FACULDADE
DE ECONOMIA
UNIVERSIDADE
DO PORTO

Detecting Suspicious Activities in Waste Transport Data via Temporal and Network-Aware Change Detection

Sara Lopes dos Santos Oliveira

Dissertation

Master in Modelling, Data Analysis and Decision Support Systems

Supervised by:

PhD Shazia Tabassum

Co-supervised by:

PhD Bruno Veloso

2025

Acknowledgments

I would like to express my sincere gratitude to Shazia Tabassum PhD and Prof. Bruno Veloso, for their guidance and support. Their feedback and expertise was essential for the development of this work.

This work was inserted in a project at INESCT-TEC - Instituto de Engenharia de Sistemas e Computadores, Tecnologia e Ciência, and funded by national funds through FCT - Fundação para a Ciência e a Tecnologia, I.P., under the support 2024.07677.IACDC (https://doi.org/10.54499/2024.07677.IACDC)

Abstract

Waste management networks are essential for environmental protection and public health, but vulnerable to regulatory evasion, fraud, and illegal trading. The regulatory authorities perform periodic inspections to detect infringements or illicit activities and promote compliance.

The limited resources available to the authorities, and the difficulty of manually overseeing large volumes of waste transport data, warrant the need for a robust anomaly detection system. However, the complex relational nature of the anomalies, often involving the collusion of several entities, and their temporal irregularity make conventional anomaly detection approaches less effective.

In this study, we introduce a dynamic network-aware framework that combines statistical change detection methods, the Page-Hinkley and CUSUM tests, alongside a deep learning model, LSTM-VAE, to detect suspicious activities in the Portuguese waste management network. Using real-world anonymised waste transfer records, we engineered temporal and network features to reveal a wide range of anomalies, including abrupt shifts in activity and unusual connectivity patterns, such as those involving collusive triangles.

The evaluation was based on four pre-labeled anomalous companies identified by regulators. Our results show that while individual methods excel at detecting certain behaviors, their combination provides robust coverage of diverse anomaly types, with each anomalous company identified by at least three of the five selected techniques. This approach demonstrates the value of integrating temporal and network-based analysis for anomaly detection in the waste management network, and lays the foundation for an efficient, data-driven tool for regulatory authorities to prioritize inspections and enhance accountability in the sector.

Keywords: anomaly detection, change detection, autoencoders, waste management.

Resumo

O setor de gestão de resíduos é fundamental para a proteção ambiental e saúde pública, mas apresenta certas vulnerabilidades, tais como fraudes, evasão fiscal e comércio de resíduos furtados. Neste sentido, as autoridades reguladoras realizam inspeções periódicas para detetar infrações ou atividades ilícitas.

Contudo, a escassez de recursos e a dificuldade em monitorizar manualmente grandes volumes de dados levam à necessidade de um sistema robusto de deteção de anomalias, para sinalizar entidades para inspeção. Por outro lado, a complexa natureza relacional destas anomalias, que muitas vezes resultam do conluio de várias entidades, bem como as suas irregularidades temporais, limitam a eficácia das abordagens convencionais de deteção.

Neste trabalho, propomos uma abordagem que combina métodos estatísticos de change detection, os testes Page-Hinkley e CUSUM, com um modelo de deep learning, o LSTM-VAE, para identificar anomalias na rede de gestão de resíduos em Portugal. Para este efeito, usamos dados reais anonimisados de transportes de resíduos para construir variáveis representativas do comportamento relacional e temporal das empresas. Deste modo, pretendemos detetar diferentes tipos de anomalias, tais como mudanças abruptas na atividade das empresas ou padrões invulgares de conectividade entre entidades, como triângulos colusivos.

A abordagem proposta foi avaliada através de quatro empresas previamente identificadas como anómalas pelas entidades reguladoras. Todas as empresas foram identificadas por, pelo menos, três das cinco técnicas selecionadas. Este resultado sugere que, embora cada método individual seja eficaz na deteção de um certo tipo de comportamento, a combinação de vários métodos permite detetar de forma robusta uma maior variedade de anomalias. Este trabalho demonstra, assim, o valor de integrar as componentes relacionais e temporais na deteção de anomalias na rede de gestão de resíduos, e estabelece as bases para desenvolver uma ferramenta analítica eficiente para priorizar inspeções e promover o bom funcionamento do setor.

Palavras-chave: deteção de anomalias, change detection, autoencoders, gestão de resíduos.

Contents

Acknowledgements . . . iii
Abstract . . . v
Resumo . . . vii

1 Introduction **1**
1.1 Motivation . . . 1
1.2 Problem Definition . . . 1
1.3 Dissertation Structure . . . 3

2 Theoretical Foundations **5**
2.1 Anomaly Detection . . . 5
2.2 Page-Hinkley Test . . . 6
2.3 Cumulative Sum Test . . . 6
2.4 LSTM-VAE . . . 7

3 Literature Review **9**
3.1 Change Detection . . . 9
3.2 Machine Learning . . . 10
3.3 Deep Learning . . . 11
3.4 Discussion . . . 12

4 Methodology **15**
4.1 Business Understanding . . . 15
4.2 Data Understanding . . . 16
4.2.1 Exploratory Data Analysis . . . 17
4.3 Data Preparation . . . 19
4.3.1 Data Cleaning . . . 19
4.3.2 Feature Creation . . . 21
4.4 Modeling . . . 26
4.4.1 Page-Hinkley and Cumulative Sum tests . . . 26
4.4.2 LSTM-VAE . . . 28
4.4.3 Anomaly Scoring . . . 30

5 Experiments and Results **33**
5.1 Page-Hinkley and CUSUM Tests . . . 33
5.2 LSTM-VAE . . . 35
5.3 Evaluation . . . 37
5.4 Analysis of Anomalous Companies . . . 40

5.5 Discussion 42

6 Conclusion **45**

Bibliography **47**

A Appendix **i**
A.1 e-GAR columns with missing values i
A.2 Distribution of type of entity by number of unique senders/receivers and number of transfers i
A.3 Total number of transfers by weekday. i

Glossary

APA - Agência Portuguesa do Ambiente
CRISP-DM - Cross- Industry Standard Process for Data Mining
CUSUM - Cumulative Sum
CV - Coefficient of Variation
IGAMAOT - Inspeção-Geral da Agricultura, do Mar, do Ambiente e do Ordenamento do Território
LSTM-VAE - Long Short-term Memory Variational Autoencoders
PH - Page-Hinkley

List of Figures

4.1 Distribution of the total number of transfers by waste receiver. 18
4.2 Daily, weekly and monthly number of transfers. 20
4.3 Temporal distribution of the features per-company. 22
4.4 Distribution of per-company percentage of zeros in the features. 23
4.5 Correlation between the features. 24
4.6 Distribution of the per-company coefficients of variations (CV) for each feature. 25
4.7 Weekly coefficient of variation (CV) for each feature. 25
4.8 Example of Page-Hinkley test with $\delta = 1$, threshold=2: Demonstrates high sensitivity, detecting minor fluctuations. 27
4.9 Example of Page-Hinkley test with $\delta = 2$, threshold=3: Demonstrates a balanced setting, detecting significant shifts. 27
4.10 Example of Page-Hinkley test with $\delta = 4$, threshold=5: Demonstrates low sensitivity, missing important changes. 28
4.11 Parameter sensitivity analysis results for LSTM-VAE-2F. 29
4.12 Parameter sensitivity analysis results for LSTM-VAE-3F. 30

5.1 Distribution of anomaly scores for Page-Hinkley and CUSUM tests. 35
5.2 Distribution of final anomaly scores for the LSTM-VAE-2F model. 35
5.3 Distribution of intra-company score coefficients of variation for the LSTM-VAE-2F model. 36
5.4 Distribution of final anomaly scores for the LSTM-VAE-3F model. 37
5.5 Distribution of intra-company score coefficients of variation for the LSTM-VAE-3F model. 37
5.6 Heatmap of anomaly scores for labeled anomalous companies across the detection methods. 38
5.7 Heatmap of correlation between anomaly scores obtained by the detection methods. 39
5.8 Distribution of the final aggregated anomaly scores across all companies. . . 39
5.9 Anomaly score profile for the top-scoring company (Score: 0.982). 41
5.10 Anomaly score profile for a high-scoring company (Score: 0.947). 42

List of Tables

5.1 Summary of Page-Hinkley and CUSUM test results, including companies removed by the Minimum Activity Filter, total changes detected, and number of companies with detected changes. 34
5.2 Summary statistics of the final aggregated anomaly scores. 40
5.3 Final anomaly scores and percentiles for the four labeled anomalous companies. 40

Chapter 1

Introduction

In this chapter, the overall context, motivation, and objectives of this dissertation are presented, establishing a foundation of the research problem guiding this work. We begin by defining the motivation of the work, providing its context within the waste management sector, its vulnerabilities and its implications for environmental protection and public health. Then, the problem definition section specifies the problem encountered by the regulatory authorities of the waste management sector, and defines how our work will address these challenges, stating the main research objectives. Finally, we conclude with an outline of how this dissertation is structured.

1.1 Motivation

Waste management is an essential sector, as it can have a critical impact on environmental protection and public health. Therefore, waste management is subject to strict regulations to ensure its adequate functioning and minimal environmental and social costs. Waste management regulations often impose financial and administrative burdens, and some entities may engage in falsifying, misreporting, or omitting waste transports to avoid regulatory obligations, reduce costs, and evade taxes or fees associated with proper waste disposal. Transporters may misreport the type of waste being moved to bypass restrictions on hazardous materials or to avoid higher disposal costs. Additionally, organized criminal schemes can arise: some types of waste, such as scrap metals and car parts, are significantly valuable and can be easily taken from construction sites, public infrastructures or vehicles. Since these types of waste can be sold to recycling facilities, they become attractive targets for theft and illegal trading. Other entities may engage in waste trafficking operations, which take advantage of regulatory inconsistencies across countries to turn a profit. Such activities can have a severe environmental and public health impact, and thus warrant intervention and preventative action from the responsible authorities.

1.2 Problem Definition

Given the vulnerability of the waste management sector to illicit activities and regulatory non-compliance, the regulatory authorities in Portugal, namely the inspectors of Inspeção-Geral da Agricultura, do Mar, do Ambiente e do Ordenamento do Território (IGAMAOT), run periodic inspections on the waste operating companies, with the aim to detect illegal and

criminal practices and to ensure transparency and regulatory compliance in the Portuguese waste management network.

However, to optimize the utilization of limited resources, it is necessary to improve the precision of inspections by focusing on companies that exhibit suspicious behavior and are more likely to have committed regulatory infringements. Nevertheless, monitoring and analyzing extensive waste transport records to identify anomalous patterns presents a significant challenge, which requires a systematic, accurate, and efficient approach.

For the purpose of developing such an approach, the Portuguese Environmental Agency (Agência Portuguesa do Ambiente - APA) has provided access to the waste transportation records from 2022 of the Portuguese waste management network. These transportation records consist of electronic documents (e-GAR) that are required to accompany each instance of waste transport within the national territory. This dataset encompasses rich information on the dynamics of the waste management network, such as the date of the transfer, the identification of the receiver and sender, and type and quantity of waste.

Traditional fraud detection methods, on one hand, often fail to capture the full temporal and structural complexity of dynamic network data. Aggregating data over time often masks short-term fluctuations or irregularities that may indicate suspicious behavior. For instance, a company might show irregular monthly patterns as a result of fraudulent transactions, a signal which would get smoothed out in aggregation. Similarly, traditional tabular or non-network models ignore the structure of interactions between actors. In the waste management context, the relationships between senders and receivers can reveal critical patterns such as connections with suspicious entities or sudden changes in network connectivity that flat-data models cannot capture. Therefore, actors involved in irregular activities often mimic normal behavior when observed through tabular or static approaches.

Therefore, it is necessary to develop an approach that can capture the complex temporal and structural patterns in the data, and detect deviations that may indicate the occurrence of illicit activities. The temporal perspective is particularly important in uncovering such activities, as they tend to happen sporadically rather than continuously, and therefore can be detected by significantly deviating from the own entity's usual activity patterns. On the other hand, the network perspective is also essential as these illicit activities and schemes often stem from the collaboration of several entities.

Therefore, in this work we propose to develop an anomaly detection framework that combines two change detection techniques, the Page-Hinkley (PH) test and the Cumulative Sum (CUSUM) test, and a deep learning approach based on Long Short-term Memory Variational Autoencoders (LSTM-VAE), to detect significant deviations or unusual patterns in the temporal and network behavior of the Portuguese waste management companies.

Specifically, we aim to achieve this goal by:

- Interpreting and understanding the rich information contained in the electronic waste transport records.

- Creating novel features representing temporal, operational, and structural behavior of waste operating companies, including network-based metrics, that are suggestive of suspicious behavior.

- Adapting statistical change detection (Page-Hinkley and CUSUM tests) and deep learning techniques (LSTM-VAE) for anomaly detection in the waste management network.

- Combining the approaches developed to obtain a robust anomaly detection framework.
- Evaluating the anomaly detection framework through a set of known anomalous companies.

1.3 Dissertation Structure

In terms of structure, the dissertation is organized in six chapters, from context and problem definition to methodology, presentations of results, and final conclusions. Following this introductory chapter, Chapter 2 presents the theoretical foundations necessary to understand both the core concepts of anomaly detection in dynamic network data and the specific statistical and machine learning techniques utilized in this work. Chapter 3 reviews the existing literature on anomaly detection in sequential and network-based data, highlighting current advances and gaps in applications to waste management networks. Chapter 4 describes the methodology employed, corresponding to the CRISP-DM framework, and details every stage from business and data understanding to feature engineering and model development. The subsequent Chapter 5 presents the experimental results, and provides an evaluation of the proposed methods against a set of pre-labeled anomalous companies. Finally, Chapter 6 concludes the dissertation by summarizing the main findings, discussing the identified limitations and the potential directions of future work.

Chapter 2

Theoretical Foundations

In this chapter, we provide the theoretical foundations underlying our work. We begin by introducing the concept of anomaly detection and its specific challenges in the context of dynamic graphs, including the use of change detection for this specific problem. We then present the main techniques used in our proposed framework, which include the Page-Hinkley test, the Cumulative Sum (CUSUM) test, and the Long Short-Term Memory Variational Autoencoder (LSTM-VAE).

2.1 Anomaly Detection

A classic and commonly used definition of anomaly is the one given by Hawkins (1980): "An outlier is an observation that differs so much from other observations as to arouse suspicion that it was generated by a different mechanism.". In practical terms, an anomaly is treated as a data object or a group of objects that is rare (e.g, rare combination of categorical attribute values), isolated (e.g., far-away points in n-dimensional spaces), and/or surprising (e.g., data instances that do not fit well in our mental/statistical model) (Akoglu, Tong, & Koutra, 2015). The general definition of anomaly is always vague, as its meaning is highly dependent on the context: the type of data and problem and the purpose of identifying the anomalies.

Anomaly Detection is the branch of Data Mining that is concerned with discovering anomalies. Anomaly Detection can be done as part of data cleaning, or it can be used to identify anomalies that are relevant in themselves: for example, fraud or other criminal activities, diseases, online network intrusions, machine failures (Akoglu et al., 2015).

Detecting anomalies in graph data is a significantly different task than that of detecting anomalies in a multi-dimensional feature space, due to the presence of interdependencies between the data points (Akoglu et al., 2015). These interdependencies make the anomaly detection a more complex and challenging task, and render traditional anomaly detection less effective, as they typically assume the data points to be independent and identically distributed (Shaji et al., 2025). However, the structuring of data in an interconnected graph can also bring additional benefits to anomaly detection, as it can help the abnormality of a data point to stand out. For instance, in fraud detection, behavioral clues may be faked or altered by the fraudster, but it may be more difficult for a fraudster to blend seamlessly into an entire network, as they cannot know its entire characteristic structure and dynamic operations (Akoglu et al., 2015). On the other hand, some anomalies are themselves relational. For example, organized crime

that is performed by a group of related subjects, or machine failure that causes a chain reaction in other machines, or that is caused by environmental factors that therefore makes failure of other machines more likely (Akoglu et al., 2015).

Anomaly Detection in dynamic graphs is a yet more challenging problem, due to the coupling of attribute and structural information with temporal information. Some dynamic graph anomaly detection approaches focus on detecting anomalies at each snapshot of an evolving graph. This can be interesting for some applications, for example, for real-time detection anomalies, such as online network intrusions. However, taking into account the structural and attribute information in relation to its past instances can provide valuable insights for anomaly detection. Thus, in dynamic graphs, detecting anomalies often becomes a change detection problem.

Change detection is the area of Data Mining which aims to discover significant alterations or patterns in data that is evolving over time. Change detection has several important applications, such as in social network analysis, financial market analysis, network traffic monitoring, and healthcare monitoring. It can be used to detect early signs of events that can have serious impact, and that can therefore be prevented or handled more promptly (Hewapathirana, 2019). Most change detection methods consist in monitoring a representative summary extracted from the overall graph, or from subgraphs, nodes or edges to characterize the patterns of behavior in the recent past and analyse deviations (Hewapathirana, 2019).

2.2 Page-Hinkley Test

The Page-Hinkley test is a change detection technique that monitors sudden changes in the mean of a time series. For a sequence of observations $x_1, x_2, ..., x_t$, the test computes the cumulative sum of deviations from a reference value (typically the running mean). The core statistic, PH_t, is defined as:

$$PH_t = \sum_{i=1}^{t} (x_i - \mu_t - \delta) \tag{2.1}$$

where μ_t represents the mean of the process up to time t and δ is a sensitivity parameter for small changes. A change is detected when the difference between the cumulative sum and its historical minimum exceeds a predefined threshold:

$$PH_t - \min_{1 \leq j \leq t} PH_j > \text{threshold} \tag{2.2}$$

At each change detected, the cumulative sum is reset. The standard Page-Hinkley test is one-sided, monitoring for increases in the mean. While there is no predefined measure of the strength of a detected change, the following measure can be used as a proxy:

$$\text{Change Strength}_{\text{PH}} = PH_t - \min_{1 \leq j \leq t} PH_j \tag{2.3}$$

The Page-Hinkley's test parameters are therefore δ and the detection threshold.

2.3 Cumulative Sum Test

The Cumulative Sum (CUSUM) test is a change detection technique designed to detect gradual and sustained changes in the mean of a time series. It maintains two cumulative

sums to monitor both positive and negative deviations. For each time point t, these sums are updated as:

$$S_t^+ = \max\left(0, S_{t-1}^+ + x_t - \mu - k\right) \tag{2.4}$$

$$S_t^- = \max\left(0, S_{t-1}^- - x_t + \mu - k\right) \tag{2.5}$$

where μ is the expected value of the process and k is the drift parameter that allows for minor fluctuations. A change is detected when either cumulative sum exceeds a specified threshold.

The strength of a detected change is calculated as:

$$\text{Change Strength}_{\text{CUSUM}} = \max\left(S_t^+, S_t^-\right) \tag{2.6}$$

The key parameters of the CUSUM test are the drift (k), the detection threshold, and the reference (target) value μ.

2.4 LSTM-VAE

The Long Short-Term Memory Variational Autoencoder is a deep generative model that combines the ability of LSTM networks to model sequential data with the probabilistic latent variable framework of Variational Autoencoders. While the Variational Autoencoder learns a compressed, probabilistic representation of normal activity that is robust to noise and variability in the data, the LSTM component captures long-term dependencies and temporal dynamics in the companies' behavior, which is critical for distinguishing between normal fluctuations and anomalies.

The model consists of an encoder LSTM that processes input windows of length t. Each input has shape (t, n), where t is the window size and n is the number of features. The encoder outputs two vectors, the mean and the log-variance, which define the approximate posterior distribution over the latent variables. A latent vector of dimension d is then sampled from this distribution using the reparameterization trick. This sampled latent vector is passed to a decoder LSTM, which reconstructs the original input sequence. The model is trained to minimize a loss function that combines two terms: the reconstruction loss and the Kullback–Leibler divergence between the approximate posterior and a standard normal prior. The impact of the KL divergence term on the total loss is controlled by the parameter β, which regulates how much weight is given to the KL divergence term in relation to the reconstruction loss, forcing the model to focus more on either accurate reconstruction or on learning a more regularized latent space.

The reconstruction error for an input sequence x and its reconstruction $\hat{x}$ is defined as the mean square error:

$$\text{Reconstruction Error} = \frac{1}{T \cdot N} \sum_{i=1}^{T} \sum_{j=1}^{N} (x_{i,j} - \hat{x}_{i,j})^2$$

where $x_{t,n}$ is the original value at time t for feature n, and $\hat{x}_{t,n}$ is the reconstructed value. High reconstruction errors indicate that the observed sequence deviates from the patterns seen during training, and can therefore be used for anomaly detection.

The main parameters of the model are therefore the dimension of the latent representations (d), the size of the input windows (t), the weight of the KL divergence term (β), and the number of hidden units in the LSTM layers, which determines the dimensionality of the

LSTM's hidden state and thus the model's capacity to capture temporal dependencies. Other hyperparameters include the dropout and recurrent dropout rates, which control the fraction of units randomly dropped during training to prevent overfitting, the batch size (number of samples processed per update), the number of epochs (iterations over the full training dataset), the random seeds used to control stochasticity in training, and the choice of the optimizer, which is the algorithm that adjusts the model's internal weights during training to minimise the loss function.

Chapter 3

Literature Review

Given our previously defined anomaly detection problem, we performed a review of existing approaches for anomaly detection in sequential and graph-based data. We explore statistical change detection methods, traditional machine learning techniques, and advanced deep learning approaches. Furthermore, we examine frameworks specifically designed for anomaly detection within the waste management network and in logistics or transportation networks. The goal is to understand the existing anomaly detection approaches and to contextualize the novel contributions of this work in addressing the unique challenges of detecting suspicious activities in the waste management network.

3.1 Change Detection

Change detection in data streams is an important area of research, with broad applications in fields such as industrial monitoring, finance, and cybersecurity, as well as in adaptive learning systems and real-time analytics. The core goal is to identify points in time where the underlying data distribution changes, which can signal important events, anomalies, or emerging patterns that require attention. Change detection techniques specifically target the identification of significant shifts in data distributions over time. Notable change detection methods include:

- **CUSUM and Page Hinkley Tests:** The CUSUM test monitors the cumulative sum of deviations from a target value and is effective for detecting small, persistent changes in mean or variance. The Page-Hinkley test, on the other hand, is designed to detect abrupt changes in the average of a data stream and is known for its robustness and low false alarm rates (Kifer, Ben-David, & Gehrke, 2004). Both methods have been extensively applied in real-time anomaly detection scenarios, such as industrial process control, financial fraud detection, and event detection in social networks (Gama, 2012; Kifer et al., 2004; Pereira, Tabassum, Gama, de Amo, & Oliveira, 2018).

- **Sliding Window Approaches:** These methods operate by comparing the data distributions within two windows: a reference window (containing historical, normal data) and a current window (containing the most recent data). A change is detected when a significant difference is observed between the distributions in these two windows (Fathy, Barnaghi, & Tafazolli, 2019; Kifer et al., 2004; Lu et al., 2019).

- **Non-parametric Techniques:** Unlike parametric methods, they do not rely on assumptions on the underlying data distribution, and often rely on statistical measures like distance or discrepancy between data samples in different windows to detect changes, making them versatile for various data types (Faithfull, Rodríguez, & Kuncheva, 2019; Frittoli, Carrera, & Boracchi, 2023).

- **Martingale-based Methods:** These techniques convert nonconformity measures from data into a "martingale," which is a sequence of random variables representing a process where the expected future value, given past information, is equal to the current value. Deviations in the martingale's behavior signal a change or anomaly, and are often used in online settings (Calikus, Nowaczyk, Sant'Anna, & Dikmen, 2020; Namoano, Emmanouilidis, Ruiz-Carcel, & Starr, 2020).

- **Bayesian Approaches:** These methods incorporate prior knowledge about potential changes and data distributions into a probabilistic model. They continuously update the probability of a change occurring as new data arrives, providing a measure of confidence in the detection. This framework allows for identifying changes even with limited data and can handle multiple changepoints (Cano, Gómez-Olmedo, & Moral, 2019; X. Chen & Wu, 2025; Namoano et al., 2020; Sebastiao & Gama, 2009).

- **Principal Component Analysis-based Frameworks:** PCA is a dimensionality reduction technique that transforms high-dimensional data into a lower-dimensional space while retaining most of the variance. In change detection, PCA can be used to monitor the changes in the underlying structure of multivariate data. Anomalies or shifts in the data distribution are detected when new data points deviate significantly from the learned principal components or when the principal components themselves change over time (Qahtan, Alharbi, Wang, & Zhang, 2015).

These statistical techniques are valued for their simplicity, low computational requirements, and ability to provide statistical guarantees on the reliability of detected changes Nogara Notarianni, Leveni, Stucchi, Frittoli, and Boracchi (2025). However, they may struggle with non-linear or highly complex temporal dependencies, and they are sensitive to parameter choices and noise Gama (2012).

3.2 Machine Learning

Traditional machine learning models are frequently used for anomaly detection, offering a diverse set of techniques to identify deviations from normal behavior. However, their direct application to raw sequential data often requires significant pre-processing or feature engineering to transform the dynamic sequences into static feature vectors. This process, which can involve extracting statistical summaries over time windows, frequency-domain features, or aggregated counts, means that these approaches primarily operate on snapshots or derived representations of the sequential data (Davis, Raina, & Jagannathan, 2020). Consequently, they frequently lose crucial temporal dependencies inherent in dynamic systems, which are vital for understanding and detecting complex anomalies (Mejri et al., 2024).

The main traditional machine learning approaches used include k-Means Clustering (Kiss, Genge, Haller, & Sebestyen, 2014; Yairi, Kato, & Hori, 2001), Density-Based Spatial Cluster-

ing of Applications with Noise (DBSCAN) (Jain, Bajpai, & Pamula, 2022; Liu, 2021), Isolation Forest (Li et al., 2021) and Local Outlier Factor (Belay, Blakseth, Rasheed, & Rossi, 2023; Oehmcke, Zielinski, & Kramer, 2015; Wei & Wang, 2022).

In k-Means clustering, data points are partitioned into clusters based on their similarity, and anomalous points are identified as those that are far from any cluster centroid or that form very small, isolated clusters. DBSCAN instead defines clusters as dense regions of data and treats points in low-density areas as anomalies, which makes it effective for identifying irregular patterns without needing the number of clusters in advance. Local Outlier Factor (LOF) focuses on the local density of each point relative to its neighbors, flagging anomalies as those with significantly lower density. Isolation Forest takes a different approach by recursively partitioning the data, and anomalous points are those that can be isolated with fewer splits, making it efficient at identifying rare and irregular events.

In addition to these techniques, machine learning methods have also been applied to dynamic graph structures, which are increasingly relevant in applications such as networked systems. These approaches extract node or edge representations and create anomaly metrics that are updated incrementally and tracked to detect anomalies (Guo, Zhou, & Skiena, 2022; Yoon, Hooi, Shin, & Faloutsos, 2020).

Despite their utility and adaptability through feature engineering, these traditional machine learning approaches often fall short in fully capturing the complex non-linear relationships and intricate temporal dependencies like the ones found in the waste management network. Subtle anomalies that manifest as changes in the temporal sequence or network structure, rather than simple deviations in aggregated values, can be easily missed. This highlights the need for more sophisticated models that can inherently learn and process the rich temporal information within these complex data types, directly motivating the development and application of deep learning approaches.

3.3 Deep Learning

Deep learning approaches represent a subfield of machine learning that uses artificial neural networks with multiple layers (deep neural networks) to learn complex patterns and representations from large-scale data. Unlike traditional machine learning methods, deep learning models can automatically extract hierarchical features from raw data, making them particularly powerful for complex, high-dimensional, and sequential datasets. In anomaly detection, deep learning models are used to learn the intricate normal behavior of a system, and then identify deviations from this learned normal state as anomalies. This can involve learning representations that make anomalies easily distinguishable, or reconstructing normal data and flagging instances with high reconstruction errors. Recent advancements in deep learning have significantly improved anomaly detection in time series and dynamic graph data:

- **LSTM-based Autoencoders**: Long Short-Term Memory networks are a type of recurrent neural network used for modeling sequential data and capturing long-term dependencies. An Autoencoder is a neural network trained to reconstruct its input. An LSTM-based autoencoder combines these by using LSTMs in both the encoder and decoder components. The model is trained on normal data to learn its typical patterns and reconstruct them accurately. During detection, if the model struggles to reconstruct a new data point, leading to a high reconstruction error, that point is flagged as

an anomaly, indicating it deviates significantly from the learned normal behavior. For time series, LSTM-based autoencoders have shown promise (R.-Q. Chen, Shi, Zhao, & Liang, 2019; Maleki, Maleki, & Jennings, 2021). While powerful, prediction-based LSTM networks can face challenges like gradient explosion in long time series and low detection efficiency due to sequential processing, and reconstruction-based VAEs may neglect sequential relationships and time intervals (Wang & Zhang, 2025). Niu et al. addressed some of these issues by proposing an LSTM-based VAE-GAN method for time series anomaly detection, leveraging the mapping ability of the encoder and the discrimination ability of the discriminator (Niu, Yu, & Wu, 2020).

- **Graph Neural Networks (GNNs)**: GNNs are deep learning models designed to process data structured as graphs. They work by iteratively aggregating information from a node's neighbors to learn a rich representation (embedding) for each node and/or the entire graph. This allows them to capture complex relational information and structural patterns that traditional methods might miss. In anomaly detection, GNNs can identify anomalous nodes, edges, or subgraphs by recognizing deviations from learned normal graph structures or behaviors (Deng & Hooi, 2021; Wu, Dai, & Tang, 2022; Zhao, Sawlani, Srinivasan, & Akoglu, 2022).

3.4 Discussion

Overall, the importance of change detection in data or network streams lies in its ability to provide early warnings of system failures, fraud, or other critical events, to enable timely intervention and adaptation. While traditional statistical methods offer interpretability and efficiency, they struggle with understanding complex temporal patterns. On the other hand, deep learning models provide greater flexibility and performance in complex data. A combination of these approaches is often beneficial, combining the strengths of each to achieve robust and accurate anomaly detection in sequential data.

For instance, change detection techniques, such as CUSUM and PH tests, complement other methods by monitoring cumulative deviations in data streams, helping to identify distributional shifts that signal potential anomalies, in offline or real-time applications. Further network-aware change detection exploits structural information from connections, interaction patterns, and paths between entities, enabling the detection of relational anomalies that are not apparent in isolated temporal analyses. Incorporating such network topology dynamics has been shown to significantly improve the accuracy and robustness of detecting malicious activities in complex, interconnected systems (Tabassum, Azad, & Gama, 2020).

In the context of waste management, literature on anomaly detection remains sparse. A notable exception is the work of Shaji et al. (2025), who proposed a graph-based framework for analyzing the e-GAR waste transportation data. The authors employ a hybrid approach combining network-based analysis and machine learning to detect anomalies in the waste transportation data. Their method integrates sub-graph detection, community structure analysis, and centrality measures to extract structural features, along with custom edge weight imbalance metrics. Machine learning techniques, including clustering, autoencoders, and density-based methods, are then applied for detecting anomalies. Their approach demonstrated the value of network analysis over traditional transaction-level methods, particularly for detecting edge weight imbalances and unusual network structures. Our work extends this

approach by introducing temporal analysis through change detection and deep learning, to detect anomalies in the network and operational behavior over time.

Beyond waste management, Davis et al. (2020) also developed an end-to-end deep learning-based anomaly detection model for temporal data in transportation networks. Their proposed model adapts the Long Short-Term Memory network and combines it with an objective function based on Extreme Value Theory, obtaining improved anomaly detection performance when compared to various established statistical, machine learning, and hybrid deep learning baselines (Davis et al., 2020).

In conclusion, while significant progress has been made in anomaly detection across various fields, a clear gap exists in applying advanced temporal and network-aware change detection techniques and deep learning models to the unique challenges of the waste management sector. Existing work in this specific domain relies on static or aggregated network analysis, or performs temporal analysis without considering the network aspect, therefore failing to capture both the sporadic nature and the relational nature of illicit activities. This work addresses this gap by adapting and evaluating change detection methods and deep learning architectures, such as LSTM-VAEs, in conjunction with novel temporal, operational, and network features derived from real-world waste transport data. The change detection techniques aim to detect suspicious deviations in the temporal evolution of the companies' operational and network behavior, while the LSTM-VAE model will be employed to capture more complex unusual patterns over time. By integrating these approaches, we aim to provide a robust and precise framework for flagging entities with suspicious behaviors, to be inspected by the regulatory authorities.

Chapter 4

Methodology

In this work, the adopted methodological approach is based on the CRISP-DM (Cross-Industry Standard Process for Data Mining) framework, a widely recognized standard for structuring data science projects. CRISP-DM divides the analytical process into distinct, phases: business understanding, data understanding, data preparation, modeling, evaluation, and, when applicable, deployment. This framework offers a clear and systematic approach from problem definition to deployable results, making it particularly suitable for complex, real-world applications such as our own anomaly detection problem.

Accordingly, the following sections of this chapter detail the phases of business understanding, data understanding, data preparation, and modeling, while the evaluation phase will be presented and discussed separately in the chapter 5. While this project is being developed with the purpose of deployment, in this study we are only demonstrating and evaluating our proposed approach, which will still see further developments and the addition of more data before its actual deployment.

4.1 Business Understanding

The effective regulation of waste management operations is essential for environmental protection and public health. The waste management sector presents inherent vulnerabilities to various forms of illicit activities, including fraud, regulatory evasion, and organized criminal schemes. These activities stem from the significant financial and administrative burdens imposed by waste management regulations, which can incentivize entities to misreport or omit waste transports to avoid obligations, reduce costs, evade taxes or fees. For example, transporters may intentionally misclassify waste types to avoid restrictions on hazardous materials or to avoid the higher costs associated with their proper disposal. Furthermore, the valuable nature of certain types of waste, such as scrap metals and vehicle components, makes them attractive targets for theft and illegal trading. Such materials can be stolen and then illegally sold to recycling facilities. Additionally, some entities may engage in waste trafficking, exploiting inconsistencies in regulations across different countries to generate profits. These activities have significant detrimental impacts on environmental quality and public health, warranting proactive intervention from the responsible authorities.

In response to these problems, Portuguese regulatory authorities, namely the inspectors of Inspeção-Geral da Agricultura, do Mar, do Ambiente e do Ordenamento do Território (IGAMAOT), conduct periodic inspections of waste operating companies. The main objec-

tive of these inspections is to detect illicit practices or infringements, to promote transparency and regulatory compliance across the Portuguese waste management network.

For an efficient allocation of the limited resources available to the authorities, they perform a risk assessment analysis on the waste operating companies, to define their inspection priorities, based on impact factors such as size of the companies, level of hazardnousness of the waste, or previous infringement history. However, this analysis is done manually and does not necessarily capture the companies that are most likely to be engaged in illicit activities. Therefore, the authorities seek a systematic, accurate and efficient approach to monitor and analyze the extensive volume of waste transport records, and detect companies with suspicious behavior, that are consequently at a higher probability of having committed regulatory infringements, in order to increase the precision of the inspections.

For the development of such an approach, the Portuguese Environmental Agency (Agência Portuguesa do Ambiente – APA) has provided access to the 2022 waste transportation records from the Portuguese waste management network. These records, consisting of electronic documents (e-GARs) mandatory for each instance of waste transport within the national territory, contain rich information, from the transfer dates, the identification details of both receivers and senders, and the specific type and quantity of waste being transported.

With this data, we intend to develop an anomaly detection framework that is able to detect deviations from normal patterns of activity that are indicative of illicit activities. In the specific context of waste management, the relational dynamics between senders and receivers can reveal critical patterns, as illicit activities are often a result of organized schemes or collusion between companies and/or individuals. Furthermore, these entities will engage in these activities occasionally, rather than as part of their normal routine, as an attempt to avoid detection, or due to the nature of these opportunities.

Given these specificities, traditional methodologies for fraud detection may fail to detect these events. Aggregation of data over time, for instance, often smooths out short-term fluctuations or irregularities that indicate suspicious behaviour. On the other hand, conventional tabular or non-network models neglect the crucial structural interactions between entities within the network.

Therefore, in this work, we intend to develop a dedicated anomaly detection approach for the waste management sector, that is capable of discerning and capturing the complex temporal and structural patterns within the waste management network data. The structural/network dimension will allow to uncover unusual patterns of connectivity between the entities, suggestive of suspicious activities, while the temporal dimension will identify significant deviations from an entity's established normal activity patterns. The intended outcome is to empower regulatory authorities with an efficient analytical tool that detects companies engaged in suspicious practices to be flagged for inspection, ultimately increasing the precision of the inspections and promoting transparency and accountability in the Portuguese waste management sector.

4.2 Data Understanding

The dataset used in this study comprises a vast collection of anonymized electronic Waste Transport Guide (e-GAR) records, documenting waste transportation activities conducted in Portugal in the year of 2022. Initially containing over 4.5 million transfer records, the dataset was subjected to data preprocessing, which will be described further in the section 4.3,

resulting in a set of more than 3.5 million transfer records. This dataset includes over 111 000 unique waste senders and 1 591 unique waste receivers. Each transfer record corresponds to a document generated within the e-GAR platform, a system developed by the Portuguese Agency for the Environment. The e-GAR document is mandatory for every waste transfer, and it is primarily filled by the waste sender, though in certain instances this responsibility may be transferred to the waste transporter or receiver. In either case, each document must always be validated by the counterparty to ensure accountability.

The data embedded within these documents is highly granular, containing rich information for data analysis, and supporting the identification of irregular or suspicious patterns in waste transports.

To inform the subsequent Data Preparation phase, we performed an Exploratory Data Analysis on the raw e-GAR dataset, providing insights into its structure, content, and the characteristics of its variables.

4.2.1 Exploratory Data Analysis

The original dataset comprised 39 features and 4 555 887 rows. Of these features, 24 were categorical, identifying aspects such as the producer and receiver of waste (by name and VAT ID), the type of operation, and specific waste categories. The remaining features included four datetime variables, two numerical variables, and a row identifier.

Before presenting the analyses of the features, we performed some data preprocessing to remove irrelevant instances. Firstly, 184 496 rows with an invalid state ('Nullified', 'Rejected', 'Deleted' or 'Emitted') were removed from the dataset, as these records either contained incorrect data or represented transfers that did not materialize. The remaining states ('Concluded', 'Accepted', and 'Correction Denied') were retained. Although the dataset was primarily limited to documents created in 2022, some transport and state date variables extended into early 2023, reflecting transfers scheduled for the following year. Transfers not taking place in 2022 (4 620 instances) were removed, restricting the analysis to waste transfers occurring within 2022. Additionally, operations related to civil construction, identified by the produtor_origem code 'OBRAS_RCD', were excluded based on expert guidance, as they were deemed irrelevant to the study's objectives.

No duplicate rows were found, and no zero values were present in any of the columns. Missing values were observed in only seven features, with three being justifiably sparse for specific use cases, three serving as descriptors, and one being irrelevant for the analysis. A detailed table of missing values can be found in the appendix A.1.

Categorical Variables

Among the categorical variables, the identifiers of waste producers and waste receivers are particularly important, as they will be the agents for the network construction. The dataset contains 1 759 unique waste receivers, of which 1 591 are companies. These companies are responsible for receiving approximately 99% of all waste transfers and constitute the primary entities for our anomaly detection framework.

Conversely, there are 113 739 unique waste producers. Companies account for 57.5% of these producers, contributing to 92% of waste transfers, while individual persons comprise 41.8% and are responsible for 6.8% of transfers. Other entity types represent only about

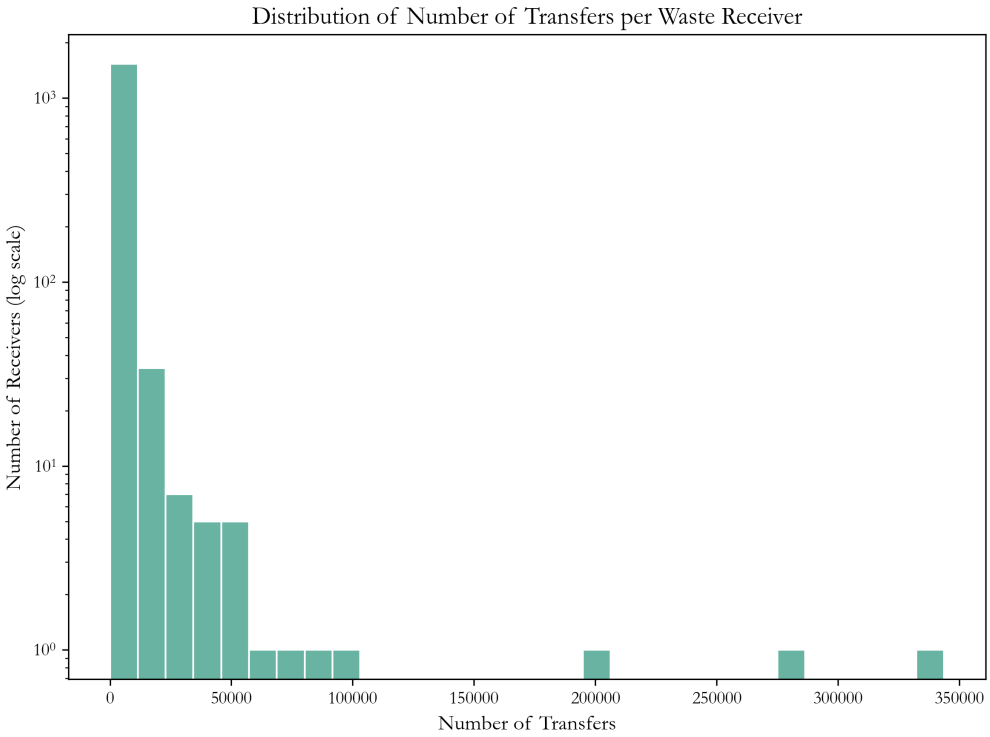


Figure 4.1: Distribution of the total number of transfers by waste receiver.

1% of transfers and unique entities. Plots displaying these distributions can be found in the appendix A.2

Focusing on the 1 591 waste-receiving companies, the targets of our anomaly detection, it is relevant to note that they present widely disparate scales of operation. While the mean number of transfers received by companies is 2236.53, the median is significantly lower at 156. This indicates a highly skewed distribution, where a large proportion of companies exhibit sparse activity, while a small set of companies handle a disproportionately high volume of transfers, with the maximum reaching 343 779 transfers in 2022. Specifically, less than 25% of companies processed at least 1 000 transfers. This distribution is shown in figure 4.1.

Numerical Variables

The dataset includes two numerical variables: quantidade (quantity), typically completed by the document-issuing entity, and quantidade_recebida (quantity received), generally filled by the waste-receiving company.

Both quantidade and quantidade_recebida exhibit highly skewed distributions, with skewness values of 773.69 and 926.42, respectively, and exceptionally high kurtosis values (708 129.4 and 1 120 436, respectively), indicative of heavy tails. This skewness is partially due to the varying operational scales of companies and the diverse nature of waste materials (for example, bulk concrete versus small amounts of toxic waste). However, it also suggests the presence of data entry errors. Quantities in the millions are exceptionally rare and typically feasible only for transport via trains or ships. Inspection revealed that most extreme values were likely data entry errors. For truck transport, which constitutes the majority of transfers, quantities rarely exceed 40 000 kg, with some exceptions such as trucks designed for transport of concrete

and construction materials. Calculating the difference between the two quantity fields further confirmed these errors; only 45.3% of transfers showed a zero quantity difference. While 99% of instances had differences under 1 000 kg, the remaining 1% likely represent errors. Of the 93 transfers exceeding 1 million in one of the quantity fields, only 3 had consistent values across both features, with 90 exhibiting differences of 1 million, unequivocally indicating data entry errors. For transfers exceeding 50 000 kg (an accommodating threshold for truck capacity), 46% displayed quantity differences exceeding 10 000 kg, further supporting the presence of errors. Consequently, the observed skewness in quantity variables is partly driven by these data inaccuracies, which are addressed in the data preprocessing phase.

Date Variables

Among the four datetime features available (emission date of the document, start of transport, and end of transport), the date corresponding to the beginning of the transport was selected as the primary time reference for feature creation. This choice was made because it most accurately reflects the real-world operational activity of the companies.

The distribution of transfers by the beginning of transport date was analyzed at daily, weekly, and monthly granularities, and is shown in figure 4.2. These plots reveal a strong weekly seasonality, which is caused by a consistent drop in activity during weekends, which is exhibited in Appendix A.3. Both daily and monthly plots also indicate slight reductions in activity in mid-August and around the beginning and end of the year, which align with typical holiday periods in Portugal. However, this holiday-related drop is largely attenuated when viewed at the monthly level.

4.3 Data Preparation

The Data Preparation phase consisted of two parts, the Data Cleaning and Feature Creation. To guarantee the quality of the data for our analysis, we firstly perform data cleaning on the original e-GAR data. After that, we develop the features that will be entered into our models. To fully understand the features and ensure them and the models are mutually adequate, we also perform an additional exploratory data analysis and stability analysis.

4.3.1 Data Cleaning

In the data cleaning stage, several data preprocessing steps were taken to ensure the quality and suitability of the data, such as removal of irrelevant data, and treatment of errors in the numerical variables.

Removal of Data

Firstly, all e-GAR documents with the states 'Nullified', 'Rejected', or 'Deleted' were removed. This ensures that only records corresponding to real waste transfers are included in our analysis. Subsequently, residual transfers from 2021 and 2023 were excluded, restricting the analysis to waste transfers occurring strictly within 2022. Additionally, operations related to civil construction, identified by the 'produtor_origem' code 'OBRAS_RCD', were removed based on guidance from regulatory inspectors who deemed these records irrelevant to the study's objectives.

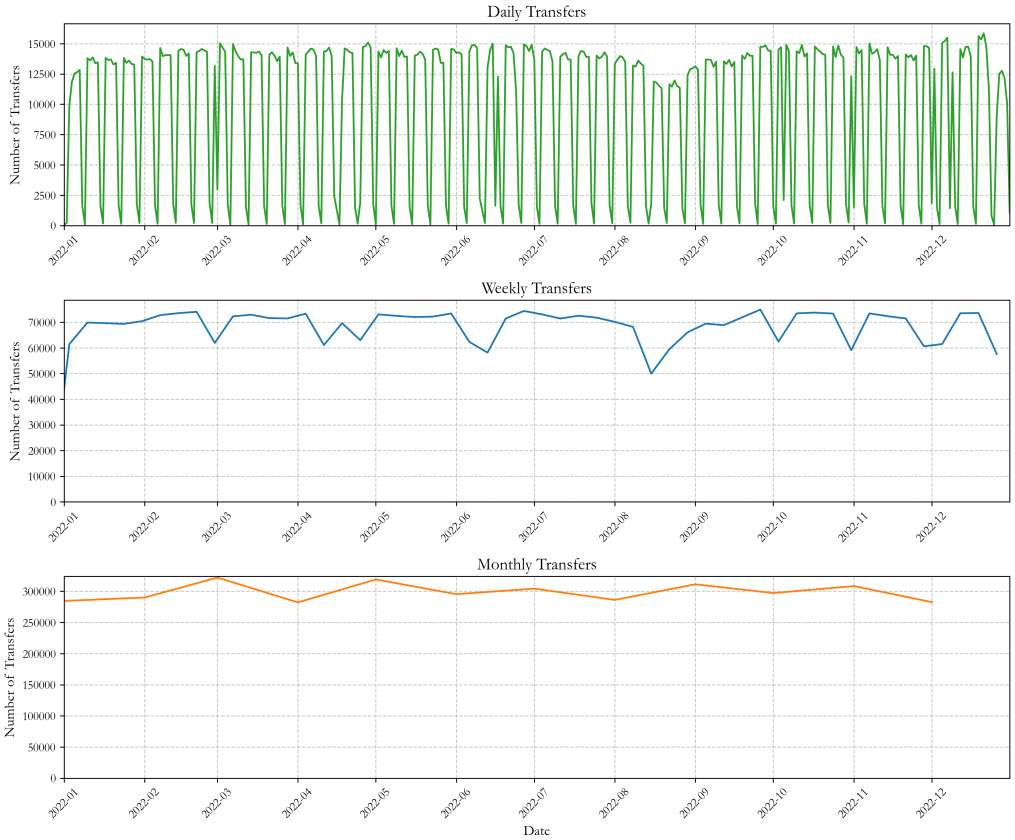


Figure 4.2: Daily, weekly and monthly number of transfers.

Error Handling

Regarding the quantity aspects of the data, two features were available: quantidade, filled by the entity emitting the document, and quantidade_recebida, filled by the company receiving the waste. There were significant inconsistencies between these features, described in the section 4.2.1. Since quantidade_recebida is generally completed by waste operator companies, which have greater experience and accuracy in this process, this value was adopted as the primary reference for waste quantity. To handle the inconsistencies identified in the section that suggested data entry errors, we took the following approach: in cases where quantidade_recebida was higher than 50 000 kg and the quantidade field was not, the value from quantidade was used instead. This decision was based on regulatory inspectors' guidance, indicating that such high values are exceptional and likely represent errors if not corroborated by both features.

Outlier Treatment

The dataset exhibits a large number of outliers and a highly skewed distribution, as presented in the exploratory data analysis (Section 4.2.1). This is inherent to the nature of the data, which describes a network of waste management companies with naturally varying scales of operation. No explicit outlier treatment was performed at this stage, as the chosen anomaly detection methods are designed to handle different magnitudes through internal scaling mechanisms. Furthermore, company-specific outliers are often crucial signals for detecting suspi-

cious changes in activity and were therefore intentionally retained.

After these cleaning steps, the resulting dataset contained over 3.5 million transfer records, with more than 111 000 unique senders and 1 591 unique waste receivers. Using this cleaned data, a network representing the waste transfers between entities was constructed. The nodes in the network correspond to the senders and receivers of waste (individuals, waste recycling companies, or other organizations). Directed edges represent the flow of waste from sender to receiver, attributed with the quantity of waste and the date of transfer. This network structure provided the foundation for the subsequent creation of informative features.

4.3.2 Feature Creation

To perform anomaly detection on the waste operating companies, a set of time series features was constructed from the e-GAR records and its constructed network. These features were designed to capture both the structural characteristics of the network, using Social Network Analysis metrics, and the activity patterns of each company, with the goal of revealing unusual dynamics indicative of illicit activities.

The following features were created for all companies receiving waste in e-GAR, with weekly and monthly periodicity, as justified below:

- **Number of Individual Senders:** This feature represents the count of unique individuals who sent waste to each company in a given time period. While companies can receive waste from other companies, individuals, and other types of organizations such as public institutions, individual senders are frequently involved in irregular or illegal transactions, such as the sale of stolen or improperly disposed valuable waste, or can otherwise represent informal waste management activities. A sudden increase in the number of individual senders may point to potential involvement in these illicit transactions. Individuals were identified by NIF numbers starting in 1, 2, or 3, or ending in 45, as per official classification.

- **Waste Quantity Received:** This feature consists of the total amount of waste, in kilograms, received by each company during a given time period. Monitoring the temporal evolution of this quantity helps identify unusual spikes or drops in a company's activity that could signal suspicious behavior when combined with other features.

- **Waste Quantity Received from Individuals:** This feature consists of the total amount of waste, in kilograms, received by each company during a given time period, specifically originating from individuals. Monitoring the temporal evolution of this quantity further helps identify suspicious patterns related to individual senders.

- **Betweenness Centrality:** A measure of a node's importance in a network, specifically how often it lies on the shortest paths between other nodes. This feature is thus the number of shortest paths between all pairs of nodes that pass through the company at a given time period. Spikes in this feature indicate companies suddenly becoming intermediaries, which may suggest participation in illicit or waste laundering schemes.

- **Ego Network Density:** Ego network density is a measure of how interconnected the neighbors of a focal node are within its immediate network. Formally, it is the ratio of the number of actual edges among the company's neighbors to the maximum

possible number of edges those nodes could share, producing a value between 0 and 1. For this application, the company itself was not included as one of its own neighbors, and whenever the company had no neighbors or only one at a given time period, the ego network density was defined as zero. Spikes in ego network density may indicate participation in collusive activities.

- **Number of Triangles:** This feature consists of the sum of unique network triangles each company participates in at a given time period. Triangles are sets of three nodes in a graph where each node is directly connected to the other two, forming a closed loop. Triangles are highly suggestive of collusive activities and schemes in the network, a notion confirmed by the experience of regulatory inspectors.

- **Connectivity with Individuals in Triangles:** This feature consists of the sum of the number of triangles that the individual waste senders connected to a company participate in, at a given time period. It measures a company's indirect exposure to suspicious behavior in the network, based on the assumption that individual senders involved in many triangles are more likely to be participating in illicit activities. This feature thus highlights companies that may be receiving stolen waste from individuals or cooperating in illicit schemes with them.

Exploratory Data Analysis of the Features

We will now proceed with the exploratory data analysis of the features that have been created, to ensure their suitability and inform the preparation of the models. Figure 4.3 displays the per-company temporal distribution of the features.

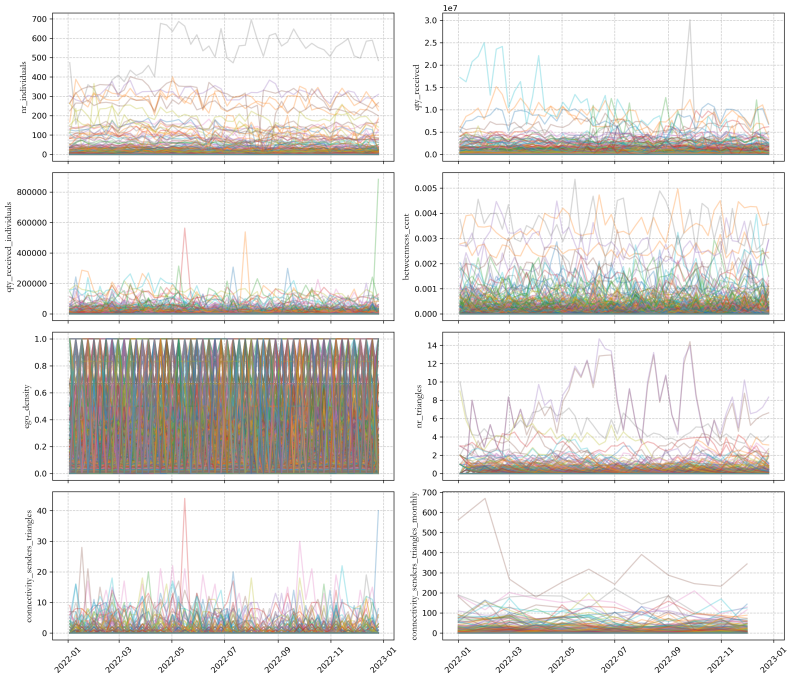


Figure 4.3: Temporal distribution of the features per-company.

While some slight seasonality was observed in August in the aggregated number of transfers, this seasonality does not manifest significantly in the features per company. The quantity

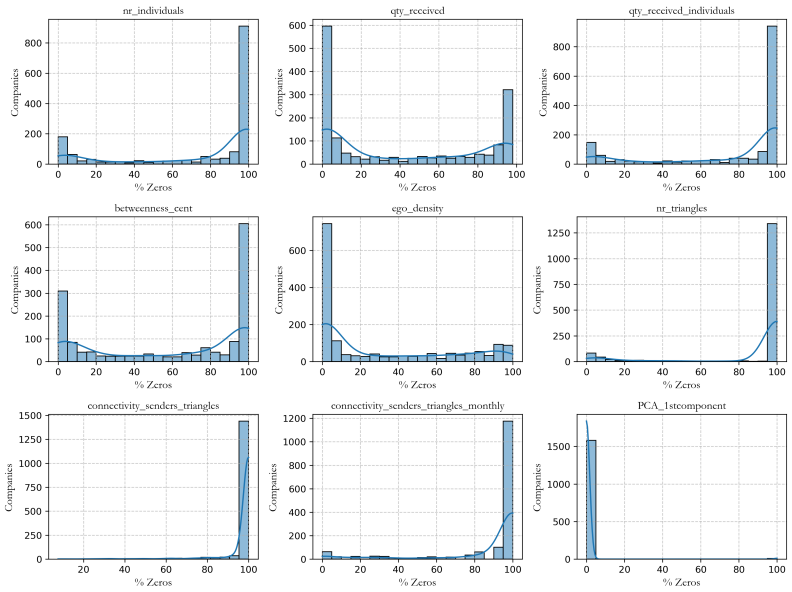


Figure 4.4: Distribution of per-company percentage of zeros in the features.

features, on the other hand, exhibit some severely high values. Manual inspection of corresponding e-GAR instances confirmed these to be true outliers or legitimate transfers, not errors. The ego_density feature is less visually prominent in plots due to many companies exhibiting an oscillating pattern, generally stemming from a low number of neighbors.

Zeros

Figure 4.4 displays the per-company prevalence of zeros for each feature.

All features have a significant percentage of zero values. For instance, the nr_individuals feature has a total zero percentage of 74.61%, and 50.41% of companies exhibit 100% zero values for this feature. Similarly, qty_received_individuals has 76.63% total zeros and 51.79% of companies with all zero values. Betweenness centrality and ego density also show substantial zero percentages (59.84% and 29.87% respectively). The triangle-related features (nr_triangles at 87.83% and connectivity_senders_triangles at 97.21%) have the highest percentages of zeros, which is expected as connection to few or no triangles is indicative of normal activity. The high percentage of zeros in features related to individuals is also justified by the fact that some companies do not receive waste from individuals in their normal course of business.

Correlation Analysis

Figure 4.5 displays the correlation between the features.

As would be expected, the features qty_received, qty_received_individuals, nr_individuals and connectivity_senders_in_triangles generally present higher levels of correlation among each other. Particularly, qty_received_individuals, and nr_individuals present the highest level of correlation (0.83). betweenness_centrality shows the lowest levels of correlation overall. All features demonstrate positive correlation with each other, as they are all jointly correlated to the general activity level of the companies.

Stability Analysis

To support the viability of our proposed approach, a stability analysis was conducted to check if the data exhibits significant oscillations detectable by our change detection methods, or that might be challenging for the LSTM-VAE model to reconstruct. This was performed using the coefficient of variation (CV) for both horizontal (across time for each company) and vertical (across companies for each time period) stability checks. The coefficient of variation is defined as $\text{CV} = \frac{\sigma}{\mu} \times 100\%$, where σ is the standard deviation and μ is the mean.

In the horizontal stability check, we analyzed the stability of values for each company over time. As is visible in figure 4.6, the CV's were overall high for most features. More than half of the companies had a CV greater than 1 for most features, with the exception of ego_density, where only 32.73% of companies had CV's higher than 1. For instance, the mean CV for nr_individuals was 2.08, qty_received was 2.31, and connectivity_senders_triangles was 4.19, indicating substantial temporal variability within individual companies.

In the vertical stability check, we analyzed the stability for each time period across all companies (Figure 4.7). The CV levels remained high, typically ranging from 4 to 10, except for ego_density, which remained around 1 across all periods. The lower CV for ego_density is due to this variable being bounded between 0 and 1. The high CVs for other variables are largely attributable to the differing scales of operation among companies, indicating that the features are vertically unstable.

Overall, the features have demonstrated high instability, both horizontally and vertically, with CVs consistently above 1 dominating the distributions.

Principal Component Analysis

Given that the change detection tests, by definition, can process only one feature at a time, it was beneficial to employ a dimensionality reduction technique to combine several features into a single composite one. This composite feature aims to preserve, combine, and relate the

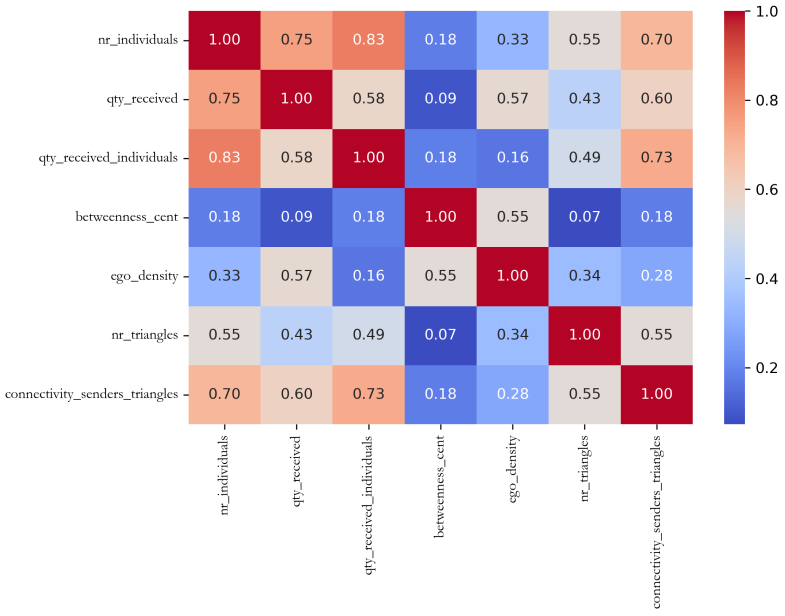


Figure 4.5: Correlation between the features.

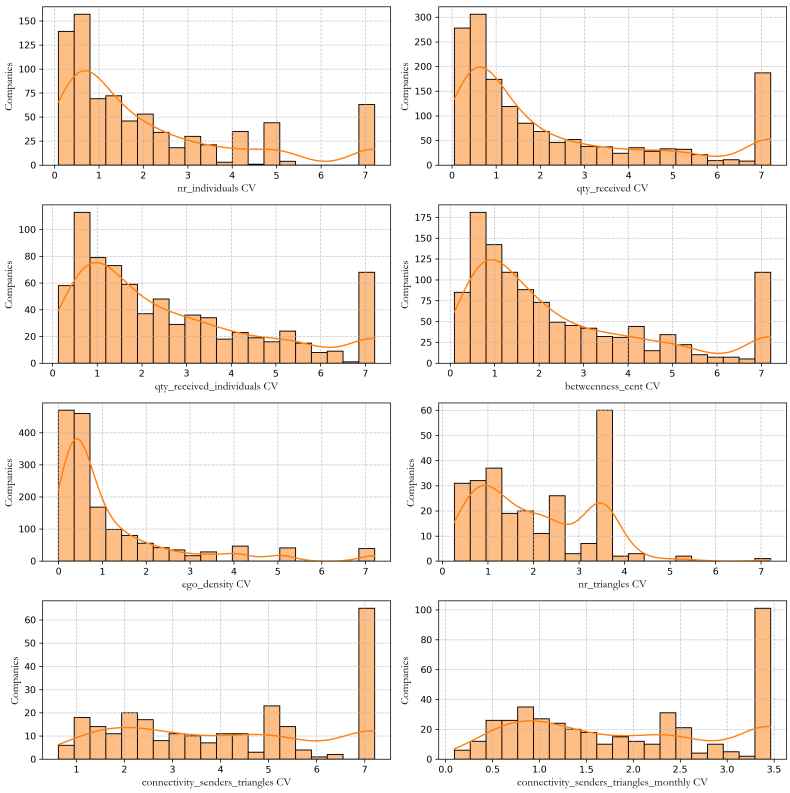


Figure 4.6: Distribution of the per-company coefficients of variations (CV) for each feature.

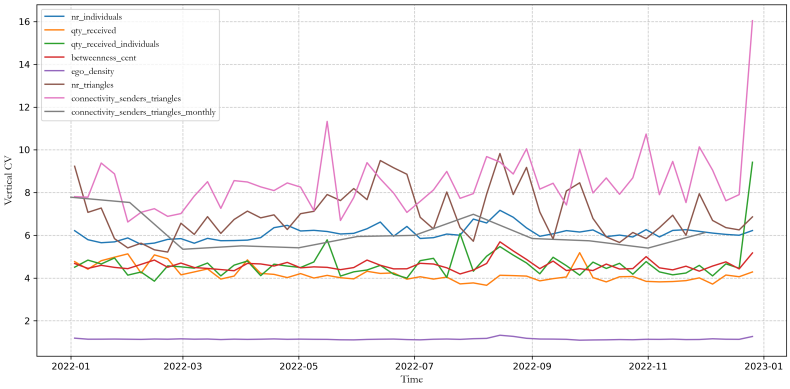


Figure 4.7: Weekly coefficient of variation (CV) for each feature.

information from multiple individual features. Incremental PCA was identified as the most appropriate technique for this purpose, as it allows to obtain a single feature that captures the majority of the variance across the original set of features, while the incremental adaptation is able to process large datasets efficiently.

The features were firstly standardized for each company using its mean and standard deviation, as they possessed substantially different scales. Subsequently, Incremental PCA was performed. The first principal component, which captures the dominant direction of variance, is then used as a new feature. This approach helps reduce dimensionality and highlights composite activity patterns that may not be evident from any single feature alone.

We obtained the following PCA-based feature:

- **First Principal Component of Activity Features:** The features selected for this component were betweenness centrality, ego network density, number of triangles, number of individual senders, and quantity received from individuals. These features were chosen because they widely cover different aspects of the companies' network activity. Consequently, their joint variance can potentially manifest suspicious patterns in their behavior.

4.4 Modeling

At this phase, we will combine two change detection techniques, the Page-Hinkley test and the Cumulative Sum (CUSUM) test, with a deep learning approach, the LSTM-VAE. With this framework, we intend to detect suspicious deviations or unusual complex patterns in the temporal evolution and behavior of the waste operating companies, by applying these approaches on a set of meaningful features representing their operational and network behavior, as described in the Section 4.3.2.

4.4.1 Page-Hinkley and Cumulative Sum tests

The Page-Hinkley and Cumulative Sum tests were applied to the following features:

- Connectivity with senders in triangles (monthly periodicity)
- First principal component of PCA of weekly betweenness centrality, ego network density, number of triangles, number of individual senders, and quantity received from individuals (weekly periodicity)
- Ego network density (weekly periodicity)

Monthly periodicity was used for connectivity with senders in triangles as it allowed for uncovering more patterns (triangles formed over a month) compared to weekly periodicity, which yielded fewer detected triangles.

Before applying the tests, a minimum activity filter was implemented. This filter excludes companies with extremely sparse data from being subjected to the change detection test. Specifically, companies with less than 20% of non-zero values for a given feature were excluded. In the case of the PCA features, companies with less than 20% of non-zero values in at least 3 of the 5 constituent features were excluded. This is because the tests do not function correctly in companies with extremely sparse activity, as any non-zero value might be incorrectly detected as a spike. The exception to this filter was for the feature of connectivity with senders in triangles, as any spike in this feature is inherently interesting and suspicious. This filter results in the tests being biased towards companies with significant activity, which aligns with the inspectors' interest in flagging larger companies for inspection.

To allow for comparison between companies with vastly different scales, the features were standardized prior to applying the test. A description of the functioning of these tests can be found in Section 2.

Specifically for the CUSUM test, the target mean was set as the rolling mean of the time series, with windows of 16 weeks for weekly data and 4 months for monthly data. While in the PH test the target is typically the mean of observed values up to the current point, in CUSUM

it is typically set as the expected mean of the process. However, as the data is not necessarily stationary for an individual company, using a rolling mean was deemed more appropriate.

After the tests were applied to the previosuly standardised features, each company obtained an anomaly score for the corresponding test and feature combination. For both the Page-Hinkley and the Cumulative Sum tests, the company's score was defined as the sum of its 3 highest registered detected change strengths. This metric was preferred over the sum of all detected changes because, while it is rare for a company to have more than 3 changes detected, when it does occur, using the sum of all changes would disproportionately inflate the score. The presence of 3 changes already signals a strong, reoccurring pattern within the year, making additional changes less incrementally meaningful for an anomaly score.

Parameter Tuning

For tuning the parameters of the Page-Hinkley and CUSUM tests, a visual inspection of the detected changes in company-specific graphs was performed. This ensured that the test was neither overly sensitive nor insensitive, preventing misdetection of changes. The following figures demonstrate this inspection.

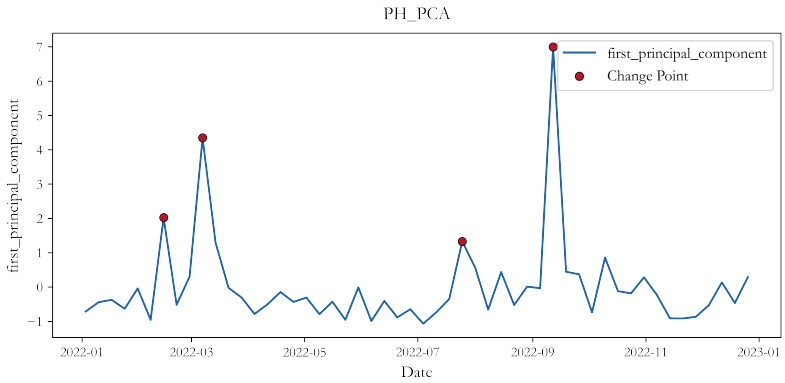


Figure 4.8: Example of Page-Hinkley test with $\delta = 1$, threshold=2: Demonstrates high sensitivity, detecting minor fluctuations.

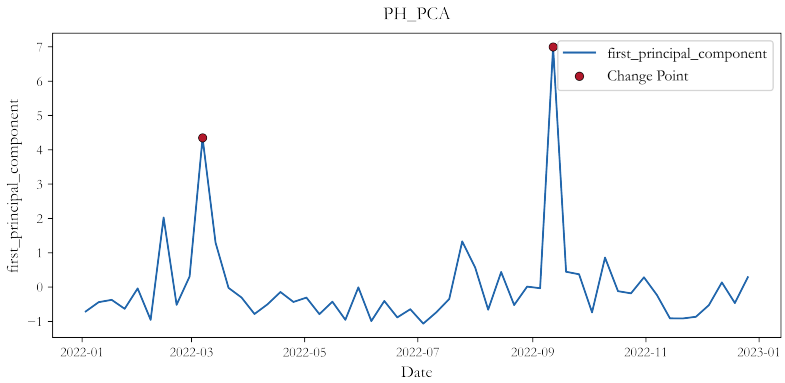


Figure 4.9: Example of Page-Hinkley test with $\delta = 2$, threshold=3: Demonstrates a balanced setting, detecting significant shifts.

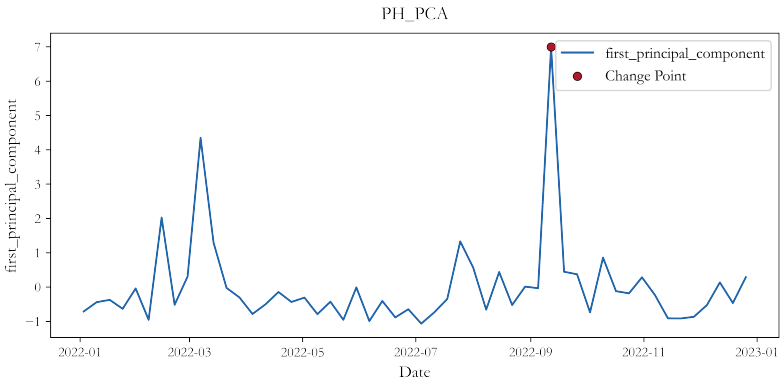


Figure 4.10: Example of Page-Hinkley test with $\delta = 4$, threshold=5: Demonstrates low sensitivity, missing important changes.

4.4.2 LSTM-VAE

Th Long Short-Term Memory Variational Autoencoder (LSTM-VAE) model, for which a detailed description can be found in Section 2.4, was applied on two sets of features:

- **Weekly betweenness centrality and quantity received (LSTM-VAE-2F):** Specifically, unusually high patterns in betweenness centrality relative to waste quantities received indicate that a company is acting as an intermediary within the network while transferring inconsistent volumes of waste, suggesting that the company may be facilitating waste laundering or other collusive activities.

- **Weekly connectivity with senders in triangles, betweenness centrality, and quantity received from individuals (LSTM-VAE-3F):** This set of features covers different network aspects from the companies, in order to jointly manifest unusual patterns in the general network and operational behavior of the companies, through the reconstruction errors.

The model was implemented using the Keras library in Python. For model training, sliding windows were constructed over each company's feature time series. Prior to training, each feature was scaled using a standard scaler fitted on non-zero values, except for features where zeros can be consistent with a pattern of normal activity, in which case the zeros were included in the scaling process. The companies labeled as anomalous were also excluded from the training set as they do not represent normal activity patterns.

In our application, the number of hidden units in the LSTM layers was fixed at 32. Both the input dropout and recurrent dropout were set to 0, meaning no units were randomly dropped during training. The models were trained for 50 epochs with a batch size of 32, using the Adam optimizer with its default settings (learning rate = 0.001, $\beta_1 = 0.9$, $\beta_2 = 0.999$, and $\epsilon = 10^{-7}$).

Considering the stochastic nature of the LSTM-VAE, the model was run 100 times with different random initializations, to obtain a robust final model for each set of features. Then, the anomaly score for each company was defined as the mean of its 10 highest normalized reconstruction errors. The errors were normalized per company to maintain the distances between errors but make them more comparable between companies. The top 10 errors were selected instead of the mean of all errors because the objective is to find spikes in periods

where it was especially difficult for the model to reconstruct the pattern, as the premise of our approach is that suspicious activities happen sporadically rather than continuously. On the other hand, using the mean reconstruction error would dilute the signal from isolated suspicious spikes.

Parameter Tuning

To tune the parameters, a parameter sensitivity analysis was performed, testing different combinations of latent dimension (latent_dim), window size (window_size), and beta (β), with 2 runs for each combination. The same random seeds were used for each parameter setting to ensure comparability across models by maintaining consistent initialization.

Unlike the change detection tests, where parameters could be tuned by visual inspection, this was not feasible for the LSTM-VAE model. Therefore, the set of labeled anomalous companies was used to evaluate performance at each parameter setting.

LSTM-VAE-2F

We tested 15 different parameter settings for the LSTM-VAE-2F model, combining latent dimensions of 4 and 6, window sizes of 6, 8, and 10, and beta values of 0.1, 0.5, and 1. For each setting, the model was run twice.

As shown in figure 4.11, the parameter settings delivering the best results were with latent_dim=4 and window_size=10. This combination yielded the lowest minimum rank and lowest mean rank for the anomalous companies. As for β, values of 0.1 and 1 delivered slightly better results, but $\beta = 0.1$ was deemed most appropriate for our use case. In an anomaly detection setting, the primary objective is not to learn a highly disentangled or smooth latent representation, as the generative capabilities of the model will not be utilized. Instead, the aim is to maximize the model's sensitivity to unusual reconstruction patterns. A lower β reduces the weight of the KL divergence term, allowing the model to prioritize accurate reconstruction over enforcing a tightly constrained latent space. This results in sharper reconstruction errors when anomalies occur, making the top-spike ranking metric more reliable. Therefore, the final setting chosen for this model was latent_dim=4, window_size=10, and $\beta = 0.1$.

LSTM-VAE-3F

We tested 13 different parameter settings for the LSTM-VAE-3F model, combining latent dimensions of 4 and 6, window sizes of 6, 8, and 10, and beta values of 0.1, 0.5, and 1. For

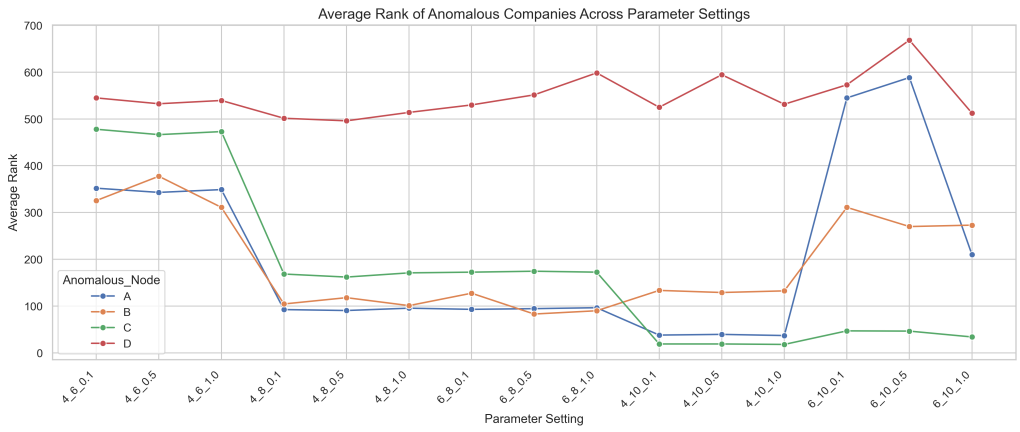


Figure 4.11: Parameter sensitivity analysis results for LSTM-VAE-2F.

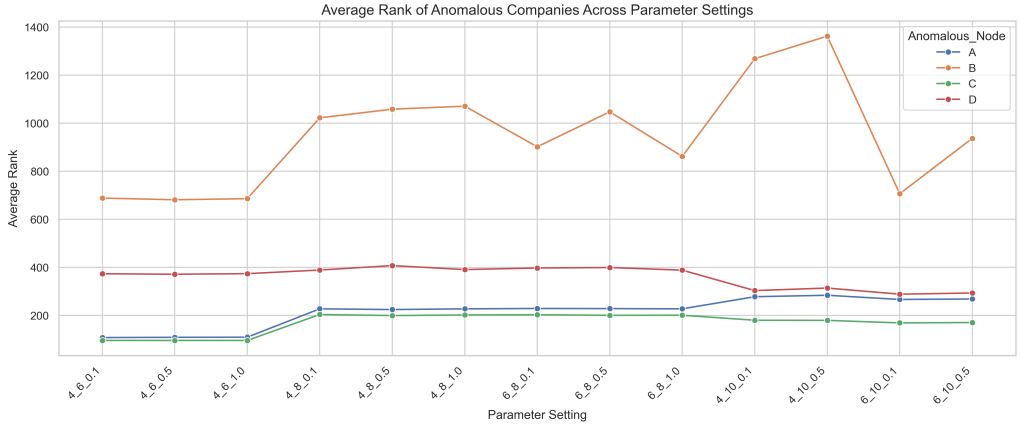


Figure 4.12: Parameter sensitivity analysis results for LSTM-VAE-3F.

each setting, the model was run twice.

The figure 4.12 shows that the parameter settings delivering the best results for this model were with latent_dim=4 and window_size=6. This combination yielded the lowest minimum rank and lowest mean rank. The beta values showed very similar performance across different settings. To maintain consistency with the LSTM-VAE-2F model, $\beta = 0.1$ was selected. Therefore, the final setting chosen for this model was latent_dim=4, window_size=6, and $\beta = 0.1$.

4.4.3 Anomaly Scoring

To consolidate the findings from the various change detection methods and LSTM-VAE models into a singular anomaly score, a ranking-based system was utilized, as it robustly handles the heterogeneity of the score distributions obtained by the different methods.

As previously defined, for the Page-Hinkley and CUSUM tests, the anomaly score for each company was calculated as the sum of its top 3 detected change strengths over time, and for the LSTM-VAE models, the score was defined as the mean of the top 10 normalized reconstruction errors. Companies were subsequently ranked for each method based on the defined score, with higher values indicating stronger evidence of anomalous behavior.

For easier interpretability and consistent comparison across different methods, these ranks were converted into standardized scores using the following formula:

$$\text{Score} = \frac{\text{Total Number of Companies - Rank of Company}}{\text{Total Number of Companies - 1}} \tag{4.1}$$

This standardization yields scores ranging between 0 and 1, where 1 represents the highest rank (most anomalous) and 0 represents the lowest rank (least anomalous).

Finally, the results of a selected set of five methods were combined into an overall anomaly score. This aggregate score was determined by taking the median of the individual standardized scores. Consequently, a company exhibiting a high anomaly score in at least three of the five selected tests would receive a high final anomaly score, indicating consistent suspicious behavior across multiple detection perspectives, but allowing for detection failure in at most two of the methods. Since different companies may exhibit distinct anomalous patterns, some may be flagged by certain methods but not others, highlighting the complementarity of the approaches to detect diverse forms of irregular behavior.

The specific selection criteria for these five methods are elaborated in the Evaluation phase, presented in Section 5.3.

Chapter 5

Experiments and Results

In this chapter, we present the results of the experiments done on the 2022 anonymized e-GAR data for our proposed anomaly detection framework. With these experiments, we aim to evaluate the functioning of the framework and its efficacy in detecting the relevant anomalies. To this end, our experiments were guided by the following research questions:

- Can the integration of statistical change detection methods (Page-Hinkley and CUSUM tests) with deep learning approaches (LSTM-VAE model) effectively identify companies performing illicit activities in the waste management network?
- Do the features subject to our models provide an informative signal for the detection power of our framework?
- What are the limitations of the proposed framework?
- To what extent can the proposed framework support regulatory authorities in the waste management network?

In the following sections, we present the results of the application of the several methods, evaluate the results, analyse examples of high anomaly scoring companies, and finally discuss the findings.

5.1 Page-Hinkley and CUSUM Tests

We now present the results for the Page-Hinkley (PH) and Cumulative Sum (CUSUM) tests, which were applied to various temporal and network-based features. Specifically, we present the following tests:

- PH test on monthly connectivity with senders in triangles (PH-Triangles)
- PH test on the first principal component derived from the PCA of weekly betweenness centrality, ego network density, number of triangles, number of individual senders, and quantity received from individuals (PH-PCA)
- PH test on weekly ego network density (PH-EgoDensity)
- CUSUM test on monthly connectivity with senders in triangles (CUSUM-Triangles)

- CUSUM test on the first principal component derived from the PCA of weekly betweenness centrality, ego network density, number of triangles, number of individual senders, and quantity received from individuals (CUSUM-PCA)

- CUSUM test on weekly ego network density (CUSUM-EgoDensity)

Table 5.1 provides a summary of the results for each of these tests.

The 'Companies excluded' column in Table 5.1 indicates the count or percentage of companies excluded by the Minimum Activity Filter, relative to the full dataset before filtering. It is evident that both Page-Hinkley and CUSUM tests produced comparable results when applied to the same feature sets, with changes being detected in approximately 10% to 29% of the companies.

A notable observation is the higher number of companies removed by the Minimum Activity Filter for the Principal Component Analysis feature. This is attributed to the more strict filtering criteria applied to the PCA-based feature: it required at least three of its five constituent features (betweenness centrality, ego network density, number of triangles, number of individual senders, and quantity received from individuals) to exhibit a minimum of 20% non-zero activity. In contrast, other features only required a 20% non-zero activity within their individual time series. The PCA feature was designed to integrate diverse aspects of company activity and network behavior, aiming to uncover more complex and intricate suspicious patterns, which naturally leads to the exclusion of companies with minimal activity across these combined dimensions.

Figure 5.1 presents the distribution of anomaly scores obtained from the PH and CUSUM tests, where each company's score is quantified as the sum of its three highest change strengths.

In line with expectations for anomaly detection, a significant proportion of companies yielded a score of zero, indicating either the absence of detected changes or their exclusion by the Minimum Activity Filter.

Consequently, the distributions of positive scores are highly positively skewed, with values extending up to 7. These raw scores are subsequently standardized during the anomaly scoring phase to enable consistent comparison across different detection methodologies.

Table 5.1: Summary of Page-Hinkley and CUSUM test results, including companies removed by the Minimum Activity Filter, total changes detected, and number of companies with detected changes.

Method	Companies excluded (%)	Changes Detected (%)	Companies Detected (%)
PH-Triangles	0.00	0.20	9.55
PH-PCA	66.31	2.71	23.95
PH-EgoDensity	16.84	0.50	23.51
CUSUM-Triangles	0.00	0.20	11.13
CUSUM-PCA	66.31	2.61	28.79
CUSUM-EgoDensity	16.84	0.20	23.00

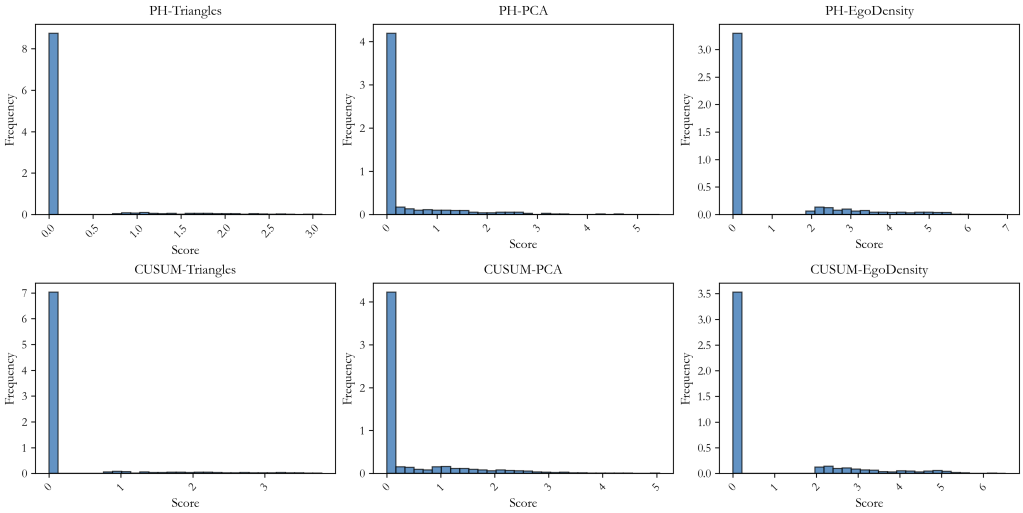


Figure 5.1: Distribution of anomaly scores for Page-Hinkley and CUSUM tests.

5.2 LSTM-VAE

We now move on to present the results obtained from the two Long Short-Term Memory Variational Autoencoder models.

LSTM-VAE-2F

The LSTM-VAE-2F model was trained using weekly betweenness centrality and quantity received as input features. The optimal parameters identified for this model were a latent dimension of 4, a window size of 10, and a beta value of 0.1. To ensure the robustness and reliability of the anomaly scores, the model was executed 100 times, each with a different random seed. The final anomaly score for each company was then computed by averaging the scores across all 100 runs.

Figure 5.2 illustrates the distribution of the final anomaly scores generated by the LSTM-VAE-2F model.

The distribution of anomaly scores exhibits a positive skewness of 0.615. This indicates that while the majority of companies display relatively lower anomaly scores, a smaller subset

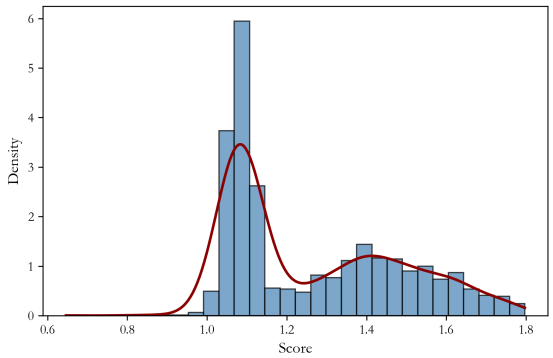


Figure 5.2: Distribution of final anomaly scores for the LSTM-VAE-2F model.

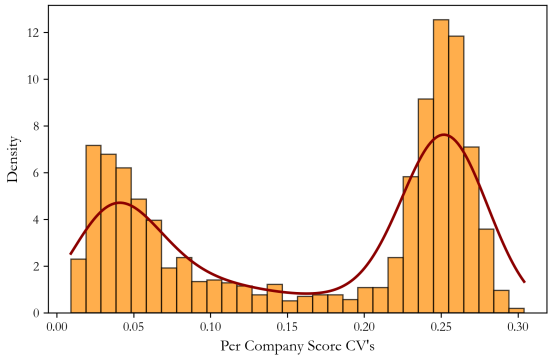


Figure 5.3: Distribution of intra-company score coefficients of variation for the LSTM-VAE-2F model.

presents substantially higher scores. This asymmetric distribution is consistent with the typical characteristics of anomaly detection tasks, where only a minor proportion of entities are expected to deviate significantly from normal behavior.

To evaluate the stability of the anomaly scores, the coefficient of variation (CV) for each company's score across the 100 runs was calculated. Figure 5.3 depicts the distribution of these intra-company score coefficients of variation.

The intra-company score coefficients of variation range from 0.0092 to 0.3040, with a mean of 0.1662. The distribution suggests the presence of two groups: one exhibiting very stable CVs, and another with moderately variable CVs. These low CV values collectively indicate that the anomaly scores generated by the LSTM-VAE-2F model are generally stable and robust across multiple runs, even for the most variable companies, therefore enhancing the reliability of the model.

LSTM-VAE-3F

The LSTM-VAE-3F model was trained on weekly betweenness centrality, connectivity with individual senders in triangles, and quantity received from individuals. The optimal parameters for this model were a latent dimension of 4, a window size of 6, and a beta value of 0.1. Consistent with the LSTM-VAE-2F model, it was run 100 times with varying random seeds, and the final anomaly score was computed as the average across all runs. Figure 5.4 illustrates the distribution of the final anomaly scores from the LSTM-VAE-3F model.

The distribution of anomaly scores exhibits a moderate positive skewness of 0.4738. This implies that while the majority of companies display lower anomaly scores, a smaller subset presents notably higher scores, which is consistent with the expectation that anomalies are rare events.

The stability of the scores was also evaluated by examining the intra-company coefficients of variation across runs. Figure 5.5 displays the distribution of these CVs.

The intra-company score coefficients of variation range from 0.4201 to 0.5269, with a mean of 0.4660. These values indicate that the anomaly scores generated by the LSTM-VAE-3F model exhibit greater variability across runs compared to the LSTM-VAE-2F model. Nevertheless, these fluctuations remain within a reasonable range, supporting the robustness of the model.

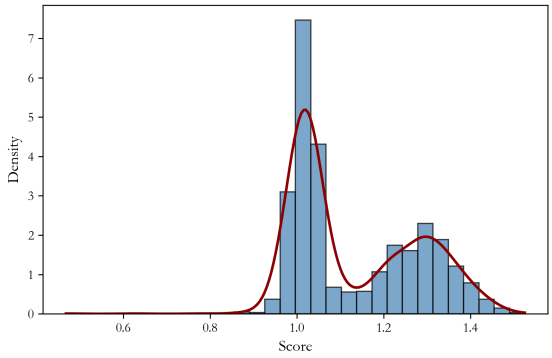


Figure 5.4: Distribution of final anomaly scores for the LSTM-VAE-3F model.

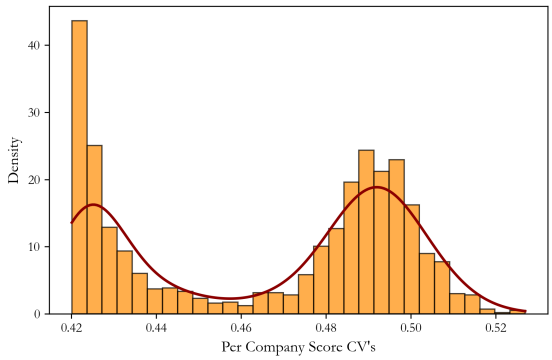


Figure 5.5: Distribution of intra-company score coefficients of variation for the LSTM-VAE-3F model.

5.3 Evaluation

To assess the efficacy of our proposed anomaly detection framework, we evaluated the performance of the individual methods against a small set of four pre-labeled anomalous companies. These companies, identified by regulatory authorities as having engaged in illicit practices, served as a ground-truth benchmark for validation. Figure 5.6 displays the anomaly scores for each company across the different methods.

The heatmap generally indicates that the anomaly scores for the labeled companies were consistently very high across the methods. The change detection techniques, particularly those based on triangle features, consistently identified all anomalous companies with high scores (above 0.9). The LSTM-VAE models, however, showed a limitation by failing to detect Company D. Nevertheless, all four labeled companies consistently showed a very high score in at least six of the employed methods.

The correlation between the scores obtained by the different methods for all companies was also analyzed. Figure 5.7 presents a heatmap illustrating these correlations.

As shown in Figure 5.7, the Page-Hinkley and CUSUM tests, when applied to the same features, exhibited high correlations, ranging from 0.76 to 0.9. Additionally, a moderate correlation was observed between the two LSTM-VAE models (0.47). All methods demonstrated a positive correlation with each other, indicating that they generally align in their identification of anomalous patterns.

Method Selection for Final Anomaly Score

Considering the high similarity and correlation between the Page-Hinkley and CUSUM approaches when applied to the same features, and given that the final anomaly score is derived from the median of selected methods, it was necessary to select a subset of approaches to prevent the change detection methods from disproportionately influencing the final score over the LSTM-VAE models. To address this, for each feature set where both PH and CUSUM tests were applied, the method that yielded slightly less favorable scores for the labeled anomalous companies (as indicated in Figure 5.6) was excluded. Specifically, PH-Triangles, CUSUM-PCA, and CUSUM-EgoDensity were removed, as they generally performed marginally worse than their counterparts in detecting the known anomalous companies. This selection ensures a balanced representation of both statistical change detection and deep learning approaches in the final aggregated score.

Therefore, to compute the final combined anomaly scores, the following five methods were selected:

- Cumulative Sum test on the connectivity with individuals in triangles (CUSUM-Triangles)
- Page-Hinkley test on the first principal component of weekly betweenness centrality, ego network density, number of triangles, number of individual senders, and quantity received from individuals (PH-PCA)
- Page-Hinkley test on the ego network density (PH-EgoDensity)
- LSTM-VAE on the weekly betweenness centrality and quantity received (LSTM-VAE-2F)
- LSTM-VAE on the weekly connectivity with individuals in triangles, quantity received from individuals, and betweenness centrality (LSTM-VAE-3F)

This combination of methods was chosen as it offers complementary perspectives, encompassing sudden shifts and complex behavioral changes in different aspects of the companies

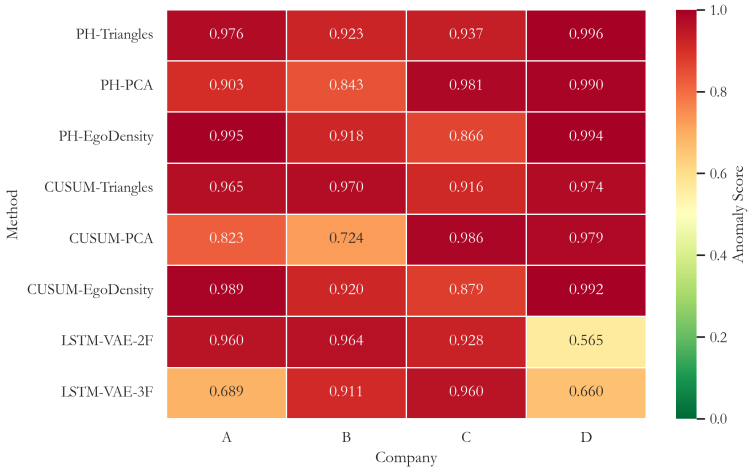


Figure 5.6: Heatmap of anomaly scores for labeled anomalous companies across the detection methods.

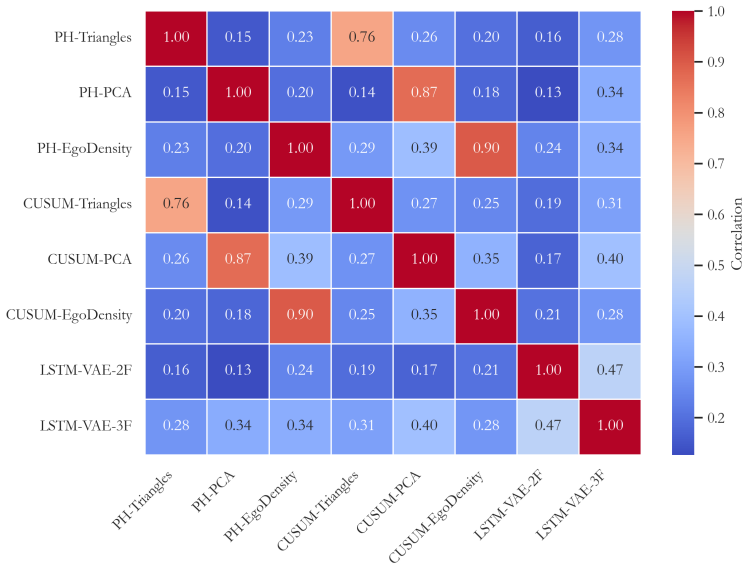


Figure 5.7: Heatmap of correlation between anomaly scores obtained by the detection methods.

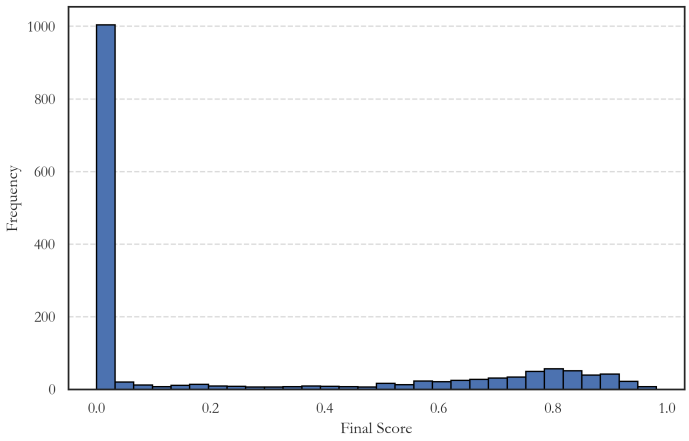


Figure 5.8: Distribution of the final aggregated anomaly scores across all companies.

behavior and network activity, such as their connection and participation in triangles, their role as intermediary in the network, and general temporal patterns of their activity.

Figure 5.8 displays the distribution of the final aggregated anomaly scores, and Table 5.2 provides their summary statistics.

The distribution of the final scores is highly positively skewed, as is expected in an anomaly score. Approximately 61.66% of the companies received a final score of zero. This is primarily a consequence of having no detected changes across all three selected statistical change detection methods, which contribute significantly to the median score. Considering an anomaly threshold of the top 5%, any company with a score of at least 0.878 would be classified as anomalous.

Table 5.3 presents the final anomaly scores and corresponding percentiles for the four

Table 5.2: Summary statistics of the final aggregated anomaly scores.

Statistic	Value
Minimum	0.000
Median	0.000
Mean	0.237
90th percentile	0.819
95th percentile	0.878
Maximum	0.982
Standard deviation	0.345
Skewness	0.938
% of zeros	61.659

Table 5.3: Final anomaly scores and percentiles for the four labeled anomalous companies.

Company	Final Score	Percentile of Final Score
A	0.960	99.81
B	0.918	98.33
C	0.928	98.81
D	0.974	99.94

labeled anomalous companies.

All four labeled anomalous companies achieved very high final anomaly scores. Company D registered the highest score at 0.974, placing it in the 99.94 percentile, indicating it was among the most strongly detected anomalies. The lowest-scoring labeled company was B, with a score of 0.918, corresponding to the 98.33 percentile. This means that all labeled anomalous companies were ranked within the top 2.67% of all companies, thus confirming they were strongly detected by our anomaly detection framework. Despite the severely limited validation dataset, these results suggest that incorporating temporal analysis on structural and network-based signals provides a valuable framework for identifying anomalous activity within the waste management network, thereby supporting regulatory inspections.

5.4 Analysis of Anomalous Companies

To demonstrate the practical application of our framework, we now analyze two examples of companies identified as highly anomalous.

The company presented in Figure 5.9 achieved the highest final anomaly score of 0.982.

This company was strongly flagged by both the change detection methods and the LSTM-VAE-3F model, while the LSTM-VAE-2F model did not identify significant anomalous patterns for this entity. The high score from the CUSUM test on connectivity with individuals in triangles indicates a period of significantly increased connectivity with individual senders participating in network triangles. This suggests potential involvement in informal or illicit waste trades, as individuals involved in such triangles can be indicative of collusive or suspicious activities, as explained in Section 4.3.2. Additionally, a notable spike was observed in

the ego network density, signifying a period of unusually high interconnectedness among the entities this company interacted with directly. This could point to participation in collusive arrangements. The PH-PCA test also identified two distinct spikes, suggesting moments of unusual variance in the company's overall behavior and network activity. The highest score was obtained from the LSTM-VAE-3F, which implies that the company exhibited high reconstruction errors during a specific period, while showing more typical behavior during the rest of the year. This pattern suggests that the company's activity was irregular, characterized by intermittent periods of highly unusual operational or network patterns that deviate significantly from learned normal behavior.

The company shown in Figure 5.10 received a final anomaly score of 0.947.

This company was strongly detected by the PH-PCA and both LSTM-VAE models. However, it had no connectivity with senders in triangles, and the PH-EgoDensity test only detected a moderate spike. The high score from the PH-PCA test indicates a significant spike in the company's overall behavior and network activity, suggesting a sudden deviation from its established patterns. Both LSTM-VAE models, with scores of 0.974 (LSTM-VAE-2F) and 0.952 (LSTM-VAE-3F), detected a period where the company's activity was particularly difficult to reconstruct, signaling an unusual pattern that falls outside the models' learned representations of normal behavior. This implies that the company engaged in activities that were structurally or temporally anomalous.

These examples highlight several critical points. First, the diverse methods employed in our framework offer complementary perspectives, each capable of detecting unusual patterns in different aspects of companies' operational and network activities. This multi-faceted approach enables the detection of diverse types of operational and network anomalies. Second, the observed oscillating patterns of connectivity and reconstruction errors suggest that suspicious activity is not necessarily continuous but may occur in bursts, possibly as a deliberate strategy to evade detection.

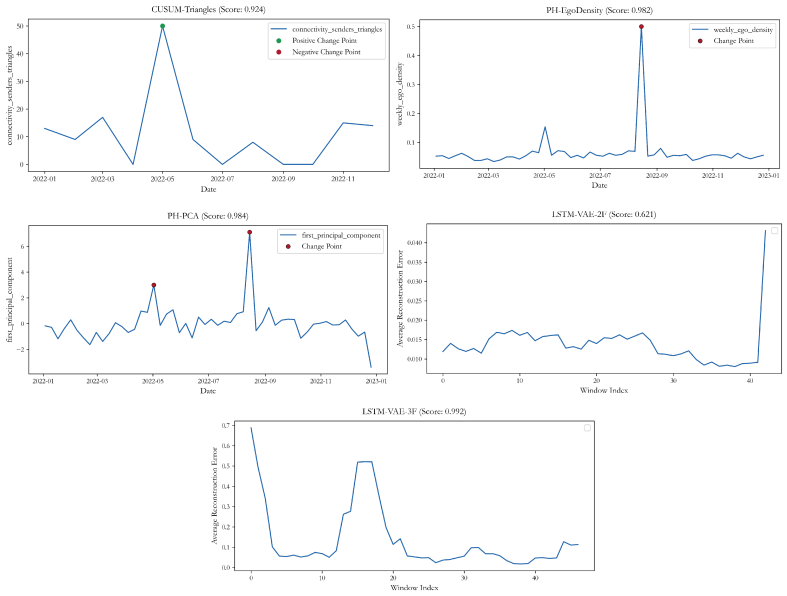


Figure 5.9: Anomaly score profile for the top-scoring company (Score: 0.982).

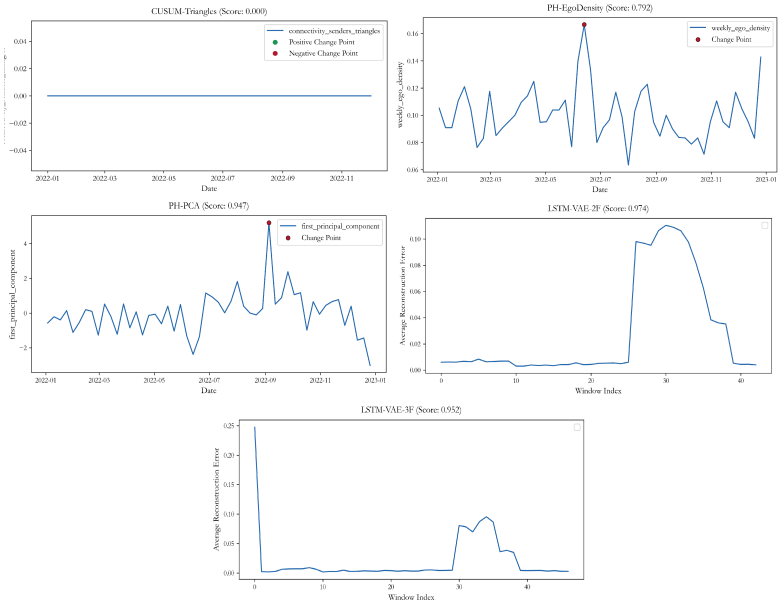


Figure 5.10: Anomaly score profile for a high-scoring company (Score: 0.947).

5.5 Discussion

Considering the experimental results presented in this chapter, we are now ready to discuss the findings, directly addressing the research questions initially posed.

- Can the integration of statistical change detection methods (Page-Hinkley and CUSUM tests) with deep learning approaches (LSTM-VAE model) effectively identify companies performing illicit activities in the waste management network?

Our experimental results demonstrate that combining statistical change detection techniques with a deep learning model provides a robust strategy for identifying anomalous companies in the waste management network. The parallel use of Page-Hinkley, CUSUM tests and LSTM-VAE models, to detect both abrupt temporal changes and more complex behavioral deviations, lead to all labeled anomalous companies being detected by at least three of the five methods employed. Additionally, the anomaly scores obtained by the change detection methods presented low, although positive, correlation (between 0.13 and 0.34) with the deep learning models, meaning that the combination of the methods allows for the detection of a more diverse group of anomalies, highlighting the complementarity of the approaches.

- Do the features subject to our models provide an informative signal for the detection power of our framework?

The features we introduced were informative when applied to the different methods. In these features, all labeled anomalous companies presented significant deviations detected by the Page-Hinkley and CUSUM tests. Connectivity with individuals in triangles provided the clearest signal of anomalous activity, successfully suggesting potential collusive behavior or informal waste trading practices. Changes in ego network density served as another strong indicator, with sharp increases suggesting unusually intense levels of interaction among neighbors that may correspond to coordinated activities. Betweenness centrality, used in the PCA

feature and in the LSTM-VAE models, further signaled companies occasionally positioning themselves as intermediaries. The input of several features representing complex activity patterns, through PCA or in the LSTM-VAE models, complemented the signal of the single-feature methods, by providing a more general anomaly detection approach.

- What are the limitations of the proposed framework?

In terms of performance, the detection methods presented high anomaly scores for the labeled entities overall. However, the LSTM-VAE models failed to identify one of the anomalous companies, which points to limitations in its sensitivity for certain behavioral patterns, and calls for a refinement of the approach. Furthermore, the main limitation of our framework is the scarcity of labeled data. Expanding and diversifying the set of labeled companies, especially including both anomalous and non-anomalous companies, would strengthen the validation of the framework, or further enable the implementation of a supervised framework.

- To what extent can the proposed framework support regulatory authorities in the waste management network?

The success in detecting the set of labeled anomalous companies suggests the feasibility of the approach for flagging companies with suspicious patterns for inspection by the regulatory authorities, but further validation is necessary to ensure its reliability and performance before deployment. Nevertheless, the proposed anomaly detection framework offers a more efficient and data-driven monitoring process for the regulatory authorities to detect suspicious companies and prioritize inspections, optimizing the use of their limited resources and enhancing regulatory compliance and environmental protection.

Chapter 6

Conclusion

In this work, we successfully developed and evaluated a framework for detecting suspicious activities within the waste management network, integrating change detection techniques, namely the Page-Hinkley and CUSUM tests, with a deep learning approach, the LSTM-VAE model. By combining these methods, we aimed to detect both significant changes in the activity of the companies and complex temporal patterns that may indicate fraud or non-compliance.

Our findings demonstrate that, while statistical change detection applied to the "connectivity with senders in triangles" feature yielded the most consistent and effective performance in flagging anomalies, the combination of different detection methodologies provided a comprehensive framework for detecting a wider range of anomalies. Specifically, the combination of five distinct detection methods ensured that all available labeled anomalous companies were detected by at least three of the five techniques. This highlights the value of integrating temporal and network-based analyses for effective detection of suspicious activities in complex, real-world waste management networks. By moving beyond static, aggregated data analysis to dynamic, network-aware change detection, we provide a more sensitive and comprehensive anomaly detection mechanism. The success in identifying known anomalous companies validates the methodology's potential for building a scalable analytical tool to monitor large-scale waste transport records, identify companies with suspicious behaviors and thus support regulatory authorities in prioritizing company inspections in the waste management network.

While promising, this research also encountered certain limitations. The primary limitation was the scarcity of labeled anomalous data, a common challenge in anomaly detection tasks within real-world domains. Although our approach was designed to operate effectively in an unsupervised or weakly supervised setting, a larger, more diverse set of ground-truth entities, including both anomalous and non-anomalous companies, would enhance model validation and generalizability, or further enable the implementation of a supervised framework.

This project is currently still in progress, and the future work will include collaborating with regulatory authorities to obtain a more comprehensive set of labeled data, which would allow for more rigorous validation of the model or the development of a supervised framework. In addition, obtaining and incorporating transport records from 2023 and 2024 will enable more robust modeling of temporal patterns. Finally, we plan to refine the deep learning approach and explore alternative architectures, such as Graph Neural Networks, to further improve detection capabilities.

Bibliography

Akoglu, L., Tong, H., & Koutra, D. (2015). Graph based anomaly detection and description: a survey. *Data mining and knowledge discovery*, *29*(3), 626–688.

Belay, M. A., Blakseth, S. S., Rasheed, A., & Rossi, P. S. (2023). Unsupervised anomaly detection for iot-based multivariate time series: Existing solutions, performance analysis and future directions. *Sensors (Basel, Switzerland)*, *23*. Retrieved from https://api.semanticscholar.org/CorpusID:257387130

Calikus, E., Nowaczyk, S., Sant'Anna, A., & Dikmen, O. (2020). No free lunch but a cheaper supper: A general framework for streaming anomaly detection. *Expert Systems with Applications*, *155*, 113453. Retrieved from https://www.sciencedirect.com/science/article/pii/S0957417420302773 doi: https://doi.org/10.1016/j.eswa.2020.113453

Cano, A., Gómez-Olmedo, M., & Moral, S. (2019). A bayesian approach to abrupt concept drift. *Knowledge-Based Systems*, *185*, 104909. Retrieved from https://www.sciencedirect.com/science/article/pii/S095070511930365X doi: https://doi.org/10.1016/j.knosys.2019.104909

Chen, R.-Q., Shi, G.-H., Zhao, W.-L., & Liang, C.-H. (2019). Sequential vae-lstm for anomaly detection on time series. *arXiv preprint arXiv:1910.03818*.

Chen, X., & Wu, W. (2025). *Bayesian online collective anomaly and change point detection in fine-grained time series.* Retrieved from https://arxiv.org/abs/2508.06385

Davis, N., Raina, G., & Jagannathan, K. (2020). A framework for end-to-end deep learning-based anomaly detection in transportation networks. *Transportation research interdisciplinary perspectives*, *5*, 100112.

Deng, A., & Hooi, B. (2021). Graph neural network-based anomaly detection in multivariate time series. In *Proceedings of the aaai conference on artificial intelligence* (Vol. 35, pp. 4027–4035).

Faithfull, W. J., Rodríguez, J. J., & Kuncheva, L. I. (2019). Combining univariate approaches for ensemble change detection in multivariate data. *Information Fusion*, *45*, 202-214. Retrieved from https://www.sciencedirect.com/science/article/pii/S1566253517301239 doi: https://doi.org/10.1016/j.inffus.2018.02.003

Fathy, Y., Barnaghi, P., & Tafazolli, R. (2019). An online adaptive algorithm for change detection in streaming sensory data. *IEEE Systems Journal*, *13*(3), 2688-2699. doi: 10.1109/JSYST.2018.2876461

Frittoli, L., Carrera, D., & Boracchi, G. (2023). Nonparametric and online change detection in multivariate datastreams using quanttree. *IEEE Transactions on Knowledge and Data Engineering*, *35*(8), 8328-8342. doi: 10.1109/TKDE.2022.3201635

Gama, J. (2012). A survey on learning from data streams: current and future trends. *Progress in Artificial Intelligence*, *1*, 45–55.

Guo, X., Zhou, B., & Skiena, S. (2022, August). Subset node anomaly tracking over large

dynamic graphs. In *Proceedings of the 28th acm sigkdd conference on knowledge discovery and data mining* (p. 475–485). ACM.

Hawkins, D. M. (1980). Multivariate outlier detection. In *Identification of outliers* (pp. 104–114). Springer.

Hewapathirana, I. U. (2019). Change detection in dynamic attributed networks. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, *9*(3), e1286.

Jain, P. K., Bajpai, M., & Pamula, R. (2022). A modified dbscan algorithm for anomaly detection in time-series data with seasonality. *Int. Arab J. Inf. Technol.*, *19*, 23-28. Retrieved from https://api.semanticscholar.org/CorpusID:245464781

Kifer, D., Ben-David, S., & Gehrke, J. (2004). Detecting change in data streams. In *Vldb* (Vol. 4, pp. 180–191).

Kiss, I. Z., Genge, B., Haller, P., & Sebestyen, G. (2014). Data clustering-based anomaly detection in industrial control systems. *2014 IEEE 10th International Conference on Intelligent Computer Communication and Processing (ICCP)*, 275-281. Retrieved from https://api.semanticscholar.org/CorpusID:6803659

Li, N., Liu, X., Liu, Z., Mao, L., Zhao, L., & Wang, X. (2021). Anomaly detection in power grid iot system based on isolated forest. *IEEE/WIC/ACM International Conference on Web Intelligence and Intelligent Agent Technology*. Retrieved from https://api.semanticscholar.org/CorpusID:248087172

Liu, X. (2021). A real-time detection method for abnormal data of internet of things sensors based on mobile edge computing. *Mathematical Problems in Engineering*. Retrieved from https://api.semanticscholar.org/CorpusID:233829519

Lu, J., Liu, A., Dong, F., Gu, F., Gama, J., & Zhang, G. (2019). Learning under concept drift: A review. *IEEE Transactions on Knowledge and Data Engineering*, *31*(12), 2346-2363. doi: 10.1109/TKDE.2018.2876857

Maleki, S., Maleki, S., & Jennings, N. (2021). Unsupervised anomaly detection with lstm autoencoders using statistical data-filtering. *Appl. Soft Comput.*, *108*, 107443. doi: 10.1016/j.asoc.2021.107443

Mejri, N., Lopez-Fuentes, L., Roy, K., Chernakov, P., Ghorbel, E., & Aouada, D. (2024). Unsupervised anomaly detection in time-series: An extensive evaluation and analysis of state-of-the-art methods. *Expert Systems with Applications*, *256*, 124922.

Namoano, B., Emmanouilidis, C., Ruiz-Carcel, C., & Starr, A. G. (2020). Change detection in streaming data analytics: A comparison of bayesian online and martingale approaches. *IFAC-PapersOnLine*, *53*(3), 336–341.

Niu, Z., Yu, K., & Wu, X. (2020). Lstm-based vae-gan for time-series anomaly detection. *Sensors*, *20*(13), 3738.

Nogara Notarianni, M. O., Leveni, F., Stucchi, D., Frittoli, L., & Boracchi, G. (2025). Change detection in multivariate data streams: Online analysis with kernel-quanttree. In *Advanced analytics and learning on temporal data* (p. 115–130). Springer Nature Switzerland.

Oehmcke, S., Zielinski, O., & Kramer, O. (2015). Event detection in marine time series data. In *Deutsche jahrestagung für künstliche intelligenz*. Retrieved from https://api.semanticscholar.org/CorpusID:37482719

Pereira, F. S., Tabassum, S., Gama, J., de Amo, S., & Oliveira, G. M. (2018). Processing evolving social networks for change detection based on centrality measures. In *Learning from data streams in evolving environments: Methods and applications* (pp. 155–176). Springer.

Qahtan, A. A., Alharbi, B., Wang, S., & Zhang, X. (2015). A pca-based change detection

framework for multidimensional data streams: Change detection in multidimensional data streams. In *Proceedings of the 21th acm sigkdd international conference on knowledge discovery and data mining* (pp. 935–944).

Sebastiao, R., & Gama, J. (2009). A study on change detection methods. In *Progress in artificial intelligence, 14th portuguese conference on artificial intelligence, epia* (pp. 12–15).

Shaji, N., Tabassum, S., Ribeiro, R. P., Gama, J., Santana, P., & Garcia, A. (2025). Network-based anomaly detection in waste transportation data. In H. Cherifi, M. Donduran, L. M. Rocha, C. Cherifi, & O. Varol (Eds.), *Complex networks & their applications xiii* (pp. 273–284). Cham: Springer Nature Switzerland.

Tabassum, S., Azad, M. A., & Gama, J. (2020). Profiling high leverage points for detecting anomalous users in telecom data networks. *Annals of Telecommunications*, *75*(9), 573–581.

Wang, H., & Zhang, H. (2025). An anomaly detection method for multivariate time series data based on variational autoencoders and association discrepancy. *Mathematics*, *13*(7), 1209.

Wei, Z., & Wang, F. (2022). Detecting anomaly data for iot sensor networks. *Scientific Programming*. Retrieved from https://api.semanticscholar.org/CorpusId:248972748

Wu, Y., Dai, H.-N., & Tang, H. (2022). Graph neural networks for anomaly detection in industrial internet of things. *IEEE Internet of Things Journal*, *9*(12), 9214-9231. doi: 10.1109/JIOT.2021.3094295

Yairi, T., Kato, Y., & Hori, K. (2001). Fault detection by mining association rules from housekeeping data.. Retrieved from https://api.semanticscholar.org/CorpusID:17400289

Yoon, M., Hooi, B., Shin, K., & Faloutsos, C. (2020). *Fast and accurate anomaly detection in dynamic graphs with a two-pronged approach.*

Zhao, L., Sawlani, S., Srinivasan, A., & Akoglu, L. (2022). Graph anomaly detection with unsupervised gnns. *arXiv preprint arXiv:2210.09535*.

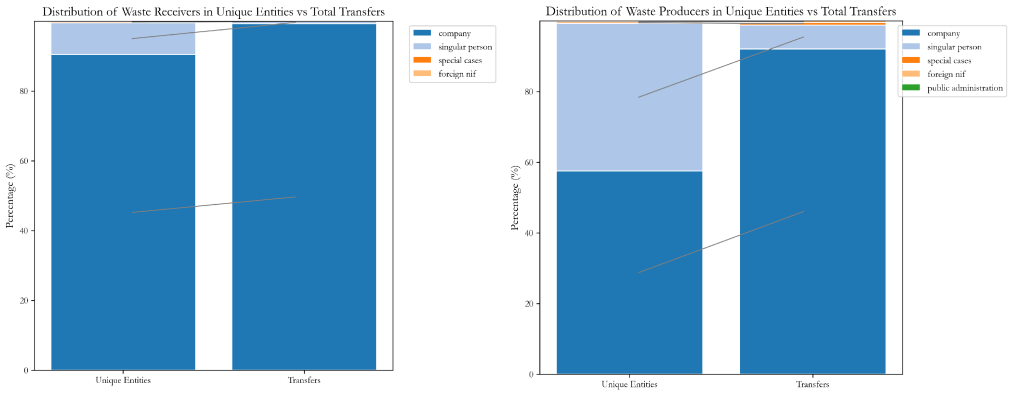
Appendix A

Appendix

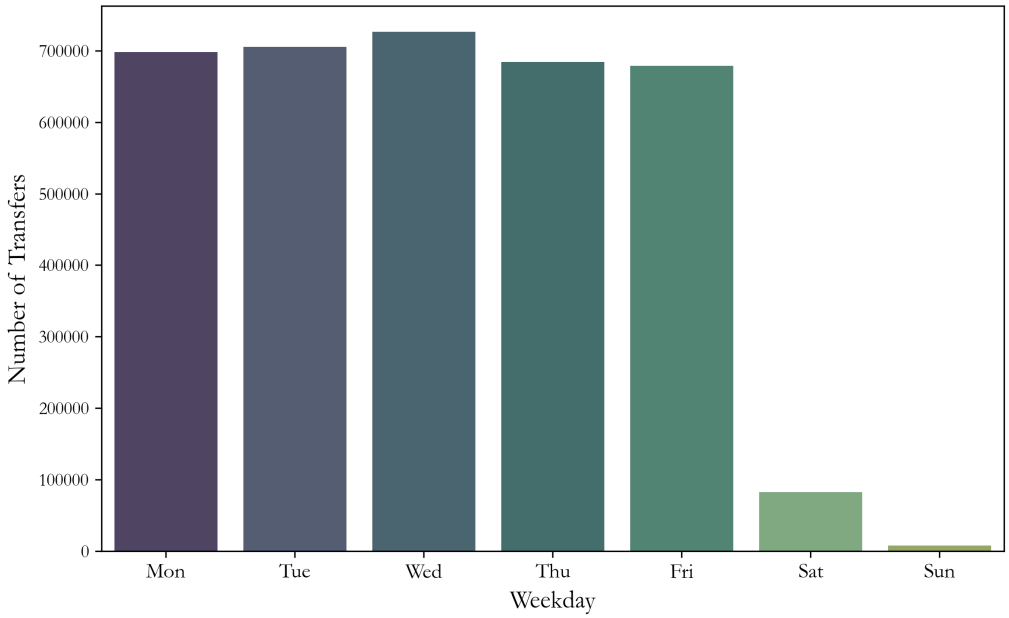
A.1 e-GAR columns with missing values

Variable	Value
num_plano_gestao_lamas_recebido	99.936897
num_plano_gestao_lamas	99.936897
grupo_hospitalar	84.306174
produtor_local	25.980237
produtor_nome	0.729121
dados_transportadores	0.002878
qualidade	0.000576

A.2 Distribution of type of entity by number of unique senders/receivers and number of transfers



A.3 Total number of transfers by weekday.


Number of Transfers
700000
600000
500000
400000
300000
200000
100000
0
Mon
Tue
Wed
Thu
Fri
Sat
Sun
Weekday