Σχολή Θετικών Επιστημών και Τεχνολογίας

Πληροφορική

Πτυχιακή Εργασία

Ανίχνευση Οικονομικής Απάτης μέσω Ανάλυσης Γράφων και Τεχνητής Νοημοσύνης

Φίλιππος Μανώλης

Επιβλέπων καθηγητής: Βασίλειος Ταμπακάς

Πάτρα, Ιούλιος 2026

Ανίχνευση Οικονομικής Απάτης μέσω Ανάλυσης Γράφων και Τεχνητής Νοημοσύνης

Φίλιππος Μανώλης

Επιτροπή Επίβλεψης Πτυχιακής Εργασίας

Επιβλέπων Καθηγητής:
Δρ. Βασίλειος Ταμπακάς
Καθηγητής,
Τμήμα Ηλεκτρολόγων Μηχανικών και Μηχανικών Υπολογιστών
Πανεπιστήμιο Πελοποννήσου

Συν-Επιβλέπων Καθηγητής:
Δρ. Χρήστος Κούτσικας
Εντεταλμένος Διδάσκων
Τμήμα Πληροφορικής
Οικονομικό Πανεπιστήμιο Αθηνών

Συν-Επιβλέπων Καθηγητής:
Δρ. Βάιος Παπαϊωάννου
Μέλος ΕΔΙΠ
Τμήμα Μηχανικών Ηλεκτρονικών Υπολογιστών και Πληροφορικής
Πανεπιστήμιο Πατρών

Πάτρα, Ιούλιος 2026

Στη γυναίκα μου, Κορίνα,

και τον γιο μου, Αλέξανδρο

Θερμές ευχαριστίες στον επιβλέποντα καθηγητή μου, κ. Βασίλη Ταμπακά,

για τη βοήθεια και την υποστήριξή του.

Περίληψη

Η παρούσα πτυχιακή εργασία πραγματεύεται την ανίχνευση οικονομικής απάτης μέσω της ανάλυσης γράφων και της τεχνητής νοημοσύνης, με έμφαση στην αξιοποίηση των σχέσεων μεταξύ οντοτήτων που συμμετέχουν σε ύποπτες συναλλαγές. Σε αντίθεση με τις παραδοσιακές προσεγγίσεις που εξετάζουν κάθε συναλλαγή ως μεμονωμένη εγγραφή, η εργασία υιοθετεί μια δικτυοκεντρική προσέγγιση, στην οποία συναλλαγές, κάρτες, email domains και συσκευές αναπαρίστανται ως κόμβοι ενός ετερογενούς γράφου, ώστε, μέσω των ακμών που τους ενώνουν, να αποκαλύπτονται συνδεδεμένα μοτίβα απάτης, κοινότητες υψηλού κινδύνου, και έμμεσες συσχετίσεις που δεν είναι εύκολα ορατές σε συμβατικά tabular δεδομένα.

Για την υλοποίηση του προτεινόμενου πλαισίου αναπτύχθηκε ένα ολοκληρωμένο pipeline, το οποίο περιλαμβάνει προεπεξεργασία και εμπλουτισμό χαρακτηριστικών, κατασκευή γράφου σε Neo4j, εξαγωγή γραφοαναλυτικών μετρικών μέσω της βιβλιοθήκης Neo4j GDS, δημιουργία embeddings με FastRP και εκπαίδευση ενός Graph Neural Network (Heterogeneous GraphSAGE). Παράλληλα, υλοποιήθηκε baseline μοντέλο XGBoost, το οποίο αξιοποιεί συνδυασμό tabular χαρακτηριστικών, graph analytics και embeddings, ώστε να καταστεί δυνατή η συγκριτική αξιολόγηση της ενδεχόμενης προστιθέμενης αξίας των τοπολογικών χαρακτηριστικών του γράφου σε περιβάλλοντα ανίχνευσης απάτης.

Η πειραματική αξιολόγηση πραγματοποιήθηκε σε έντονα ανισορροπημένα δεδομένα, με κύριο πεδίο εφαρμογής το IEEE-CIS Fraud Detection dataset, και βασίστηκε σε μετρικές κατάλληλες για το πεδίο του προβλήματος και την ανισορροπία των δεδομένων. Τα αποτελέσματα έδειξαν ότι το HeteroGraphSAGE παρουσίασε ελαφρώς υψηλότερο AUC-PR (0.6317), σημαντικά υψηλότερο Precision (0.7569) και αισθητά βελτιωμένο F1-score (0.6078), όμως περιορίστηκε σε μέτριο Recall (0.5077), ενώ το XGBoost εμφάνισε καλύτερο AUC-ROC (0.9318) και διατήρησε αρκετά υψηλότερο Recall, με αντάλλαγμα σημαντικά περισσότερα false positives.

Επιπλέον, η εργασία ενσωματώνει τεχνικές XAI, και συγκεκριμένα τις SHAP και GNNExplainer, με στόχο την ερμηνεία τόσο της συνολικής συμπεριφοράς των μοντέλων όσο και επιμέρους προβλέψεων σε επίπεδο υπογράφων. Με τον τρόπο αυτό, εξετάζεται το πώς μπορεί να ενισχυθεί η διαφάνεια, η αξιοπιστία και η πρακτική αξιοποίηση των προτεινόμενων μεθόδων σε πραγματικά περιβάλλοντα λήψης αποφάσεων.

Συνολικά, τα ευρήματα της εργασίας υποδεικνύουν ότι, από τη μία πλευρά, η συνδυαστική χρήση graph databases, graph analytics, GNNs και XAI συνιστά μια ιδιαίτερα υποσχόμενη προσέγγιση για την αποτελεσματικότερη ανίχνευση σύνθετων μορφών οικονομικής απάτης, ενώ από την άλλη, καθιερωμένα μοντέλα ML όπως το XGBoost, διατηρούν την αξία τους, προσφέροντας συγκεκριμένα πλεονεκτήματα.

Λέξεις – Κλειδιά

Ανίχνευση οικονομικής απάτης, Ανάλυση γράφων, GNN, Neo4j, XAI, Heterogeneous GraphSAGE

Abstract

This thesis addresses the detection of financial fraud through graph analysis and artificial intelligence, with an emphasis on leveraging the relationships between entities involved in suspicious transactions. Unlike traditional approaches that treat each transaction as an isolated record, this work adopts a network-centric perspective, in which transactions, cards, email domains, and devices are represented as nodes in a heterogeneous graph. Through the edges connecting them, it becomes possible to uncover connected fraud patterns, high-risk communities, and indirect associations that are difficult to capture in conventional tabular data.

To implement the proposed framework, a comprehensive end-to-end pipeline was developed, including data preprocessing and feature enrichment, graph construction in Neo4j, extraction of graph analytics metrics using the Neo4j GDS library, embedding generation with FastRP, and training of a Graph Neural Network (heterogeneous GraphSAGE). In parallel, an XGBoost baseline model was trained using a combination of tabular features, graph analytics, and embeddings, allowing a comparative assessment of the added value of graph features in fraud detection environments.

The experimental evaluation was conducted on highly imbalanced data, primarily focusing on the IEEE-CIS Fraud Detection dataset, and was based on metrics suitable for the problem domain and the data imbalance. The results showed that HeteroGraphSAGE exhibited a slightly higher AUC-PR (0.6317), significantly higher Precision (0.7569), and a noticeably improved F1-score (0.6078); however, it was limited to a moderate Recall (0.5077). On the other hand, XGBoost achieved a better AUC-ROC (0.9318) and maintained a considerably higher Recall, at the cost of significantly more false positives.

Additionally, the work incorporates Explainable AI (XAI) techniques, specifically SHAP and GNNExplainer, aiming to interpret both the overall behavior of the models and individual predictions at the subgraph level. In doing so, it examines how to enhance transparency, reliability, and the practical deployment of the proposed methods in real-world decision-making environments.

Overall, the findings suggest that while the combined use of graph databases, graph analytics, GNNs, and XAI represents a highly promising approach for more effectively

detecting complex forms of financial fraud, established ML models such as XGBoost retain their value by offering distinct advantages.

Keywords

Financial fraud detection, Graph analysis, GNN, Neo4j, XAI, Heterogeneous GraphSAGE

Περιεχόμενα

Κατάλογος Εικόνων / Σχημάτων

Κατάλογος Πινάκων

Συντομογραφίες & Ακρωνύμια

ΒΔ	Βάση Δεδομένων
ΒΔΓ	Βάση Δεδομένων Γράφων
ΝΔΓ	Νευρωνικό Δίκτυο Γράφων
ΤΝ	Τεχνητή Νοημοσύνη
ADASYN	Adaptive Synthetic
AI	Artificial Intelligence
GAT	Graph Attention Network
GCN	Graph Convolution Network
GNN	Graph Neural Network
GraphSAGE	Graph Sample and Aggregate
HGNN	Heterogeneous Graph Neural Network
LIME	Local Interpretable Model-agnostic Explanations
ML	Machine Learning
PyG	PyTorch Geometric
RNN	Recurrent Neural Networks
SHAP	Shapley Additive Explanations
SMOTE	Synthetic Minority Oversampling Technique
SVM	Support Vector Machines
XAI	Explainable AI
XGBoost	Extreme Gradient Boosting

1. Εισαγωγή

1.1 Το πρόβλημα και η σημασία του

Η οικονομική απάτη αποτελεί ένα κρίσιμο πρόβλημα για τον χρηματοπιστωτικό τομέα και τις σύγχρονες οικονομίες παγκοσμίως. Σύμφωνα με τα στοιχεία του Οργανισμού Ηνωμένων Εθνών, το 2-5% του παγκόσμιου ΑΕΠ, που αντιστοιχεί σε 0,8 έως 2 τρισεκατομμύρια δολάρια, ξεπλένεται ετησίως μέσω παράνομων δραστηριοτήτων.[1] Πιο συγκεκριμένα, σύμφωνα με πρόσφατες εκτιμήσεις, οι απώλειες από απάτη πιστωτικών καρτών αναμένεται να φτάσουν τα 403,88 δισεκατομμύρια δολάρια τα επόμενα 10 χρόνια. (Naim κ.ά., 2025) Η εξέλιξη της ψηφιακής οικονομίας και η ευρεία υιοθέτηση των ηλεκτρονικών συναλλαγών έχουν διευρύνει τις ευκαιρίες για εγκληματικές δραστηριότητες, καθιστώντας επιτακτική την ανάγκη για προηγμένα συστήματα ανίχνευσης.

Οι παραδοσιακές μέθοδοι ανίχνευσης απάτης, όπως τα συστήματα βασισμένα σε κανόνες (rule-based systems) και οι στατιστικές προσεγγίσεις, εμφανίζουν σημαντικούς περιορισμούς. Αυτά τα συστήματα αντιμετωπίζουν δυσκολίες στην προσαρμογή τους σε νέα μοτίβα απάτης, παρουσιάζουν υψηλά ποσοστά ψευδώς θετικών αποτελεσμάτων (false positives), και αδυνατούν να εντοπίσουν εξελιγμένα σχήματα που εκτείνονται σε πολλαπλούς λογαριασμούς και συναλλαγές. Αν και απλά στην υλοποίηση, απαιτούν συνεχείς ενημερώσεις των κανόνων καθώς οι απατεώνες προσαρμόζουν και εξελίσσουν τις τακτικές τους.

Η θεωρία γράφων, καθώς επίσης και εφαρμογές τεχνητής νοημοσύνης (ΤΝ) όπως είναι τα νευρωνικά δίκτυα γράφων (Graph Neural Networks - GNNs) προσφέρουν μια επαναστατική προσέγγιση στην ανίχνευση οικονομικής απάτης. Αντί να αναλύουν τις συναλλαγές μεμονωμένα, αντιμετωπίζουν το χρηματοπιστωτικό σύστημα ως ένα πολύπλοκο δίκτυο όπου οι λογαριασμοί, οι συναλλαγές, οι συσκευές, τα άτομα και άλλες οντότητες αποτελούν κόμβους, ενώ οι σχέσεις μεταξύ τους αναπαρίστανται ως ακμές. Αυτή η αναπαράσταση επιτρέπει την αναγνώριση δομικών μοτίβων που χαρακτηρίζουν δόλιες δραστηριότητες, όπως δίκτυα ξεπλύματος χρημάτων, συνθετικές ταυτότητες και οργανωμένα κυκλώματα απάτης. Τα GNNs έχουν δείξει εξαιρετική απόδοση στην

[1] https://www.unodc.org/unodc/en/money-laundering/overview.html

ανίχνευση απάτης, με ορισμένες εφαρμογές να επιτυγχάνουν ακρίβεια έως 98,5% σε ανάλυση συναλλαγών Bitcoin. (Asiri & Somasundaram, 2025)

1.2 Στόχοι και ερευνητικά ερωτήματα

Η παρούσα πτυχιακή εργασία έχει τους ακόλουθους στόχους:

Κύριος Στόχος:

Ανάπτυξη ενός ολοκληρωμένου συστήματος ανίχνευσης οικονομικής απάτης, το οποίο μοντελοποιεί τα δεδομένα κάνοντας χρήση βάσης δεδομένων γράφων, και για την επεξεργασία τους συνδυάζει τεχνικές ανάλυσης γράφων με αλγορίθμους τεχνητής νοημοσύνης, επιτυγχάνοντας αποδεκτές επιδόσεις όσον αφορά την προγνωστική του ικανότητα.

Δευτερεύοντες Στόχοι:

1. Συγκριτική αξιολόγηση των μεθόδων βασισμένων σε γράφους με παραδοσιακές τεχνικές μηχανικής μάθησης σε πραγματικά δεδομένα οικονομικών συναλλαγών.
2. Ενσωμάτωση τεχνικών Explainable AI (XAI), όπως SHAP και GNNExplainer, για τη διασφάλιση της διαφάνειας και της συμμόρφωσης με κανονιστικές απαιτήσεις.
3. Σχεδιασμός αρχιτεκτονικής που να υποστηρίζει την επεξεργασία μεγάλου όγκου συναλλαγών και να επιτρέπει την εύκολη επέκταση με νέες λειτουργίες.

Για την επίτευξη των παραπάνω στόχων, η έρευνα θα επικεντρωθεί στα ακόλουθα κρίσιμα ερευνητικά ερωτήματα:

1. **Πώς μπορούν τα Graph Neural Networks (GNNs) να βελτιώσουν την ακρίβεια ανίχνευσης οικονομικής απάτης σε σύγκριση με τις παραδοσιακές μεθόδους;** Το ερώτημα αυτό αποσκοπεί να διερευνήσει την συγκριτική αποδοτικότητα των GNNs έναντι των κλασικών μεθόδων μηχανικής μάθησης σε σημαντικά μεγέθη δεδομένων.
2. **Ποιος είναι ο ρόλος των βάσεων δεδομένων γράφων (graph databases) στον εντοπισμό πολύπλοκων μοτίβων απάτης που εκτείνονται σε πολλαπλές οντότητες και συναλλαγές;** Η έρευνα αυτή θα εξετάσει πώς η χρήση τεχνολογιών

και λογισμικού όπως η Neo4j μπορεί να υποστηρίξει αποτελεσματικά την ανίχνευση δικτύων απάτης.

3. **Πώς μπορούν οι τεχνικές Τεχνητής Νοημοσύνης και Μηχανικής Μάθησης (AI/ML) να μειώσουν τα ψευδώς θετικά αποτελέσματα διατηρώντας υψηλά ποσοστά ανίχνευσης;** Το ερώτημα αυτό αναφέρεται σε ένα κρίσιμο πρόβλημα στην πράξη, όπου τα συστήματα ανίχνευσης παράγουν ένα μεγάλο αριθμό ψευδών συναγερμών που απαιτούν χρονοβόρα χειροκίνητη ανάλυση.
4. **Ποιες μέθοδοι graph analytics (κεντρικότητα, ανίχνευση κοινοτήτων, ανάλυση μονοπατιών) είναι πιο αποτελεσματικές για τον εντοπισμό διαφορετικών τύπων οικονομικής απάτης;** Η έρευνα αυτή θα εξετάσει τα σχετικά πλεονεκτήματα διαφορετικών αλγορίθμων γράφων για ειδικά σενάρια απάτης.

1.3 Πεδίο εφαρμογής και περιορισμοί

Οι πιο διαδεδομένοι τύποι οικονομικής απάτης, οι οποίοι παρουσιάζονται συνοπτικά στο θεωρητικό μέρος της εργασίας, είναι οι εξής:

Απάτη σε Πληρωμές με Κάρτες (Payment Card Fraud): Εντοπισμός παράνομων συναλλαγών με πιστωτικές και χρεωστικές κάρτες.

Ξέπλυμα Χρημάτων (Money Laundering): Αναγνώριση μοτίβων που υποδηλώνουν απόκρυψη παράνομων κεφαλαίων μέσω πολύπλοκων ροών συναλλαγών.

Συνθετική Ταυτότητα (Synthetic Identity Fraud): Ανίχνευση ψευδών ταυτοτήτων που δημιουργούνται συνδυάζοντας πραγματικά και πλαστά προσωπικά στοιχεία.

Η παρούσα εργασία ωστόσο θα επικεντρωθεί στην απάτη σε πληρωμές με κάρτες, και συγκεκριμένα στις συναλλαγές ηλεκτρονικού εμπορίου (e-commerce) χωρίς φυσική παρουσία της κάρτας (Card-not-present).

Λόγω της φύσης και του περιορισμένου εύρους της εργασίας, ισχύουν οι ακόλουθοι περιορισμοί:

Δεδομένων: Λόγω της εμπιστευτικότητας των χρηματοπιστωτικών δεδομένων, η έρευνα θα βασιστεί κυρίως σε δημόσια διαθέσιμα datasets, όπου ένα μεγάλο μέρος των πληροφοριών είναι ανωνυμοποιημένο.

Τεχνικοί: Η υλοποίηση θα περιοριστεί σε περιβάλλον ανάπτυξης με συγκεκριμένες υπολογιστικές δυνατότητες, χωρίς πρόσβαση σε διακομιστές προορισμένους για εργασίες ML ή σε συστήματα παραγωγής χρηματοπιστωτικών ιδρυμάτων.

Επεκτασιμότητας: Αν και το πρωτότυπο θα σχεδιαστεί με γνώμονα την επεκτασιμότητα, η αξιολόγηση σε πολύ μεγάλης κλίμακας δεδομένα θα είναι εκτός του πεδίου της παρούσας εργασίας.

Χρονικοί: Η υλοποίηση θα επικεντρωθεί στα πιο σημαντικά χαρακτηριστικά του συστήματος, ενώ προηγμένες λειτουργίες, όπως η ανίχνευση σε πραγματικό χρόνο σε streaming δεδομένα, είναι εκτός πεδίου.

1.4 Δομή της πτυχιακής εργασίας

Η πτυχιακή εργασία θα οργανωθεί στα ακόλουθα κεφάλαια:

Κεφάλαιο 1 - Εισαγωγή: Παρουσίαση του προβλήματος, των στόχων και των ερευνητικών ερωτημάτων, καθώς και των περιορισμών της μελέτης.

Κεφάλαιο 2 - Ανασκόπηση Βιβλιογραφίας: Επισκόπηση της υφιστάμενης έρευνας στους τομείς της οικονομικής απάτης, των παραδοσιακών μεθόδων ανίχνευσής της και στις τεχνικές με χρήση ΤΝ.

Κεφάλαιο 3 - Θεωρητικό Πλαίσιο: Ανάλυση των βασικών εννοιών και αλγορίθμων της θεωρίας γράφων, των νευρωνικών δικτύων γράφων και των ΒΔ γράφων.

Κεφάλαιο 4 - Μεθοδολογία: Περιγραφή της ερευνητικής προσέγγισης, της συλλογής δεδομένων, του σχεδιασμού ανάπτυξης, καθώς και των τεχνικών ανάλυσης.

Κεφάλαιο 5 - Υλοποίηση: Λεπτομερής παρουσίαση της αρχιτεκτονικής του συστήματος, του κώδικα, καθώς και των τεχνολογιών που χρησιμοποιήθηκαν.

Κεφάλαιο 6 - Αποτελέσματα και Ανάλυση: Συγκριτική αξιολόγηση και ανάλυση της απόδοσης των μοντέλων.

Κεφάλαιο 7 - Συμπεράσματα και Μελλοντική Εργασία: Σύνοψη των συνεισφορών και προτάσεις για μελλοντική έρευνα.

1.5 Χρονοδιάγραμμα

Ο καταμερισμός συγγραφής των κεφαλαίων και υλοποίησης των πρακτικών θεμάτων της πτυχιακής εργασίας στις τέσσερεις γραπτές εργασίες, σύμφωνα με τη μεθοδολογία του ΕΑΠ, έγινε ως εξής:

- *1η Γραπτή Εργασία* – **Ανάλυση** – *Νοέμβριος 2025*:
  - Κεφάλαια 1, 2, 3
- *2η Γραπτή Εργασία* – **Σχεδίαση** – *Ιανουάριος 2026*
  - Κεφάλαιο 4
- *3η Γραπτή Εργασία* – **Υλοποίηση** – *Μάρτιος 2026*
  - Κεφάλαια 5, 6
- *4η Γραπτή Εργασία* – **Ολοκλήρωση** – *Μάιος 2026*
  - Κεφάλαιο 7

2. Βιβλιογραφική Ανασκόπηση

2.1 Είδη οικονομικής απάτης και προκλήσεις

Η οικονομική απάτη εμφανίζεται σε πολλές μορφές, καθεμία με τα δικά της χαρακτηριστικά και προκλήσεις ανίχνευσης. Εδώ αναφέρουμε τρεις βασικές κατηγορίες:

2.1.1 Απάτη σε πληρωμές με κάρτες

Η απάτη σε πληρωμές με κάρτες συμβαίνει όταν μη εξουσιοδοτημένα άτομα χρησιμοποιούν κλεμμένα ή πλαστά στοιχεία καρτών για να πραγματοποιήσουν αγορές. Σύμφωνα με πρόσφατες μελέτες, παγκοσμίως περισσότερα από 28 δισεκατομμύρια δολάρια χάνονται ετησίως από αυτού του είδους την απάτη. (Hilal κ.ά., 2022) Οι κύριες μορφές απάτης με κάρτες περιλαμβάνουν: (Bolton & Hand, 2002; Hilal κ.ά., 2022)

Πρώτον, την **απλή κλοπή κάρτας**, όπου ο απατεώνας τυπικά ξοδεύει όσο το δυνατόν περισσότερα χρήματα σε όσο το δυνατόν λιγότερο χρόνο, προτού η κλοπή ανιχνευθεί και η κάρτα απενεργοποιηθεί. Προφανώς, η ταχεία ανίχνευση της κλοπής μπορεί να αποτρέψει μεγάλες απώλειες.

Δεύτερον, την **απάτη αίτησης** (application fraud), η οποία προκύπτει όταν άτομα λαμβάνουν νέες πιστωτικές κάρτες από εκδότριες αρχές χρησιμοποιώντας ψευδή προσωπικά δεδομένα. Καθώς δεν απαιτούν ταχεία δράση από τον απατεώνα – σε αντίθεση με την άμεση κλοπή – οι ψευδείς αιτήσεις μπορεί να καθυστερήσουν να γίνουν αντιληπτές. Συνήθως η απάτη γίνεται εμφανής σε μεταγενέστερα στάδια του κύκλου ζωής του λογαριασμού, όπως την αποστολή της κάρτας και κυρίως τη μη έγκαιρη εξόφληση των οφειλών, αφού οι απατεώνες δεν κάνουν βεβιασμένες κινήσεις, όπως π.χ. αλλεπάλληλες συναλλαγές σε μικρό χρονικό διάστημα, ακριβώς γιατί κάτι τέτοιο θα ήγειρε υποψίες.

Τρίτον, την **απάτη "cardholder-not-present"** (CNP), όπου η συναλλαγή πραγματοποιείται εξ αποστάσεως και χρειάζονται μόνο τα στοιχεία της κάρτας – χωρίς να είναι απαραίτητη η φυσική κατοχή δηλαδή. Αυτό περιλαμβάνει τηλεφωνικές πωλήσεις και διαδικτυακές συναλλαγές, και αποτελεί μεγάλο ποσοστό των απωλειών. Η απάτη αυτή είναι δυνατή όταν τα στοιχεία της κάρτας αποκτώνται χωρίς γνώση του κατόχου, μέσω τεχνικών όπως το "skimming" (αντιγραφή της μαγνητικής λωρίδας με φορητό αναγνώστη), το

"shoulder surfing" (παρατήρηση κατά την εισαγωγή των στοιχείων) και η απάτη ταυτοπροσωπίας από δήθεν εκπροσώπους εταιρειών πιστωτικών καρτών.

2.1.2 Ξέπλυμα χρημάτων

Το ξέπλυμα χρημάτων (money laundering) αποτελεί την διαδικασία απόκρυψης της προέλευσης παράνομων κεφαλαίων μέσω πολύπλοκων χρηματοοικονομικών συναλλαγών. Περιλαμβάνει συνήθως τρία διακριτά στάδια: τοποθέτηση (placement), διαστρωμάτωση (layering) και ενσωμάτωση (integration). (Bolton & Hand, 2002; Usman κ.ά., 2023)

Κατά το στάδιο **τοποθέτησης**, παράνομα κεφάλαια εισάγονται στο χρηματοπιστωτικό σύστημα με τρόπο που δεν ελκύει την προσοχή των αρχών. Αυτό μπορεί να περιλαμβάνει την εναπόθεση μετρητών, τις μεταφορές κεφαλαίων σε τράπεζες του εξωτερικού ή άλλες μεθόδους που κάνουν τα χρήματα να φαίνονται νόμιμα.

Κατά το στάδιο **διαστρωμάτωσης**, πολλαπλές συναλλαγές και μεταφορές διεξάγονται για να αποκρύψουν τα παράνομα κεφάλαια και να περιπλέξουν την ιχνηλάτησή τους – πολύ συχνά «σπάζοντας» μεγάλα ποσά σε πολλά μικρότερα. Αυτό το στάδιο έχει γίνει ιδιαίτερα γρήγορο μετά την εισαγωγή και καθιέρωση των ψηφιακών συναλλαγών, όπου πολλές συναλλαγές και μεταφορές μπορούν να συμβούν σε σύντομο χρονικό διάστημα.

Κατά το στάδιο **ενσωμάτωσης**, τα χρήματα ενσωματώνονται στο νόμιμο χρηματοπιστωτικό σύστημα αναμιγνύοντάς τα με άλλα διαθέσιμα κεφάλαια μέσω θεμιτών χρηματοοικονομικών δραστηριοτήτων. Σε αυτό το σημείο, είναι εξαιρετικά δύσκολο για τις αρχές να εντοπίσουν την παράνομη προέλευση των κεφαλαίων.

2.1.3 Συνθετικές ταυτότητες

Η απάτη συνθετικής ταυτότητας (synthetic identity fraud) αναφέρεται στη δημιουργία ψεύτικων ταυτοτήτων συνδυάζοντας πραγματικά και πλαστά προσωπικά στοιχεία. Οι συνθετικές ταυτότητες χρησιμοποιούν συχνά έναν πραγματικό αριθμό κοινωνικής ασφάλισης (ή κάτι παρεμφερές) σε συνδυασμό με ονόματα και διευθύνσεις που δημιουργούνται τεχνητά. Αυτή η μορφή απάτης αποτελεί μία από τις ταχύτερα αναπτυσσόμενες απειλές στον χρηματοπιστωτικό τομέα, ιδιαίτερα καθώς οι απατεώνες βελτιώνουν τις τακτικές τους για να αποφύγουν την ανίχνευση. (Hodler, 2021; Leidig, 2025)

2.2 Παραδοσιακές μέθοδοι ανίχνευσης απάτης

Τα παραδοσιακά συστήματα ανίχνευσης απάτης βασίζονται κυρίως σε συστήματα κανόνων και στατιστικές μεθόδους. Ωστόσο, αυτές οι προσεγγίσεις αντιμετωπίζουν σημαντικές προκλήσεις.

2.2.1 Συστήματα βασισμένα σε κανόνες

Τα rule-based συστήματα χρησιμοποιούν προκαθορισμένους κανόνες για τον εντοπισμό ύποπτων συναλλαγών. Αυτοί οι κανόνες είναι συνήθως της μορφής: «Για μια συναλλαγή, αν ικανοποιούνται ορισμένες συνθήκες, τότε σημείωσέ την ως ύποπτη». Ένα παράδειγμα θα ήταν ο κανόνας: "Αν η συναλλαγή υπερβαίνει τα $10.000 και προέρχεται από χώρα υψηλού κινδύνου, σηματοδότησέ την". Αυτή η προσέγγιση μπορεί να είναι σχετικά απλή στην υλοποίηση και πολύ αποτελεσματική σε γνωστές, καλώς καθορισμένες απειλές. (Bolton & Hand, 2002; Usman κ.ά., 2023)

2.2.2 Στατιστικές και αναλυτικές μέθοδοι

Υπάρχουν συστήματα που χρησιμοποιούν στατιστικά μοντέλα για τον εντοπισμό ανωμαλιών στα δεδομένα συναλλαγών. Αυτά τα μοντέλα έχουν αποδειχθεί πιο ευέλικτα από τα συστήματα βασισμένα σε κανόνες. Παραδείγματα στατιστικών τεχνικών περιλαμβάνουν τη γραμμική διακριτική ανάλυση (Linear Discriminant Analysis), τη λογιστική παλινδρόμηση (Logistic Regression) και άλλες κλασικές μεθόδους ταξινόμησης, όπως τα Δέντρα Απόφασης (Decision Trees). (Bolton & Hand, 2002)

2.2.3 Περιορισμοί και Αδυναμίες

Ενώ τα συστήματα βασισμένα σε κανόνες είναι διαφανή και ερμηνεύσιμα (Liu κ.ά., 2021), έχουν σημαντικούς περιορισμούς. Πρώτον, αδυνατούν να προσαρμοστούν σε νέα μοτίβα απάτης και απαιτούν συνεχείς ενημερώσεις των κανόνων καθώς οι απατεώνες προσαρμόζουν τις τακτικές τους. Κάθε φορά που ανακαλύπτεται μια νέα τακτική, πρέπει να ενημερωθούν οι κανόνες του συστήματος. (Hilal κ.ά., 2022; Motie & Raahemi, 2024) Δεύτερον, παράγουν υψηλά ποσοστά ψευδώς θετικών αποτελεσμάτων, με αποτέλεσμα η ανάλυση και ο προσδιορισμός των πραγματικών απειλών να καθίσταται χρονοβόρα και δαπανηρή. (Usman κ.ά., 2023) Τρίτον, δεν μπορούν να εντοπίσουν εξελιγμένα σχήματα που

εκτείνονται σε πολλαπλούς λογαριασμούς και συναλλαγές, εφόσον αναλύουν κάθε συναλλαγή ή λογαριασμό ξεχωριστά, αγνοώντας τη δομή του δικτύου που αυτά σχηματίζουν. (X. Zhang, 2025) Μελέτες πάνω στην κοινωνική δικτύωση δείχνουν ότι είναι δυνατό να γίνουν καλύτερες προβλέψεις σχετικά με τους ανθρώπους, εξετάζοντας όλες τις πληροφορίες από τους φίλους τους και τους φίλους των φίλων τους, παρά εξετάζοντας τις πληροφορίες για τους ίδιους. (Hodler, 2021)

2.3 Μηχανική μάθηση στην ανίχνευση απάτης

2.3.1 Μάθηση με επίβλεψη (Supervised learning)

Η μάθηση με επίβλεψη είναι μια κατηγορία αλγορίθμων μηχανικής μάθησης που χρησιμοποιούν σημειωμένα δεδομένα (π.χ. με ετικέτες ή βαθμολογίες) για την εκπαίδευση μοντέλων. Στο πλαίσιο της ανίχνευσης απάτης, τα μοντέλα εκπαιδεύονται σε ήδη γνωστά παραδείγματα τόσο νόμιμων όσο και δόλιων συναλλαγών, και στη συνέχεια χρησιμοποιούνται για την ταξινόμηση νέων συναλλαγών. (Bolton & Hand, 2002)

Τα rule-based μοντέλα τα οποία είδαμε στην προηγούμενη ενότητα, βασίζονται στην πραγματικότητα σε supervised learning τεχνικές. Άλλα δημοφιλή μοντέλα αυτής της κατηγορίας που έχουν χρησιμοποιηθεί για ανίχνευση απάτης είναι:

Random Forest: Ένα ensemble μοντέλο (*βλ. υποενότητα 2.3.4*) που δημιουργεί πολλαπλά δέντρα απόφασης και συνδυάζει τις προβλέψεις τους. Έχει αποδειχθεί ιδιαίτερα αποτελεσματικό στην ανίχνευση απάτης καθώς μπορεί να χειριστεί μη γραμμικές σχέσεις και να κρατήσει υψηλή ακρίβεια ακόμη και με ανισόρροπα δεδομένα. (Bin Sulaiman κ.ά., 2022; Chen κ.ά., 2025)

XGBoost (Extreme Gradient Boosting): Επίσης ensemble μοντέλο που δημιουργεί διαδοχικά νέα δέντρα που διορθώνουν τα σφάλματα των προηγούμενων δέντρων. Το XGBoost είναι γνωστό για την υψηλή ακρίβεια και την ικανότητά του να χειρίζεται μη γραμμικές σχέσεις και αλληλεπιδράσεις χαρακτηριστικών. (Hilal κ.ά., 2022; X. Zhang, 2025)

Support Vector Machines (SVM): Ένας αλγόριθμος που προσπαθεί να βρει το κατάλληλο υπερεπίπεδο που διαχωρίζει τις νόμιμες και δόλιες συναλλαγές στον χώρο χαρακτηριστικών. (Bin Sulaiman κ.ά., 2022; Hilal κ.ά., 2022)

Νευρωνικά Δίκτυα: Μοντέλα που μιμούνται την δομή του ανθρώπινου εγκεφάλου και μπορούν να μάθουν πολύπλοκα πρότυπα στα δεδομένα. Τα νευρωνικά δίκτυα έχουν αποδειχθεί ιδιαίτερα δυνατά στη μάθηση σύνθετων μοτίβων. (Hilal κ.ά., 2022)

Οι μέθοδοι αυτές απαιτούν μεγάλα σύνολα σημειωμένων δεδομένων και συχνά αντιμετωπίζουν προβλήματα με ανισορροπημένα datasets (όπου δηλαδή η πλειονότητα των συναλλαγών είναι νόμιμες). Η ανισορροπία κλάσεων αποτελεί σημαντική πρόκληση, καθώς τα μοντέλα τείνουν να δίνουν προτεραιότητα στην πλειονοτική κλάση, οδηγώντας σε χαμηλή απόδοση στην ανίχνευση απάτης. (Ounacer κ.ά., 2018; Hilal κ.ά., 2022)

2.3.2 Μάθηση χωρίς επίβλεψη (Unsupervised Learning)

Η μάθηση χωρίς επίβλεψη αναζητά κρυφά πρότυπα σε δεδομένα που δεν έχουν σημειωθεί. Δεν απαιτεί ετικέτες εκπαίδευσης, αλλά αναζητά μοτίβα και δομές εγγενείς στα δεδομένα. Στο πλαίσιο της ανίχνευσης απάτης, αυτή η προσέγγιση χρησιμοποιείται για τον εντοπισμό ανωμαλιών ή ασυνήθιστων μοτίβων που ενδέχεται να υποδηλώνουν δόλιες δραστηριότητες. (Ounacer κ.ά., 2018; Hilal κ.ά., 2022)

Κοινές μέθοδοι unsupervised learning για ανίχνευση απάτης αποτελούν:

K-Means Clustering: Ένας αλγόριθμος που ομαδοποιεί τα δεδομένα σε *k* διακριτές ομάδες βάσει της ομοιότητας. Οι συναλλαγές που δεν ταιριάζουν σε καμιά ομάδα μπορούν να θεωρηθούν ως ανωμαλίες. (Ounacer κ.ά., 2018)

Isolation Forest: Ένας αλγόριθμος που απομονώνει τις ανωμαλίες με τη δημιουργία δέντρων απόφασης που χωρίζουν επανειλημμένα το σύνολο δεδομένων. Οι ανωμαλίες απομονώνονται συνήθως πιο γρήγορα από τα κανονικές δεδομένα. (Hilal κ.ά., 2022)

One-Class SVM: Μια παραλλαγή των SVM που προσπαθεί να βρει το όριο που περικλείει τα «κανονικά» δεδομένα, θεωρώντας τα δεδομένα έξω από αυτό το όριο ως ανωμαλίες. (Hilal κ.ά., 2022)

Οι μέθοδοι unsupervised learning είναι ιδιαίτερα χρήσιμες όταν δεν υπάρχουν διαθέσιμα σημειωμένα δεδομένα ή/και όταν απαιτείται η ανίχνευση νέων, άγνωστων τύπων απάτης.

2.3.3 Βαθιά μάθηση (Deep Learning)

Η βαθιά μάθηση αναφέρεται σε νευρωνικά δίκτυα με πολλαπλά στρώματα που μπορούν να μάθουν πολύπλοκες αναπαραστάσεις δεδομένων. Στον τομέα της ανίχνευσης απάτης, τα μοντέλα deep learning έχουν δείξει εξαιρετική απόδοση. Βασικές αρχιτεκτονικές deep learning για ανίχνευση απάτης περιλαμβάνουν:

Recurrent Neural Networks (RNNs): Νευρωνικά δίκτυα που έχουν "μνήμη" και μπορούν να επεξεργάζονται ακολουθιακά δεδομένα. Αυτά είναι ιδιαίτερα χρήσιμα για τη μοντελοποίηση χρονοσειρών (time-series) συναλλαγών. (Chen κ.ά., 2025; Wang κ.ά., 2022)

Autoencoders: Νευρωνικά δίκτυα που μαθαίνουν να συμπιέζουν τα δεδομένα σε μια χαμηλότερης διάστασης αναπαράσταση και στη συνέχεια να τα ανακατασκευάζουν. Οι ανωμαλίες συχνά έχουν υψηλό σφάλμα ανακατασκευής, και σύμφωνα με αυτό, λαμβάνουν αντίστοιχη βαθμολογία για την κατάταξή τους ως τέτοιες. (Hilal κ.ά., 2022)

Graph Neural Networks (GNNs): Νευρωνικά δίκτυα που λειτουργούν σε δεδομένα γράφων, ιδιαίτερα χρήσιμα για τη μοντελοποίηση του δικτύου συναλλαγών και την ανίχνευση ύποπτων δεδομένων δικτύου απάτης. (Wang κ.ά., 2022) Παρ΄ ότι αποδεικνύονται ιδιαίτερα αποτελεσματικά, τα μοντέλα βαθιάς μάθησης χαρακτηρίζονται από έλλειψη ερμηνευσιμότητας (αποτελούν «μαύρο κουτί»), κάτι που αποτελεί μειονέκτημα για τη χρήση τους σε ρυθμιζόμενα περιβάλλοντα. (Hilal κ.ά., 2022) Η δυσκολία εξήγησης των αποφάσεών τους, μπορεί να δημιουργήσει προβλήματα συμμόρφωσης με κανονιστικές απαιτήσεις, όπως είναι το GDPR. (Chen κ.ά., 2025) Παρουσιάζονται αναλυτικότερα στο επόμενο κεφάλαιο.

2.3.4 Ensemble Methods

Οι ensemble methods είναι μέθοδοι μηχανικής μάθησης που συνδυάζουν πολλαπλά μοντέλα για να επιτύχουν καλύτερη απόδοση από ότι ένα μεμονωμένο μοντέλο θα μπορούσε να επιτύχει. Η βασική ιδέα είναι ότι τα διαφορετικά μοντέλα κάνουν διαφορετικά σφάλματα, και η άθροισή τους μέσω διαφόρων τεχνικών (όπως π.χ. ψηφοφορία, μέσος όρος, χρήση μετα-μοντέλου), προκαλεί τη διόρθωση πολλών από αυτά τα σφάλματα. Χρησιμοποιούνται ευρέως επειδή μπορούν να μειώσουν τη διακύμανση (variance) και την

προκατάληψη (bias), να βελτιώσουν τη γενίκευση (generalization) και γενικά να επιτύχουν στην πράξη σε πολύπλοκες εργασίες του πραγματικού κόσμου.

Τα σύνολα αυτά (ensembles) μπορούν να είναι ομοιογενή, δηλαδή να περιλαμβάνουν ίδιο βασικό τύπο μοντέλου (base learner), ή ετερογενή, να συνδυάζουν δηλαδή διάφορους τύπους μοντέλων/αλγορίθμων. (Mohammed & Kora, 2023)

2.3.5 Σύγκριση μετρικών

Η αξιολόγηση της απόδοσης των μοντέλων ανίχνευσης απάτης απαιτεί τη χρήση κατάλληλων μετρικών. Οι πιο κοινές μετρικές για ανίχνευση απάτης, οι οποίες συναντώνται συχνότερα στη βιβλιογραφία, είναι οι εξής: (Chen κ.ά., 2025)

Precision (Ακρίβεια): Το ποσοστό των προβλεπόμενων θετικών αποτελεσμάτων (ανιχνευμένες απάτες) που είναι πραγματικά θετικά.

Recall (Ανάκληση): Το ποσοστό των πραγματικών θετικών (πραγματικές απάτες) που ανιχνεύθηκαν σωστά.

F1-Score: Ο αρμονικός μέσος της ακρίβειας και της ανάκλησης, που παρέχει μια ισορροπημένη μέτρηση απόδοσης.

AUC-ROC (Area Under the Receiver Operating Characteristic Curve): Μια μέτρηση που αναπαριστά την ικανότητα ενός μοντέλου να διαχωρίσει τις νόμιμες από τις δόλιες συναλλαγές σε διαφορετικά κατώφλια ταξινόμησης.

AUC-PR (Area Under the Precision/Recall Curve): Είναι πιο κατάλληλη μέτρηση για ανισόρροπα δεδομένα, καθώς αφορά την κλάση των θετικών αποτελεσμάτων (απάτες).

3. Θεωρητικό Πλαίσιο

3.1 Βασικές έννοιες θεωρίας γράφων

Η θεωρία γράφων είναι ένας κλάδος των διακριτών μαθηματικών που ασχολείται με τη μελέτη των γράφων, οι οποίοι είναι μαθηματικές δομές αποτελούμενες από κόμβους και ακμές. Οι γράφοι μοντελοποιούν με πολύ βολικό τρόπο τις συσχετίσεις μεταξύ αντικειμένων και έχουν ευρεία εφαρμογή στην Επιστήμη των Υπολογιστών, καθώς και σε άλλα πεδία και επιστήμες. Ουσιαστικά, μέσω της θεωρίας γράφων μπορεί να αναλυθεί οποιοδήποτε δίκτυο, αφού στην πράξη οι έννοιες γράφος και δίκτυο μπορούν να θεωρηθούν ταυτόσημες. Στο πλαίσιο της οικονομικής απάτης, οι κόμβοι αναπαριστούν οντότητες (λογαριασμούς, χρήστες, συσκευές), ενώ οι ακμές αναπαριστούν σχέσεις (συναλλαγές, κοινά χαρακτηριστικά). Η αναπαράσταση χρηματοπιστωτικών δεδομένων ως γράφου, επιτρέπει την ανάλυση της δομής του δικτύου και την αποκάλυψη κρυφών σχέσεων.

3.1.1 Ορισμός και Αναπαράσταση Γράφου

Ένας **γράφος** G ορίζεται ως ένα διατεταγμένο ζεύγος G = (V, E), όπου:

- V είναι ένα σύνολο **κόμβων** (nodes)
- E είναι ένα σύνολο **ακμών** (edges) μεταξύ ζευγών κορυφών

Κάθε ακμή $e \in E$ μπορεί να αναπαρασταθεί ως $e = (u, v)$, όπου $u, v \in V$.

Επιπλέον, οι κόμβοι και οι ακμές μπορούν να έχουν σχετικά χαρακτηριστικά ή ιδιότητες, που ονομάζονται **χαρακτηριστικά κόμβων** (node features) και **χαρακτηριστικά ακμών** (edge features). Τα χαρακτηριστικά κόμβων μπορούν να περιλαμβάνουν πληροφορίες όπως ο τύπος λογαριασμού, η ιστορία συναλλαγών ή το υπόλοιπο λογαριασμού. Τα χαρακτηριστικά ακμών μπορούν να περιλαμβάνουν πληροφορίες όπως το ποσό της συναλλαγής, ο τύπος συναλλαγής ή η χρονική σφραγίδα.

Μια συχνή αναπαράσταση είναι μέσω του πίνακα γειτνίασης. Ο **πίνακας γειτνίασης** A είναι ένας $n \times n$ πίνακας όπου $A[i,j] = 1$ αν υπάρχει ακμή μεταξύ των κόμβων i και j, και 0 διαφορετικά (για απλούς γράφους). Για γράφους με βάρη (βλ. επόμενη υποενότητα), το $A[i,j]$ περιέχει το βάρος της ακμής. (Huang κ.ά., 2009; Wang κ.ά., 2022)

3.1.2 Τύποι Γράφων

Μη Κατευθυνόμενοι Γράφοι (Undirected Graphs)

Σε έναν μη κατευθυνόμενο γράφο, οι ακμές δεν έχουν κατεύθυνση. Δηλαδή, η ακμή (u, v) είναι ίδια με την ακμή (v, u). Αυτό σημαίνει ότι η σχέση μεταξύ των κορυφών είναι αμφίδρομη. (Wang κ.ά., 2022) Οι μη κατευθυνόμενοι γράφοι χρησιμοποιούνται για να αναπαραστήσουν αμφίδρομες σχέσεις, όπως π.χ. η φιλία στα κοινωνικά δίκτυα.

Κατευθυνόμενοι Γράφοι (Directed Graphs)

Σε έναν κατευθυνόμενο γράφο, οι ακμές έχουν κατεύθυνση, που υποδηλώνεται με ένα βέλος. Η ακμή (u, v) είναι διαφορετική από την ακμή (v, u). Αυτό σημαίνει ότι η σχέση μεταξύ των κορυφών έχει μια συγκεκριμένη κατεύθυνση. (Wang κ.ά., 2022) Οι κατευθυνόμενοι γράφοι είναι ιδανικοί για την αναπαράσταση της κατεύθυνσης των χρηματοοικονομικών ροών, όπου η ακμή (u, v) αναπαριστά τη ροή χρημάτων από τον λογαριασμό u στον λογαριασμό v. (Motie & Raahemi, 2024)

Γράφοι με Βάρη (Weighted Graphs)

Αυτοί οι γράφοι έχουν βάρη που σχετίζονται με τις ακμές, που αντιπροσωπεύουν το κόστος, το μήκος ή την ικανότητα της σύνδεσης. Για παράδειγμα, τα βάρη των ακμών μπορούν να αναπαριστούν τα ποσά των συναλλαγών. (Motie & Raahemi, 2024)

Ετερογενείς Γράφοι (Heterogeneous Graphs)

Ετερογενείς γράφοι είναι γράφοι με πολλαπλούς τύπους κόμβων και ακμών. Σε έναν ετερογενή γράφο, υπάρχουν οι συναρτήσεις χαρτογραφίας τύπου κόμβου $\theta : V \rightarrow TV$ και τύπου ακμής $\xi : E \rightarrow TE$, όπου TV και TE είναι τα σύνολα τύπων κόμβων και ακμών, αντίστοιχα. (Motie & Raahemi, 2024) Στο πλαίσιο της ανίχνευσης απάτης, διαφορετικοί τύποι κόμβων μπορούν να αναπαριστούν διαφορετικές οντότητες (π.χ., πελάτες, λογαριασμοί, συσκευές, έμποροι), και διαφορετικοί τύποι ακμών διαφορετικούς τύπους σχέσεων (π.χ., συναλλαγές, κοινή συσκευή, κοινή διεύθυνση IP). (Patil κ.ά., 2024)

Διμερείς Γράφοι (Bipartite Graphs)

Ένας ειδικός τύπος ετερογενούς γράφου, όπου οι κόμβοι χωρίζονται σε δύο διακριτά σύνολα (U, W), και όπου ακμές υπάρχουν μόνο μεταξύ κόμβων διαφορετικών τύπων (π.χ. χρήστης – προϊόν ή λογαριασμός – συσκευή). (Wang κ.ά., 2022)

Χρονικοί και δυναμικοί γράφοι (Temporal and Dynamic Graphs)

Οι χρονικοί γράφοι έχουν χαρακτηριστικά που αλλάζουν με το χρόνο, αλλά η δομή του γράφου (κόμβοι, ακμές) παραμένει σταθερή. Οι δυναμικοί γράφοι αλλάζουν τόσο τη δομή τους (νέοι κόμβοι και ακμές προστίθενται ή αφαιρούνται), όσο και τα χαρακτηριστικά τους, με την πάροδο του χρόνου. (Motie & Raahemi, 2024)

3.1.3 Μέτρα κεντρικότητας

Τα μέτρα κεντρικότητας είναι μαθηματικά μέτρα που ποσοτικοποιούν τη σημαντικότητα ή την «επιρροή» ενός κόμβου (ή ακμής) σε έναν γράφο. Ακολουθούν μερικά από τα πιο σημαντικά:

Κεντρικότητα Βαθμού (Degree Centrality)

Ένα απλό μέτρο που μετρά τον αριθμό των απευθείας συνδέσεων ενός κόμβου. (J. Zhang & Luo, 2017) Στο πλαίσιο της ανίχνευσης απάτης, ένα υψηλό degree centrality μπορεί να υποδηλώνει έναν κόμβο που είναι συνδεδεμένος με πολλούς άλλους κόμβους, πιθανώς υποδηλώνοντας σύνδεσμο σε ένα δίκτυο απάτης. (Leidig, 2025)

Κεντρικότητα Ενδιάμεσης Θέσης (Betweenness Centrality)

Ένα μέτρο που ποσοτικοποιεί τον αριθμό των συντομότερων μονοπατιών που περνούν μέσα από ένα κόμβο. (J. Zhang & Luo, 2017) Ένα κόμβος με υψηλή κεντρικότητα ενδιάμεσης είναι ένας σημαντικός ενδιάμεσος που συνδέει διαφορετικές ομάδες κόμβων. Στο πλαίσιο ανίχνευσης απάτης, αυτό μπορεί να υποδηλώνει ένα "mule" λογαριασμό που μεσολαβεί στη μεταφορά χρημάτων μεταξύ δύο απατηλών δικτύων. (Hodler, 2021; Leidig, 2025)

PageRank Centrality

Ένας αλγόριθμος που υπολογίζει τη σημαντικότητα ενός κόμβου βάσει του αριθμού και της ποιότητας των κόμβων που συνδέουν με αυτό. Ο PageRank χρησιμοποιήθηκε αρχικά από τη Google για την κατάταξη σελίδων ιστού, αλλά έχει εφαρμοστεί επιτυχώς στην ανίχνευση απάτης για τον εντοπισμό κεντρικών κόμβων σε απατηλά δίκτυα. (Molloy κ.ά., 2017)

Κεντρικότητα Ιδιοδιανύσματος (Eigenvector Centrality)

Ένα μέτρο που ποσοτικοποιεί την σημαντικότητα ενός κόμβου βάσει του αριθμού και της σημαντικότητας των κόμβων με τους οποίους συνδέεται. Ένας κόμβος έχει υψηλή κεντρικότητα ιδιοδιανύσματος εάν συνδέεται με άλλους κόμβους που έχουν επίσης υψηλή κεντρικότητα ιδιοδιανύσματος. (Modepalli, 2025)

3.1.3 Ανίχνευση κοινοτήτων

Η ανίχνευση κοινοτήτων (community detection) είναι μια διαδικασία που στοχεύει στον εντοπισμό ομάδων κόμβων που είναι πυκνά συνδεδεμένες μεταξύ τους, αλλά χαλαρά συνδεδεμένες με άλλες ομάδες. Στο πλαίσιο της ανίχνευσης απάτης είναι κρίσιμη, καθώς οι κοινότητες μπορούν να αναπαριστούν δίκτυα απάτης ή ομάδες συναλλαγών που σχετίζονται μεταξύ τους. (Hodler, 2021) Ορισμένοι δημοφιλείς αλγόριθμοι ανίχνευσης κοινοτήτων είναι οι εξής:

Louvain Algorithm

Ένας αλγόριθμος που βελτιστοποιεί την «αρθρωτότητα» (modularity) του γράφου για τον εντοπισμό κοινοτήτων. Ξεκινά με κάθε κόμβο ως μια ξεχωριστή κοινότητα και στη συνέχεια συνδυάζει κοινότητες αυξάνοντας/μεγιστοποιώντας τη δομή. Ο Louvain έχει δείξει εξαιρετική απόδοση σε εφαρμογές ανίχνευσης απάτης, με ορισμένες μελέτες να δείχνουν ακρίβεια έως 92% όταν συνδυάζεται με μοντέλα μηχανικής μάθησης (XGBoost). (X. Zhang, 2025)

Label Propagation

Ένας πολύ απλός αλγόριθμος που επιτρέπει στους κόμβους να υιοθετήσουν την πιο κοινή ετικέτα μεταξύ των γειτόνων τους. Αυτό έχει ως αποτέλεσμα τον σχηματισμό κοινοτήτων με έναν «φυσικό» τρόπο. Είναι υπολογιστικά αποδοτικός και λειτουργεί καλά σε μεγάλης κλίμακας δίκτυα. (Sherwin Akshay κ.ά., 2024)

Girvan-Newman Algorithm

Ένας αλγόριθμος που χρησιμοποιεί την κεντρικότητα ενδιάμεσης θέσης της ακμής για να αφαιρεί επαναληπτικά ακμές που συνδέουν διαφορετικές κοινότητες μεταξύ τους, σχηματίζοντας προοδευτικά μικρότερες συστάδες (clusters). Αν και υπολογιστικά πιο ακριβός από τον Louvain, παρέχει υψηλής ποιότητας ανίχνευση κοινοτήτων. (Sherwin Akshay κ.ά., 2024)

3.1.4 Ανάλυση διαδρομών

Η ανάλυση διαδρομών αποτελεί κεντρικό αντικείμενο στη θεωρία γράφων, καθώς επιτρέπει τον εντοπισμό βέλτιστων (ή συγκεκριμένων) μονοπατιών μεταξύ ζευγών κόμβων σε δίκτυα με πολύπλοκη δομή. Στο πεδίο της ανίχνευσης οικονομικής απάτης, μπορούν να αξιοποιηθούν για τον εντοπισμό λογαριασμών που βρίσκονται «κοντά» ή διαμεσολαβούν σε δίκτυα ύποπτης δραστηριότητας, συμβάλλοντας στον εντοπισμό κρυφών σχέσεων και μοτίβων ξεπλύματος χρήματος. (Hodler, 2021)

Μερικοί σημαντικοί αλγόριθμοι είναι οι παρακάτω: (Huang κ.ά., 2009)

Αναζήτηση κατά πλάτος (Breadth-First Search – BFS)

Η BFS βρίσκει το συντομότερο μονοπάτι σε μη σταθμισμένους γράφους ή σε γράφους όπου όλα τα βάρη μπορούν να θεωρηθούν ίσα. Σε έναν γράφο συναλλαγών, για παράδειγμα, μπορεί να δώσει ένα γρήγορο δείκτη εγγύτητας προς ύποπτα δίκτυα. (‘Using Graph Machine Learning to Improve Fraud Detection Rates’, 2023)

Dijkstra

Ο αλγόριθμος του Dijkstra υπολογίζει τη συντομότερη διαδρομή από μία πηγή προς όλους τους κόμβους, σε γράφους με μη αρνητικά βάρη, με υψηλή αποδοτικότητα σε μεγάλες δομές. Σε περιβάλλοντα οικονομικής απάτης, τα βάρη μπορούν να κωδικοποιούν ρίσκο, ποσά ή χρονικές καθυστερήσεις, επιτρέποντας τον ποσοτικό υπολογισμό του «κοντινότερου» μονοπατιού μεταξύ ενός λογαριασμού και μιας οντότητας σε “blacklist”, κάτι που τροφοδοτεί χαρακτηριστικά κινδύνου σε μοντέλα ανίχνευσης. (DataWalk, 2022)

Bellman–Ford

Ο Bellman–Ford επιλύει το πρόβλημα της συντομότερης διαδρομής από μία πηγή προς τους υπόλοιπους κόμβους, σε γράφους με θετικά ή αρνητικά βάρη, και μπορεί επίσης να ανιχνεύσει αρνητικούς κύκλους. Είναι ωστόσο υπολογιστικά «ακριβότερος» από τον Dijkstra. Σε χρηματοοικονομικά δίκτυα, η δυνατότητα χειρισμού «ανώμαλων» βαρών και εντοπισμού αρνητικών κύκλων τον καθιστά κατάλληλο για σενάρια όπου υπάρχουν αλυσιδωτές συναλλαγές με μη φυσιολογικές αποδόσεις, τα οποία συχνά συνδέονται με ύποπτη ή καταχρηστική συμπεριφορά. (Arnau κ.ά., 2024)

3.2 Graph Neural Networks (GNNs): Θεωρία και Αρχιτεκτονικές

Τα Νευρωνικά Δίκτυα Γράφων αποτελούν μια κατηγορία βαθιάς μάθησης που επεκτείνει τις παραδοσιακές νευρωνικές αρχιτεκτονικές για την επεξεργασία δεδομένων δομημένων σε μορφή γράφου. Η σημαντικότερη καινοτομία των GNNs είναι η ικανότητά τους να μοντελοποιούν ευέλικτα τις μη-Ευκλείδειες σχέσεις ανάμεσα στα στοιχεία δεδομένων, κάτι που τα παραδοσιακά νευρωνικά δίκτυα δεν μπορούν να κάνουν αποτελεσματικά.

3.2.1 Message Passing

Τα GNNs λειτουργούν χρησιμοποιώντας ένα πλαίσιο που ονομάζεται "message passing". Σε αυτό το πλαίσιο, κάθε κόμβος επικοινωνεί με τους γείτονές του, περνώντας μηνύματα που περιέχουν πληροφορίες σχετικές με το κόμβο. (Motie & Raahemi, 2024) Περιλαμβάνει τρία κύρια στάδια:

1. **Message Construction**: Για κάθε ακμή (u, v), δημιουργείται ένα μήνυμα που συνδυάζει τα χαρακτηριστικά του κόμβου u (το "source" του μηνύματος), τα χαρακτηριστικά του κόμβου v (το "target") και τα χαρακτηριστικά της ακμής.
2. **Message Aggregation**: Τα μηνύματα από όλους τους γείτονες ενός κόμβου συνδυάζονται σε μια ενιαία αναπαράσταση. Αυτό μπορεί να γίνει χρησιμοποιώντας διαφορές συναρτήσεις συνάθροισης όπως άθροιση, μέσο ή μέγιστο.
3. **Node Update**: Το συναθροισμένο μήνυμα χρησιμοποιείται για να ενημερώσει τα χαρακτηριστικά του κόμβου. Αυτή η διαδικασία επαναλαμβάνεται για πολλές επαναλήψεις, επιτρέποντας στο δίκτυο να συλλέξει πληροφορίες από απομακρυσμένους κόμβους.

3.2.2 Graph Convolutional Networks (GCNs)

Τα Συνελικτικά Δίκτυα Γράφων αποτελούν τη πιο θεμελιώδη και ευρέως χρησιμοποιούμενη κατηγορία GNN. Το βασικό μοντέλο GCN προτάθηκε από τους Kipf και Welling το 2017 και έχει καταστεί ένα βασικό εργαλείο για εργασίες κατάταξης κόμβων σε γράφους. (Motie & Raahemi, 2024)

Ο κανόνας διάδοσης πληροφοριών σε κάθε επίπεδο ενός GCN περιγράφεται μαθηματικά ως:

$$H^{(l+1)} = \sigma\left(\hat{D}^{-\frac{1}{2}}\hat{A}\,\hat{D}^{-\frac{1}{2}}H^{(l)}W^{(l)}\right)$$

όπου $\hat{A} = A + I_N$ είναι ο κανονικοποιημένος πίνακας γειτνίασης με προσθήκη αυτο-βρόχων (self-loops), $\hat{D}$ είναι ο πίνακας βαθμών της $\hat{A}$, $H^{(l)}$ είναι ο πίνακας χαρακτηριστικών κόμβων στο επίπεδο l, $W^{(l)}$ είναι ένας πίνακας βαρών προς εκπαίδευση, και σ είναι μια συνάρτηση ενεργοποίησης (συνήθως ReLU).

Τα GCNs είναι αποδοτικά στη μάθηση χαρακτηριστικών και αναγνωρίζουν μοτίβα σε οικονομικά δεδομένα, μέσω της αξιοποίησης της τοπικής συνδεσιμότητας και της αυτόματης διάκρισης των χαρακτηριστικών κόμβων και ακμών. Ωστόσο, τα GCNs μπορεί να υποφέρουν από over-smoothing, με αποτέλεσμα την απώλεια των διακριτών χαρακτηριστικών των κόμβων, και είναι επιρρεπή σε overfitting (υπερεκπαίδευση) όταν τα διαθέσιμα δεδομένα είναι περιορισμένα. (Cheng κ.ά., 2025)

3.2.3 Graph Attention Networks (GATs)

Τα Δίκτυα Προσοχής σε Γράφους εισάγουν τον μηχανισμό της προσοχής (attention mechanism) στο πεδίο των GNNs. Αντί να χρησιμοποιούν ομοιόμορφα βάρη για όλους τους γείτονες ενός κόμβου, τα GATs μαθαίνουν να προσαρμόζουν δυναμικά τα βάρη προσοχής ανάλογα με τη σχετικότητα κάθε γείτονα. (Cheng κ.ά., 2025)

Οι συντελεστές προσοχής υπολογίζονται ως:

$$\alpha_{ij} = \frac{\exp\left(\text{LeakyReLU}\left(a^T\left[Wh_i \parallel Wh_j\right]\right)\right)}{\sum_{k\in N_i}\exp\left(\text{LeakyReLU}\left(a^T\left[Wh_i \parallel Wh_k\right]\right)\right)}$$

όπου α_{ij} αναπαριστά τη σημασία των χαρακτηριστικών του κόμβου j ως προς τον κόμβο i, h_i και h_j είναι οι ενσωματώσεις (embeddings) των κόμβων, N_i είναι το σύνολο των γειτόνων του i, W είναι ένας πίνακας βαρών προς εκπαίδευση, και $\parallel$ συμβολίζει τη συνένωση.

Το σημαντικό πλεονέκτημα αυτής της προσέγγισης είναι ότι ο χαρακτήρας κάθε γείτονα συνεισφέρει διαφορετικά στη συνολική αναπαράσταση ενός κόμβου, και έτσι το δίκτυο μπορεί να μάθει ποιοι γείτονες είναι σημαντικότεροι για την πρόβλεψη, ανάλογα την περίπτωση. (Motie & Raahemi, 2024)

3.2.4 Heterogeneous GNNs (HGNNs)

Τα Ετερογενή Νευρωνικά Δίκτυα Γράφων έχουν σχεδιαστεί για να χειρίζονται γραφήματα που περιέχουν πολλούς τύπους κόμβων και άκρων, επιτρέποντας στο μοντέλο να λαμβάνει ένα πλουσιότερο σύνολο σχέσεων και ιδιοτήτων από τα δεδομένα.

Για την επεξεργασία ετερογενών γραφημάτων, τα HGNNs συχνά χρησιμοποιούν έναν μηχανισμό μετασχηματισμού ξεχωριστό για κάθε τύπο, ο οποίος μπορεί να διατυπωθεί μαθηματικά ως:

$$H_a^{(k+1)} = \sigma\left(\sum_{r\in\mathcal{R}}\sum_{b\in\text{A}} W_{r,a,b}^{(k)} \cdot \text{AGG}\left(H_b^{(k)}, E_{r,a,b}\right)\right)$$

όπου $H_a^{(k)}$ αντιπροσωπεύει τα χαρακτηριστικά κόμβων τύπου a στο επίπεδο k, $W_{r,a,b}^{(k)}$ είναι ο πίνακας βαρών για τη σχέση r μεταξύ των τύπων κόμβων a και b, $E_{r,a,b}$ υποδηλώνει τις

ακμές τύπου r μεταξύ των τύπων κόμβων a και b, και AGG() είναι μια συνάρτηση συσσώρευσης (aggregation function) που συνδυάζει χαρακτηριστικά από γειτονικούς κόμβους. Η σ είναι μια μη-γραμμική συνάρτηση ενεργοποίησης. Συνήθεις επιλογές για το AGG() περιλαμβάνουν άθροισμα, μέσο όρου και μηχανισμούς προσοχής, καθένας από τους οποίους παρέχει διαφορετικούς τρόπους για να τονίσει τις συνεισφορές των γειτόνων.

Αυτός ο μηχανισμός διασφαλίζει ότι διαφορετικοί τύποι σχέσεων μπορούν να συμβάλλουν διαφορετικά στη σχηματισμό των αναπαραστάσεων κόμβων, κάτι πολύ σημαντικό στα χρηματοπιστωτικά δίκτυα, τα οποία, όπως έχουμε αναφέρει παραπάνω, μπορεί να περιλαμβάνουν πολλαπλούς τύπους οντοτήτων και πολλαπλούς τύπους σχέσεων.

Θα πρέπει να αναφερθεί ωστόσο, ότι η υψηλή πολυπλοκότητα των HGNNs τα καθιστά απαιτητικά σε πόρους, κάτι που μπορεί να επηρεάσει την απόδοση και εν γένει την καταλληλότητά τους όσον αφορά εφαρμογές εστιασμένες στην επεκτασιμότητα (scalability). (Cheng κ.ά., 2025)

3.2.5 GraphSAGE

Το GraphSAGE (Graph Sample and Aggregate) είναι ένα GNN που χρησιμοποιεί δειγματοληψία στα χαρακτηριστικά της γειτονιάς ενός κόμβου και τα συναθροίζει για να δημιουργήσει μια αναπαράσταση. Αντί να κοιτάει όλους τους γείτονες, το μοντέλο δειγματοληπτεί ένα σταθερό αριθμό γειτόνων σε κάθε επανάληψη. (Hamilton κ.ά., 2017)

Ο κανόνας διάδοσης του GraphSAGE είναι:

$$h_v^{(k)} = \sigma\left(W^{(k)} \cdot \text{CONCAT}\left(h_v^{(k-1)}, \text{AGG}\left(\left\{h_u^{(k-1)}, \forall u \in \mathcal{N}(v)\right\}\right)\right)\right)$$

όπου $h_v^{(k)}$ είναι η αναπαράσταση του κόμβου v στο επίπεδο k, $\mathcal{N}(v)$ το σύνολο γειτόνων του v, και $W^{(k)}$ ο πίνακας βαρών του επιπέδου k.

Εκεί όπου υστερούν άλλες αρχιτεκτονικές, όπως τα HGNNs που είδαμε παραπάνω, το GraphSAGE πλεονεκτεί – καθώς δεν απαιτεί τη γνώση του ολόκληρου γράφου, παρέχει επεκτασιμότητα, γεγονός που το καθιστά καλή επιλογή για μεγάλα δίκτυα. Επιπλέον, και για τον ίδιο λόγο, είναι ιδιαίτερα χρήσιμο για συστήματα ανίχνευσης απάτης σε πραγματικό χρόνο, όπου νέα δεδομένα προστίθενται συνεχώς στο δίκτυο.

3.3 Βάσεις Δεδομένων Γράφων (Graph Databases)

Οι βάσεις δεδομένων γράφων είναι βάσεις δεδομένων που έχουν σχεδιαστεί ειδικά για την αποθήκευση, τη διαχείριση, την ανάκτηση και την εκτέλεση πολύπλοκων αλγορίθμων σε μεγάλα σύνολα δεδομένων δομημένων σε μορφή γράφων, όπως αυτά που συναντώνται στις οικονομικές συναλλαγές. Μία από τις πιο δημοφιλείς βάσεις δεδομένων γράφων είναι η Neo4j.

3.3.1 Neo4j, Cypher, Graph Data Science Library

Η Neo4j είναι μια ισχυρή και ευέλικτη βάση δεδομένων γράφων που επιτρέπει την αποθήκευση σχέσεων και τη διεξαγωγή πολύπλοκων αναζητήσεων στο δίκτυο. Ένα σημαντικό πλεονέκτημα της Neo4j είναι ότι υποστηρίζει την γλώσσα ερωτημάτων Cypher, που είναι ειδικά σχεδιασμένη για τη διεξαγωγή ερωτημάτων (queries) σε δεδομένα γράφων. Τα ερωτήματα της Cypher είναι συχνά πολύ πιο φυσικά και κατανοητά από τα ερωτήματα σε SQL για τέτοιου είδους δεδομένα. (Hodler, 2021)

Η Neo4j διαθέτει επίσης τη Neo4j Graph Data Science Library, μια ισχυρή βιβλιοθήκη που υλοποιεί σημαντικούς αλγορίθμους ανάλυσης γράφων. Αυτοί οι αλγόριθμοι περιλαμβάνουν μέτρα κεντρικότητας, αλγόριθμους ανίχνευσης κοινοτήτων, εύρεση διαδρομών, κ.α. (Hodler, 2021)

3.3.2 Σύγκριση Neo4j και TigerGraph

Η TigerGraph είναι μια άλλη αναδυόμενη βάση δεδομένων γράφων που είναι γνωστή για την υψηλή απόδοσή της σε αναλύσεις γράφων μεγάλης κλίμακας. Η TigerGraph είναι ιδιαίτερα ισχυρή για ανάλυση σε πραγματικό χρόνο και μπορεί να χειριστεί πολύ μεγαλύτερα δίκτυα από τη Neo4j με ταχύτητα. Από την άλλη, η Neo4j είναι βελτιστοποιημένη για περιπτώσεις όπου τα δεδομένα είναι σε μεγάλο βαθμό διασυνδεδεμένα, ενώ είναι ευκολότερη στη χρήση και καλύτερα τεκμηριωμένη για τις περισσότερες εφαρμογές. (Tanner, 2025)

3.4 Ειδικά θέματα

3.4.1 Knowledge Graphs

Τα knowledge graphs ορίζονται στη βιβλιογραφία ως δομές δεδομένων σε μορφή γράφου, που αναπαριστούν οντότητες (entities) του πραγματικού κόσμου και τις μεταξύ τους σχέσεις (relations) ως κόμβους και ακμές αντίστοιχα, επιτρέποντας την αποτύπωση τόσο των επιμέρους ιδιοτήτων όσο και του σημασιολογικού πλαισίου συνδεσιμότητας. Σε αντίθεση με τις παραδοσιακές tabular αναπαραστάσεις σε μορφή πίνακα, όπου τα δεδομένα εξετάζονται συχνά απομονωμένα, τα knowledge graphs διατηρούν ρητά τη δομική σχέση μεταξύ των στοιχείων, καθιστώντας εφικτή την εξαγωγή νέας γνώσης μέσω συλλογισμού πάνω σε δίκτυα σχέσεων. Η εμπλουτισμένη αυτή δομή υποστηρίζει βασικές λειτουργίες όπως knowledge completion, entity resolution και relational reasoning, καθιστώντας τα knowledge graphs βασικό εργαλείο στην σύγχρονη έρευνα γύρω από την αναπαράσταση γνώσης. (Ji κ.ά., 2022; Peng κ.ά., 2023)

Στο πεδίο της ανίχνευσης οικονομικής απάτης, τα knowledge graphs αναδεικνύονται ως ιδανικό πλαίσιο για την μοντελοποίηση της απάτης ως δικτυακού φαινομένου, όπου οι συσχετίσεις μεταξύ οντοτήτων (π.χ. λογαριασμοί, συναλλαγές, συσκευές) υπερβαίνουν τα όρια της μεμονωμένης ανάλυσης. (Mao κ.ά., 2022; Wen κ.ά., 2022) Η αναπαράσταση σε γράφο επιτρέπει την αποκάλυψη κρυμμένων μοτίβων, όπως fraudulent communities, collusive networks και shared identities, τα οποία παραμένουν αόρατα σε κλασικές στατιστικές μεθόδους. (Madduru & Janvekar, 2023; Mao κ.ά., 2022) Ειδικότερα, η ετερογενής φύση των knowledge graphs με πολλαπλούς τύπους κόμβων και ακμών, διευκολύνει την ενσωμάτωση τεχνικών συλλογισμού και μάθησης για εξαγωγή γνώσης, ενισχύοντας την ακρίβεια ανίχνευσης σε πολύπλοκα οικονομικά περιβάλλοντα. (Madduru & Janvekar, 2023; Majumder κ.ά., 2025) Συνεπώς, τα knowledge graphs γεφυρώνουν την αναπαράσταση γνώσης με την πρακτική εφαρμογή σε προβλήματα όπου η δομή των δεδομένων είναι εξίσου κρίσιμη με τα χαρακτηριστικά τους.

3.4.2 Explainable AI (XAI)

Σε ρυθμιζόμενα περιβάλλοντα όπως ο χρηματοπιστωτικός τομέας, είναι κρίσιμο τα μοντέλα AI να μπορούν να εξηγήσουν τις αποφάσεις τους. Κανονισμοί όπως το GDPR (Ευρωπαϊκή

Ένωση) και το CCPA (Καλιφόρνια, ΗΠΑ) απαιτούν διαφάνεια στις αυτοματοποιημένες αποφάσεις, κάτι που σημαίνει πως η έλλειψη ερμηνευσιμότητας μπορεί να οδηγήσει σε νομικά προβλήματα και απώλεια εμπιστοσύνης από τους πελάτες. (Chen κ.ά., 2025; Liu κ.ά., 2021)

Οι άξονες στους οποίους βασίζεται η XAI είναι οι ακόλουθοι: (Σπανάκη κ.ά., 2024)

- Χρήση «απλών» και διαφανών μοντέλων (π.χ. δέντρα απόφασης, γραμμικά μοντέλα, κανόνες) ώστε το ίδιο το μοντέλο να είναι άμεσα ερμηνεύσιμο χωρίς πρόσθετο εργαλείο.
- Εφαρμογή τεχνικών επεξήγησης πάνω σε οποιοδήποτε μοντέλο μηχανικής μάθησης, ακόμα και αν αυτό χαρακτηρίζεται ως «μαύρο κουτί». (ενδεικτικές τεχνικές αποτελούν οι SHAP και LIME που θα δούμε στη συνέχεια)
- Διαδικασίες επεξήγησης τόσο σε μεμονωμένες αποφάσεις όσο και στη συνολική συμπεριφορά των μοντέλων.
- Μετατροπή των διαδικασιών λήψης αποφάσεων σε μορφή κατανοητή από τον άνθρωπο μέσω κανόνων.
- Τρόποι παρουσίασης και οπτικοποίησης των πληροφοριών με τρόπο που να γίνονται κατανοητοί οι συσχετισμοί των δεδομένων και των ενεργειών του μοντέλου.

Χαρακτηριστικά παραδείγματα τεχνικών XAI είναι οι παρακάτω:

SHAP (Shapley Additive Explanations)

Αποκαλύπτει ποιοι παράγοντες επηρέασαν περισσότερο την πρόβλεψη του μοντέλου. Το SHAP βασίζεται στη θεωρία παιγνίων για να αποδώσει τη συνεισφορά κάθε χαρακτηριστικού στην τελική πρόβλεψη. (Lundberg & Lee, 2017)

LIME (Local Interpretable Model-agnostic Explanations)

Αποδομεί τις αποφάσεις σε απλούστερες εξηγήσεις για μεμονωμένα παραδείγματα. Το LIME δημιουργεί τοπικά, απλούστερα και ερμηνεύσιμα μοντέλα γύρω από συγκεκριμένες προβλέψεις, τα οποία προσομοιώνουν το αρχικό μη-ερμηνεύσιμο μοντέλο, έτσι ώστε να

γίνει δυνατή η ανίχνευση της συμβολής των χαρακτηριστικών στην πρόβλεψη. (Ribeiro κ.ά., 2016)

3.4.3 Τεχνικές για Imbalanced Datasets

Τα σύνολα οικονομικών δεδομένων είναι χαρακτηριστικά ανισόρροπα (imbalanced), με τις περιπτώσεις απάτης να είναι πολύ λιγότερες από τις νόμιμες συναλλαγές. (Cheng κ.ά., 2025) Για την αντιμετώπιση αυτής της πρόκλησης, χρησιμοποιούνται τεχνικές κατά την προεπεξεργασία των datasets, οι οποίες εξισορροπούν την κατανομή των παραδειγμάτων στις κλάσεις. (Bin Sulaiman κ.ά., 2022) Υπάρχουν διάφορα είδη τεχνικών, με πιο σημαντικές τις τεχνικές επαναδειγματοληψίας (resampling), οι οποίες διακρίνονται στις παρακάτω βασικές κατηγορίες: (Jiang κ.ά., 2025)

- **Υπερδειγματοληψία (Oversampling):** Αύξηση των δειγμάτων της μειονοτικής κλάσης (minority class).

  Ενδεικτικά: Random Over-sampling, SMOTE, ADASYN.

- **Υποδειγματοληψία (Undersampling):** Μείωση των δειγμάτων της πλειονοτικής κλάσης (majority class).

  Ενδεικτικά: Random Under-sampling, Tomek Links, NearMiss

Δύο από τις πιο δημοφιλείς τεχνικές είναι οι ακόλουθες:

SMOTE (Synthetic Minority Oversampling Technique)

Παράγει συνθετικά δείγματα της μειονοτικής κλάσης χρησιμοποιώντας τα ήδη υπάρχοντα δεδομένα, με την εξής διαδικασία: Για κάθε παράδειγμα μειονοτικής κλάσης, επιλέγει τυχαία *k* γείτονες (*k*NN – *k* nearest neighbors) και ενοποιεί τα χαρακτηριστικά τους. Αυτή η προσέγγιση διατηρεί σε ικανοποιητικό βαθμό την κατανομή των χαρακτηριστικών και αντιμετωπίζει την υπερεκπαίδευση. (Chawla κ.ά., 2002)

ADASYN (Adaptive Synthetic)

Χρησιμοποιεί μια σταθμισμένη κατανομή για διαφορετικά παραδείγματα μειονοτικών τάξεων, ανάλογα με το επίπεδο δυσκολίας μάθησης που έχουν. Δηλαδή, παράγονται

περισσότερα συνθετικά δεδομένα για παραδείγματα μειονοτικών τάξεων που είναι πιο δύσκολο να τα μάθουν, σε σύγκριση με αυτά που είναι πιο εύκολο. Ως αποτέλεσμα, η ADASYN βελτιώνει τη μάθηση με δύο τρόπους: πρώτον, μειώνοντας την προκατάληψη (bias) που εισάγεται από την ανισορροπία των τάξεων και δεύτερον, μετατοπίζοντας προσαρμοστικά το όριο απόφασης ταξινόμησης προς τα δύσκολα παραδείγματα. (Haibo He κ.ά., 2008)

4. Μεθοδολογία Σχεδιασμού και Ανάπτυξης Συστήματος Ανίχνευσης Οικονομικής Απάτης

4.1 Εισαγωγή στη Μεθοδολογική Προσέγγιση

4.1.1 Γενικό πλαίσιο

Η παρούσα πτυχιακή εργασία εστιάζει στην ανάπτυξη ενός συστήματος για τον εντοπισμό και την πρόληψη της οικονομικής απάτης, αξιοποιώντας τη συνέργεια μεταξύ των ΒΔΓ (Graph Databases) και των ΝΔΓ (GNNs). Η μεθοδολογική προσέγγιση που υιοθετείται στο παρόν κεφάλαιο αποτελεί μια θεμελιώδη μετατόπιση από την παραδοσιακή, «εγγραφοκεντρική» (record-centric) ανάλυση συναλλαγών, σε μια «δικτυοκεντρική» (network-centric) θεώρηση.

Στην παραδοσιακή προσέγγιση, κάθε συναλλαγή εξετάζεται ως μεμονωμένο γεγονός, αποκομμένο από το ευρύτερο πλαίσιο των αλληλεπιδράσεων του χρήστη. Αντιθέτως, η προτεινόμενη μεθοδολογία αντιμετωπίζει το χρηματοπιστωτικό σύστημα ως έναν ζωντανό, εξελισσόμενο οργανισμό διασυνδεδεμένων οντοτήτων, όπου η δομή των σχέσεων φέρει εξίσου κρίσιμη πληροφορία με τα ατομικά χαρακτηριστικά των συναλλαγών.

Η επιλογή αυτής της προσέγγισης υπαγορεύεται από τη φύση των σύγχρονων σχημάτων απάτης, όπως τα κυκλώματα ξεπλύματος χρήματος (money laundering rings) και οι συνθετικές ταυτότητες (synthetic identities), τα οποία σχεδιάζονται ακριβώς για να παρακάμπτουν τους κανόνες που εξετάζουν μεμονωμένα συμβάντα. Οι δόλιες οντότητες σπάνια δρουν απομονωμένες· αντιθέτως, σχηματίζουν πολύπλοκα δίκτυα συνεργατών, μοιράζονται πόρους (όπως διευθύνσεις IP, συσκευές ή αριθμούς τηλεφώνου) και δημιουργούν μοτίβα ροής χρήματος που είναι ορατά μόνο μέσω της τοπολογικής ανάλυσης του γράφου.

4.1.2 Ειδικό πεδίο εφαρμογής

Το πρακτικό μέρος της παρούσας εργασίας εστιάζει στον εντοπισμό απάτης σε περιβάλλον ηλεκτρονικού εμπορίου (e-commerce), όπου η απουσία φυσικής παρουσίας της κάρτας (Card-Not-Present) αυξάνει κατακόρυφα το επίπεδο κινδύνου. Σε αυτό το πλαίσιο, η «ταυτότητα» ενός χρήστη δεν ορίζεται μόνο από το ονοματεπώνυμό του, αλλά από ένα

σύνθετο ψηφιακό αποτύπωμα που περιλαμβάνει τεχνικές παραμέτρους, όπως η διεύθυνση IP, ο τύπος της συσκευής, το λειτουργικό σύστημα και το domain του ηλεκτρονικού ταχυδρομείου. Η πρόκληση έγκειται στο γεγονός ότι οι δράστες συχνά χρησιμοποιούν πραγματικά στοιχεία πελατών (account takeover) ή δημιουργούν συνθετικές ταυτότητες που μοιάζουν με νόμιμες, καθιστώντας τα παραδοσιακά φίλτρα ανεπαρκή.

Η μεθοδολογία μας στοχεύει ακριβώς στην αποκάλυψη αυτών των έμμεσων συσχετίσεων. Εστιάζει στην αναγνώριση μοτίβων όπου διαφορετικές ταυτότητες χρηστών συγκλίνουν σε κοινά τεχνικά χαρακτηριστικά (όπως το ID μιας συσκευής ή ένα domain email), δημιουργώντας «κοινότητες» υψηλού κινδύνου. Με την εφαρμογή του GNN πάνω σε αυτό το δίκτυο σχέσεων, το σύστημα δεν αξιολογεί μόνο τα χαρακτηριστικά της ίδιας της συναλλαγής, αλλά «δανείζεται» πληροφορία από τη συμπεριφορά των γειτονικών κόμβων στον γράφο, επιτρέποντας τον εντοπισμό δόλιων ενεργειών που θα παρέμεναν αόρατες σε οποιαδήποτε παραδοσιακή ανάλυση πίνακα.

4.1.3 Διάρθρωση

Η μεθοδολογία διαρθρώνεται σε μια ολοκληρωμένη ροή επεξεργασίας των δεδομένων (pipeline), η οποία ξεκινά από την ακατέργαστη πληροφορία (raw dataset) και καταλήγει στην εξαγωγή γνώσης και την λήψη αποφάσεων. Συγκεκριμένα, η διαδικασία περιλαμβάνει τη συλλογή και προεπεξεργασία δεδομένων, τον μετασχηματισμό τους σε δομή γράφου, την εφαρμογή γραφοαναλυτικών αλγορίθμων για την εξαγωγή τοπολογικών χαρακτηριστικών, την εκπαίδευση ενός μοντέλου βαθιάς μάθησης και ενός ensemble ταξινομητή (classifier) βασισμένου σε δέντρα απόφασης (decision trees), και τέλος, την αξιολόγηση και ερμηνεία των αποτελεσμάτων. Κάθε στάδιο έχει σχεδιαστεί με γνώμονα την επεκτασιμότητα (scalability), την ακρίβεια (precision) και την κανονιστική συμμόρφωση (compliance), λαμβάνοντας υπόψη το πεδίο εφαρμογής της εργασίας και τους περιορισμούς που τέθηκαν στα προηγούμενα κεφάλαια.

4.2 Αρχιτεκτονική Συστήματος και Ροή Δεδομένων

Η αρχιτεκτονική του συστήματος είναι σπονδυλωτή (modular), επιτρέποντας την ανεξάρτητη ανάπτυξη και βελτιστοποίηση κάθε επιμέρους υποσυστήματος. Αυτή η προσέγγιση εξασφαλίζει ευελιξία, καθώς νέα μοντέλα ή πηγές δεδομένων μπορούν να ενσωματωθούν χωρίς να απαιτείται ριζικός ανασχεδιασμός του συστήματος.

4.2.1 Ολοκληρωμένη Αρχιτεκτονική Pipeline

Το σύστημα αποτελείται από έξι κύρια στάδια, τα οποία λειτουργούν διαδοχικά:

1. ***Συλλογή και Προεπεξεργασία Δεδομένων (Data Ingestion & Preprocessing)***

   Στο πρώτο στάδιο, τα ακατέργαστα δεδομένα συναλλαγών και ταυτοτήτων εισάγονται στο σύστημα. Εφαρμόζονται τεχνικές καθαρισμού για την αντιμετώπιση ελλιπών τιμών και θορύβου, καθώς και τεχνικές κανονικοποίησης και κωδικοποίησης για την προετοιμασία των χαρακτηριστικών. Γίνεται διαχωρισμός των δεδομένων (Train-Validation-Test) για την επικείμενη εκπαίδευση των μοντέλων με ειδική πρόνοια για την αποφυγή διαρροής πληροφορίας (data leakage).

2. ***Μοντελοποίηση και Κατασκευή Γράφου (Graph Construction)***

   Τα προεπεξεργασμένα δεδομένα «πινακοειδούς» (tabular) μορφής μετασχηματίζονται σε έναν ετερογενή γράφο εντός της Neo4j. Ορίζονται οι κόμβοι (Συναλλαγές, Κάρτες, Email, Συσκευές) και οι ακμές συσχέτισης και συν-εμφάνισης, δημιουργώντας μια πλούσια αναπαράσταση των αλληλεπιδράσεων.

3. ***Ανάλυση Γράφων (Graph Analytics)***

   Πριν την εφαρμογή βαθιάς μάθησης, ο γράφος αναλύεται με αλγορίθμους της Neo4j GDS βιβλιοθήκης. Υπολογίζονται μέτρα κεντρικότητας (Centrality Measures) και ανιχνεύονται κοινότητες (Community Detection) για τον εμπλουτισμό των δεδομένων με τοπολογική πληροφορία.

4. ***Εξαγωγή Χαρακτηριστικών και Embeddings (Feature Engineering)***

   Συνδυάζονται τα αρχικά χαρακτηριστικά των συναλλαγών με τα αποτελέσματα της ανάλυσης γράφων. Επιπλέον, παράγονται διανυσματικές αναπαραστάσεις των

κόμβων (node embeddings) μέσω της τεχνικής FastRP, η οποία κωδικοποιεί τη δομική θέση κάθε κόμβου σε μορφή κατάλληλη για νευρωνικά δίκτυα.

5. ***Εκπαίδευση Μοντέλου GNN και Αξιολόγηση (GNN Training)***

   Ο ετερογενής γράφος τροφοδοτείται σε ένα μοντέλο Heterogeneous GraphSAGE. Το μοντέλο εκπαιδεύεται να διακρίνει μεταξύ νόμιμων και δόλιων κόμβων-συναλλαγών μέσω συλλογής πληροφοριών (neighbor sampling) από τις συνδεδεμένες οντότητες. Η απόδοση του συστήματος αξιολογείται με όλες τις βασικές καθιερωμένες μετρικές (AUC-PR, F1-Score, κλπ.) και συγκρίνεται με το μοντέλο αναφοράς (baseline) XGBoost.

6. ***Ερμηνευσιμότητα Αποτελεσμάτων (XAI & Interpretability)***

   Γίνεται μια πρώιμη απόπειρα εφαρμογής τεχνικών XAI (SHAP, GNNExplainer), με σκοπό την ανάδειξή τους ως βασικά συστατικά μέρη ενός ολοκληρωμένου συστήματος που σέβεται τις επιταγές των σύγχρονων κανονιστικών πλαισίων.

4.2.2 Τεχνολογίες και Λογισμικό (Technology Stack)

Η υλοποίηση βασίζεται σε ένα σύνολο εργαλείων ανοικτού κώδικα και εξειδικευμένου λογισμικού, επιλεγμένα για την απόδοση και τη συμβατότητά τους:

- **Γλώσσα Προγραμματισμού:** *Python (3.12)*

  Η Python κυριαρχεί στην επιστήμη δεδομένων και διαθέτει πληθώρα διαθέσιμων βιβλιοθηκών.

- **Βάση Δεδομένων Γράφων:** *Neo4j (2026.02.2)*

  Η Neo4j επιλέχθηκε ως η κορυφαία πλατφόρμα γράφων, προσφέροντας την απαραίτητη αποθήκευση (native graph storage) και την ισχυρή βιβλιοθήκη αλγορίθμων GDS για μαζική ανάλυση.

- **Πλαίσιο Βαθιάς Μάθησης:** *PyTorch / PyTorch Geometric (PyG)*

  Παρέχει βελτιστοποιημένες υλοποιήσεις των GNN αλγορίθμων καθώς και έτοιμες δομές για ετερογενείς γράφους, αναλαμβάνει την εκπαίδευση των νευρωνικών δικτύων και επιτρέπει την εύκολη διασύνδεση με τη Neo4j.

- **Επεξεργασία Δεδομένων:** *Pandas / NumPy / scikit-learn*

Για τη διαχείριση (φόρτωση, καθαρισμό, μετασχηματισμό, κωδικοποίηση, κανονικοποίηση) των δεδομένων πριν την εισαγωγή στον γράφο.

- **Οπτικοποίηση:** *Seaborn / Matplotlib*

  Οπτική διερεύνηση κατανομών χαρακτηριστικών, αποδόσεων μοντέλων (ROC/PR curves, learning curves) και υπογράφων απάτης στη Neo4j.

4.3 Συλλογή και Προεπεξεργασία Δεδομένων

4.3.1 Σύνολα Δεδομένων (Datasets)

Λόγω περιορισμών εμπιστευτικότητας σε πραγματικά τραπεζικά δεδομένα, η υλοποίηση βασίζεται σε δημόσια διαθέσιμα σύνολα δεδομένων. Προτείνονται δύο – είναι απαραίτητο ωστόσο να διευκρινιστεί ότι στο τελικό, υλοποιημένο pipeline χρησιμοποιείται μόνο το IEEE-CIS Fraud Detection dataset. Το Elliptic Bitcoin dataset δεν ενσωματώνεται στην παρούσα υλοποίηση, παρουσιάζεται όμως ως θεωρητικά κατάλληλο για μελλοντική επέκταση.

IEEE-CIS Fraud Detection Dataset

Το dataset αυτό, που διατέθηκε από την Vesta Corporation, αποτελεί ένα από τα πλουσιότερα σύνολα δεδομένων για ανίχνευση απάτης σε διαδικτυακές συναλλαγές (e-commerce). Περιλαμβάνει περίπου 590.000 συναλλαγές, εκ των οποίων το 3.5% έχει χαρακτηριστεί ως απάτη. Η δομή του αποτελείται από δύο κύριους πίνακες που συνδέονται μέσω ενός πρωτεύοντος κλειδιού.

Η πολυπλοκότητα αυτού του dataset έγκειται στην υψηλή ετερογένεια και την παρουσία πεδίων ταυτότητας (Identity), τα οποία είναι ιδανικά για τη δημιουργία ενός ετερογενούς γράφου συναλλαγών-οντοτήτων. Για παράδειγμα, διαφορετικές συναλλαγές που μοιράζονται την ίδια συσκευή ή διεύθυνση IP μπορούν να συνδεθούν, αποκαλύπτοντας πιθανώς έναν κοινό δράστη πίσω από πολλαπλούς λογαριασμούς.

Elliptic Bitcoin Dataset

Το Elliptic dataset είναι ένας έτοιμος γράφος συναλλαγών Bitcoin, που χαρτογραφεί τη ροή κεφαλαίων μεταξύ οντοτήτων. Περιλαμβάνει 203.769 κόμβους (συναλλαγές) και 234.355 ακμές (ροές πληρωμών). Κάθε κόμβος έχει χαρακτηριστεί σε μία από τρεις κλάσεις: "Illicit" (παράνομη - 2%), "Licit" (νόμιμη - 21%) και "Unknown" (άγνωστη - 77%).

Αυτό το dataset είναι βολικό για την αξιολόγηση των GNNs, καθώς η δομή του γράφου δίνεται άμεσα και το πρόβλημα είναι ο εντοπισμός των παράνομων κόμβων με βάση τα μοτίβα ροής.

4.3.2 Καθαρισμός και Επιμέλεια Δεδομένων (Data Cleaning)

Η διαδικασία καθαρισμού είναι κρίσιμη, δεδομένου ότι τα χρηματοοικονομικά δεδομένα περιέχουν συχνά θόρυβο και ελλείψεις.

Διαχείριση Ελλιπών Τιμών (Missing Values)

Το IEEE-CIS dataset χαρακτηρίζεται από υψηλή αραιότητα (sparsity). Για τα αριθμητικά χαρακτηριστικά χρησιμοποιείται η μέθοδος *imputation* με τη διάμεσο (median) αντί του μέσου όρου, ώστε να αποφευχθεί η στρέβλωση από ακραίες τιμές. Για τα κατηγορικά χαρακτηριστικά, οι ελλιπείς τιμές δεν απορρίπτονται αλλά αντικαθίστανται με μια νέα κατηγορία "Unknown". Αυτή η επιλογή βασίζεται στην παραδοχή ότι η σκόπιμη απόκρυψη πληροφορίας (π.χ. χρήση proxy που κρύβει την IP) μπορεί να αποτελεί από μόνη της ένδειξη κινδύνου (risk signal). Στήλες με μεγάλο ποσοστό ελλιπών τιμών (π.χ. άνω του 80%) αφαιρούνται, καθώς δεν παρέχουν επαρκή πληροφορία για γενίκευση.

Αφαίρεση Διπλοεγγραφών και Θορύβου

Ελέγχονται οι συναλλαγές για διπλοεγγραφές που μπορεί να προκύψουν από σφάλματα συστήματος. Επιπλέον, εξετάζονται τα labels: συναλλαγές που αρχικά επισημάνθηκαν ως απάτη αλλά αργότερα αναιρέθηκαν (chargeback reversals) πρέπει να αντιμετωπίζονται με προσοχή.

4.3.3 Κανονικοποίηση και Κωδικοποίηση (Normalization & Encoding)

Τα νευρωνικά δίκτυα απαιτούν αριθμητικά δεδομένα σε συγκεκριμένες κλίμακες για τη βέλτιστη σύγκλιση των αλγορίθμων βελτιστοποίησης (optimizers).

Στα αριθμητικά χαρακτηριστικά (continuous features) εφαρμόζεται *λογαριθμικός μετασχηματισμός (Log-transformation)* για να μειωθεί η επίδραση των ακραίων τιμών και να προσεγγιστεί η κανονική κατανομή. Στη συνέχεια, ακολουθεί κανονικοποίηση *Z-score (Standard Scaling),* ώστε τα δεδομένα να έχουν μέση τιμή 0 και τυπική απόκλιση 1.

Όσον αφορά τα κατηγορικά χαρακτηριστικά (categorical features), χρησιμοποιείται Target Encoding, όπου κάθε κατηγορία αντιστοιχίζεται σε αριθμητική τιμή που αντανακλά τη συσχέτισή της με την κλάση απάτης.

4.3.4 Αντιμετώπιση Ανισόρροπων Δεδομένων (Imbalanced Data Handling)

Η ανίχνευση απάτης αποτελεί κλασικό παράδειγμα προβλήματος με εξαιρετικά ανισόρροπες κλάσεις. Στο IEEE-CIS dataset, οι απάτες είναι μόλις το ~3.5%, ενώ στο Elliptic το ~2% των κόμβων είναι "illicit". Η εκπαίδευση σε τέτοια δεδομένα ενέχει τον κίνδυνο να οδηγήσει σε μοντέλα που προβλέπουν συνεχώς τη πλειονοτική κλάση (νόμιμες συναλλαγές) επιτυγχάνοντας υψηλή ακρίβεια αλλά ελάχιστη ανάκληση (recall) στις απάτες. Για την αντιμετώπιση του προβλήματος, υιοθετούνται οι εξής στρατηγικές:

- **Stratified Train-Validation-Test Split:** Το σύνολο δεδομένων διασπάται σε train-validation-test (80%-10%-10%) με στρωματοποιημένη δειγματοληψία, ώστε το ποσοστό απάτης να διατηρείται σταθερό σε κάθε υποσύνολο και να αποφεύγεται το data leakage.
- **Αποφυγή Συνθετικών Δειγμάτων:** Αν και τεχνικές όπως το SMOTE εξετάστηκαν σε θεωρητικό επίπεδο, στην τελική έκδοση του pipeline δεν χρησιμοποιούνται, ώστε να διατηρηθεί η ρεαλιστική κατανομή και να αποφευχθεί η εισαγωγή τεχνητών μοτίβων απάτης.
- **Κατάλληλες Μετρικές και Κατώφλια:** Η εκπαίδευση και η αξιολόγηση εστιάζουν σε μετρικές ευαίσθητες στην ανισορροπία (AUC-PR, Recall, F1), ενώ το κατώφλι ταξινόμησης (threshold) του GNN ρυθμίζεται με βάση την καμπύλη Precision-Recall στο validation set, ώστε να επιτευχθεί καλύτερη ισορροπία μεταξύ εντοπισμού απάτης και ψευδώς θετικών αποτελεσμάτων.

- **Κλιμάκωση Βαρών στο XGBoost:** Στο baseline μοντέλο δίνεται μεγαλύτερη έμφαση στην κλάση απάτης, βελτιώνοντας την ανάκληση χωρίς υπερβολική υποβάθμιση της ακρίβειας.

4.4 Κατασκευή και Μοντελοποίηση Γράφου

Η μετατροπή των δεδομένων από μορφή πίνακα σε μορφή γράφου είναι ο θεμέλιος λίθος της προσέγγισής μας. Στόχος είναι η δημιουργία μιας δομής που να αναδεικνύει τις κρυφές σχέσεις μεταξύ των οντοτήτων.

4.4.1 Σχεδιασμός Ετερογενούς Γράφου

Σε αντίθεση με τους ομοιογενείς γράφους όπου όλοι οι κόμβοι είναι ίδιου τύπου, η πολυπλοκότητα των δεδομένων του IEEE-CIS απαιτεί έναν ετερογενή γράφο. Σε αυτόν, διακρίνουμε διαφορετικούς τύπους κόμβων και ακμών.

Ορισμός Κόμβων (Nodes)

1. Transaction Node: Ο κεντρικός κόμβος που αντιπροσωπεύει κάθε μεμονωμένη συναλλαγή. Φέρει χαρακτηριστικά όπως το ποσό, την ώρα, καθώς και το label απάτης.
2. Identity Nodes: Κόμβοι που αντιπροσωπεύουν οντότητες οι οποίες μπορούν να συνδέσουν πολλαπλές συναλλαγές: *Card, Device, EmailDomain*

Ορισμός Ακμών (Edges)

Οι ακμές ορίζουν τις σχέσεις μεταξύ των κόμβων και – στο επόμενο στάδιο – επιτρέπουν τη ροή πληροφορίας (message passing) κατά την εκπαίδευση του GNN. Θα ορίσουμε πολλαπλούς τύπους ακμών, ανάλογα με τον τύπο κόμβου με τον οποίο συσχετίζεται η συναλλαγή (π.χ. *Η συναλλαγή πραγματοποιήθηκε με τη συγκεκριμένη κάρτα / έγινε από τη συγκεκριμένη συσκευή / συνδέεται με το συγκεκριμένο email.*)

Επιπρόσθετα, κατασκευάζεται και μια ακμή «συνεμφάνισης» (co-occurrence) μεταξύ καρτών που έχουν χρησιμοποιηθεί με το ίδιο email σε διαφορετικές συναλλαγές, βοηθώντας στον δυνητικό εντοπισμό οργανωμένων κυκλωμάτων (fraud rings).

Αυτή η δομή «αστέρα» (star-schema topology) γύρω από κάθε συναλλαγή επιτρέπει στον αλγόριθμο να εντοπίσει έμμεσες συνδέσεις. Για παράδειγμα, δύο συναλλαγές που φαίνονται άσχετες μπορεί να συνδέονται επειδή πραγματοποιήθηκαν από την ίδια συσκευή, υποδεικνύοντας πιθανή απάτη.

4.4.2 Υλοποίηση στη Βάση Δεδομένων Neo4j

Η κατασκευή του γράφου υλοποιείται στη Neo4j. Με τις κατάλληλες εντολές στη γλώσσα Cypher, τα δεδομένα φορτώνονται μαζικά και, για τη διασφάλιση της απόδοσης κατά την ανάλυση, δημιουργούνται ευρετήρια (indexes) και περιορισμοί μοναδικότητας (constraints). Η Neo4j επιτρέπει την αποθήκευση εκατομμυρίων κόμβων και ακμών και την ταχύτατη διάσχισή τους (traversal), κάτι που είναι αδύνατο με τις παραδοσιακές SQL βάσεις.

4.5 Ανάλυση Γράφων και Εξαγωγή Χαρακτηριστικών

Πριν την εισαγωγή των δεδομένων στα μοντέλα μηχανικής μάθησης, εκτελείται μια σειρά αλγορίθμων ανάλυσης γράφων μέσω της βιβλιοθήκης Neo4j Graph Data Science (GDS). Σκοπός είναι ο εμπλουτισμός των δεδομένων με τοπολογικά χαρακτηριστικά που ποσοτικοποιούν τη σημασία και τη δομική θέση κάθε κόμβου. Αυτή η διαδικασία ονομάζεται *graph feature engineering*.

4.5.1 Μέτρα Κεντρικότητας (Centrality Measures)

Τα μέτρα κεντρικότητας θα μας βοηθήσουν στον εντοπισμό κόμβων που παίζουν κρίσιμο ρόλο στο δίκτυο. Επιλέγουμε τα ακόλουθα:

1. **Degree Centrality:** Μετρά τον αριθμό των συνδέσεων ενός κόμβου. Στο πλαίσιο της απάτης, μια συσκευή ή μια IP με εξαιρετικά υψηλό βαθμό είναι ύποπτη, καθώς μπορεί να υποδηλώνει bot ή κέντρο συντονισμού επιθέσεων.

2. **PageRank:** Μετρά την «επιρροή» ενός κόμβου. Υψηλό PageRank σε έναν κόμβο μπορεί να υποδηλώνει ότι μέσω αυτού διακινείται μεγάλος όγκος κεφαλαίων, κάτι που συχνά παρατηρείται σε κόμβους «ξεπλύματος».
3. **Betweenness Centrality:** Εντοπίζει κόμβους που λειτουργούν ως «γέφυρες» μεταξύ διαφορετικών τμημάτων του δικτύου. Αυτοί οι κόμβοι είναι κρίσιμοι για τη ροή της πληροφορίας (ή του παράνομου χρήματος) και συχνά αντιστοιχούν σε ενδιάμεσους λογαριασμούς ("mules") που χρησιμοποιούνται για τη διαστρωμάτωση (layering) των κεφαλαίων.

4.5.2 Ανίχνευση Κοινοτήτων (Community Detection)

Συχνά, η απάτη δεν είναι ατομική υπόθεση, αλλά οργανωμένη. Οι αλγόριθμοι ανίχνευσης κοινοτήτων στοχεύουν στον εντοπισμό ομάδων κόμβων που έχουν πυκνές συνδέσεις μεταξύ τους. Θα εφαρμόσουμε τους εξής αλγόριθμους:

- **Louvain Modularity:** Εφαρμόζεται για τον εντοπισμό κοινοτήτων (clusters). Αν ένας ικανός αριθμός μελών μιας κοινότητας είναι απάτες, τότε όλα τα υπόλοιπα μέλη της κοινότητας θεωρούνται υψηλού κινδύνου («ενοχή λόγω συσχέτισης» - “guilt by association”).
- **Weakly Connected Components (WCC):** Χρησιμοποιείται για την εύρεση απομονωμένων υπογράφων. Στα δίκτυα απάτης, συχνά παρατηρούνται μικρές, απομονωμένες νησίδες αλληλεπιδράσεων (π.χ. μια ομάδα καρτών που χρησιμοποιούνται αποκλειστικά σε συγκεκριμένους εμπόρους για ξέπλυμα).

Τα αποτελέσματα αυτών των αλγορίθμων (π.χ. PageRank score, Community ID) αποθηκεύονται ως νέα χαρακτηριστικά στους κόμβους και χρησιμοποιούνται ως είσοδος στο επόμενο στάδιο.

4.6 Εξαγωγή Embeddings και Σύνθεση Διανύσματος Εισόδου

Πέρα από τα υπολογιστικά χαρακτηριστικά, η σύγχρονη προσέγγιση απαιτεί την αυτόματη εκμάθηση χαρακτηριστικών μέσω graph embeddings. Τα embeddings είναι διανύσματα χαμηλής διάστασης που «συμπιέζουν» την πληροφορία της γειτονιάς ενός κόμβου.

4.6.1 Graph Embeddings

Χρησιμοποιούνται αλγόριθμοι που βασίζονται σε τυχαίους περιπάτους (random walks) στον γράφο. Επιλέγεται ο FastRP (Fast Random Projection), για την ταχύτητά του και την ικανότητά του να κλιμακώνεται σε πολύ μεγάλους γράφους εντός της Neo4j. Το FastRP παράγει embeddings που διατηρούν τις αποστάσεις μεταξύ των κόμβων, έτσι ώστε κόμβοι με παρόμοια δομική θέση να έχουν κοντινά διανύσματα. Τα παραγόμενα embeddings προστίθενται στο διάνυσμα χαρακτηριστικών του κόμβου.

4.6.2 Τελική Σύνθεση Χαρακτηριστικών

Η στρατηγική διαχωρίζεται ανάλογα με το μοντέλο:

Για κάθε κόμβο συναλλαγής, το τελικό διάνυσμα εισόδου x που τροφοδοτείται στο XGBoost, αλλά και αργότερα για ανάλυση σημασίας χαρακτηριστικών (SHAP), αποτελείται από τη συνένωση (concatenation) τριών πηγών πληροφορίας:

$$x = f_{\text{tabular}} \,||\, f_{\text{graph\_metrics}} \,||\, f_{\text{embeddings}}$$

Όπου:

- f_{tabular}: Τα αρχικά δεδομένα εμπλουτισμένα, μετά την κανονικοποίηση.
- $f_{\text{graph\_metrics}}$: Οι τιμές κεντρικότητας (PageRank, Degree) και το Community ID.
- $f_{\text{embeddings}}$: Το διάνυσμα που παρήχθη από το FastRP.

Η επιλογή αυτή είναι στοχευμένη· σκοπός είναι να αξιολογηθεί η επίδοση ενός GNN (το οποίο τροφοδοτείται μόνο με tabular features και μαθαίνει τη δομή δυναμικά μέσω message passing) έναντι του κορυφαίου αλγόριθμου Gradient Boosting, όταν όμως σε αυτόν δίνονται έτοιμα (hand-crafted) γραφικά χαρακτηριστικά (graph features).

Στο GNN από την άλλη, οι κόμβοι συναλλαγών φέρουν αποκλειστικά τα tabular χαρακτηριστικά στο διάνυσμα εισόδου x, ώστε το νευρωνικό δίκτυο να ανακαλύψει από μόνο του την τοπολογική συσχέτιση, χωρίς διαρροή της ετικέτας. Αντίστοιχα, στους βοηθητικούς κόμβους (Card, Email, Device) τοποθετούνται ντετερμινιστικά δομικά χαρακτηριστικά (π.χ. λογάριθμος του πλήθους συναλλαγών ανά οντότητα) για να λειτουργήσουν ως σταθερή είσοδος προς μάθηση.

4.7 Εκπαίδευση Μοντέλων GNN

4.7.1 Αρχιτεκτονική GNN

Ως κύρια αρχιτεκτονική επιλέγεται το Heterogeneous GraphSAGE. Η επιλογή του GraphSAGE έναντι των παραδοσιακών GCN είναι στρατηγική: πρόκειται για έναν inductive αλγόριθμο, ικανό να γενικεύει σε νέους, αόρατους κόμβους δημιουργώντας αναπαραστάσεις μέσω δειγματοληψίας των γειτόνων του (neighbor sampling). Η ετερογενής παραλλαγή του επιτρέπει στο δίκτυο να επεξεργάζεται ξεχωριστά τις ακμές διαφορετικού τύπου, μαθαίνοντας εξειδικευμένα βάρη για κάθε είδος οντότητας (Card, Email, Device).

4.7.2 Στρατηγική Εκπαίδευσης και Βελτιστοποίηση

Ακολουθούνται οι παρακάτω σχεδιαστικές και τεχνικές επιλογές:

- **Συνάρτηση Απώλειας (Loss Function):** Λόγω του δυαδικού χαρακτήρα της ταξινόμησης, χρησιμοποιείται η *BCEWithLogitsLoss* (Binary Cross Entropy with Logits) χωρίς ρητά class weights, με το πρόβλημα της ανισορροπίας να αντιμετωπίζεται κυρίως μέσω δυναμικής επιλογής κατωφλίου και μετρικών αξιολόγησης.
- **Βελτιστοποιητής (Optimizer):** Χρησιμοποιείται ο *AdamW (Adam with Weight Decay)* για σταθερή εκπαίδευση και αποφυγή overfitting.
- **Μηχανισμοί Mini-batch:** Για τη διαχείριση του τεράστιου γράφου και της υπολογιστικής πολυπλοκότητας, η εκπαίδευση γίνεται μέσω του *NeighborLoader* της PyG, το οποίο αντλεί τοπικά υπογραφήματα συγκεκριμένου βάθους με περιορισμένο αριθμό γειτόνων για κάθε batch. Παράλληλα, ενσωματώνεται μηχανισμός Early Stopping βασισμένος στο Average Precision (AP) του Validation set.

4.8 Πλαίσιο Αξιολόγησης (Evaluation Framework)

Η αξιολόγηση του συστήματος δεν μπορεί να βασιστεί στην απλή ακρίβεια (Accuracy), καθώς σε ένα dataset με 97% νόμιμες συναλλαγές, όπως είναι το IEEE-CIS, ένα «τυφλό» μοντέλο που το μόνο που κάνει είναι να μαρκάρει όλες τις συναλλαγές ως νόμιμες, θα είχε 97% ακρίβεια, αλλά θα ήταν στην ουσία άχρηστο. (Valero, 2025)

Τα κριτήρια αξιολόγησης είναι:

1. **Precision:** Δείχνει πόσες από τις συναλλαγές που το μοντέλο χαρακτήρισε ως απάτη ήταν όντως απάτη. Υψηλό Precision σημαίνει λιγότερα false positives.
2. **Recall:** Δείχνει πόσες από τις πραγματικές περιπτώσεις απάτης κατάφερε να βρει το μοντέλο. Υψηλό Recall σημαίνει λιγότερα false negatives.
3. **AUC-PR (Area Under the Precision-Recall Curve):** Δείχνει πόσο καλά το μοντέλο ξεχωρίζει τις απάτες από τις νόμιμες συναλλαγές σε όλα τα πιθανά thresholds, δίνοντας έμφαση στη θετική κλάση, δηλαδή την απάτη. Είναι πολύ σημαντική σε imbalanced datasets, γιατί αποτυπώνει καλύτερα την πρακτική απόδοση.
4. **F1-score:** Είναι ο συνδυασμός Precision και Recall σε μία μόνο μετρική (ο αρμονικός τους μέσος). Χρησιμοποιείται όταν θέλουμε ισορροπία ανάμεσα στο να βρίσκουμε αρκετές απάτες και στο να μη χτυπάμε υπερβολικά πολλές νόμιμες συναλλαγές.
5. **AUC-ROC (Area Under the Receiver Operating Characteristic Curve):** Μετρά τη συνολική ικανότητα του μοντέλου να ξεχωρίζει τις δύο κλάσεις για όλα τα πιθανά thresholds. Είναι χρήσιμη ως συμπληρωματική εικόνα, αλλά σε πολύ ανισόρροπα δεδομένα συνήθως είναι λιγότερο ενδεικτική από την AUC-PR.

4.9 Ερμηνευσιμότητα και Διαφάνεια

Σε περιβάλλοντα υψηλής ρύθμισης (π.χ. GDPR), η ικανότητα εξήγησης των αποφάσεων ενός μοντέλου «μαύρου κουτιού» είναι απαραίτητη. Η μεθοδολογία ενσωματώνει εργαλεία XAI για την παροχή διαφάνειας.

- **SHAP:** Εφαρμόζεται στο μοντέλο XGBoost για την εξαγωγή της καθολικής σπουδαιότητας των χαρακτηριστικών (Global Feature Importance). Αποκαλύπτει εάν οι ταξινομήσεις βασίζονται κυρίως στα παραδοσιακά (tabular), στα γραφηματικά (graph analytics) ή στα embedding χαρακτηριστικά.
- **GNNExplainer:** Ένα εξειδικευμένο εργαλείο για GNNs που εντοπίζει το υπογράφημα (subgraph) που επηρέασε περισσότερο την πρόβλεψη. Μπορεί να αναδείξει οπτικά το «μονοπάτι» της απάτης μέσα στον γράφο, προσφέροντας πολύτιμες πληροφορίες για τη δράση των κυκλωμάτων.
- **Ερωτήματα στη ΒΔΓ:** Επιπλέον, εκτελούνται δυναμικά queries στη Neo4j για τον μακροαναλυτικό εντοπισμό κυκλωμάτων (Fraud Rings), ύποπτων καρτών και παραβιασμένων domains.

4.10 Περιορισμοί

Η μεθοδολογία αναγνωρίζει ορισμένους εγγενείς περιορισμούς:

- Η χρήση συνθετικών ή «ανωνυμοποιημένων» (anonymized) δεδομένων ενδέχεται να μην αποτυπώνει πλήρως την πολυπλοκότητα των πραγματικών επιθέσεων και μπορεί να περιορίσει την αποτελεσματικότητα του συστήματος σε πραγματικό περιβάλλον.
- Ενώ ο σχεδιασμός έχει γνώμονα την επεκτασιμότητα, η υλοποίηση και αξιολόγηση έγινε σε ένα περιορισμένο μέγεθος δεδομένων και προσαρμόστηκε λόγω των περιορισμών υπολογιστικής ισχύος και υλικού (hardware).
- Τα δεδομένα συναλλαγών διαθέτουν χρονική διάσταση, αλλά το μοντέλο δεν την λαμβάνει υπόψη. Αυτό θα μπορούσε να αντιμετωπιστεί με τη χρήση πιο προηγμένων αρχιτεκτονικών, όπως είναι τα Temporal GNNs.
- Το μοντέλο θα εκπαιδευτεί και θα αξιολογηθεί στα ίδια datasets. Δεν θα ελεγχθεί η ικανότητά του να γενικεύεται σε διαφορετικές πηγές δεδομένων.

4.11 Επισκόπηση

Η προτεινόμενη μεθοδολογία παρουσιάζει μια ολοκληρωμένη προσέγγιση για τη δημιουργία ενός συστήματος ανίχνευσης οικονομικής απάτης, που συνδυάζει τη δύναμη της θεωρίας γράφων με την προγνωστική ικανότητα της τεχνητής νοημοσύνης, προσφέροντας μια σύγχρονη λύση στο πρόβλημα της ανίχνευσης οικονομικής απάτης. Ο σχεδιασμός του pipeline επιτρέπει τη σταδιακή δημιουργία και δοκιμή των διαφόρων συστατικών, ενώ ταυτόχρονα διατηρεί την ευελιξία και την επεκτασιμότητα που απαιτείται για πραγματικές εφαρμογές. Στα επόμενα κεφάλαια θα παρουσιάσουμε την τεχνική υλοποίηση, καθώς και τα αποτελέσματα αυτής της προσέγγισης.

5. Υλοποίηση

5.1 Περιβάλλον Ανάπτυξης και Προετοιμασία Συστήματος

5.1.1 Οργάνωση Περιβάλλοντος και Δεδομένων

Για την εύρυθμη λειτουργία του κώδικα και τη διαχείριση του μεγάλου όγκου δεδομένων, υιοθετήθηκε μια αυστηρή ιεραρχική δομή φακέλων, η οποία διασφαλίζει την ακεραιότητα των δεδομένων και την ευκολία αναπαραγωγής των πειραμάτων.

Η κεντρική ρίζα (root folder) του έργου (`fraud_detection`) περιλαμβάνει τους εξής καταλόγους:

- Φάκελος `data/raw/ieee_cis/`: Φιλοξενεί τα πρωτογενή δεδομένα (IEEE-CIS datasets).
- Φάκελος `data/processed/`: Χρησιμοποιείται για την αποθήκευση των ενδιάμεσων επεξεργασμένων αρχείων, των παραγόμενων γραφημάτων και των τελικών αποτελεσμάτων αξιολόγησης.
- Φάκελος `data/neo4j_imports/`: Προορίζεται για τα αρχεία CSV που εξάγονται από την Python και εισάγονται στη βάση δεδομένων γράφων.
- Φάκελος `models/`: Περιέχει την προσαρμοσμένη υλοποίηση του GNN `hetero_graphsage.py`.
- Φάκελος `notebooks/`: Περιέχει τα Python scripts που υλοποιούν το pipeline της ανάλυσης.
- Φάκελος `utils/`: Περιέχει το Python script `neo4j_connector.py` που υλοποιεί τη σύνδεση με τη βάση.

5.1.2 Συλλογή και Εισαγωγή Δεδομένων

Τα δεδομένα που χρησιμοποιήθηκαν για την εκπαίδευση και αξιολόγηση του συστήματος προέρχονται από το dataset "IEEE-CIS Fraud Detection" της πλατφόρμας Kaggle. Η διαδικασία λήψης και εισαγωγής περιλαμβάνει τα εξής στάδια:

1. **Λήψη Δεδομένων**: Πρόσβαση στον διαγωνισμό της IEEE-CIS στο Kaggle και λήψη των συμπιεσμένων αρχείων.
2. **Επιλογή Αρχείων**: Χρήση των αρχείων `train_transaction.csv` και `train_identity.csv`, τα οποία περιλαμβάνουν τις απαραίτητες ετικέτες (labels) για την επίβλεψη της μάθησης.
3. **Ενσωμάτωση στο Project**: Αποσυμπίεση και τοποθέτηση των αρχείων στον υποφάκελο `data/raw/ieee_cis/`, διασφαλίζοντας τη συμβατότητα με τις προκαθορισμένες διαδρομές (paths) του κώδικα.

5.1.3 Διαμόρφωση Βάσης Δεδομένων Γράφων (Neo4j)

Η διαχείριση των σχέσεων μεταξύ των οντοτήτων πραγματοποιείται μέσω του Neo4j DBMS. Για τις ανάγκες της υλοποίησης, χρησιμοποιήθηκε η εφαρμογή Neo4j Desktop (έκδοση 2.1.3). Η παραμετροποίηση περιλαμβάνει:

1. **Δημιουργία Τοπικού Instance**: Δημιουργία DBMS με την ονομασία `eapGraphFraud` (έκδοση 2026.02.2).
2. **Εγκατάσταση Plugins**:
    1. **Graph Data Science (GDS)**: Απαραίτητο για την εκτέλεση προηγμένων αλγορίθμων γράφων και την παραγωγή node embeddings.
    2. **APOC (Awesome Procedures on Cypher)**: Χρησιμοποιείται για πολύπλοκους μετασχηματισμούς δεδομένων, μαζικές εισαγωγές και βοηθητικές λειτουργίες της Cypher.
3. **Κατάσταση Λειτουργίας**: Το DBMS πρέπει να βρίσκεται σε κατάσταση ενεργής λειτουργίας (*Running*) στο παρασκήνιο για την επικοινωνία μέσω του Bolt driver.
4. **Διαπιστευτήρια Σύνδεσης**: Παραμετροποίηση του `neo4j_connector.py` με Username: `neo4j` και Password τον κωδικό που ορίστηκε κατά τη δημιουργία του instance.

5.1.4 Περιβάλλον Ανάπτυξης και Εγκατάσταση Βιβλιοθηκών

Χρησιμοποιήθηκε η έκδοση Python 3.12 σε εικονικό περιβάλλον (Virtual Environment). Με την εκτέλεση των εντολών δημιουργίας και εγκατάστασης στο τερματικό (Windows Terminal – Command Prompt), δημιουργείται αυτόματα ο φάκελος `venv` εντός του κεντρικού καταλόγου του project. Ο φάκελος αυτός περιέχει το τοπικό αντίγραφο του διερμηνέα της Python καθώς και όλες τις βιβλιοθήκες που εγκαθιστούμε, εξασφαλίζοντας ότι το περιβάλλον είναι πλήρως αυτοτελές και ανεξάρτητο από το υπόλοιπο λειτουργικό σύστημα. Οι εντολές που εκτελέστηκαν στο τερματικό είναι οι εξής:

```
# Πλοήγηση στον φάκελο του project
cd Desktop/fraud_detection

# Δημιουργία εικονικού περιβάλλοντος
python -m venv venv

# Ενεργοποίηση εικονικού περιβάλλοντος
venv\Scripts\activate

# Εγκατάσταση βασικών βιβλιοθηκών επεξεργασίας και μηχανικής μάθησης
pip install pandas numpy matplotlib seaborn scikit-learn
category_encoders neo4j tqdm xgboost shap networkx

# Εγκατάσταση βιβλιοθηκών Deep Learning (PyTorch & PyTorch Geometric)
pip install torch torchvision torchaudio
pip install torch_geometric

# Εγκατάσταση εξειδικευμένων βιβλιοθηκών υποστήριξης (Scatter & Sparse)
για το PyTorch Geometric
# (!) Η συγκεκριμένη εντολή είναι συμβατή με το συγκεκριμένο περιβάλλον
που έγινε η εκτέλεση (!)
python -c "import torch, os; v = torch.__version__.split('+')[0];
os.system(f'pip install torch_scatter torch_sparse -f
[https://data.pyg.org/whl/torch-](https://data.pyg.org/whl/torch-
){v}+cpu.html')"
```

Μετά την ολοκλήρωση όλων των παραπάνω βημάτων, η συνολική δομή (directory tree) του φακέλου εργασίας – πριν την εκτέλεση των notebooks – αποτυπώνεται παρακάτω:

```
fraud_detection/
├── data/
│   ├── raw/
│   │   └── ieee_cis/
│   │       ├── train_transaction.csv
│   │       └── train_identity.csv
│   ├── processed/
│   └── neo4j_imports/
├── models/
│   ├── hetero_graphsage.py
├── notebooks/
│   ├── NOTEBOOK 1 - IEEE-CIS DATA PREPROCESSING...py
│   ├── NOTEBOOK 2 - GRAPH CONSTRUCTION IN NEO4J.py
│   ├── ...
│   └── NOTEBOOK 6 - XAI & INTERPRETABILITY.py
├── utils/
│   └── neo4j_connector.py
└── venv/
```

5.1.5 Ροή Εκτέλεσης (Pipeline)

Η υλοποίηση ολοκληρώνεται σε έξι διακριτές φάσεις, εκτελώντας τις εντολές στο terminal σειριακά:[2]

i. **Notebook 1 – Προεπεξεργασία Δεδομένων**: Καθαρισμός, διαχείριση ελλιπών τιμών και μηχανική χαρακτηριστικών (feature engineering) στα αρχικά δεδομένα.

```
python "notebooks/NOTEBOOK 1 - IEEE-CIS DATA PREPROCESSING & FEATURE ENGINEERING.py"
```

ii. **Notebook 2 – Κατασκευή Γράφου**: Μεταφορά των επεξεργασμένων δεδομένων στο Neo4j και δημιουργία των κόμβων και των σχέσεων ενός ετερογενούς γράφου.

```
python "notebooks/NOTEBOOK 2 - GRAPH CONSTRUCTION IN NEO4J.py"
```

iii. **Notebook 3 – Ανάλυση Γράφου (GDS)**: Εκτέλεση αλγορίθμων μέσω της βιβλιοθήκης GDS για την παραγωγή graph-native χαρακτηριστικών (graph features).

```
python "notebooks/NOTEBOOK 3 - GRAPH ANALYTICS WITH NEO4J GDS.py"
```

iv. **Notebook 4 – Προετοιμασία Χαρακτηριστικών (HeteroData)**: Δημιουργία node embeddings και κατασκευή ετερογενών δομών δεδομένων για το GNN μοντέλο.

```
python "notebooks/NOTEBOOK 4 - NODE EMBEDDINGS & FEATURE PREPARATION (HETERODATA).py"
```

v. **Notebook 5 – Εκπαίδευση Μοντέλου GNN**: Εκπαίδευση του ετερογενούς GraphSAGE και σύγκριση της απόδοσης με το μοντέλο αναφοράς XGBoost.

```
python "notebooks/NOTEBOOK 5 - GNN MODEL TRAINING (HETEROGENEOUS GS & BCE).py"
```

vi. **Notebook 6 – Επεξηγησιμότητα (XAI)**: Ανάλυση της συνεισφοράς των χαρακτηριστικών και ερμηνεία των προβλέψεων μέσω SHAP και GNNExplainer.

```
python "notebooks/NOTEBOOK 6 - XAI & INTERPRETABILITY.py"
```

[2] Σημειώνεται ότι το τελικό μέγεθος του συνολικού directory μετά την εκτέλεση όλων των notebooks, την εξαγωγή όλων των παραγόμενων αρχείων, και συμπεριλαμβανομένων των αρχείων εισόδου (raw dataset files), είναι ~12,3GB.

5.2 Ανάλυση Πηγαίου Κώδικα

Στην ενότητα αυτή πραγματοποιείται λεπτομερής ανάλυση των επιμέρους τμημάτων του κώδικα, ακολουθώντας τη ροή του pipeline. Κάθε υποενότητα εστιάζει στις βασικές λειτουργίες, τις βιβλιοθήκες και τους αλγορίθμους που χρησιμοποιήθηκαν σε κάθε στάδιο της υλοποίησης.

5.2.1 Notebook 1

Το πρώτο στάδιο της υλοποίησης επικεντρώνεται στην εισαγωγή, τον καθαρισμό και τον μετασχηματισμό των πρωτογενών δεδομένων του IEEE-CIS, καθώς και στην προετοιμασία τους για την κατασκευή του ετερογενούς γράφου. Οι βασικές βιβλιοθήκες που αξιοποιούνται είναι το pandas για τον χειρισμό των δεδομένων, η scikit-learn για τον διαχωρισμό και την κανονικοποίηση, και η category_encoders για την κωδικοποίηση των κατηγορικών μεταβλητών. Οι κύριες λειτουργίες του script αναλύονται στα εξής στάδια:

Ενοποίηση Δεδομένων (Data Merging)

Τα αρχεία train_transaction.csv και train_identity.csv φορτώνονται και ενοποιούνται μέσω μιας πράξης αριστερής συνένωσης (left join) με βάση το μοναδικό αναγνωριστικό TransactionID. Το τελικό σύνολο περιλαμβάνει τόσο τις λεπτομέρειες των συναλλαγών (π.χ. ποσό, χρόνος, κάρτα) όσο και τις διαθέσιμες πληροφορίες ταυτοποίησης (π.χ. email, συσκευή).

Διαχείριση Ελλιπών Τιμών (Missing Value Imputation)

Αρχικά, πραγματοποιείται ανάλυση του ποσοστού των ελλιπών τιμών (missing values) για κάθε στήλη του συνόλου δεδομένων. Μεταβλητές οι οποίες παρουσιάζουν περισσότερο από 80% ελλιπείς τιμές απορρίπτονται, καθώς δεν θεωρούνται αξιοποιήσιμες και ενδέχεται να εισαγάγουν θόρυβο.

Για τις εναπομείνασες στήλες, εφαρμόζεται συμπλήρωση τιμών (imputation). Συγκεκριμένα, χρησιμοποιείται η διάμεσος (median) για τις αριθμητικές μεταβλητές προκειμένου να αποφευχθεί η ευαισθησία σε ακραίες τιμές (outliers), ενώ οι κατηγορικές μεταβλητές συμπληρώνονται με τη συμβολοσειρά 'Unknown'.

Μηχανική Χαρακτηριστικών (Feature Engineering)

Για την περαιτέρω ενίσχυση της προγνωστικής ικανότητας των μοντέλων, παράγονται νέα χαρακτηριστικά:

- **Ποσό Συναλλαγής**:

  Δημιουργείται μια λογαριθμική αναπαράσταση (TransactionAmt_log) για την εξομάλυνση της ασύμμετρης κατανομής των ποσών, καθώς και εξαγωγή του καθαρού δεκαδικού μέρους (TransactionAmt_decimal).

  Αυτή η επιλογή βασίζεται στην παρατήρηση ότι οι δόλιες συναλλαγές συχνά παρουσιάζουν ασυνήθιστα δεκαδικά μοτίβα. Για παράδειγμα, οι επιτιθέμενοι (fraudsters) ενδέχεται να δοκιμάζουν ακέραια ποσά (π.χ. 100.00$) ή τυχαία δεκαδικά σε ξένα νομίσματα (π.χ. 43.71$), σε αντίθεση με τις νόμιμες συναλλαγές που συχνά καταλήγουν σε συνηθισμένες τιμές λιανικής (π.χ. .99 ή .50).

- **Χρονικά Χαρακτηριστικά**:

  Από τη μεταβλητή TransactionDT (η οποία αντιπροσωπεύει δευτερόλεπτα από μια σταθερή χρονική στιγμή αναφοράς), εξάγονται κυκλικά χαρακτηριστικά που υποδηλώνουν την ώρα της ημέρας (hour), την ημέρα της εβδομάδας (day) και την εβδομάδα (week).

- **Συχνότητα Κάρτας**:

  Υπολογίζεται η συχνότητα εμφάνισης (frequency count) κάθε κάρτας στο dataset (card1_count), προσθέτοντας έμμεση πληροφορία για την ένταση της συναλλακτικής δραστηριότητας της εκάστοτε κάρτας.

Διαχωρισμός Δεδομένων και Αποφυγή Διαρροής (Train-Val-Test Split & Data Leakage)

Το σύνολο δεδομένων χωρίζεται σε δεδομένα εκπαίδευσης (80%), επικύρωσης (10%) και δοκιμής (10%) με τη συνάρτηση train_test_split.[3] Χρησιμοποιείται στρωματοποιημένη

[3] Είναι θεμελιώδους σημασίας ότι ο διαχωρισμός αυτός πραγματοποιείται πριν από οποιονδήποτε μετασχηματισμό των δεδομένων. Αν τεχνικές όπως το Target Encoding εφαρμοστούν σε ολόκληρο το dataset πριν τον διαχωρισμό, οι τιμές της μεταβλητής-στόχου από το test set θα ενσωματώνονταν στις εκτιμήσεις του train set, δημιουργώντας «διαρροή» πληροφορίας (data leakage) και οδηγώντας σε υπερεκτίμηση της απόδοσης του μοντέλου.

δειγματοληψία (stratify=y) βάσει της μεταβλητής-στόχου isFraud για τη διατήρηση του ποσοστού απάτης σε όλα τα υποσύνολα.

Επιπροσθέτως, τα αναγνωριστικά συναλλαγών για κάθε υποσύνολο αποθηκεύονται σε ξεχωριστά αρχεία (train_ids.csv, val_ids.csv, test_ids.csv), προκειμένου να εγγυηθεί η συνέπεια των διαχωρισμών όταν το dataset μετατραπεί σε γράφο (στα επόμενα notebooks).

Κωδικοποίηση και Κανονικοποίηση (Encoding & Scaling)

- Εφαρμόζεται Target Encoding (TargetEncoder(smoothing=1.0)) στις κατηγορικές μεταβλητές. Ο κωδικοποιητής προσαρμόζεται (fit) αποκλειστικά στα δεδομένα εκπαίδευσης και στη συνέχεια εφαρμόζεται στα σύνολα επικύρωσης και δοκιμής, αντικαθιστώντας τις αρχικές κατηγορίες με την αναμενόμενη πιθανότητα απάτης τους.
- Οι αριθμητικές μεταβλητές κανονικοποιούνται με χρήση της κλάσης StandardScaler, η οποία κλιμακώνει τα χαρακτηριστικά ώστε να έχουν μηδενική μέση τιμή και μοναδιαία διακύμανση, και πάλι με παραμέτρους (mean, std) υπολογισμένες αυστηρά μόνο από το training set.

Προετοιμασία Εξαγωγής για το Neo4j

Το τελευταίο στάδιο ανασυνθέτει το επεξεργασμένο dataset (αποθηκεύοντάς το ως ieee_cis_processed_full.csv) και εξάγει τις οντότητες του γράφου και τις σχέσεις τους σε μορφή ξεχωριστών αρχείων CSV. Αυτά τα αρχεία αποθηκεύονται στον φάκελο data/neo4j_imports/ για να καταναλωθούν από τη βάση δεδομένων:

- **Αρχεία Κόμβων (Nodes)**:

Εξάγονται οι Συναλλαγές με την ετικέτα απάτης (nodes_transactions.csv), οι Κάρτες (nodes_cards.csv), τα Email Domains (nodes_emails.csv), και οι Συσκευές (nodes_devices.csv).

- **Αρχεία Ακμών (Edges)**:

Δημιουργούνται λίστες ακμών για να αποτυπώσουν ποιες οντότητες συσχετίζονται με ποιες συναλλαγές, παράγοντας τα αρχεία edges_transaction_card.csv (σχέση

USED_CARD), edges_transaction_email.csv (σχέση USED_EMAIL) και edges_transaction_device.csv (σχέση USED_DEVICE).

5.2.2 Notebook 2

Το δεύτερο στάδιο της ροής εκτέλεσης αφορά τη μετάβαση από τα παραδοσιακά σχεσιακά tabular δεδομένα σε δομή γράφου, φορτώνοντας τις πληροφορίες στη βάση δεδομένων Neo4j. Η επικοινωνία με τη βάση επιτυγχάνεται μέσω του επίσημου Neo4j Python Driver (ο οποίος έχει ενθυλακωθεί στην προσαρμοσμένη κλάση Neo4jConnector).

Οι βασικές αρχιτεκτονικές επιλογές και λειτουργίες αυτού του σταδίου περιλαμβάνουν:

Εκκαθάριση και Επιβολή Περιορισμών (Constraints & Indexes)

Πριν από την εισαγωγή οποιουδήποτε στοιχείου, η βάση καθαρίζεται από προηγούμενα δεδομένα για την αποφυγή διπλοεγγραφών. Στη συνέχεια, δημιουργούνται μοναδικοί περιορισμοί (Unique Constraints) για όλα τα αναγνωριστικά (π.χ. CREATE CONSTRAINT ... REQUIRE t.transaction_id IS UNIQUE).

Η επιλογή αυτή είναι στρατηγικής σημασίας· πέρα από τη διασφάλιση της ακεραιότητας των δεδομένων (ώστε να μην υπάρξουν ποτέ δύο κόμβοι για την ίδια συναλλαγή ή κάρτα), η δημιουργία constraint αυτόματα δημιουργεί και δομές ευρετηρίου (B-Tree indexes) στη βάση. Αυτό επιταχύνει εκθετικά την εντολή MERGE στα επόμενα βήματα, καθώς η βάση μπορεί να εντοπίσει ακαριαία αν ένας κόμβος υπάρχει ήδη, χωρίς να σαρώνει ολόκληρο τον γράφο.

Μαζική Εισαγωγή Κόμβων και Ακμών (Batch Ingestion)

Τα CSV αρχεία (π.χ. nodes_transactions.csv, edges_transaction_card.csv) φορτώνονται στη μνήμη της Python μέσω pandas και αποστέλλονται στο Neo4j. Αντί η εισαγωγή να γίνει εγγραφή-προς-εγγραφή, χρησιμοποιείται η εντολή UNWIND $nodes AS node της Cypher, στέλνοντας λίστες των 1.000 (για κόμβους) ή 5.000 (για ακμές) εγγραφών ανά συναλλαγή.

Η μαζική εισαγωγή (batching) είναι απολύτως απαραίτητη σε δεδομένα αυτής της κλίμακας, καθώς αποτρέπει την εξάντληση της μνήμης της βάσης δεδομένων

(OutOfMemory errors) και ελαχιστοποιεί τον χρόνο (network overhead) που απαιτείται για την επικοινωνία μεταξύ της Python και του Neo4j server.

Δημιουργία Ακμών Συνεμφάνισης (Card Co-occurrence Edges)

Το σημαντικότερο βήμα μηχανικής χαρακτηριστικών σε επίπεδο γράφου είναι η δημιουργία έμμεσων σχέσεων SHARED_EMAIL_WITH. Μέσω αυτού του ερωτήματος (query), ανιχνεύονται διαφορετικές κάρτες (c1, c2) που χρησιμοποιήθηκαν σε συναλλαγές με την ίδια ακριβώς διεύθυνση email (e). Για κάθε τέτοια ταύτιση, προστίθεται μια ακμή μεταξύ των δύο καρτών, με ένα βάρος (r.weight) που αυξάνεται όσο συχνότερα μοιράζονται το ίδιο email.

Στον πραγματικό κόσμο, τα κυκλώματα απάτης (fraud rings) συχνά χρησιμοποιούν κλεμμένες κάρτες διαφόρων χρηστών, αλλά χρησιμοποιούν έναν κοινό παραλήπτη email ή κοινή συσκευή για τον έλεγχο των αγορών. Δημιουργώντας απευθείας ακμές μεταξύ των καρτών, «συντομεύουμε» τη διαδρομή (path) στον γράφο. Αυτό επιτρέπει στους αλγορίθμους ανίχνευσης κοινοτήτων (όπως ο Louvain στο επόμενο notebook) να αναγνωρίσουν αυτά τα συνδεδεμένα δίκτυα εξαιρετικά γρήγορα και με μεγαλύτερη ακρίβεια.

Για την αποφυγή υπερφόρτωσης κατά τον υπολογισμό όλων των πιθανών συνδυασμών (Cartesian product), αξιοποιείται η συνάρτηση apoc.periodic.iterate της βιβλιοθήκης APOC.

Επαλήθευση και Εξαγωγή Μετρικών

Το notebook ολοκληρώνεται με την εκτέλεση επιλεγμένων ερωτημάτων (Cypher queries) που επιβεβαιώνουν τη δομή του γράφου (π.χ. αναζήτηση καρτών που συνδέονται με πολλαπλές δόλιες συναλλαγές).

Εξάγεται το αρχείο 02_graph_construction_summary.png, το οποίο αποτυπώνει οπτικά την κατανομή των ετικετών (Fraud/Legitimate) και τις ποσότητες κόμβων και σχέσεων που έχουν πλέον δημιουργηθεί επιτυχώς στη βάση δεδομένων.

5.2.3 Notebook 3

Το τρίτο στάδιο αξιοποιεί την ισχυρή βιβλιοθήκη Graph Data Science (GDS) του Neo4j για τον υπολογισμό graph-native χαρακτηριστικών (τοπολογικών μετρικών). Τα χαρακτηριστικά αυτά εξάγονται από τη δομή του δικτύου και στοχεύουν στον εμπλουτισμό της περιγραφικής ικανότητας των αλγορίθμων Μηχανικής Μάθησης. Η υλοποίηση ακολουθεί συγκεκριμένα τεχνικά βήματα για τη βελτιστοποίηση της ταχύτητας και της διαχείρισης μνήμης:

Δημιουργία In-Memory Προβολής (Graph Projection)

Το πρώτο τεχνικό βήμα περιλαμβάνει την εκτέλεση της εντολής gds.graph.project.cypher. Αυτή η διαδικασία προβάλλει τα δεδομένα από την κανονική αποθήκευση της βάσης (transactional database) στη μνήμη RAM του συστήματος, σε έναν in-memory γράφο με το όνομα fraud-graph. Οι αλγόριθμοι γράφων απαιτούν επαναληπτικές (iterative) διελεύσεις σε όλο το δίκτυο, κάτι που θα ήταν απαγορευτικά αργό μέσω τυπικών I/O λειτουργιών δίσκου. Η in-memory προβολή (fraud-graph) εξασφαλίζει εξαιρετικά γρήγορη εκτέλεση των αλγορίθμων. Στα ερωτήματα προβολής (nodeQuery και relQuery), διατηρούνται κρίσιμες ιδιότητες (properties), όπως το is_fraud (μετατρέποντάς το σε float για συμβατότητα με το GDS) και το βάρος (weight) της σχέσης SHARED_EMAIL_WITH.

Υπολογισμός Μετρικών Κεντρικότητας (Centrality Metrics)

Υπολογίζονται τρεις βασικές μετρικές κεντρικότητας για κάθε κόμβο, καθεμία εκ των οποίων αναδεικνύει διαφορετική συμπεριφορά απάτης:

- **Degree Centrality**

  Εφαρμόζεται καλώντας τη διαδικασία gds.degree.write. Αποθηκεύει το πλήθος των ακμών κάθε οντότητας στην ιδιότητα degree_centrality.

- **PageRank**

  Καλείται μέσω της gds.pageRank.write. Στον κώδικα, παραμετροποιείται ρητά με συντελεστή απόσβεσης dampingFactor: 0.85 και μέγιστο αριθμό επαναλήψεων maxIterations: 20 για να εξισορροπηθεί η ακρίβεια με τον χρόνο εκτέλεσης. Η μετρική αποθηκεύεται ως pagerank.

- **Betweenness Centrality**

  Η εντολή gds.betweenness.write έχει υπολογιστική πολυπλοκότητα $O(N^3)$ στο χειρότερο σενάριο, καθιστώντας την απαγορευτική για εκατοντάδες χιλιάδες κόμβους. Γι' αυτόν τον λόγο, στο script ορίζεται υποχρεωτικά τεχνική δειγματοληψίας μέσω των παραμέτρων samplingSize: 5000 και samplingSeed: 42.[4] Έτσι, υπολογίζεται μια προσεγγιστική, αλλά ικανοποιητική για το μοντέλο, τιμή της ιδιότητας betweenness_centrality.

Ανίχνευση Κοινοτήτων (Community Detection)

Χρησιμοποιούνται δύο κλασικοί αλγόριθμοι για ανίχνευση κοινοτήτων:

- **Αλγόριθμος Louvain**

  Εφαρμόζεται για τον διαχωρισμό του γράφου σε διακριτές, πυκνά συνδεδεμένες κοινότητες (clusters). Εφαρμόζεται με τη διαδικασία gds.louvain.write. Για να αποφευχθεί η υπερβολική διάσπαση του γράφου, τίθενται τα όρια maxLevels: 10 και maxIterations: 10,[5] ενώ η παράμετρος includeIntermediateCommunities: false διασφαλίζει ότι αποθηκεύεται μόνο το τελικό ID της κοινότητας (community_id) για κάθε κόμβο.

- **Ασθενώς Συνδεδεμένες Συνιστώσες (Weakly Connected Components - WCC)**

  Η μέθοδος gds.wcc.write αναγνωρίζει τα αποκομμένα «νησιά» συναλλαγών στον γράφο (δηλαδή αποκομμένους υπογράφους) και δημιουργεί την ιδιότητα component_id, η οποία αποθηκεύεται ως ιδιότητα κόμβου (όπως και η community_id).

[4] Η επιλογή της τιμής 5000 ως sampling size (αντιστοιχεί σε περίπου 0,8% του συνολικού πλήθους των κόμβων) θεωρήθηκε αρκετή για να προσφέρει μια στατιστικά αξιόπιστη προσέγγιση, ενώ ταυτόχρονα διατηρεί τον χρόνο εκτέλεσης σε διαχειρίσιμα επίπεδα για το διαθέσιμο hardware, αποτρέποντας την εξάντληση μνήμης. Το 42 είναι μια αυθαίρετη τιμή seed για την ψευδοτυχαία γεννήτρια αριθμών, ώστε να επιλέγει το ίδιο ακριβώς δείγμα κόμβων σε κάθε μελλοντική επανεκτέλεση του κώδικα, καθιστώντας έτσι το πείραμα επαληθεύσιμο και εξασφαλίζοντας την αναπαραγωγιμότητα (reproducibility) της έρευνας.

[5] Με το maxLevels περιορίζονται τα ιεραρχικά επίπεδα συγχώνευσης του αλγορίθμου, αποτρέποντας τη δημιουργία υπερβολικά μεγάλων και ετερογενών (μακροσκοπικών) κοινοτήτων που θα «έκρυβαν» τα μικρότερα, διακριτά κυκλώματα απάτης. Το maxIterations λειτουργεί ως μηχανισμός εξοικονόμησης υπολογιστικών πόρων αποτρέποντας τις περιττές επαναλήψεις. Η τιμή 10 είναι η προεπιλογή (default) και για τις δύο παραμέτρους.

Εξαγωγή Τοπολογικών Χαρακτηριστικών

Μετά τον υπολογισμό των παραπάνω, εκτελείται ένα Cypher query (features_query) που συγκεντρώνει όλες τις μετρικές. Για κάθε συναλλαγή (Transaction), εξάγεται όχι μόνο το δικό της PageRank ή η δική της κοινότητα, αλλά και οι αντίστοιχες μετρικές της συσχετιζόμενης κάρτας (card), του email και της συσκευής (device).

Στατιστική Ανάλυση, Οπτικοποίηση και Αποθήκευση

Τα εξαγόμενα δεδομένα μορφοποιούνται μέσω pandas και οι κενές τιμές (για συναλλαγές που π.χ. δεν είχαν συνδεδεμένο email) καλύπτονται με το 0 (fillna(0)). Δημιουργείται μια σειρά από ιστογράμματα που συγκρίνουν τις κατανομές των μετρικών ανάμεσα σε νόμιμες και δόλιες συναλλαγές, επιβεβαιώνοντας τη χρησιμότητά τους.

Μετά την εκτέλεση του script, θα έχουν παραχθεί τα ακόλουθα αρχεία:

- ✓ Ο πίνακας με τα νέα χαρακτηριστικά αποθηκεύεται ως graph_features.csv, ο οποίος θα ενωθεί αργότερα με τα tabular δεδομένα.
- ✓ Οι οπτικοποιήσεις αποθηκεύονται στο αρχείο 03_graph_analytics_features.png.

5.2.4 Notebook 4

Το τέταρτο στάδιο λειτουργεί ως γέφυρα μεταξύ της ΒΔΓ (Neo4j) και του πλαισίου Βαθιάς Μάθησης (PyTorch / PyG). Αναλαμβάνει την παραγωγή διανυσματικών αναπαραστάσεων (embeddings), τη συγχώνευση όλων των δεδομένων και την κατασκευή του αντικειμένου HeteroData, το οποίο θα τροφοδοτήσει το ΝΔΓ. Η υλοποίηση περιλαμβάνει τις εξής αρχιτεκτονικές και τεχνικές αποφάσεις:

Υπολογισμός Διανυσματικών Αναπαραστάσεων (FastRP Embeddings)

Μέσω της GDS, καλείται η διαδικασία gds.fastRP.write για την παραγωγή node embeddings 128 διαστάσεων (embeddingDimension: 128). Ο αλγόριθμος FastRP (Fast Random Projection) επιλέχθηκε επειδή είναι εξαιρετικά επεκτάσιμος (scalable) και ικανός να συμπυκνώσει την τοπολογική δομή του γράφου σε πυκνά διανύσματα γρήγορα. Τα

διανύσματα αυτά διαχωρίζονται σε 128 ξεχωριστές στήλες (emb_0 έως emb_127) και προστίθενται στο συνολικό σύνολο χαρακτηριστικών.

Συγχώνευση Χαρακτηριστικών και Στρατηγική Διαχωρισμού (Feature Strategy)

Φορτώνονται και ενοποιούνται (μέσω της μεθόδου merge του pandas) τρεις διαφορετικές πηγές:

α) Τα tabular features (από το Notebook 1),

β) Τα graph features (από το Notebook 3),

γ) Τα FastRP embeddings.

Στον κώδικα ορίζεται η μεταβλητή gnn_feature_cols = tabular_features. Αυτό σημαίνει ότι ο πίνακας χαρακτηριστικών X που θα δοθεί στο GNN περιέχει μόνο τα tabular χαρακτηριστικά. Αντιθέτως, τα γραφικά χαρακτηριστικά (graph features) και τα embeddings παραλείπονται επίτηδες από την είσοδο του GNN.

Αυτό συμβαίνει διότι το GNN εκπαιδεύεται για να μάθει από μόνο του τη δομή του γράφου (μέσω message passing). Αν του δίναμε έτοιμα graph features (π.χ. PageRank), θα προκαλούσαμε πλεονασμό πληροφορίας (redundancy) και κίνδυνο υπερπροσαρμογής (overfitting).

Ωστόσο, όλα τα χαρακτηριστικά (tabular + graph + embeddings) διατηρούνται στον πίνακα final_feature_matrix.csv, προκειμένου να δοθούν στο XGBoost (βλ. Notebook 5) ώστε να λειτουργήσει ως ένα εξίσου ισχυρό μοντέλο αναφοράς (baseline).

Κατασκευή Αντικειμένου HeteroData

Αυτή είναι η πιο κρίσιμη τεχνικά διαδικασία για τη μετάβαση σε Heterogeneous GNN. Η δομή HeteroData του PyG απαιτεί τη δημιουργία ξεχωριστών πινάκων (tensors) για κάθε τύπο κόμβου και κάθε τύπο ακμής.

- **Χαρτογράφηση Δεικτών (Index Mappings)**

  Οι συμβολοσειρές αναγνωριστικών (π.χ. string UUIDs του Neo4j) μετατρέπονται σε συνεχείς ακέραιους δείκτες (0, 1, ..., N-1) μέσω λεξικών της Python

(trans_id_to_idx, cardid_to_idx, κλπ.), καθώς οι πίνακες τανυστών (tensors) απαιτούν αυστηρά ακέραια ευρετηρίαση.

- **Χαρακτηριστικά Βοηθητικών Κόμβων (Auxiliary Node Features)**

  Ενώ οι συναλλαγές έχουν πλούσια χαρακτηριστικά, οι βοηθητικοί κόμβοι (Card, Email, Device) στερούνται εγγενών χαρακτηριστικών. Για να ικανοποιηθεί η απαίτηση του PyG ότι κάθε κόμβος πρέπει να έχει έναν πίνακα χαρακτηριστικών X, υπολογίζονται ντετερμινιστικά στατιστικά, όπως ο λογάριθμος του πλήθους των συναλλαγών (np.log1p(tx_count)), και συνδυάζονται με έναν συνεχή αριθμό αναγνώρισης (node_id).

- **Ακμές και Αντίστροφες Ακμές (Reverse Edges)**

  Δημιουργούνται τα edge_index tensors μεγέθους [2, E]. Πέρα από τις κατευθυνόμενες ακμές (π.χ. Transaction -> USED_CARD -> Card), προστίθενται ρητά και οι αντίστροφες (π.χ. Card -> rev_used_card -> Transaction) μέσω της εντολής edge_index.flip(0). Αυτό επιτρέπει την αμφίδρομη ροή μηνυμάτων (bidirectional message passing) κατά τη διάρκεια της εκπαίδευσης του GraphSAGE.

Επιβολή των Μασκών Διαχωρισμού (Masking)

Φορτώνονται τα αρχεία train_ids.csv, val_ids.csv, και test_ids.csv (που εξήχθησαν στο Notebook 1) και δημιουργούνται οι boolean τανυστές train_mask, val_mask, και test_mask. Αυτή η διαδικασία εγγυάται απολύτως ότι το GNN θα αξιολογηθεί ακριβώς στο ίδιο υποσύνολο συναλλαγών με το XGBoost, διασφαλίζοντας την αποφυγή οποιασδήποτε μορφής data leakage.

Εξαγωγή Αρχείων

Με την ολοκλήρωση της διαδικασίας παράγονται και αποθηκεύονται:

- ✓ final_feature_matrix.csv: Ο πλήρης πίνακας χαρακτηριστικών για τα παραδοσιακά ML μοντέλα.
- ✓ feature_lists.json: Λεξικό με τις ονομασίες των στηλών (για να γνωρίζει το XAI ποια στήλη αντιστοιχεί σε ποιο feature).

- ✓ pyg_heterodata.pt: Το δυαδικό αρχείο που περιέχει την πλήρη και έτοιμη προς εκπαίδευση ετερογενή δομή γράφου.
- ✓ Οπτικοποιήσεις των συσχετίσεων και της διακύμανσης των features, στα αρχεία 04_feature_correlations.png και 05_feature_groups_analysis.png.

5.2.5 Notebook 5

Στο πέμπτο στάδιο πραγματοποιείται ο πυρήνας της εκπαίδευσης: η εφαρμογή του αλγορίθμου Heterogeneous GraphSAGE πάνω στο αρχείο pyg_heterodata.pt και η άμεση σύγκρισή του με ένα παραδοσιακό μοντέλο Gradient Boosting (XGBoost).

Η αρχιτεκτονική της εκπαίδευσης δομήθηκε γύρω από τις εξής τεχνικές επιλογές:

Δειγματοληψία Γειτονιάς (Neighbor Sampling)

Αντί να φορτωθεί ολόκληρος ο γράφος στη μνήμη (full-batch training) – κάτι που θα οδηγούσε μαθηματικά σε εξάντληση μνήμης (Out-Of-Memory) λόγω του φαινομένου της «έκρηξης γειτονιάς» (neighborhood explosion)[6] – χρησιμοποιείται η κλάση NeighborLoader.

Το NeighborLoader υλοποιεί εκπαίδευση σε mini-batches (batch_size: 1024), εκτελώντας τυχαία δειγματοληψία σε σταθερό αριθμό γειτόνων ανά επίπεδο (hop). Στον κώδικα ορίστηκε num_neighbors=[5, 5], που σημαίνει ότι για κάθε συναλλαγή του batch, το δίκτυο "κοιτάει" το πολύ 5 άμεσους γείτονες (1ο hop) και το πολύ 5 γείτονες του κάθε γείτονα (2ο hop).[7] Έτσι, η υπολογιστική πολυπλοκότητα παραμένει σταθερή και επεκτάσιμη, ενώ το GraphSAGE μαθαίνει γενικεύσιμα πρότυπα.

[6] Το φαινόμενο αυτό παρατηρείται στα GNNs όταν, για να υπολογιστεί η αναπαράσταση ενός κόμβου σε βάθος πολλαπλών επιπέδων (hops), το δίκτυο πρέπει να συλλέξει αναδρομικά πληροφορίες από όλους τους γείτονές του, τους γείτονες των γειτόνων του, κοκ. Σε πυκνά συνδεδεμένους γράφους, όπως ο δικός μας, ο αριθμός των απαιτούμενων κόμβων αυξάνεται εκθετικά με κάθε βήμα, φτάνοντας γρήγορα στο σημείο όπου ένας και μόνο κόμβος απαιτεί τη φόρτωση ολόκληρου σχεδόν του δικτύου.

[7] Η αρχική δοκιμή έγινε με num_neighbors=[15, 10], ωστόσο (λόγω αδύναμου hardware) ο χρόνος που απαιτούνταν για κάθε epoch ήταν απαγορευτικός. Έπρεπε επομένως να γίνει ένας συμβιβασμός και να επιλεγούν λιγότεροι γείτονες για το υποδέντρο γύρω από κάθε κόμβο.

Διαχείριση Ανισορροπίας Κλάσεων & Δυναμικό Κατώφλι (Dynamic Thresholding)

Στο dataset μας, όπως και στα προβλήματα ανίχνευσης απάτης γενικότερα, η θετική κλάση (Απάτη) συνιστά ένα ελάχιστο ποσοστό των συνολικών συναλλαγών. Το GNN εκπαιδεύεται χρησιμοποιώντας τη συνάρτηση απώλειας BCEWithLogitsLoss(), η οποία συνδυάζει ένα sigmoid layer και την απώλεια (loss) Binary Cross-Entropy για μεγαλύτερη αριθμητική σταθερότητα κατά το backpropagation.

Η χρήση ενός σταθερού κατωφλίου (threshold) πιθανότητας (π.χ. η προεπιλογή threshold = 0.5) για την ταξινόμηση των προβλέψεων δεν λειτουργεί καλά σε μη ισορροπημένα δεδομένα. Για τον λόγο αυτό, υλοποιήθηκε η συνάρτηση get_best_f1_threshold, η οποία υπολογίζει την καμπύλη Ακρίβειας-Ανάκλησης (Precision-Recall Curve) σε κάθε εποχή και βρίσκει το κατώφλι πιθανότητας που μεγιστοποιεί το F1-Score στο σύνολο επικύρωσης (validation set).

Στρατηγική Αξιολόγησης και Early Stopping

Η διαδικασία βελτιστοποίησης παρακολουθείται κατά κύριο λόγο μέσω της μετρικής Average Precision (AP) / AUC-PR, και όχι μέσω της απλής Ακρίβειας (Accuracy), η οποία είναι παραπλανητική σε ανισόρροπα σύνολα. Εφαρμόζεται μηχανισμός Early Stopping με patience = 5, όπου αν η μετρική Validation AP δεν βελτιωθεί για 5 συνεχόμενες εποχές, η εκπαίδευση διακόπτεται για να αποτραπεί η υπερπροσαρμογή (overfitting). Το μοντέλο με την υψηλότερη απόδοση αποθηκεύεται στον δίσκο.

Σύγκριση με Μοντέλο Αναφοράς (XGBoost Baseline)

Για την επαλήθευση της αξίας του ΝΔΓ, εκπαιδεύεται παράλληλα ένας ταξινομητής (classifier) XGBoost. Το XGBoost εκπαιδεύεται στον πλήρη πίνακα χαρακτηριστικών (final_feature_matrix.csv), ο οποίος περιλαμβάνει όχι μόνο τα tabular features, αλλά και τα graph analytics (PageRank, Degree, Communities) και τα FastRP Embeddings.

Το XGBoost χρησιμοποιεί την παράμετρο scale_pos_weight για να εξισορροπήσει την ποινή των σφαλμάτων στις δύο κλάσεις.

Δεν έχει υλοποιηθεί μηχανισμός για δυναμικό ορισμό threshold – χρησιμοποιείται η προεπιλεγμένη του τιμή (0.5).

Εξαγωγή Αποτελεσμάτων και Αρχείων

Στο τέλος της αξιολόγησης στο Test Set, εκτυπώνονται οι Πίνακες Σύγχυσης (Confusion Matrices) και οι μετρικές Precision, Recall, F1, και AUC-PR.

Το notebook καταλήγει στη δημιουργία και αποθήκευση:

- ✓ heterographsage_best.pt: Τα βελτιστοποιημένα βάρη του δικτύου PyTorch.
- ✓ xgboost_baseline.json: Το εκπαιδευμένο μοντέλο XGBoost.
- ✓ 06_hetero_model_comparison.png: Μια σύνθετη οπτικοποίηση (4-panel) που παρουσιάζει την εξέλιξη της απώλειας (Training Loss), τη σύγκριση των καμπυλών Precision-Recall (PR Curves) και ROC, καθώς και ένα ραβδόγραμμα (bar chart) με τις αναλυτικές μετρικές σύγκρισης μεταξύ HeteroGraphSAGE και XGBoost.

5.2.6 Notebook 6

Το τελευταίο στάδιο του pipeline αφορά το XAI. Στόχος αυτής της φάσης είναι μία προσπάθεια – έστω και σε πρωτόλειο επίπεδο – για ερμηνεία των αποφάσεων του συστήματος και κατανόηση της λογικής του.

Η ανάλυση κινείται σε δύο άξονες: την καθολική ερμηνεία (global interpretability), που εξετάζει ποια χαρακτηριστικά επηρεάζουν γενικά τις προβλέψεις απάτης, και την τοπική ερμηνεία (local interpretability), που εξηγεί τους συγκεκριμένους λόγους για τους οποίους μια μεμονωμένη συναλλαγή θεωρήθηκε ύποπτη.

Αξιοποιούνται οι βιβλιοθήκες shap, torch_geometric.explain και networkx.

Καθολική Επεξηγησιμότητα (Global XAI) μέσω SHAP

Αρχικά, φορτώνεται το μοντέλο XGBoost και αναλύεται με τη χρήση του shap.TreeExplainer. Η επιλογή του XGBoost για την καθολική επεξηγησιμότητα δεν είναι τυχαία. Οι Tree-based Explainers της βιβλιοθήκης SHAP είναι ταχύτατοι και υπολογίζουν ακριβείς τιμές Shapley σε τεράστια σύνολα δεδομένων, προσφέροντας μια αξιόπιστη πανοραμική εικόνα. Τα χαρακτηριστικά ομαδοποιούνται ρητά σε τρεις κατηγορίες: “Tabular”, “Graph Analytics”, και “Embedding”. Ο σκοπός αυτής της ομαδοποίησης είναι να καταδειχθεί ποσοτικά η προστιθέμενη αξία του γράφου.

Τα αποτελέσματα εξάγονται στο αρχείο 07_shap_analysis.png.

Τοπική Επεξηγησιμότητα Γράφου (Local XAI) μέσω GNNExplainer

Ενώ το SHAP εξηγεί τα χαρακτηριστικά (features), δεν εξηγεί τη δομή του δικτύου. Για τον λόγο αυτό, επιστρατεύεται η κλάση GNNExplainer του PyG πάνω στο εκπαιδευμένο HeteroGraphSAGE. Ο κώδικας εντοπίζει ένα μοναδικό περιστατικό απάτης με πολύ υψηλή βεβαιότητα (True Positive) και απομονώνει μια εστιασμένη τοπική γειτονιά δύο (2) επιπέδων (2-hop local neighborhood) γύρω από αυτή τη συναλλαγή, μέσω ενός NeighborLoader (με μέγεθος batch = 1).

Η εκτέλεση του GNNExplainer σε ολόκληρο τον ετερογενή γράφο απαιτεί τεράστιους πόρους και χρόνους εκτέλεσης. Λόγω της περιορισμένης υπολογιστικής ισχύος του διαθέσιμου hardware, η καθολική εφαρμογή του αλγορίθμου κρίθηκε απαγορευτική. Ως εκ τούτου, επιλέχθηκε η υλοποίηση μέσω δειγματοληψίας υπογράφου (subgraph sampling). Απομονώνοντας τον υπογράφο (subgraph) γύρω από τη συναλλαγή-στόχο, ο αλγόριθμος μπορεί να εκπαιδεύσει γρήγορα (σε 200 epochs) μια μάσκα ακμών (edge_mask) και μια μάσκα κόμβων (node_mask), εντοπίζοντας ακριβώς ποιες γειτονικές κάρτες/συσκευές/emails και ποια χαρακτηριστικά συνέβαλαν περισσότερο στην πρόβλεψη.

Οπτικοποίηση Υπογράφου Απάτης (NetworkX & Neo4j)

Για να καταστεί η επεξήγηση κατανοητή από τον τελικό χρήστη, οι εσωτερικοί δείκτες του PyTorch (tensor indices) πρέπει να μεταφραστούν ξανά σε πραγματικές οντότητες (π.χ. το πραγματικό ID της κάρτας). Ο κώδικας αντλεί τα πραγματικά αναγνωριστικά (reverse mapping) απευθείας από τη βάση Neo4j. Στη συνέχεια, χρησιμοποιείται η βιβλιοθήκη networkx για να σχεδιαστεί ένα κατευθυνόμενο γράφημα (DiGraph) που δείχνει τη συναλλαγή και τις οντότητες με τη μεγαλύτερη συμβολή (βάσει της βαθμολογίας του GNNExplainer).

Neo4j Fraud Patterns

Το τελικό βήμα του notebook μεταβαίνει από τα μοντέλα μηχανικής μάθησης στην ανάλυση του ίδιου του γράφου, εκτελώντας απευθείας Cypher queries στο Neo4j. Τα μοντέλα XAI

εξηγούν τις παραμέτρους, αλλά ένας οργανισμός χρειάζεται και πρακτική γνώση. Μέσω σύνθετων ερωτημάτων (queries), αντλούνται αυτόματα τα fraud rings (κοινότητες του Louvain με περισσότερες από 5 απάτες), οι κάρτες υψηλού κινδύνου (High-Risk Cards) και τα ύποπτα Email Domains (με βάση το fraud rate και το PageRank).

Εξαγόμενα Αρχεία

Η ολοκλήρωση της ανάλυσης παράγει τα εξής αρχεία τεκμηρίωσης:

- ✓ 07_shap_analysis.png: Καθολική ανάλυση σημαντικότητας χαρακτηριστικών (Global SHAP).
- ✓ 09_gnnexplainer_single_case_subgraph.png: Η τοπολογική αναπαράσταση του υπογράφου (τοπική επεξήγηση).
- ✓ gnnexplainer_single_case_features.csv, 08_gnnexplainer_single_case_features.png: Η αναλυτική λίστα των σημαντικότερων χαρακτηριστικών (node attributes) για τη συγκεκριμένη τοπική συναλλαγή-απάτη.

6. Παρουσίαση και Αξιολόγηση Αποτελεσμάτων

Στο παρόν κεφάλαιο παρουσιάζονται τα αποτελέσματα[8] που προέκυψαν από την εκτέλεση του pipeline ανίχνευσης απάτης. Η αξιολόγηση περιλαμβάνει τη στατιστική ανάλυση του γράφου, τη σύγκριση των επιδόσεων μεταξύ του μοντέλου βαθιάς μάθησης (GNN) και του μοντέλου αναφοράς (XGBoost), καθώς και την ερμηνεία των αποφάσεων μέσω τεχνικών Επεξηγήσιμης Τεχνητής Νοημοσύνης (XAI).

6.1 Προεπεξεργασία και Στατιστική Ανάλυση

Στο πρώτο στάδιο της ανάλυσης (όπως υλοποιήθηκε στο Notebook 1), εξετάστηκαν τα πρωτογενή δεδομένα που παραχωρήθηκαν από τον διαγωνισμό της IEEE-CIS. Η ενδελεχής κατανόηση της κατανομής και της ποιότητας αυτών των δεδομένων αποτέλεσε το θεμέλιο λίθο για τις επακόλουθες φάσεις της μοντελοποίησης. Το αρχικό ενιαίο σύνολο δεδομένων, το οποίο προέκυψε μετά τη συνένωση των αρχείων Συναλλαγών (Transactions) και Ταυτότητας (Identity), περιλάμβανε 590.540 διακριτές συναλλαγές και 434 χαρακτηριστικά. (*Εικόνα 6.1*)

6.1.1 Περιγραφική Στατιστική και Ανισορροπία Κλάσεων

Από το συνολικό πλήθος των συναλλαγών, μόλις οι 20.663 χαρακτηρίστηκαν ως επιβεβαιωμένα δόλιες. Το γεγονός αυτό καταδεικνύει ένα έντονα ανισορροπημένο σύνολο (highly imbalanced dataset), με το ποσοστό απάτης (fraud rate) να ανέρχεται σε μόλις 3,50%. (*Εικόνα* 6.2)

6.1.2 Διαχείριση Ελλιπών Τιμών (Missing Data)

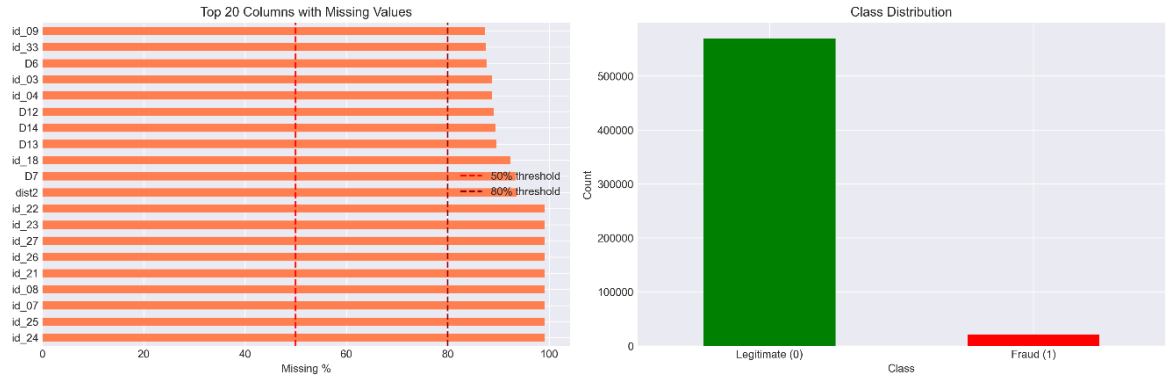
Η αρχική διερεύνηση του συνόλου δεδομένων ανέδειξε ένα σημαντικό πρόβλημα αραιότητας (sparsity). Συγκεκριμένα, 414 στήλες παρουσίαζαν ελλιπή δεδομένα, εκ των

[8] Περιλαμβάνονται αποσπάσματα – σε μορφή screenshots – από την παραγόμενη έξοδο στο τερματικό (terminal output) κατά την εκτέλεση των script, όπως επίσης και αυτούσιες οι παραγόμενες εικόνες με σχήματα και διαγράμματα. Τα πλήρη αποτελέσματα της εξόδου στο τερματικό βρίσκονται στο παράρτημα Α.

οποίων 74 στήλες κατέγραφαν ποσοστό απουσίας πληροφορίας άνω του 80%. Η διατήρηση αυτών των στηλών θα εισήγαγε θόρυβο και θα αύξανε άσκοπα την υπολογιστική πολυπλοκότητα, συνεπώς απορρίφθηκαν. Μετά από αυτό το δραστικό φιλτράρισμα, παρέμειναν 26 κατηγορικά και 332 αριθμητικά χαρακτηριστικά. (*Εικόνα* 6.1) Με την προσθήκη των 6 νέων παραγόμενων μεταβλητών από τη διαδικασία μηχανικής χαρακτηριστικών (TransactionAmt_log, TransactionAmt_decimal, hour, day, week, card1_count), το τελικό προεπεξεργασμένο σύνολο διαμορφώθηκε σε 364 χαρακτηριστικά.

```
🔍 Missing Data Summary:
   Columns with missing data: 414
   Columns with >50% missing: 214
   Columns with >80% missing: 74

[3/8] Feature Selection & Cleaning...
✓ Dropped 74 columns with >80% missing

📋 Feature Types:
   Categorical: 26
   Numerical: 332

🔧 Handling Missing Values...
✓ Missing values handled
   Remaining nulls: 0
```

6.1 Διαχείριση ελλιπών τιμών – Πλήθος εναπομεινάντων χαρακτηριστικών

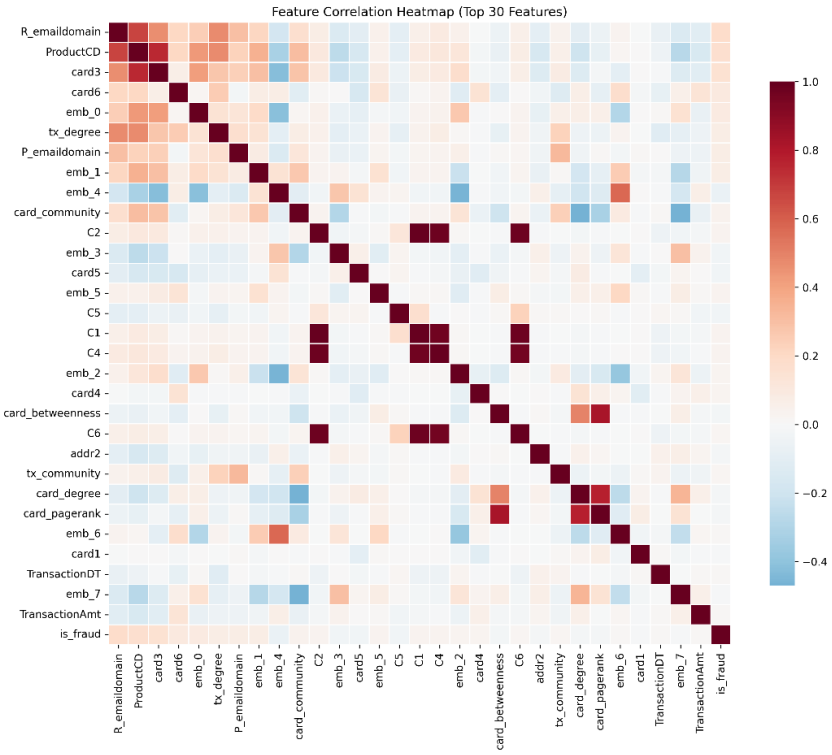


6.2 Αριστερά, οι 20 στήλες με το μεγαλύτερο ποσοστό ελλιπών τιμών. Δεξιά, η κατανομή των δύο κλάσεων.

6.1.3 Ανάλυση Συσχετίσεων (Correlations) και Ενδείξεις Απάτης

Όπως προέκυψε από τη στατιστική ανάλυση στο Notebook 4, ορισμένα χαρακτηριστικά παρουσίασαν έντονη στατιστική εξάρτηση με τη μεταβλητή-στόχο. Τα χαρακτηριστικά με την ισχυρότερη θετική συσχέτιση ήταν το R_emaildomain (0,1828), το ProductCD (0,1684) και το card3 (0,1540). (*Εικόνες 6.3, 6.4*)

Μια ερμηνεία αυτού του αποτελέσματος, σε πρακτικό επίπεδο, θα ήταν ότι ο πάροχος ηλεκτρονικού ταχυδρομείου (R_emaildomain) και η συγκεκριμένη κατηγορία προϊόντος (ProductCD) αποτελούν χαρακτηριστικά της συμπεριφοράς των δραστών, οι οποίοι συχνά στοχεύουν σε εύκολα ρευστοποιήσιμα αγαθά χρησιμοποιώντας συγκεκριμένες ανώνυμες ή προσωρινές υπηρεσίες email.



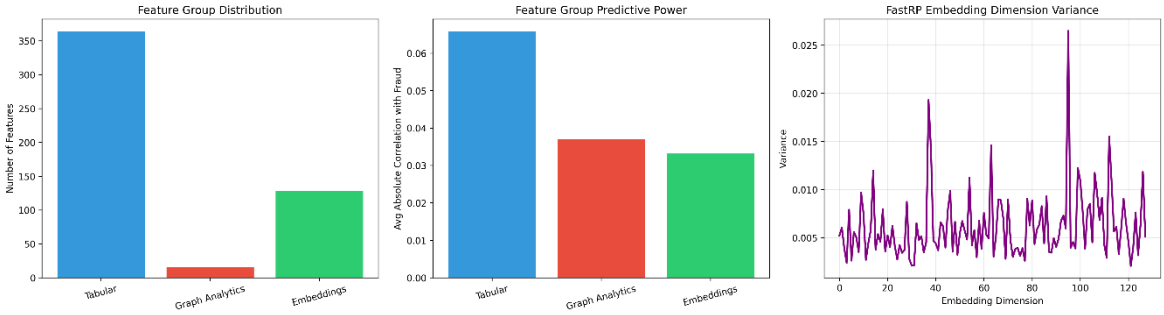
6.3 Θερμικός χάρτης (Heatmap) συσχέτισης χαρακτηριστικών - απάτης

```
🔍 Top 20 Features Most Correlated with Fraud:
Feature                                  Correlation
------------------------------------------------------------
 1. R_emaildomain                            0.1828
 2. ProductCD                                0.1684
 3. card3                                    0.1540
 4. card6                                    0.1005
 5. emb_0                                    0.0905
 6. tx_degree                                0.0806
 7. P_emaildomain                            0.0790
 8. emb_1                                    0.0722
 9. card_community                           0.0439
10. C2                                       0.0372
11. emb_5                                    0.0319
12. C1                                       0.0306
13. C4                                       0.0304
14. emb_2                                    0.0263
15. card4                                    0.0250
16. C6                                       0.0209
17. tx_community                             0.0188
18. emb_6                                    0.0154
19. TransactionDT                            0.0131
20. TransactionAmt                           0.0113
```

6.4 Χαρακτηριστικά συσχετιζόμενα με απάτη – Top 20

6.1.4 Προγνωστική Ισχύς Χαρακτηριστικών

Όπως παρατηρείται από ανάλυση των κατηγοριών χαρακτηριστικών, τα παραδοσιακά οικονομικά δεδομένα (Tabular) κατέχουν την υψηλότερη μέση προγνωστική ισχύ. Ωστόσο, οι μετρικές γράφων (Graph Analytics) και τα διανύσματα ενσωμάτωσης (FastRP Embeddings) συνεισφέρουν συμπληρωματικά κρίσιμη τοπολογική πληροφορία. (*Εικόνα 6.5*) Δεν επικαλύπτονται με τα οικονομικά μεγέθη, επιτρέποντας στο μοντέλο να αντλεί πολύτιμα, «κρυφά» σήματα απάτης από τις πολλαπλές διαστάσεις του δικτύου.



6.5 Κατανομή και προγνωστική ισχύς χαρακτηριστικών - Διακύμανση embeddings

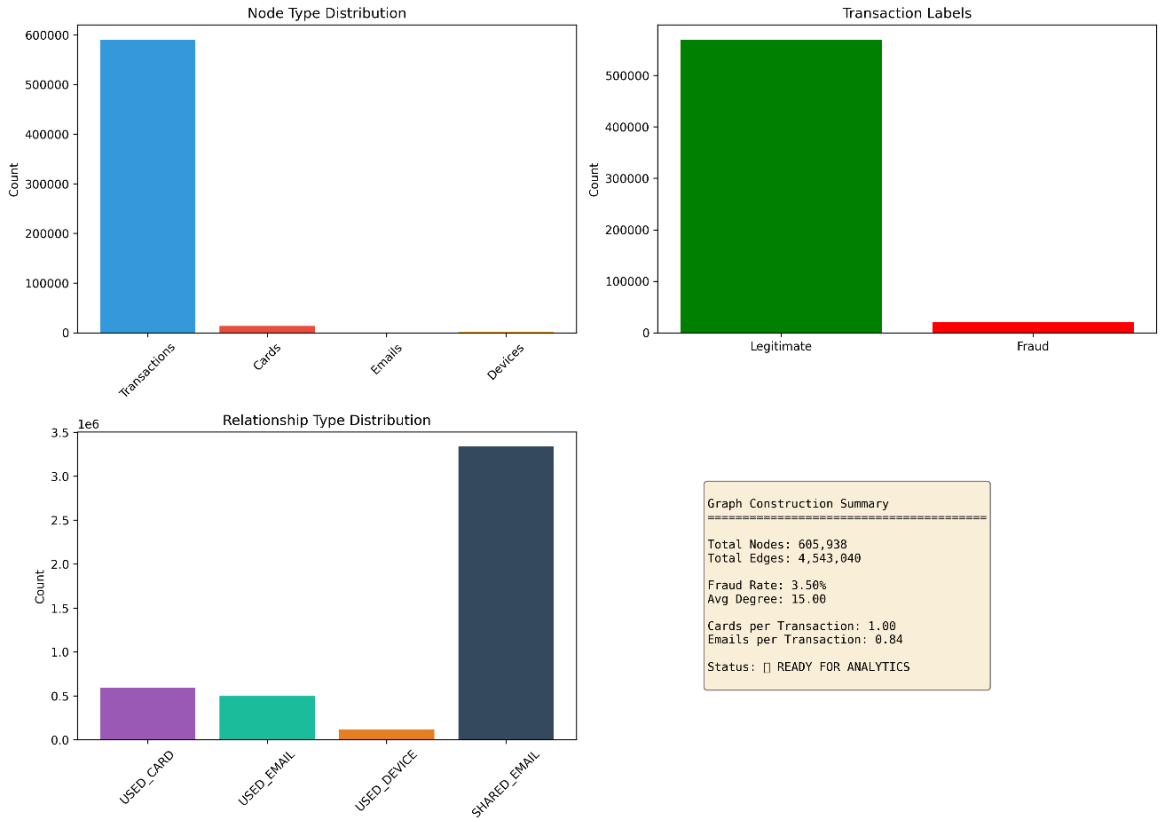
6.2 Τοπολογικά Χαρακτηριστικά και Δομή Γράφου

Η μετάβαση από τον παραδοσιακό δισδιάστατο πίνακα δεδομένων σε μια πλούσια πολυδιάστατη δομή γράφου (Notebook 2) και η επακόλουθη ανάλυσή τους μέσω της βιβλιοθήκης GDS (Notebook 3) αποκάλυψε την πολύπλοκη τοπολογία των συναλλαγών.

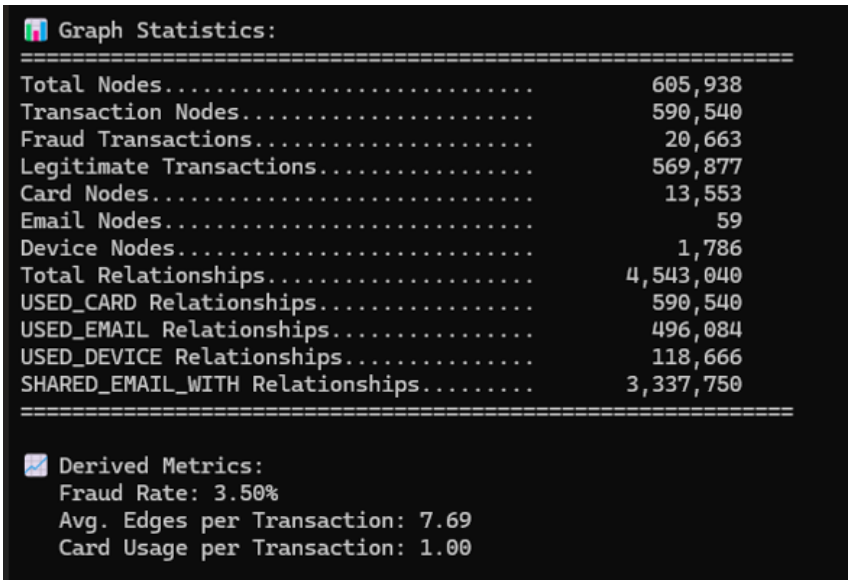
6.2.1 Μετρικές Δικτύου και Ετερογένεια

Η διαδικασία κατασκευής απέδωσε έναν ετερογενή γράφο με 605.938 συνολικούς κόμβους και 4.543.040 σχέσεις (ακμές). Αναλυτικότερα, ο γράφος δεν αποτελείται μόνο από τις 590.540 συναλλαγές, αλλά εμπλουτίζεται περιφερειακά με 13.553 κόμβους καρτών, 59 email domains και 1.786 διακριτές συσκευές. (*Εικόνες 6.6, 6.7*)

Ο μέσος αριθμός ακμών ανά συναλλαγή διαμορφώθηκε στο 7,69. (*Εικόνα 6.7*) Αυτή η συνδεσιμότητα παρέχει το αναγκαίο υπόβαθρο ώστε το μοντέλο να μπορεί να πλοηγηθεί, από μια ύποπτη συναλλαγή, πίσω στην κάρτα που χρησιμοποιήθηκε, και από εκεί σε όλες τις προηγούμενες συναλλαγές της ίδιας κάρτας.



6.6 Κατανομές κόμβων, ακμών και βασικά στοιχεία γράφου



6.7 Στατιστικά γράφου

6.2.2 Ανάλυση Κεντρικότητας

Μια ενδιαφέρουσα παρατήρηση αφορά τη σύγκριση των μετρικών κεντρικότητας μεταξύ νόμιμων και δόλιων συναλλαγών. Παρατηρήθηκε ότι οι δόλιες συναλλαγές τείνουν σταθερά να έχουν υψηλότερο βαθμό κεντρικότητας (μέσο tx_degree 2,28 έναντι 2,03 των νόμιμων). (*Εικόνα 6.8*) Αν και η διαφορά φαντάζει μικρή, σε ένα δίκτυο εκατομμυρίων ακμών μπορεί να είναι στατιστικά κρίσιμη.

```
Feature comparison (Fraud vs Legitimate):
Feature              Fraud Mean      Legit Mean      Ratio
------------------------------------------------------------
tx_degree              2.281518        2.032275      1.12x
tx_pagerank            0.150000        0.150000      1.00x
card_degree         2403.075255     2569.577588      0.94x
card_pagerank        278.536958      310.829956      0.90x
```

6.8 Μέσος όρος βαθμού και PageRank για συναλλαγές και κάρτες σε δόλια και νόμιμα παραδείγματα

6.2.3 Ανάλυση Κοινοτήτων

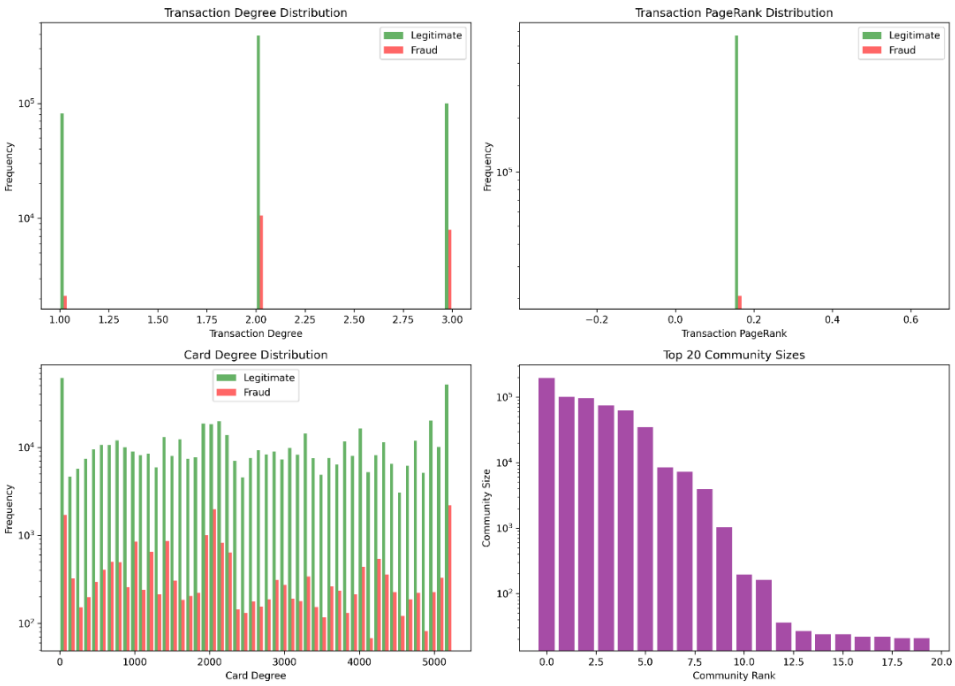
Η εφαρμογή του αλγορίθμου Louvain για την ανίχνευση κοινοτήτων αποκάλυψε 403 διακριτές κοινότητες, παρουσιάζοντας συνολικό βαθμό σπονδυλωτής δομής (modularity) 0.299. (*Εικόνα 6.9*) Το modularity αυτό επιβεβαιώνει ότι ο γράφος διαθέτει μια υπαρκτή, αν και περίπλοκη, δομή υποομάδων.

Ενδιαφέρον εύρημα ήταν ο εντοπισμός μεγάλων κοινοτήτων με υψηλή συγκέντρωση απάτης. Για παράδειγμα, η κοινότητα "598965" (με 34.673 μέλη) εμφάνισε fraud rate 9,3%, ενώ η κοινότητα "605081" (με 74.628 μέλη) είχε fraud rate 7,3%. (*Εικόνα 6.9*)

```
[6/10] Running Community Detection (Louvain Algorithm)...
✓ Louvain clustering completed:
   Communities detected: 403
   Modularity: 0.2990
   Nodes clustered: 605,938
   Time: 24.35s

   Top 20 communities by fraud rate (size >= 10):
   Community    Size     Frauds   Fraud Rate
   ------------------------------------------------
   598965       34673    3210          9.3%
   605081       74628    5422          7.3%
   604170       36       2             5.6%
   604155       1024     51            5.0%
   604165       7214     251           3.5%
   604434       197786   5221          2.6%
   590977       102159   2673          2.6%
   590900       62819    1564          2.5%
   605517       3984     90            2.3%
   604095       96116    2029          2.1%
   604153       8534     129           1.5%
   604399       163      1             0.6%
   604179       196      1             0.5%
```

6.9 Ανάλυση κοινοτήτων (Louvain) – Top 13

Μια συγκεντρωτική εικόνα των αναλύσεων του γράφου φαίνεται παρακάτω:



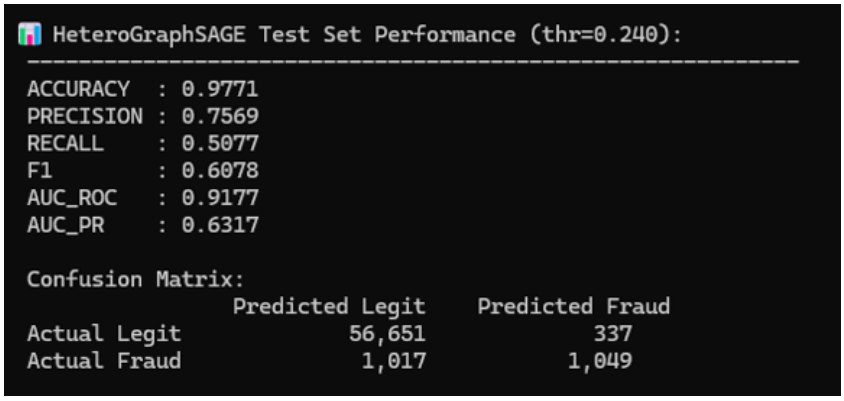
6.10 Κατανομές αλγορίθμων graph analytics

6.3 Αποτελέσματα Εκπαίδευσης

6.3.1 Συγκριτική Αξιολόγηση

Ο πυρήνας της πειραματικής αξιολόγησης (όπως υλοποιήθηκε στο Notebook 5) εστιάζει στην αντιπαραβολική σύγκριση ανάμεσα στο προηγμένο ετερογενές μοντέλο ΝΔΓ (HeteroGraphSAGE) και σε ένα βελτιστοποιημένο μοντέλο αναφοράς XGBoost. Η αξιολόγηση πραγματοποιήθηκε στο ανεξάρτητο Test Set των 59.054 συναλλαγών. Όπως θα δούμε παρακάτω, το ΝΔΓ επικράτησε σε περισσότερες μετρικές, όμως και το XGBoost είχε τα δικά του δυνατά σημεία.

Λόγω της προαναφερθείσας ακραίας ανισορροπίας κλάσεων, μετρικές όπως η απλή Ακρίβεια (Accuracy) θεωρούνται παραπλανητικές. Συνεπώς, η κύρια μετρική αξιολόγησης και βελτιστοποίησης ορίστηκε η Επιφάνεια κάτω από την Καμπύλη Ακρίβειας-Ανάκλησης (AUC-PR) καθώς και το F1-Score, το οποίο βελτιστοποιήθηκε δυναμικά μέσω αλγορίθμου στο validation set, καταλήγοντας σε ιδανικό threshold (κατώφλι) 0,240 για το GNN. (*Εικόνα 6.10*)



6.11 Η απόδοση του ΝΔΓ – Το κατώφλι για το F1 ορίστηκε στο 0.24 μετά από δυναμική βελτιστοποίηση κατά την εκπαίδευση.

Η σύγκριση των επιδόσεων των δύο μοντέλων αποτυπώνεται στον παρακάτω πίνακα:

	XGBoost	**HeteroGraphSAGE**
AUC-PR	0.6248	**0.6317**
AUC-ROC	**0.9318**	0.9177
Precision	0.2330	**0.7569**
Recall	**0.8107**	0.5077
F1-Score	0.3620	**0.6078**
Accuracy	0.9000	**0.9771**

6.1 Συγκριτικός Πίνακας Επιδόσεων

6.3.2 Ανάλυση Συμπεριφοράς και Επιχειρηματικός Αντίκτυπος

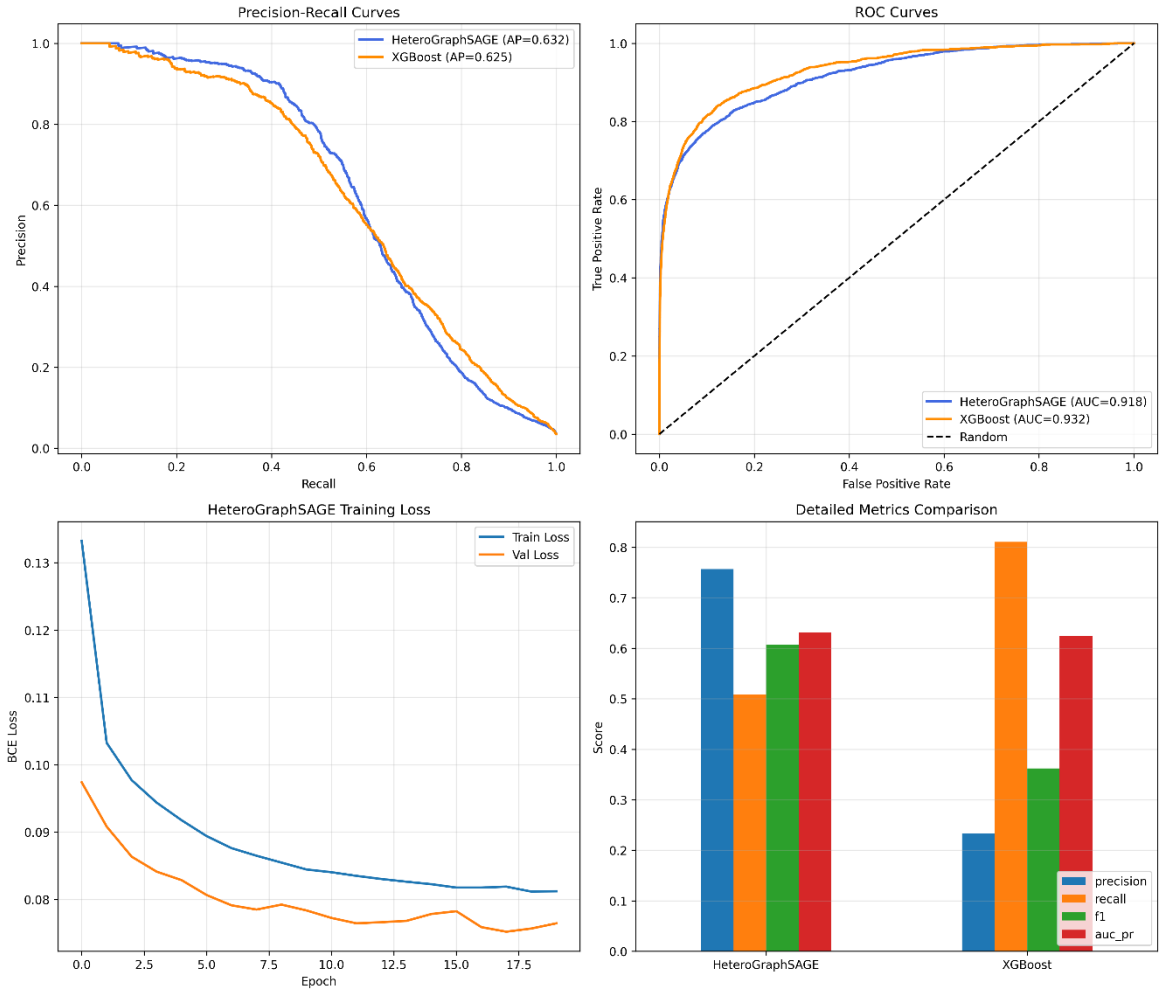
Η μελέτη του πίνακα 6.1 αποκαλύπτει δύο θεμελιωδώς διαφορετικές προσεγγίσεις στην πρόβλεψη. Από τη μία, το XGBoost πέτυχε ένα υψηλό Recall (0,8107), γεγονός που σημαίνει ότι ο αλγόριθμος μπόρεσε να «πιάσει» το 81% περίπου των πραγματικών κρουσμάτων απάτης. Ωστόσο, το τίμημα για αυτή την ευαισθησία ήταν η ακρίβειά του (Precision), η οποία έπεσε στο 0,2330. Υπερτέρησε επίσης στο AUC-ROC με 0,9318 έναντι 0,9177 του GNN, όμως το F1-Score έμεινε χαμηλά (0,3620).

Στον αντίποδα, το HeteroGraphSAGE (GNN) περιόρισε δραστικά τα false positives (σημειώθηκαν μόλις 337 λάθη σε σύνολο 56.651 νόμιμων συναλλαγών). (*Εικόνα 6.10*) Αν και το Recall ήταν χαμηλότερο (0,5077), το Precision ανέβηκε στο 0,7569, επιτρέποντας

στο GNN να κυριαρχήσει στο F1-Score με 0,6078 (μια βελτίωση της τάξης του 68% συγκριτικά με το XGBoost). Όσον αφορά την βασική μετρική AUC-PR της έρευνας, το GNN επικράτησε οριακά (0,6317 έναντι 0,6248 του XGBoost).

Τα παραπάνω αποτελέσματα μας οδηγούν στο εξής συμπέρασμα: το GNN υπερέχει όταν προτεραιότητα είναι η μείωση των false positives, ενώ το XGBoost παραμένει ισχυρότερο όταν στόχος είναι ο εντοπισμός όσο το δυνατόν περισσότερων απατηλών συναλλαγών.

Η σύγκριση αποτυπώνεται οπτικά στα παρακάτω διαγράμματα:



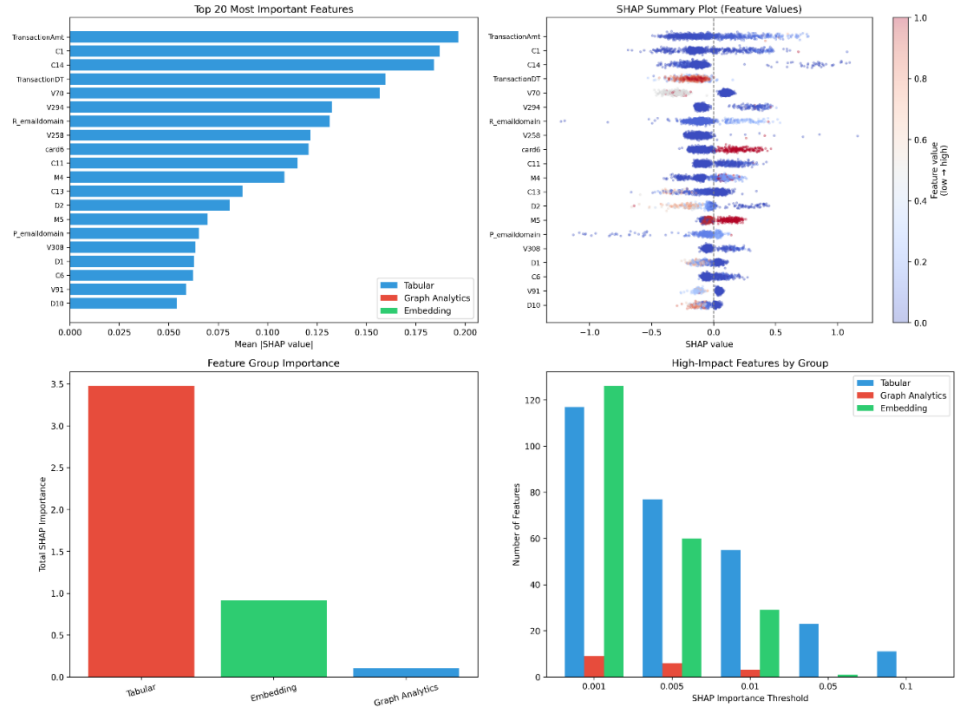
6.12 Σύνθετη οπτικοποίηση (4-panel) που περιλαμβάνει τις Καμπύλες Precision-Recall και ROC, την ομαλή εξέλιξη της καμπύλης εκπαίδευσης του GNN, καθώς και τη γραφική σύγκριση των μετρικών

6.4 Ερμηνευσιμότητα/Επεξηγησιμότητα (XAI)

Μία από τις πιο συνήθεις κριτικές στα σύγχρονα μοντέλα βαθιάς μάθησης είναι η αδιαφάνειά τους. Στο πλαίσιο αυτής της πρόκλησης, η ανάλυση στο Notebook 6 εστίασε στην Επεξηγήσιμη Τεχνητή Νοημοσύνη (XAI), προσεγγίζοντας το πρόβλημα τόσο σε μακροσκοπικό (καθολικό) όσο και σε μικροσκοπικό (τοπικό) επίπεδο.

6.4.1 Καθολική Επεξηγησιμότητα (Global Importance)

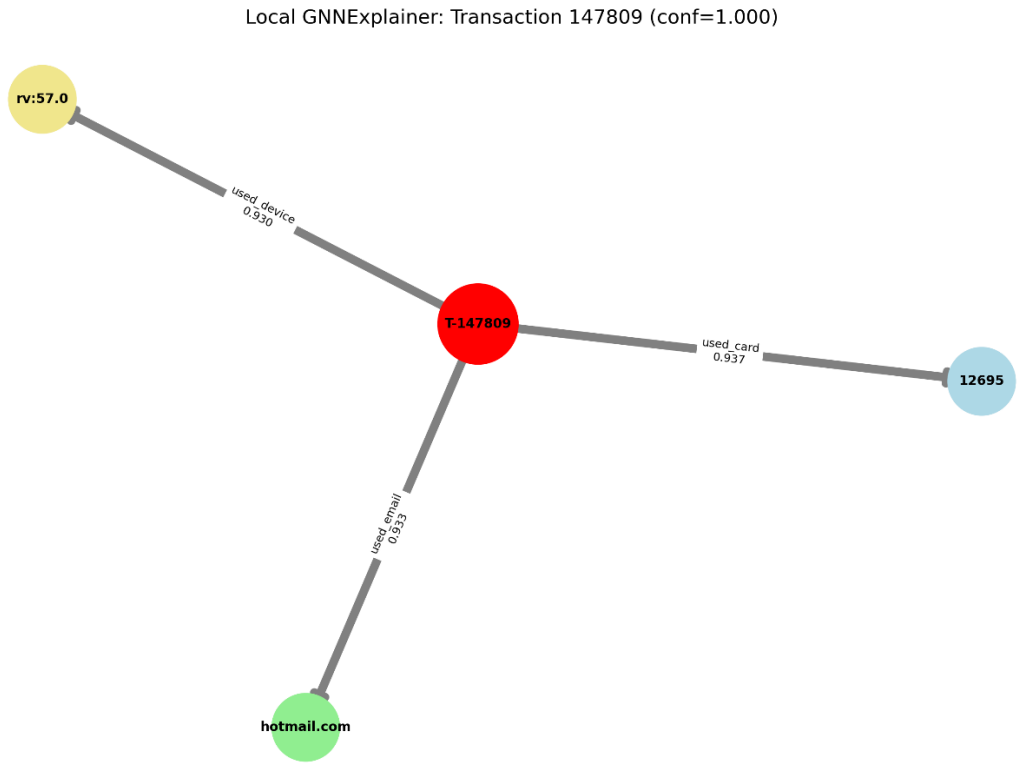
Αναλύοντας το μοντέλο XGBoost με τον αλγόριθμο TreeExplainer της βιβλιοθήκης SHAP, κατασκευάστηκε μια πανοραμική εικόνα της συμπεριφοράς του συστήματος. Η ανάλυση έδειξε ότι τα παραδοσιακά (tabular) χαρακτηριστικά – όπως το ποσό της συναλλαγής (TransactionAmt), τα χαρακτηριστικά C1 και C14, καθώς και η χρονική σήμανση TransactionDT – είχαν κατά μέσο όρο τη μεγαλύτερη βαρύτητα στη λήψη των αποφάσεων. (*Εικόνα 6.12*) Το XGBoost, επομένως, βάσισε κατά κύριο λόγο τις προβλέψεις του στα tabular features, ενώ τα embeddings πρόσθεσαν συμπληρωματική πληροφορία, και τα graph analytics είχαν πολύ μικρότερη συνολική επίδραση. Η τελική απόφαση επηρεάστηκε κυρίως από άμεσες αριθμητικές ενδείξεις συναλλαγής και δευτερευόντως από δομικά σήματα του γράφου.



6.13 Σύνθετη οπτικοποίηση (4-panel) που περιλαμβάνει τα βασικά αποτελέσματα της ανάλυσης SHAP

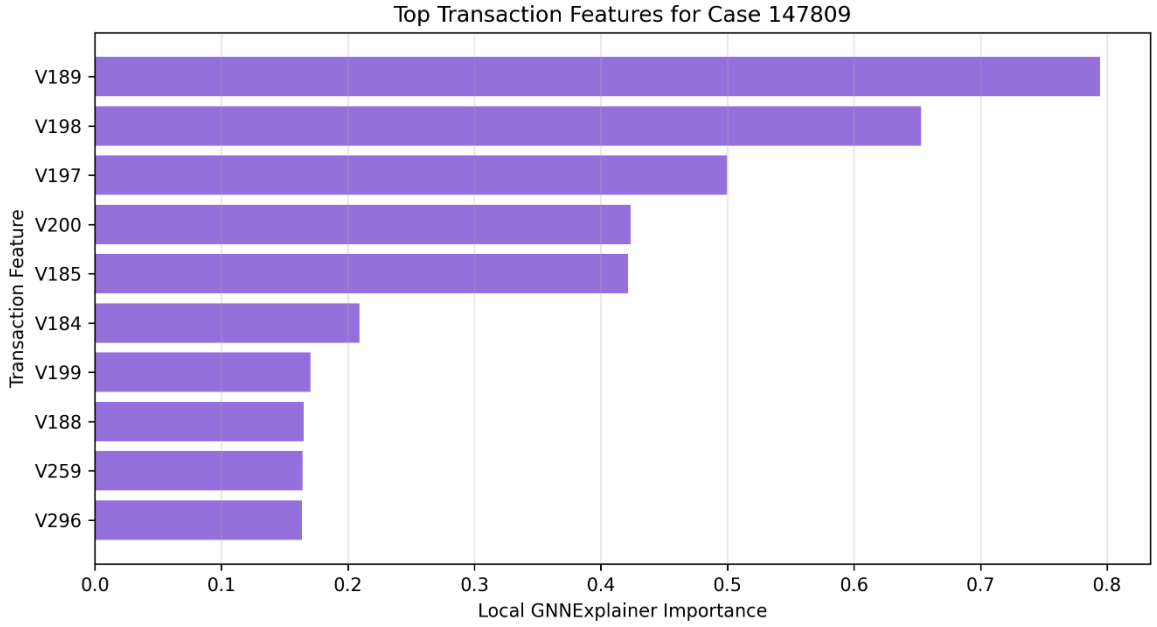
6.4.2 Τοπική Ερμηνευσιμότητα Γράφου (Local Interpretability)

Για να κατανοήσουμε ακριβώς πώς το HeteroGraphSAGE εκμεταλλεύεται τον γράφο, εξετάσαμε μια συγκεκριμένη μελέτη περίπτωσης. Το σύστημα απομόνωσε μια πραγματική δόλια συναλλαγή (Case ID: 147809) την οποία το GNN είχε ταξινομήσει με 100% βεβαιότητα (confidence score = 1.0). Εκπαιδεύοντας το GNNExplainer αποκλειστικά στον τοπικό υπογράφο της συγκεκριμένης συναλλαγής (ένα δίκτυο 2 επιπέδων με 13 συναλλαγές, 5 κάρτες, 1 email, και 1 συσκευή), αναδείχθηκε ο μηχανισμός απόφασης. Το οπτικό αποτέλεσμα έδειξε ότι η πρόβλεψη του μοντέλου στηρίχθηκε κυρίως στη σύνδεση με την Κάρτα 12695 και το email domain "hotmail.com". (*Εικόνα 6.14*)[9] Παράλληλα, τα χαρακτηριστικά του κόμβου στα οποία βασίστηκε για την απόφαση, ήταν οι μεταβλητές V189, V198 και V197. (*Εικόνα 6.15*)



6.14 Επεξήγηση της συναλλαγής 147809 (κόκκινος κόμβος) με χρήση GNNExplainer

[9] Να σημειωθεί εδώ ότι η οπτικοποίηση δεν περιλαμβάνει ολόκληρη τη γειτονιά που χρησιμοποιήθηκε ως δείγμα, αλλά μόνο τους κόμβους κάθε κατηγορίας οντοτήτων που είχαν τη μεγαλύτερη συνεισφορά στην απόφαση.



6.15 Τα χαρακτηριστικά με τη μεγαλύτερη σημαντικότητα για τη συναλλαγή 147809

6.4.3 Ανακάλυψη Μοτίβων Απάτης (Business Insights & Intelligence)

Η δημιουργία του γράφου και η χρήση της βάσης Neo4j δεν περιορίζεται αποκλειστικά στην υποστήριξη των μοντέλων Μηχανικής Μάθησης. Ο ίδιος ο γράφος λειτουργεί ως ένα δυναμικό, αυτόνομο εργαλείο Επιχειρηματικής Ευφυΐας (Business Intelligence), επιτρέποντας την εκτέλεση εξειδικευμένων ερωτημάτων (Cypher queries) για την ανάκληση στρατηγικών πληροφοριών σε πραγματικό χρόνο.

Οργανωμένα Κυκλώματα Απάτης (Fraud Rings)

Εντοπίστηκαν 10 οργανωμένα κυκλώματα απάτης (ως διακριτές κοινότητες Louvain με ≥ 5 καταγεγραμμένες απάτες έκαστην). Ο μεγαλύτερος όγκος απατών, σε απόλυτα νούμερα, καταγράφηκε στην Κοινότητα 605081 (5.422 δόλιες συναλλαγές, ποσοστό 7,3%), ενώ το υψηλότερο ποσοστό επικινδυνότητας/απάτης σημειώθηκε στην Κοινότητα 598965 (9,3%, 3.210 απάτες σε 34.673 συναλλαγές). (*Εικόνα 6.16*) Η παρακολούθηση αυτών των κυκλωμάτων επιτρέπει ενδεχόμενες προληπτικές ενέργειες σε ολόκληρες παρτίδες από κάρτες προτού χρησιμοποιηθούν ξανά.

🚨 Top 10 Detected Fraud Rings:

Community	Size	Frauds	Fraud Rate
605081	74628	5422	7.3%
604434	197786	5221	2.6%
598965	34673	3210	9.3%
590977	102159	2673	2.6%
604095	96116	2029	2.1%
590900	62819	1564	2.5%
604165	7214	251	3.5%
604153	8534	129	1.5%
605517	3984	90	2.3%
604155	1024	51	5.0%

6.16 Fraud rings – Top 10

Κάρτες Υψηλού Κινδύνου (High-Risk Cards)

Μέσω της ανάλυσης βαθμού (degree analysis), ταυτοποιήθηκαν 20 κάρτες που χρησιμοποιούνται κατ' επανάληψη σε απάτες. Η χαρακτηριστικότερη περίπτωση είναι η Κάρτα 9633, η οποία με συνολικό βαθμό συνδέσεων (degree) 2012, χρησιμοποιήθηκε σε 742 επιβεβαιωμένες απάτες. (*Εικόνα 6.17*) Μια τέτοια οντότητα – ενδεχομένως προϊόν κλοπής ή διαρροής δεδομένων στο dark web – είναι πιθανό να δρα ως κεντρικός κόμβος αναδιανομής (hub) σε επιθέσεις.

Αξίζει να σημειωθεί και η εμφάνιση της Κάρτας 12695 σε αυτήν τη λίστα, η οποία εμπλέκεται στην μελέτη περίπτωσης εφαρμογής GNNExplainer που είδαμε πιο πάνω (βλ. υποενότητα 6.4.2).

💳 Top 20 High-Risk Cards:

Card ID	Frauds	PageRank	Degree
9633	742	333.484440	2012
9500	528	1338.762788	5211
15885	444	678.661432	2141
9026	397	222.131441	1925
15063	319	375.128332	4279
5812	314	286.340207	2064
2616	314	591.574211	5178
15066	313	829.429859	5155
9917	306	122.184788	1247
6019	294	630.682658	5210
3154	286	367.600261	2211
17188	278	1015.019899	5135
7585	263	602.880039	5239
14276	252	146.317708	1452
2256	235	171.425242	1896
4504	210	150.712813	2009
10616	202	614.153393	5227
12695	201	754.523568	4396
4461	199	259.736809	2273
8755	199	157.170193	2025

6.17 High-risk cards – Top 20

Ύποπτα Email Domains

Η ανάλυση ανέδειξε ποιοι πάροχοι υπηρεσιών email συγκέντρωσαν τα μεγαλύτερα ποσοστά απατηλών συναλλαγών. Το protonmail.com, μια υπηρεσία γνωστή για την ισχυρή κρυπτογράφηση και την ανωνυμία της, βρέθηκε στην κορυφή της λίστας με 40,8% ποσοστό απάτης (31 frauds σε 76 συναλλαγές συνολικά), με το mail.com να ακολουθεί με 19% (106 frauds σε 559 συναλλαγές). (*Εικόνα 6.18*) Τα ευρήματα αυτά θα μπορούσαν, έπειτα από περεταίρω ανάλυση, να τροφοδοτήσουν συστήματα βασισμένα σε κανόνες (rule-based engines) για άμεση σήμανση (flagging).

Top 15 Suspicious Email Domains:

Email Domain	Total	Frauds	Fraud Rate	PageRank
protonmail.com	76	31	40.8%	4.336250
mail.com	559	106	19.0%	33.746250
outlook.es	438	57	13.0%	22.080000
aim.com	315	40	12.7%	19.275000
outlook.com	5096	482	9.5%	286.026250
hotmail.es	305	20	6.6%	15.938750
live.com.mx	749	41	5.5%	38.506250
hotmail.com	45250	2396	5.3%	2489.077500
gmail.com	228355	9943	4.4%	13611.986250
yahoo.fr	143	5	3.5%	7.566250
embarqmail.com	260	9	3.5%	15.237500
mac.com	436	14	3.2%	21.888750
icloud.com	6267	197	3.1%	386.071250
comcast.net	7888	246	3.1%	457.662500
charter.net	816	25	3.1%	47.346250

6.18 Suspicious email domains – Top 15

7. Συμπεράσματα και Μελλοντική Εργασία

7.1 Σύνοψη ευρημάτων

Η παρούσα πτυχιακή εργασία είχε ως κύριο στόχο την ανάπτυξη ενός πρωτοτύπου για την ανίχνευση οικονομικής απάτης, το οποίο υπερβαίνει τις παραδοσιακές μεθόδους ανάλυσης και προσεγγίζει τις συναλλαγές υπό το πρίσμα της θεωρίας γράφων. Αξιοποιώντας το σύνολο δεδομένων IEEE-CIS Fraud Detection, το πρόβλημα μοντελοποιήθηκε ως ένας ετερογενής γράφος εντός της ΒΔΓ Neo4j, ενσωματώνοντας οντότητες όπως συναλλαγές, κάρτες, διευθύνσεις ηλεκτρονικού ταχυδρομείου και συσκευές.

Η συνδυαστική αξιολόγηση των αποτελεσμάτων του κεφαλαίου 6 επιβεβαιώνει την κεντρική υπόθεση και τον πυρήνα της ερευνητικής μας προσέγγισης και συμφωνεί με τη διεθνή βιβλιογραφία, ότι δηλαδή η τοπολογική πληροφορία και οι σχέσεις μεταξύ των οντοτήτων, όπως αυτά εξάγονται από έναν γράφο αλληλοσυνδεόμενων συναλλαγών και αξιοποιούνται από ένα GNN, προσφέρουν μια αξιόλογη ικανότητα διάκρισης μεταξύ νόμιμων και δόλιων συναλλαγών, και παρέχουν συγκεκριμένα πλεονεκτήματα στον κρίσιμο τομέα της ανίχνευσης οικονομικής απάτης. Το προτεινόμενο μοντέλο Heterogeneous GraphSAGE κατόρθωσε να φιλτράρει ένα μεγάλο μέρος του θορύβου και αποδείχθηκε ικανό να εντοπίζει μοτίβα, αξιολογώντας το ευρύτερο ψηφιακό αποτύπωμα των εμπλεκόμενων οντοτήτων, φτάνοντας σε ένα πρακτικά βιώσιμο επίπεδο ικανότητας διάκρισης. Ως αδύναμο σημείο, εντοπίζεται η μέτρια επίδοσή του ως προς το ποσοστό των συνολικών απατών που κατάφερε να ανιχνεύσει.

Από τη μεριά τους, οι παραδοσιακοί αλγόριθμοι εξακολουθούν να αποτελούν αξιόπιστα, ισχυρά εργαλεία, τα οποία μπορούν να επιτύχουν υψηλά ποσοστά ανάκλησης (recall). Βέβαια, κάτι τέτοιο μπορεί να συνοδεύεται με κόστος σε ακρίβεια, παράγοντας έναν σημαντικό όγκο ψευδώς θετικών συναγερμών (false positives). Επιπρόσθετα, η τοπολογική, «γραφηματική» πληροφορία η οποία διοχετεύτηκε στο XGBoost, είχε μάλλον επικουρικό χαρακτήρα, καθώς δεν υπήρξαν ισχυρές ενδείξεις ότι προσέφερε σε μεγάλο βαθμό στην απόδοση του μοντέλου. Σε αυτό το σημείο, θα ήταν πολύ χρήσιμη μία αντιπαραβολική μελέτη αφαίρεσης (ablation study) όπου θα εξεταζόταν η απόδοση του XGBoost με είσοδο που να μην περιέχει τέτοιου είδους πληροφορία.

Συμπερασματικά με βάση τα παραπάνω, προκύπτει ότι η κάθε προσέγγιση και συμπεριφορά των μοντέλων που εξετάσαμε – συντηρητική για το HeteroGraphSAGE, «επιθετική» για το XGBoost – αντικατοπτρίζει έναν διαφορετικό συμβιβασμό (trade-off) μεταξύ Precision και Recall, ανάλογα με τον επιχειρησιακό στόχο. Με άλλα λόγια, η επιλογή του μοντέλου για μια εργασία εξαρτάται από το ποια είναι η προτεραιότητα μας – θέλουμε να επικεντρωθούμε στην ανίχνευση όσο το δυνατόν περισσότερων απατών, ή μας ενδιαφέρει πιο πολύ η ισορροπία και ο περιορισμός των false positives σε χαμηλά επίπεδα;

Τέλος, και σε ότι αφορά την ερμηνευσιμότητα των αποφάσεων των μοντέλων, έγινε μια προσπάθεια διερεύνησης του πώς εφαρμόζονται σύγχρονες μέθοδοι XAI, οι οποίες στοχεύουν στο να παρέχουν κατανοητές, οπτικοποιημένες και επιχειρηματικά αξιοποιήσιμες ενδείξεις. Γενικά, η ικανότητα να παρέχουμε ακριβείς και τοπικές εξηγήσεις έχει τεράστια αξία· ωστόσο, είναι πολύ σημαντικό να τονίσουμε το εξής: οι εξηγήσεις αυτές είναι μεν ενδεικτικές του μηχανισμού απόφασης του μοντέλου, δεν τεκμηριώνουν όμως από μόνες τους αιτιώδη σχέση ή οριστικό χαρακτηρισμό της συναλλαγής.

7.2 Απαντήσεις στα ερευνητικά ερωτήματα

Η αξιολόγηση του συστήματος απαντά άμεσα στα τέσσερα βασικά ερευνητικά ερωτήματα που τέθηκαν κατά τον σχεδιασμό της έρευνας.

7.2.1 Ερώτημα 1

Πώς μπορούν τα GNN να βελτιώσουν την ακρίβεια ανίχνευσης οικονομικής απάτης σε σύγκριση με τις παραδοσιακές μεθόδους;

Το HeteroGraphSAGE υπερτέρησε του παραδοσιακού μοντέλου αναφοράς XGBoost – ενώ το δεύτερο ήταν ενισχυμένο με επιπλέον πληροφορία (graph metrics και embeddings) – ως προς την Ακρίβεια (Precision) και το F1-Score. Ενώ το XGBoost εντόπισε περισσότερες απάτες (υψηλότερο Recall), το πέτυχε παράγοντας έναν δύσκολα διαχειρίσιμο όγκο ψευδώς θετικών συναγερμών. Το GNN εκμεταλλεύτηκε τη δομή της γειτονιάς κάθε κόμβου, ανεβάζοντας το Precision στο 0.7569 και το συνολικό F1-Score στο 0.6078 (αύξηση σχεδόν 68% έναντι του baseline), επιδεικνύοντας ανώτερη ικανότητα διάκρισης.

7.2.2 Ερώτημα 2

Ποιος είναι ο ρόλος των graph DBs στον εντοπισμό πολύπλοκων μοτίβων απάτης που εκτείνονται σε πολλαπλές οντότητες και συναλλαγές;

Η χρήση της Neo4j ήταν καθοριστική για τη μοντελοποίηση πολύπλοκων σχέσεων που θα ήταν υπολογιστικά ασύμφορες σε σχεσιακές βάσεις δεδομένων. Η δυνατότητα δημιουργίας έμμεσων ακμών (όπως η κοινή χρήση του ίδιου email από διαφορετικές κάρτες), συνέδεσε διαφορετικά τμήματα του γράφου δημιουργώντας «γέφυρες» μεταξύ φαινομενικά ασύνδετων συναλλαγών. Ο υπολογισμός όλων αυτών των καρτεσιανών γινομένων σε μια παραδοσιακή βάση SQL θα ήταν απαγορευτικός σε κόστος πόρων και χρόνο εκτέλεσης.

Παράλληλα, επέτρεψε την εφαρμογή «γραφοαναλυτικών» αλγορίθμων για την εξαγωγή πληροφορίας που προκύπτουν από μετρικές που αφορούν τους κόμβους και τη συνδεσιμότητα μεταξύ τους. Η χρήση της Neo4j, επομένως, δεν αποτελεί απλώς μια αρχιτεκτονική επιλογή αποθήκευσης, αλλά την απαραίτητη προϋπόθεση για τη δημιουργία της ίδιας της γνώσης και της πληροφορίας.

7.2.3 Ερώτημα 3

Πώς μπορούν οι τεχνικές AI/ML να μειώσουν τα ψευδώς θετικά αποτελέσματα διατηρώντας υψηλά ποσοστά ανίχνευσης;

Το πρόβλημα των ανισόρροπων δεδομένων (με την απάτη να αποτελεί μόλις το 3,5% στη συγκεκριμένη μελέτη περίπτωσης) αντιμετωπίστηκε επιτυχώς χωρίς την τεχνητή παραγωγή δεδομένων (όπως το SMOTE), αλλά μέσω της χρήσης δυναμικού κατωφλίου (dynamic thresholding) και βελτιστοποίησής του με βάση τη μετρική AUC-PR, με σκοπό την μεγιστοποίηση του F1-Score στο σύνολο επικύρωσης. Αυτό οδήγησε στον περιορισμό των λανθασμένων συναγερμών σε μόλις 337 επί συνόλου 57.000 περίπου νόμιμων συναλλαγών.

Αυτό το ποσοστό σφάλματος, σε ένα σύστημα μικρής εμβέλειας που διεξήχθη εντός ενός υποτυπώδους πειραματικού περιβάλλοντος, είναι πολλά υποσχόμενο και δείχνει ότι τα GNNs, όταν συνδυάζονται με κατάλληλες τεχνικές βελτιστοποίησης μετρικών ευαίσθητων στην ανισορροπία (όπως η AUC-PR εδώ), διαθέτουν την εγγενή ικανότητα να ξεχωρίζουν τον θόρυβο από το πραγματικό σήμα απάτης.

7.2.4 Ερώτημα 4

Ποιες μέθοδοι graph analytics είναι πιο αποτελεσματικές για τον εντοπισμό διαφορετικών τύπων οικονομικής απάτης;

Η εργασία ανέδειξε την προστιθέμενη αξία της αναπαράστασης των δεδομένων ως γράφου για την εξαγωγή επιχειρησιακών συμπερασμάτων. Μέσω της Neo4j GDS βιβλιοθήκης, εντοπίστηκαν υποψήφια fraud rings με βάση τις Louvain communities και τα υψηλά PageRank scores, αναδείχθηκαν «κόμβοι-μεσάζοντες» (υποψήφια mule accounts) με υψηλή betweenness centrality, και καταγράφηκαν κάρτες και email domains με ιδιαίτερα υψηλό βαθμό και αυξημένο fraud rate.

Από τα μέτρα κεντρικότητας, το degree centrality έδειξε να έχει τη μεγαλύτερη χρησιμότητα, συνηγορώντας στο ότι οι δράστες τείνουν να μοιράζονται και να ανακυκλώνουν υποδομές που παρουσιάζουν υψηλότερο μέσο βαθμό συνδεσιμότητας. Παράλληλα, ο αλγόριθμος ανίχνευσης κοινοτήτων Louvain εντόπισε συστάδες με αξιοπρόσεκτα ποσοστά απάτης (έως και 9,3%), στοιχείο που μπορεί – με περεταίρω ανάλυση – να οδηγήσει στον εντοπισμό "fraud rings". Τέλος, η συνεισφορά του αλγορίθμου WCC περιορίστηκε κυρίως στην εμπέδωση της πυκνής συνδεσιμότητας του γράφου.

7.3 Κριτική αξιολόγηση και περιορισμοί

Η παρούσα εργασία υπόκειται σε ορισμένους περιορισμούς, οι οποίοι συνδέονται κυρίως με τη φύση των δεδομένων και τους διαθέσιμους υπολογιστικούς πόρους.

7.3.1 Δεδομένα

Πρώτον, τόσο το σύνολο δεδομένων που χρησιμοποιήθηκε (IEEE-CIS), όσο και αυτό που προτείνεται ως επέκταση (Elliptic), είναι δημόσια διαθέσιμα και σε μεγάλο βαθμό «ανωνυμοποιημένα», με αποτέλεσμα να μην αποτυπώνουν πλήρως τις ιδιαιτερότητες, την πολυπλοκότητα και τις πρακτικές διαδικασίες ενός πραγματικού χρηματοπιστωτικού ιδρύματος. Η απουσία πραγματικών παραγωγικών (production-grade) δεδομένων περιορίζει την εξαγωγή συμπερασμάτων ως προς την απόδοση σε «ζωντανό» (live) περιβάλλον. Επιπλέον, αυτή η «ανωνυμοποίηση» έχει και πρακτικές συνέπειες: ένα μεγάλο

μέρος από τα πρωτογενή χαρακτηριστικά αντιπροσωπεύουν συγκαλυμμένες εσωτερικές μετρικές, των οποίων η πραγματική σημασιολογία παρέμεινε άγνωστη. Αυτό δυσχέρανε την πλήρη ερμηνεία των αποτελεσμάτων κατά τη διαδικασία της καθολικής ανάλυσης SHAP, αναγκάζοντας το μοντέλο να εξάγει συμπεράσματα βασισμένο σε αφηρημένες μαθηματικές κατανομές, και όχι σε φυσικά και άμεσα κατανοητά χαρακτηριστικά.

Δεύτερον, το IEEE-CIS dataset περιορίζεται αυστηρά σε απάτες που αφορούν συναλλαγές ηλεκτρονικού εμπορίου χωρίς φυσική παρουσία της κάρτας (Card-Not-Present e-commerce fraud). Αυτό σημαίνει ότι τα συμπεράσματα που εξήχθησαν σχετικά με την τοπολογία της απάτης, αφορούν αποκλειστικά τη συμπεριφορά των δόλιων υποκειμένων σε αυτού του είδους την απάτη. Η αρχιτεκτονική που αναπτύχθηκε δεν μπορεί να εγγυηθεί, χωρίς την κατάλληλη παραμετροποίηση, την ίδια επιτυχία απέναντι σε άλλες κατηγορίες απάτης (π.χ. ξέπλυμα χρημάτων – AML) που έχουν τα δικά τους χαρακτηριστικά γνωρίσματα και μεθόδους δράσης.

7.3.2 Υπολογιστικοί Πόροι

Λόγω του μεγάλου υπολογιστικού κόστους που απαιτεί η εκπαίδευση μοντέλων βαθιάς μάθησης σε γράφους κλίμακας εκατομμυρίων ακμών, ήταν αναγκαίο να γίνουν κάποιοι συμβιβασμοί. Το γεγονός ότι η υλοποίηση του pipeline πραγματοποιήθηκε σε περιβάλλον με πολύ περιορισμένους πόρους υλικού (hardware)[10], επέβαλε συγκεκριμένες επιλογές στην αρχιτεκτονική του συστήματος, όπως mini-batch training με neighbor sampling, sampling-based betweenness centrality, περιορισμό του μεγέθους του in-memory γράφου της GDS, δειγματοληπτική εφαρμογή του GNNExplainer.

Αυτές οι επιλογές είναι ρεαλιστικές για ένα proof of concept[11], αλλά βέβαια σε επίπεδο prototype έχουν ένα τίμημα όσον αφορά τις δυνατότητες των μοντέλων. Για παράδειγμα, η υλοποίηση του μηχανισμού NeighborLoader με τόσο αυστηρές παραμέτρους δειγματοληψίας, στερεί από το GNN μια πολύ ευρύτερη εικόνα του γράφου, την οποία ενδεχομένως να είχε ανάγκη ώστε να είναι σε θέση να ανακαλύψει κυκλώματα απάτης που

[10] Συγκεκριμένα, η εκτέλεση των notebooks έγινε σε φορητό υπολογιστή με 16GB RAM και ενσωματωμένα Intel γραφικά – δηλαδή η εκπαίδευση των μοντέλων έγινε σε CPU mode. Η ύπαρξη discrete GPU (ειδικότερα της Nvidia, ώστε να γίνει χρήση του CUDA framework) θα βελτίωνε θεαματικά τις δυνατότητες και θα οδηγούσε σε διαφορετικές σχεδιαστικές επιλογές.

[11] https://www.geeksforgeeks.org/software-engineering/proof-of-concept-vs-prototype-vs-minimum-viable-product/

μπορεί να εκτείνονται σε περισσότερες από δύο (2) μεταβάσεις μεταξύ κόμβων (hops) – και που στην παρούσα έκδοση του συστήματος παραμένουν αόρατα.

Επιπρόσθετα, η κατασκευή του γράφου βασίστηκε σε συγκεκριμένες αυθαίρετες – αν και τεκμηριωμένες – επιλογές σχετικά με τους τύπους κόμβων και ακμών (Transactions, Card, EmailDomain, Device, USED_* και SHAREDEMAILWITH), οι οποίες δεν εξαντλούν τον χώρο των πιθανών μοντελοποιήσεων. Η μη ενσωμάτωση επιπλέον οντοτήτων που θα μπορούσαν να εξαχθούν από τις διαθέσιμες στήλες του dataset, όπως merchants, γεωγραφικές τοποθεσίες, IP διευθύνσεις κ.α., σημαίνει ότι ο γράφος εστιάζει σε ένα υποσύνολο των ενδεχόμενων διαστάσεων κινδύνου. Ο λόγος για τα παραπάνω ήταν και πάλι οι απαιτήσεις υλικού.

Τέλος, οι υπερπαράμετροι (hyperparameters) των αλγορίθμων της GDS (π.χ. damping factor και αριθμός επαναλήψεων για τον αλγόριθμο PageRank, resolution στον αλγόριθμο Louvain, διάσταση και παράμετροι του FastRP) επιλέχθηκαν από τη μία εμπειρικά, χωρίς να πραγματοποιηθεί πλήρης βελτιστοποίηση ή συστηματική μελέτη της ευαισθησίας τους, και από την άλλη με γνώμονα το υπολογιστικό κόστος.

7.3.3 Χρονική και δυναμική διάσταση

Στην παρούσα υλοποίηση, παρότι η χρονική μεταβλητή (TransactionDT) υφίστατο στο dataset και αποδομήθηκε σε στατικά χαρακτηριστικά (όπως η ώρα της ημέρας ή η εβδομάδα) κατά τη φάση του Feature Engineering στο Notebook 1, η τελική αρχιτεκτονική του γράφου στη βάση Neo4j και το ίδιο το μοντέλο HeteroGraphSAGE αντιμετώπισαν το δίκτυο ως μια σταθερή, αμετάβλητη δομή, ενώ επίσης ο αλγόριθμος FastRP παρήγαγε στατικά embeddings, χωρίς να λαμβάνει υπόψη τη χρονική εξέλιξη του κόμβου. Εν ολίγοις, το σύστημα δεν διαθέτει εγγενή μηχανισμό για να αντιληφθεί τη σειρά (sequence) και την ακριβή χρονική απόσταση με την οποία δημιουργήθηκαν οι ακμές (συναλλαγές). Η απουσία αυτής της χρονικής διάστασης στερεί από το μοντέλο τη δυνατότητα να κατανοήσει τα πρότυπα συμπεριφοράς που βασίζονται στην ταχύτητα (με την οποία γίνονται οι συναλλαγές), κάτι το οποίο είναι κρίσιμο για την ανίχνευση απάτης σε πραγματικό χρόνο. (Alessi & Fugini, 2026; Valero, 2025; Zakaria κ.ά., 2025)

Επιπλέον, η παρούσα προσέγγιση θεωρεί τα δεδομένα ως σταθερή κατανομή και δεν αντιμετωπίζει ρητά το φαινόμενο του concept drift, δηλαδή την εξέλιξη στο χρόνο της

σχέσης μεταξύ εισόδων και στόχου, που μπορεί να καταστήσει σταδιακά ένα εκπαιδευμένο μοντέλο λιγότερο αντιπροσωπευτικό και αποτελεσματικό. Στο πλαίσιο της ανίχνευσης οικονομικής απάτης, οι τακτικές των απατεώνων και οι κανονιστικές απαιτήσεις αλλάζουν διαρκώς, με αποτέλεσμα οι στατικοί, offline εκπαιδευμένοι ταξινομητές (classifiers) – όπως το μοντέλο που αναπτύξαμε – να υποβαθμίζονται σε απόδοση αν δεν συνοδεύονται από μηχανισμούς ανίχνευσης και προσαρμογής σε επικαιροποιημένα δεδομένα. (Shahapurkar & Patil, 2023; Valero, 2025; Yelleti, 2025)

7.4 Μελλοντικές κατευθύνσεις

Οι προαναφερθέντες περιορισμοί δεν συνιστούν αδιέξοδα της τεχνολογίας, αλλά φυσιολογικά στάδια εξέλιξης σε ένα ταχέως αναπτυσσόμενο ερευνητικό πεδίο. Η πρακτική εφαρμογή και η ακαδημαϊκή ωρίμανση της παρούσας μεθοδολογίας ανοίγουν τον δρόμο για στοχευμένες επεκτάσεις στο μέλλον, οι οποίες μπορούν να αντιμετωπίσουν τα κενά και να διευρύνουν τις δυνατότητες του συστήματος ανίχνευσης. Οι μελλοντικές προσπάθειες προτείνεται να κινηθούν κατά μήκος τριών κεντρικών αξόνων.

7.4.1 Εμπλουτισμός του Γράφου και Επέκταση σε Άλλες Κατηγορίες Απάτης

Ο πρώτος προφανής άξονας αφορά τόσο την κάθετη εμβάθυνση της παρούσας μοντελοποίησης, όσο και την οριζόντια επέκτασή της σε νέες κατηγορίες οικονομικής παρανομίας. Ως προς την εμβάθυνση, η ενσωμάτωση επιπλέον τύπων οντοτήτων (όπως merchants, γεωγραφικές τοποθεσίες και IP διευθύνσεις) θα εμπλούτιζε τον χώρο των διαστάσεων κινδύνου που καλύπτει ο γράφος, δίνοντας την δυνατότητα να εξετάσουμε την αποδοτικότητα των μοντέλων σε αυτήν την ενισχυμένη δομή.

Ως προς την οριζόντια επέκταση, η εφαρμογή της ίδιας μεθοδολογικής προσέγγισης στο Elliptic dataset – που αφορά το οικοσύστημα των κρυπτονομισμάτων – ή η παραμετροποίησή της για σενάρια AML, αποτελούν ιδιαίτερα ελκυστικές κατευθύνσεις, καθώς αυτές οι κατηγορίες απάτης παρουσιάζουν εγγενώς «γραφοκεντρικά» μοτίβα συναλλαγών. (Oztas κ.ά., 2024)

7.4.2 Δυναμική Μοντελοποίηση Γράφου και Χρονική Εξέλιξη

Ο δεύτερος άξονας αφορά τη μετάβαση από έναν στατικό, αμετάβλητο γράφο σε ένα δυναμικό, χρονικά εξελισσόμενο μοντέλο. Η ενσωμάτωση αρχιτεκτονικών Temporal Graph Networks (TGN) ή Temporal Graph Attention Networks (TGAT), οι οποίες μοντελοποιούν ρητά τη σειρά και τη χρονική απόσταση μεταξύ ακμών, θα επέτρεπε στο σύστημα να ανιχνεύει μοτίβα συμπεριφοράς που βασίζονται στην ταχύτητα εκτέλεσης των συναλλαγών (βλ. υποενότητα 7.3.3).

Αυτή η επέκταση είναι από μόνη της απαιτητική, καθώς οι παραπάνω αρχιτεκτονικές είναι πιο σύνθετες από την υφιστάμενη – εάν δε, προσθέσουμε και την εφαρμογή μηχανισμών ανίχνευσης και προσαρμογής σε concept drift (για παράδειγμα μέσω online ή incremental learning), η πολυπλοκότητα αυξάνεται σημαντικά.

7.4.3 Βελτιστοποίηση Αρχιτεκτονικής και Επεκτασιμότητα

Ο τρίτος άξονας εστιάζει στην τεχνική ενίσχυση του ίδιου του συστήματος. Μια ισχυρότερη υπολογιστική υποδομή, θα επέτρεπε την εκπαίδευση με βαθύτερες αρχιτεκτονικές GNN (περισσότερα hops) και πλουσιότερα neighbor sampling σχήματα, εξασφαλίζοντας στο μοντέλο ορατότητα σε κυκλώματα απάτης που εκτείνονται πέρα από τις δύο μεταβάσεις κόμβων. Επιπρόσθετα, η διενέργεια συστηματικής βελτιστοποίησης υπερπαραμέτρων (hyperparameter tuning) για τους αλγορίθμους της GDS βιβλιοθήκης, καθώς και η επιλογή της βέλτιστης αρχιτεκτονικής – ενδεχομένως με την εφαρμογή κάποιου πλαισίου AutoML[12] – θα αντιμετώπιζαν τις εμπειρικές και συχνά αυθαίρετες επιλογές που επιβλήθηκαν από τους υπολογιστικούς περιορισμούς της παρούσας υλοποίησης.

Κάπως πιο φιλόδοξα, η ανάπτυξη ενός πλήρως αυτοματοποιημένου inference pipeline – ικανού να λειτουργεί σε streaming mode με την ενσωμάτωση πλατφορμών όπως το Apache Kafka – θα αποτελούσε ένα μεγάλο βήμα για τη μετάβαση από ένα proof of concept σε ένα σύστημα παραγωγικής κλίμακας. (Chhaparia κ.ά., 2025)

Τέλος, θα είχε ιδιαίτερο ενδιαφέρον η μετατροπή του συστήματος σε ένα ενιαίο, συνδυαστικό pipeline, το οποίο θα κάνει χρήση και των δύο υφιστάμενων μοντέλων (π.χ. μέσω ensemble ή two-stage inference), προκειμένου να καταλήξει σε μία τελική απόφαση για κάθε παράδειγμα. Το σκεπτικό σε αυτό είναι η αξιοποίηση των επιμέρους

[12] https://en.wikipedia.org/wiki/Automated_machine_learning

χαρακτηριστικών και πλεονεκτημάτων των δύο μοντέλων, με σκοπό την αναζήτηση της χρυσής τομής ανάμεσα στην προσπάθεια βελτιστοποίησης αφενός του Precision, και αφετέρου του Recall.

Βιβλιογραφία

Σπανάκη, Κ., Ζοπουνίδης, Κ., & Λεμονάκης, Χ. (2024, Φεβρουάριος 3). *Επεξηγηματική Τεχνητή Νοημοσύνη (ETN) – Explainable Artificial Intelligence (XAI) στη διαχείριση Εφοδιαστικής Αλυσίδας και Πιστωτικού Κινδύνου*. Οικονομικός Ταχυδρόμος - ot.gr. https://www.ot.gr/2024/02/03/apopseis/epeksigimatiki-texniti-noimosyni-etn-explainable-artificial-intelligence-xai-sti-diaxeirisi-efodiastikis-alysidas-kai-pistotikou-kindynou/

Alessi, G., & Fugini, M. (2026). Adaptive Real-Time Financial Fraud Detection with Explainable AI Tools. *Digital Threats*, *7*(1), 7:1-7:31. https://doi.org/10.1145/3794859

Arnau, R., Calabuig, J. M., García-Raffi, L. M., Sánchez Pérez, E. A., & Sanjuan, S. (2024). A Bellman–Ford Algorithm for the Path-Length-Weighted Distance in Graphs. *Mathematics*, *12*(16), 2590. https://doi.org/10.3390/math12162590

Asiri, A., & Somasundaram, K. (2025). Graph convolution network for fraud detection in bitcoin transactions. *Scientific Reports*, *15*(1), 11076. https://doi.org/10.1038/s41598-025-95672-w

Bin Sulaiman, R., Schetinin, V., & Sant, P. (2022). Review of Machine Learning Approach on Credit Card Fraud Detection. *Human-Centric Intelligent Systems*, *2*(1–2), 55–68. https://doi.org/10.1007/s44230-022-00004-0

Bolton, R. J., & Hand, D. J. (2002). Statistical Fraud Detection: A Review. *Statistical Science*, *17*(3). https://doi.org/10.1214/ss/1042727940

Chawla, N. V., Bowyer, K. W., Hall, L. O., & Kegelmeyer, W. P. (2002). SMOTE: Synthetic Minority Over-sampling Technique. *Journal of Artificial Intelligence Research*, *16*, 321–357. https://doi.org/10.1613/jair.953

Chen, Y., Zhao, C., Xu, Y., Nie, C., & Zhang, Y. (2025). *Year-over-Year Developments in Financial Fraud Detection via Deep Learning: A Systematic Literature Review* (arXiv:2502.00201). arXiv. https://doi.org/10.48550/arXiv.2502.00201

Cheng, D., Zou, Y., Xiang, S., & Jiang, C. (2025). Graph Neural Networks for Financial Fraud Detection: A Review. *Frontiers of Computer Science*, *19*(9), 199609. https://doi.org/10.1007/s11704-024-40474-y

Chhaparia, N., Ajay, R., Makhija, K., Rathor, K., & Vankayalapati, H. D. (2025). Real Time UPI Fraud Detection Using GNNs. *2025 IEEE Pune Section International Conference (PuneCon)*, 1–6. https://doi.org/10.1109/PuneCon67554.2025.11379772

DataWalk. (2022, Απρίλιος 8). Graph Analytics: The New Game-Changer For Anti-Fraud. *DataWalk*. https://datawalk.com/whitepaper-graph-analytics-the-new-game-changer-for-anti-fraud/

Haibo He, Yang Bai, Garcia, E. A., & Shutao Li. (2008). ADASYN: Adaptive synthetic sampling approach for imbalanced learning. *2008 IEEE International Joint Conference on Neural Networks (IEEE World Congress on Computational Intelligence)*, 1322–1328. https://doi.org/10.1109/IJCNN.2008.4633969

Hamilton, W., Ying, Z., & Leskovec, J. (2017). Inductive Representation Learning on Large Graphs. *Advances in Neural Information Processing Systems*, *30*. https://proceedings.neurips.cc/paper_files/paper/2017/hash/5dd9db5e033da9c6fb5ba83c7a7ebea9-Abstract.html

Hilal, W., Gadsden, S. A., & Yawney, J. (2022). Financial Fraud: A Review of Anomaly Detection Techniques and Recent Advances. *Expert Systems with Applications*, *193*, 116429. https://doi.org/10.1016/j.eswa.2021.116429

Hodler, A. (2021). *Financial Fraud Detection with Graph Data Science*.

Huang, C.-Y. (Ric), Lai, C.-Y., & Cheng, K.-T. (Tim). (2009). Fundamentals of algorithms. Στο *Electronic Design Automation* (σελ. 173–234). Elsevier. https://doi.org/10.1016/B978-0-12-374364-0.50011-4

Ji, S., Pan, S., Cambria, E., Marttinen, P., & Yu, P. S. (2022). A Survey on Knowledge Graphs: Representation, Acquisition and Applications. *IEEE Transactions on Neural Networks and Learning Systems*, *33*(2), 494–514. https://doi.org/10.1109/TNNLS.2021.3070843

Jiang, J., Zhang, C., Ke, L., Hayes, N., Zhu, Y., Qiu, H., Zhang, B., Zhou, T., & Wei, G.-W. (2025). A review of machine learning methods for imbalanced data challenges in chemistry. *Chemical Science*, *16*(18), 7637–7658. https://doi.org/10.1039/D5SC00270B

Leidig, P. (2025, Σεπτέμβριος 18). Why Graph Centrality Measures Detect Fraud Faster. *TigerGraph*. https://www.tigergraph.com/blog/why-graph-centrality-measures-detect-fraud-faster/

Liu, H., Wang, Y., Fan, W., Liu, X., Li, Y., Jain, S., Liu, Y., Jain, A. K., & Tang, J. (2021). *Trustworthy AI: A Computational Perspective* (arXiv:2107.06641). arXiv. https://doi.org/10.48550/arXiv.2107.06641

Lundberg, S. M., & Lee, S.-I. (2017). A Unified Approach to Interpreting Model Predictions. *Advances in Neural Information Processing Systems*, *30*.

https://proceedings.neurips.cc/paper/2017/hash/8a20a8621978632d76c43dfd28b67767-Abstract.html

Madduru, P. R., & Janvekar, N. (2023). *A Heterogeneous Graph-based Framework for Scalable Fraud Detection*.

Majumder, S., Singh, M. T., Prasad, R. K., Boruah, A., Michael, G. R., Kaphungkui, N. K., & Singh, N. H. (2025). Heterogeneous Graph Auto-Encoder for Credit Card Fraud Detection. *IJCA*, *32*(2).

Mao, X., Sun, H., Zhu, X., & Li, J. (2022). Financial fraud detection using the related-party transaction knowledge graph. *Procedia Computer Science*, *199*, 733–740. https://doi.org/10.1016/j.procs.2022.01.091

Modepalli, U. P. (2025). Network Analytics for Identifying Fraud Rings and Systemic Risk. *Journal of Information Systems Engineering and Management*, *10*(58s), 616–624. https://doi.org/10.52783/jisem.v10i58s.12641

Mohammed, A., & Kora, R. (2023). A comprehensive review on ensemble deep learning: Opportunities and challenges. *Journal of King Saud University - Computer and Information Sciences*, *35*(2), 757–774. https://doi.org/10.1016/j.jksuci.2023.01.014

Molloy, I., Chari, S., Finkler, U., Wiggerman, M., Jonker, C., Habeck, T., Park, Y., Jordens, F., & van Schaik, R. (2017). Graph Analytics for Real-Time Scoring of Cross-Channel Transactional Fraud. Στο J. Grossklags & B. Preneel (Επιμ.), *Financial Cryptography and Data Security* (σελ. 22–40). Springer. https://doi.org/10.1007/978-3-662-54970-4_2

Motie, S., & Raahemi, B. (2024). Financial fraud detection using graph neural networks: A systematic review. *Expert Systems with Applications*, *240*, 122156. https://doi.org/10.1016/j.eswa.2023.122156

Naim, Rees, B., & Liu, S. (2025, Ιούνιος 3). *Supercharging Fraud Detection in Financial Services with Graph Neural Networks (Updated)*. NVIDIA Technical Blog. https://developer.nvidia.com/blog/supercharging-fraud-detection-in-financial-services-with-graph-neural-networks/

Ounacer, S., Ait El Bour, H., Oubrahim, Y., Ghoumari, M. Y., & Azzouazi, M. (2018). Using isolation forest in anomaly detection: The case of credit card transactions. *Periodicals of Engineering and Natural Sciences (PEN)*, *6*(2), 394. https://doi.org/10.21533/pen.v6i2.533

Oztas, B., Cetinkaya, D., Adedoyin, F., Budka, M., Aksu, G., & Dogan, H. (2024). Transaction monitoring in anti-money laundering: A qualitative analysis and points of view from industry. *Future Generation Computer Systems*, *159*, 161–171. https://doi.org/10.1016/j.future.2024.05.027

Patil, A., Mahajan, S., Menpara, J., Wagle, S., Pareek, P., & Kotecha, K. (2024). Enhancing fraud detection in banking by integration of graph databases with machine learning. *MethodsX*, *12*, 102683. https://doi.org/10.1016/j.mex.2024.102683

Peng, C., Xia, F., Naseriparsa, M., & Osborne, F. (2023). Knowledge Graphs: Opportunities and Challenges. *Artificial Intelligence Review*, *56*(11), 13071–13102. https://doi.org/10.1007/s10462-023-10465-9

Ribeiro, M. T., Singh, S., & Guestrin, C. (2016). 'Why Should I Trust You?': Explaining the Predictions of Any Classifier. *Proceedings of the 22nd ACM SIGKDD International*

Conference on Knowledge Discovery and Data Mining, 1135–1144. https://doi.org/10.1145/2939672.2939778

Shahapurkar, A., & Patil, R. (2023). Concept drift and machine learning model for detecting fraudulent transactions in streaming environment. *International Journal of Electrical and Computer Engineering (IJECE)*, *13*(5), 5560. https://doi.org/10.11591/ijece.v13i5.pp5560-5568

Sherwin Akshay, J. G., Vinusha, T., Sharon Bianca, R., Sarath Krishna, C. K., & Radhika, G. (2024). Enhancing Credit Card Fraud Detection with Deep Learning and Graph Neural Networks. *2024 15th International Conference on Computing Communication and Networking Technologies (ICCCNT)*, 1–6. https://doi.org/10.1109/ICCCNT61001.2024.10725042

Tanner, M. (2025, Νοέμβριος 28). *TigerGraph vs Neo4j: How to Choose for Your Workload*. https://www.puppygraph.com/blog/tigergraph-vs-neo4j

Using Graph Machine Learning to Improve Fraud Detection Rates. (2023, Αύγουστος 30). *TigerGraph*. https://www.tigergraph.com/blog/using-graph-machine-learning-to-improve-fraud-detection-rates/

Usman, A., Naveed, N., & Munawar, S. (2023). Intelligent Anti-Money Laundering Fraud Control Using Graph-Based Machine Learning Model for the Financial Domain: *Journal of Cases on Information Technology*, *25*(1), 1–20. https://doi.org/10.4018/JCIT.316665

Valero, C. S. (2025, Απρίλιος 3). *From Classical ML to DNNs and GNNs for Real-Time Financial Fraud Detection*. César Soto Valero.

https://www.cesarsotovalero.net/blog/from-classical-ml-to-dnns-and-gnns-for-real-time-financial-fraud-detection.html

Wang, J., Zhang, S., Xiao, Y., & Song, R. (2022). *A Review on Graph Neural Network Methods in Financial Applications* (arXiv:2111.15367). arXiv. https://doi.org/10.48550/arXiv.2111.15367

Wen, S., Li, J., Zhu, X., & Liu, M. (2022). Analysis of financial fraud based on manager knowledge graph. *Procedia Computer Science*, *199*, 773–779. https://doi.org/10.1016/j.procs.2022.01.096

Yelleti, V. (2025). *ROSFD: Robust Online Streaming Fraud Detection with Resilience to Concept Drift in Data Streams* (arXiv:2504.10229). arXiv. https://doi.org/10.48550/arXiv.2504.10229

Zakaria, R. M., Rahman, M. M., Choudhury, M. T. H., Rahman, H., Rafi, M. A., Minto, A., Hossain, M. S., & Saimon, S. I. (2025). Detecting Financial Fraud in Real-Time Transactions Using Graph Neural Networks and Anomaly Detection Techniques. *Journal of Economics, Finance and Accounting Studies*, *7*(6), 01–13. https://doi.org/10.32996/jefas.2025.7.6.1

Zhang, J., & Luo, Y. (2017). Degree Centrality, Betweenness Centrality, and Closeness Centrality in Social Network. *Proceedings of the 2017 2nd International Conference on Modelling, Simulation and Applied Mathematics (MSAM2017)*. 2017 2nd International Conference on Modelling, Simulation and Applied Mathematics (MSAM2017). https://doi.org/10.2991/msam-17.2017.68

Zhang, X. (2025). Automobile Finance Credit Fraud Risk Early Warning System based on Louvain Algorithm and XGBoost Model. *2025 3rd International Conference on*

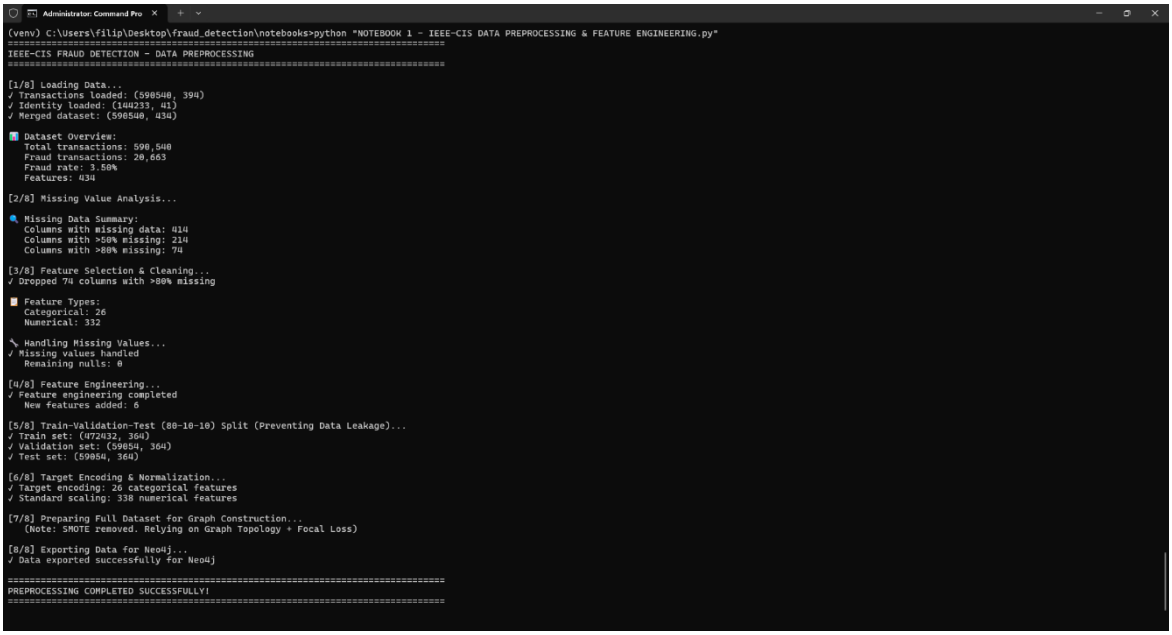
Data Science and Information System (ICDSIS), 1–7.

https://doi.org/10.1109/ICDSIS65355.2025.11070425

Παράρτημα Α: Στιγμιότυπα Οθόνης (Screenshots)

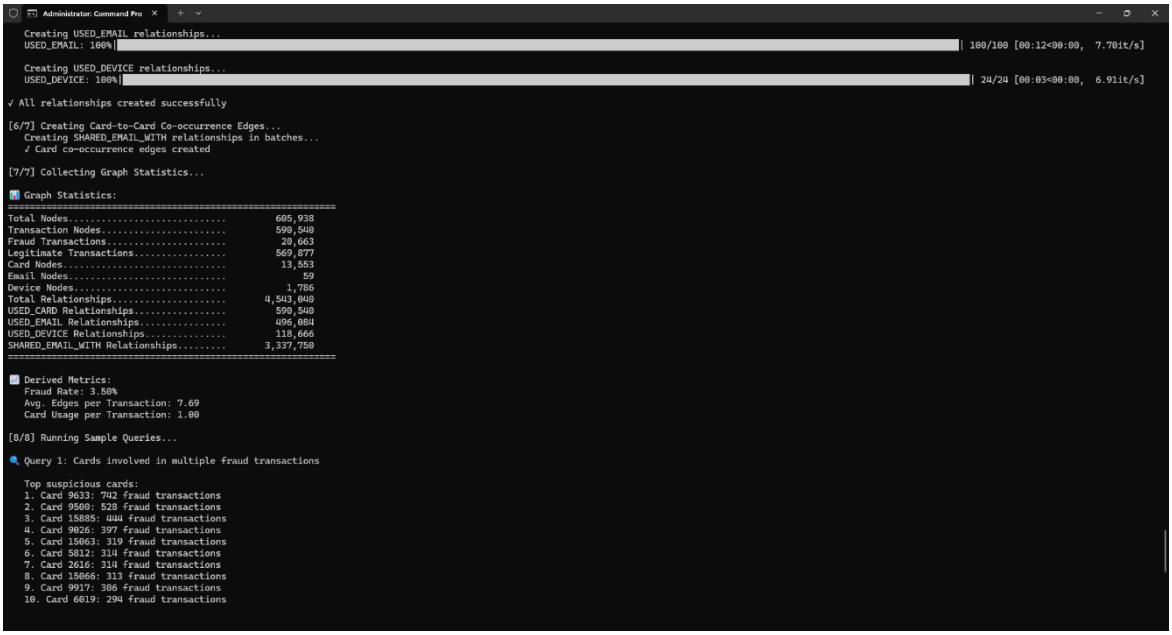
i. Εκτέλεση Pipeline – Έξοδος τερματικού

Εδώ παρουσιάζονται – κατά σειρά εκτέλεσης – τα αποτελέσματα (παραγόμενη έξοδος) του κάθε Notebook, όπως αυτά εμφανίστηκαν στο τερματικό (Windows Terminal).

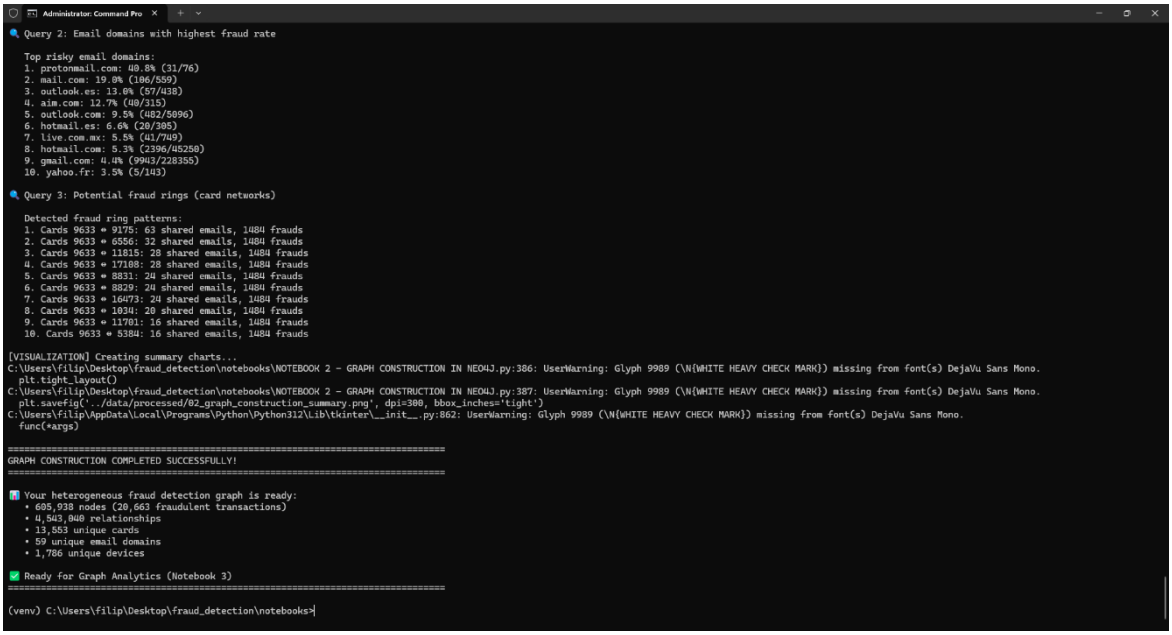


1. Notebook 1

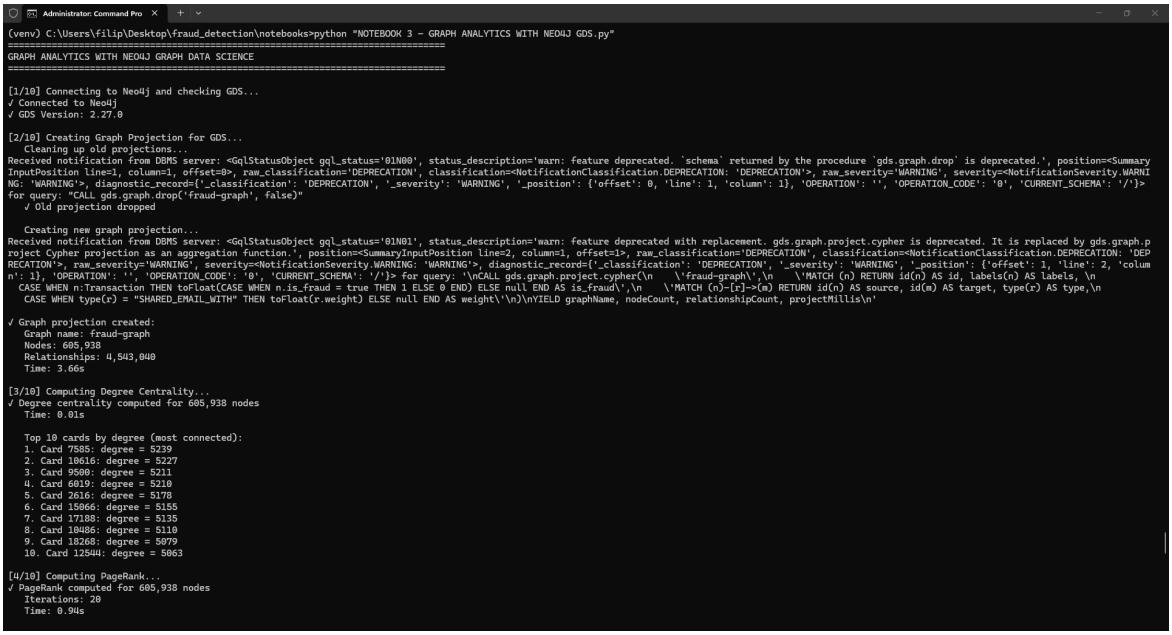
2. Notebook 2 - a



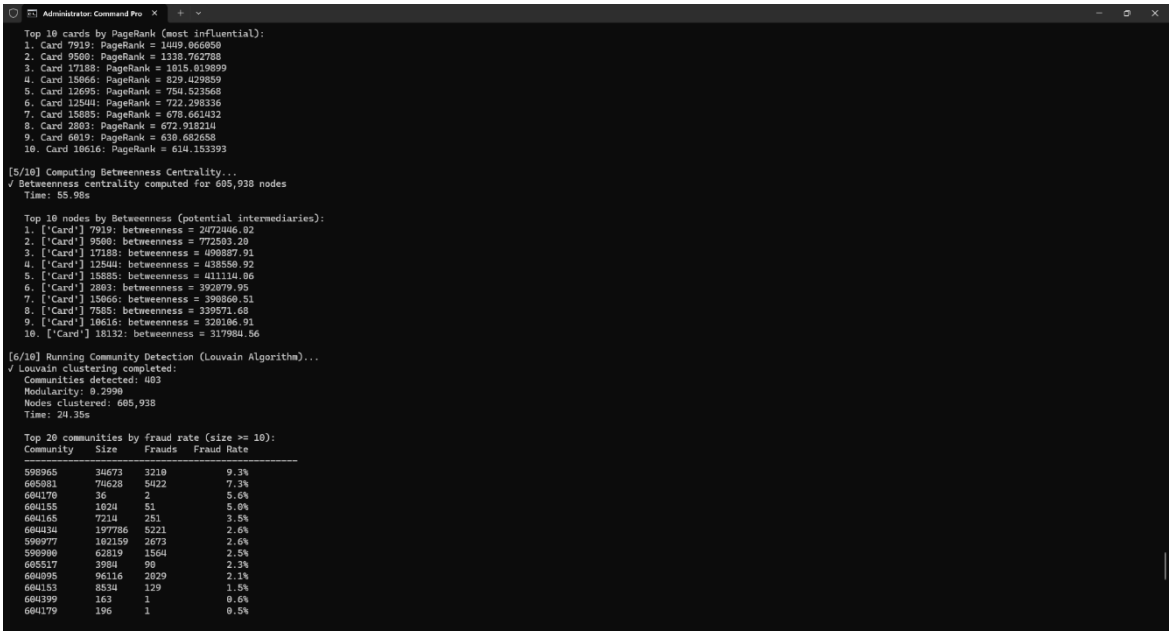
3. Notebook 2 - b



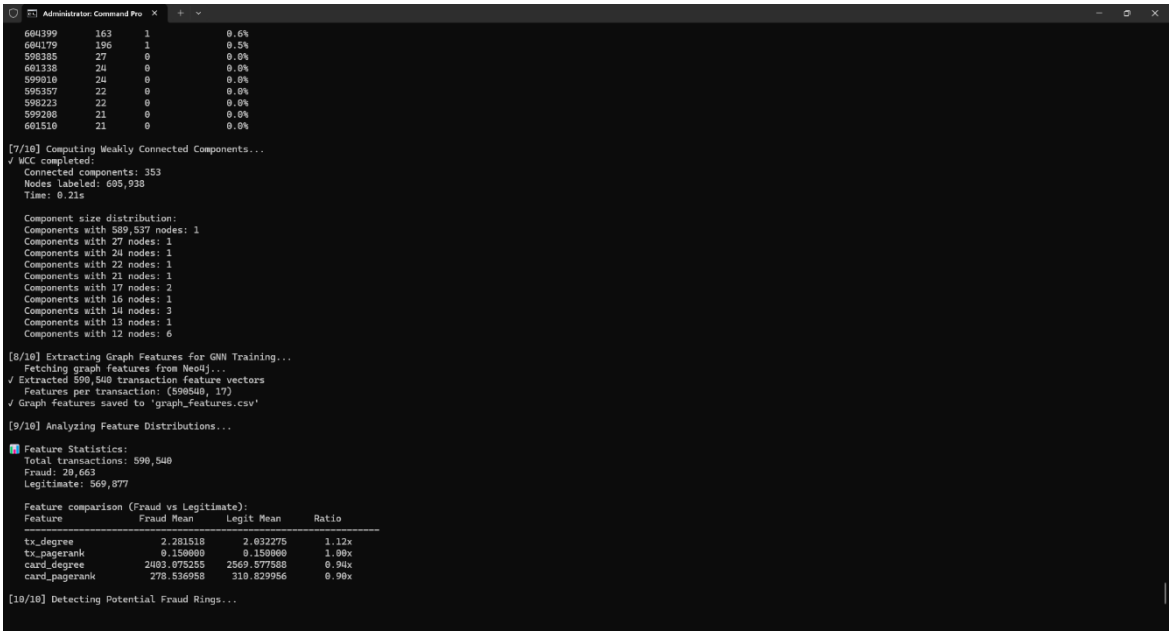
4. Notebook 2 - c



5. Notebook 3 - a

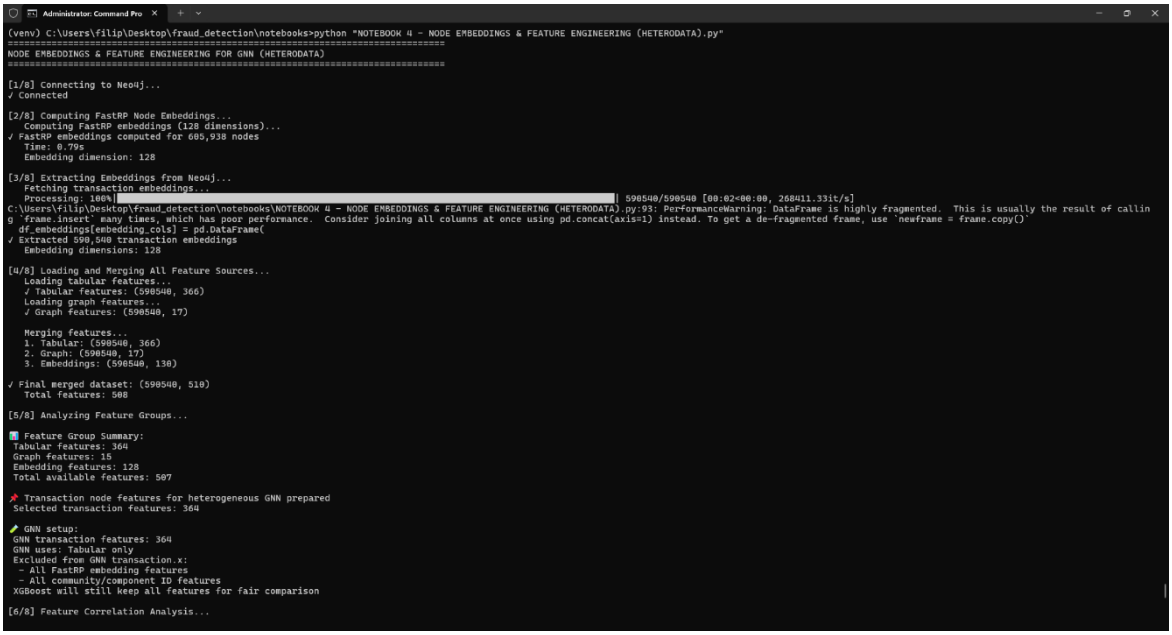


6. Notebook 3 - b

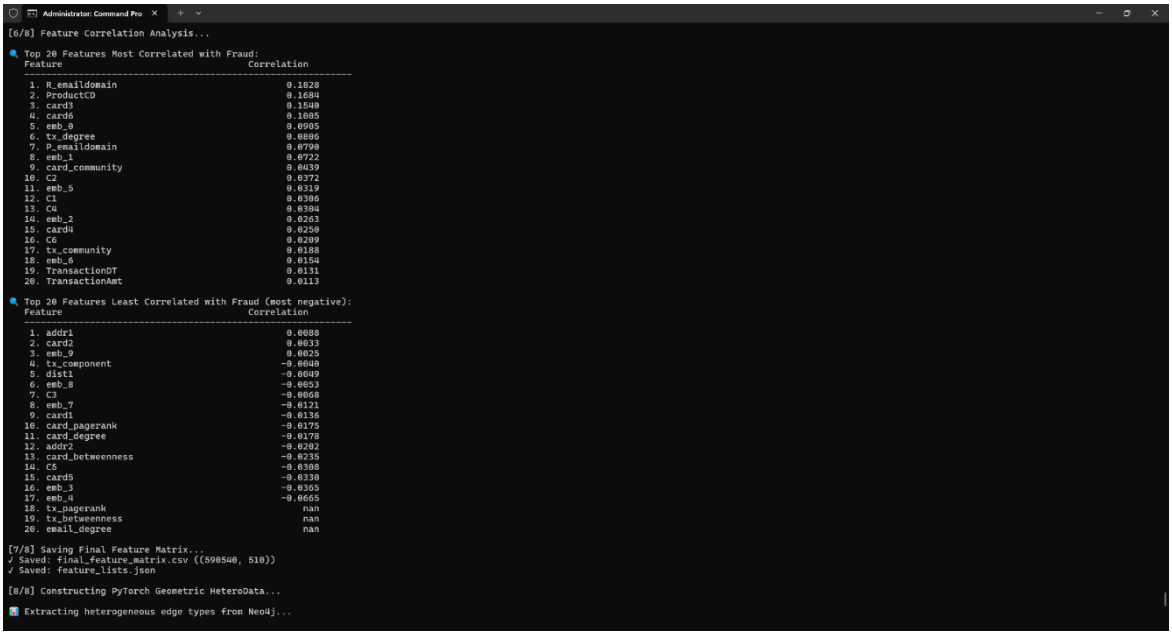


7. Notebook 3 - c

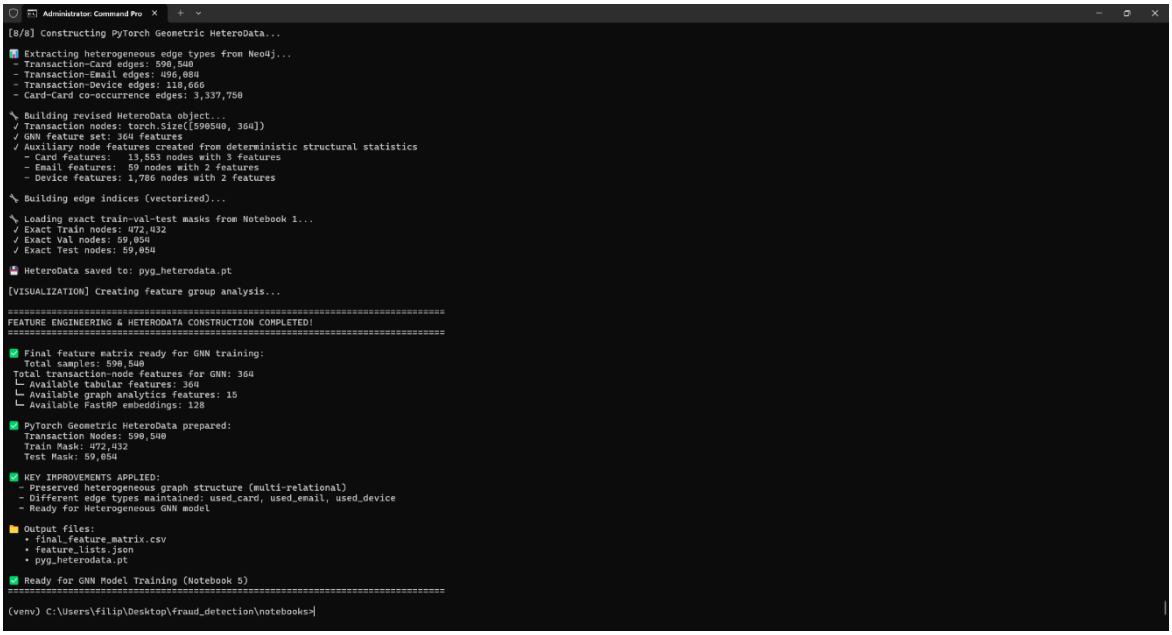
8. Notebook 3 - d



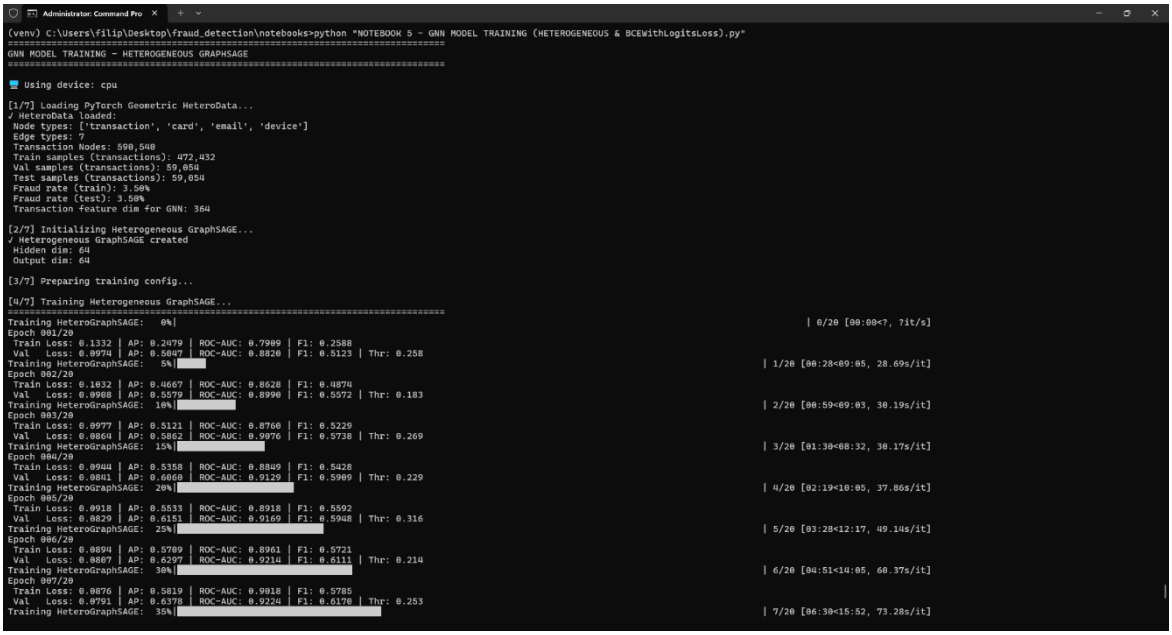
9. Notebook 4 - a



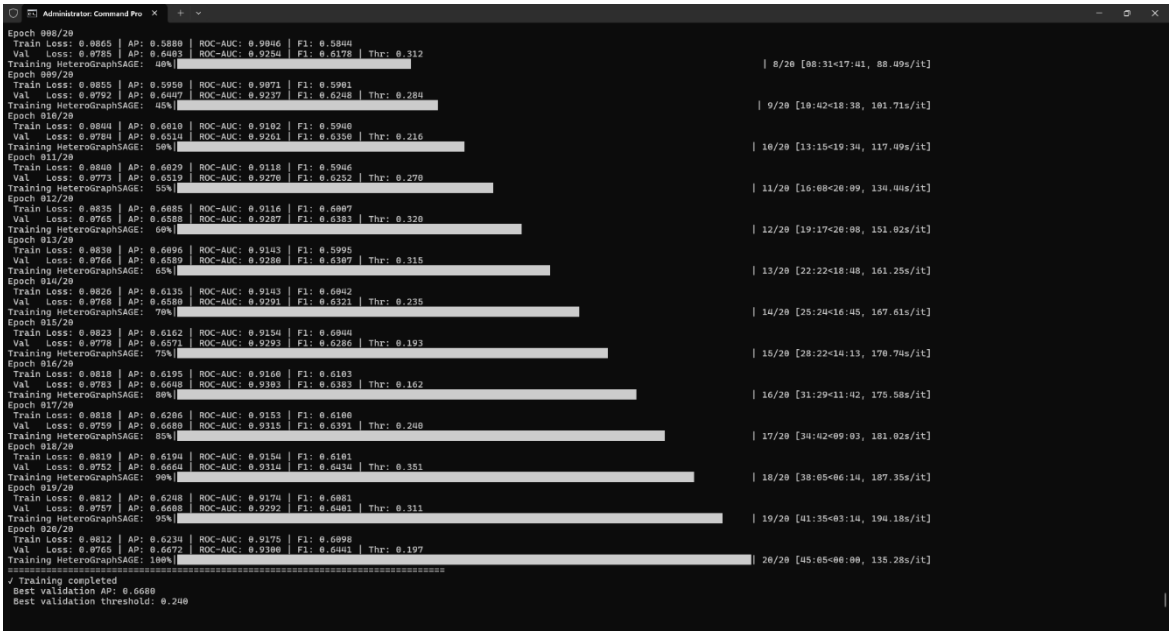
10. Notebook 4 - b



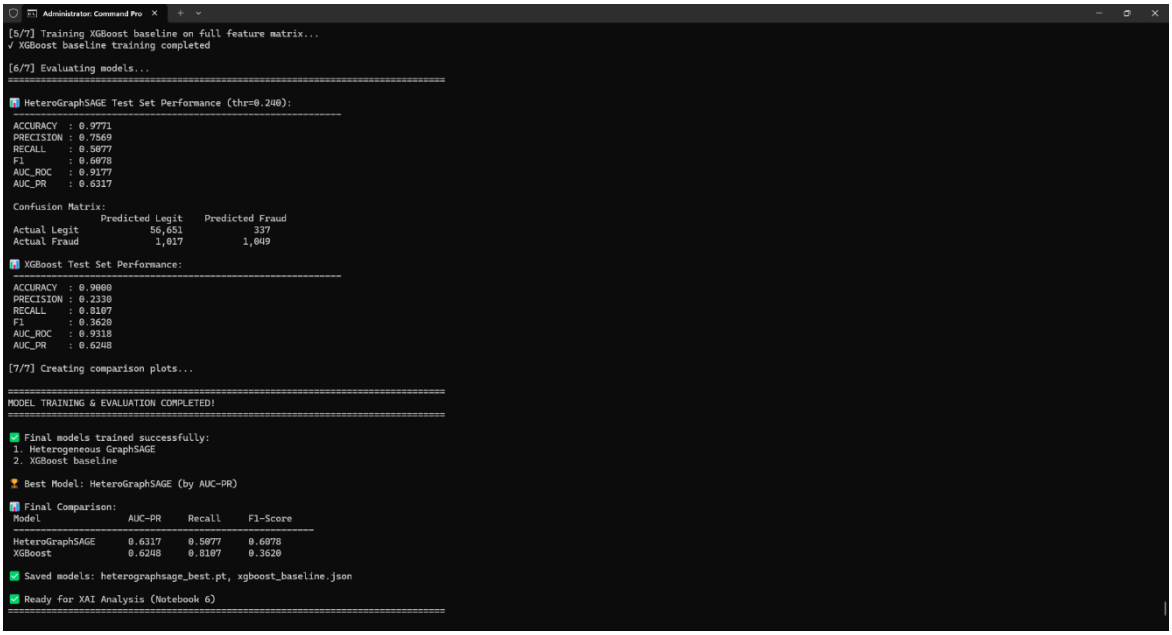
11. Notebook 4 - c



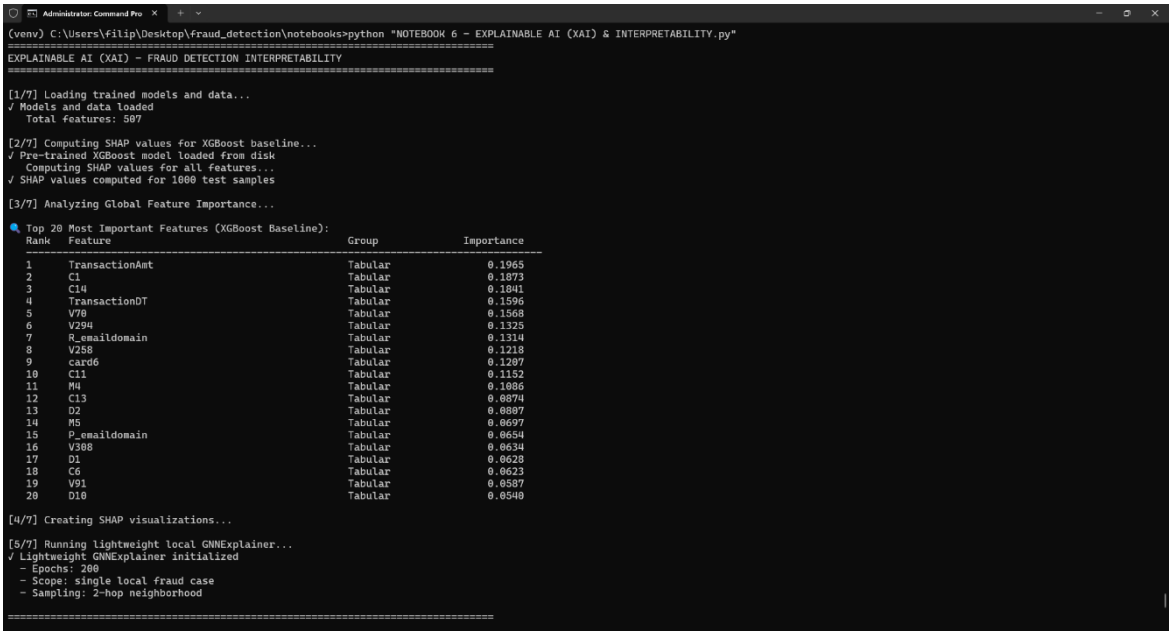
12. Notebook 5 - a



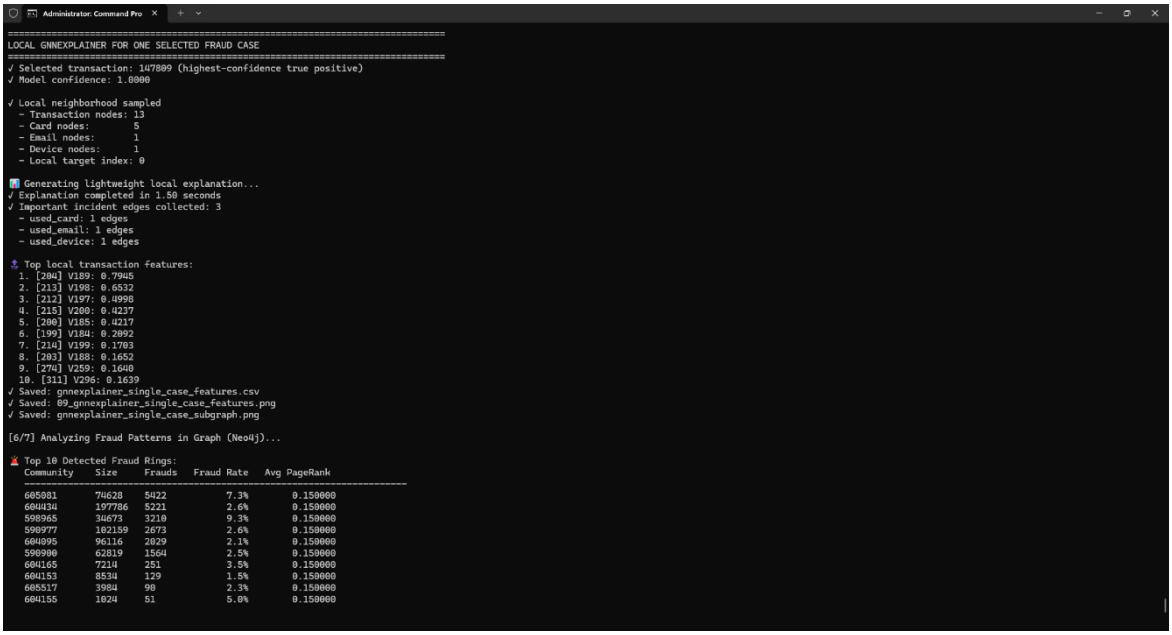
13. Notebook 5 - b



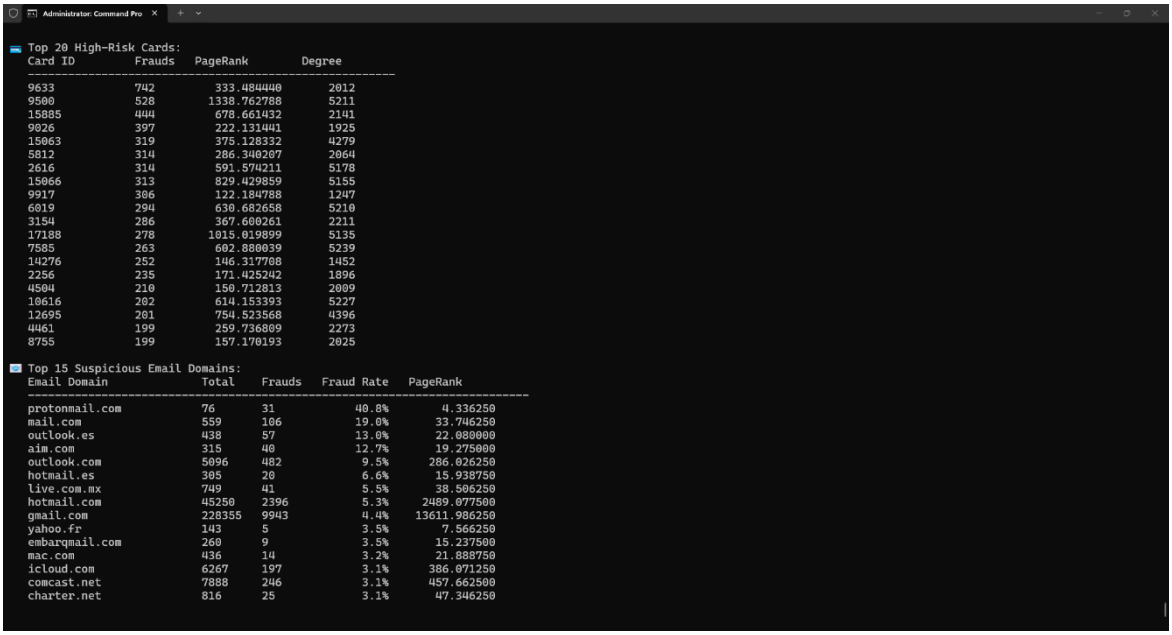
14. Notebook 5 - c



15. Notebook 6 - a

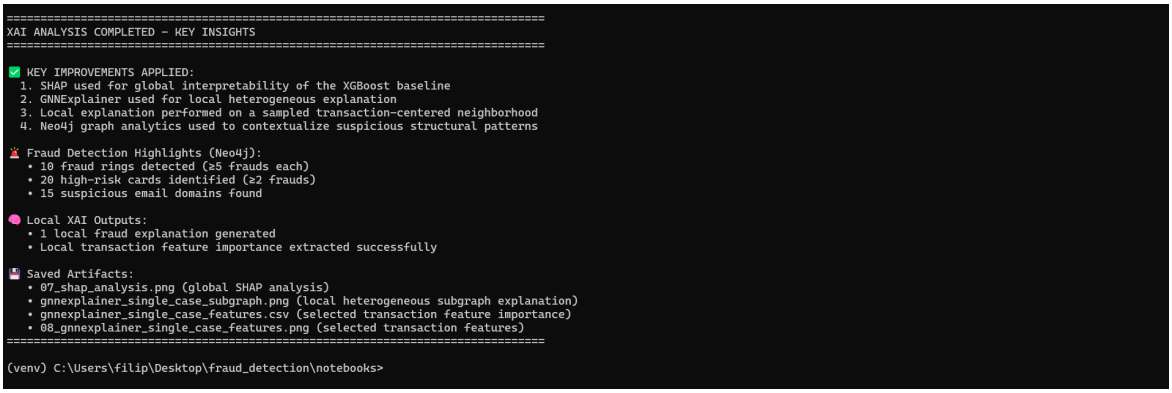


16. Notebook 6 - b



```
Administrator: Command Pro

Top 20 High-Risk Cards:
Card ID         Frauds   PageRank        Degree
------------------------------------------------------------
9633            742         333.484440       2012
9500            528        1338.762788       5211
15885           444         678.661432       2141
9026            397         222.131441       1925
15063           319         375.128332       4279
5812            314         286.340207       2064
2616            314         591.574211       5178
15066           313         829.429859       5155
9917            306         122.184788       1247
6019            294         630.682658       5210
3154            286         367.600261       2211
17188           278        1015.019899       5135
7585            263         602.880039       5239
14276           252         146.317708       1452
2256            235         171.425242       1896
4504            210         150.712813       2009
10616           202         614.153393       5227
12695           201         754.523568       4396
4461            199         259.736809       2273
8755            199         157.170193       2025

Top 15 Suspicious Email Domains:
Email Domain              Total    Frauds   Fraud Rate   PageRank
--------------------------------------------------------------------------------
protonmail.com            76       31           40.8%       4.336250
mail.com                  559      106          19.0%      33.746250
outlook.es                438      57           13.0%      22.080000
aim.com                   315      40           12.7%      19.275000
outlook.com               5096     482           9.5%     286.026250
hotmail.es                305      20            6.6%      15.938750
live.com.mx               749      41            5.5%      38.506250
hotmail.com               45250    2396          5.3%    2489.077500
gmail.com                 228355   9943          4.4%   13611.986250
yahoo.fr                  143      5             3.5%       7.566250
embarqmail.com            260      9             3.5%      15.237500
mac.com                   436      14            3.2%      21.888750
icloud.com                6267     197           3.1%     386.071250
comcast.net               7888     246           3.1%     457.662500
charter.net               816      25            3.1%      47.346250
```

17. Notebook 6 - c

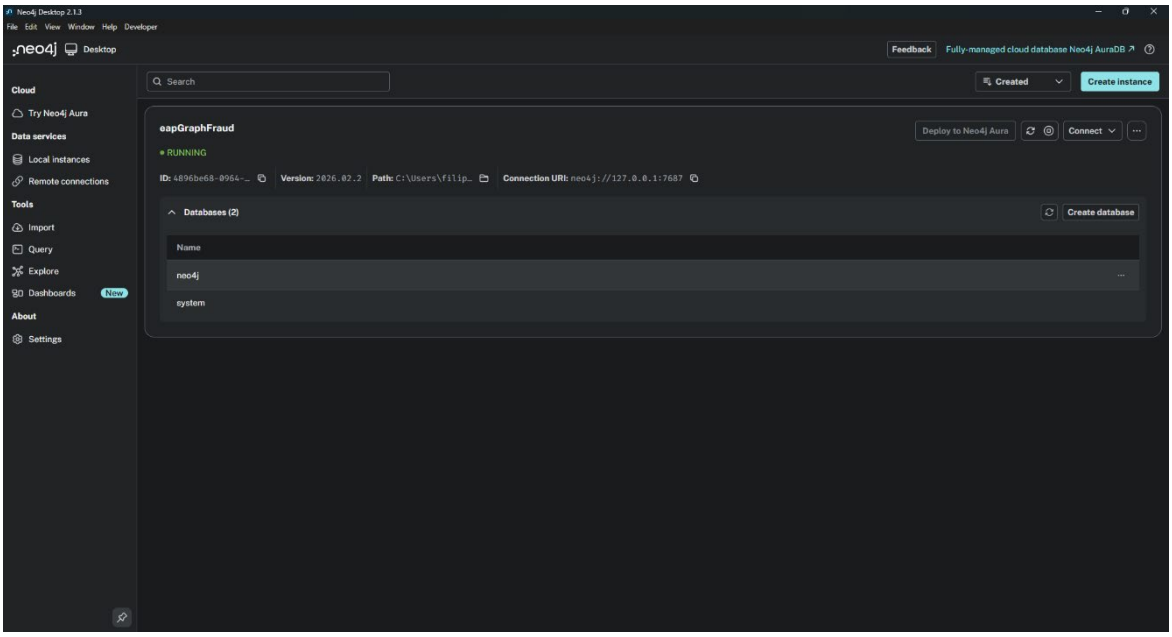


```
================================================================================
XAI ANALYSIS COMPLETED - KEY INSIGHTS
================================================================================

KEY IMPROVEMENTS APPLIED:
  1. SHAP used for global interpretability of the XGBoost baseline
  2. GNNExplainer used for local heterogeneous explanation
  3. Local explanation performed on a sampled transaction-centered neighborhood
  4. Neo4j graph analytics used to contextualize suspicious structural patterns

Fraud Detection Highlights (Neo4j):
   • 10 fraud rings detected (≥5 frauds each)
   • 20 high-risk cards identified (≥2 frauds)
   • 15 suspicious email domains found

Local XAI Outputs:
   • 1 local fraud explanation generated
   • Local transaction feature importance extracted successfully

Saved Artifacts:
   • 07_shap_analysis.png (global SHAP analysis)
   • gnnexplainer_single_case_subgraph.png (local heterogeneous subgraph explanation)
   • gnnexplainer_single_case_features.csv (selected transaction feature importance)
   • 08_gnnexplainer_single_case_features.png (selected transaction features)
================================================================================

(venv) C:\Users\filip\Desktop\fraud_detection\notebooks>
```

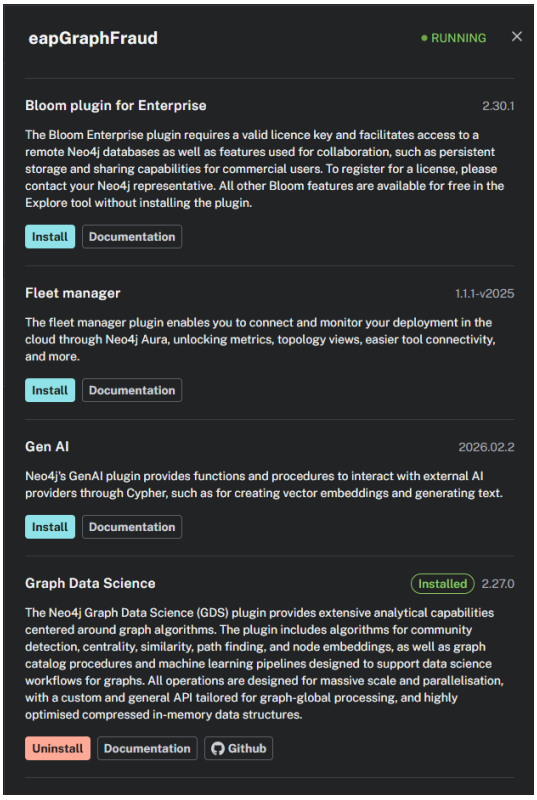
18. Notebook 6 - d

ii. Βάση Δεδομένων Γράφων – Neo4j Desktop

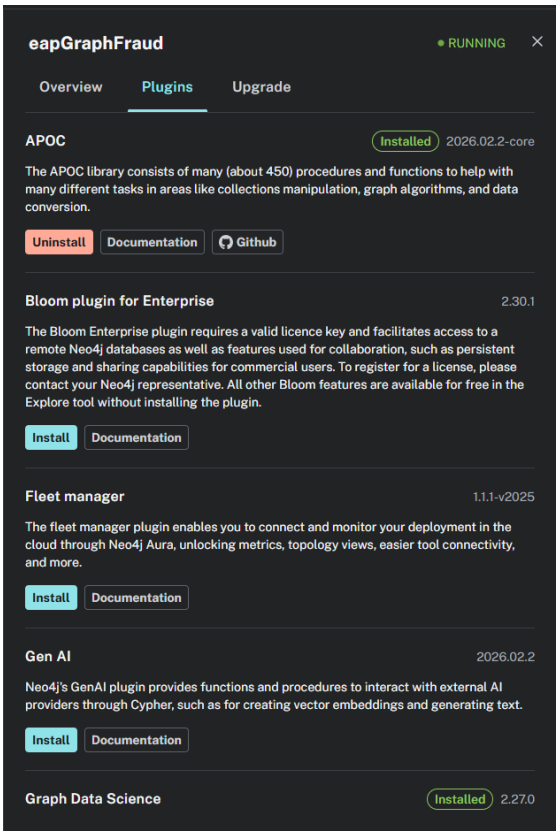
Εδώ φαίνεται το περιβάλλον διαχείρισης της ΒΔΓ, μετά τη δημιουργία της, καθώς και τα εγκατεστημένα plugins.



19. Neo4j Desktop - a



20. Neo4j Desktop - b



21. Neo4j Desktop - c

Παράρτημα Β: Κώδικας Python

Ο πλήρης πηγαίος κώδικας με τον οποίο υλοποιήθηκε το σύστημα/pipeline της παρούσας πτυχιακής εργασίας, είναι δημοσίως διαθέσιμος στο GitHub στον ακόλουθο σύνδεσμο: *https://github.com/fil-man/graph-fraud-detection-gnn.git*

Υπεύθυνη Δήλωση Συγγραφέα:

Δηλώνω ρητά ότι, σύμφωνα με το άρθρο 8 του Ν.1599/1986, η παρούσα εργασία αποτελεί αποκλειστικά προϊόν προσωπικής μου εργασίας, δεν προσβάλλει κάθε μορφής δικαιώματα διανοητικής ιδιοκτησίας, προσωπικότητας και προσωπικών δεδομένων τρίτων, δεν περιέχει έργα/εισφορές τρίτων για τα οποία απαιτείται άδεια των δημιουργών/δικαιούχων και δεν είναι προϊόν μερικής ή ολικής αντιγραφής, οι πηγές δε που χρησιμοποιήθηκαν περιορίζονται στις βιβλιογραφικές αναφορές και μόνον και πληρούν τους κανόνες της επιστημονικής παράθεσης.