Unsupervised Graph Anomaly Detection in a Large Payment Transaction Network

Graduate

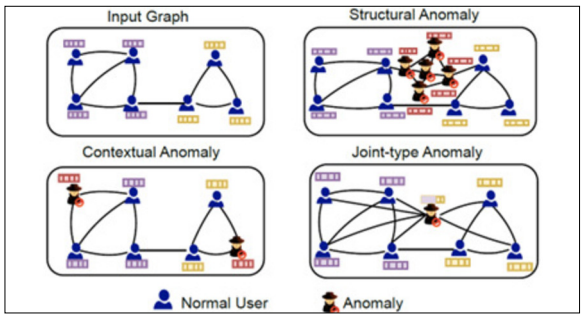
Roman Loop

Introduction: In the ever-changing world of financial transactions, identifying and interpreting anomalies is crucial for maintaining the integrity, security, and stability of financial systems. Anomaly detection systems have traditionally been rule-based or machine learning based, trained on tabular data. However, these systems are constrained in their capacity to identify complex relationships within a financial transaction network. In recent years, Graph Neural Networks (GNNs) have redefined anomaly detection by harnessing the relational structures present in graph data. This inherent capability to model complex relationships and dependencies empowers GNNs as potent tools for anomaly detection.

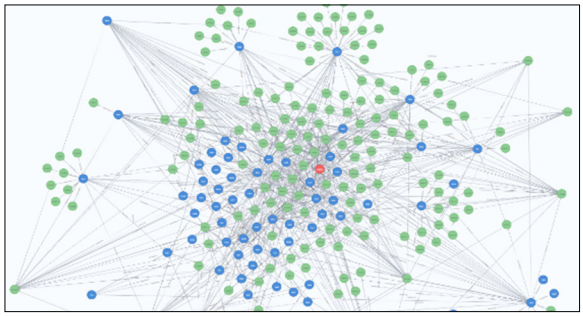
Approach / Technology: This study involves transforming a comprehensive payment transaction dataset into a graph, where nodes represent entities such as accounts or transactions, and edges represent the relationships between them. Many novel graph anomaly detection models have been published recently. This thesis focuses on two of these models: DOMINANT (Deep Anomaly Detection on Attributed Networks) and CONAD (Contrastive Anomaly Detection). Both models utilize an encoder-decoder architecture, using reconstruction errors as anomaly scores. However, CONAD incorporates well-known anomaly patterns, transforming the unsupervised learning task into a self-supervised one.

Conclusion: Although the research field of GNN-based anomaly detection is highly active and has made significant progress in recent years, it is still in its infancy regarding industrial adoption. The defined objective was to apply GNN-based anomaly detection models to a large transaction graph comprising millions of transactions, aiming to detect anomalies within this network. This thesis overcame the inherent scalability limitations of the DOMINANT and CONAD models by enabling them for outer loop mini-batch training. However, the study clearly demonstrated the limitations of these approaches. The two most significant issues are the lack of explainability in the results, and the constrained adaptability of state-of-the-art GNN-based approaches to the underlying graph.

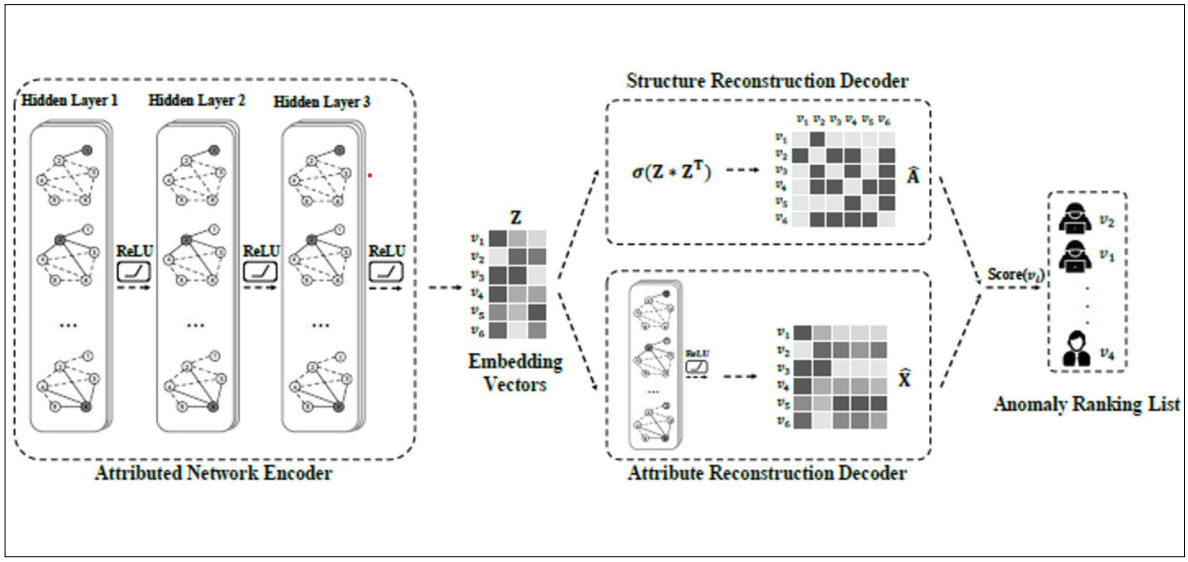
Anomaly Types
A. Roy, et al, GAD-NR, 2024



Detected Anomaly
Own presentment



DOMINANT Framework
K. Ding, et al, DOMINANT, 2019



Advisor
Dr. Shao Jü Woo

Co-Examiner
Christian Büchel, LGT Financial Services AG, FL-9490 Vaduz

Subject Area
Data Science, Computer Science