A Survey on Visual Analysis of Financial Data

Author:
Xuanwu YUE

Advisor:
Prof. Huamin QU

10 October 2017

TABLE OF CONTENTS

A SURVEY ON VISUAL ANALYSIS OF FINANCIAL DATA

by

XUANWU YUE

Department of Computer Science and Engineering

The Hong Kong University of Science and Technology

ABSTRACT

Market participants and businesses continually spare no efforts towards making a smarter decision during the past decades. From the perspective of data statistics and data analysis, the progress that based on the enormous financial data has drawn a lot of attention not only in the industry but also the academic. However, owing to the high diversity, large volume, and unstructured characteristic, the visual analysis has great potential among data exploration progress. The tasks summarized by financial practitioners provide a more clear direction for computer science fields. However, there are limited visual analysis systems developed to solve those tasks efficiently. Proper summarization combining main functions and applications in financial areas should be emphasized to get an overview of state-of-art work and future potential research fields.

In this survey, we first introduce the background and motivation of visual analysis on financial data and give a comprehensive review of typical tasks like anomaly detection, predictive analysis, correlative analysis, and application in FinTech. Finally, we conclude the survey with a discussion of future research directions.

CHAPTER 1

INTRODUCTION

1.1 Background and Motivation

Due to the unpredictable economic and business environment, business and market participants must make the best decisions promptly. The decision-making process based on financial data has become a hot topic in the industry. However, due to the diversity of financial data, the analysis of large amounts of financial data is an important task. This led to the rapid research and development of visualization and visual analysis systems for financial data exploration. The availability of financial data from different sources provides market analysts and investors with the opportunity to gain new insights to make informed decisions. The analysis can guide them when developing the best investment and risk management strategies. To facilitate this analysis, analysts traditionally use visual analysis tools that are typically built on standard statistical methods such as moving averages and regression. Examples of these tools include popular line charts and candlestick (Edwards et al., 2007) for decision-making missions in financial market professionals (Sorenson & Brath, 2013).

Although these technologies are ubiquitous, these tools are often insufficient to deal with problems that arise in large data. With the advent of the digital age, different financial sectors are producing countless real-time financial data, including asset transactions, news, and economic indicators. A large number of stocks with multiple attributes (such as price, time, volume) are exchanged every few seconds, and a large number of reports on the stock market and economic indicators are published on a regular basis for the benefit of stakeholders. In the era of big data (Zarate Santovena, 2013; Cook & Thomas, 2005), conventional methods that using standard analytical techniques are often limited. Line graphs are still very popular among financial analysts because they incorporate different aggregation methods to reduce overplotting issues (Sorenson & Brath, 2013). However, such aggregations may lead to unacceptable loss of information in financial data analysis (for example, sales and transaction comparisons). In the decision-making process, the analyst not only needs to study the aggregated results but also to investigate and consider detailed and scattered information.

From the financial point of view, some urgent tasks can be resolved through computer science and practice. As a result, financial institutions deal with millions of transactions each year. While most of these transactions are legal, but a few of them are likely to cause serious harm to the client or the financial institution itself. Therefore, the organization must assess the credibility of each transaction. However, due to complex multidimensional data, financial fraud detection (FFD) is a daunting task. Also, Kielman et al. (Kielman et al., 2009) portray fraud detection as an open visual analysis problem that requires the visual inspection, disclosure, and analysis.

Information visualization and visual analysis communities have paid attention to these financial datasets due to their diversity, complexity, and volume, leading to the rapid development of many visual analysis systems and interactive methods (such as bank wire transactions (Chang et al., 2007) and corporate sales (Ko et al., 2012)). The development of such systems and methods often requires visual analysis researchers to work with financial experts to assess their needs and challenges and to design visual analysis solutions accordingly to address these issues. The high-bandwidth information channels and rapid iterative analysis of the flexibility and powerful features is the core advantage of visualization and visual analysis. This combination is particularly valuable in the context of the financial sector, which increasingly dominated by a large number of dynamic and heterogeneous data.

To meet this need, the researchers conducted several surveys, outlining the existing work in the financial sector. For example, Pryke (Pryke, 2010)) discusses the role of visualization in financial industry applications. Schwabish's (Schwabish, 2014) survey provided an introduction and guidance to economists. Recently, Floods (Flood et al., 2016) survey focuses on applications aimed at monitoring financial stability. We focus on research publications, as well as visual analysis tools in the industry to obtain enough information of this field. We investigate research papers in the field of information visualization and visual analysis for financial markets.

Overall, this survey attempts to categorize and characterize financial systems regarding three tasks: anomaly detection, predictive analysis, correlative analysis and their implementation in FinTech arena, which are also highly relevant topics in the visualization and visual analysis fields. They are also the increasing demands requested from the industries. This survey also illustrates how visual analysis can better leverage the human ability to gain insights from data, including identifying the underlying trends, anomalies, and correlations, through an interactive and undirected search (Keim et al., 2006).

1.2 Overview

Since the finance field is a complicated and well-developed sphere, we do not aim to survey all the models and systems related to knowledge discovery and machine learning.

We focus on how visual analysis affects tasks, such as anomaly detection, predictive analysis, correlative analysis, and their applications in FinTech. Our goal is twofold: 1) to review the tasks from the perspective of data processing, data mining or machine learning methods. 2) to emphasize the role of visual analysis and illustrate its benefits for exploring data. The rest of this survey is organized as follows.

Chapter 2 introduces the taxonomy of this survey regarding the visual analysis of financial data.

Chapter 3 describes anomaly detection methods from both data mining and visual analysis aspects. More specifically, fraud detection and market abuse detection are given closer attention.

Chapter 4 presents the previous research of predictive analysis towards stock prices, foreign exchange and bankruptcy, which has been a popular topic in both the business arena and computer science. How humans engages in the prediction process relies on the essential role of visualization.

Chapter 5 investigates how the correlation provides broader and different perspectives for gaining insights into the data, including different data sources and their correlations. Besides, the various interactions also facilitate exploration, which may find existing or potential time-series patterns or spatial patterns.

Chapter 6 discusses the FinTech arena, which is currently riding the crest of the wave. The issues existing in Blockchain analysis and Bitcoin are prices that similar to stocks but substantially different. These can be transferred to typical and classical visualization research topics, which have been demonstrated and well-developed in the visualization field.

Chapter 7 summarizes the current situation and lists some future research directions of the visual financial analysis area.

CHAPTER 2

TAXONOMY

In this chapter, we discuss related surveys towards visual analysis of financial data. Tegarden's (Tegarden, 1999) survey is the first work where information visualization techniques have been emphasized in financial analysis systems. The contribution relies on the 16 typical business application domains that are summarized as shown in Table 2.1, which provided a more clear direction for researchers to move forward.

Financial Risk Management	Industrial Process Control
Operations Planning	Capital Markets Management
Military Strategic Planning	Network Monitoring
Market Analysis	Derivatives Trading
Fraud/Surveillance Analysis	Portfolio Management
Actuarial Modeling	Customer/Product Analysis
Budget Planning	Operations Management
Economic Analysis	Fleet/Shipping Admin

Table 2.1: 16 typical business application domains (Tegarden, 1999).

There are two surveys from the field of economics (Pryke, 2010; Schwabish, 2014). Pryke (Pryke, 2010) conducted interviews with practitioners from both tech and financial companies to have a more clear comprehension of financial applications. Based on that, Pryke concludes that visual approaches should be further emphasized for financial market analysis in the whole industry. The survey issued by Flood et al. (Flood et al., 2016) mainly focuses on systems for monitoring financial stability and benefits of Visual Analytics (Cook & Thomas, 2005) in monitoring stability monitoring tasks. Dumas et al. (Dumas et al., 2014) launch an online visual overview website with filtering options to show financial visualization systems till now.

Compared with the surveys above, our work puts more emphasis on the tasks that that financial systems aim to solve. We categorize existing works into four classes, namely anomaly detection, predictive analysis, correlation analysis and applications in FinTech, which are shown in Figure 2.1.

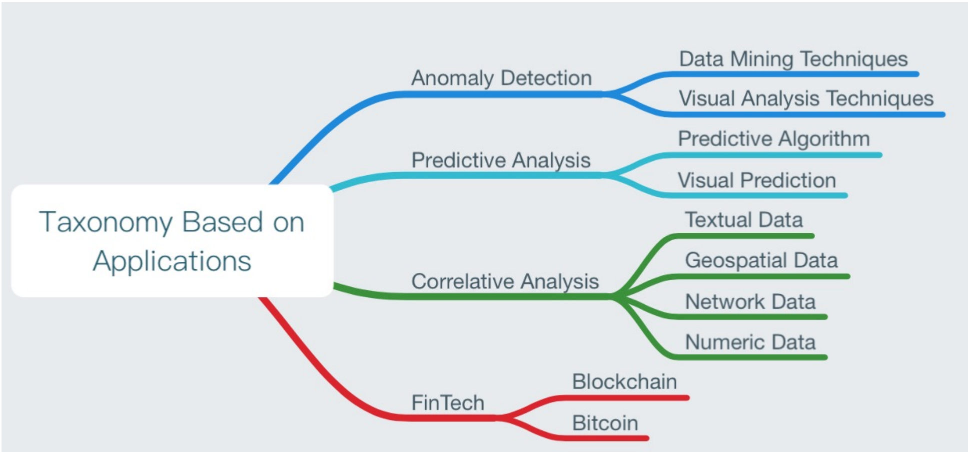

Figure 2.1: The Taxonomy.

Anomaly Detection: Data analysis towards detecting transaction anomalies is an essential financial task. Visual analysis tools with high interactive functions can facilitate the investigators in identifying fraudulent transactions more efficiently and effectively.

Predictive Analysis: Stock price prediction is a topic which attracts not only the financial practitioners but also the data mining and deep learning fields to provide more efficient and accurate algorithms. In the visualization arena, Hao et al. (Hao et al., 2009) illustrated their method, visual prediction: the act of visually predicting a time-series variable by observing the predictions from a computational model, shown alongside the time-series representations, which offers a combination of visualization and prediction models.

Correlation Analysis: With the advent of different accessible sources of data, not only time-series, but also textual data, such as financial news and announcements, multivariate spatio-temporal data or social network relationships can boost the correlation analysis towards a better understanding of a corresponding scenario.

FinTech: 'FinTech' refers to technology enabled financial solutions. More specifically, in our survey, it refers to 'BlockChain' and 'BitCoin' fields. The existing knowledge and methodologies can be transferred into this undiscovered but potentially valuable field.

CHAPTER 3

ANOMALY DETECTION

Anomaly detection, as one of the essential tasks for finding new insights from large datasets, has been widely explored in various research and application domains. According to its definition in the data mining domain, anomaly detection refers to "the problem of finding patterns in data that do not conform with expected behavior". In different domains, researchers have different names for these non-conforming patterns, such as anomalies, outliers, exceptions, aberrations, surprises, peculiarities, and so on. When referring to the financial area in particular, it facilitates the search process and identification of trends to analyze these anomalies in a temporal-related scenario. In addition, the examination of anomalies enables analysts to identify risks, extreme changes, or rare incidents. In this chapter, we concentrate on the identification of anomalies in the finance domain, which is commonly known as the financial fraud detection (FFD).

Financial fraud often leads to immeasurable losses and destructive results, which is why FFD has become so important in this arena. FFD includes identifying fake financial information from real data, thereby revealing fraud and allowing decision makers to develop appropriate strategies to minimize the impact of fraud. In the economy, financial fraud is an increasingly mature issue. A representative example is the former Nasdaq market president Bernard Madoff implementation of the Ponzi scheme, resulting in a huge global loss of about US$50 billion US dollars (FBI, 2008).

According to the (FBI, 2007), the proposed financial fraud classification framework depends on the FBI's financial crime framework on the grounds that it is a well-established FFD framework. Financial fraud (FF) categories cover bank fraud, insurance fraud, securities and merchandise fraud, and other related financial fraud. More specifically, bank fraud further includes money laundering, credit card and mortgage fraud; insurance fraud can occur in different steps in insurance procedures, including applications, qualifications, ratings, settlement and claims, and by consumers and practitioners, insurance company staff and health care providers; securities fraud through market control, theft of security accounts and telephone fraud and other means; Finally, sorts of financial fraud types which fall in other than those that previously mentioned categories, corporate fraud as an example.

In order to complete the classification of financial fraud, the first level of FFD is divided into bank fraud, insurance fraud, securities and commodity fraud and other related financial fraud. Then they are further classified according to the second level of fraud, as shown in Figure 3.1.

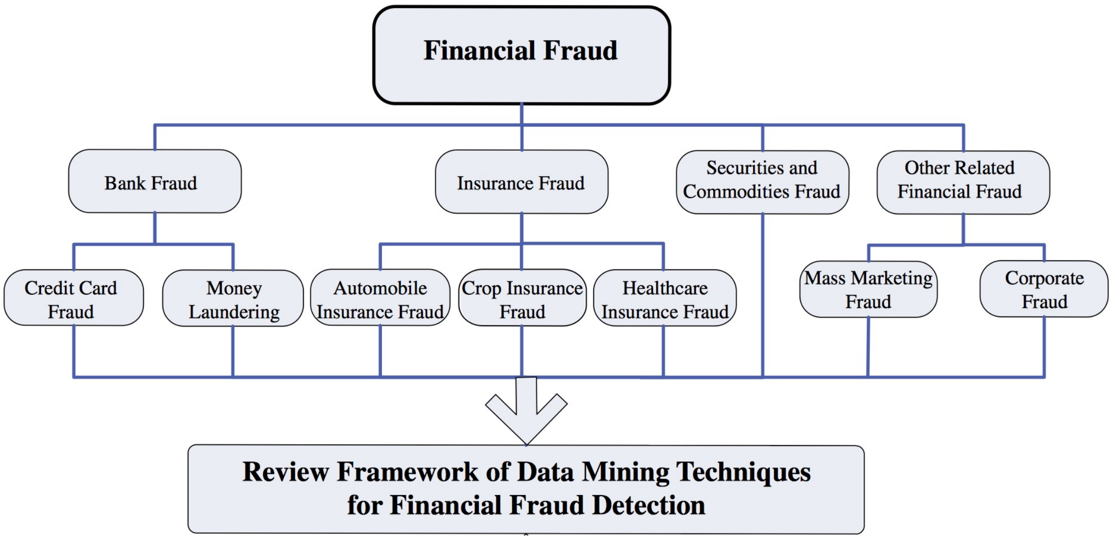

Figure 3.1: Analysis of FFD research based on the proposed classification framework (FBI, 2007).

Meanwhile, Turban et. al (Turban et al., 2011) define data mining as a process of using statistical, mathematical, and machine learning techniques to extract valuable information and then gain insight from large databases. Again, fraud investigators have recently realized that data visualization is becoming increasingly important for fraud detection (Clopton et al., 2014) and has begun to actually implement a variety of visualization technologies in the real world. Data visualization is particularly important at the beginning of fraud detection. At this stage, the analyst's goal is to better understand the possible relationships in complex data sets. Here, we illustrate some of these two aspects of the typical works.

3.1 Data Mining Techniques

Among the data mining techniques, logistic models, neural networks, the Bayesian belief network, and decision trees are four mostly popular ones in classification category. Of these techniques, logistic models are the most popular, followed by neural networks, and then the Bayesian belief network and decision trees. (Ngai et al., 2011) We will give a brief introduction of these four techniques.

Logistic Model: The logistic model is a generalized linear model that is used in regression progress (Spathis, 2002; Yeh & Lien, 2009). The predictor variables can be numerical or categorical. It is mainly used to solve automobile insurance and corporate fraud problems.

Neural Networks: Artificial neural networks (ANNs), which is based on a collection of connected units called artificial neurons, are computing systems inspired by the biological neural networks (Ghosh & Reilly, 1994; Yeh & Lien, 2009). They are used among classification and clustering categories. The advantages of neural network are: 1) It is adaptive. 2) It can generate robust models. 3) The classification process can be modified based on the new weights or parameters. Neural networks are mainly used in credit card, automobile insurance and corporate fraud issues.

Bayesian Belief Network: The Bayesian Belief Network (BBN) using a directed acyclic graph (DAG) to represent random variables and their conditional interdependencies. In the DAG, nodes are random variables and missing edges mean conditional independencies between the variables (Kirkos et al., 2007; Pearl, 2014). The Bayesian belief network is also mainly used in credit card, automobile insurance, and corporate fraud detection.

Decision Trees: Decision trees are adopted in the predictive decision that creates mapping from observations to possible consequences (Han et al., 2011; Li et al., 2008). These trees can implement machine-learning-based algorithms, like ID3, CART and C4.5. Leaves encode prediction outcomes, and branches mean features conjunctions. Decision trees are often applied in credit card, automobile insurance, and corporate fraud issues.

3.2 Visual Analysis Techniques

Kirkland et al. (Kirkland et al., 1999) published the first work on financial fraud detection using visual techniques in the AAAI Conference. Regulatory analysis, fraud detection alerts, and knowledge discovery processes are solved by the combination of visualization, pattern recognition, data mining and artificial intelligence. The architecture of ADS is shown in Fig. 3.2.

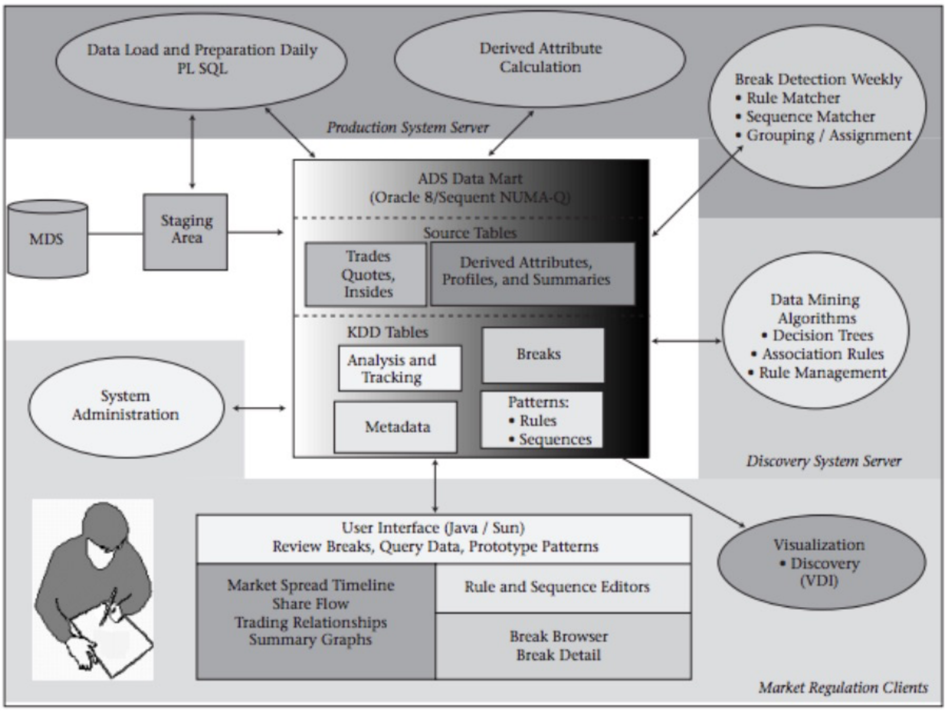

Figure 3.2: The Advanced-Detection System Architecture (Kirkland et al., 1999).

The main idea of WireVis is to use multiple coordinated views to explore a large amount of transaction data (Chang et al., 2007). They highlight account similarities based on defined keywords. WireVis combines a keyword network view, a heatmap, a sample search example tool and a string & bead view to provide a highly interactive tool. The four views together detect the relationship between the account in the transaction, the time range, and the keywords. A global overview of the data provides aggregated ability and organizes groups of transactions for deeper investigation and analysis. However, WireVis only focuses on individual accounts or specific cluster analysis based on the frequency of occurrence of keywords. The system screenshots are shown in Fig. 3.3.

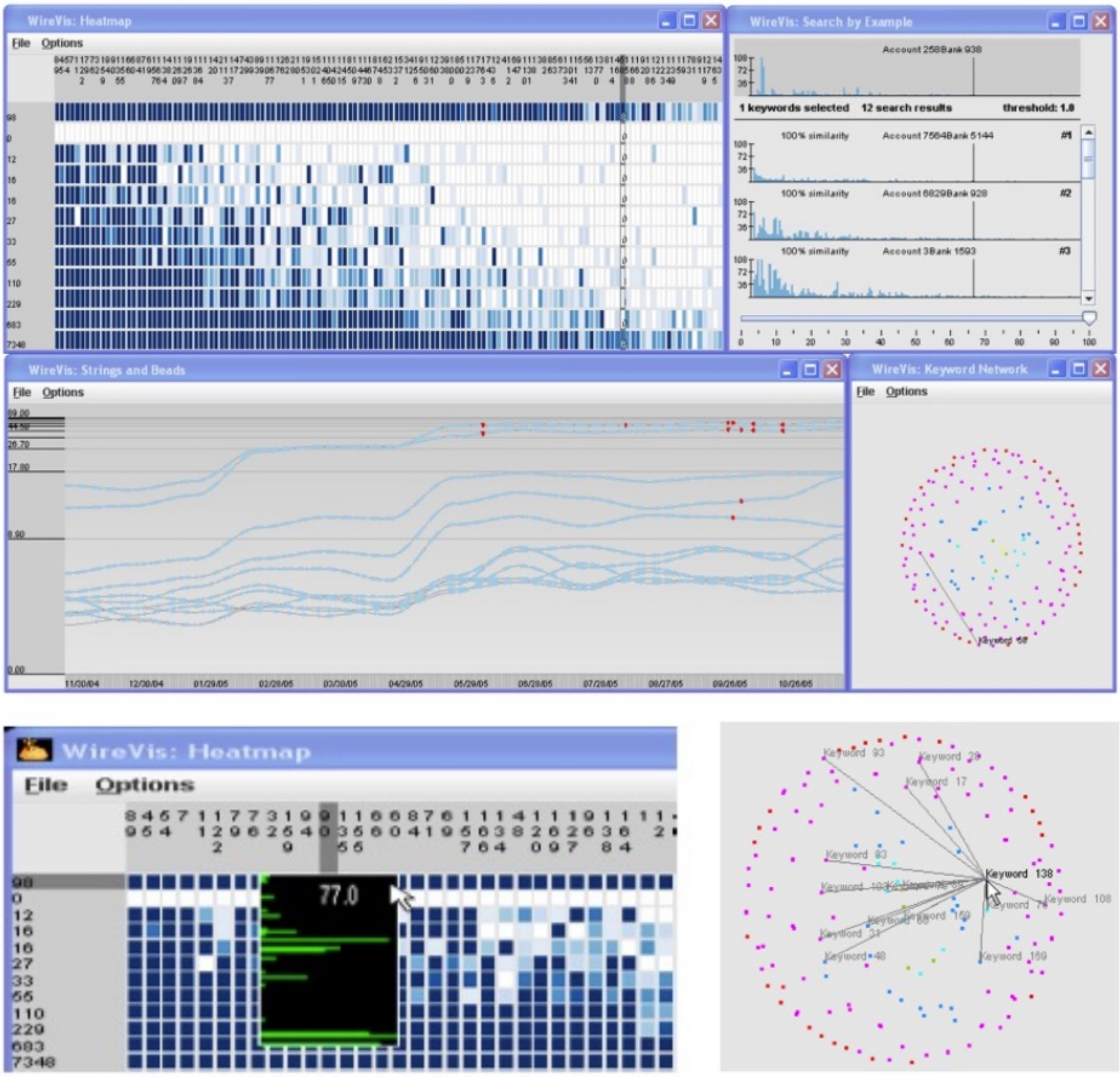

Figure 3.3: Information visualization tools from WireVis (Chang et al., 2007).

Huang et al. (Huang et al., 2009) proposed a Visual Analysis framework for stock market security towards globally event monitoring. His work presents a 3D tree map and a node-link diagram for market performance analysis and network analysis to reduce the number of false alarms caused by traditional AI algorithms.

When it comes to "unauthorized trading" testing, Leite et al. (Leite et al., 2017) develop interactive visualization system, EVA (using Visual Analytics for event detection), which collaborates with the domain experts. They are used to automatically calculate the score based on the profile in the backend model, which helps to monitor the behavior of the customer (especially in multivariate and time-specific payment transactions). The EVA interface is shown in Fig. 3.4.

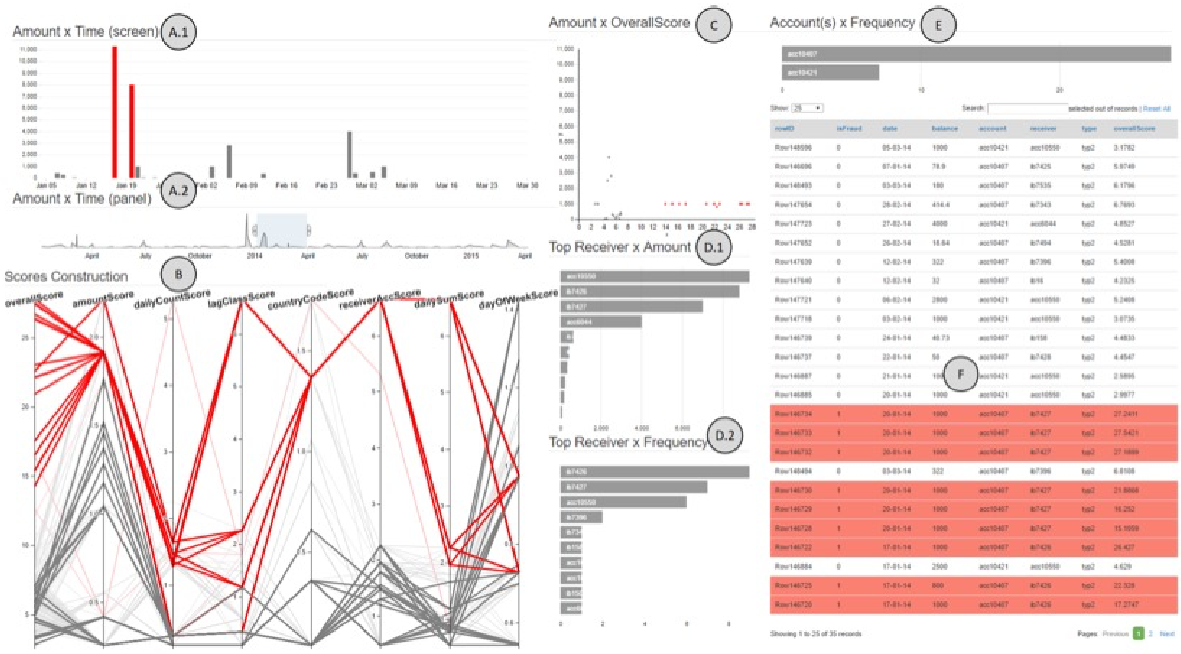

Figure 3.4: Screenshots of EVA(Event detection with Visual Analytics) (Leite et al., 2017).

The theoretical framework structure provided by Dilla et al. (Dilla & Raschke, 2015) implies the current needs in FFD. They illustrate when and how investigators should apply visual analysis. They evaluated various visualization techniques. In addition, the authors also suggest that future challenges in this research area are visualizations support different cognitive processes, which facilitate fraud detection visualization system. The framework, as shown in Fig. 3.5, are mentioned relationships between task and fraud investigator characteristics, data visualization characteristics, and decision outcomes.

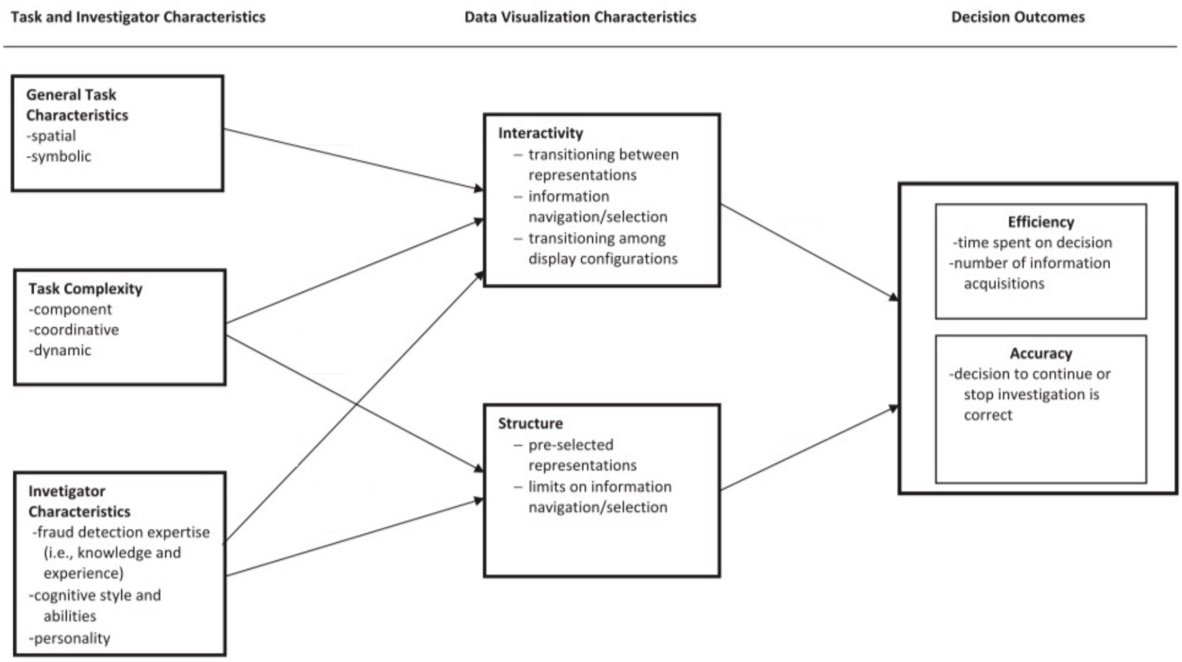

Figure 3.5: Relationships between task, investigator, and data visualization tool characteristics with implications for decision outcomes.

The current interactive data visualizations and visual analysis systems towards FFD are extremely limited. This is caused by the limited professionals judgment and domain expertise experience. There are also close collaboration with expert investigators to improve research. The purpose of the study is to test the basic proposition of the impact of domain knowledge on the use of interactive data visualization tools in low- and moderate-complexity fraud detection tasks. In low complexity fraud detection tasks such as the purchase example, the investigator can check the exception pattern in the data (the space task supported by the graphical representation) and drill down into the data to select a single item for further inspection (supported by the table Symbolic characterization). In a more complex detection task, they may be more dependent on the use of graphical representation of the pattern detection, rather than the analysis of a single transaction in order to minimize cognitive effort (Vessey, 2006).

Successful fraud detection depends on the investigator's ability to detect patterns that suggest fraudulent transactions. Interactive data visualization tools have the potential to make fraudulent transaction detection processes more efficient and efficient. Interactive visualization allows the researcher to more easily select a representation that has a cognitive fit with the different steps of the task, thereby more accurately detecting fraudulent transactions in a given set of data.

CHAPTER 4

PREDICTIVE ANALYSIS

In general, financial data such as stock price or currency exchange is the price data of discrete time series, like a day's closing price, for a certain time interval (day, week, month, etc.). Financial time series usually move in a fluctuating way because the advantages of trading are constantly changing. At the same time, because of emergency or daily changes in operations and many other random factors, this volatility movement is not regular.

The motivation for financial time series data analysis is the desire to predict future possibilities based on historical data. Financial time series forecasting is important for making business decisions. Over the past few years, a lot of forecasting technologies have been developed. Due to the chaos and uncertainty of the system, the prediction of the financial market is quite challenging. In the past, many statistical and artificial intelligence tools were applied to market analysis.

In this chapter, we talk about the financial predictive analysis from two aspects, Prediction Algorithm and Visual Prediction.

4.1 Predictive Algorithm

Due to the chaos and uncertainty of the system, the prediction of the financial market is quite challenging. In the past, many statistical and artificial intelligence tools have been used in market analysis. In Kovalerchuks and Vityaev's book (Kovalerchuk & Vityaev, 2000), they divide the current method into three categories: numerical model (ARIMA models, instance-based learning, neural networks, etc.), rule-based model (decision tree and DNF learning, Leaf classifier, and hidden Markov model, etc.) and relational data mining (induction logic programming). In these methods, soft computing techniques such as fuzzy logic, neural networks, and probability reasoning have attracted much attention due to their ability to deal with stock market uncertainty and noise (Vanstone & Tan, 2008). These technologies have the value of reducing the uncertainty of the financial system by extracting global features or dynamically adjusting historical data. In financial markets, neural networks have succeeded in predicting stocks based on past price trends (Venugopal et al., 2009). However, most of the stock forecasts through automation technology are limited to individual stocks, without the need for promotion.

In classical finance theory, the main idea of homo economics suggests that investors are determined to maximize their personal utility functions in a logically understandable and predictable way. Their asset price perception is equal to the discounted value of the expected cash flow in the future, and is not affected by any emotional irrationality. In behavioral finance, investors are no longer rational, but decisions are influenced by irrational behavior (Schwartz, 2007). Theoretical evidence shows that personal emotions affect human decision making (Peterson, 2011). Various studies use this hypothesis and try to find meaningful relationships with financial markets.

Earlier studies of stock market forecasts were based on the random walk theory and effective market hypothesis (EMH) (Fama, 1965). According to the EMH stock market, prices are mainly driven by new information, that is, news updates, rather than the price of the present and the past. As the news unpredictable, the stock market price to follow the random trend mode, can not be more than 50% accuracy prediction. Gabriel et al. (Fung et al., 2003) proposed a system framework that simultaneously exploited multiple time series and used text documents as a source of prediction. As news articles affect behavior and stock price volatility is caused by the tender and inquiry decisions, it seems logical to say that the press release indirectly affects the stock market.

Although the news will definitely affect the stock market, but the public's mood or mood may also play an equally important role. We know from the psychological research, in addition to information, the emotions play an important role in human decision-making. Behavioral finance provides further evidence that financial decisions are driven by emotions and emotions (Nofsinger, 2005). However, a large number of public surveys of representative population samples are often expensive and time-consuming, such as Gallup's opinion polls and various consumer and welfare indices. However, the accuracy of these methods is limited by the lower level of the selected indicators associated with the public mood. As a result, according to the (Bollen et al., 2011) survey, the announcement of the large-scale Twitter on twitter.com will affect the stock market, and the results show that changes in public mood can indeed be tracked from large-scale content, which is a simple text processing technology for large-scale Twitter feeds, and respond to a variety of socio-cultural drivers in a highly differentiated way. As can be seen from Figure 4.1, the time series often overlap or point in the same direction. The change in the past calm value (t-3) predicts the rise or fall of the DJIA value (t-0). The calm mood dimension is predictive of DJIA.

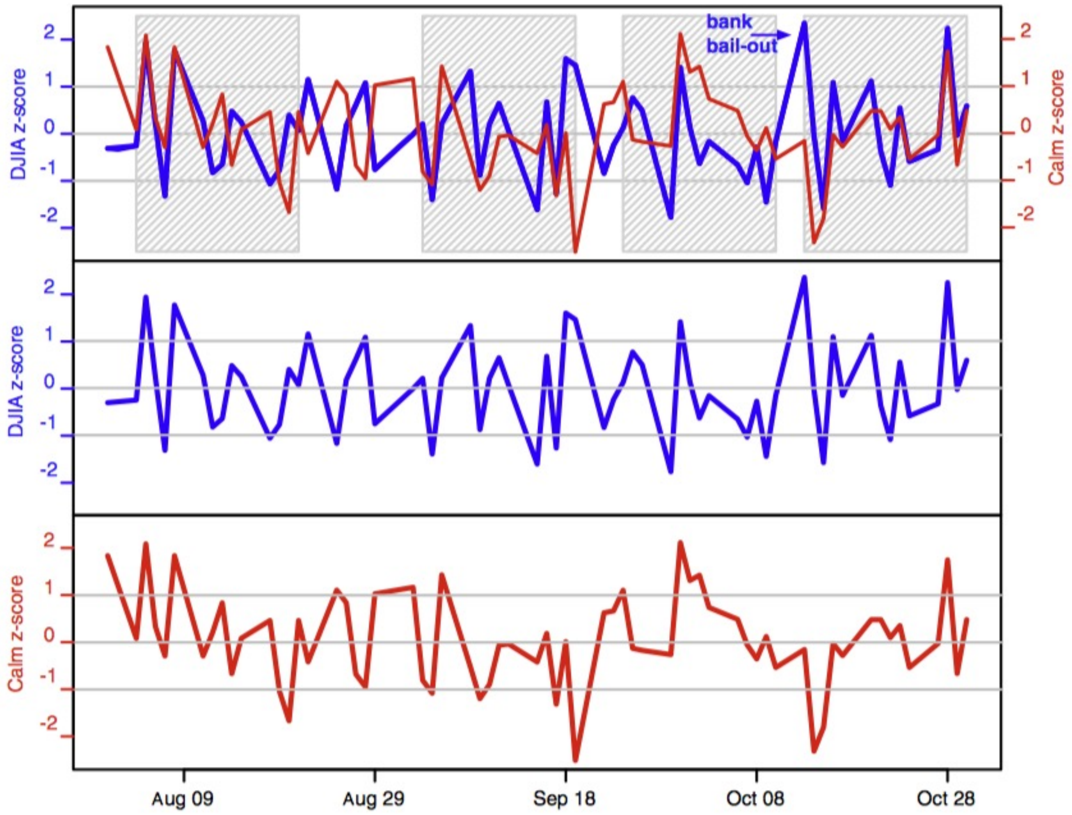

Figure 4.1: The overlap of the day-to-day difference of DJIA values(blue) with the Calm time series(red) that has lagged for three days (Bollen et al., 2011).

Another approach, deep learning, has drawn considerable attention in recent years. Some models which use deep learning to shed light on financial market have achieved relatively promising results. Ding et al. (Ding et al., 2015) proposed a novel model combining natural language processing and deep learning to predict stock price movement. Events were extracted from news text and a deep convolutional neural network was constructed to train the model. This new method has achieved 6% improvement compared with the state-of-the-art baseline algorithm and is more profitable than time-series based prediction models.

4.2 Visual Prediction

The forecast information has been integrated into the time series representation. Such a method of visual prediction was defined by Hao et al. (Hao et al., 2009): by observing the prediction from the calculation model to visual predict the behavior of time series variables, together with the time series representation. For visual prediction, we can start from the general time series data.

Broberg (Broberg et al., 1999) applies the Kalman filter algorithm to predict the behavior and execution of multithreaded programs using line graphs. Ichikawa (Ichikawa et al., 2002) introduces a visual environment that allows users to view a large number of stock price forecasts using different types of line charts, textures, colors, and 3D graphs. Croker (Croker, 2007) declares a forecast scenario by displaying predictive predictions and confidence levels using color areas. SAS Predictive System (Croker, 2007) provides automatic model fitting and prediction of confidence limits for time series data. SAS compares the actual and predicted values with the prediction errors using different color time series. The screenshots of the system are shown in Fig. 4.2.

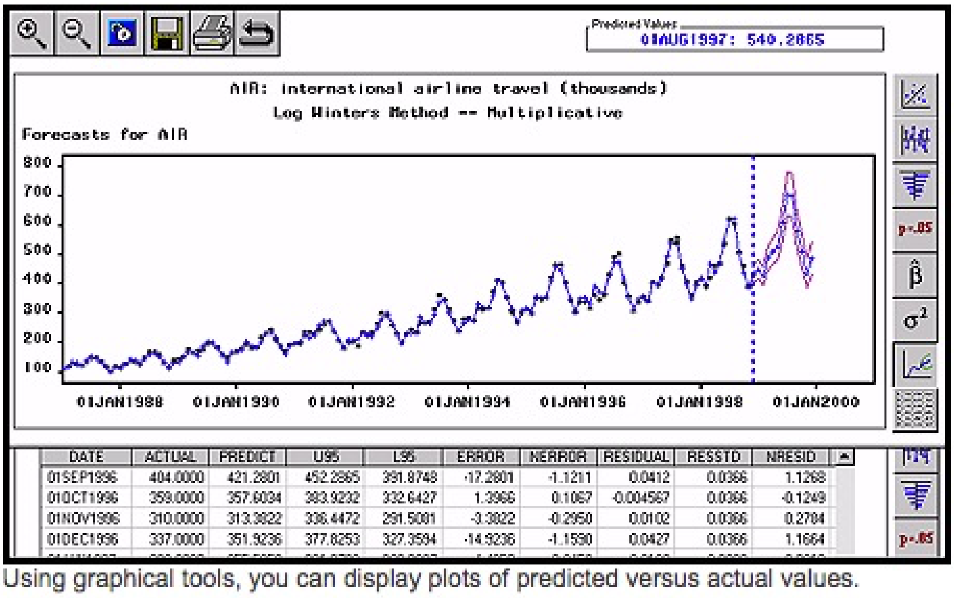

Figure 4.2: The Interface of SAS Time Series Forecasting System (Croker, 2007).

TimeSearcher3 (Buono et al., 2007) visualizes the predictions in the time series visualization based on the data-driven prediction method. The system estimates the time series data by looking for similar past sequences, while allowing the user to control the similarity metric. Hao et al.'s support for visual predictions goes beyond TimeSearcher3, which supports the integration of multiple predictive models, including autoregressive integrated moving averages (ARIMA (Box et al., 2015)), Holt Winters (seasonal methods), and similarity-based models. They extend this approach to visualize peak predictions of seasonal trends(Hao et al., 2011), interact with the model to adjust smoothing parameters, and connect predictions to similar past trends by brushing-and-linking, which is shown in Fig. 4.3.

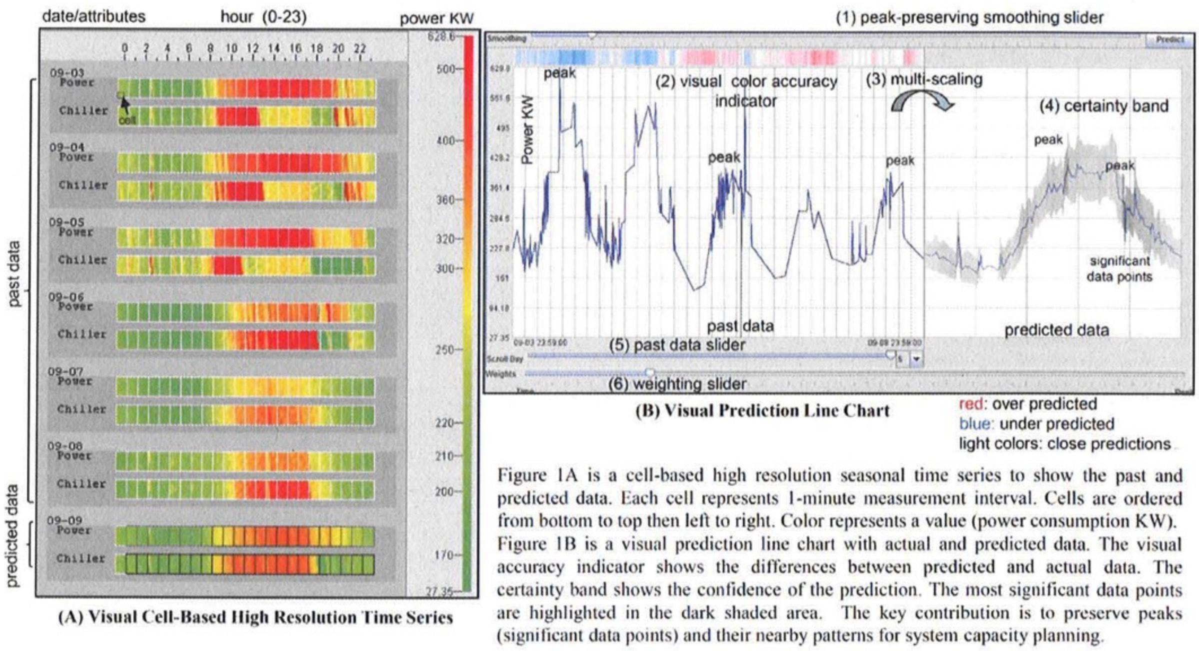

Figure 4.3: Visual prediction of next day's power consumption KW from historical data in a data center (Hao et al., 2011).

However, traditional methods cannot fully support complex visual datasets (such as stock markets) for visual exploration of future trends, mainly due to the lack of consideration of the relationship between variables (e.g., if A increases, B decreases). Stock traders can get stock prices and trading volume information, and they may also be interested in news, Twitter and the company's interest in their earnings reports. Visualization can help them understand the data, such as the stock over time, the status of each stock, and how people react to Twitter. However, after understanding this information, they must estimate the future of each stock investment, even if they may not fully understand how best to interpret their observations in the visualization, during which events (such as product launches, mergers and acquisitions) as well as the dynamic relationship between the stock.

By using computer-supported visualization techniques, traders can interact with the computer to show them multiple predictions (based on different historical data), explaining different "what-if" scenarios (e.g., what if stock A is due to increased incomes Increase) and share a set of predictions with the computer through visual exploration. Exploring these relationships through "what-if" can help us better judge the future, rather than blindly trusting the computational models that are lack of background information (such as fracturing legislation that affects oil stock prices).

TimeFork (Badam et al., 2016) technology combines computational models with interactive visual analysis, which is designed to support prediction of multivariate time series data interactively. TimeFork combines the visual representation of multiple time series with the prediction information generated by the calculation model. Using this method, the analyst manually predicts future changes through interaction and automatically determines the most likely results through interactive simulations, and they eventually use the model for a common prediction to talk to the computational model.

Each time step will produce conditional predictions and link to longer periods of time. The conditional prediction can be combined with the visualization time prediction of the previous step or as a new band / line. This step, therefore, allows the computer to evaluate its own stock based on analyst forecasts, enabling visual exploration of predictions and maintaining ongoing dialogue (Fig. 4.4). The analyst gets enough information to make an investment decision.

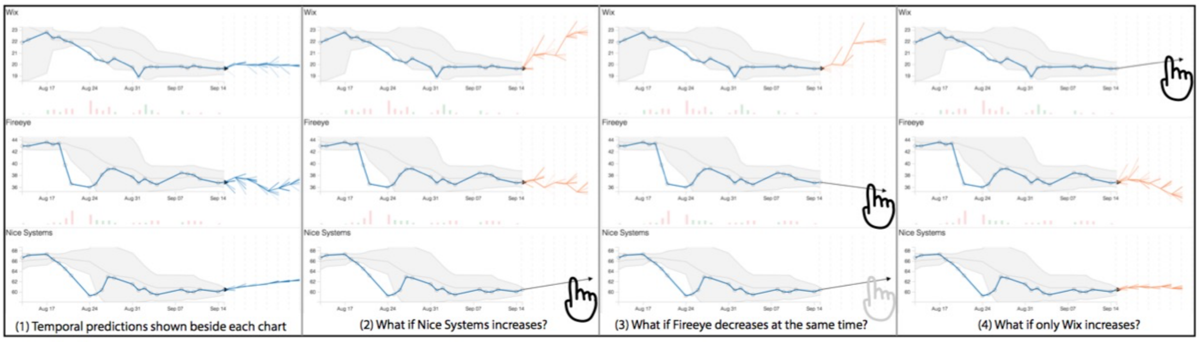

Figure 4.4: A typical dialogue in TimeFork (Badam et al., 2016): the analyst interacts and the computer updates the predictions in the visualizations for further exploration.

In general, effective predictive visualization is a daunting task, mainly because the audience is usually a nontechnical organization. Non-technical viewers attach great importance to any visualization, so it is necessary to be as simple as possible. Even so, multivariate time series of visual predictions are very common in the financial sector and is still a challenge because of the lack of support for visual exploration of variable futures.

CHAPTER 5

CORRELATIVE ANALYSIS

As data has become easier to collect and cheaper to store, there is an increasing number of agencies and organizations gathering data. However, another challenge appears with more complex and larger datasets, which makes it difficult to handle the exploration procedure. Therefore, a need for facilitating the exploration and analysis emerges. Among all the opportunities, a particular one made from such multivariate and multivariate data is the possibility to search for correlations and investigate potential links from these datasets, either causally or in a predictive way. With such analysis, more difficulties appear. On the one hand, most end users do not have much background on the data analytics, which might encounter obstacles when understanding the statistical algorithms; one the other hand, as different algorithms perform differently under various circumstances, how to choose the best appropriate one for the matched temporal and spatial scale is also a challenging task. What is more, in the real-world scenario, many datasets have large noise and we sometimes have to deal with them. Hence, a visual analytics approach would be a good choice for analysts to investigate and further understand the potential correlations at various scales and different levels of detail, which can be helpful for effective decision making (Malik et al., 2012).

As mentioned earlier, it is essential for analysts to understand relationships among various spatial trends and temporal patterns from multivariate spatio-temporal datasets. Such analyses often perform as a precursor, which helps make the predictive models from the original datasets; such analysis can also help analysts identify more possibilities in hypothesis generation and decision makers gain insights in the exploration process in the real world practice. However, the later two challenges mentioned in the previous paragraph, data noise and data complexity, still remain, which make it difficult to identify potential relationships.

Then again, big data has developed as a hot topic and public spotlight in data analytics for the past few years. Research issues for big data analytics emerge from all perspectives. The majority deal with massive volumes of data, rapid and real-time generation of data, broadly changing forms of data. In terms of data forms, especially for unstructured textual data, the data themselves make it even more challenging for the effective extraction of meaningful information.

In addition, quantitative network analysis assumes a key part in measuring specific aspects of interest in a precise way and makes it easier to compare among different designs.

However, it still has shortcomings which network visualization can overcome. By the interaction, the gap towards understanding data models between domain experts and end users can be facilitated.

With the potential power of correlative visual analysis, not only the researchers from computer science but also the practitioners can combine different sources of data, like geospatial data, network data, text data and so on. to gain more insights from the data. We represent some typical works, both academic papers and online tools, from the angle of correlative analysis to demonstrate its usefulness in financial markets.

5.1 Geospatial Data

MarketAnalyzer (Ko et al., 2012) allows analysts to explore current sales volumes, trends, and time market share growth rates using a series of linked views, including pixel-based visualization matrices, line graphs, stacked bar charts, and choropleth graphs. By observing all the characteristics of the point of sale data, the analyst can investigate the market situation. In addition, competitive conditions, point of sale, trends and growth rates are expected to be analyzed on the map for regional markets. It uses statistical models such as linear trend estimates and ARIMA (Automated Regression Integrated Moving Average)(Box et al., 2015) to provide forecasts for individual products and different stores. In order to reduce the inherent perceptual difficulties of pixel-based visualization techniques, a localized zoom is also provided for focus + context analysis. It also supports market forecasts or geographic analysis, discussing design strategies with market analysts. It has been pointed out that predicting future trends and understanding the superior of geographical location in market analysis is important.

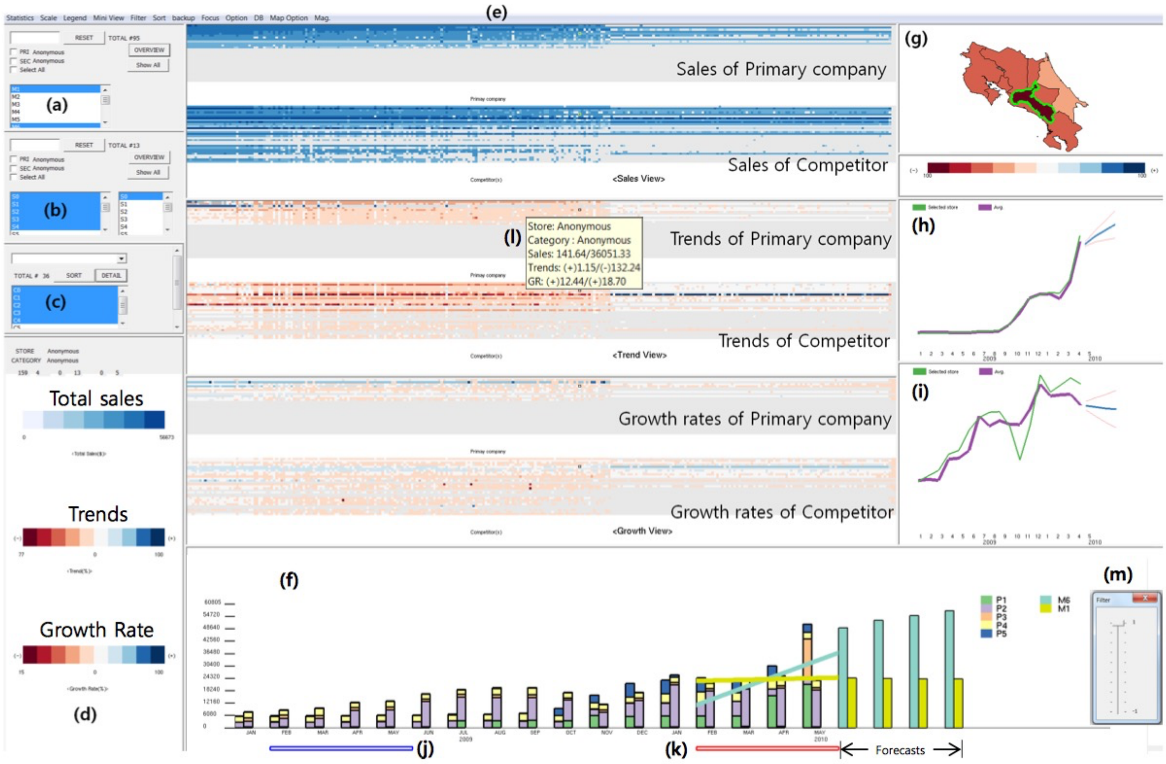

Figure 5.1: The MarketAnalyzer (Ko et al., 2012) Interface.

MarketAnalyzer, as shown in Fig. 5.1., consists of multiple coordinated views linked with interactive filters: (a) Company filter, (b) Store filter, (c) Products filter, (d) Legend view, (e) (Sorted) Matrix view for sales, trends, and growth rates. (f) Stacked bar view, (g) Geographical view, (h) and (i) Line graph small multiples views, (j) and (k) Time slider widgets and aggregation tools for temporal comparison. (l) Tooltip. (m) Filter. In the legends, the blue indicates positive and the red represents negative measurements in sales, trends, or growth rates.

Overall, MarketAnalyzer is a new visual analytics tool, which aims at market and competitive intelligence analysis with multiple tightly integrated interactive visualizations views and trend analysis techniques. Scalable and sortable matrix displays provide enhanced pixel-based visualization of sales, trends and growth rates, while linked line graph views and stacked bar charts can help analyze and sense access to global and specific product and information. The link to the choropleth map makes geospatial, temporal and regional competition analysis, making multivariable spatio-temporal data more realistic.

5.2 Social Network

The global financial crisis has brought several banks facing the brink of collapse. This not only causes the loss of the investor, but also the cost of the real economy and the overall welfare. The most common sources used to describe bank interdependencies and networks are based on data, such as equity and debt exposure or payment flows, and common changes in market data (for example, stock price, CDS spreads and bond spreads). Although these direct and indirect links complement each other, they show a series of restrictions. Even in the ideal world banking network, direct and real links should be used to assess most of the interbank data between bank balance sheets without disclosure.

Malik et. al. (Malik et al., 2012) presents a visual analysis method to explore the relevance of multivariate spatio-temporal data. However, the visual prediction of multivariate time series is very common in the financial field-it is still a challenge due to the lack of support for visual exploration of variable futures variables. StockFork (Fig. 5.2.) contains an overview + detail layout to follow Shneiderman's visual exploratio, including guide-first, zoom and filter, and on-demand details. The overview outlines the overall pattern of stock market data, and the detail view can display the selected time period.

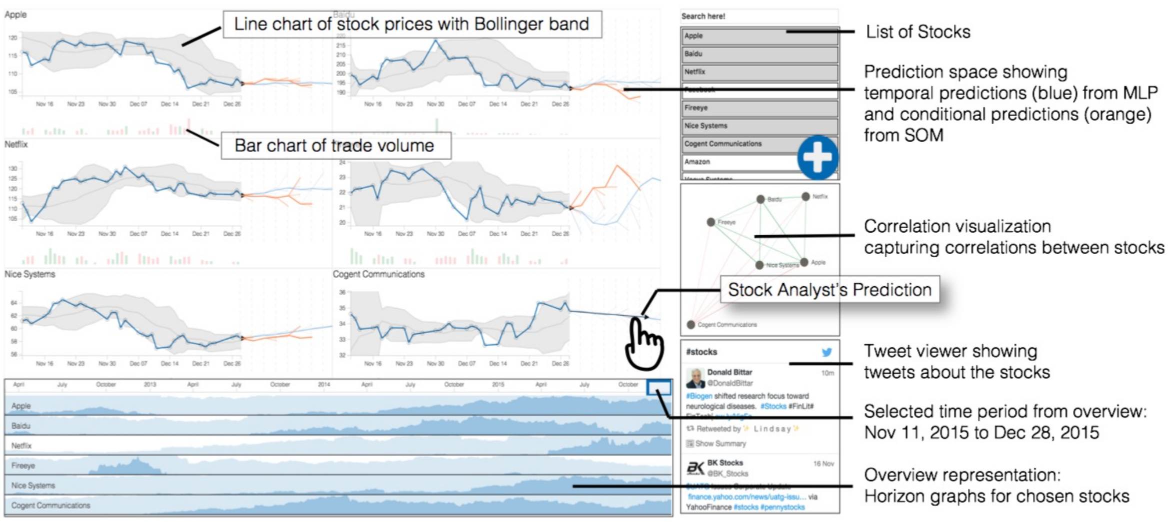

Figure 5.2: StockFork (Badam et al., 2016) presents line charts of stock prices along with other visualizations and prediction space to support visual prediction.

StockFork's overview captures the overall pattern of multiple shares of interest, selected from the list of available stocks in the interface. The overview also supports selecting the time period on the timeline to display specific data in the detail view. This helps analysts look at different time scales (such as past weeks and months) that are common in stock trading and also test real ground data over the past time period.

StockFork has three detailed views: 1) Small multi-graphs include the trade volume of the stock price and the histogram, as they are common representations of time series data. The line graph also shows that the Bollinger bands captures the moving average and moving the standard difference, indicating that the stock price is bullish / bearish. 2) Relevant visualization can show the correlation between stocks to see which stocks show similar trends (tasks pursued by multiple stock analysts). 3) In addition to stock market data, public opinion and external information reported by the company are often used as a qualitative source of information for forecasting. The browser uses Twitter views to streamline stock-related tweets to indicate such information.

The StockFork interface is designed to visualize stock prices and trading volumes, which are the main quantitative information that stock traders and analysts follow, as well as through the derivative attributes (such as cross-correlation) for stock analysis. Through the relevant point of view, experts or practitioners can assume the existence of relevance, and then potentially predict the emergence of certain patterns of the phenomenon.

5.3 Textual Data

Samuel et al. (Rönnqvist & Sarlin, 2015) focuses on the text mining method because it presents a way to analyze how to relate relationships between banks, such as news, official reports, forums, and so on, by analyzing how financial discourse is mentioned. The idea of analyzing the relationship in the text itself is simple but widely applicable. Their contribution is to present this text-based approach to bank-based interrelationships, focusing on analyzing the resulting banking network model and ultimately quantifying the importance or central position of the bank (Fig. 5.3).

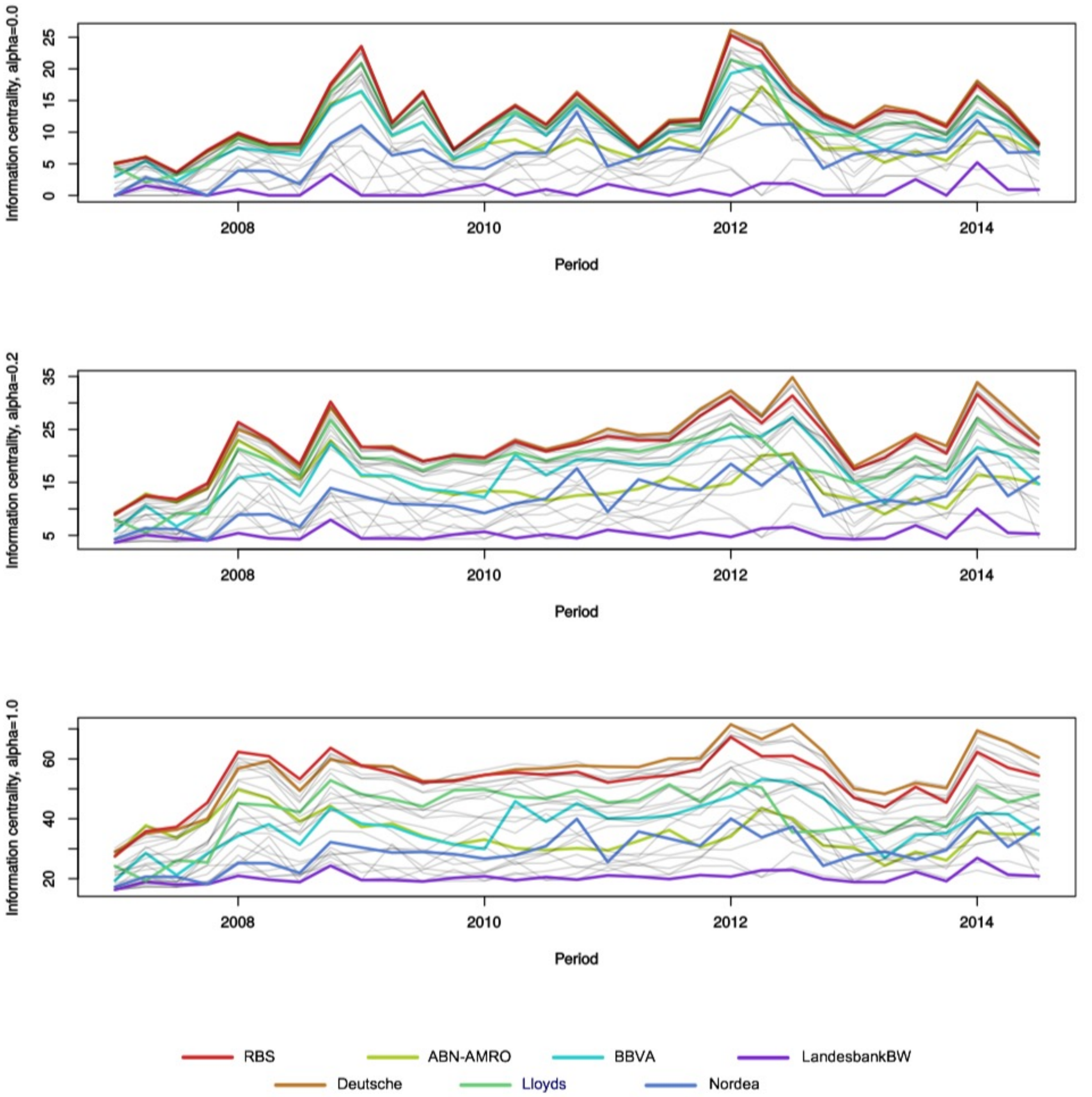

Figure 5.3: Information centrality for banks over time. The charts show dierent levels of smoothing: none (= 0.0), little (= 0.2) and moderate (= 1.0). A few example banks are highlighted (bank labels are described in Table A.1 in the Appendix).

Chan et al. (Chan & Qu, 2016) present FinaVistory (Fig. 5.4.), a visual narrative system that uses visualization to produce narratives that help readers understand financial news from different perspectives. While reading financial news, although critics explain the volatility of economic indicators in the article every day, the news usually prejudices the author's favorite views. On the other hand, the number of financial news released these days is shocking, the same problem also has a variety of views. Unless the diversity of the entire public environment, the audience will be difficult to maintain an objective from the mass media to identify useful information. If the computer is given the ability to analyze all the news and produce a narrative of all the issues related to the news, various readers can get more compelling stories. Nowadays, narrative visualization is a popular technology in the news media that provides reader-driven stories, as well as potential areas of application. In this article, they first identify financial news, with two characteristics: the polarity associated with each message, the impact of indicators or factors affected by the indicators. We then use these features to explain price fluctuations and explain the socio-economic relationship between prices and different topics.

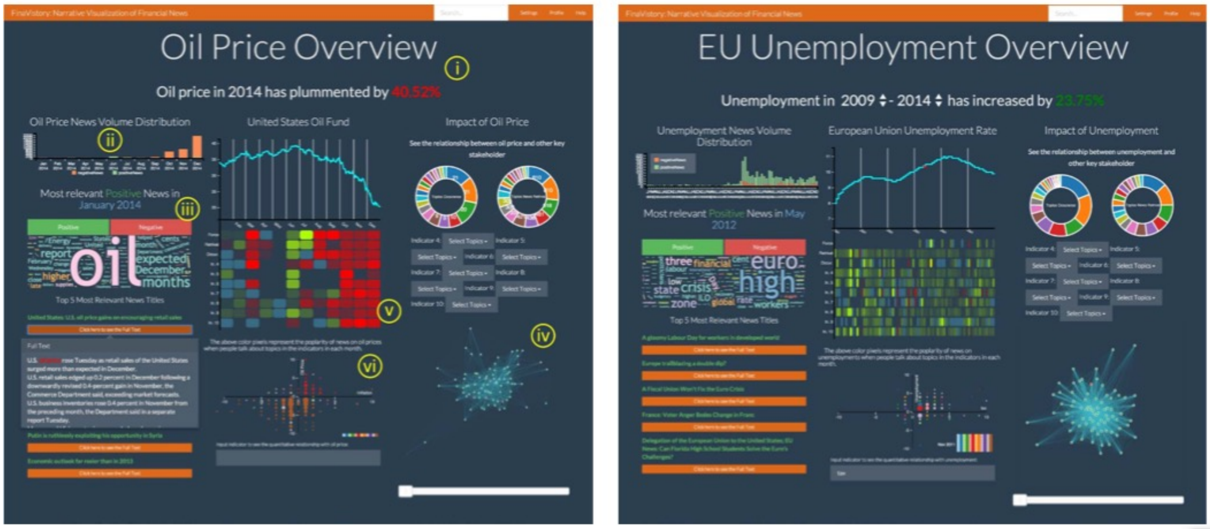

Figure 5.4: Sample interface of FinaVistory visualizing oil price news in 2014 and EU Unemployment during 2008-14 (Chan & Qu, 2016).

5.4 Online Tools

For engineering systems, unlike the financial system, functions are carefully assigned to different components with clearly defined interrelations and interdependencies, each with a dedicated instrument, such as fuel and cooling tables on a dashboard-reports and its relationship to these clearly defined states of physical and functional. Since the officially designed system is more closely controlled than the financial system, the financial dashboard requires additional creativity to ensure that all important points are effectively captured. For example, as shown in Fig. 5.5, between 2000 and 2014, 519 US banks' financial reports showed three different views - geography, time slider and economic significance. The dashboard supports research analyst's sensory exercises. Loan defaults and other credit losses are the main source of risk for most banks and financial institutions, and credit assessment is critical to maintaining the bank's financial stability.

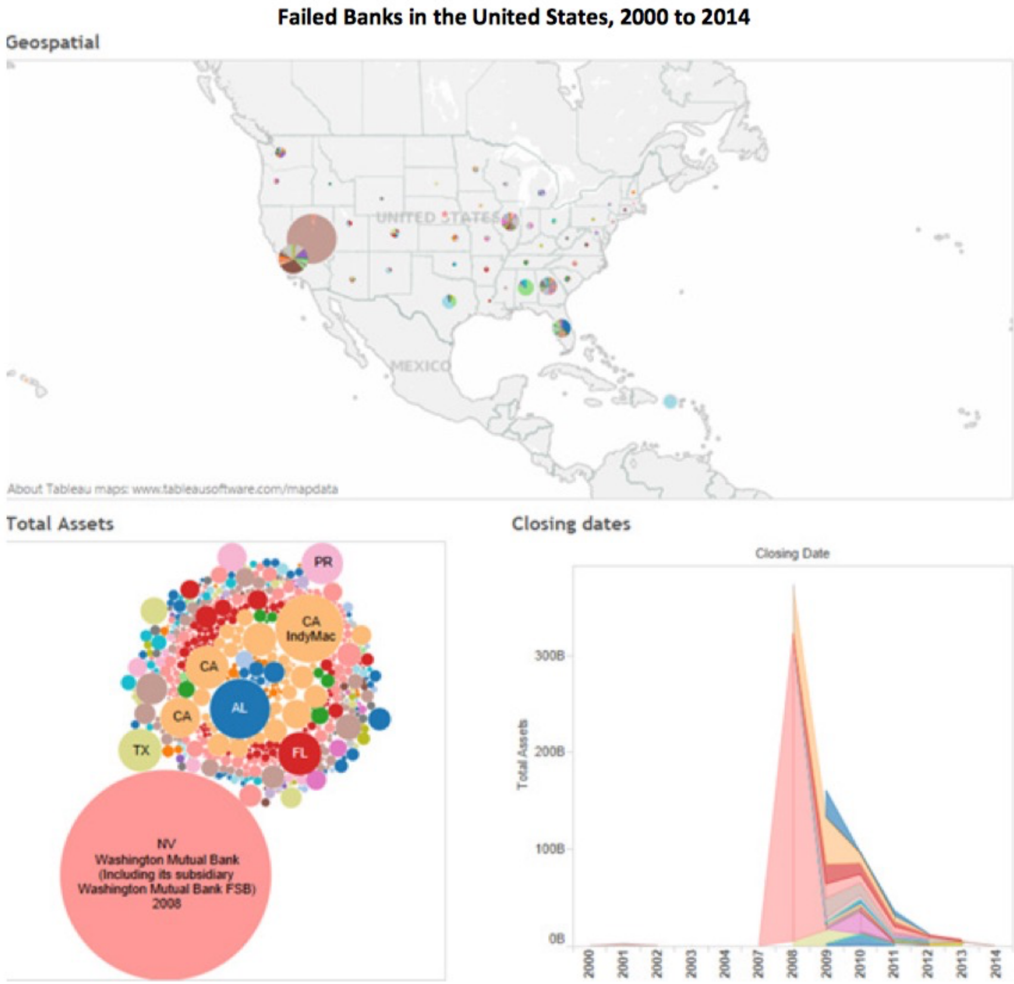

Figure 5.5: Interactive System Dashboard (FDIC Data).

CrisisMetrics (Infolytica, 2015)(http://crisismetrics.com/) is an interactive interface that utilizes the Bank's (BIS) quarterly bank statistics to explore bank cross-border risks. The tool provides regulators and bank risk managers with a way to monitor the risk of banking sector spillovers. CrisisMetrics offers two optional network displays, namely, force-directed layout and chord layout. The left graph shows the force-directed layout in the fourth quarter of 2003. This allows the user to understand the network structure, so as to understand the strength, vulnerability and relationship and so on. The chord is shown on the right side panel of Fig. 5.6. The same data is presented, but the immediate node neighborhood is displayed in a circular fashion. The user can zoom and pan to browse the data.

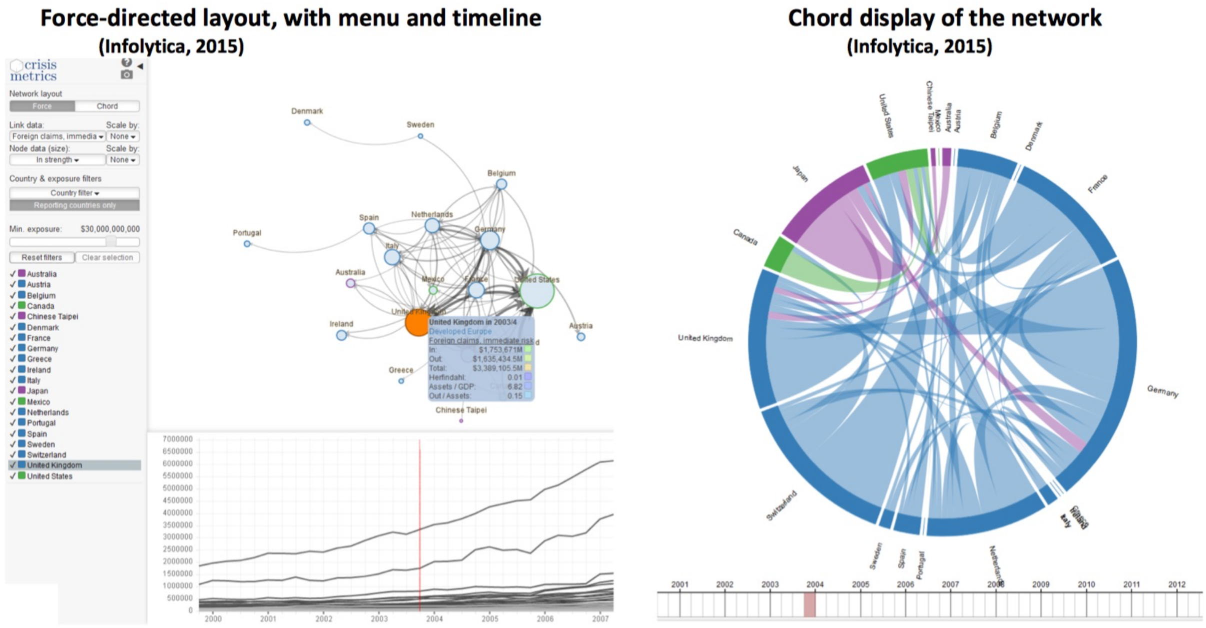

Figure 5.6: CrisisMetrics cross-border network of banking exposures (BIS data)

CHAPTER 6

FINTECH

FinTech, a.k.a Financial Technologies, typically, represents the technologies used and applied in the financial fields with financial institutions themselves on the back end. However with the rapid development of digital techniques nowadays, FinTech is evolving from traditional financial technologies into modern financial services, including mobile payments, money transfers, loans, fund raising, and asset management.

FinTech is broadly taken as the combination of financial services and information technology. In fact, the intertwinement and mutual reinforcement of financial and technological development have long been observed. The Global Financial Crisis of 2008 was a tipping point and it partially explains why the evolution of FinTech is now two-folded. This is the new challenge for market participants, particularly in balancing the potential benefits and risks of innovation. The challenge is shown to have more presence in the developing world, particularly Asia.

From the perspective of academic research, the business or mathematic domain experts are eager to demonstrate the theory behind this phenomenon. However, as computer science researchers, we prefer to use the knowledge and methodology from the computer science domain to solve problems. In our survey, we focus more particularly on the 'Blockchain' and 'Bitcoin' instead of the whole FinTech arena.

A blockchain is a digital, decentralized, all implicit transaction of the public ledger. Growing 'Completed' blocks (the most recent transactions) are recorded and added in chronological order, allowing market participants to track digital currency transactions without the need for central record keeping. Each node (connected to a network computer) gets a copy of an automatically downloaded blockchain. Using distributed ledger technology (DLT), it was originally designed as an accounting method for bit currency (virtual currency). Then the technology often appears in different types of commercial applications. Today, digital transactions are primarily verified by this technique, although it is possible to digitize, encode, and insert almost any document into the block chain. This creates an indelible unchanged record; in addition, the authenticity of the record can be verified by the entire community using a block chain rather than a centralized authority.

The block may be the main technological innovation of Bitcoin. Bitcoin is not subject to the central authority's control. On the contrary, when a person pays a payee for goods or services, the user decides and verifies the transaction, thus eliminating the need for a third party to process or store the payment. The completed transaction is publicly logged into the block and eventually enters the block chain, where it is verified and relayed by other Bitcoin users. On average, a new block is added to the block chain every 10 minutes by mining.

The combination of open and transparent data provided by various interfaces or APIs combined with the network data structure should be a cumulative bonus for data visualization and information designers. We will list some visual analysis systems and visualization works to highlight potential research directions in this chapter.

6.1 Academic Works

Moser et al. (Moser et al., 2013) systematically introduced the money-laundering tools available in the Bitcohol ecosystem, understood their mode of operation, and concluded the effectiveness of the counter-money laundering work. More specifically, they evaluated the current mix of bitmaps that can be used to increase the anonymity of their users; if so, the cost and the risk. They tested three services and tried to track anonymous transactions in public accounts and found that the service BitLaundry (Fig. 6.1.) did not provide sufficient anonymity, and Blockchain.info Send Shared service (Fig. 6.2) makes it impossible to find any related connection in the graph directly.

They pay 0.33 BTC in service to transfer and share two transactions in one day. They found a direct connection between the input t14 and the first output t15 in the transaction graph (Fig. 6.1), resulting in a high level of contamination. The second output t16 is not connected to the input. In this experiment, the service uses half of the input transactions directly to create an output transaction. Although the sample is not large, it shows that the service does not provide a very good anonymity. The reason for this may be that the service usage is low and there is a lack of technical measures to ensure that the user does not receive their input coins.

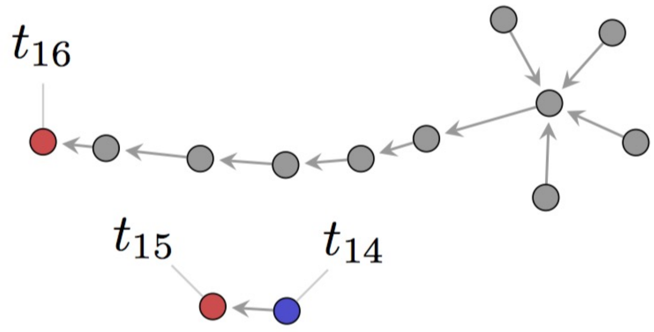

Figure 6.1: Transaction graph of the BitLaundry experiment.

As for Blockchain.info Send Shared service, they send 0.4 BTC to the online wallet, and after six minutes, use the shared wallet function to send them to another address. Since they want to detect any of the special patterns in the transaction graph, 11 additional transactions were created to increase the chances of receiving multiple coins from the same address. They can not find any direct connection between input and output transactions. However, only eight separate graphs (Fig. 6.2) are created, which means that there is a connection between multiple outputs. In addition, there are some hubs that bundle a lot of transactions into a transaction. They only find some coin transactions, which mainly uses other Bitcoins. A transaction that connects to multiple output transactions indicates that the transaction is bound to a larger transaction and then split into payments again. Although the service has already used the input, it is not possible to find any direct connection between the input and output transactions. The service will be a large number of small transactions bundled into a larger amount of transactions, and then split, it is difficult to infer the origin of Bitcoin.

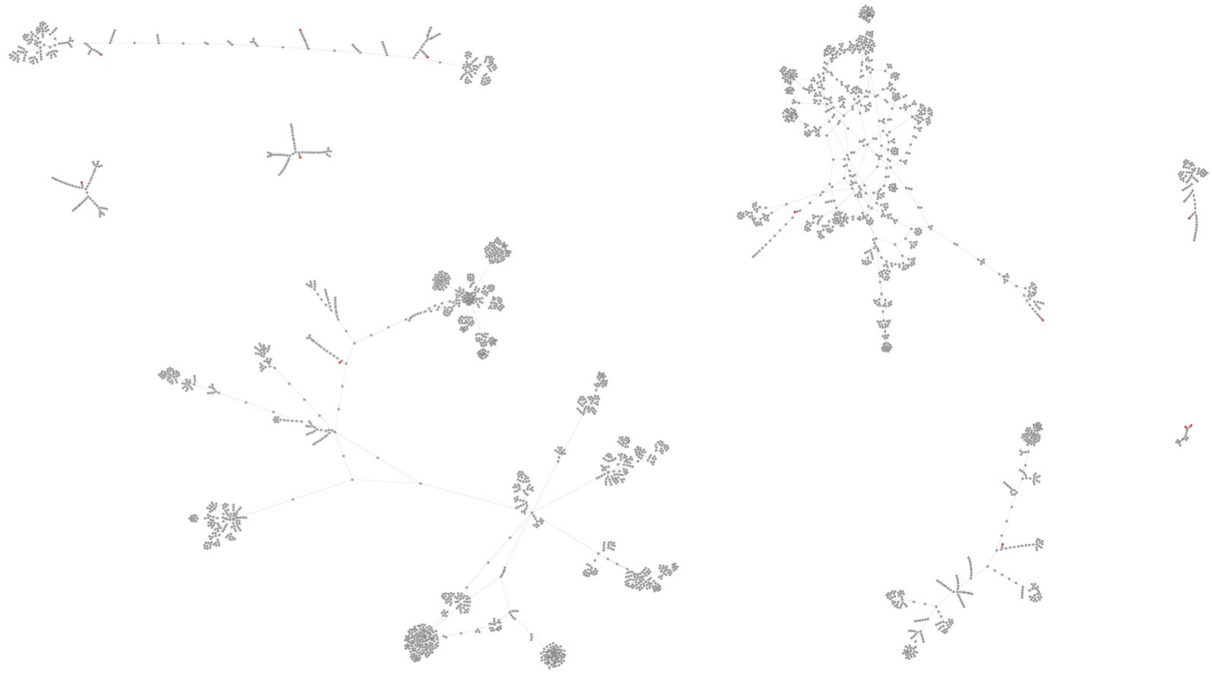

Figure 6.2: Partial transaction graphs of the Blockchain.info Send Shared experiment.

McGinn et al. (McGinn et al., 2016) introduced development tools to explore patterns of relevant patterns of behavior in datasets that are closely linked to all Bitcoin transactions. This paper generates a top-down system visualization function that allows pattern to drill through details to enter any transaction, as compared to the bottom-up approach, like Battista et. al (Di Battista et al., 2015), previously from a single source transaction data. In addition, they have demonstrated how to combine transactions and address maps into an associated high-fidelity visualization. Effective data visualization in large-scale data observation facilities is used to accelerate this data exploration and gain useful insights for domain experts and the general public (Fig. 6.3).

The input is an orange node whose size is proportional to its value. They are associated with the orange edges that contain them. The output is a blue node whose size is also proportional to its value. They are associated with the transactions they contain through the blue edges, and if the output is referenced as input in subsequent transactions within the visualization range, they are connected to the transaction through the orange input side, resulting in a series of costs. The system provides translation, scaling and interaction to show other attributes, such as transaction references and address information. The ability of transferring this formatted subcomponent data to the hand-held tablet display by highlighting the connected components and by PeerJS allows further detailed data analysis to be more intuitively associated with the Bitcoin browsing tool (e.g., Blockchain.info). The visualization data set can also be filtered from the hand-held tablet screen by amount, address, or reference value.

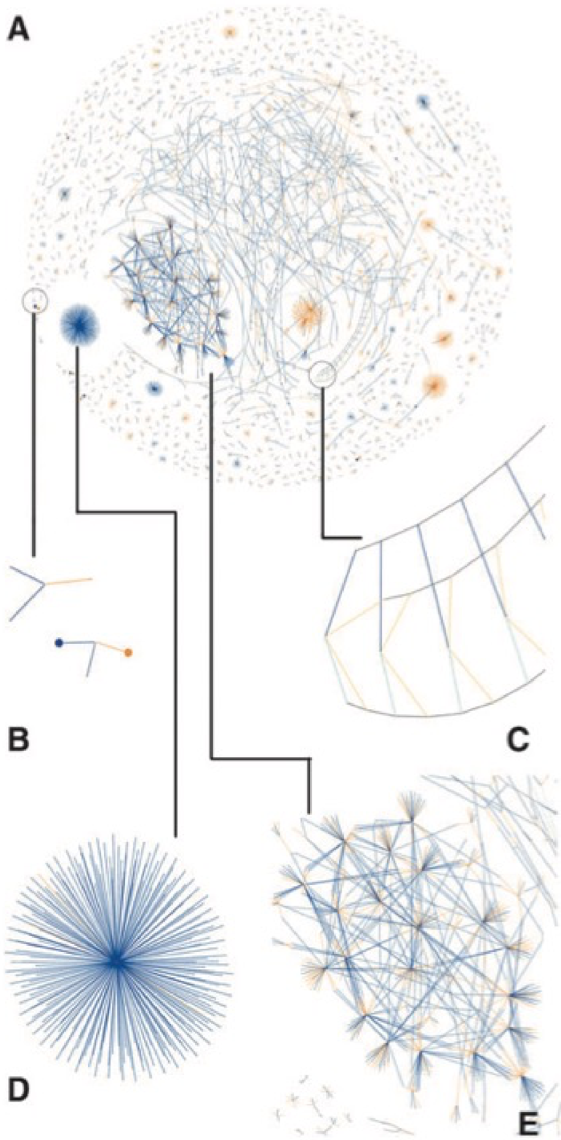

Figure 6.3: (A) High-resolution (8k) visualization of a standard block; (B) detail of both a low (small node) and a high (large node) value transaction, (C) known and linked Bitcoin addresses, (D) a payout system, and (E) a highly associated disconnected component believed to be a coin-tumbling service to move amounts rapidly between addresses, obfuscating the source and destination of funds.

Bitcoin community has a long inconsistent debate that if the block size should be 1MB limited. Initially implemented to prevent certain denial of service attacks, it prevents the system from exceeding the transaction rate of only about four transactions per second. In 2015, unknown actors themselves automatically generate economically insignificant spam transactions, which artificially increase the data rate, and seem to need to increase the 1MB limit. By visualizing the events that these transactions are excavated during this time, some interesting observations can be made. Attack begins with a sudden increase in transaction speed, in the visualization of the formation of "parasitic worm" structure, which is due to the high frequency segmentation algorithm to a small amount to the same group of addresses, as shown in Figure 6.4(a).

The second attack occurs at two stages, as shown by the gradient change in the number of records in the UTXOs. Attacks have limited impact on the backlog of transactions in the plot, but have a very detrimental effect on the number of UTXOs. By studying the block visualization of this period, we can see the use of very different algorithms that produce a "cancer" structure. This attack reveals at the data density rather than the transaction rate, possibly by a completely independent second party. Also note that the simple constant parameters in the algorithm are modified to increase the attack data density for the second phase, as shown in Fig 6.4(b).

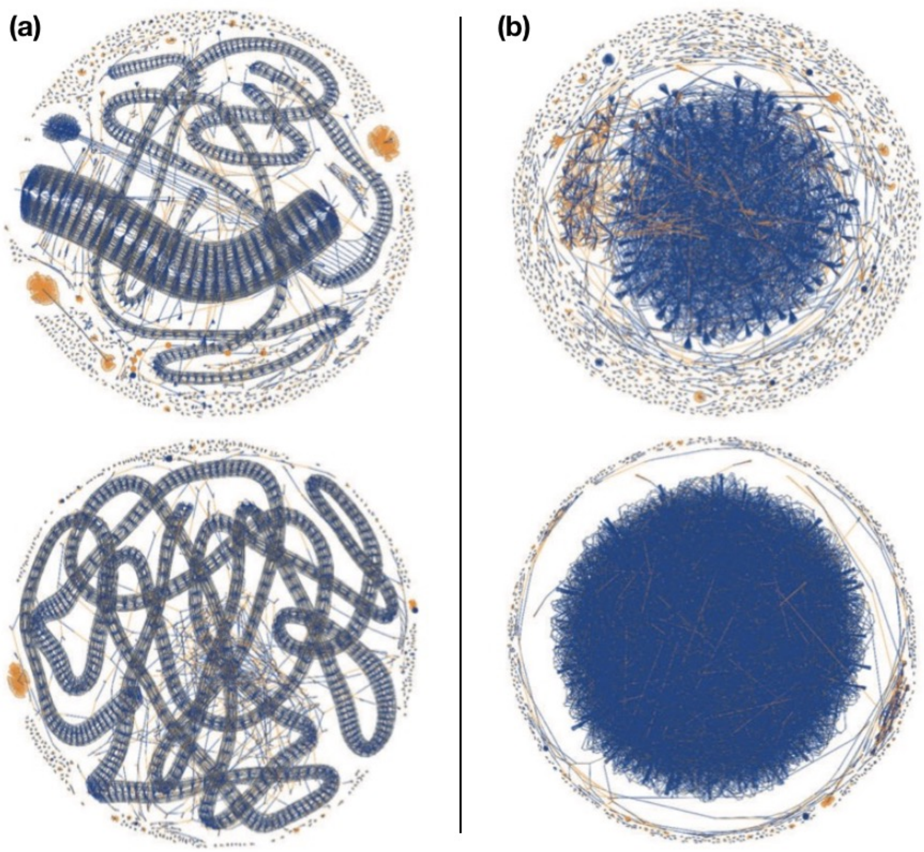

Figure 6.4: (a):Blocks #364133,364618: Initial "parasitic worm" transaction rate attack. (b):Blocks #367409,368580 from 29th July to 6th August 2015 show two distinct phases of the second data density-based "tumor" attack, note obvious change in algorithm parameterization to increase density.

So far, the use of visualization has been limited to the extent that only the results of these bottom-up methods are presented. Di Battista et. al (Di Battista et al., 2015) present the first interesting deployment of small-scale visualization to directly analyze the transaction data in the blockchain, which reveals a tool for the bottom-up of the effects of the selected source transaction on subsequent flows in the transaction graph.

6.2 Online Tools

Here comes to the visualization tools and demos. Interaqt(http://bitcoin.interaqt.nl/) and Daily Blockchain (http://dailyblockchain.github.io/) display the networked nature of Bitcoin. We can see Bitcoin transaction happening in real-time and the evolving hubs of the Bitcoin network. Screenshots are shown in Fig. 6.5.

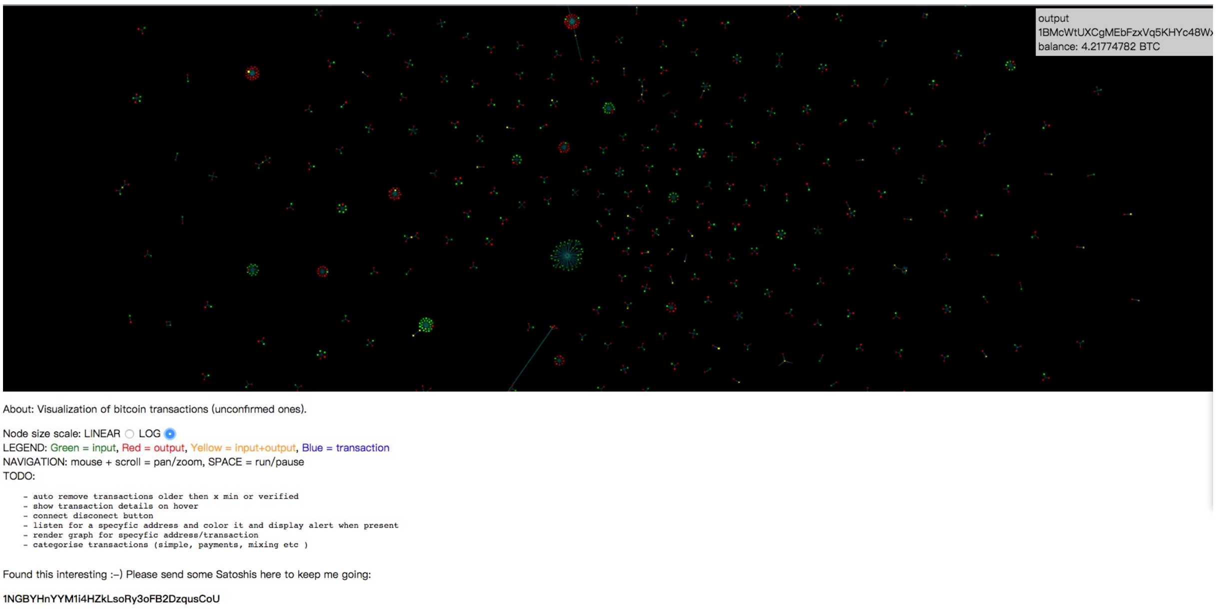

Figure 6.5: Daily Blockchain Interface & Description.

The designers prefer to discover patterns from geospatial perspective by Bitcoin transactions visualization. There are three typical works from BitTicker (http://bitcointicker.co/transactions/), FiatLeak (http://fiatleak.com/) and Bitnodes (https://bitnodes.21.co/), which can watch the world currencies flow into BTC and transactions location real time (Fig. 6.6). Also the Bitcoin Globe (http://bitcoinglobe.com/) describes the volume of transactions form the global view (Fig. 6.7).

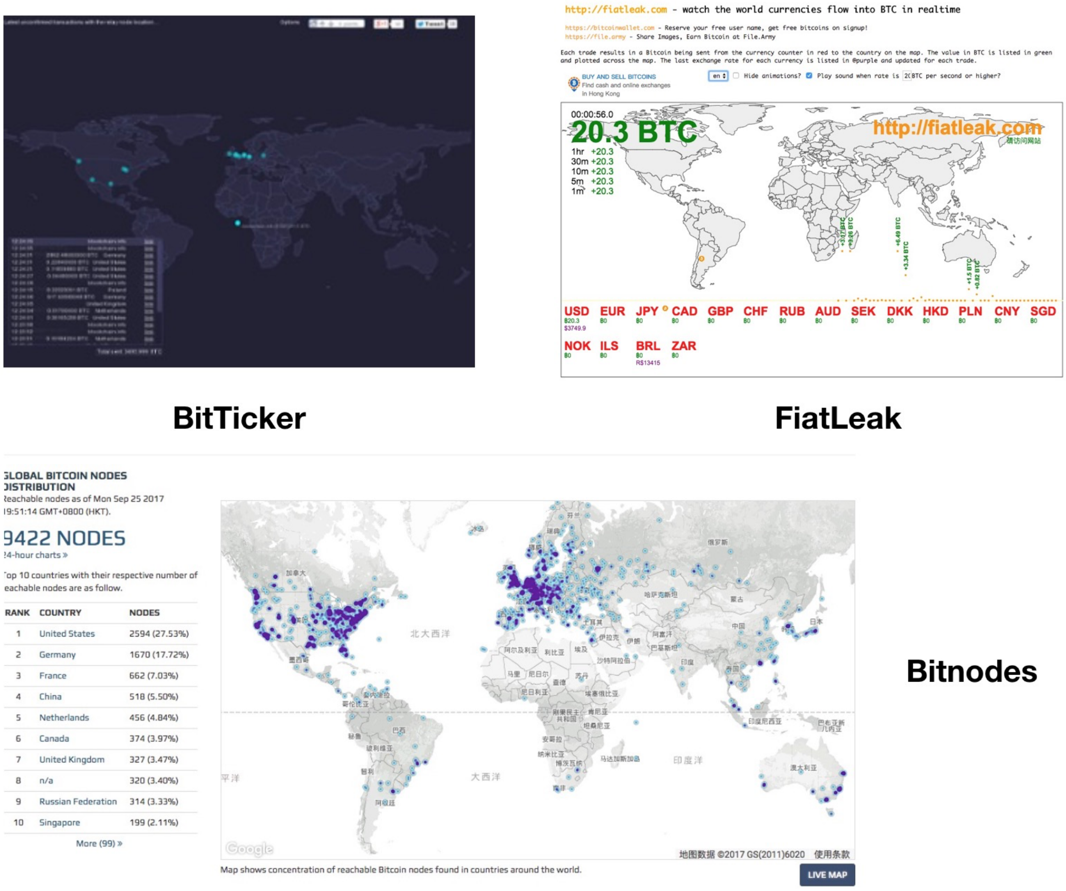

Figure 6.6: Screenshots from BitTicker, FiatLeak and Bitnodes.

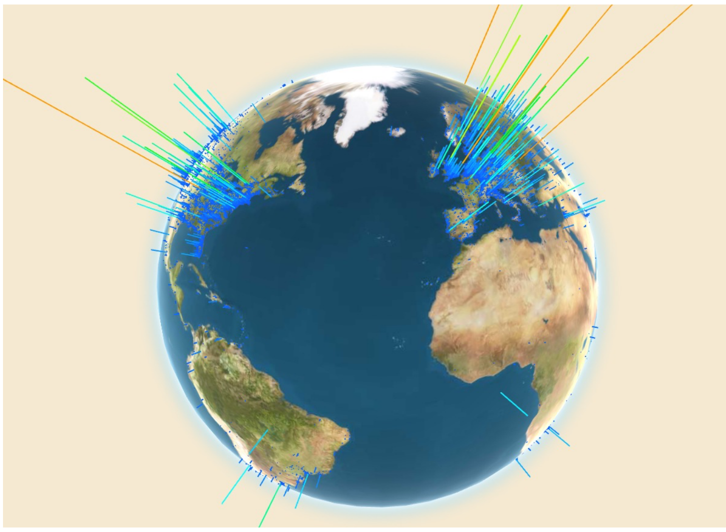

Figure 6.7: Bitcoin Globe View.

Elliptic (https://www.elliptic.co/), a London-based Bitcoan analysis and storage company, has announced an interactive block-chain visualization tool that shows the history of all bitmaps since its inception in 2009. The company is expected to use the "Bitcoin Big

Bang" (Fig. 6.8) for new tools to promote the adoption of digital currency. The security currency has been associated with anonymity to prevent the company from embracing digital money. In addition, Bitcoin is often involved in illegal activities such as drug trade or money laundering. This is why it is possible to give the financial institution a criminal Bitcoin transaction map . The new tool will help Bitcoin community to ensure that they are using implicit currencies that have not yet been used for criminal purposes.

"If the digital money in the enterprise to occupy the legal status, then it is bound to out of the shadow of the black network. Our technology enables us to track the history and real-time process, on behalf of the business use of Bitcoin turning point. We have developed this technology, Not the crime, but to support corporate anti-money laundering obligations." Elliptic CEO James Smith said that compliance officials can finally be assured that they have carried out a real, law-abiding efforts to determine. The Bitcoin Big Bang displays a web of more than 250 blockchain entities and shows transactions between them. The company's team has been working on the development of the 35 GB map of the bitcoin transactions over the last year.

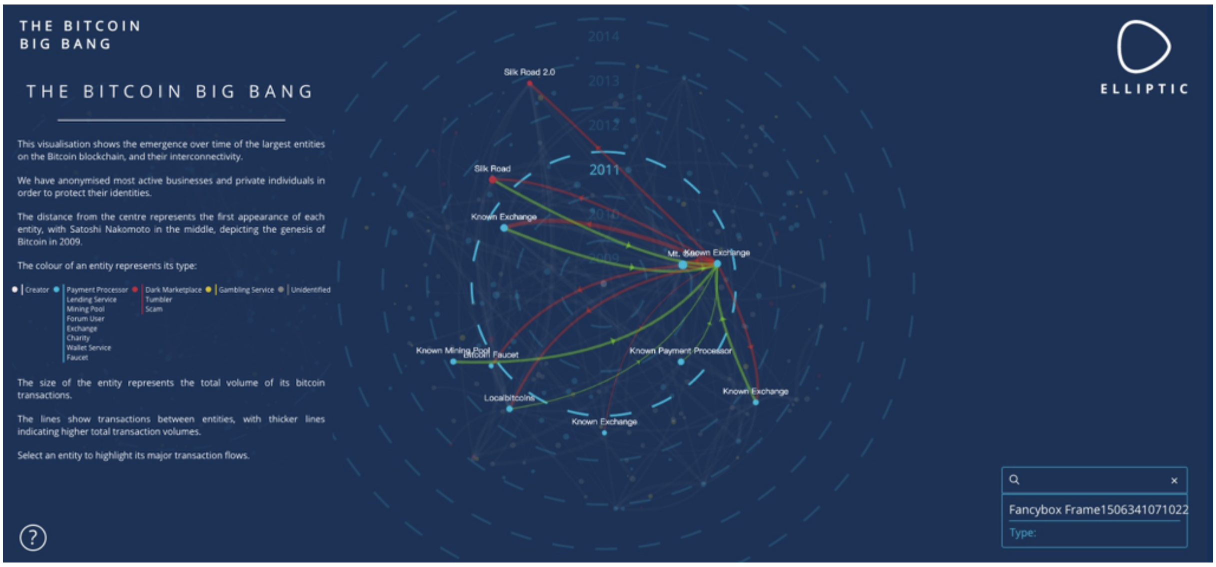

Figure 6.8: The Bitcoin Big Bang.

This cryptographic system consists primarily of an unassigned public database, and anyone with a tokenized and pseudo-anonymous identity can write data that conforms to the protocol. Because identity is obfuscated by using tagged addresses, the ability to identify and classify patterns of abnormal behavior in data can be used by many parties, such as financial regulators (such as money laundering) or protocol developers (in the case of attack system resilience). Performing initial graphical observations is a useful first step in the data analysis workflow to investigate the structural nature of this repetitive behavior. The visualization is shown in Fig. 6.4. enabling the discovery of unexpected trading patterns such as money laundering, as well as the observation of several apparent denial of service attacks on Bitcoin networks. This allows researchers to quickly understand the structure of this behavioral pattern for accelerated analysis and classification surveys.

CHAPTER 7

CONCLUSION AND FUTURE WORK

In this survey, we first review the background and motivation of visual analysis towards financial data. We then discuss the three aspects of typical financial tasks and their applications in FinTech. Specifically, we introduce how to solve financial fraud detection using data mining techniques and visual analysis techniques. In addition, we present the typical predictive algorithm and the widely emergence of visual prediction. Then, we discuss how correlative analysis makes a great contribution to gaining insights from different sources of data. Finally, we investigate the current trends of FinTech with a focus on the Blockchain and Bitcoin analysis and underlying issues.

However, scholarly research into the application of visualizations in finance has been relatively sparse. One of the major reasons for this has been the difficulty in obtaining real, internal data and detailed problem descriptions from financial organizations, and conducting research on the inner workings of the firms. The field-defining document for Visual Analytics (VA), "Illuminating the Path" (Cook & Thomas, 2005), pointed out that visualizations cannot exist in a vacuum, and that the development of new visualization techniques should be guided by the needs of "customers", their analytics problems, workflow, organization, and data (Cook & Thomas, 2005). However, to date this has apparently not been taken seriously in the area of finance, and only a close discussion can boost the outcomes of the finance field. The computer science methodology and knowledge can be used to solve tasks like anomaly detection or prediction.

Before the formal creation of VA as a distinct enterprise, the greatest strides in related research have been in the development of graphical presentations, interaction techniques, and visual metaphors (as a whole, the "visualization") of information. The field of VA inherited this foundation. Many of these visualizations have been applied to financial data: variations of candlestick and line charts, stacked and bullet graphs, scatterplots, parallel coordinate plots, bead cluster diagrams, treemaps and heatmaps, node-link diagrams and so on. Modern visualization techniques, even those developed during the past decade, are still not widely employed in financial analytics products. We need to facilitate the power of advanced visualization to improve the accuracy of analysis and efficiency of partitioners.

As for the increasing popularity of Blockchain and Bitcoin, which have similar attributes as typical financial markets. For example, 1) Bitcoin prediction is almost the most popular topic among the community. When the factors influence the fluctuation of price in the near future. Or when is the perfect time to fork a block. 2) Anomaly detection in Bitcoin transaction can contribute to a better platform for the whole blockchain community. 3) Is there any real-world connection among those transactions? We can start digging by comparison analysis of Bitcoin network and Social network, which is a typical task in visualization named "Graph Mining" or "Social Network Analysis".

In the future, I will focus on the wide use of visual analysis in the financial area. Specifically, for the deciders, I will try to facilitate the power of visual analysis for some typical areas, like anomaly detection and predictive analysis. As for the dealers, the proper improvement of their daily working interfaces should be emphasized, let alone the enhancement of their efficiency.

Bibliography

(2007). Financial crimes report 2007..

(2008). Financial crimes 2008..

Badam, S. K., Zhao, J., Sen, S., Elmqvist, N., & Ebert, D. (2016). Timefork: Interactive prediction of time series. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*, pp. 5409–5420. ACM.

Bollen, J., Mao, H., & Zeng, X. (2011). Twitter mood predicts the stock market. *Journal of computational science*, *2*(1), 1–8.

Box, G. E., Jenkins, G. M., Reinsel, G. C., & Ljung, G. M. (2015). *Time series analysis: forecasting and control*. John Wiley & Sons.

Broberg, M., Lundberg, L., & Grahn, H. (1999). Visualization and performance prediction of multithreaded solaris programs by tracing kernel threads. In *Parallel Processing, 1999. 13th International and 10th Symposium on Parallel and Distributed Processing, 1999. 1999 IPPS/SPDP. Proceedings*, pp. 407–413. IEEE.

Buono, P., Plaisant, C., Simeone, A., Aris, A., Shmueli, G., & Jank, W. (2007). Similarity-based forecasting with simultaneous previews: A river plot interface for time series forecasting. In *Information Visualization, 2007. IV'07. 11th International Conference*, pp. 191–196. IEEE.

Chan, Y.-Y., & Qu, H. (2016). Finavistory: Using narrative visualization to explain social and economic relationships in financial news. In *Big Data and Smart Computing (BigComp), 2016 International Conference on*, pp. 32–39. IEEE.

Chang, R., Ghoniem, M., Kosara, R., Ribarsky, W., Yang, J., Suma, E., Ziemkiewicz, C., Kern, D., & Sudjianto, A. (2007). Wirevis: Visualization of categorical, time-varying data from financial transactions. In *Visual Analytics Science and Technology, 2007. VAST 2007. IEEE Symposium on*, pp. 155–162. IEEE.

Clopton, J., Heitger, L., & Morrow, L. (2014). Adding augmented intelligence and data visualization to the mix. *Fraud Magazine*.

Cook, K. A., & Thomas, J. J. (2005). Illuminating the path: The research and development agenda for visual analytics..

Croker, S. T. (2007). Effective forecast visualization with sas/graph. In *SAS Global Forum*.

Di Battista, G., Di Donato, V., Patrignani, M., Pizzonia, M., Roselli, V., & Tamassia, R. (2015). Bitconeview: visualization of flows in the bitcoin transaction graph. In *Visualization for Cyber Security (VizSec), 2015 IEEE Symposium on*, pp. 1–8. IEEE.

Dilla, W. N., & Raschke, R. L. (2015). Data visualization for fraud detection: Practice implications and a call for future research. *International Journal of Accounting Information Systems*, *16*, 1–22.

Ding, X., Zhang, Y., Liu, T., & Duan, J. (2015). Deep learning for event-driven stock prediction.. In *Ijcai*, pp. 2327–2333.

Dumas, M., McGuffin, M. J., & Lemieux, V. L. (2014). Financevis. net: A visual survey of financial data visualizations. In *Poster Abstracts of IEEE Conference on Visualization*, Vol. 2.

Edwards, R. D., Bassetti, W., & Magee, J. (2007). *Technical analysis of stock trends*. CRC press.

Fama, E. F. (1965). The behavior of stock-market prices. *The journal of Business*, *38*(1), 34–105.

Flood, M. D., Lemieux, V. L., Varga, M., & Wong, B. W. (2016). The application of visual analytics to financial stability monitoring. *Journal of Financial Stability*, *27*, 180–197.

Fung, G. P. C., Yu, J. X., & Lam, W. (2003). Stock prediction: Integrating text mining approach using real-time news. In *Computational Intelligence for Financial Engineering, 2003. Proceedings. 2003 IEEE International Conference on*, pp. 395–402. IEEE.

Ghosh, S., & Reilly, D. L. (1994). Credit card fraud detection with a neural-network. In *System Sciences, 1994. Proceedings of the Twenty-Seventh Hawaii International Conference on*, Vol. 3, pp. 621–630. IEEE.

Han, J., Pei, J., & Kamber, M. (2011). *Data mining: concepts and techniques*. Elsevier.

Hao, M. C., Janetzko, H., Mittelstädt, S., Hill, W., Dayal, U., Keim, D. A., Marwah, M., & Sharma, R. K. (2011). A visual analytics approach for peak-preserving prediction of large seasonal time series. In *Computer Graphics Forum*, Vol. 30, pp. 691–700. Wiley Online Library.

Hao, M. C., Janetzko, H., Sharma, R. K., Dayal, U., Keim, D. A., & Castellanos, M. (2009). Poster: visual prediction of time series. In *Visual Analytics Science and Technology, 2009. VAST 2009. IEEE Symposium on*, pp. 229–230. IEEE.

Huang, M. L., Liang, J., & Nguyen, Q. V. (2009). A visualization approach for frauds detection in financial market. In *Information Visualisation, 2009 13th International Conference*, pp. 197–202. IEEE.

Ichikawa, Y., Tsunawaki, T., Fujishiro, I., & Yoon, H. (2002). A visualization environment for multiple daytime stock price predictions. In *Proceedings of the 2nd VIIP International Conferences on Visualization, Imaging and Image Processing, Malaga, Spain*.

Keim, D. A., Mansmann, F., Schneidewind, J., & Ziegler, H. (2006). Challenges in visual data analysis. In *Information Visualization, 2006. IV 2006. Tenth International Conference on*, pp. 9–16. IEEE.

Kielman, J., Thomas, J., & May, R. (2009). Foundations and frontiers in visual analytics. *Information Visualization*, *8*(4), 239–246.

Kirkland, J. D., Senator, T. E., Hayden, J. J., Dybala, T., Goldberg, H. G., & Shyr, P. (1999). The nasd regulation advanced-detection system (ads). *AI Magazine*, *20*(1), 55.

Kirkos, E., Spathis, C., & Manolopoulos, Y. (2007). Data mining techniques for the detection of fraudulent financial statements. *Expert systems with applications*, *32*(4), 995–1003.

Ko, S., Maciejewski, R., Jang, Y., & Ebert, D. S. (2012). Marketanalyzer: an interactive visual analytics system for analyzing competitive advantage using point of sale data. In *Computer Graphics Forum*, Vol. 31, pp. 1245–1254. Wiley Online Library.

Kovalerchuk, B., & Vityaev, E. (2000). *Data mining in finance: advances in relational and hybrid methods*, Vol. 547. Springer Science & Business Media.

Leite, R. A., Gschwandtner, T., Miksch, S., Kriglstein, S., Pohl, M., Gstrein, E., & Kuntner, J. (2017). Eva: Visual analytics to identify fraudulent events. *IEEE Transactions on Visualization and Computer Graphics*.

Li, J., Huang, K.-Y., Jin, J., & Shi, J. (2008). A survey on statistical methods for health care fraud detection. *Health care management science*, *11*(3), 275–287.

Malik, A., Maciejewski, R., Elmqvist, N., Jang, Y., Ebert, D. S., & Huang, W. (2012). A correlative analysis process in a visual analytics environment. In *Visual Analytics Science and Technology (VAST), 2012 IEEE Conference on*, pp. 33–42. IEEE.

McGinn, D., Birch, D., Akroyd, D., Molina-Solana, M., Guo, Y., & Knottenbelt, W. J. (2016). Visualizing dynamic bitcoin transaction patterns. *Big data*, *4*(2), 109–119.

Moser, M., Bohme, R., & Breuker, D. (2013). An inquiry into money laundering tools in the bitcoin ecosystem. In *eCrime Researchers Summit (eCRS), 2013*, pp. 1–14. IEEE.

Ngai, E., Hu, Y., Wong, Y., Chen, Y., & Sun, X. (2011). The application of data mining techniques in financial fraud detection: A classification framework and an academic review of literature. *Decision Support Systems*, *50*(3), 559–569.

Nofsinger, J. R. (2005). Social mood and financial economics. *The Journal of Behavioral Finance*, *6*(3), 144–160.

Pearl, J. (2014). *Probabilistic reasoning in intelligent systems: networks of plausible inference*. Morgan Kaufmann.

Peterson, R. L. (2011). *Inside the investor's brain: the power of mind over money*, Vol. 295. John Wiley & Sons.

Pryke, M. (2010). Money's eyes: the visual preparation of financial markets. *Economy and Society*, *39*(4), 427–459.

Rönnqvist, S., & Sarlin, P. (2015). Bank networks from text: interrelations, centrality and determinants. *Quantitative Finance*, *15*(10), 1619–1635.

Schwabish, J. A. (2014). An economist's guide to visualizing data. *The Journal of Economic Perspectives*, *28*(1), 209–233.

Schwartz, H. H. (2007). A introduction to behavioral economics: The complicating but sometimes critical considerations..

Sorenson, E., & Brath, R. (2013). Financial visualization case study: Correlating financial timeseries and discrete events to support investment decisions. In *Information Visualisation (IV), 2013 17th International Conference*, pp. 232–238. IEEE.

Spathis, C. T. (2002). Detecting false financial statements using published data: some evidence from greece. *Managerial Auditing Journal*, *17*(4), 179–191.

Tegarden, D. P. (1999). Business information visualization. *Communications of the AIS*, *1*(1es), 4.

Turban, E., Sharda, R., & Delen, D. (2011). *Decision support and business intelligence systems*. Pearson Education India.

Vanstone, B., & Tan, C. (2008). Artificial neural networks in financial trading. In *Intelligent Information Technologies: Concepts, Methodologies, Tools, and Applications*, pp. 1758–1764. IGI Global.

Venugopal, K., Srinivasa, K., & Patnaik, L. (2009). Fuzzy based neuro-genetic algorithm for stock market prediction. In *Soft Computing for Data Mining Applications*, pp. 139–166. Springer.

Vessey, I. (2006). The theory of cognitive fit. *Human-computer interaction and management information systems: Foundations*, 141–183.

Yeh, I.-C., & Lien, C.-h. (2009). The comparisons of data mining techniques for the predictive accuracy of probability of default of credit card clients. *Expert Systems with Applications*, *36*(2), 2473–2480.

Zarate Santovena, A. (2013). *Big data: evolution, components, challenges and opportunities*. Ph.D. thesis, Massachusetts Institute of Technology.