

# A network approach to detect Value Added Tax fraud

Angelos Alexopoulos<sup>1</sup>, Petros Dellaportas<sup>2,3</sup> , Stanley Gyoshev<sup>4</sup>, Christos Kotsogiannis<sup>5,6</sup> , Sofia C. Olhede<sup>7</sup> and Trifon Pavkov<sup>8</sup>

<sup>1</sup>Department of Economics, Athens University of Economics and Business, Athens, Greece

<sup>2</sup>Department of Statistical Science, University College London, London, UK

<sup>3</sup>Department of Statistics, Athens University of Economics and Business, Athens, Greece

<sup>4</sup>Department of Finance, University of Exeter Business School, Exeter, UK

<sup>5</sup>Tax Administration Research Centre (TARC), University of Exeter, Exeter, UK

<sup>6</sup>CESifo, Munich, Germany

<sup>7</sup>Institute of Mathematics, Ecole Polytechnique Federale de Lausanne, Lausanne, Switzerland

<sup>8</sup>Faculty of Economics and Business Administration, Sofia University 'St Kliment Ohridski', Sofia, Bulgaria

*Address for correspondence:* Angelos Alexopoulos, Department of Economics, Athens University of Economics and Business, Patision 74-76, Athens 10434, Greece. Email: [angelos@auerb.gr](mailto:angelos@auerb.gr)

## Abstract

Value Added Tax (VAT) fraud erodes public revenue and puts legitimate businesses at a disadvantaged position thereby exacerbating inequality. This article develops scalable algorithms to detect fraudulent transactions by leveraging the rich information embedded in the complex, high-dimensional VAT network structure. Supervised methods are not always suitable for VAT fraud detection, as issues in the auditing process—such as selection bias and audit quality—can seriously affect the labelling of businesses as fraudsters or not. Therefore, both supervised and unsupervised techniques in which VAT fraud detection is implemented through a suitably constructed Laplacian matrix informed by business-specific covariates. The developed methods are applied to the universe of Bulgarian VAT data and detect around 50% of the VAT fraud, outperforming well-known techniques that ignore the information provided by the transactional network structure. The proposed methods are automated and can be implemented following taxpayers' submission of their VAT returns, thus allowing the authorities to prevent large revenue losses.

**Keywords:** anomaly detection, big data, heterogeneous data sources, information systems, tax evasion

## 1 Introduction

The collection and analysis of network data play a key role in a wide range of fields. Examples include, but are not limited to, applications in biology, computer science, sociology, and economics (Kolaczyk & Csárdi, 2014; Newman, 2012). A particularly important question, which network data and techniques can address more efficiently than traditional approaches, is the identification of anomalies in large and complex systems such as credit card and business-to-business (B2B) transactions, health insurance claims, computer security, and biological or genetic data sets; see, for example, Akoglu et al. (2015) for a survey. In fact, anomaly detection methods that utilize data network structures are very useful in cases when supervised classification is infeasible or inappropriate.

---

Received: September 11, 2024. Revised: July 18, 2025. Accepted: October 31, 2025

© The Royal Statistical Society 2025.

This is an Open Access article distributed under the terms of the Creative Commons Attribution-NonCommercial License (<https://creativecommons.org/licenses/by-nc/4.0/>), which permits non-commercial re-use, distribution, and reproduction in any medium, provided the original work is properly cited. For commercial re-use, please contact [journals.permissions@oup.com](mailto:journals.permissions@oup.com)

Network anomaly detection is typically a big data problem, and its complex structure requires the use of advanced data analytical methods. The objective of this article is to develop fraud detection algorithms for Value Added Tax (VAT), a tax base which constitutes a major source of revenue for over 165 countries, but also one that suffers from significant fraud. VAT is a consumption tax in the sense that the VAT collected throughout the supply chain is ultimately paid by the final consumer when the good is consumed. At the core of VAT lies an ‘invoice-credit’ mechanism, whereby the net tax liability of a business is calculated by subtracting from the VAT on sales the aggregate VAT paid on invoices for inputs used in production. This mechanism requires sellers along the production chain (B2B transactions) to provide invoices to their buyers showing the amount of VAT that was paid on each transaction. Any fractional revenue collection on the value added that is generated at each stage of the production chain must be remitted to the revenue authority. The B2B transactions and the VAT invoice-credit mechanism together create a *network* through which businesses interact within and across production sectors and along the supply chain. Throughout, the terms trader, business, taxpayer are used interchangeably. In order to claim VAT credits, businesses must be registered for VAT with the revenue authority.

Despite its widespread adoption as a major tax innovation, the VAT system is widely acknowledged—by both policymakers and scholars—to suffer from inherent weaknesses and vulnerabilities (Ebrill et al., 2001; Keen & Lockwood, 2010; Keen & Smith, 2006). A central vulnerability of the VAT system lies in its invoice-credit mechanism, which, while fundamental to its design, creates systematic opportunities for fraud and abuse. This structural weakness has become a major policy concern across numerous jurisdictions, including the European Union (EU) which in a Communication in 2016 recognized that ‘[t]he current VAT system, which was intended to be a transitional system, is fragmented, complex for the growing number of businesses operating cross-border and leaves the door open to fraud...’, p. 3, COM(2016). Combating VAT fraud was also designated as a strategic priority by the European Union for the 2018–2021 period, as part of its broader efforts to combat organized crime.<sup>1</sup> Among the measures implemented was the establishment of the European Public Prosecutor’s Office (EPPO), which began operations in 2021. The EPPO is tasked with ensuring the criminal law protection of the EU’s financial interests, including those threatened by cross-border VAT fraud.

There is a growing recognition that effectively combating VAT fraud requires tax administrations to match the sophistication of fraudsters. This entails both the design of more efficient tax structures—supported by improved technological capabilities and a deeper understanding of VAT evasion and enforcement dynamics (see Ainsworth & Madzharova, 2012; Shah, 2021; Waseem, 2023)—and the strategic use of data analytics for the detection (and prevention) of non-compliance. While a considerable body of research has focused on the former, work in the latter remains limited, particularly in exploiting the rich informational content embedded in the *network* structure of B2B transactions. This article helps to close the gap by developing flexible and scalable machine learning algorithms tailored to fraud detection in VAT networks. Detecting fraud in VAT networks presents significant challenges due to the high dimensionality and heterogeneity of observed B2B transaction data. Supervised learning methods are often ill-suited to this context, as labelling businesses as fraudulent is inherently problematic: audit-based classifications performed by tax authorities are subject to selection bias, as well as unintentional (or even intentional) mis-identification of under-reported tax liabilities. Furthermore, as discussed in more detail below, VAT fraud is often a coordinated effort involving networks of firms—some genuinely legitimate, others only appearing so—that engage in complex transactional schemes. Consequently, identifying a single fraudulent entity is rarely sufficient for effective detection or meaningful prevention.

To address these challenges, this article develops scalable algorithms that analyse the community structure of observed B2B transaction networks, drawing on recent advances in network analysis (see, for example, Binkiewicz et al., 2017; Chaudhuri et al., 2012). In addition, the approach incorporates established machine learning techniques to enhance the analysis by integrating

<sup>1</sup> VAT fraud, in addition to distorting market competition, leads to significant compliance costs for legitimate traders who are required to exercise due diligence in ensuring the legitimacy of their suppliers. There is also the risk of even face bankruptcy as a result of fraudulent actions committed by others. Consumers are of course not insulated from VAT fraud either, as trading outside the formal supply chain might result in higher VAT gap (defined as the difference between what the government could collect and what it actual collects in revenues). VAT fraud exacerbates this difference and meeting the revenue target might necessitate a VAT rate increase to compensate for lost tax revenues.

business-specific characteristics that are typically informative in identifying VAT fraud. The resulting framework enables both the detection of latent communities within the network and the estimation of fraud probabilities for each VAT-registered business. Notably, the proposed methods can be implemented in both supervised and unsupervised settings: when reliable labels are available, the algorithms can learn patterns associated with known fraud cases; when such labels are absent or unreliable, the same methods can operate in an unsupervised manner to uncover potentially fraudulent structures.

The developed fraud methodologies is tested on the universe of the Bulgarian administrative data, which include output and input VAT, sales transactions across all businesses and sectors, and detailed businesses characteristics for the years 2016 and 2017. Importantly, the methods developed have broader applicability, and can be applied to any fraud detection problem where network information is available. To summarize, this article addresses the pressing and persistent challenge of VAT fraud by developing scalable and data-driven detection tools. It does so by,

- developing scalable network anomaly detection methods that can be applied both in a supervised and unsupervised manner to detect fraud in observed VAT networks. Importantly, the use of unsupervised methods is often inevitable in tax fraud detection applications. This is because: (i) the availability of labelled data is limited, as tax authorities (given capacity constraints) cannot perform more than a few audits per year and (ii) even for audited cases in which no fraudulent activity has been detected, there might be uncertainty regarding their label, since fraud may have gone undetected during the auditing procedure.
- The proposed methods integrate transactional network data with business-specific characteristics to classify firms as potentially fraudulent and to identify clusters of firms likely involved in VAT fraud schemes.
- The algorithms are automated and can be implemented upon receipt of purchase and sales declarations, requiring minimal additional investment. This offers significant benefits to tax authorities, including lower administrative costs, greater transparency, and enhanced reproducibility. The empirical application confirms substantial gains in fraud detection at a fixed false positive rate.
- Applied to real-world data, the network-based methods demonstrate superior performance in identifying anomalies compared to traditional classification models that rely solely on firm-level attributes.

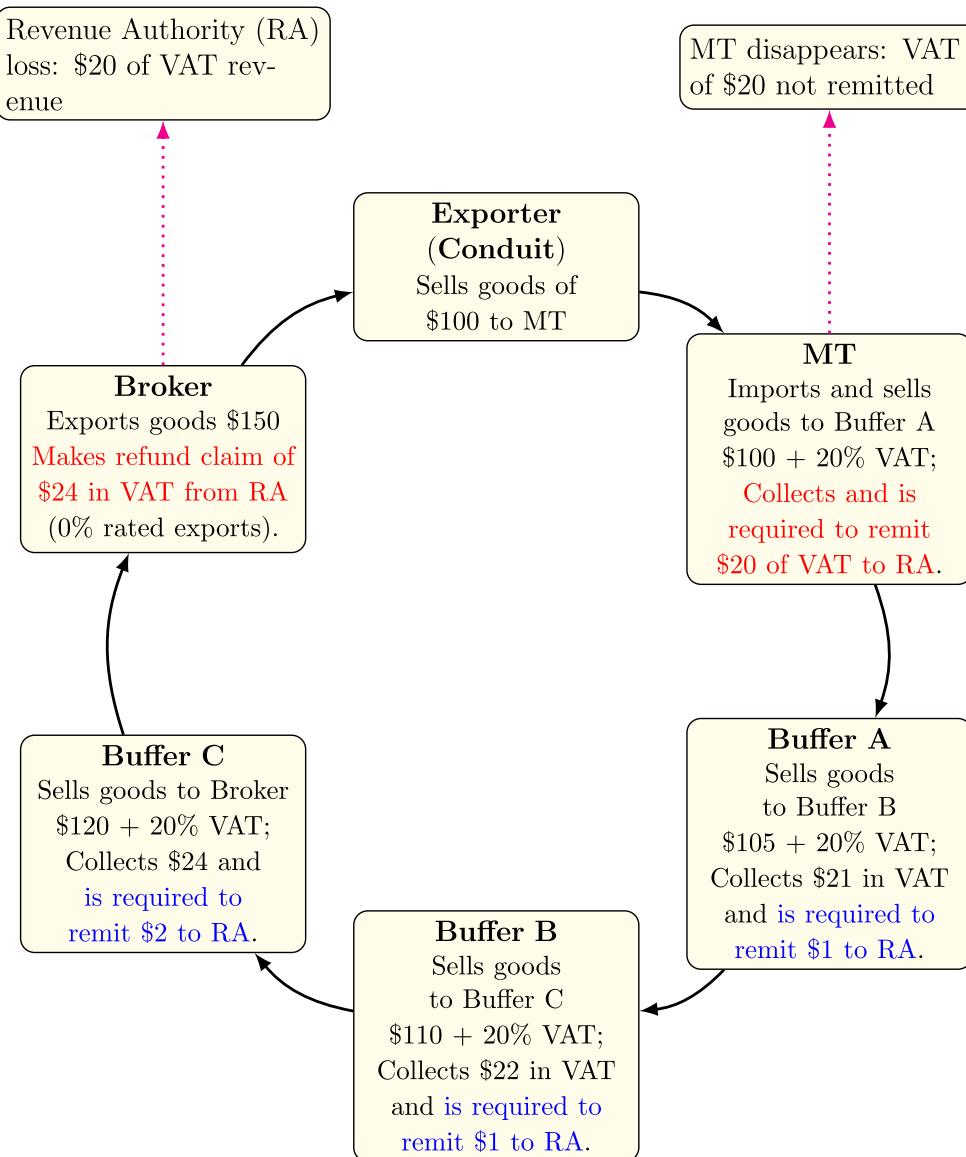
The article is structured as follows. Section 2 provides a general overview of the forms and mechanisms of VAT fraud, highlighting the inherent complexity and diversity of fraudulent schemes that motivate the modelling and methodological contributions that follow. Section 3 reviews the relevant literature on VAT fraud detection. Section 4 introduces the proposed fraud detection methodology, while Section 5 presents the results of its application to real-world data. Section 6 concludes.

## 2 Fraud in the VAT network

There are many forms of VAT fraud, ranging from fictitious trading of invoices to circular transactions involving fraudulent activities known as ‘carousel’ or, more formally, the ‘missing trader’ (MT) fraud. Two key characteristics of VAT fraud are: (i) it requires the interaction of multiple B2B traders and thus reflects communal behaviour among group of nodes and (ii) not all B2B transactions are real; some are fictitious. Figure 1 illustrates the MT fraud in its simplest form. The scheme has four ‘types’ of firms: The ‘Conduit’ (a trader that partakes in a transaction that is connected with the fraudulent evasion of VAT), the ‘MT’ (a firm that will go missing without remitting to the revenue authority any VAT collected), the ‘Buffer’ (firms that could be part of the fraud fulfilling the role of concealing the identity of the MT) and the ‘Broker’ the firm that has orchestrated the fraud.

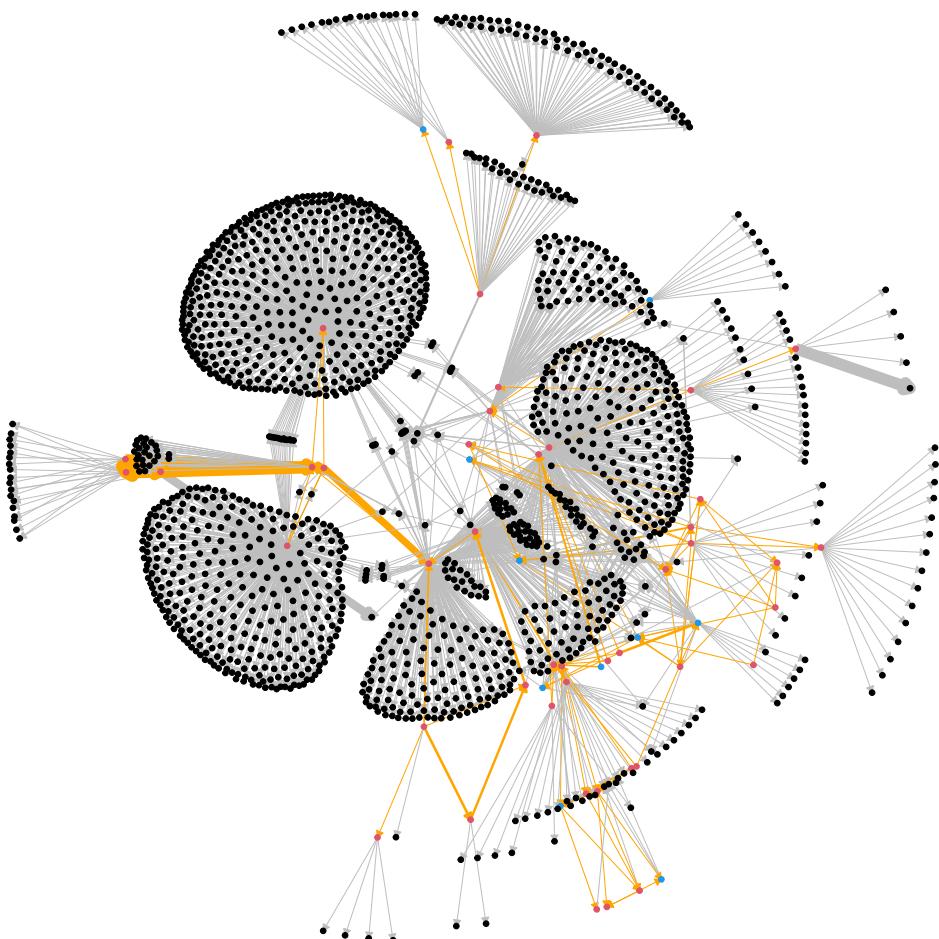
The fraud involves the MT importing goods with an invoiced value of US\$100.<sup>2</sup> MT then sells these goods to Buffer A, charging US\$20 in VAT. Since MT has paid no VAT on the acquired

<sup>2</sup> The monetary values used are illustrative. In actual VAT fraud schemes, the transaction values often run into the millions of US dollars. While all transactions formally comply with VAT law—that is, they meet the documentary and legal requirements—they may be purely fictitious, involving no physical movement of goods or services.



**Figure 1.** Illustration of the simplest form of Missing Trader (MT) VAT fraud. The scheme involves a Broker seeking a refund for input VAT that was never actually remitted to the tax authority, because the MT—having collected VAT from Buffer A—has subsequently disappeared from the market. The fraud exploits the structure of the invoice-credit and refund mechanism, as well as the timing mismatch between the entitlement to input VAT refunds and the actual remittance of output VAT further upstream in the transaction chain. Transactions follow the black arrows and may be repeated in a carousel structure, enabling sustained fraud until discovery.

goods, the full amount of US\$20 collected should be remitted to the revenue authority. However, MT disappears without making this payment. Buffer A, having paid US\$100 plus US\$20 in VAT to MT, resells the goods to Buffer B for US\$105, charging US\$21 in VAT. It then offsets the US\$20 it paid as input VAT against the US\$21 collected, and remits the difference—US\$1—to the revenue authority. This process, based on the ‘invoice-credit’ mechanism, continues through Buffers B and C, with each subsequent trader reclaiming input VAT and remitting only the net amount. In the final stage of the transaction chain, the Broker purchases the goods from Buffer C, paying a 20% VAT on their value—amounting to US\$24. The Broker then re-exports the goods to the



**Figure 2.** Directed weighted network of Bulgarian B2B VAT amounts (output and input VAT). Each node corresponds to a VAT-registered business and the width of the edges represents the amounts of VAT associated to sales transactions and in the direction of the edge. The network depicts VAT amounts between businesses that have been identified as Missing Traders and Buffers/Brokers (lightly shaded nodes) in VAT missing trader fraud with legitimate businesses (black nodes). The edges in orange highlight transactions between VAT fraudsters.

Conduit firm. As exports are zero-rated under VAT rules, the Broker is entitled to claim a refund for the input VAT paid, even though the corresponding output VAT has never been remitted to the revenue authority. This discrepancy arises because the MT, who originally charged VAT to Buffer A, has since disappeared without remitting the US\$20 collected to the government. One can imagine this process continuing in a ‘carousel’ fashion and with the goods being re-exported and re-imported with refund claims being accumulated until the fraud is discovered. The invoice-refund mechanism is a structural element of VAT that has been eloquently described as VAT’s Achilles heel (Keen & Smith, 2006).<sup>3</sup>

While the preceding example is deliberately simplified for illustrative purposes, it captures the core mechanics of VAT fraud. In reality, however, such schemes tend to be far more complex, often involving dozens or even hundreds of firms operating across multiple sectors and jurisdictions and engaging in sophisticated transactional arrangements specifically designed to obscure fraudulent behaviour and evade detection. VAT fraud is therefore best understood as a

<sup>3</sup> This fraud is not of course unique to the European Union but it is also of relevance to countries where fiscal checks at the physical borders have been relaxed following trade agreements.

coordinated, network-based phenomenon rather than the isolated action of a single firm. This degree of complexity is evident in [Figure 2](#), which presents the structure of an actual transaction network in which MT fraud has been identified by the Bulgarian National Revenue Agency (BNRA). Nodes represent VAT-registered businesses, while directed edges indicate sales relationships between them. The width of each edge reflects the volume of VAT involved, visually conveying the scale in terms of transactions of the fraud.

The information provided by businesses to BNRA is very rich and includes comprehensive data from VAT returns and VAT ledgers covering all purchases and sales transactions, including intra-community within the EU. The displayed network consists of 1,697 nodes and represents a small part of a much larger network comprising transactions among over 300,000 businesses in Bulgaria. Within this sub-network there are 32 missing traders and 22 brokers; both MTs and Brokers are considered as fraudsters in an MT fraud scheme, transacting with the remaining 1,642 businesses identified as legitimate. [Figure 2](#) clearly indicates a pattern in the transactions made by the VAT fraudsters: a few of them transact with a large number of legitimate businesses (indicated by the black nodes) and make no transactions with other fraudsters (indicated by the lack of connection across red and blue nodes) whereas the majority of the fraudsters have no (or limited) transactions with legitimate businesses. This observation motivates the study of business interactions within a given network to extract important information on potential fraudulent behaviour. These interactions are weighted and directed, and are methodologically incorporated into the analysis in addition to any other information on node-specific (estimated) covariates.

The objective of this article is to develop methodologies that identify communities of taxpayers whose transaction patterns resemble those of traders participating in VAT fraud, and to estimate the probability that each taxpayer participates in a VAT fraud scheme. The proposed methods incorporate characteristic (i) discussed above and are robust with respect to characteristic (ii). As such, they offer an advantage over existing approaches that ignore VAT network interactions and rely only on local characteristics—such as nodes degree, strength, and/or the number of triangles—as will be discussed later on and in Section 3. The methodological approach integrates the universe of observed B2B transactions into a network framework, where each node represents a VAT-registered business, and an edge between nodes  $i$  and  $j$  exists if the corresponding businesses have conducted at least one transaction recorded in their latest VAT returns. This network-based representation is incorporated into fraud detection techniques that rely on scalable analytics, leveraging both the connectivity structure of network and node-specific information. The information obtained from the network of transactions is projected into suitably constructed low-dimensional vectors that preserve the key network properties, and these properties are then utilized within machine learning methods to identify aberrant edges, nodes and sub-networks. Although VAT fraud takes many forms—with the MT scheme discussed earlier being among the most prevalent and consequential in terms of revenue loss—these schemes tend to share key structural features. Most notably, regardless of the specific type, the transactional patterns involved are typically anomalous. It is this anomaly that the proposed detection methods are designed to identify. The algorithms developed here are flexible enough to uncover a wide range of VAT fraud schemes embedded within observed transaction networks.

As previously noted, the methods can be implemented in both supervised and unsupervised modes. In the supervised setting, the algorithm relies on historical information about fraud cases—such as prior classifications of businesses as fraudulent—to guide learning and identify similar patterns within the network. For instance, if tax authorities provide a binary vector indicating which businesses were previously identified as MT fraudsters, the method will estimate fraud probabilities for each taxpayer in the dataset and classify them accordingly. Moreover, the algorithm produces a clustering of the observed population, enabling the identification of groups of taxpayers likely to be involved in MT fraud, including both primary fraudsters and their potential collaborators.

Since VAT fraud schemes often share common characteristics, applying the developed methods to detect fraudsters in MT fraud does not preclude the identification of other schemes, such as circular virtual transactions within a country (another form of fraud). This flexibility also characterizes the proposed fraud detection techniques when applied in an unsupervised manner. Rather than being restricted to a specific type of VAT fraud, these methods analyse taxpayer interactions

through transactions and, in combination with their specific attributes, enable the classification and clustering of taxpayers with a high likelihood of participating in VAT fraud.

The remainder of this section formally defines, for expositional completeness, the network of transactions associated with VAT-registered businesses, and then reviews existing work on anomaly and fraud detection, placing our contribution within the broader literature.

## 2.1 Network modelling

The B2B transactions underlying the VAT system naturally give rise to a network  $G$ , which can be represented by an adjacency matrix  $\mathbf{A}$ , constructed from the (weighted) edge set  $E$  of interactions between nodes (taxpayers)  $V$  with weights  $\{w_{ij}\}$ . The elements of the adjacency matrix  $\mathbf{A} = (a_{ij})$  are defined as

$$a_{ij}(G) = \begin{cases} w_{ij} & \text{if } (i, j) \in E(G) \quad i, j \in \{1, \dots, N\} \\ 0 & \text{otherwise.} \end{cases} \quad (1)$$

$V$  is the set of the *vertices* or *nodes* of the network, and  $|V| = N \in \mathbb{N}$  is the total number of nodes.  $E$  denotes the edge set of directed, weighted edges such that of the network that is,  $E = \{(i, j) \in V \times V : a_{ij} > 0\}$ . The pair  $G(\mathbf{A}) = (V, E)$  denotes the weighted network that corresponds to the adjacency matrix  $\mathbf{A}$ . If  $\mathbf{A}$  is symmetric, and so  $\mathbf{A} = \mathbf{A}^\top$ , then  $G(\mathbf{A})$  is called undirected whereas if  $\mathbf{A}$  is not symmetric then  $G(\mathbf{A})$  is called directed (the Appendix presents networks of real VAT transactions based on the data used in the algorithms).

## 2.2 Our contribution

VAT fraud is characterized by two key features. First, it gives rise to B2B transactions that, in the absence of fraud, would likely not occur, creating anomalous patterns in the transaction network. Second, it typically involves coordinated behaviour among multiple entities, as executing such schemes requires interaction among several VAT-registered traders. While the intensity and structure of coordination may vary across fraud types, these shared behavioural patterns are central to the detection strategies proposed in this article. The methodology is designed to capture both anomalous transactional activity and community-level behaviour indicative of fraud.

VAT-registered businesses generate substantial volumes of transactional and firm-level information. This places additional demands on any fraud detection method: it must be not only sensitive and robust in identifying fraudulent behaviour, but also computationally efficient and scalable to real-world datasets. Addressing this challenge of high-dimensional data is a central focus of this article (an issue that is taken up in Section 4).

To identify communities whose members are likely to be involved of VAT-fraud, the proposed approach constructs a corrected version of the Laplacian matrix. This correction incorporates information from both the node-specific structure and the interaction patterns across businesses, reflecting the fact that the anatomy of VAT fraud combines individual characteristics with community-level dynamics. Two algorithms are developed, both of which can be applied without labelled data in an unsupervised manner, and their output is the classification of the businesses in the dataset in distinct clusters. Because the clustering procedures incorporate firm-specific features indicative of VAT fraudulent behaviour, they are designed to produce a small number of large clusters composed primarily of legitimate businesses, along with a smaller number of clusters that are more likely to contain fraudulent traders. When labelled data are available, the same methodology can be extended to estimate the probability that each business in the network is engaged in VAT fraud.

For both algorithms, the key point of departure is to map the observed network on to low-dimensional Euclidean vector space so it preserves the original connectivity structure of their nodes. The spectral analysis of networks is a well-documented technique for classifying the nodes of large networks in distinct clusters; see for example Chung and Graham (1997) and Ng et al. (2001) for more details. Notice also that the spectral clustering methods are closely related to the so-called eigenmap technique appeared in the graph embedding approach as early as in the contribution of Belkin and Niyogi (2003). Here, the spectral clustering approach is extended by considering the eigendecomposition of a *risk-corrected* Laplacian that maps the observed networks, together

with node-specific information related to their risk profile, on to a Euclidean vector space. Thus, in contrast with network analytics methods for fraud detection (Subelj et al., 2011; Van Vlasselaer et al., 2017)—where only the so-called direct network features (for example, node degrees and number of triangles) are employed to discover fraudulent activity—the approach follows recent advances which rely on graph representation learning (Gao et al., 2021; Pourhabibi et al., 2020) to study the interactions recorded by the observed VAT networks to develop scalable machine learning algorithms with the aim to classify businesses as fraudulent or not.

### 3 Related literature

The possibility of automating the detection of VAT fraud is part of a larger current international research theme seeking to utilize large scale data sets to improve tax (and social) policy (Athey, 2017; Baesens et al., 2003; De Bacco et al., 2023; D. Lazer et al., 2009; D. M. Lazer et al., 2020) as well as to provide a better understanding of interactions (Barons & Shenvi, 2024; Fritz et al., 2023; Jackson & Wolinsky, 1996; Margetts & Dorobantu, 2019; Pillinger et al., 2024). Early contributions to the problem of tax evasion detection, such as VAT fraud, have been made by Gupta and Nagadevara (2007) and Basta et al. (2009), where traditional statistical methods such as logistic regression and discriminant function analysis have been adopted to detect VAT evasion. Other contributions (González & Velásquez, 2013; Wu et al., 2012) have utilized data mining methods, such as clustering and decision trees, to achieve the same aim. Within this context, classification methods have been combined with dimensionality reduction methods and, particularly, principal component analysis and singular value decomposition (Matos et al., 2015). Finally, machine learning methods have been applied to the tax fraud detection problem by Cecchini et al. (2010), Kleanthous and Chatzis (2020), Vanhoeyveld et al. (2020), Gao et al. (2021), and Savić et al. (2022).

As noted already, what distinguishes VAT fraud with other tax fraud is that VAT fraud—by the very nature of the mechanism underlying the VAT system—is typically not conducted by a single business but is a group (or community) activity. This necessitates that any detection model incorporate the network structure, an element that is central to the present contribution.

Over the last decade, methods related to anomalous detection in networks have been increasingly considered for uncovering fraud. In particular, Chiu et al. (2011) and Subelj et al. (2011) address online auction fraud and insurance fraud, respectively, by using social network analysis, while Van Vlasselaer et al. (2017) utilize network information to detect fraud in social security systems, whereas (Baghdasaryan et al., 2022) develop a network-informed fraud detection technique applied to tax data. Interestingly, Van Vlasselaer et al. (2017) show that incorporating network information allows well-known classification algorithms, such as the random forest, to achieve more accurate fraud detection; they report approximately 7% increase in the area under the curve (AUC) of the receiving operating characteristic (ROC) curve. Additionally, Baghdasaryan et al. (2022) show that historical audit and fraud information for taxpayers can be replaced by features of the observed network of their transactions without reducing significantly classification metrics such as the AUC. However, in contrast with the methodologies developed here, these approaches use only local, node-specific, characteristics of the observed networks to construct covariates (e.g. degree, triangles, and quadrangles) or explore the connectivity of neighbouring nodes in networks with special structure (for example, bipartite).

Closer to the focus of the contribution of this article are contributions which aim to detect anomalies (not specifically fraud) in networks by using the graph embedding approach mapping a network on a vector space which preserves the network structure properties; see for example Cai et al. (2018) and Xu (2021) for recent reviews on the existing graph embedding methods and Ma et al. (2021) for their use on network anomaly detection. Indeed, as demonstrated in Section 5 (Table 1), incorporating network information via graph Laplacian embeddings—derived from its eigenstructure—leads to an approximate 19% improvement in AUC compared to classification models that do not use such network-based features.

### 4 The methodological approach

This section presents the anomaly detection methods for VAT fraud detection. As noted earlier, there are two main challenges in an anomaly detection problem of the type investigated here.

**Table 1.** Sensitivity analysis for XGBoost with and without network features across thresholds.

| Model                | AUC   | Threshold | Accuracy | Sensitivity | Specificity |
|----------------------|-------|-----------|----------|-------------|-------------|
| XGBoost with network | 0.953 | 0.300     | 0.772    | 0.951       | 0.771       |
|                      |       | 0.500     | 0.958    | 0.723       | 0.960       |
|                      |       | 0.700     | 0.991    | 0.514       | 0.994       |
| XGBoost              | 0.802 | 0.300     | 0.938    | 0.415       | 0.941       |
|                      |       | 0.500     | 0.981    | 0.298       | 0.986       |
|                      |       | 0.700     | 0.992    | 0.228       | 0.997       |

The first refers to the classification of ‘normal’ and ‘anomalous’ behaviour, whereas the second relates to the scalability of the problem and the necessity to make the algorithm computationally efficient and robust. The analysis deals with these challenges as follows.

In network analysis, communities are typically identified using the Laplacian matrix, which is derived from the (weighted) adjacency matrix (Merris, 1994). Since the anatomy of VAT fraud involves both individual propensity and community opportunity, the approach adjusts the Laplacian to capture these two dimensions of behaviour. This adjustment is achieved through a global or local spectral decomposition of a corrected Laplacian. In the global approach, singular value decomposition is applied to the regularized Laplacian of the entire network, which consists of hundreds of thousands of vertices. The resulting decomposition is used to update a vertex-specific binary vector based on estimated anomaly probabilities. This process encapsulates both (a) across nodes information and (b) node-specific details. This leads to the development of a new *graph-informed* classifier designed to separate anomalous nodes, such as VAT missing traders, from all the other taxpayers. In the local approach, the corrected Laplacian matrix is embedded into the hierarchical clustering technique recently developed by Li, Lei et al. (2020). In contrast to global clustering methods, which generate a single partition of the network into a fixed number of clusters, this approach constructs a hierarchical tree of communities by recursively dividing larger groups into smaller ones. The process of cluster identification is therefore conducted in a local manner facilitating the challenging task, due to the large number of non-fraudulent nodes, of detecting VAT fraudulent clusters.

The developed methods provide tax authorities with two tools that can be used either separately or in combination in order to enhance their ability to quickly identify VAT MT fraud schemes. In the following subsections, the incorporation of covariate information into the observed VAT network is described, along with the proposed algorithms for community detection and classification. Finally, the implementation of these methods is demonstrated on a population-sized data set covering the entire universe of VAT-registered business in Bulgaria.

#### 4.1 A risk-informed network Laplacian

To model the group structure of activities, it is necessary to detect groups of taxpayers that are more likely to be involved in VAT fraudulent behaviour. This can be achieved by fitting a group model that identifies the nodes belonging to each group. Such a fit can be implemented either under the assumption that there are true blocks in the data (see Newman, 2012) or there is a propensity of a range of nodes to behave like a grouping (as in Olhede & Wolfe, 2014). The most common approach for extracting community structure from a network is *spectral clustering* (Chung & Graham, 1997), which relies on a spectral partitioning of the network’s Laplacian matrix. There are multiple ways to define the Laplacian, both in terms of the Laplacian itself and the adjacency matrix; see, for example, Priebe et al. (2019) for a discussion. This analysis adopts spectral clustering based on a doubly regularized Laplacian, constructed to account for both the community structure underlying VAT fraud and the firm-specific risk of fraudulent behaviour.

In the first level of regularization, following Chaudhuri et al. (2012) and Qin and Rohe (2013), the strong degree heterogeneity (reflecting significant differences in business sizes) of VAT networks is addressed by considering the normalized Laplacian

$$L_\tau = D_\tau^{-1/2} \tilde{A} D_\tau^{-1/2}, \quad (2)$$

where  $\tilde{\mathbf{A}} = \mathbf{A} + \mathbf{A}^T$  is a symmetric matrix,  $\mathbf{D} = \text{diag}\{d_1, \dots, d_N\}$  is a diagonal matrix consisted of the node degrees  $d_i = \sum_{j=1}^N w_{ij}$  and  $\mathbf{D}_\tau = \mathbf{D} + \tau \mathbf{I}$ . Notice that in (2) the transformed adjacency matrix  $\tilde{\mathbf{A}}$  is used rather than the observed adjacency  $\mathbf{A}$ . This choice is convenient because  $\tilde{\mathbf{A}}$  is a symmetric matrix, making its spectral decomposition more interpretable, as it avoids complex eigenvalues that can arise with a non-symmetric  $\mathbf{A}$ . Moreover, it is computationally more efficient to calculate the eigenvalues and eigenvectors of square matrices even in large dimensions (Baglama & Reichel, 2005). Importantly,  $\tilde{\mathbf{A}}$  preserves the directionality of the edges in the observed graph (Malliaros & Vazirgiannis, 2013; Satuluri & Parthasarathy, 2011). Specifically,  $\tilde{\mathbf{A}}$  is the adjacency matrix of an undirected network with the same number of edges, where every directed edge is replaced by an undirected edge whose weight is the sum of the weights of the corresponding edges in the original graph. It is important to note, however, that the transformation of  $\mathbf{A}$  does not account for node similarity based on ingoing and outgoing edges. As a result, clustering approaches based on  $\tilde{\mathbf{A}}$  may fail to group together nodes that are directly connected even if they exhibit similar in- and out-links. Nonetheless, this characteristic is unlikely to substantially affect VAT fraud detection, since the primary goal is to cluster fraudsters who engage in sophisticated, coordinated transactions designed to conceal fraudulent activity.

The parameter  $\tau \geq 0$  introduced by Chaudhuri et al. (2012), corrects for the so-called poor concentration properties caused by large heterogeneity in nodes degrees. In the presence of nodes with very high or very low degrees the spectral analysis of the usual Laplacian  $\mathbf{L}_\tau = \mathbf{D}^{-1/2} \tilde{\mathbf{A}} \mathbf{D}^{-1/2}$  is mainly affected by the highest degree nodes; see for example Mihail and Papadimitriou (2002) for more details. By including  $\tau$  a suitable normalization is achieved, thereby mitigating degree heterogeneity.

To also account for the individual risk of each node being involved in fraud, we follow the approach of Binkiewicz et al. (2017) and introduce a second level of regularization into  $\mathbf{L}_\tau$  by defining the Laplacian

$$\begin{aligned} \mathbf{L}(\alpha, \tau) &= \mathbf{L}_\tau + \alpha \hat{\mathbf{p}} \hat{\mathbf{p}}^T \\ &= \mathbf{D}_\tau^{-1/2} \tilde{\mathbf{A}} \mathbf{D}_\tau^{-1/2} + \alpha \hat{\mathbf{p}} \hat{\mathbf{p}}^T, \end{aligned} \quad (3)$$

where  $\hat{\mathbf{p}}$  is of fraud probabilities estimated using node-specific covariates  $\mathbf{X}$ , with dimensions  $N \times R$ , and scalable machine learning methods described in Section 4.2. Notice that the Laplacian in (3) can also be constructed directly from the covariates in  $\mathbf{X}$  by replacing  $\hat{\mathbf{p}} \hat{\mathbf{p}}^T$  with the matrix  $\mathbf{X} \mathbf{X}^T$ . However, in applications where the number of vertices  $N$  is quite large using  $\hat{\mathbf{p}}$  instead of  $\mathbf{X}$  in (3) can significantly reduce the computational cost of the proposed method. In fraud detection problems there is typically a large imbalance between the fraud and non-fraud cases and most elements of  $\hat{\mathbf{p}}$  are very close to zero. More precisely, if we denote by  $S$  the number of entries in  $\hat{\mathbf{p}}$  exceeding user-specified threshold, we typically expect  $S < N$ . By setting the  $N - S$  values below the threshold to zero the complexity of calculating  $\hat{\mathbf{p}} \hat{\mathbf{p}}^T$  becomes  $\mathcal{O}(S)$ ; much smaller than the  $\mathcal{O}(NR)$  cost of computing  $\mathbf{X} \mathbf{X}^T$  or the  $\mathcal{O}(N)$  complexity for calculating  $\hat{\mathbf{p}} \hat{\mathbf{p}}^T$ . Overall, the estimated probabilities  $\hat{\mathbf{p}}$  summarize the node-specific information contained in each row of  $\mathbf{X}$  in a computationally efficient manner.

It is clear from equation (3) that if the adjacency matrix is zero—that is, there is no network structure in the data—then clustering would rely solely on the values of the vector  $\hat{\mathbf{p}}$ . Moreover, if  $\tau$  was set to zero, then there would be no regularization when inverting the degree matrix. What this means in practice, and for the issues at hand, is that the Laplacian matrix defined in equation (3) effectively accounts for the presence of many low-degree nodes (that is, businesses with few B2B transactions) alongside a few businesses that may have a large number of such transactions. In fact, it is a *risk-corrected Laplacian* designed to improve spectral clustering performance by taking into account the node-specific covariates  $\mathbf{X}$  or the fraud probabilities  $\hat{\mathbf{p}}$ .

Equipped with the risk-corrected Laplacian  $\mathbf{L}(\alpha, \tau)$ , we proceed by calculating eigenvalues  $\lambda_j \in \mathbb{R}$  and eigenvectors  $\mathbf{u}_j \in \mathbb{R}^N$  satisfying

$$\mathbf{L}(\alpha, \tau) \mathbf{u}_j = \lambda_j \mathbf{u}_j. \quad (4)$$

To maintain scalability, we employ the implicitly restarted Lanczos bidiagonalization algorithm (Baglama & Reichel, 2005) to compute the first  $K < N$  eigenvalues and eigenvectors of the (significantly large) dimensional matrix  $L(\alpha, \tau)$  (for example, in equation (4) we have that  $j = 1, \dots, K$ ). This algorithm requires  $O(|E| + NR)K$  operations to compute the top  $K$  eigenvectors of  $L(\alpha, \tau)$ , as it only needs to calculate products of the form  $L(\alpha, \tau)\mathbf{u}$ —where  $\mathbf{u}$  is an arbitrary vector—at each iteration. Furthermore, the method becomes even more computationally efficient by noting that  $L(\alpha, \tau)\mathbf{u} = L_\tau(L_\tau\mathbf{u}) + \alpha\widehat{\mathbf{p}}(\widehat{\mathbf{p}}^\top\mathbf{u})$  and thus the sparsity of  $L_\tau$  and the low rank structure of  $\widehat{\mathbf{p}}\widehat{\mathbf{p}}^\top$  are taken into account. Additionally, by using the vector  $\widehat{\mathbf{p}}$  instead of the covariates matrix  $\mathbf{X}$ , the scalability of the method is further enhanced, reducing the cost for the calculation of computing  $K$  eigenvectors of  $L(\alpha, \tau)$  to  $O(|E| + N)K$ . Next, we discuss how  $\widehat{\mathbf{p}}$  can be estimated efficiently.

## 4.2 Scalable estimation of fraud probabilities

Assume access to an  $N$ -dimensional binary vector  $\mathbf{Y}$ , where the  $i$ th element equals 1 if the business that corresponds to the  $i$ th vertex has engaged in fraudulent activity in the past. Fraud probabilities, independent of the network structure, are then estimated using covariates  $\mathbf{X}$  via a scalable XGBoost binary classification (Chen & Guestrin, 2016), resulting in node-specific risk probabilities probabilities  $\widehat{\mathbf{p}}(\mathbf{X}) \equiv \widehat{\mathbf{p}} = (\widehat{p}_1, \dots, \widehat{p}_N)$ .

Let  $y_i$  denote the  $i$ th element of  $\mathbf{Y}$  and  $\mathbf{x}_i$  the  $i$ th row of  $\mathbf{X}$ . The XGboost algorithm is a regularized version of the well-known gradient boosting method in which an ensemble of decision trees is employed to construct a prediction model for a target variable of interest. Gradient boosting<sup>4</sup> is an iterative algorithm that adopts a gradient descent approach to minimize a loss function using the prediction errors at each data point. At each iteration, the predictions are updated by fitting a new decision tree that aims to reduce the loss function further. More precisely, after training the XGboost algorithm we obtain node-specific predictions

$$\widehat{z}_i = \log\left(\frac{\widehat{p}_i}{1 - \widehat{p}_i}\right) = \sum_{s=1}^S f_s(\mathbf{x}_i), f_s \in \mathcal{F}, \quad (5)$$

where  $\mathcal{F}$  is the space of decision classification trees. Each  $f_s$  corresponds to an independent tree structure  $q_s : \mathbb{R}^p \rightarrow T$  with leaf weights  $\mathbf{v}_s \in \mathbb{R}^T$  where  $T$  is the number of leaves of the tree. Thus, equation (5) implies that the  $i$ th observation is classified by using the decision rules specified by  $q_s$  and by summing up the leaf weights  $\mathbf{v}_s$ . To determine the functions  $f_1, \dots, f_s$ , the objective function

$$\mathcal{L} = \sum_{i=1}^N \ell(y_i, \widehat{z}_i) + \sum_{s=1}^S \Omega(f_s), \quad (6)$$

is minimized, where  $\Omega(f_s) = \gamma T + v \sum_{j=1}^T v_{s,j}^2$  is a regularization term that prevents over-fitting and  $\gamma$  and  $v$  are tuning parameters whereas

$$\ell(y_i, \widehat{z}_i) = \log(1 + e^{\widehat{z}_i}) - y_i \widehat{z}_i,$$

is a differentiable convex loss function which is typically chosen for binary classification (see for example Murphy, 2012 for more details). By noting that the objective in equation (6) includes functions as parameters, and thus its minimization cannot be achieved by using traditional methods, Chen and Guestrin (2016) suggest to perform an additive optimization in the sense that  $f_s(\mathbf{x}_i)$  is added to the prediction  $\widehat{z}_i^{(s-1)}$  obtained in the  $(s-1)$ th iteration of the algorithm. Moreover, since the space of tree structures  $\mathcal{F}$ , is very large, Chen and Guestrin (2016) develop a scalable technique to conduct the required calculations.

<sup>4</sup> See James et al. (2013) for a detailed description of gradient boosting methods.

**Algorithm 1** Network Informed Multiscale Anomaly Detector (NIMAD)

---

**Input:**  $N \times N$  network adjacency matrix  $\mathbf{A}$ ;  $N$ -dimensional vertex specific binary vector  $\mathbf{Y}$  (optional);  $N \times R$  matrix  $\mathbf{X}$  with vertex specific covariates; tuning constant  $\alpha > 0$ ; positive integer  $K$ .

```

1: if  $\mathbf{A}$  symmetric then
2:   Set  $\tilde{\mathbf{A}} = \mathbf{A}$ 
3: else
4:   Set  $\tilde{\mathbf{A}}$  to be the symmetric matrix obtained after suitable transformation on  $\mathbf{A}$ .
5: end if
6: (Optional) Predict anomaly probabilities  $\hat{\mathbf{p}}$  by first training XGboost on responses  $\mathbf{Y}$  and covariates  $\mathbf{X}$ .
7: Calculate  $\mathbf{L}(\alpha, \tau)$  defined by equation (3) if the optional step 6 is implemented or by replacing  $\hat{\mathbf{p}}\hat{\mathbf{p}}^T$  with  $\mathbf{X}\mathbf{X}^T$  otherwise.
8: Compute the eigendecomposition  $\mathbf{L}(\alpha, \tau)$  and form the  $N \times K$  matrix  $\mathbf{U}$  with columns the eigenvectors that correspond to the  $K$  largest eigenvalues.
9: Normalize each row in  $\mathbf{U}$  to have unit length and form the  $N \times K$  matrix  $\mathbf{W}$  with  $w_{ik} = u_{ik}\sqrt{\lambda_k}$ .
10: if  $\mathbf{Y}$  is not provided then
11:   Apply the k-means algorithm to the rows of  $\mathbf{W}$  with  $k = 2$  to obtain vectors with clusters memberships that divide two sub-networks.
12:   Create an  $N$ -dimensional binary vector  $\mathbf{Y}$  where its labels correspond to the memberships of the sub-networks in 11.
13: end if
14: Estimate anomaly probabilities  $\tilde{\mathbf{p}}$  by using XGboost with responses  $\mathbf{Y}$  and features  $\mathbf{W}$ .
```

**Output:**  $N$ -dimensional vector  $\tilde{\mathbf{p}}$  with vertex specific anomaly probabilities.

---

### 4.3 Classification and clustering

Let  $\mathbf{U}$  be the  $N \times K$  matrix whose columns are the eigenvectors  $\mathbf{u}_i$ , which can also be interpreted as a network-informed feature matrix. Two anomaly detection methods are developed based on this structure. The first relies on the global spectral decomposition of  $\mathbf{L}(\alpha, \tau)$  in equation (3) from which  $\mathbf{U}$  is obtained. Then, using the XGboost algorithm again,  $\hat{\mathbf{p}}$  is updated to  $\tilde{\mathbf{p}}$ . A threshold is then selected to separate the businesses into two clusters: those with  $\tilde{p}_i$  below the threshold, and considered as legitimate, and those with  $\tilde{p}_i$  above the threshold, for which further investigation regarding potential participation in VAT fraud is needed; see [Algorithm 1](#) below for a detailed description of the steps in the proposed method.

The second proposed method constructs a hierarchical tree of communities by utilizing the spectral decomposition of  $\mathbf{L}(\alpha, \tau)$  locally within each tree as suggested recently by [Li, Lei, et al. \(2020\)](#). This approach is based on recursive bi-partitioning whereby any given sub-network is divided into two parts. A stopping rule can be incorporated to determine whether a sub-network can be further subdivided into two. In practice, various partitioning methods and stopping rules can be employed; see for example [Li, Levina, et al. \(2020\)](#) for choices on both. In this framework, each sub-network is partitioned using the spectral decomposition of the risk-informed Laplacian  $\mathbf{L}(\alpha, \tau)$  and the corresponding feature of matrix  $\mathbf{U}$ . More precisely, following [Li, Lei et al. \(2020\)](#), the procedure begins by fixing  $K = 2$  and splitting the initial network in two clusters by applying the  $k$ -means algorithm with  $k = 2$  applied to the feature matrix  $\mathbf{U}$ . The same procedure is then recursively applied to each of the two sub-networks obtained in the first step, resulting in 4 clusters after the end of the second iteration. The process continues until the desired number of clusters is identified. [Algorithm 2](#) below outlines the steps of the proposed hierarchical clustering method.

#### 4.3.1 Anomaly detection algorithms

This section provides the algorithmic steps of the two methods developed to detect anomalies in the network constructed from the universe of VAT transactions in Bulgaria. Both proposed algorithms require as inputs the network structure (given by the adjacency matrix) and a node specific set of covariates. Notice also that the first 6 steps of [Algorithms 1](#) and [2](#) are identical.

**Algorithm 2** Hierarchical Anomalous Cluster Identifier (HACI)

---

**Input:**  $N \times N$  network adjacency matrix  $\mathbf{A}$ ;  $N$ -dimensional vertex specific binary vector  $\mathbf{Y}$  (optional);  $N \times p$  matrix  $\mathbf{X}$  with vertex specific covariates; tuning constant  $\alpha > 0$ ; positive integer  $K$ .

- 1: if  $\mathbf{A}$  symmetric then
- 2: Set  $\tilde{\mathbf{A}} = \mathbf{A}$
- 3: else
- 4: Set  $\tilde{\mathbf{A}}$  to be the symmetric matrix obtained after suitable transformation on  $\mathbf{A}$ .
- 5: end if
- 6: (Optional) Predict anomaly probabilities  $\hat{\mathbf{p}}$  by first training XGboost on responses  $\mathbf{Y}$  and covariates  $\mathbf{X}$ .
- 7: Calculate  $\mathbf{L}(\alpha, \hat{\tau})$  defined by equation (3) if the optional step 6 is implemented or by replacing  $\hat{\mathbf{p}}\hat{\mathbf{p}}^T$  with  $\mathbf{X}\mathbf{X}^T$  otherwise.
- 8: Calculate the eigenvectors of  $\mathbf{L}(\alpha, \hat{\tau})$  that correspond to the two largest eigenvalues and form the  $N \times 2$  matrix  $\mathbf{U}$ ; apply the  $k$ -means algorithm to the row of  $\mathbf{U}$  with  $k = 2$  to obtain an  $N$ -dimensional vector  $\mathbf{C}_1$  that separates the  $N$  vertices in two clusters.
- 9: for  $i = 2, \dots, K$  do
- 10: Calculate  $\mathbf{L}(\alpha, \hat{\tau})$  defined by equation (3) for each of the sub-networks specified by  $\mathbf{C}_{i-1}$ .
- 11: For each sub-network and its corresponding matrix  $\mathbf{L}(\alpha, \hat{\tau})$  calculate its eigenvectors and form the  $N \times 2$  matrix  $\mathbf{U}$ ; apply the  $k$ -means algorithm to the row of  $\mathbf{U}$  with  $k = 2$  to obtain vectors with clusters memberships that divide each sub-network in two smaller sub-networks.
- 12: Form the  $N$ -dimensional vector  $\mathbf{C}_i$  that separates the  $N$  vertices into the sub-networks identified in the previous step.
- 13: end for
- 14: Set  $\mathbf{C} = \mathbf{C}_K$  **Output:**  $N$ -dimensional vector  $\mathbf{C}$  with vertex specific cluster memberships.

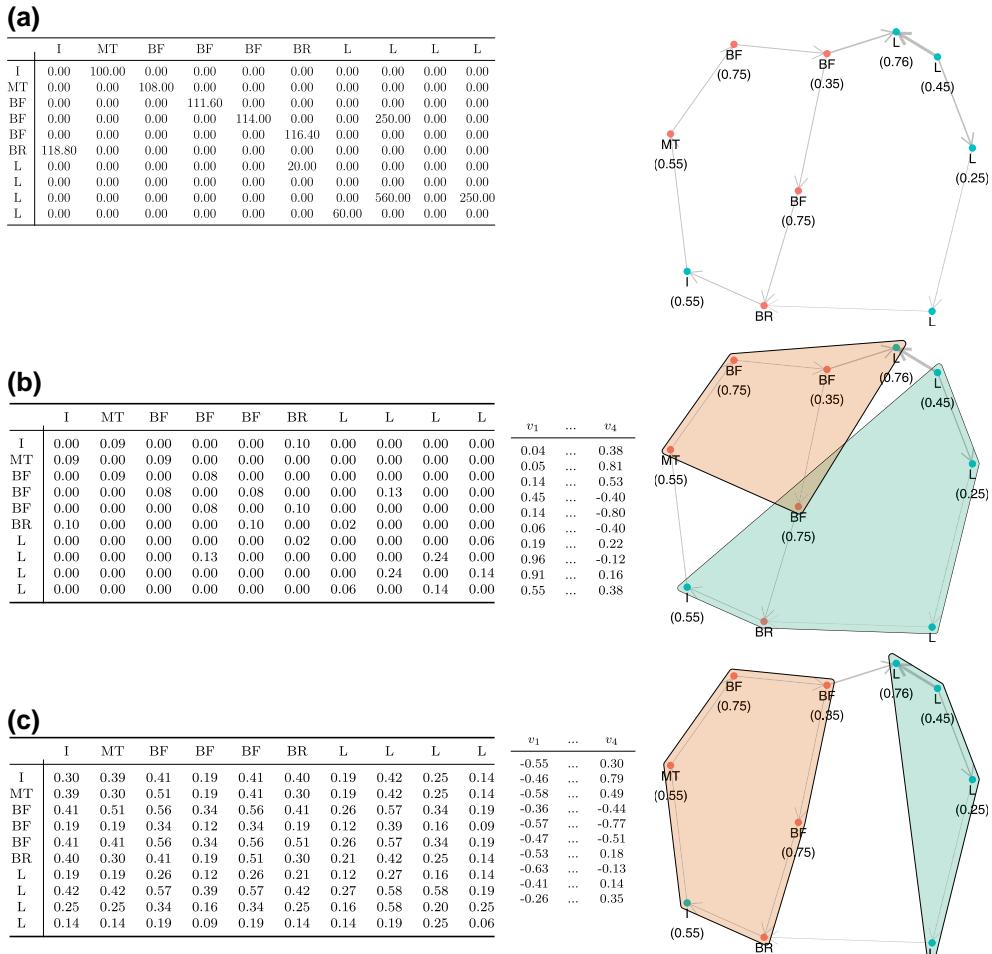
---

**Algorithm 1**—Network Informed Multiscale Anomaly Detector (NIMAD)—summarizes the steps of the network anomaly detection technique developed to classify the network vertices as anomalous or not. In particular, **Algorithm 1** classifies taxpayers as high- or low-risk by applying a classification method to the vertices of the observed network. As a by-product, it also enables clustering of the network’s nodes by leveraging the spectral decomposition of the Laplacian matrix defined in equation (3), which captures the structure of the entire network. The output of the algorithm consists of a vector with estimated anomaly probabilities for each vertex and a vector indicating cluster memberships.

**Algorithm 2**—Hierarchical Anomalous Cluster Identifier (HACI)—summarises the steps of the anomaly detection technique developed to identify anomalous clusters in a network. In contrast to 1, **Algorithm 2** is explicitly cluster-oriented and thus its primary objective is the hierarchical grouping of taxpayers based on shared transaction patterns. This is achieved by recursively applying the spectral decomposition of the Laplacian matrix (from equation (3)) to each branch (or ‘leaf’) of a clustering tree, thereby constructing a hierarchical taxonomy of taxpayer clusters. Although the inputs of **Algorithm 2** are the same as those required by **Algorithm 1**, the positive integer  $K$  which specifies the depth of the constructed hierarchical tree of clusters can be estimated automatically by the algorithm; see for example [Li, Levina, et al. \(2020\)](#) and [Li, Lei, et al. \(2020\)](#). In this article, since the analysis involves a population-scale network consisting of more than 300, 000 vertices,  $K$  is chosen in advance to maintain scalability. More precisely, a small sensitivity analysis indicates that any integer between 5 and 10 does not affect the resulting identification of anomalous clusters and vertices in the real data application. The output of **Algorithm 2** is a vector consisting of vertex specific cluster memberships.

#### 4.4 A toy example

To highlight the advantages of the proposed methodology over traditional network and machine learning methods, an example is developed using simulated data that mimic the simple case of MT fraud discussed in Section 2 and illustrated in [Figure 1](#). In particular, the directed network of



**Figure 3.** (a) Simulated directed network of VAT transactions and fraud probabilities for each vertex where the width of the edges is proportional to the VAT amount exchanged between the businesses represented by each vertex (right) and the adjacency matrix of the simulated network (left). (b) the normalized network Laplacian defined in equation (2) (left), its eigenvectors (middle) and the clusters identified using these eigenvectors (polygons in the right). (c) the risk-corrected network Laplacian defined in equation (3) (left), its eigenvectors (middle) and the clusters identified using these eigenvectors (polygons in the right). The simulated vertices correspond to an importer (I), a missing trader (MT), brokers (BR), buffers (BF), and legitimate (L) taxpayers.

business invoices data (input/output VAT) presented in Figure 3(a) is simulated. It is assumed that 6 out of the  $N = 10$  VAT traders in this network are involved in an MT fraud whereas the remaining 4 vertices correspond to VAT-registered traders who may or may not have transactions with traders. To simplify matters (and save space), it is also assumed that the initial fraud probability for each trader is known, so there is no need to implement the first step of the proposed methodology (Line 6 of Algorithm 1), where the initial fraud probabilities are estimated using the XGboost algorithm. With the simulated network and the known vertex specific fraud probabilities at hand, the traditional normalized (Line 7 of the Algorithm) Laplacian matrix  $L_\tau$ , defined in equation (2), as well as the proposed risk-informed Laplacian  $L(\alpha, \tau)$ , defined in equation (3), are calculated with their numerical values being presented in Figure 3(b) and (c), respectively. Since both of the proposed fraud detection algorithms rely on the eigenvalues and the eigenvectors of the network Laplacian matrix, the eigendecomposition of  $L(\alpha, \tau)$  is calculated (Step 8 of the Algorithm) to compare their efficiency when using either version. Next, the  $k$ -means algorithm is applied on the first  $K = 4$  (Line 11 of the Algorithm), normalized to have unit length, from each matrix to

partition the observed network in two sub-networks (clusters). Figure 3(b) and (c) displays the identified clusters using  $L_\tau$  and  $L(\alpha, \tau)$ , respectively.

In this application, the XGBoost algorithm is implemented using the r-package xgboost (Chen et al., 2019). The described steps correspond to the application of Algorithm 2 for  $K = 2$  in the simulated data set by calculating either  $L_\tau$  or  $L(\alpha, \tau)$  in the 8th step with  $\tau = 0.01$  and  $\alpha = 1$ .

Close inspection of the identified clusters reveals that by relying on the eigendecomposition of the proposed risk-informed Laplacian  $L(\alpha, \tau)$  enables detection of all businesses involved in the simulated MT fraud, whereas clustering based on  $L_\tau$  is less effective in identifying the fraudulent clusters. This example, therefore, provides clear evidence that the outputs from both Algorithms 1 and 2—namely, the estimated fraud probabilities  $\tilde{p}$  and cluster membership  $C$ , respectively, which rely on the eigendecomposition of  $L(\alpha, \tau)$ —are more accurate than the outputs of the same algorithms when using  $L_\tau$  instead. Furthermore, attempting to identify the members of the MT scheme solely relying on the initial fraud probabilities would also result in less accurate fraud detection than that illustrated in Figure 3(c). The implication of all this is clear: fraud detection approaches that combine the structure of the VAT network with risk information about individual businesses (vertices) should be preferred over methods that utilize only a single source of information. Section 5 provides similar supporting evidence based on real data from the Bulgarian tax authorities.

#### 4.5 Sensitivity analysis

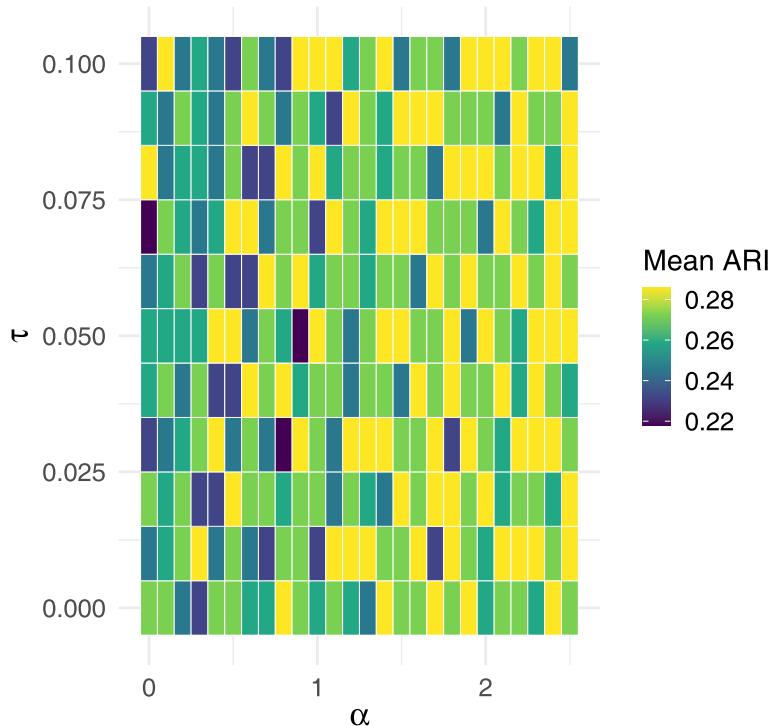
Noting that both of the proposed fraud detection algorithms depend on tuning parameters, namely  $\tau$  and  $\alpha$ , the following sensitivity analysis is conducted. A grid of values for  $\tau$  and  $\alpha$  is considered and, for each parameter combination, 30 networks of VAT transactions are simulated as follows. The base adjacency matrix is taken from Figure 3(a) perturbed with Gaussian noise to introduce variability, while each node is labelled as fraud or not by simulating binary random variables with success probabilities given in Figure 3. Then, for each simulated network, Algorithm 2 is applied again with  $K = 2$  and the adjusted rand index (ARI) is calculated. The ARI is a statistical measure that is commonly employed to compare different clustering assignments. In particular, ARI is a measure of similarity between clustering from two different methods corrected for random clustering, ranging from  $-1$  to  $1$ . A value greater than zero indicates perfect agreement between the clustering from the methods under comparison, a value close to zero implies that the predicted clusters are no better than randomly assigning nodes to groups while negative ARI implies that even random clustering would be more accurate; for more details see the Appendix as well as Zhang et al. (2012) for a comprehensive discussion.

In the application, the true clustering of the nodes in the simulated VAT networks—defined by the binary labels assigned to the nodes—is compared with the clusters identified through the application of Algorithm 2. Figure 4 shows, for each combination of  $\tau$  and  $\alpha$ , the mean ARI across the 30 simulated VAT networks. It is clear that the mean ARI values remain relatively stable within a broad region of the parameter space, particularly for  $\alpha$  values between 0 and 2.5 and  $\tau$  between 0 and 0.1, where the mean ARI ranges from 0.22 to 0.28. This indicates that the clustering results are robust to moderate changes in both parameters. Only for some extreme values in the parameter grid the performance of the clustering technique begins to degrade, suggesting that the algorithm does not require precise tuning to achieve good results. Overall, the developed clustering method demonstrates strong robustness across a reasonable range of hyperparameters.

The code to replicate the described examples can be found online at <https://gitlab.com/aggelisalexopoulos/vat-fraud>.

### 5 Real data analysis

The proposed algorithms, NIMAD and HACI, are applied to the universe of VAT returns provided by the Bulgarian National Revenue Authority (BNRA) for the years 2016–2017, along with ledger data for all  $N = 312,762$  VAT-registered taxpayers in Bulgaria in 2017. An out-of-sample exercise is also conducted, in which the models are trained using networks constructed from monthly VAT returns submitted by taxpayers between January 2016 and November 2017. The objective of this exercise is to probabilistically predict the illegitimate taxpayers for December 2017. The results are compared with those from classification methods



**Figure 4.** The mean adjusted rand index (ARI) of the clustering detected by using [Algorithm 2](#) across a grid of values for the parameters  $\tau$  and  $\alpha$ .

that rely solely on covariates describing taxpayer profiles without taking into account the network structure of the data. This out-of-sample exercise demonstrates that network information plays a key role in the efficient detection of anomalous vertices. Finally, the results from the two anomaly detection methods are analysed, classifying groups of fraudsters and legitimate taxpayers, and identifying clusters of taxpayers with characteristics similar to known fraudsters.

Both of the proposed algorithms rely on the prediction of probabilities of risky VAT taxpayers. This is achieved by first training the XGboost algorithm with inputs a binary response vector  $\mathbf{Y}$  and the  $N \times R$  matrix  $\tilde{\mathbf{X}}$  consisted of the available covariates which include the number of employees, the labour cost, and other records that taxpayers declare with their VAT returns. In particular, the types of covariates used are a subset of the risk-based criteria which the BNRA employs in order to prioritize the taxpayers with respect to their riskiness of being involved in a VAT missing trader fraud. Covariates are also constructed using the characteristics of the 23 observed networks, corresponding to the VAT returns submitted monthly between January 2016 and November 2017. For each vertex, the mean degree, strength, and centrality across the observed networks are calculated. The resulting matrix has  $R = 49$  columns. The  $N \times R$  matrix  $\mathbf{X}$  consisting of these covariates for December 2017 (the month for which risk probabilities are to be predicted) is then used to obtain the vector  $\hat{\mathbf{p}}$  appearing in equation (3).

The input adjacency matrix  $\mathbf{A}$ , required by both of the developed anomaly detection methods, corresponds to the adjacency matrix of a directed weighted network, constructed by the VAT returns submitted in December 2017. In this case,  $\mathbf{A}$  is an asymmetric matrix reflecting the fact that relationships between taxpayers are not necessarily reciprocal. To address this, a symmetric matrix is constructed as  $\tilde{\mathbf{A}} = \mathbf{A} + \mathbf{A}^T$ . The undirected network represented by  $\tilde{\mathbf{A}}$  retains the same edges as the original network, but directed edges are replaced with undirected edges whose weights equal the sum of the original directed weights; that is, each pair of nodes  $i, j$  is connected by an undirected edge with weight  $\tilde{A}_{ij} = A_{ij} + A_{ji}$  associated with the edge in question. Community detection methods that are based on  $\tilde{\mathbf{A}}$  tend to group nodes that share similar incoming and outgoing edges ([Satuluri & Parthasarathy, 2011](#)). Arguably, this symmetrization is reasonable since malicious behaviour often manifests through anomalous connectivity patterns that are not strictly dependent

on directionality, and thus the structure of the undirected network still captures essential signals of suspicious activity. However, the effectiveness of this approach may be limited when the direction of transactions or interactions carries key asymmetrical information—such as deliberate imbalances in money flow or one-sided interactions—which are purposefully introduced to conceal fraudulent behaviour. Finally, since both of the anomaly detection algorithms rely on the eigenvalues and eigenvectors (and so on the spectral decomposition of the matrix  $L(\alpha, \hat{\tau})$  in equation (3)) it is necessary to choose the tuning parameters  $\alpha$  and  $\tau$  carefully. Following the approach in [Qin and Rohe \(2013\)](#),  $\tau$  is set to the average degree that is,  $\hat{\tau} = N^{-1} \sum_{i=1}^n d_{ii} = \bar{d}$ . The parameter  $\alpha$  can be determined from the eigenvectors of  $D_{\tau}^{-1/2} \tilde{A} D_{\tau}^{-1/2}$  and  $\hat{p}$  (see [Binkiewicz et al., 2017](#) where they show how to set  $\alpha$  such that the information contained in  $D_{\tau}^{-1/2} \tilde{A} D_{\tau}^{-1/2}$  as well as in  $\hat{p}$  is captured in the leading eigenspace of  $L(\alpha, \hat{\tau})$ ).

## 5.1 Out-of-sample detection

To evaluate the performance of the anomaly detection algorithms, an out-of-sample detection exercise is designed by constructing a time series of networks based on 24 months of data, corresponding to the monthly observations from 2016 and 2017.

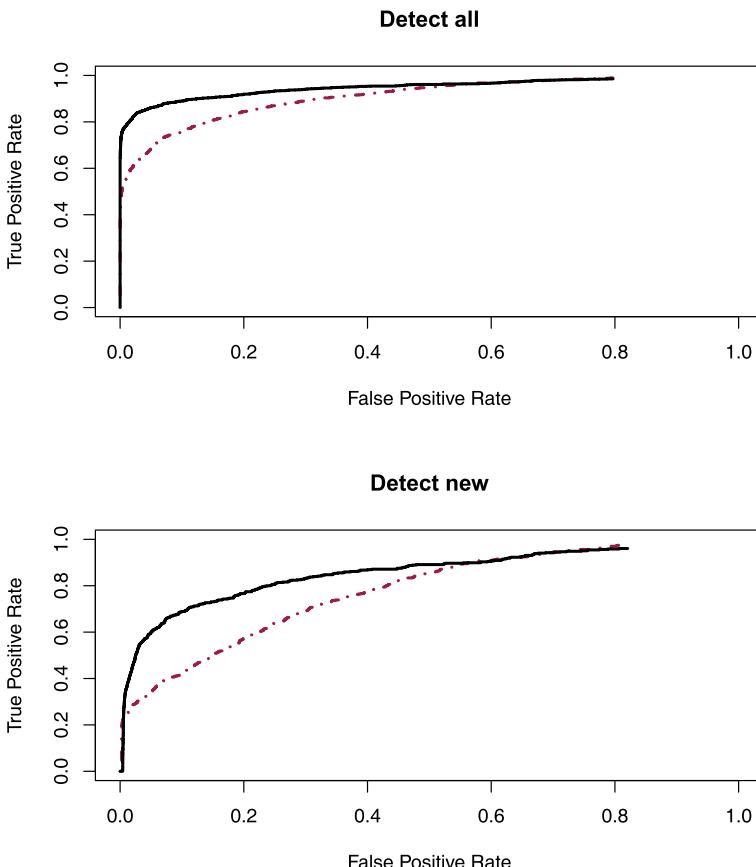
The first step, for both methods, involves classifying the 24th month of observations (December 2017) using information from the preceding 23 months. This setup requires a binary vector indicating the anomalous vertices of ‘high-risk’ taxpayers, a matrix of covariates, and an adjacency matrix. The binary vector  $Y$  represents a classification of ‘high-risk’ and ‘low-risk’ taxpayers, as determined by the BNRA up to November 2017. This constitutes an unbalanced classification problem, as the proportion of fraudulent nodes is unlikely to approach one half ([Hand & Vinciotti, 2003](#)), implying that different types of misclassification are associated with different losses. To address the class imbalance problem random oversampling is applied by re-sampling the set of ‘high-risk’ taxpayers to construct a balanced data set. This oversampling technique is chosen among other possible approaches to keep the method simple while preserving all information contained in the original; see for example [Menon et al. \(2013\)](#) for a comparison of various methods that have been developed to deal with data imbalance problems. For the out-of-sample analysis, the weighted directed network constructed from the VAT returns submitted in December 2017 is used. The tuning parameter  $\alpha$  is selected to balance the contribution of the network structure, as captured by  $\tilde{A}$ , and the individual probabilities  $\hat{p}$ . Sensitivity analysis indicates that a value of 0.01 for  $\alpha$  is appropriate.

Finally, both of the developed anomaly detection methods rely on the calculation of the spectral decomposition of the matrix  $L(0.01, \hat{\tau})$  in equation (3), which is computed using the Lanczos bi-diagonalization method ([Baglama & Reichel, 2005](#)). In the case of NIMAD, which uses the eigenvectors computed globally from the matrix  $L(0.01, \hat{\tau})$  corresponding to the entire observed network, the algorithm is stopped after calculating the first  $K = 200$  eigenvalues and eigenvectors, as the eigenvalues beyond this point were largely similar. Their values are reported in [Figure A1](#) in the Appendix. The implementation of HACI requires only the  $K = 2$  largest eigenvalues and corresponding eigenvectors of the matrix  $L(0.01, \hat{\tau})$ , which pertains to each local ‘leaf’ in the hierarchical tree of communities under construction. Notably, applying [Algorithm 1](#) to the dataset took approximately 3 h on a laptop with a 1.6 GHz dual-core Intel Core i5 CPU running R 4.0.0 ([R Core Team, 2021](#)), whereas [Algorithm 2](#) completed in under an hour.

## 5.2 Determining the accuracy of the proposed methods

[Algorithm 1](#) is evaluated by assessing its ability to predict the provided list of risky taxpayers as of December 2017. From this list, it is observed that 64% of the ‘high-risk’ registrations of taxpayers in December 2017 had in fact been identified as ‘high-risk’ already in November 2017. The remaining 36% were registered for the first time as ‘high-risk’ in December 2017. Accordingly, two prediction tasks are considered: (a) identifying all high-risk VAT registrations in 2017 and (b) identifying only the newly classified high-risk VAT registrations in 2017.

To assess the performance of the proposed methodology, ROC curves (as in [Hsieh et al., 1996](#)) are compared between the proposed approach and an XGBoost classifier that excludes network information. [Figure 5](#) illustrates that the proposed algorithm outperforms the standard XGboost classifier for both existing and newly identified high-risk taxpayers in December

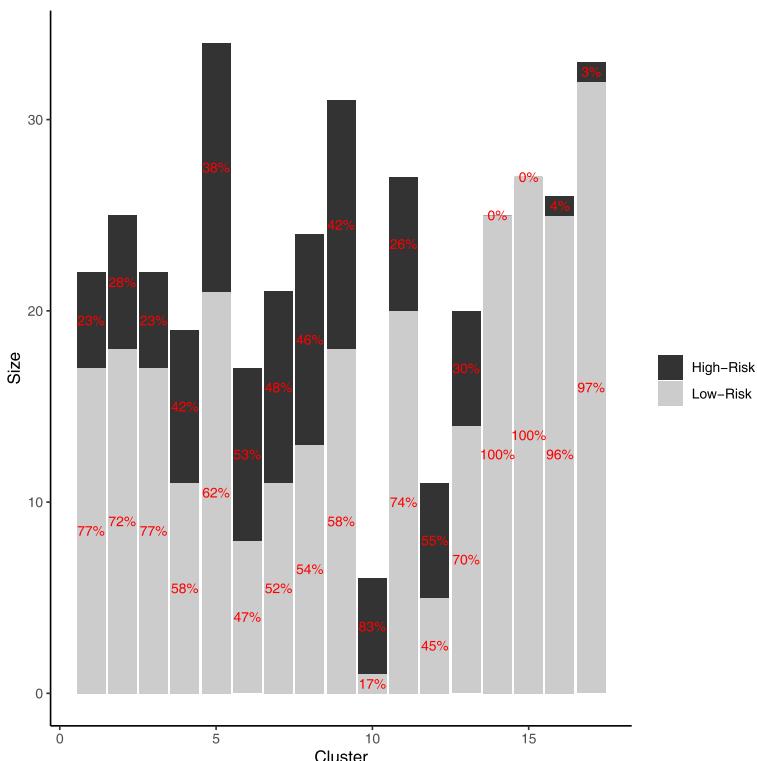


**Figure 5.** ROC curves comparing the out-of-sample classification performance of Algorithm 1 (solid line) with the performance of a classifier that does not utilize network information (dotted line). Top panel: results for detecting all ‘high-risk’ taxpayers of December 2017. Bottom panel: results for detecting taxpayers newly added to the risk registration list of the Bulgarian National Revenue Agency in December 2017.

2017. This provides strong evidence of the value of combining both individual and group-level patterns to detect fraud. [Table 1](#) presents a sensitivity analysis of XGBoost models, comparing versions with and without network features across a range of classification thresholds. The inclusion of network information significantly improves performance—most notably in the most challenging case: predicting newly registered risky VAT taxpayers in 2017. In this setting, the model’s area under the ROC curve (AUC) increases markedly from 0.802 to 0.953, as shown in the bottom panel of [Figure 5](#).

To assess the output of [Algorithm 2](#), the analysis focuses on the identification of taxpayers included for a first time in December 2017 in the list of ‘high-risk’ registrations. [Figure 6](#) displays the sizes of the 17 clusters, each containing more than 10% known (up to November 2017) ‘high-risk’ taxpayers, as well as the proportion of ‘high-’ and ‘low-risk’ taxpayers within each cluster. Visual inspection of [Figure 6](#) reveals that the hierarchical construction of clusters enables the identification of groups that rarely exceed 30 members, with the proportion of target ‘high-risk’ taxpayers in most clusters ranging between 23% and 83%.

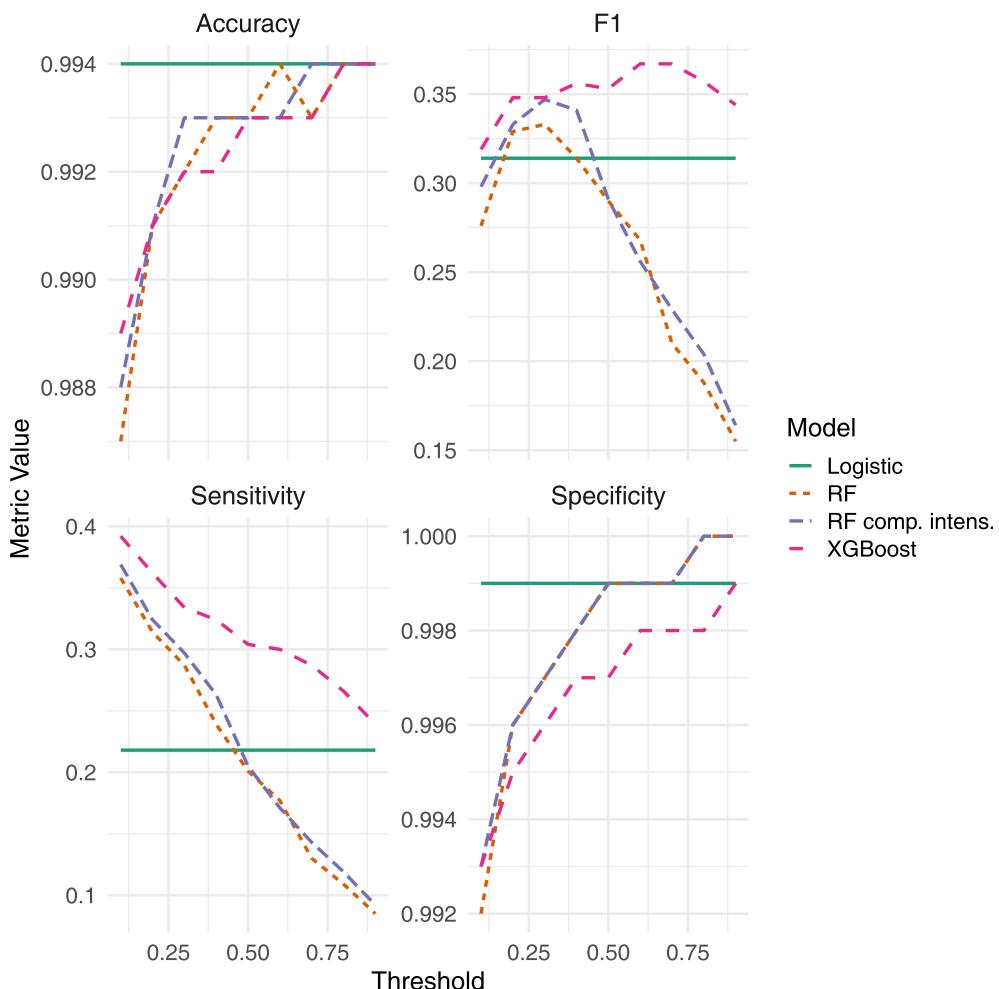
Recognizing that both proposed fraud detection methods rely on XGBoost for classification, its performance is compared against two widely used alternatives: logistic regression and random forests. Logistic regression serves as a benchmark due to its interpretability and long-standing use in classification problems. Random forests, on the other hand, provide a strong nonparametric alternative capable of capturing complex interactions, while requiring less intensive hyperparameter tuning than XGBoost.



**Figure 6.** Size of VAT fraudulent clusters identified using [Algorithm 2](#) which constructs an hierarchical tree of clusters by using recursive bi-partitioning of the observed VAT network. Each bar displays the proportion of 'high-' and 'low-risk' taxpayers included in the corresponding cluster.

To examine whether added model complexity improves predictive performance, a computationally intensive version of the random-forest classifier is also evaluated. This comparative analysis provides insights into whether the results presented in [Figure 5](#) are model-dependent or robust across different classification methods. [Figure 7](#) shows that XGBoost clearly outperforms the alternatives. Specifically, it achieves the highest sensitivity and F1 score across a broad range of decision thresholds, indicating strong performance in identifying fraudulent cases while balancing precision and recall. The specificity and overall accuracy of XGBoost are also competitive, suggesting that this improvement in sensitivity does not come at the expense of overall correctness. Notably, logistic regression lags behind in most metrics, particularly in F1 and sensitivity, likely due to its linear nature. Although both versions of the random forest surpass logistic regression, they still fall short of XGBoost, underscoring the XGBoost's superior ability to model nonlinear interactions and deliver consistent predictive gains.

To compare the two proposed anomaly detection methods, it is first important to note that they are primarily distinguished by the output they provide. [Algorithm 1](#) classifies the taxpayers as 'high-' and 'low-risk' by conducting the corresponding classification in the vertices of the observed network. As a by-product, it also enables clustering of the network nodes using the spectral decomposition of the Laplacian matrix in (3), computed across the entire network. [Algorithm 2](#), in contrast to 1, can be considered as a *cluster-oriented* algorithm since its main aim is the hierarchical identification of groups of taxpayers with common patterns of transactions. It achieves this by recursively applying the spectral decomposition of the Laplacian matrix in (3) to each leaf of a hierarchical clustering tree. Consequently, [Algorithm 1](#) is expected to deliver a more accurate classification of 'high-' and 'low-risk' taxpayers, while [Algorithm 2](#) is expected to be more effective in identifying sizeable fraudulent clusters that merit further investigation by tax authorities, particularly when targeting groups of taxpayers involved in coordinated illegal activities.

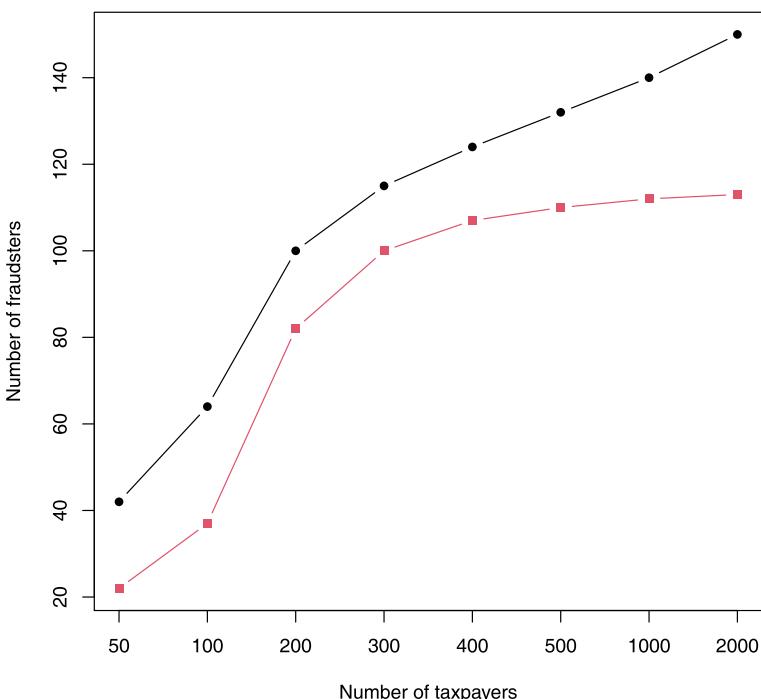


**Figure 7.** Performance metrics (sensitivity, specificity, accuracy, and F1 score) against classification thresholds for logistic regression, random forest (standard and computationally intensive) and XGBoost.

### 5.3 Policy evaluation of the algorithmic outputs

The benefit derived from the automated detection algorithms proposed in this article is evident. Currently, BNRA applies risk-based rules to all submitted tax returns and each month prioritizes 15,000 returns as ‘high-risk’. Through additional selection criteria this number is reduced to 500 and, ultimately, audits identify 100 taxpayers as having participated in VAT fraud. The methods proposed in this contribution offer fully automated mechanisms for identifying VAT fraudsters streamlining and potentially improving this multi-stage process.

Automation has a number of clearly established advantages: it reduces costs, increases transparency and reproducibility, and explicitly balances information obtained from a single taxpayer with that provided by the population-scale data. The out-of-sample exercise demonstrates a clear improvement in identification for a fixed false positive rate. In particular, the proposed method identified 200 taxpayers with the highest estimated fraud probabilities (using [Algorithm 1](#)) of whom 100 had been flagged as high-risk for VAT fraud for the first time in December 2017. By automating the process, the set of potentially fraudulent taxpayer is reduced from the set of 500 identified through BNRA’s human-driven selection procedure. Moreover, the hierarchical clustering provided by [Algorithm 2](#) facilitates quick identification of relatively small groups of taxpayers exhibiting similar fraudulent behaviour. Interestingly, BNRA, as a response to the results presented in



**Figure 8.** The x-axis shows the number of taxpayers that need to be reported for auditing to identify the number of taxpayers that have entered the risk-list of the Bulgarian National Revenue Agency for a first time in December 2017 (y-axis). The dotted line corresponds to [Algorithm 1](#) and the squared line corresponds to [Algorithm 2](#).

this article, has already begun automating and strengthening further their auditing function to fully leverage the benefits from detecting multiple members of VAT fraud schemes. This work has also appeared as a case study in [OECD \(2022\)](#).

Finally, [Figure 8](#) displays the number of new entries in the risky taxpayers list that can be identified for a given number of taxpayers using either [Algorithm 1](#) or [2](#). The figure shows that reducing the number of reported taxpayers from 200 to 50 minimizes the false positive rate, since 40 of them indeed entered the BNRA's list of risky VAT-registered taxpayers in December 2017. Allowing for more false positives—by increasing the number of reported taxpayers from 200 to 500 (the number currently audited by the BNRA)—enables the prediction of more than 120 'high-risk' taxpayers. This number further increases to 140 if 2,000 VAT-registered taxpayers are reported for auditing. Overall, [Figure 8](#) confirms the superior performance of [Algorithm 1](#) in classifying 'high-' and 'low-risk' taxpayers compared to [Algorithm 2](#).

## 6 Concluding remarks

This article contributes to the emerging literature focused on developing novel and efficient tools for fraud detection. With VAT fraud in mind—a form of fraud with significant revenue consequences—it develops fraud detection tools that leverage advanced quantitative, statistical, and machine learning methods. Importantly, the analysis explicitly accounts for inherent issues in the fraud detection process, such as the non-random nature of the audits and the quality of the audit execution. Accordingly, the methods are designed to operate either in a supervised manner—using historical audit information from the Revenue Agency—or in an unsupervised manner, enabling fraud detection without relying on prior audit labels. Importantly, unlike traditional data mining and machine learning approaches, the proposed methods draw on tools from network science to integrate business-specific characteristics with insights derived from businesses' interactions transactional—specifically, through analysis of the VAT transaction network. This enables Revenue Authorities to more effectively and efficiently identify VAT fraudsters, who often

rely on complex transaction structures to obscure fraudulent behaviour and hinder detection efforts. Application of the developed methods to real-world data demonstrates their effectiveness. In particular, incorporating network structure to model VAT transactions significantly enhances the performance of standard approaches that rely exclusively on business-specific information.

Though the algorithms are general enough to capture fraud within a broad range of VAT systems (e.g. simplified tax regime), and as long as the incentives for some form of evasion or misreporting remains—such as underreporting sales, inflating input claims, or exploiting network structures to obscure liability—a more formal treatment of institutional variations of VAT offers a promising direction for future research and a natural extension of the current work. Arguably, detecting anomalies in the VAT network is not solely a cross-sectional problem but it has an inter-temporal dimension. Fraudulent taxpayers learn from interacting with the Revenue Authority, just as the Revenue Authority learns from uncovering fraud. For this, the adjustment of the compliance strategy requires to be appropriately adjusted (Black & Baldwin, 2012). The VAT networks analysed are static, in the sense that the changes in their structure through time are assumed to convey no additional information. For the time horizon of the data set this is not a significant omission, since B2B interactions are not expected to vary significantly within a period of 23 months. For longer time horizons, however, this might matter. For this, it will be interesting to extend the developed fraud detection methodology to *multi-layer* networks which can also incorporate the time dimension of the observed networks. Multi-layer networks can capture different types of relationships between businesses—such as transactional links and shared board membership—which may evolve over time. These structures allow for the detection of communities exhibiting abnormal connectivity patterns across multiple layers. A detailed exploration of this approach is left for future research.

Nevertheless, it is hoped that the results presented in this article will prove instructive and underscore the value of developing algorithms designed to support the effective functioning of economic systems.

## Acknowledgments

We are grateful to the Management of the Bulgarian National Revenue Agency for supporting this research and to the many officials who have provided feedback through extensive discussions and, in particular, Petya Staneva, Albena Nikolova, Zheko Zhelev, Mariela Zarkova, and Yuliana Velichkova for providing us with their insightful operational knowledge. The views expressed in the article are those of the authors and do not necessarily reflect the views of the Bulgarian National Revenue Agency and its Management. Dellaportas, Gyoshev, and Kotsogiannis acknowledge financial support from HSBC-Alan Turing Institute under grant TEDSA2/100056. Kotsogiannis also acknowledges support from ESRC (Grant ES/S00713X/1 and ES/X003973/1). Alexopoulos and Olhede acknowledge support from the 7th European Community Framework Programme (Grant CoG 2015-682172NETS). Part of the work was completed during a Post-Doctoral Fellowship of Alexopoulos at TARC (University of Exeter). The research project is implemented in the framework of H.F.R.I call ‘Basic research Financing (Horizontal support of all Sciences)’ under the National Recovery and Resilience Plan ‘Greece 2.0’ funded by the European Union – NextGenerationEU (H.F.R.I. Project Number: 15973). An earlier version of the article was circulated as ‘Detecting Anomalies in Heterogeneous Population-Scale VAT Networks’. Discussions with Michael Veale (UCL) are gratefully acknowledged.

*Conflicts of interest:* The authors declare that they have no conflicts of interest.

## Funding

P.D., S.G., and C.K. acknowledge financial support from HSBC-Alan Turing Institute under grant TEDSA2/100056. C.K. also acknowledges support from ESRC (Grant ES/S00713X/1 and ES/X003973/1). A.A. and S.C.O. acknowledge support from the 7th European Community Framework Programme (Grant CoG 2015-682172NETS). The research project implemented in the framework of H.F.R.I. call ‘Basic Research Financing (Horizontal Support of all Sciences)’

under the National Recovery and Resilience Plan 'Greece 2.0' funded by the European Union - NextGenerationEU (H.F.R.I. Project Number: 15973).

## Data availability

Code for the replication of all plots based on simulated data is available at <https://gitlab.com/aggelisalexopoulos/vat-fraud>. The real data used in the empirical analysis can be made available upon request.

## Appendix

### Economic sectors

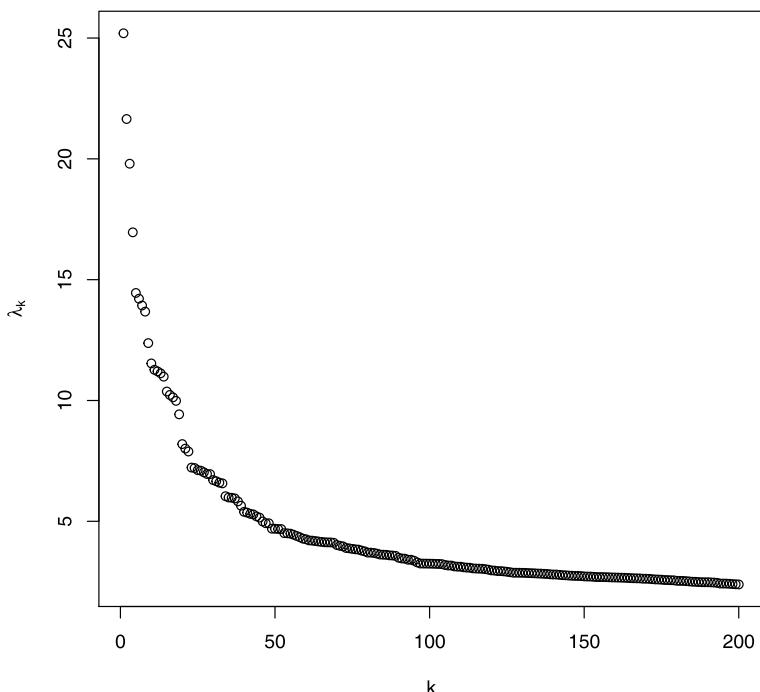
**Table A1** displays the codes of the economic sectors in Bulgaria classified according to the Nomenclature of Economic Activities (NACE) system.

**Table A1.** Sector codes according to the Nomenclature of Economic Activities (NACE) classification system.

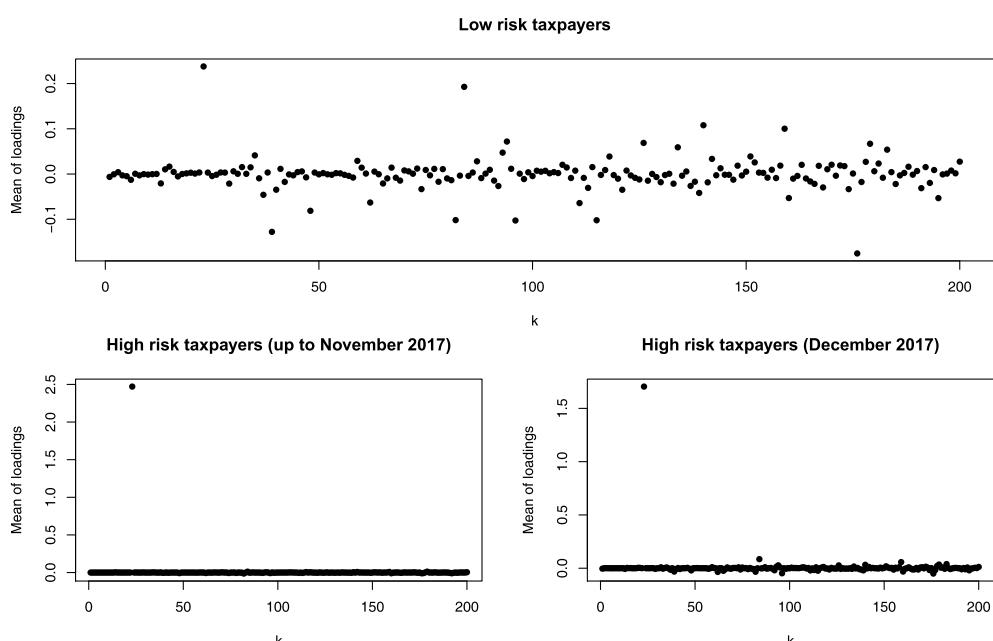
| Code | Sector   |
|------|--|
| A    | Agriculture, forestry, and fishing                                   |
| B    | Mining and quarrying   |
| C    | Manufacturing  |
| D    | Electricity, gas, steam, and air conditioning supply                 |
| E    | Water supply; sewerage; waste management and remediation activities  |
| F    | Construction   |
| G    | Wholesale and retail trade; repair of motor vehicles and motorcycles |
| H    | Transporting and storage   |
| I    | Accommodation and food service activities                            |
| J    | Information and communication  |
| K    | Financial and insurance activities                                   |
| L    | Real estate activities   |
| M    | Professional, scientific, and technical activities                   |
| N    | Administrative and support service activities                        |
| O    | Public administration and defence; compulsory social security        |
| P    | Education  |
| Q    | Human health and social work activities                              |
| R    | Arts, entertainment, and recreation                                  |
| S    | Other services activities  |
| NA   | Not available information of the economic activity                   |

### Results from the spectral decomposition

**Figure A1** displays the first 200 eigenvalues of the matrix  $L(0.01, \hat{\tau})$  computed by using the Lanczos bidiagonalization algorithm. **Figure A2** shows the mean of each loading vector, separately for 'low-risk' taxpayers, the 'high-risk' taxpayers used to train NIMAD and the 'high-risk' taxpayers targeted for detection. Close inspection of the figure reveals that for the 'high-risk' taxpayers there exists one eigenvector whose mean loading is substantially higher than those corresponding to the remaining eigenvectors. In contrast, for 'low-risk' taxpayers, the mean loadings are relatively uniform across all eigenvectors. This suggests that using the columns of matrix  $\mathbf{W}$  as features in the XGBoost algorithm at Step 9 of the **Algorithm 1** enables an accurate classification between 'high-' and 'low-risk' taxpayers.



**Figure A1.** The first  $K = 200$  eigenvalues of the matrix  $\mathbf{L}(0.01, \hat{\tau})$  computed by using the Lanczos bidiagonalization algorithm, [Baglama and Reichel \(2005\)](#).



**Figure A2.** Mean of the loadings that correspond to the first  $K = 200$  eigenvalues of the matrix  $\mathbf{L}(0.01, \hat{\tau})$ . The x-axis indicates the loading that corresponds to the  $k$ th eigenvalue.

## References

Ainsworth R., & Madzharova B. (2012). Real time collection of value-added taxes: Some business and legal implications. *Boston University School of Law Working Paper* No 12–51.

Akoglu L., Tong H., & Koutra D. (2015). Graph based anomaly detection and description: A survey. *Data Mining and Knowledge Discovery*, 29(3), 626–688. <https://doi.org/10.1007/s10618-014-0365-y>

Athey S. (2017). Beyond prediction: Using big data for policy problems. *Science*, 355(6324), 483–485. <https://doi.org/10.1126/science.aal4321>

Baesens B., Setiono R., Mues C., & Vanthienen J. (2003). Using neural network rule extraction and decision tables for credit-risk evaluation. *Management Science*, 49(3), 312–329. <https://doi.org/10.1287/mnsc.49.3.312.12739>

Baghdasaryan V., Davtyan H., Sarikyan A., & Navasardyan Z. (2022). Improving tax audit efficiency using machine learning: The role of Taxpayer's network data in fraud detection. *Applied Artificial Intelligence*, 36(1), Article 2012002. <https://doi.org/10.1080/08839514.2021.2012002>

Baglama J., & Reichel L. (2005). Augmented implicitly restarted Lanczos bidiagonalization methods. *SIAM Journal on Scientific Computing*, 27(1), 19–42. <https://doi.org/10.1137/04060593X>

Barons M. J., & Shenvi A. (2024). Where the bee sucks: A dynamic Bayesian network approach to decision support for pollinator abundance strategies. *Journal of the Royal Statistical Society Series A: Statistics in Society*, 187(2), 393–409. <https://doi.org/10.1093/jrsssa/qnac126>

Basta S., Fassetti F., Guarascio M., Manco G., Giannotti F., Pedreschi D., Spinsanti L., Papi G., & Pisani S. (2009). High quality true-positive prediction for fiscal fraud detection. In *2009 IEEE International Conference on Data Mining Workshops* (pp. 7–12). IEEE.

Belkin M., & Niyogi P. (2003). Laplacian eigenmaps for dimensionality reduction and data representation. *Neural Computation*, 15(6), 1373–1396. <https://doi.org/10.1162/089976603321780317>

Binkiewicz N., Vogelstein J. T., & Rohe K. (2017). Covariate-assisted spectral clustering. *Biometrika*, 104(2), 361–377. <https://doi.org/10.1093/biomet/asw008>

Black J., & Baldwin R. (2012). When risk-based regulation aims low: A strategic framework. *Regulation & Governance*, 6(2), 131–148. <https://doi.org/10.1111/rego.2012.6.issue-2>

Cai H., Zheng V. W., & Chang K. C.-C. (2018). A comprehensive survey of graph embedding: Problems, techniques, and applications. *IEEE Transactions on Knowledge and Data Engineering*, 30(9), 1616–1637. <https://doi.org/10.1109/TKDE.69>

Cecchini M., Aytug H., Koehler G. J., & Pathak P. (2010). Detecting management fraud in public companies. *Management Science*, 56(7), 1146–1160. <https://doi.org/10.1287/mnsc.1100.1174>

Chaudhuri K., Chung F., & Tsitsas A. (2012). Spectral clustering of graphs with general degrees in the extended planted partition model. In *Conference on Learning Theory, JMLR Workshop and Conference Proceedings* (pp. 35–1).

Chen T., & Guestrin C. (2016). Xgboost: A scalable tree boosting system. In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* (pp. 785–794).

Chen T., He T., Benesty M., Khotilovich V., Tang Y., Cho H., Chen K., Mitchell R., Cano I., Zhou T., Li M., Xie J., Lin M., Geng Y., & Li Y. (2019). *xgboost: Extreme gradient boosting*. R package version 0.90.0.2.

Chiu C., Ku Y., Lie T., & Chen Y. (2011). Internet auction fraud detection using social network analysis and classification tree approaches. *International Journal of Electronic Commerce*, 15(3), 123–147. <https://doi.org/10.2753/JEC1086-4415150306>

Chung F. R., & Graham F. C. (1997). *Spectral graph theory*. American Mathematical Soc, Number 92.

De Bacco C., Contisciani M., Cardoso-Silva J., Safdari H., Borges G. L., Baptista D., Sweet T., Young J.-G., Koster J., Ross C. T., McElreath R., Redhead D., & Power E. A. (2023). Latent network models to account for noisy, multiply reported social network data. *Journal of the Royal Statistical Society Series A: Statistics in Society*, 186(3), 355–375. <https://doi.org/10.1093/jrsssa/qnac004>

Ebrill L. P., Keen M., & Perry V. P. (2001). *The modern VAT*. International Monetary Fund.

Fritz C., De Nicola G., Kevork S., Harhoff D., & Kauermann G. (2023). Modelling the large and dynamically growing bipartite network of German patents and inventors. *Journal of the Royal Statistical Society Series A: Statistics in Society*, 186(3), 557–576. <https://doi.org/10.1093/jrsssa/qnac009>

Gao Y., Shi B., Dong B., Wang Y., Mi L., & Zheng Q. (2021). Tax evasion detection with FBNE-PU algorithm based on PnGCN and PU learning. *IEEE Transactions on Knowledge and Data Engineering*, 35(1), 931–944. <https://doi.org/10.1109/TKDE.2021.3090075>

González P. C., & Velásquez J. D. (2013). Characterization and detection of taxpayers with false invoices using data mining techniques. *Expert Systems with Applications*, 40(5), 1427–1436. <https://doi.org/10.1016/j.eswa.2012.08.051>

Gupta M., & Nagadevara V. (2007). Audit selection strategy for improving tax compliance: Application of data mining techniques. In *Foundations of Risk-Based Audits. Proceedings of the eleventh International Conference on e-Governance, Hyderabad, India, December* (pp. 28–30). Citeseer.

Hand D. J., & Vinciotti V. (2003). Choosing  $k$  for two-class nearest neighbour classifiers with unbalanced classes. *Pattern Recognition Letters*, 24(9–10), 1555–1562. [https://doi.org/10.1016/S0167-8655\(02\)00394-X](https://doi.org/10.1016/S0167-8655(02)00394-X)

Hsieh F., & Turnbull B. W. (1996). Nonparametric and semiparametric estimation of the receiver operating characteristic curve. *Annals of Statistics*, 24(1), 25–40. <https://doi.org/10.1214/aos/1033066197>

Jackson M. O., & Wolinsky A. (1996). A strategic model of social and economic networks. *Journal of Economic Theory*, 71(1), 44–74. <https://doi.org/10.1006/jeth.1996.0108>

James G., Witten D., Hastie T., & Tibshirani R. (2013). *An introduction to statistical learning* (Vol. 112). Springer.

Keen M., & Lockwood B. (2010). The value added tax: Its causes and consequences. *Journal of Development Economics*, 92(2), 138–151. <https://doi.org/10.1016/j.jdeveco.2009.01.012>

Keen M., & Smith S. (2006). VAT fraud and evasion: What do we know and what can be done? *National Tax Journal*, 51(4), 861–887. <https://doi.org/10.17310/ntj.2006.4.07>

Kleanthous C., & Chatzis S. (2020). Gated mixture variational autoencoders for value added tax audit case selection. *Knowledge-Based Systems*, 188(10), Article 105048. <https://doi.org/10.1016/j.knosys.2019.105048>

Kolaczyk E. D., & Csárdi G. (2014). *Statistical analysis of network data with R* (Vol. 65). Springer.

Lazer D., Pentland A. S., Adamic L., Aral S., Barabasi A. L., Brewer D., Christakis N., Contractor N., Fowler J., Gutmann M., Jebara T., King G., Macy M., Roy D., & Van Alstyne M. (2009). Life in the network: The coming age of computational social science. *Science (New York, NY)*, 323(5915), 721. <https://doi.org/10.1126/science.1167742>

Lazer D. M., Pentland A., Watts D. J., Aral S., Athey S., Contractor N., Freelon D., Gonzalez-Bailon S., King G., Margetts H., Nelson A., Salganik M. J., Strohmaier M., Vespignani A., & Wagner C. (2020). Computational social science: Obstacles and opportunities. *Science*, 369(6507), 1060–1062. <https://doi.org/10.1126/science.aaz8170>

Li T., Lei L., Bhattacharyya S., den Berge K. V., Sarkar P., Bickel P. J., & Levina E. (2020). Hierarchical community detection by recursive partitioning. *Journal of the American Statistical Association*, 117(538), 951–968. <https://doi.org/10.1080/01621459.2020.1833888>

Li T., Levina E., & Zhu J. (2020). Network cross-validation by edge sampling. *Biometrika*, 107(2), 257–276. <https://doi.org/10.1093/biomet/asa006>

Ma X., Wu J., Xue S., Yang J., Zhou C., Sheng Q. Z., Xiong H., & Akoglu L. (2021). A comprehensive survey on graph anomaly detection with deep learning. *IEEE Transactions on Knowledge and Data Engineering*, 35(12), 12012–12038. <https://doi.org/10.1109/TKDE.2021.3118815>

Malliaros F. D., & Vazirgiannis M. (2013). Clustering and community detection in directed networks: A survey. *Physics Reports*, 533(4), 95–142. <https://doi.org/10.1016/j.physrep.2013.08.002>

Margetts H., & Dorobantu C. (2019). Rethink government with AI.

Matos T., de Macedo J. A. F., & Monteiro J. M. (2015). An empirical method for discovering tax fraudsters: A real case study of Brazilian fiscal evasion. In *IDEAS '15: Proceedings of the 19th International Database Engineering & Applications Symposium*, 41–48. <https://doi.org/10.1145/2790755.2790759>

Menon A., Narasimhan H., Agarwal S., & Chawla S. (2013). On the statistical consistency of algorithms for binary classification under class imbalance. *Proceedings of Machine Learning Research*, 28(3), 603–611. <https://proceedings.mlr.press/v28/menon13a.html>

Merris R. (1994). Laplacian matrices of graphs: A survey. *Linear Algebra and Its Applications*, 197, 143–176. [https://doi.org/10.1016/0024-3795\(94\)90486-3](https://doi.org/10.1016/0024-3795(94)90486-3)

Mihail M., & Papadimitriou C. (2002). On the eigenvalue power law. In *Randomization and Approximation Techniques in Computer Science: 6th International Workshop, RANDOM 2002 Cambridge, MA, USA, September 13–15, 2002 Proceedings* 5 (pp. 254–262). Springer.

Murphy K. P. (2012). *Machine learning: A probabilistic perspective*. MIT Press.

Newman M. E. (2012). Communities, modules and large-scale structure in networks. *Nature Physics*, 8(1), 25–31. <https://doi.org/10.1038/nphys2162>

Ng A., Jordan M., & Weiss Y. (2001). *Advances in Neural Information Processing Systems*, 14.

OECD (2022). *Tax Administration 2022*, p.105, Box 6.3.

Olhede S. C., & Wolfe P. J. (2014). Network histograms and universality of blockmodel approximation. *Proceedings of the National Academy of Sciences*, 111(41), 14722–14727. <https://doi.org/10.1073/pnas.1400374111>

Pillinger R., Steele F., Leckie G., & Jenkins J. (2024). A dynamic social relations model for clustered longitudinal dyadic data with continuous or ordinal responses. *Journal of the Royal Statistical Society Series A: Statistics in Society*, 187(2), 338–357. <https://doi.org/10.1093/jrsssa/qnab115>

Pourhabibi T., Ong K.-L., Kam B. H., & Boo Y. L. (2020). Fraud detection: A systematic literature review of graph-based anomaly detection approaches. *Decision Support Systems*, 133(4), Article 113303. <https://doi.org/10.1016/j.dss.2020.113303>

Priebe C. E., Park Y., Vogelstein J. T., Conroy J. M., Lyzinski V., Tang M., Athreya A., Cape J., & Bridgeford E. (2019). On a two-truths phenomenon in spectral graph clustering. *Proceedings of the National Academy of Sciences*, 116(13), 5995–6000. <https://doi.org/10.1073/pnas.1814462116>

Qin T., & Rohe K. (2013). Regularized spectral clustering under the degree-corrected stochastic blockmodel. In *Advances in neural information processing systems* (pp. 3120–3128).

R Core Team (2021). *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing.

Satuluri V., & Parthasarathy S. (2011). Symmetrizations for clustering directed graphs. In *Proceedings of the 14th International Conference on Extending Database Technology* (pp. 343–354).

Savić M., Atanasićević J., Jakovetić D., & Krejić N. (2022). Tax evasion risk management using a hybrid unsupervised outlier detection method. *Expert Systems with Applications*, 193(1), Article 116409. <https://doi.org/10.1016/j.eswa.2021.116409>

Shah J. (2021). Using computerized information to enforce vat: Evidence from pakistan. *Mimeo, University of Oxford*.

Subelj L., Furlan S., & Bajec M. (2011). An expert system for detecting automobile insurance fraud using social network analysis. *Expert Systems with Applications*, 38(1), 1039–1052. <https://doi.org/10.1016/j.eswa.2010.07.143>

Vanhoeyveld J., Martens D., & Peeters B. (2020). Value-added tax fraud detection with scalable anomaly detection techniques. *Applied Soft Computing*, 86(4), Article 105895. <https://doi.org/10.1016/j.asoc.2019.105895>

Van Vlasselaer V., Eliassi-Rad T., Akoglu L., Snoeck M., & Baesens B. (2017). Gotcha! network-based fraud detection for social security fraud. *Management Science*, 63(9), 3090–3110. <https://doi.org/10.1287/mnsc.2016.2489>

Waseem M. (2023). Overclaimed refunds, undeclared sales, and invoice mills: Nature and extent of noncompliance in a value-added tax. *Journal of Public Economics*, 218(1–2), Article 104783. <https://doi.org/10.1016/j.jpubeco.2022.104783>

Wu R.-S., Ou C.-S., Lin H.-y., Chang S.-I., & Yen D. C. (2012). Using data mining technique to enhance tax evasion detection performance. *Expert Systems with Applications*, 39(10), 8769–8777. <https://doi.org/10.1016/j.eswa.2012.01.204>

Xu M. (2021). Understanding graph embedding methods and their applications. *SIAM Review*, 63(4), 825–853. <https://doi.org/10.1137/20M1386062>

Zhang S., Wong H.-S., & Shen Y. (2012). Generalized adjusted rand indices for cluster ensembles. *Pattern Recognition*, 45(6), 2214–2226. <https://doi.org/10.1016/j.patcog.2011.11.017>