Visual Informatics 2 (2018) 198–212

Contents lists available at ScienceDirect

Visual Informatics

journal homepage: www.elsevier.com/locate/visinf

Visual analytics for event detection: Focusing on fraud

Roger A. Leite [a,*], Theresia Gschwandtner [a], Silvia Miksch [a], Erich Gstrein [b], Johannes Kuntner [c]

[a] Vienna University of Technology, Austria
[b] s IT Solutions AT Spardat GmbH, Austria
[c] Erste Group IT International, Austria

ARTICLE INFO

Article history:
Received 5 November 2018
Accepted 27 November 2018
Available online 5 December 2018

Keywords:
Visual knowledge discovery
Time series data
Business and finance visualization
Financial fraud detection

ABSTRACT

The detection of anomalous events in huge amounts of data is sought in many domains. For instance, in the context of financial data, the detection of suspicious events is a prerequisite to identify and prevent attempts to defraud. Hence, various financial fraud detection approaches have started to exploit Visual Analytics techniques. However, there is no study available giving a systematic outline of the different approaches in this field to understand common strategies but also differences. Thus, we present a survey of existing approaches of visual fraud detection in order to classify different tasks and solutions, to identify and to propose further research opportunities. In this work, fraud detection solutions are explored through five main domains: banks, the stock market, telecommunication companies, insurance companies, and internal frauds. The selected domains explored in this survey were chosen for sharing similar time-oriented and multivariate data characteristics. In this survey, we (1) analyze the current state of the art in this field; (2) define a categorization scheme covering different application domains, visualization methods, interaction techniques, and analytical methods which are used in the context of fraud detection; (3) describe and discuss each approach according to the proposed scheme; and (4) identify challenges and future research topics.

1. Introduction

The detection of events is an important task in many domains such as detecting interesting changes in stock markets, detecting problems in health parameters, or detecting financial fraud. Analyzing these events in a temporal context fosters further insights such as frequency, trends, and changes. While event detection is aimed at identifying any type of event (not necessarily anomalous events), outlier detection focuses on patterns and samples that do not conform to expected behavior, i.e. anomalies or outliers (Chandola et al., 2009). Once an event is identified it can be classified, which is usually done in a domain-specific way. In the case of fraud detection, we are interested in events that could be classified as fraud. For instance, the purchase of a car might cause the transaction of a high amount of money that is unusual for the respective bank account, and thus, it might be classified as an outlier transaction. Such transactions of high amounts of money require special attention before being executed in order to avoid fraudulent schemes such as hackers trying to use someone else's credit card for their own benefit. However, not every fraudulent event can be classified as an outlier, sometimes attacks are hidden in known patterns in order to avoid detection by simple rule-based approaches. The well renowned Oxford Dictionary defines fraud as "wrongful or criminal deception intended to result in financial or personal gain".[1] However, in this survey we focus on fraudulent events of different applications domains that have direct financial impact on a person or an institution, and all data necessary to identify this fraud is electronically processed. Hence, we do not address types of fraud and data that are not directly related to financial loss such as malicious events identified in unstructured text or malware traffic in system networks, which are largely covered in other surveys (Wanner et al., 2014; Wagner et al., 2015).

The domains we consider in this survey share similar data characteristics. For example, monitoring calls, stocks, bank transactions, as well as employees are tasks that involve data with multivariate and time-oriented aspects. Both aspects require sophisticated exploration techniques and, by consequence, are subject of interest to the Visual Analytics (VA) community. In

* Corresponding author.
E-mail addresses: roger.leite@tuwien.ac.at (R.A. Leite), gschwandtner@ifs.tuwien.ac.at (T. Gschwandtner), miksch@ifs.tuwien.ac.at (S. Miksch), Erich.Gstrein@erstegroup.com (E. Gstrein), Johannes.Kuntner@s-itsolutions.at (J. Kuntner).

Peer review under responsibility of Zhejiang University and Zhejiang University Press.

[1] http://www.oxforddictionaries.com/definition/english/fraud (accessed January 4, 2017).

https://doi.org/10.1016/j.visinf.2018.11.001

addition, Kielman et al. (2009) describe fraud detection as an open VA problem that requires visual exploration, discovery, and analysis. However, many of the current solutions involve mainly data mining techniques. VA approaches have the potential to improve these solutions by integrating human analysis into the process by means of visual representations and interaction techniques (Keim et al., 2008). Despite that, VA approaches are barely explored in the field of fraud detection.

Besides its challenging nature, visual fraud event detection has also a strong social and financial importance. For instance, fraudulent schemes such as 'money laundering', or 'straw person' should be detected and fought as fast as possible by financial systems. Governments, banks, and other financial institutions that provide credit and money transaction services are always interested in improving operation monitoring and fraud detection. Software environments handling sensitive data such as financial operation management systems, systems for insurance evaluation, or companies' internal control systems, need to be in constant evaluation to detect ever-changing fraudulent attempts, to provide risk management, and, thus, to avoid catastrophic consequences.

In this article, we present a survey of existing visualization techniques used for fraud detection in different application domains. We describe and discuss each approach according to our categorization scheme, covering application domains, visualization methods, utilized interaction techniques, and analytical methods. Finally, we elaborate on the benefits and shortcomings of these approaches and identify open challenges and future research directions.

2. Literature research

The scope of our literature research was defined by these three instances: keywords, time period, and publication media and databases. We used the following attributes and databases to search for relevant work:

Keywords used in the search were: "fraud visualization", "visual analytics", "visualization", "visual mining", "information visualization". The keywords were searched individually ('OR') and with different arrangements ('AND').

Time period. The definition of the time range was set to the period from 1997 to the year of 2018.

Publication media and databases that were used: IEEE TVCG, VIS, VizSec, EuroVis, PacificVIS, Information Visualization (published by SAGE), Computer & Graphics (published by Elsevier), EuroVA, ABI Database, Academic Search Premier, ACM, Business Source Premier, Emerald Full text Science, and Google Scholar.

Aiming for a better understanding of the fraud detection scenario, we did not only include papers which follow a pure VA approach in this survey. We also included articles from target application domains which utilize visualization techniques as part of their solution.

2.1. Data, user, and tasks

In this section, we define our survey target studies with respect to the data, users, and tasks (Miksch and Aigner, 2014).

Data. All selected approaches tackle multivariate and time-oriented aspects in their data set(s).

User. The actual users of the selected approaches vary with the application domain, but they share similar tasks. With 'user' we refer to the person who is in charge of identifying fraudulent attempts within multivariate time-oriented data. Possible users are: investigators (Chang et al., 2008), market makers (Kirkland et al., 1999), business users (He et al., 2003), analysts (Didimo et al., 2011; Huang et al., 2009), and others.

Task. The common task is to identify fraudulent events within multivariate time-oriented data. Subsequent tasks depend on the application domain and may include stopping the fraudulent behavior in order to avoid future financial damage.

2.2. Contributions

In our survey, we intend to guide and motivate future research in the field of event detection. Our main contributions are:

- a classification of existing approaches with respect to application domains, visualization methods, interaction techniques, and analytical method;
- a brief presentation of each approach;
- a comparative assessment of these approaches;
- the identification of open challenges and possible future research directions in the field;

3. Related work

Temporal event detection is a vast subject. It is relevant in different fields, such as: biology, security, finances, sales, social networks, and disease monitoring. One recent example is the survey provided by Atefeh and Khreich (2013) that presents techniques for event detection from Twitter streams. The authors discuss the problem of analyzing Twitter content, and they classify the existent techniques by event type, detection task, and detection method. Guided by text stream visualizations, another example of an event detection survey is presented by Šilić and Bašić (2010). In this article, a new aspect of method comparison by data type, text representation, and temporal drawing approach are presented.

In 2002, Bolton and Hand (2002) published a review about fraud detection approaches. They described the available tools for statistical fraud detection and identified the most used technologies in four areas of fraud detection: credit card fraud, money laundering, telecommunication fraud, and computer intrusion. In the same sense, Kou et al. (2004) presented a survey about techniques for identifying the same types of fraud as in Bolton and Hand (2002). The techniques are classified according to the different fraud detection types. Some of the techniques described are: outlier detection, neural networks, expert systems, model-based reasoning, data mining, state transition analysis, and information visualization.

Surveys that focus specifically on data mining techniques for fraud detection research were conducted, for instance, by Ngai et al. (2011), who present a classification scheme for data mining techniques. Phua et al. (2010) formalize the main types and subtypes of known fraud. Sharma and Panigrahi (2013) not only classify data mining techniques, but also propose a framework for fraud detection data mining techniques. The survey presented by Sithic and Balasubramanian (2013) is focused on using data mining for insurance fraud detection.

When looking on surveys of visual approaches for fraud detection, we identified FinanceVis (Dumas et al., 2014) which is a browser for searching papers related to financial data visualization. More than 85 papers are integrated in the browser.[2] "Data

[2] http://financevis.net (accessed February 4, 2016).

visualization for fraud detection", by Dilla and Raschke (2015), is perhaps the most recent work that tackles fraud detection with VA. This paper presents a theoretical framework to predict when and how the analysts should apply VA techniques. They evaluated various visualization techniques and concluded that different visualizations support different cognitive processes. In addition, the authors also suggest future challenges for this research area.

Ko et al. (2016) presented a survey of visualizations and VA approaches for exploring financial data in general. Financial data experts were interviewed concerning their preferences of data sources, automated techniques, visualizations, and interaction methods. Despite presenting many event detection works, this survey does not cover any fraud detection approach. The main goal of this survey is to support researchers with designing better systems to reach dedicated goals.

In summary, the existing surveys do not tackle our overall visions and needs: On the one hand, they are mainly data mining-oriented and utilize visualization only as visual aid for input and output data. On the other hand, the more visualization-oriented surveys among them are very general from the application point view, and neglect the specific characteristics of temporal event detection. Finally, despite focusing on event detection, the text and Twitter-oriented surveys focus on totally different data types, namely text, documents, and document collections. In contrast to that, our survey is oriented towards the particular nature and characteristics of financial fraud event detection in multivariate time-oriented data. The main propose of our survey is to analyze financial fraud detection in application domains with similar characteristics (e.g., insurance data registers, bank transactions, telecommunication companies, stock market logs, and companies' internal systems), which was not tackled yet, and present an overview of the area.

4. Categorization scheme

Based on the "Visual Analytics: Definition, Process and Challenges" book (Keim et al., 2008) we categorize the existing work with respect to three aspects: supported application domains, visualization methods, and interaction techniques. The classification of fraud detection papers based on their main application domains facilitates future comparisons and trends in the fraud detection area. Methods that represent data visually are one of the core proposes of the VA area (Miksch and Aigner, 2014). As a consequence, we also classify the found approaches with respect to the chosen visualizations. One of the core aspects of VA is the combination of visualization and human factors. Thus, by classifying interaction techniques we support a better understanding of how analysts interactively explore the data to gain insights about fraudulent events.

4.1. Supported application domains

We characterize the identified application domains together with their specific tasks.

Telecommunication fraud detection. Although the percentage of fraudulent mobile usage is small with respect to the omnipresence of mobile telecommunication, the costs of these frauds are significant. There are different types of telecommunication fraud that cause enormous harm. For instance, cloned chips not only cause financial injury for the company, but also considerately harm the reputation of the company since the company could not keep the information of clients safe. In this respect, the domains of telecommunication, financial health, and public health are of similar nature.

Stock Market fraud detection. Stock market, or sharing market, is the network of economic transactions made by sellers and buyers concerning shares of the ownership of companies. NASDAQ (NASDAQ Stock Market, 0000) is a well known example of a stock market. Frauds such as late-trade report, anti market integrity behavior, and best execution beat are described by Kirkland et al. (1999). Not only well defined frauds were considered in this work. New pattern recognition and pattern changes can also lead to the identification of suspicious behavior. Therefore, they also included solutions involving pattern detection.

Insurance fraud detection. Fraudulent acts in the insurance sector usually include a malicious analyst who simulates some damage, alternates data, or conducts any other kind of fraud. However, the identification of such acts is a hard task, tackled only by a small number of approaches.

Bank fraud detection. Financial management systems need to be in constant evaluation to avoid frauds and to provide risk limitation. Fraudulent schemes such as 'money laundering', or 'straw persons' must be identified and fought as fast as possible. Thus, banks, and other financial institutions that provide credit and money transaction services have a strong interest in improving operation monitoring and fraud detection systems.

Internal fraud detection. Securing information is a critical task within companies. To sell or edit confidential data from companies usually causes severe damage. Also known as "occupational fraud", this type of fraud results not only in straight economic loss, but also harms the reputation of the company among employees, clients, and financial institutions. When it comes to companies internal access violation, identifying the person who committed the fraud is difficult, since these frauds could also be committed by people who do not belong to the company (e.g., crackers). However, in most cases, it is done by an insider employee. These internal attacks can occur by an employee that has privileged credential access, an employee that has access to someone else's credential, or an employee who hacks the system.

4.2. Visualization methods

Visualization techniques take advantage of the human perception system and allow analysts to more easily derive insights about data. For instance, instead of exhaustively looking into tables to identify data characteristics, an analyst is able to see, explore, and understand a large amount of information by using visualization techniques. However, the efficiency of these techniques varies with respect to different tasks. The following list of visualization techniques are used in the identified fraud detection approaches:

Line plots. One of the most popular visual representations – line plots – are graphs that display numerical values along continuous dimensions by using lines. For instance, in Fig. 4(C) each line represents transactions clustered by keyword variations and shows the sum of the respective transactions' amounts (y-axis) during a certain period of time (x-axis). Line plots are well suited for the detection of outliers and to analyze periodic patterns and similarities. Special attention has to be paid to the handling of missing values, to not lead to wrong impressions.

Node-link Diagrams. These diagrams illustrate relations (represented by links) between entities (represented by nodes). The same network, can be represented by different layouts. Some of them are: forced-based layout, spectral layout, layered graph drawing, arc diagrams, circular layout, and dominance drawing. The goal of this technique is to facilitate the understanding of systems and networks by representing usability, costs, flows, and connections. However, data sets that change over the time are hard to represent due to the constant deletion and creation of nodes and links that the technique would impose in some cases. A node-link diagram can be seen in Fig. 1, and in Fig. 4(D).

Bar charts. Rectangles (bars) are used to represent different entities, where the height or the width encodes quantitative values.

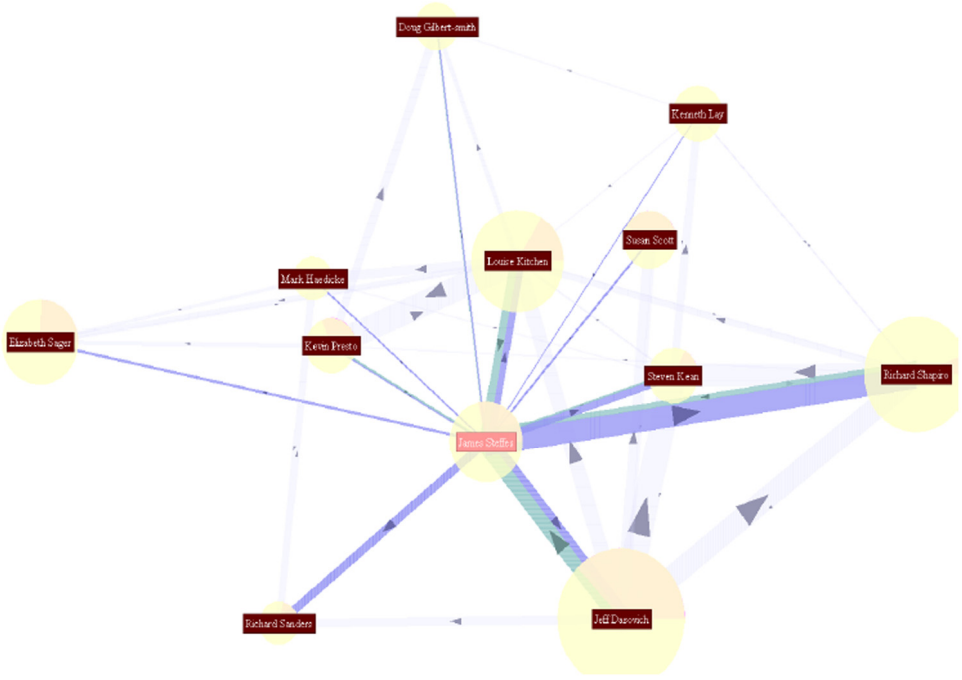

Fig. 1. This node-link diagram from Huang et al. (2009) that represents a trading pattern network. Each node is a trader and each edge encodes a trading relationship.

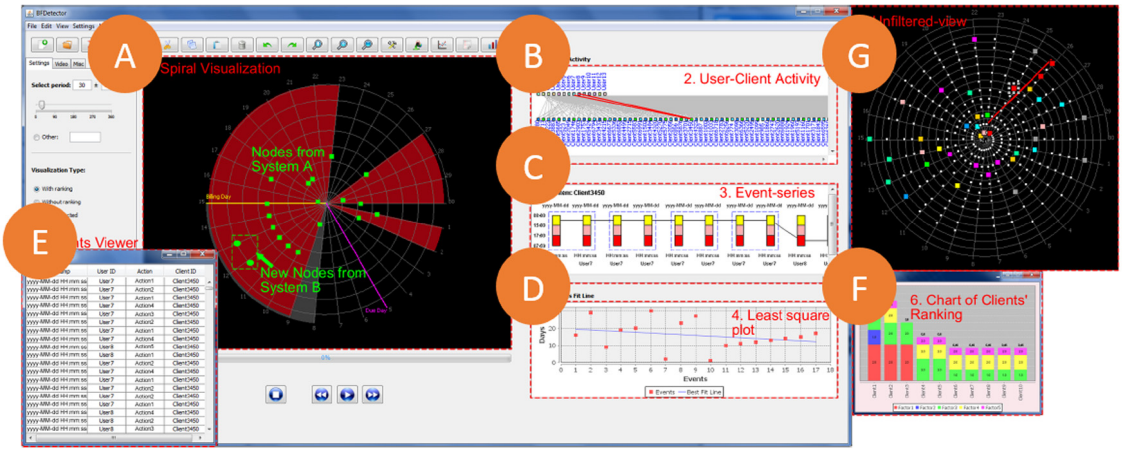

Fig. 2. This is a visualization for internal data analysis proposed by Argyriou et al. (2013). (A) and (G) are radar charts that display periodical patterns. Each spiral represents a month. (B) displays the user–client activities while (C) shows the event times. A mixture of scatter plot and line plot is presented in (D) to determine relations between days of the month and events. View (E) displays raw data, and view (F) shows a stacked bar chart that ranks clients based on the sum of predefined factors.

For example, in Fig. 2(F), we have an example of stacked bar charts that is being utilized for ranking. This visualization technique is well suited to represent relative differences. There are different types of bar charts such as horizontal bar charts, stacked bar charts, and range bar charts that were classified equally in this category.

Scatter plots. These are graphs where each sample is represented by a point or symbol. Each point or symbol position is defined according to two dimensions, or two generated features of these samples. Those graphs are useful to illustrate trends and correlations.

Pixel-oriented Diagrams. These visualizations are well suited for the exploration and analysis of massive data sets. The main idea is to map data objects to pixels in order to be able to represent as many samples as the screen resolution allows (Keim et al., 2000). One example can be seen in Fig. 3(B), each pixel encodes a sample, representing data from a period of 5 years being represented in the same visualization. However, a data item is not limited to be represented by one pixel, for instance if you zoom in the pixel-oriented diagram. This technique is usually used to highlight data clusters, patterns, and outliers. This type of visualization is sensitive to the chosen color palette.

Tree Maps. By using nested rectangles, this technique visualizes hierarchical data. In fraud detection, this visualization technique is mainly used to represent ranking relationships and to categorize 'normal' and 'suspicious' cases.

Heat Maps. This graphical representation represents values by colors. In fraud detection this technique is usually used to visually query for patterns or outliers in a large amount of data. One example of this technique can be seen in Fig. 4(A).

Radar Charts. The idea of radar charts is to display multiple related dimensions in a radial visualization. This allows the analyst to compare quantities. Radar charts can also be used to represent periodical events, as can be seen in Fig. 2(A) and Fig. 2(G).

Parallel Coordinates. This technique uses a similar concept as Radar Charts to display multiple dimensions. However, here the dimensions are laid out side-by-side, and not in a radial format.

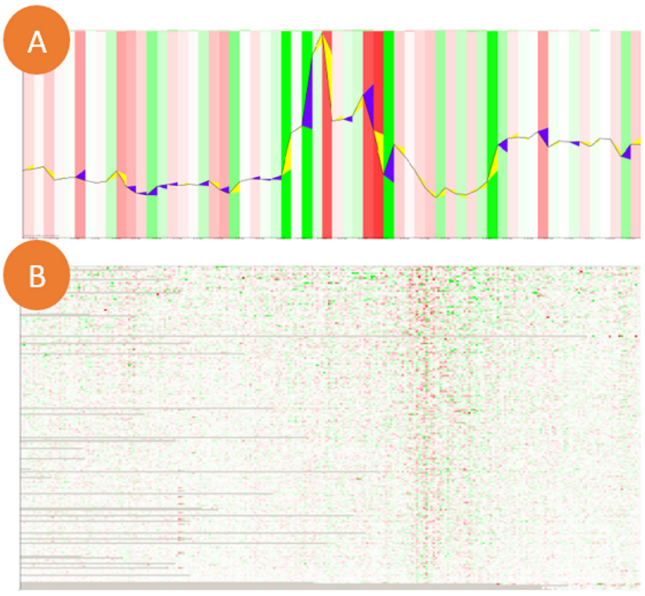

Fig. 3. Visualizations from Schaefer et al. (2011): (A) Polygons visualization for correlation analysis and pattern detection. The correlation is represented by yellow and violet "slope polygons". Positive correlations between the share and the sector index are represented by yellow polygons, while negative correlations are encoded by violet polygons. (B) represents 222 stocks of the oil sector for a period of 5 years.

Unlike radar charts, parallel coordinates are not used to display time.

Box Plots. This technique is used to represent and analyze features of groups of samples. Box Plots represent the median, upper quartiles, and lower quartiles by rectangles. They can also be extended to indicate variability through vertical lines that extend the rectangle box. Box Plots are often used to identify outliers, which are usually represented as dots outside the boxes, or the vertical lines.

Polygons. Proposed in Schaefer et al. (2011), polygons aim to visualize correlations between individual share performances. This visualization technique can be generated by using trapezoids or triangles (see Fig. 3(A)).

3D visualizations. There are quite some fraud detection approaches that use 3D visualizations for their tasks. While 3D visualizations can express an extra dimension if compared to 2D techniques, two characteristics of this type of visualization may confuse the analyst, leading him/her to wrong analysis: occlusion and perspective misunderstanding. Some 2D techniques are also just presented in 3D. For example, Fig. 5(A) is a 3D representation of Fig. 5(B). However, we also found approaches utilizing real 3D representations (see Fig. 7).

4.3. Interaction techniques

When it comes to VA, the interaction technique employed by a solution is a determinant factor. It has a strong influence on how analysts will explore the proposed technique as well as on the usability of the approach. When developing a VA solution, the interaction techniques should be chosen in accordance with the visualization techniques and tasks (Aigner et al., 2011). Determining this set of techniques is a critical task during visualization design. It impacts the quality of the analysts' insights and the efficiency of the solution. We emphasize the selection of the appropriate interaction techniques as a critical task of fraud detection VA projects.

Based on Yi's definition of interaction (Yi et al., 2007) (called users' intents) and the enhancements made by Aigner et al. (2011), we consider:

Selection. An analyst who spots an item or a temporal period of interest can select and highlight this region. Highlighting visual elements is an useful feature during data analysis, and also to aid the explanation of insights.

Exploration. This interaction technique shows extra information about the data. When it comes to large, time-oriented, and multivariate data sets, visualizations that aggregate information are common, but not always justified. For those massive (large data sets), and/or complex (multivariate) representations, analysts need to interactively explore different parts of the data set in order to have a better overview of the content. Examples of exploration usage are visiting, investigating of different time intervals, or changing the visual encoding.

Reconfigure. This interaction displays a different arrangement of data items. A distinct arrangement of the data can highlight very different features. For example, in case a analyst wants to analyze different time-oriented aspects, he or she can arrange the elements in linear time or periodical time.

Encode. By encoding the analyst gets different representations of the same data. This interaction allows the analyst to adapt the visual encoding to suit different tasks. Some data sets have many interesting dimensions to be evaluated. In order to cover multivariate data without increasing the visual complexity of the visualization, the encode technique allows the analyst to address different data dimensions to different visual features each time. For instance, an analyst can use it to verify a hypothesis when looking at the same data in another visual encoding.

Abstract/Elaborate. These are related to the aggregation level of the visualization. We group these two interaction techniques into one category due to their similar nature. Different data representation scenarios are more efficient depending on the task. For some tasks, such as short period analysis, the analyst needs to inspect certain items in detail. To do so, he or she increases the degree of data detail in the visualization. This is called data "elaboration". For other tasks, however, a schematic representation may be sufficient and could lead to faster results. For tasks when this functionality helps or is needed, the interaction works ordering information aggregation to the visualization elements. This is called data "abstraction". Switching between different time granularities is one example of these interaction techniques. Zoom in and zoom out are another example of interaction techniques within this category.

Filter. By filtering, the visualization shows only the data that satisfies a specific condition. This interaction is used when the analyst is searching for specific information in the data, or when he or she is trying to verify a certain hypothesis about the data set. By the usage of filter rules, elements out of interest are excluded from the visual representation. This feature allows the analyst to unclutter the view, and, as a consequence, to focus on the current task. In Fig. 4(B) a bar chart is displayed, which allows the user to find similar events by specifying filter rules based on already existent events.

Connect. This interaction shows or highlights items that are related to each other. For example, when an insight is found in a data set, the analysts can verify if similar or related behavior appears in other parts of the data. This feature is suited to find, compare, and evaluate similarities or relationships. In time-oriented data this interaction technique helps to discover if a pattern is seasonal (has a determined frequency) or if it rather appears at irregular intervals.

Undo/Redo. These techniques allow analysts to return to previous analysis states. We group both in the same category due to their similar nature. During a VA process, analysts have to navigate through different data dimensions, time periods, and levels of granularity. In case a hypothesis or query did not lead to the expected results, this interaction helps the analysts to return to an

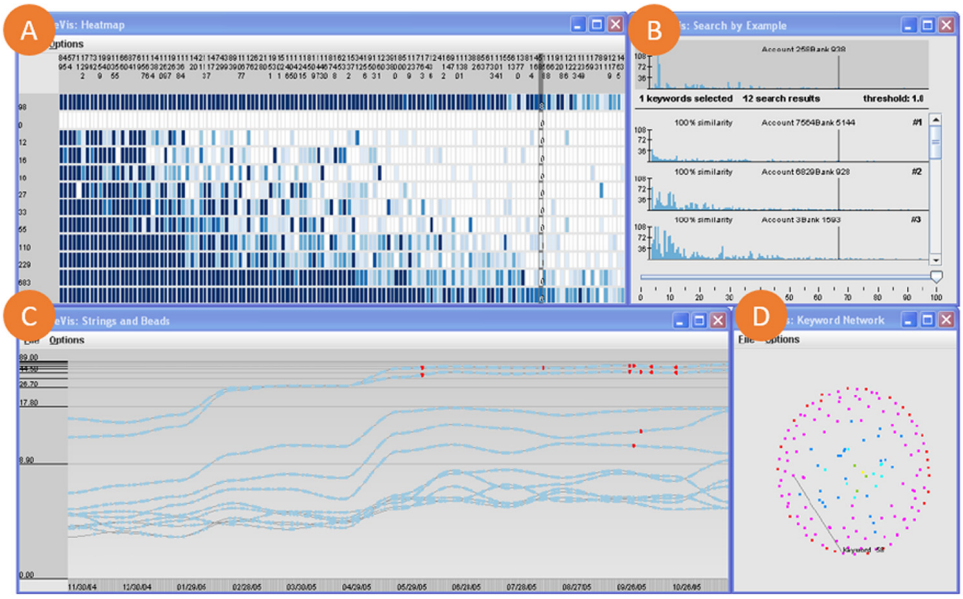

Fig. 4. WireVis (Chang et al., 2007): (A) shows a heat map that reveals relationships between accounts and keywords, (B), search by example, is a bar chart view that allows the user to select an event and filter data by similarity, (C) shows the dimension of time (x-axis) and transactions amount (y-axis) by using a line plot where each string represents a cluster of accounts, and (D) a keyword network view that shows relationships between keywords.

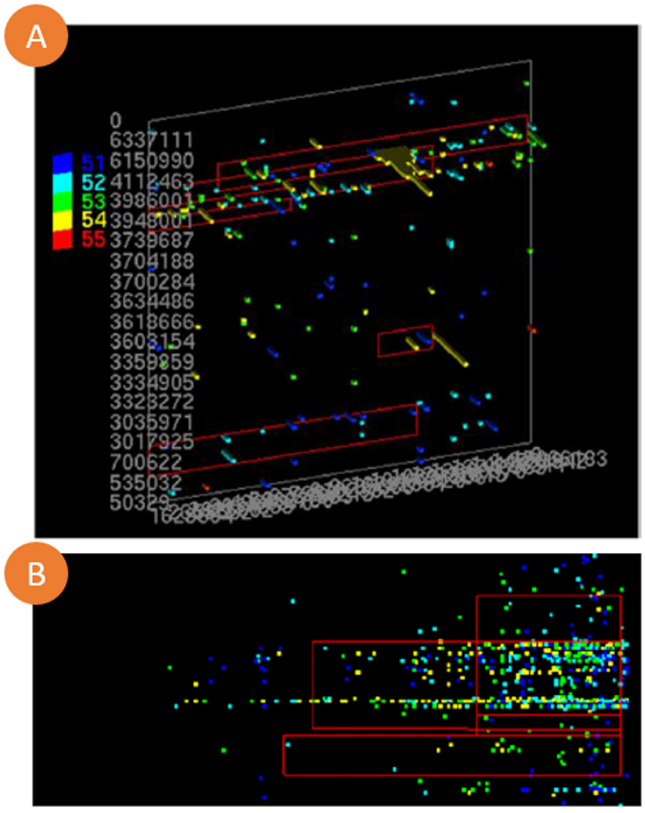

Fig. 5. For both visualizations, the x-axis encodes time, the y-axis encodes credit codes, and colors encode fraud types (Sakoda et al., 2010). In (A), the height of the colored bars refers to the amount of money in the transaction. In (B), we see a similar plot, but in 2D instead of 3D.

early state without losing time by being forced to deconstruct the scenario.

Change Configuration. This technique allows the analyst to adjust the interface. Besides adapting the visualization to the tasks and data at hand, the analyst may also want to configure the overall system that provides the visual representation. This interaction may affect various levels of the system, from "set the windows arrangement", until "manipulate the amount of memory to be used".

No Interaction. An example of visualizations without interaction is a classical static visualization. In this classification, we consider all visualization techniques that do not provide interactive features.

4.4. Analytical methods

In Keim's model (Keim et al., 2008), analytical methods are considered in parallel with interactive visualizations. Analytical methods vary a lot between the different approaches, however, we identified different ways of combining visualizations with analytical methods.

Pre-Processing. Approaches that first utilize analytical methods such as automatic algorithms of search, statistical methods, detection, or clustering, and subsequently use visualizations to present the results.

Post-Processing. Approaches that use visualization techniques in order to understand, select, or observe the data before applying an analytical method.

Integrated. Approaches that allow a 'back-and-forth' between visualization techniques and automatic analysis.

Pure Visualizations. Some approaches do not integrate any form of automatic analysis. Those approaches visualize the raw data without providing analytical methods for further analysis (in this report we do not consider approaches that do not provide any kind of visualization).

5. Approaches summary

We structure this section according to the application domains that are tackled. These are (1) telecommunication, (2) stock market, (3) insurance, (4) bank, and (5) internal fraud. For each application domain we outline existing approaches related to VA in fraud detection. It is important to highlight that, besides being organized by application domains, some approaches are hybrid, which means they belong to more than one category.

5.1. Telecommunication

Adaptive Fraud Detection (Fawcett and Provost, 1997) is one of the pioneer papers in fraud detection from 1997. This technique is considered hybrid, because it is tailored to bank fraud and telecommunication fraud detection. Moreover, it is a data mining-oriented and uses line plots during the analysis process. Using a rule-learning algorithm, a set of monitors is generated to profile legitimate customer behavior.

By using a node-link diagrams, bar charts, and line charts, Cox et al. (1997) propose group analysis as well as individual analysis for telecommunication fraud detection. This is one of the first works which use VA techniques in the context of fraud detection. The main idea is to build visual interfaces that allow for the exploration of the data. Clustering techniques are used to improve the node-link visualizations during interaction. This is the only work from the telecommunication domain that allows for an integrated visual and automated analysis.

Hollmén and Tresp (1999) present an online fraud detection system based on a hierarchical regime-switching generative model. In this paper line plots support analysts to determine the probability of detection and of false alarms. With this information, the analysts are able to decide which alarm is worth further effort, and which alarm can be discarded. The methods were developed by using and analyzing real mobile communication network data.

Hilas and Sahalos (2005), present an approach to fraud detection in telecommunication based on a machine learning method that generates user profiling. Further analysis is done by using line plots in order to compare different user profiles and, thus, identify strange behaviors.

Becker et al. (2012) present a review of the history of fraud detection from a big company. The authors also describe classes of fraud in the domain, and propose VA models that support fraud detection of in each of different classes.

The most recent work found in this domain was (Manunza et al., 2017). The authors present the design and implementation of Kerberos, a system to detect frauds over Voice over IP networks. This work is a rather analytical approach that aims for real-time detection of frauds. Kerberos allows the construction of pre-defined detection rules and the configuration of alarms. This work was experimentally evaluated using real-world data and presents good performance with different configurations.

Besides (Manunza et al., 2017), we could not find many recent papers dealing with fraud detection in the telecommunication domain. Although the major part of the selected papers from the telecommunication domain being not recent, we still consider the approaches to be relevant in this context, and thus, we include them in this survey.

5.2. Stock market

Using 2D and 3D visualization techniques, ADS (Kirkland et al., 1999) combines feature discovery and fraud detection for market data. This approach uses diverse visualizations, interaction techniques, and data mining techniques. Its main objectives are: regular monitoring of the stock market, pattern detection, generating alerts for suspicious cases, and knowledge discovery concerning the stock market transactions.

Aiming to aid traders to find trading patterns in market data, Nesbitt and Barrass (2004) use visualizations (i.e., 3D plots, bar charts, and line plots) to look for patterns in stock market data. Besides not being a fraud detection oriented article, this solution fits to the fraud detection problem and, due to that, was considered for this survey.

Merino et al. (2006) present an empirical study of five different visualization techniques for stock market data. Their study advises that line plots and recursive patterns are better suited for retrospective data analysis, but pixel techniques are also useful to find patterns in large data sets.

Huang et al. (2009) present a new VA approach for stock market security. They describe a two-stage process with each stage utilizing different visualization techniques: (1) 3D tree maps for monitoring market performance, and (2) node-link diagrams for behavior driven analysis of trading networks (see Fig. 1). This approach may also help to identify future fraud plans.

Schaefer et al. (2011) support an interactive analysis of financial data that contains the stock prices variation over a long period of time, and sector indices information. Line plots coupled with a proposed visualization called polygons aim to explore patterns and trends (see Fig. 3(A)). Pixel-based visualizations are also used to visually explore large amounts of data (see Fig. 3(B)). The proposed tool aids the analysts in detecting frauds, analyzing performances of the historical stock price, and decision making in the financial market. The main contribution of the approach is the integration of different views and the design of two new visualizations. However, the solution does not provide any automatic algorithms to aid the identification of interesting events. Thus, we classify this work as pure visualization approach.

When using node-link diagrams to visualize large entity-relationship data sets, some scalability problems such as visibility, usability, and high degree of nodes are likely to appear. Gauldie et al. (2013) propose a clustering algorithm to support these scenarios by using visual aggregation techniques and easily tailorable components.

Bitcoin is considered a currency by many, even being a digital one. On stock markets and forex markets, its price is driven by supply and demand. An anomaly detection approach is proposed by Pham and Lee (2016) in order to prevent bitcoin owners from loss. The analysis of the results from three unsupervised learning methods (K-means, Mahalanobis distance, and Unsupervised Support Vector Machine) is supported by two different types of graphs (line plots and scatter plots). Also aiming to detect suspicious activities on the bitcoin market, (Monamo et al., 2016) used k-means and trimmed k-means in combination with scatter plots in order to improve the detection rate.

Fraud detection is not the only task tackled by the major part of the papers in this domain. These works usually focus on monitoring and querying for known patterns that also aim to identify new behaviors. Once identified, patterns are interpreted and classified as suspicious or not.

5.3. Insurance

Artís et al. (2002) use a Spanish automobile insurance market database to indicate the performance of binary choice models for fraud detection. In this work, line plots are used to estimate the probability of fraud for particular types of claims. This paper not only estimates frauds, but also presents an estimation of the probability of detection efficiency of the proposed model.

He et al. (2003) presented a cluster-based local outlier detection system. They proposed an algorithm for discovering outliers which outperformed the existing methods. This work presents scatter plots in order to visualize the different clusters that are found by the algorithm (i.e., a pre-processing approach). This is a flexible technique suited to discover criminal activities in electronic commerce, credit cards, marketing, and customer segmentation. Phua et al. (2004), explore three existing classification algorithms and suggest a new hybrid solution using automatic algorithms for fraud detection by using meta-learning. Bar charts and line plots are used to assist in the comparison of the different techniques.

Furlan and Bajec (2008) present a holistic approach focused on activities of fraud management: deterrence, prevention, detection,

investigation, sanction, redress, and monitoring. This work is based on a health care data set. Bar charts are used in this work to analyze the time spent on each specific task.

Šubelj et al. (2011) use social network analysis data in order to detect automobile insurance fraud. The solution uses node-link diagrams to describe fraudulent networks. When a new fraud is detected, the proposed system allows self-calibration in order to adapt to new fraudulent schemes. The authors suggest that the system could also be used in other domains.

Yi, et al. (Sun et al., 2014) detect health insurance fraud by using a data set about medical expense in China. Besides proposing a discrete choice model to identify predictive factors of fraudulent claims, this paper also addresses limitations of using the discrete choice model. In this paper, line plots are used to describe the influence of different metric specifications on the proposed fraud detection algorithm.

A VA approach is proposed by Cheng et al. (2017) for loan guarantee network risk management. In this work, the authors present five analysis tasks that were defined in collaboration with financial experts and, for solving those, an interactive node-linked diagram. Credit risk evaluation is a major phase during an insurance evaluation and this work supports insurance managers during the decision workflow.

The fraud detection tasks in the insurance domain vary with respect to the insurance type. Most of the selected approaches are based on health care and the automobile industry. To detect fraud, this domain is more focused on analyzing data value variations and outliers rather than analyzing networks. All identified studies from the insurance domain use automated methods before representing the results visually. Thus, they all are classified as pre-processing approaches.

5.4. Bank

Kirkos et al. (2007) explore the performances of data mining classification techniques to identify fraudulent companies. The decision tree of the model is represented by using a binary tree with bar charts inside each node to represent the decision degree.

WireVis (Chang et al., 2007) is focused on detecting frauds in bank transactions, in particular in money laundering. It proposes a set of visualizations displayed as a multiview system that allows the analyst to interactively explore the data. For each account transaction, keywords are analyzed and explored by various visual means, including heat maps, bar charts, line plots, and node-link diagrams. The different views are connected so that selecting a node in the node-link diagram provokes a filter on all other views. A view of the proposed system can be seen in Fig. 4.

One year later, Chang et al. (2008) enhanced WireVis (Chang et al., 2007) by concentrating the examples and results on wire transaction data. Suspicious behavior could be found more easily as well as global trends. Multiview analysis allows analysts to get the whole picture of relationships between accounts, keywords, time, and patterns of activity. Both VA solutions were created in collaboration with Bank of America.

VisForFraud (Di Giacomo et al., 2010) is a VA system for financial crimes identification. In order to discover potential illegal actor networks, the analyst is supported by visualizations and multiple interaction techniques. The actors' network activities are analyzed by using interactions such as selection, elaboration, filtering, and connection on the top of a network diagram visualization.

Due to its similarity with bank fraud, we include credit card frauds in this section. In this context, Sakoda et al. (2010) present a VA tool for assisting rule definition for fraud detection. In this VA solution, colored scatter plots are displayed in 2D and 3D visualization, those can be seen in Fig. 5.

VISFAN (Didimo et al., 2011) is an interactive network visualization system for financial crime detection. Detecting transaction frauds, such as money laundering, is the main task of this system. The VISFAN tries to define the involvement of a certain actor in the network. The tool allows the found cluster region to be customized, using a mix of automatic and manual clustering solutions.

Fig. 6. Parallel coordinates (a), scatter plots (b), and horizontal stacked bar charts (c) are connected by brushing and linking.

VIS4AUI (Didimo et al., 2012) is a system that supports analysts during financial crime analysis, such as money laundering fraud. This system was constructed to use a touchscreen interface. Besides the tool allowing a mix of automatic and manual clustering, the visual exploration of the networks are supported by abstraction and elaboration interaction techniques.

Querying for fraud patterns in credit card transaction data, Seeja and Zareapoor (2014) use a pattern mining technique called "frequent itemset mining" to identify suspicious behavior, and a matching algorithm to compare customers. The idea is to define regular customers behavior in order to analyze strange practices in the future. Multiple line plots are used to measure the algorithm's performance.

Carminati et al. (2014) presented a semi-supervised online banking fraud analysis and decision support based on profile generation and analysis. While this approach provides only one line plot as visual support for result analysis (i.e., classified as pre-processing approach), it has a strong statistical part. The approach is divided into three steps: (1) quantification of the anomaly of each user transaction, (2) define clusters of similar spending habits, and (3) apply a temporal threshold system that measures the anomaly of the current spending pattern. The evaluation of this approach was performed using real-world data and showed that the technique could correctly identify complex frauds.

Leite et al. (2015) proposed a pipeline of multiple connected views for fraud detection and monitoring. They present a VA approach and identify challenges in the field. In this work, the analysts interact with a group of bar charts and gain insights by filtering, selecting, abstracting, elaborating, and exploring the data. Later (Leite et al., 2016), the same authors proposed a multiple coordinated view approach based on the analysis of customer profiles in order to aid financial fraud detection. Based on an automatic scoring system that evaluates each transaction, this approach also allows several interactions with parallel coordinates, a scatter plot matrix, and a bar chart (see Fig. 6). In a follow-up work (Leite et al., 2018), the authors propose EVA, a VA approach to identify fraudulent events based on bank transactions logs. EVA combines a self-adaptive, profile-based detection algorithm with well known visualization techniques. The proposed combination of automatic methods and visual exploration features proved to be efficient for the actual target users using real-world data.

Frauds in the bank domain have the potential to cause huge financial harm. Thus, not only is the detection of such frauds of great

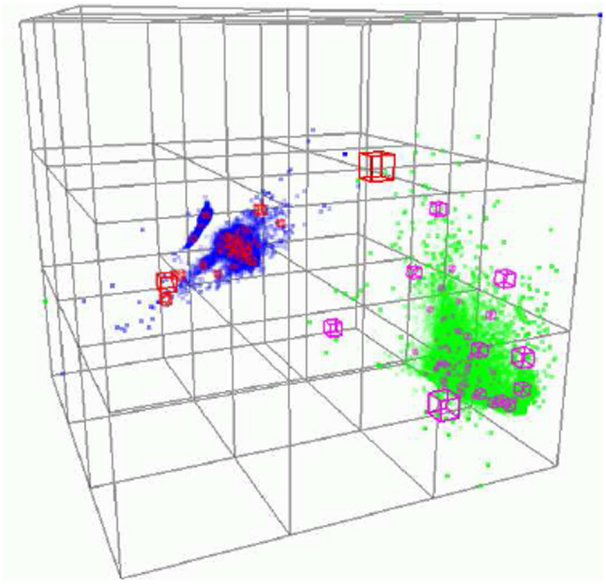

Fig. 7. 3D visualization from Keahey (1999). The screenshot shows a 3D frame visualizing 11 dimensions of a medicare data set. Each discrete point is a data record; clusters' centroids are rendered as a wire-frame cube.

importance, but also monitoring of the data and the prediction of possible fraudulent events. One of the most crucial tasks, in this respect, is the detection of money laundering. This is usually done by analyzing bank transaction data, i.e. an application of network analysis. Another important task is the identification of outliers in order to identify suspicious credit card usage.

5.5. Internal

A multivariate 3D visualization (see Fig. 7) tool to improve fraud detection in medicare programs is presented by Keahey (1999). This paper's main goal is to allow fraud analysts to get better insights from the cluster algorithms that are used during investigation.

Eberle and Holder (2009) present a graph representation approach to detect anomalies in business transactions and processes. Three algorithms are used in the pre-processing phase to aid the identification of anomalies. These assembly techniques successfully discover irregularity in data from different sizes with minimal to no false-positives.

The Ph.D. Thesis of Luell (2010) proposes a decision framework to aid fraud detection approaches. The system is constructed by a detection component and visualization. They demonstrate the potential of the proposed solution by three case studies.

Argyriou et al. (2013) propose an internal fraud detection solution. Despite developed to detect frauds in internal systems, the proposed solution should work in any system that involves relationships of two different entities, for example, client and employee. This paper presents a spiral visualization where events are displayed over time after the automatic pre-processing phase. The analyst exploits the advantages of VA by allowing the analysts to interact with a multiview representation (see Fig. 2) in order to find suspicious events.

A recent publication by Lookman and Nurcan (2015) presents a framework that aims to extend fraud detection data mining techniques with VA approaches. This work is based on social network analysis (SNA). According to the authors, features such as the analysis of hidden network connections, the dynamic adaptation of rules, and the time that can be saved if compared with manual investigation are justifications for SNA usage in fraud detection. The described solution also involves a node-link diagram combined with a semantic reasoning feature that integrates new patterns to the fraud detection engine. The visualization presented in this work is used to represent results of the pre-processing phase.

Webga and Lu (2015) proposed a real-time VA approach for the analysis of stream data to identify "rating frauds" in online e-commerce stores. Based on singular value decomposition, the system generates a pixel-based matrix that can interactively be reordered according to different dimensions during the analysis.

Another VA system for detecting fraud in health care systems was proposed by Liu et al. (2016). Node-linked diagrams were used to identify the relations between doctors, patients, and pharmacies, while a bar chart represented a financial fiscal report. A line plot was used to demonstrate geospatial anomaly between medication buyers and pharmacies. This work led to the identification of (i) patients going to multiple doctors to get narcotics, (ii) doctors prescribing an unusual high amount of medications, and (iii) pharmacies with significant narcotics earnings.

When it comes to internal fraud detection, data and tasks vary among companies. Despite Webga and Lu (2015), the major part of the selected studies present no integration of automatic algorithms and visualizations. Besides, recent works highlight that this application domain has still many open challenges (Luell, 2010; Lookman and Nurcan, 2015).

6. Comparison and findings

In this section we compare the different approaches with respect to visualization and interaction techniques.

With a total of 40 approaches, from 1997 to 2018, six were classified as being part of two or more domains. These are considered "hybrid" approaches. We categorized each of these approaches according to our interpretation.

By analyzing light-colored and dark-colored cells in Table 1 we can determine the popularity of interaction techniques in each application domain. Most of the visualizations presented in articles for insurance fraud detection and internal fraud detection do not support interaction techniques (e.g., they are static). On the other hand, implementation of interaction techniques is widely present in bank and stock market fraud detection.

6.1. Fraud aspects according to domain

When it comes to different application domains, we face different types of fraud (compare Section 5). Bank frauds are usually more network related and include the identification of "payment fraud", "money laundering", and "straw persons". Insurance frauds often are analyzed by checking if a sequence of events is plausible and if fraudulent patterns exist. While internal fraud detection often uses process data mining solutions, telecommunication frauds are tackled with the help of visual rule-based systems. Stock market fraud detection uses a wider range of techniques due to the various types of frauds that can be found in this domain.

6.2. Applied visualization techniques

Considering that fraud techniques are always changing, automated methods of detection are only temporally solutions and tend to fail with time. VA techniques are more robust to these changing conditions as they integrate human perception into the detection process, which is flexible and suited to spot many different kinds of outliers (i.e., suspicious events).

From 40 approaches studied in this survey, the most popular visualization techniques used in fraud detection are line plots (21 appearances), node-link diagrams (16 appearances), and bar charts

Table 1
This table shows the application domain classification for each of the selected papers in this survey. The columns are sorted from left to right according with the number of appearances of each classification. Despite seven articles being hybrid, papers are grouped by application domain. Darker cells represent works that did not implement interaction techniques.

	bank	stock market	internal	insurance	telecommunication	
bank	•					[CGK*07]
	•					[CLG*08]
	•					[SNI*10]
	•					[DGDLP10]
	•					[DLM12]
	•					[SZ14]
	•					[LGM*15]
	•	•				[DLMP11]
	•	•				[GLS*13]
	•					[LGM*16]
	•					[CCM*14]
	•	•	•			[KSM07]
	•					[LGM*18]
stock market		•				[KSH*99]
		•				[NB04]
		•				[MSK*06]
		•				[HLN09]
		•				[SWK*11]
		•				[PL16]
		•				[MMT16]
internal			•			[EH09]
			•			[Lue10]
			•			[ASS13]
			•			[LN15]
			•	•		[Kea99]
			•			[WL15]
			•			[LBW*16]
insurance				•		[AAG02]
	•		•	•		[HXD03]
	•			•	•	[PAL04]
				•		[FB08]
				•		[ŠFB11]
				•		[SYL14]
	•			•		[CNY*17]
telecommunication	•				•	[FB97]
					•	[CEWB97]
					•	[HT99]
					•	[HS05]
					•	[BVW12]
					•	[MMR17]
	17	10	9	8	7	sum

(19 appearances). Line plots are used for presenting a comprehensive view of one or more variables changing over time (e.g., the amount of transactions of different users). Due to the number of networking analysis tasks involved in fraud detection, also node-link diagrams are frequently used in the papers discussed in this survey. Bar charts usually represent numerical values grouped into classes and facilitate the task of visually comparing these groups.

A very interesting observation is, that the most part of the early approaches do not provide interactive features or techniques. One possible interpretation could be the design time of these approaches: in earlier years, interaction techniques were more cumbersome to be designed and implemented than now-a-days. Another interpretation could be that the focus of these approaches is more on the automatic methods than on the interactive visual means.

We also observed that 3D approaches got less popular during the last decade. This could be due to the difficulties to represent and interact with multivariate temporal data in 3D. The main problem with 3D approaches is data occlusion, which may confuse the analyst and make gaining insights difficult. Furthermore, the perception of the 3D views may cause misleading impressions that can lead the analysts to wrong data interpretations.

6.3. Interaction and exploration

Suspicious cases are either identified by automatic methods, such as machine learning algorithms, by human analysts, or by a combination of both. The latter refers to a VA approach where the results of automatic methods are visually supported for further investigation.

Any time the human perception system is involved in the analysis process, interaction techniques reinforce the identification of results as well as the user's learning curve. In this survey, we identified various interaction techniques (categorized according to Section 4) used for fraud detection. Based on Table 3, from 40 papers, abstraction (16 appearances), elaboration (16 appearances), selection (18 appearances), exploration (15 appearances), filtering (12 appearances), and connection (9 appearances) can be found in the most of the approaches. Selection, abstraction, elaboration, and exploration techniques reflect the need for managing huge and complex data sets, while connection and filtering techniques reflect the need for querying for suspicious behavior and outlier profiles.

User interactions are one of the most important elements in visualization or even the "heart" of it as Spence stated (Spence, 2007). User interaction is even more important in Visual Analytics, as studies, like the one by Saraiya et al. (2006) showed: users preferred inferior visualizations with interaction over superior static visualizations. Furthermore, abstract visual representations provide only an initial direction to the data and its meaning, but through the combination of visual representations and appropriate interaction mechanisms, the users achieve insights into the data (Saraiya et al., 2006).

Our research identifies 19 out of 40 approaches that do not offer any type of interaction. Even though the success of interaction techniques that was demonstrated in different studies (Saraiya et al., 2006; Spence, 2007), multiple approaches selected in this survey do not support them. Approaches that miss this feature are mainly based on executing algorithm techniques and represent the results by visualization techniques. Some of the algorithmic approaches used for fraud detection are: outlier detection, self-organizing maps, neural networks, Bayesian classifier, support vector machines, artificial immune systems, fuzzy systems, genetic algorithms, K-nearest neighbor, and hidden Markov models.

Approaches without interaction did also lead to some success. However, recent papers are proposing more intertwined approaches of various techniques, such as VA solutions. They support interaction techniques coupled with visualizations and automatic methods. The combination of these features increases the fraud detection system precision and efficiency, and thus, leads to better results.

6.4. Visual and interactive combinations

We constructed a heat map table (see Table 4) in order to analyze the relation between visualization techniques and interaction techniques in fraud detection scenarios.

Table 2
This table shows the visualization techniques presented in each of the selected papers in this survey. The columns are sorted from left to right according to the number of approaches that fall into each class. Despite seven articles being hybrid, papers are grouped by application domain. Darker cells in the table represent visualizations that do not provide interaction techniques.

	line plot	node-link	bar chart	scatter plot	heat map	treemap	parallel coordinates	pixel based	radar chart	box plot	polygons	3D	
bank	•	•	•		•								[CGK*07]
	•	•	•		•								[CLG*08]
				•								•	[SNI*10]
		•											[DGDLP10]
	•	•											[DLM12]
	•												[SZ14]
			•										[LGM*15]
		•											[DLMP11]
		•	•										[GLS*13]
			•	•			•						[LGM*16]
	•												[CCM*14]
		•	•									•	[KSM07]
	•		•	•			•						[LGM*18]
stock market	•	•	•									•	[KSH*99]
	•		•							•		•	[NB04]
	•				•	•	•						[MSK*06]
		•				•						•	[HLN09]
	•							•			•		[SWK*11]
	•			•									[PL16]
			•	•									[MMT16]
internal		•											[EH09]
		•	•										[Lue10]
			•	•			•		•				[ASS13]
		•											[LN15]
				•								•	[Kea99]
								•					[WL15]
	•	•	•										[LBW*16]
insurance	•												[AAG02]
				•									[HXD03]
	•		•										[PAL04]
			•									•	[FB08]
			•										[ŠFB11]
	•												[SYL14]
		•	•		•	•							[CNY*17]
telecommunication	•												[FB97]
	•	•	•										[CEWB97]
	•												[HT99]
	•												[HS05]
	•	•	•										[BVW12]
	•			•									[MMR17]
	21	16	19	9	4	3	3	2	1	1	1	6	sum

Table 3
This table shows the interaction techniques presented in each of the selected papers in this survey. The columns are sorted from left to right according to the number of papers that fall into each class. Despite seven articles being hybrid, papers are grouped by application domain. Darker rows in the table highlight papers that did not implement interaction techniques.

	abstract/elaborate	select	explore	filter	connect	reconfigure	encode	undo/redo	change configuration	none	
bank	•	•	•	•	•						[CGK*07]
	•	•	•	•	•						[CLG*08]
	•	•									[SNI*10]
	•	•		•	•						[DGDLP10]
	•	•	•		•	•		•	•		[DLM12]
										•	[SZ14]
	•	•	•	•		•					[LGM*15]
	•	•	•	•	•	•					[DLMP11]
										•	[GLS*13]
	•	•	•	•		•					[LGM*16]
											[CCM*14]
										•	[KSM07]
	•	•	•	•				•			[LGM*18]
stock market	•	•	•		•						[KSH*99]
										•	[NB04]
	•	•		•							[MSK*06]
	•		•		•		•				[HLN09]
	•	•	•	•			•				[SWK*11]
										•	[PL16]
										•	[MMT16]
internal										•	[EH09]
										•	[Lue10]
		•	•	•				•			[ASS13]
										•	[LN15]
	•	•	•	•	•	•	•				[Kea99]
		•				•					[WL15]
		•									[LBW*16]
insurance										•	[AAG02]
										•	[HXD03]
										•	[PAL04]
										•	[FB08]
										•	[ŠFB11]
										•	[SYL14]
		•	•								[CNY*17]
telecommunication										•	[FB97]
	•	•	•		•						[CEWB97]
										•	[HT99]
										•	[HS05]
	•		•	•			•				[BVW12]
										•	[MMR17]
	16	18	15	12	9	6	4	3	0	19	

Based on Table 4 we can affirm that most approaches are using interaction techniques on the top of node-link diagrams. This may be caused by the specific characteristics of these visualization technique. Node-link diagrams appear in 16 approaches out of 40 in this survey (see Table 2). This visualization technique is used most frequently in combination with selection, abstraction/elaboration, connection, and exploration. The connection (9 appearances) technique are directly related to network analysis, while selection (18 appearances) and exploration (15 appearances) are related to individual monitoring and analysis. Even being a wise choice to represent networks, this technique does not scale well, and thus, enforces the usage of techniques such as abstraction and elaboration (16 appearances).

Bar charts approaches are also often used in combination with filter and exploration techniques. With a total of 19 appearances out of 40 approaches (see Table 2), bar charts aim to represent quantitative values for different categories. In case of representing individuals, bar charts are a good tool for analyzing the behavior of these individuals. In case of representing groups of individuals, bar charts may serve as a filter to further exploration.

Besides highlighting and selection techniques, line plots surprise regarding the amount of appearances that do not use any

Table 4
Heat map table of the relationship between visualization techniques (x-axis) and interaction techniques (y-axis). This is based on the 40 approaches studied in this survey. The value inside each cell represents the amount of appearances for the respective cell combination. Below the table, we present a legend demonstrating the color encoding criteria used in our heat map table.

	line plot	node-linked	bar chart	scatter plot	pixel based	treemap	heat map	radar chart	par-coord	box plot	polygons	3D
abstract/elaborate	2	5	4	3		2	3		1			4
select	4	10	3	3	1	1	3	1	1			3
explore	2	9	5	2	1	1	2	1	2			3
filter	3	2	6	3			3	1	1			1
connect		8		1								3
reconfigure		2	2	2	2				1			1
encode			1	1							1	2
undo/redo		1				1						
change configuration		1										
none	15	6	9	5						1		3
		legend:	0	1	2	3	4	5+				

interaction technique (15 out of 21). This happens due to line plots being commonly used to represent data results from automatic methods. The usage of abstract line plots to support other visualizations is also a frequent approach in multiple view solutions, as we can see in Cox et al. (1997), Kirkland et al. (1999) and Didimo et al. (2012).

6.5. Analytical methods

With a total of 28 out of 40 approaches, we observe that the majority of fraud detection approaches use visualization as a tool to represent results of different algorithmic processing, i.e., they are pre-processing methods (see Table 5). In a good deal of these works, visualization is not only used to communicate the results of automatic methods but also to analyze the data.

We could identify only three techniques that did not present any automatic analysis method. Those were classified as pure visualization approaches. Di Giacomo et al. (2010) proposed a network visualization approach that does not involve any automatic methods. Leite et al. (2015) present a multiple coordinated view solution that filters various features from raw data to support user-queries. However, sophisticated algorithms for data analysis are not supported. Schaefer et al. (2011) proposed a new design for the visualization of fraudulent data based on triangle slope polygons. This work also did not involve any automatic algorithm for data processing.

From the papers presented in this study we could identify eight (Chang et al., 2007, 2008; Sakoda et al., 2010; Didimo et al., 2011, 2012; Gauldie et al., 2013; Webga and Lu, 2015; Cox et al., 1997) approaches that present an integrated approach, i.e., an interactive loop between visualization and automatic analysis. (See Table 6.)

7. Challenges and opportunities

In this section we present fraud detection challenges and opportunities grouped into data and task complexity, visual scalability, multi-coordinated views for interactive exploration, VA approach, and evaluation according to particular domains and tasks.

7.1. Data and task complexity

Fraud detection is not a well explored area in scientific research. One of the reasons for that is the type of data involved. It is hard to get real world bank transaction data or companies' internal information. This is mainly due to privacy and security reasons. Moreover, the data sets often have some features hidden or changed in order to preserve customers privacy (Leite et al., 2015).

One aspect that adds up to the complexity of fraud detection is that finding suited solutions for detecting suspicious cases is not enough (Dilla and Raschke, 2015). New fraud techniques are always upcoming or being re-adapted. Fraudsters can be very creative when it comes to hiding their attempts. One example is to hide attacks in known and non-suspicious patterns of events. This may cause the simple rule-based approaches to fail. In addition, another challenge in the field is to find a monitoring solution (Huang et al., 2009). The detection of already happened frauds and the prevention of future similar threats is a critical task, and so is the prediction of possible frauds. This makes the task of fraud detection complex and challenging.

Suited solutions need to avoid false-positive identification, which would burden the analysts and waste their time of analysis, as well as false-negatives, which miss actual recurrent frauds and, by consequence, result in fraudulent harm (Luell, 2010). In other words, in order to be more helpful than harmful, the solutions need to be precise in estimating possible threats fine-tuned to each application domain.

In order to analyze and estimate threats, interpreting single events does not usually lead analysts to conclusions. However, a sequence of events, or a network of events, allows the analyst to reason about suspicious behavior by comparing events within their contexts (Chang et al., 2007).

7.2. Visual scalability

In fraud detection, independent of the application domain, the data is always multivariate, temporal, and comprises huge amounts of data items. Contexts, such as daily bank transactions of a huge amount of customers (Leite et al., 2015), a bid and offer variation of the NASDAQ (Huang et al., 2009) as well as the internal operations made by all employees of a company in different systems during a certain period (Argyriou et al., 2013) are very hard to be visually represented (compare Section 5). This is partly due to the fact that the data is not only multivariate but usually also covers long periods of time. To this end, visual aggregation techniques are often needed in order to display such rich data sets. However, during analysis, the exploration of individual cases or short period analysis might still be interesting tasks. Thus, interaction techniques such as elaboration, exploration, and/or filtering are usually applied in order to support these tasks (see Section 6.4).

7.3. Multi-coordinated views for interactive exploration

Detecting suspicious cases according to particular criteria within time-oriented and multivariate data sets is a challenging

Table 5
This table shows how the visualizations are combined with automated processing for each presented approach. We could not identify any approach that uses visual means to get an overview of the data first and to aid the decision for a suited automatic analysis technique, which is interesting and indicates opportunities in the field of VA for fraud detection.

	pure visualization	pre-processing	post-processing	integrated	
bank				•	[CGK*07]
				•	[CLG*08]
				•	[SNI*10]
	•				[DGDLP10]
				•	[DLM12]
		•			[SZ14]
	•				[LGM*15]
				•	[DLMP11]
				•	[GLS*13]
		•			[LGM*16]
		•			[CCM*14]
		•			[KSM07]
		•			[LGM*18]
stock market		•			[KSH*99]
		•			[NB04]
		•			[MSK*06]
		•			[HLN09]
	•				[SWK*11]
		•			[PL16]
		•			[MMT16]
internal		•			[EH09]
		•			[Lue10]
		•			[ASS13]
		•			[LN15]
		•			[Kea99]
				•	[WL15]
		•			[LBW*16]
insurance		•			[AAG02]
		•			[HXD03]
		•			[PAL04]
		•			[FB08]
		•			[ŠFB11]
		•			[SYL14]
		•			[CNY*17]
telecommunication		•			[FB97]
				•	[CEWB97]
		•			[HT99]
		•			[HS05]
		•			[BVW12]
		•			[MMR17]
	3	28	0	9	sum

task. This process asks for an intertwined visual and automatic approach in an interactive multiple-coordinated exploration environment. However, many of the approaches we surveyed made limited use of interaction to support the analysis task. Some approaches used a loose coupling of views (compare for example (Schaefer et al., 2011) and Argyriou et al. (2013)), others a more closed coupling (compare for example WireVis (Chang et al., 2007)). Moreover, the most popular visualization methods are line plots, node-link diagrams, and bar charts. According our survey, 3D approaches are getting less popular recently (compare Section 6.2). Investigating in a systematical applicability of various interaction and visualization techniques according to particular tasks would open new possibilities to explore and analyze fraudulent behavior.

This challenge is closely related to the next one.

7.4. VA approach

Solely automated methods often fail to detect fraudulent behavior, because actors are strategically changing their behavior to mislead monitoring and detecting systems (compare Section 5). This asks for a VA approach. To our knowledge, only (Chang et al., 2007; Sakoda et al., 2010; Argyriou et al., 2013; Leite et al., 2015, 2016) are pursuing a VA approach. The user (in our case the analyst) should take an active role in selecting automatic/analytical approaches, fine-tuning the parameter settings, interactive exploration of the data set, etc. and a seamless integration thereof. In other words, there is a lot of open space to support the various steps in the knowledge generation process (Sacha et al., 2014).

7.5. Evaluation according to particular domains and tasks

Different domains and tasks in fraud detection demand similar evaluation. Evaluating VA based on fraud detection solutions is difficult since it requires experts of the respective area. Domain knowledge is crucial to perform fraud analysis. Common practice is to select a group of analysts to explore a new tool/solution and further ask for empirical feedback.

In most papers, the evaluators came from the same companies, banks, or insinuations that provided the data set (compare Section 5). However, to perform a fair evaluation and avoid previous knowledge to influence the results, the analyst should not be too familiar with the data sets. Otherwise, the evaluator could be subconsciously influenced to find a determine outlier or pattern that he or she already knew. On the other hand, it is hard to find analysts, who have a suitable background to analyze such data sets.

Approaches that address a domain with a high degree of social and financial impact, such as fraud detection, should be carefully evaluated in order to guarantee the worth of substitution of the already existed approaches and the investment of implementing these new solutions into real system. However, in the surveyed papers it is common to use a small number of evaluators (between 2 and 5), usually with previous knowledge about the data set. On the other hand, the urgent demands to detect, analyze, and monitor fraudulent behavior are constantly increasing. Therefore, we are optimistic that more fraud analysts will formulate their demands and needs and will also volunteer to participate in evaluations. There is still enough research space to conduct qualitative and quantitative evaluations to access usability and usefulness of the proposed solutions.

8. Conclusion

We have outlined similarities and differences of fraud detection tasks and approaches in financial domains that share specific characteristics. Yet, when abstracting these tasks of fraud detection they share many characteristics with other domains that deal with the detection of events, such as malware risk analysis, health parameter monitoring, terrorist detection, and governmental fraud. Thus, we believe the solutions outlined in this survey may very well be generalized and the different techniques described here can be adapted to tackle similar problems in other application domains.

Our systematic overview and comparison of different application domains, visualizations, and interaction techniques serves as a sound basis for further research in VA to aid the important task of fraud detection. Using our categorization model we could identify findings, characteristics of the area, and further challenges. Important findings where presented with respect to the applied

Table 6
Heat map table of the relationship between visualization techniques (x-axis) and automatic methods usability (y-axis). This is based on the 40 approaches studied in this survey. The value inside each cell represents the amount of appearances for the respective cell combination. Below the table, we present a legend demonstrating the color encoding.

	3D	polygons	box plot	radar chart	pixel based	par-coord	treemap	heat map	scatter plot	bar chart	node-linked	line plot
integrated	1				1			2	1	4	7	4
post-processing												
pre-processing	5		1	1		4	3	2	8	14	8	16
pure visualization		1			1					1	1	1
		legend:	0	1	2	4	6	8+				

visualization techniques, interaction and exploration, visual and interactive combinations as well as how visual and automated methods were combined. Identified challenges and opportunities include data and task complexity, visual scalability, interactive exploration, exploiting a real VA approach that successfully combines visual and algorithmic means, and domain specific evaluation. By outlining these challenges and pointing to opportunities we encourage further work in this field.

Acknowledgments

The research leading to these results has received funding from the Centre for Visual Analytics Science and Technology (CVAST), funded by the Austrian Federal Ministry of Science, Research, and Economy in the exceptional Laura Bassi Centres of Excellence initiative (#822746).

References

Aigner, Wolfgang, Miksch, Silvia, Schumann, Heidrun, Tominski, Christian, 2011. Visualization of Time-Oriented Data. Springer Science & Business Media.

Argyriou, Evmorfia N, Sotiraki, Aikaterini A, Symvonis, Antonios, 2013. Occupational fraud detection through visualization. In: Intelligence and Security Informatics (ISI), 2013 IEEE International Conference on. IEEE, pp. 4–6.

Artís, Manuel, Ayuso, Mercedes, Guillén, Montserrat, 2002. Detection of automobile insurance fraud with discrete choice models and misclassified claims. J. Risk Insur. 69 (3), 325–340.

Atefeh, Farzindar, Khreich, Wael, 2013. A survey of techniques for event detection in twitter. Comput. Intell..

Becker, Richard A., Volinsky, Chris, Wilks, Allan R., 2012. Fraud detection in telecommunications: history and lessons learned. In: Technometrics. Taylor & Francis.

Bolton, Richard J., Hand, David J., 2002. Statistical fraud detection: a review. Statist. Sci. 17 (3), 235–249.

Carminati, Michele, Caron, Roberto, Maggi, Federico, Epifani, Ilenia, Zanero, Stefano, 2014. Banksealer: an online banking fraud analysis and decision support system. In: ICT Systems Security and Privacy Protection. Springer, pp. 380–394.

Chandola, Varun, Banerjee, Arindam, Kumar, Vipin, 2009. Anomaly detection: a survey. ACM Comput. Surv. (CSUR) 41 (3), 15.

Chang, Remco, Ghoniem, Mohammad, Kosara, Robert, Ribarsky, William, Yang, Jing, Suma, Evan, Ziemkiewicz, Caroline, Kern, Daniel, Sudjianto, Agus, 2007. Wirevis: visualization of categorical, time-varying data from financial transactions. In: Visual Analytics Science and Technology. VAST. IEEE Symposium on. IEEE, pp. 155–162.

Chang, Remco, Lee, Alvin, Ghoniem, Mohammad, Kosara, Robert, Ribarsky, William, Yang, Jing, Suma, Evan, Ziemkiewicz, Caroline, Kern, Daniel, Sudjianto, Agus, 2008. Scalable and interactive visual analysis of financial wire transactions for fraud detection. Inf. Vis. 7 (1), 63–76.

Cheng, Dawei, Niu, Zhibin, Yan, Junchi, Zhang, Jiawan, Zhang, Liqing, 2017. Visual analytics for loan guarantee network risk management. arXiv preprint arXiv: 1705.02937.

Cox, Kenneth C, Eick, Stephen G, Wills, Graham J, Brachman, Ronald J, 1997. Visual data mining: recognizing telephone calling fraud. Data Min. Knowl. Discov. 1 (2), 225–231.

Di Giacomo, Emilio, Didimo, Walter, Liotta, Giuseppe, Palladino, Pietro, 2010. Visual analysis of financial crimes:[system paper]. In: Proceedings of the International Conference on Advanced Visual Interfaces. ACM, pp. 393–394.

Didimo, Walter, Liotta, Giuseppe, Montecchiani, Fabrizio, 2012. Vis4AUI: visual analysis of banking activity networks. In: Richard, Paul, Kraus, Martin, Laramee, Robert S., Braz, Jos (Eds.), GRAPP/IVAPP. SciTePress, pp. 799–802.

Didimo, Walter, Liotta, Giuseppe, Montecchiani, Fabrizio, Palladino, Pietro, 2011. An advanced network visualization system for financial crime detection. In: Pacific Visualization Symposium (PacificVis), 2011 IEEE. IEEE, pp. 203–210.

Dilla, William N., Raschke, Robyn L., 2015. Data visualization for fraud detection: practice implications and a call for future research. Int. J. Account. Inf. Syst. 16, 1–22.

Dumas, Maxime, McGuffin, Michael J., Lemieux, Victoria L., 2014. Financevis.net - A Visual Survey of Financial Data Visualizations. Poster Abstracts of IEEE VIS 2014. Poster and Extended Abstract.

Eberle, William, Holder, Lawrence, 2009. Mining for insider threats in business transactions and processes. In: Computational Intelligence and Data Mining, 2009. CIDM '09. IEEE Symposium on. IEEE, pp. 163–170.

Fawcett, Tom, Provost, Foster, 1997. Adaptive fraud detection. Data Min. Knowl. Discov. 1 (3), 291–316.

Furlan, Štefan, Bajec, Marko, 2008. Holistic approach to fraud management in health insurance. J. Inf. Organ. Sci. 32 (2), 99–114.

Gauldie, David, Langevin, Scott, Schretlen, Peter, Jonker, David, Bozowsky, Neil, Wright, William, 2013. Louvain Clustering for Big Data Graph Visual Analytics.

He, Zengyou, Xu, Xiaofei, Deng, Shengchun, 2003. Discovering cluster-based local outliers. Pattern Recognit. Lett. 24 (9), 1641–1650.

Hilas, Constantinos S., Sahalos, John N., 2005. User profiling for fraud detection in telecommunication networks. In: 5th International Conference on Technology and Automation. pp. 382–387.

Hollmén, Jaakko, Tresp, Volker, 1999. Call-based fraud detection in mobile communication networks using a hierarchical regime-switching model. Adv. Neural Inf. Process. Syst. 889–895.

Huang, Mao Lin, Liang, Jie, Nguyen, Quang Vinh, 2009. A visualization approach for frauds detection in financial market. In: Information Visualisation, 13th International Conference. IEEE, pp. 197–202.

Keahey, T.A.Ian, 1999. Visualization of high-dimensional clusters using nonlinear magnification. In: Electronic Imaging'99. International Society for Optics and Photonics, pp. 228–235.

Keim, Daniel, Andrienko, Gennady, Fekete, Jean-Daniel, Görg, Carsten, Kohlhammer, Jörn, Melançon, Guy, 2008. Visual analytics: definition, process, and challenges. Inf. Vis..

Keim, Daniel A, Mansmann, Florian, Schneidewind, Jörn, Thomas, Jim, Ziegler, Hartmut, 2008. Visual Analytics: Scope and Challenges. Springer.

Keim, Daniel, et al., 2000. Designing pixel-oriented visualization techniques: theory and applications. IEEE Trans. Vis. Comput. Graphics 6 (1), 59–78.

Kielman, Joe, Thomas, Jim, May, Richard, 2009. Foundations and frontiers in visual analytics. Inf. Vis. 8 (4), 239.

Kirkland, J Dale, Senator, Ted E, Hayden, James J, Dybala, Tomasz, Goldberg, Henry G, Shyr, Ping, 1999. The NASD Regulation advanced-detection system (ADS). AI Mag. 20 (1), 55.

Kirkos, Efstathios, Spathis, Charalambos, Manolopoulos, Yannis, 2007. Data mining techniques for the detection of fraudulent financial statements. Expert Syst. Appl. 32 (4), 995–1003.

Ko, Sungahn, Cho, Isaac, Afzal, Shehzad, Yau, Calvin, Chae, Junghoon, Malik, Abish, Beck, Kaethe, Jang, Yun, Ribarsky, William, Ebert, David S, 2016. A survey on visual analysis approaches for financial data. In: Comput. Graph. Forum. 35 (3), 599–617.

Kou, Yufeng, Lu, Chang-Tien, Sirwongwattana, Sirirat, Huang, Yo-Ping, 2004. Survey of fraud detection techniques. In: Networking, sensing and control, 2004 IEEE International Conference on, Vol. 2. IEEE, pp. 749–754.

Leite, Roger A, Gschwandtner, Theresia, Miksch, Silvia, Gstrein, Erich, Kuntner, Johannes, 2015. Visual analytics for fraud detection and monitoring. In: Visual Analytics Science and Technology (VAST), 2015 IEEE Conference on. IEEE, pp. 201–202.

Leite, Roger A, Gschwandtner, Theresia, Miksch, Silvia, Gstrein, Erich, Kuntner, Johannes, 2016. Visual analytics for fraud detection: focusing on profile analysis. In: Proceedings of the Eurographics Conference on Visualization (EuroVis) - Posters 2016. IEEE.

Leite, Roger A, Gschwandtner, Theresia, Miksch, Silvia, Kriglstein, Simone, Pohl, Margit, Gstrein, Erich, Kuntner, Johannes, 2018. EVA: visual analytics to identify fraudulent events. IEEE Trans. Vis. Comput. Graphics 24 (1), 330–339.

Liu, Juan, Bier, Eric, Wilson, Aaron, Guerra-Gomez, John Alexis, Honda, Tomonori, Sricharan, Kumar, Gilpin, Leilani, Davies, Daniel, 2016. Graph analysis for detecting fraud, waste, and abuse in healthcare data. AI Mag. 37 (2), 33–46.
Lookman, Sanni, Nurcan, Selmin, 2015. A framework for occupational fraud detection by social network analysis. In: CAISE 2015 FORUM.
Luell, Jonas, 2010. Employee Fraud Detection Under Real World Conditions. University of Zurich.
Manunza, L., Marseglia, S., Romano, S.P., 2017. Kerberos: a real-time fraud detection system for ims-enabled voip networks. J. Netw. Comput. Appl. 80, 22–34.
Merino, Carmen Sanz, Sips, Mike, Keim, Daniel A, Panse, Christian, Spence, Robert, 2006. Task-at-hand interface for change detection in stock market data. In: Proceedings of the Working Conference on Advanced Visual Interfaces. ACM, pp. 420–427.
Miksch, Silvia, Aigner, Wolfgang, 2014. A matter of time: applying a data–users–tasks design triangle to visual analytics of time-oriented data. Comput. Graph. 38, 286–290.
Monamo, Patrick, Marivate, Vukosi, Twala, Bheki, 2016. Unsupervised learning for robust bitcoin fraud detection. In: Information Security for South Africa (ISSA), 2016. IEEE, pp. 129–134.
NASDAQ Stock Market. (0000). http://www.nasdaq.com/.
Nesbitt, Keith V., Barrass, Stephen, 2004. Finding trading patterns in stock market data. IEEE Comput. Graph. Appl. 24 (5), 45–55.
Ngai, E.W.T., Hu, Yong, Wong, Y.H., Chen, Yijun, Sun, Xin, 2011. The application of data mining techniques in financial fraud detection: a classification framework and an academic review of literature. Decis. Support Syst. 50 (3), 559–569.
Pham, Thai, Lee, Steven, 2016. Anomaly detection in bitcoin network using unsupervised learning methods. arXiv preprint arXiv:1611.03941.
Phua, Clifton, Alahakoon, Damminda, Lee, Vincent, 2004. Minority report in fraud detection: classification of skewed data. ACM SIGKDD Explor. Newslett. 6 (1), 50–59.
Phua, Clifton, Lee, Vincent, Smith, Kate, Gayler, Ross, 2010. A comprehensive survey of data mining-based fraud detection research. arXiv preprint arXiv:1009.6119.
Sacha, Dominik, Stoffel, Andreas, Stoffel, Florian, Kwon, Bum Chul, Ellis, Geoffrey, Keim, Daniel A, 2014. Knowledge generation model for visual analytics. IEEE Trans. Vis. Comput. Graphics 20 (12), 1604–1613.
Sakoda, Chika, Nagasaki, Azusa, Itoh, Takayuki, Ise, Masayuki, Miyashita, Kousuke, 2010. Visualization for assisting rule definition tasks of credit card fraud detection systems. In: IIEEJ Image Electronics and Visual Computing Workshop.
Saraiya, Purvi, North, Chris, Lam, Vy, Duca, Karen A., 2006. An insight-based longitudinal study of visual analytics. IEEE Trans. Vis. Comput. Graphics 12 (6), 1511–1522.
Schaefer, Matthias, Wanner, Franz, Kahl, Roman, Zhang, Leishi, Schreck, Tobias, Keim, Daniel, 2011. A Novel Explorative Visualization Tool for Financial Time Series Data Analysis. Bibliothek der Universität Konstanz.
Seeja, K.R., Zareapoor, Masoumeh, 2014. Fraudminer: a novel credit card fraud detection model based on frequent itemset mining. Sci. World J. 2014.
Sharma, Anuj, Panigrahi, Prabin Kumar, 2013. A review of financial accounting fraud detection based on data mining techniques. arXiv preprint arXiv:1309.3944.
Šilić, Artur, Bašić, Bojana Dalbelo, 2010. Visualization of text streams: a survey. In: Knowledge-Based and Intelligent Information and Engineering Systems. Springer, pp. 31–43.
Sithic, H.Lookman, Balasubramanian, T., 2013. Survey of insurance fraud detection using data mining techniques. arXiv preprint arXiv:1309.0806.
Spence, Robert, 2007. Information Visualization: Design for Interaction, second ed.
Šubelj, Lovro, Furlan, Štefan, Bajec, Marko, 2011. An expert system for detecting automobile insurance fraud using social network analysis. Expert Syst. Appl. 38 (1), 1039–1052.
Sun, Qixiang, Yao, Yi, Lin, Shanjun, 2014. Detection of health insurance fraud with discrete choice model: evidence from medical expense insurance in China. Available at SSRN 2459343.
Wagner, Markus, Fischer, Fabian, Luh, Robert, Haberson, Andrea, Rind, Alexander, Keim, Daniel A., Aigner, Wolfgang, 2015. A survey of visualization systems for malware analysis. EG Conference on Visualization (EuroVis) - STARs. The EGA, pp. 105–125.
Wanner, Franz, Stoffel, Andreas, Jäckle, Dominik, Kwon, Bum Chul, Weiler, Andreas, Keim, Daniel A, Isaacs, Katherine E, Giménez, Alfredo, Jusufi, Ilir, Gamblin, Todd, et al., 2014. State-of-the-art report of visual analysis for event detection in text data streams. Comput. Graph. Forum 33 (3).
Webga, Kodzo, Lu, Aidong, 2015. Discovery of rating fraud with real-time streaming visual analytics. In: Visualization for Cyber Security (VizSec), 2015 IEEE Symposium on. IEEE, pp. 1–8.
Yi, Ji Soo, ah Kang, Youn, Stasko, John T., Jacko, Julie A., 2007. Toward a deeper understanding of the role of interaction in information visualization. IEEE Trans. Vis. Comput. Graphics 13 (6), 1224–1231.

Roger Almeida Leite is a Ph.D. student at Institute of Software Technology & Interactive Systems, Vienna University of Technology. His main research interests include Information Visualization and Visual Analytics.

Theresia Gschwandtner is Postdoc University Assistant at Institute of Software Technology & Interactive Systems, Vienna University of Technology.

Silvia Miksch is University Professor at Institute of Software Technology & Interactive Systems, Vienna University of Technology. Her main research interests are Visualization/Visual Analytics (in particular Focus+Context and Interaction methods) and Time.

Erich Gstrein is lead data scientist at Erste Group IT International. His current research interests are outlier detection, personalization and their application on the financial domain.

Johannes Kuntner is an IT architect in the area of Datawarehousing/BI and Big Data in Erste Group IT International, the solution provider of Erste Group. One of his main targets is the creation of an overarching architecture enabling and fostering the development and practice of advanced analytics – especially visual – methods in the bank.