millisecond-level latency. Strict temporal constraints are enforced throughout the pipeline to eliminate any future information leakage. The framework explores several configurations to handle severe class imbalance, and introduces the fraud rate ratio as a calibration metric that complements standard detection scores by measuring whether the model predicts fraud at a realistic rate.

A systematic ablation study evaluates the contribution of each design decision. The framework is evaluated on the IEEE-CIS Fraud Detection benchmark against tabular and graph-based baselines, achieving an AUPRC of 0.422, an AUROC of 0.869, and a fraud rate ratio of 0.995 with an average inference latency of 3.82 ms per transaction.

Keywords: Financial fraud detection, Graph Neural Networks, Heterogeneous temporal graphs, Lambda architecture, Real-time inference, Contrastive learning, Class imbalance, Calibration.

Résumé

La croissance rapide des systèmes de paiement numérique a considérablement augmenté le volume et la sophistication de la fraude financière à l'échelle mondiale. Les approches traditionnelles de détection traitent chaque transaction de manière indépendante, ce qui limite leur capacité à capturer les patterns relationnels et temporels complexes inhérents aux réseaux de transactions financières. Les réseaux de neurones sur graphes ont émergé comme un paradigme puissant pour pallier ces limitations, permettant la détection de schémas frauduleux invisibles aux approches conventionnelles.

Cette thèse présente un framework original pour la détection de fraude financière en temps réel sur des graphes temporels hétérogènes. Le système proposé suit une architecture Lambda qui découple le calcul profond hors ligne de l'inférence légère en temps réel : un réseau de traitement par lots construit des embeddings d'entités riches à travers des couches d'attention spatiale spécifiques aux relations et des unités récurrentes à portes, mis en cache dans Redis pour une récupération instantanée, tandis qu'un réseau d'inférence en temps réel combine ces embeddings avec les caractéristiques des transactions entrantes pour produire un score de fraude en quelques millisecondes. Des contraintes temporelles strictes sont appliquées tout au long du pipeline afin d'éliminer toute fuite d'informations futures. Le framework explore plusieurs configurations pour gérer le déséquilibre sévère des classes, et introduit le ratio de taux de fraude comme métrique de calibration qui complète les scores de détection standards en mesurant si le modèle prédit la fraude à un taux réaliste.

Une étude d'ablation systématique évalue la contribution de chaque décision de conception. Le framework est évalué sur le benchmark IEEE-CIS Fraud Detection face à des baselines tabulaires et à des méthodes basées sur les graphes, atteignant un AUPRC de 0,422, un AUROC de 0,869 et un ratio de taux de fraude de 0,995, avec une latence moyenne d'inférence de 3,82 ms par transaction.

Mots-clés : Détection de fraude financière, Réseaux de neurones sur graphes, Graphes temporels hétérogènes, Architecture Lambda, Inférence en temps réel, Apprentissage contrastif, Déséquilibre de classes, Calibration.

ملخص

أدى النمو السريع لأنظمة الدفع الرقمي إلى زيادة ملحوظة في حجم الاحتيال المالي وتطور أساليبه على المستوى العالمي. تعالج نهج الكشف التقليدية كل معاملة بشكل مستقل، مما يحد من قدرتها على استيعاب الأنماط العلائقية والزمنية المعقدة الكامنة في شبكات المعاملات المالية. وقد برزت الشبكات العصبية على الرسوم البيانية بوصفها نموذجاً فعالاً للتغلب على هذه القيود، إذ تُتيح الكشف عن أنماط الاحتيال التي تعجز الأساليب التقليدية عن رصدها.

تقدم هذه الأطروحة إطاراً جديداً للكشف عن الاحتيال المالي في الوقت الفعلي على رسوم بيانية زمنية غير متجانسة. يتبع النظام المقترح معمارية لامدا التي تفصل بين الحوسبة العميقة غير المتزامنة والاستدلال الخفيف في الوقت الفعلي: تقوم شبكة المعالجة الدفعية ببناء تمثيلات غنية للكيانات عبر طبقات انتباه مكانية خاصة بكل علاقة ووحدات تكرارية بوابية، يتم تخزينها مؤقتاً في Redis لاسترجاعها فوراً، بينما تجمع شبكة الاستدلال الفوري هذه التمثيلات مع خصائص المعاملات الواردة لإنتاج درجة احتيال في غضون ميلي ثانية. تُطبق قيود زمنية صارمة على طول خط الأنابيب للقضاء على أي تسرب للمعلومات المستقبلية. يستكشف الإطار عدة تهيئات للتعامل مع الخلط الحاد في توازن الفئات، ويُقدم نسبة معدل الاحتيال كقياس معايرة يُكَلِّل درجات الكشف المعيارية بقياس مدى واقعية معدل الاحتيال الذي يتنبأ به النموذج.

تُقيم دراسة إزالة منهجية شاملة لأربعة عشر تهيئة مساهمة كل قرار تصميمي. يُقَيَّم الإطار على معيار IEEE-CIS للكشف عن الاحتيال في مواجهة نماذج مرجعية جدولية وأخرى قائمة على الرسوم البيانية، محققاً AUPRC بقيمة 0,422، و AUROC بقيمة 0,869، ونسبة معدل احتيال تبلغ 0,995، مع متوسط زمن استدلال يبلغ 3,82 ميلي ثانية لكل معاملة.

الكلمات المفتاحية: الكشف عن الاحتيال المالي، الشبكات العصبية على الرسوم البيانية، الرسوم البيانية الزمنية غير المتجانسة، معمارية لامدا، الاستدلال في الوقت الفعلي، التعلم التبايني، الخلط في توازن الفئات، المعايرة.